ELSEVIER

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Journal of Biotechnology 115 (2005) 113–128

JOURNAL OF Biotechnology

www.elsevier.com/locate/jbiotec

# Advanced genetic strategies for recombinant protein expression in *Escherichia coli*

Hans Peter Sørensen, Kim Kusk Mortensen*

*Laboratory of BioDesign, Department of Molecular Biology, Aarhus University, Gustav Wieds Vej 10 C, DK-8000 Aarhus C, Denmark*

Received 29 March 2004; received in revised form 26 August 2004; accepted 30 August 2004

**Abstract**

Preparations enriched by a specific protein are rarely easily obtained from natural host cells. Hence, recombinant protein production is frequently the sole applicable procedure. The ribosomal machinery, located in the cytoplasm is an outstanding catalyst of recombinant protein biosynthesis. *Escherichia coli* facilitates protein expression by its relative simplicity, its inexpensive and fast high-density cultivation, the well-known genetics and the large number of compatible tools available for biotechnology. Especially the variety of available plasmids, recombinant fusion partners and mutant strains have advanced the possibilities with *E. coli*. Although often simple for soluble proteins, major obstacles are encountered in the expression of many heterologous proteins and proteins lacking relevant interaction partners in the *E. coli* cytoplasm. Here we review the current most important strategies for recombinant expression in *E. coli*. Issues addressed include expression systems in general, selection of host strain, mRNA stability, codon bias, inclusion body formation and prevention, fusion protein technology and site-specific proteolysis, compartment directed secretion and finally co-overexpression technology. The macromolecular background for a variety of obstacles and genetic state-of-the-art solutions are presented.
© 2004 Elsevier B.V. All rights reserved.

*Keywords:* *Escherichia coli*; Recombinant protein expression systems; Inclusion bodies; Fusion proteins; Rare codon tRNAs

## 1. The modern recombinant expression system

A number of central elements are essential in the design of recombinant expression systems (Baneyx, 1999; Jonasson et al., 2002). Expression is normally induced from a plasmid harboured by a system compatible genetic background. The genetic elements of the expression plasmid include origin of replication (*ori*), an antibiotic resistance marker, transcriptional promoters, translation initiation regions (TIRs) as well as transcriptional and translational terminators.

### 1.1. The replicon

The replicon of plasmids contain the origin of replication and in some cases associated *cis* acting elements (del Solar et al., 1998). Most plasmid vectors used in re-

* Corresponding author. Fax: +45 86 12 31 78.
*E-mail address:* kkm@mb.au.dk (K.K. Mortensen).

0168-1656/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.jbiotec.2004.08.004

combinant protein expression replicate by the ColE1 or the p15A replicon. Plasmid copy number is controlled by the origin of replication that preferably replicates in a relaxed fashion (Baneyx, 1999). The ColE1 replicon present in modern expression plasmids is derived from the pBR322 (copy number 15–20) or the pUC (copy number 500–700) family of plasmids, whereas the p15A replicon is derived from pACYC184 (copy number 10–12). These multi-copy plasmids are stably replicated and maintained under selective conditions and plasmid free daughter cells are rare (Summers, 1998). Plasmid incompatibility is defined as the inability of two plasmids to be stably maintained in the same cell (Hardy, 1987). Different replicon incompatibility groups and drug resistance markers are required when multiple plasmids are employed for the co-expression of gene products. Derivatives containing ColE1 and p15A replicons are often combined in this context since they are compatible plasmids (Mayer, 1995).

### *1.2. Resistance markers*

The most common drug resistance markers in recombinant expression plasmids confer resistance to ampicillin, kanamycin, chloramphenicol or tetracycline. Plasmid mediated resistance to ampicillin is accomplished by expression of β-lactamase from the *bla* gene. This enzyme is secreted to the periplasm, where it catalyse hydrolysis of the β-lactam ring. Ampicillin present in the cultivation medium is especially susceptible to degradation, either by secreted β-lactamase, or acidic conditions in high-density cultures. The latter effect can be alleviated by the less degradation susceptible ampicillin analog, carbenicillin, Kanamycin, chloramphenicol and tetracycline interfere with protein synthesis by binding to critical areas of the ribosome. Kanamycin is inactivated in the periplasm by aminoglycoside phosphotransferases and chloramphenicol by the *cat* gene product, chloramphenicol acetyl transferase. Various genes confer resistance to tetracycline (Connell et al., 2003).

### *1.3. Promoters*

Recombinant expression plasmids require a strong transcriptional promoter to control high-level gene expression. Basal transcription in the absence of inducer is minimized through the presence of a suitable repressor. Minimization of basal transcription is especially important when the expression target introduce a cellular stress situation and thereby selects for plasmid loss. Promoter induction is either thermal or chemical and the most common inducer is the sugar molecule isopropyl-beta-D-thiogalactopyranoside (IPTG) (Hannig and Makrides, 1998).

### *1.4. Messenger RNA*

Translation initiation from the translation initiation region (TIR) of the transcribed messenger RNA require a ribosomal binding site (RBS) including the Shine–Dalgarno (SD) sequence and a translation initiation codon (Sørensen et al., 2002). The Shine–Dalgarno sequence is located $7 \pm 2$ nucleotides upstream from the initiation codon, which is the canonical AUG in efficient recombinant expression systems (Ringquist et al., 1992). Optimal translation initiation is obtained from mRNAs with the SD sequence UAAGGAGG. The RBS secondary structure is highly important for translation initiation and efficiency is improved by high contents of adenine and thymine (Laursen et al., 2002). Translation initiation efficiency is in particular influenced by the codon following the initiation codon and adenine is abundant in highly expressed genes (Stenstrom et al., 2001).

A transcription terminator placed downstream from the sequence encoding the target gene, serves enhancement of plasmid stability by preventing transcription through the origin of replication and from irrelevant promoters located in the plasmid. Transcription terminators stabilize the mRNA by forming a stem loop at the three prime end (Newbury et al., 1987). Translation termination is preferably mediated by the stop codon UAA in *Escherichia coli*. Increased efficiency of translation termination is achieved by insertion of consecutive stop codons or the prolonged UAAU stop codon (Poole et al., 1995).

### *1.5. Current expression systems*

A wealth of expression systems designed for various applications and compatibilities are available. Approximately 80% of the proteins used to solve three-dimensional structures submitted to the protein data bank (PDB) in 2003 were prepared in an *E. coli* ex-

pression system. The T7 based pET expression system (commercialized by Novagen) is by far the most used in recombinant protein preparation (pET represents more than 90% of the 2003 PDB protein preparation systems). Systems using the λPL promoter/cI repressor (e.g., Invitrogen pLEX), Trc promoter (e.g., Amersham Biosciences pTrc), Tac promoter (e.g., Amersham Biosciences pGEX) and hybrid *lac*/T5 (e.g., Qiagen pQE) promoters are common (Hannig and Makrides, 1998). A radically different system is based on the *araBAD* promoter (e.g., Invitrogen pBAD). Here we review two particular systems that illustrate the most general mechanisms in current recombinant expression systems. Various expression systems and promoters have been reviewed elsewhere (Baneyx, 1999; Hannig and Makrides, 1998; Jonasson et al., 2002).

## 2. The pET expression system

Studier and colleagues first described the pET expression system, which has been developed for a variety of expression applications (Dubendorff and Studier, 1991; Studier et al., 1990). More than 40 different pET plasmids are commercially available. The system includes hybrid promoters, multiple cloning sites for the incorporation of different fusion partners and protease cleavage sites, along with a high number of genetic backgrounds modified for various expression purposes. Expression requires a host strain lysogenized by a DE3 phage fragment, encoding the T7 RNA polymerase (bacteriophage T7 gene 1), under the control of the IPTG inducible *lacUV5* promoter (Fig. 1A). LacI represses the *lacUV5* promoter and the T7/lac hybrid promoter encoded by the expression plasmid. A copy of the *lacI* gene is present on the *E. coli* genome and on the plasmid in a number of pET configurations. LacI is a weakly expressed gene and a 10-fold enhancement of the repression is achieved when the overexpressing promoter mutant $LacI^q$ is employed (Calos, 1978). T7 RNA polymerase is transcribed when IPTG binds and triggers the release of tetrameric LacI from the *lac* operator. Transcription of the target gene from the T7/lac hybrid promoter (repressed by LacI as well) is subsequently initiated by T7 RNA polymerase (Fig. 1A).

The T7 promoter is a 20-nucleotide sequence not recognized by the *E. coli* RNA polymerase. T7 RNA polymerase transcribes maximally 230 nucleotides per second and is five times faster than *E. coli* RNA polymerase (50 nucleotides per second). Background expression from pET expression plasmids is diminished by the presence of T7 lysozyme (bacteriophage T7 gene 3.5 amidase), which is a natural inhibitor of T7 RNA polymerase. Co-expression of T7 lysozyme is achieved by either plasmid pLysS or pLysE. These plasmids harbour the T7 lysozyme gene in silent (pLysS) and expressed (pLysE) orientations, with respect to the cognate tetracycline responsive (Tc) promoter (Studier, 1991). The *lacUV5* promoter is less sensitive to regulation by the cAMP-CRP (cAMP receptor protein) complex, than the *lac* promoter. However, incorporation of 1% glucose in the cultivation medium reduces cAMP levels and enhances repression of the promoter significantly (cAMP is produced as a response to low glucose levels). Graded inductions of pET vectors have recently been included in the pET system repertoire (Novagen Tuner strains). Host strains deficient in the *lacY* gene product lactose permease offers precise control of target protein expression (Khlebnikov and Keasling, 2002).

## 3. The pBAD expression system

Expression plasmids based on the *araBAD* promoter are designed for tight control of background expression and L-arabinose dependent graded expression of the target protein (Guzman et al., 1995). The latter property is in contrast to the all-or-nothing induction experienced by most other bacterial expression systems (Morgan-Kiss et al., 2002). A linear increase in gene expression with increasing inducer concentration is seen at the population level when the *araBAD* system is employed. Induction is unfortunately all-or-nothing in individual cells, which are either fully induced or uninduced (Siegele and Hu, 1997). Autocatalytic mechanisms related to the natural inducer transport systems, in concert with arabinose degradation, are responsible for all-or-nothing induction of the *araBAD* promoter. The autocatalytic effect occurs since the arabinose transporters (*araE* and *araFGH*) are under arabinose inducible control. Homogenous gene expression has been achieved in strains deficient in arabinose transport and degradation, by facilitated diffusion of arabinose, catalyzed by arabinose independent transporters supplied *in trans*

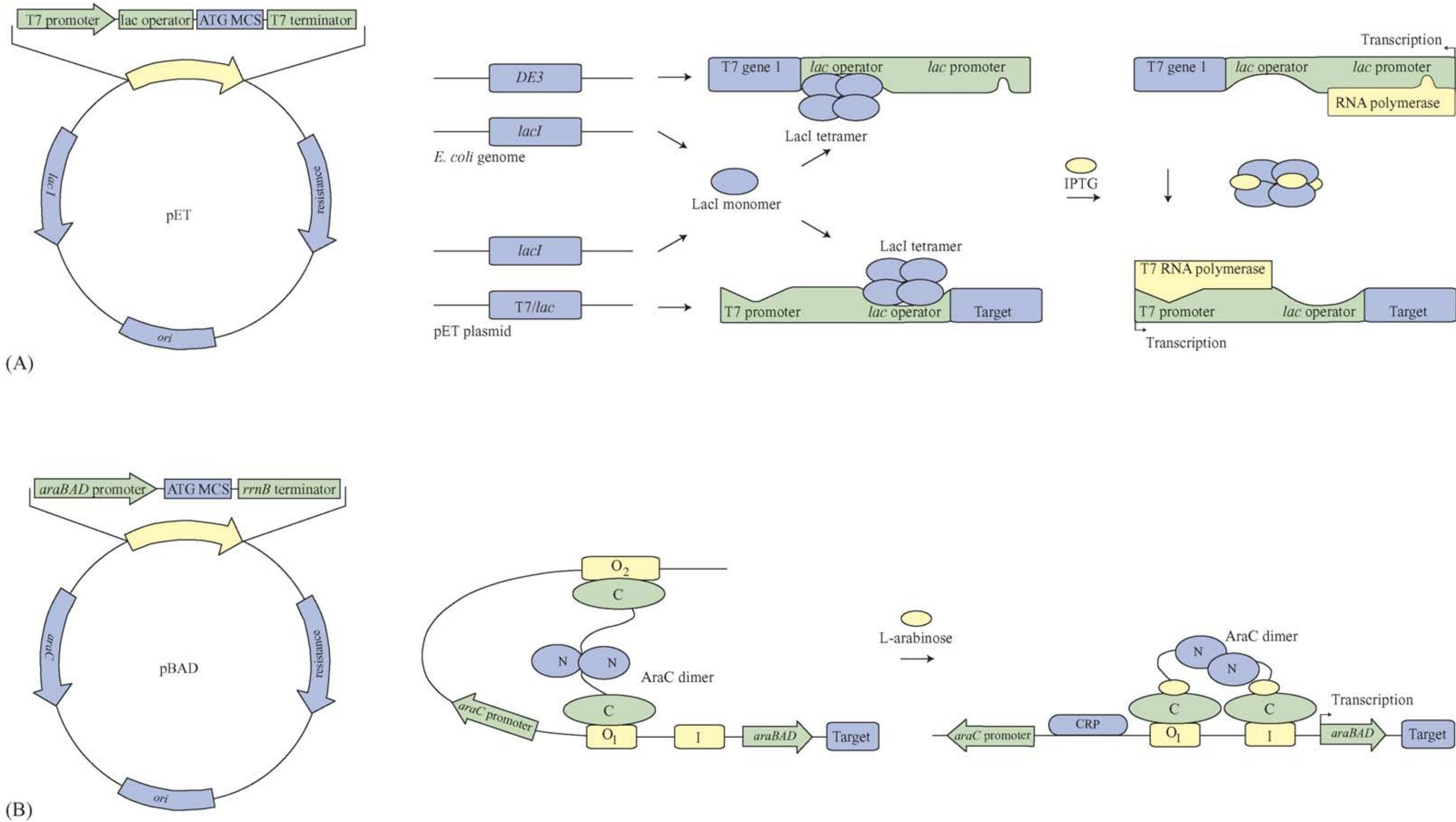


Fig. 1. Recombinant expression mechanisms. (A) The pET expression system. A general pET plasmid configuration is shown on the left. The macromolecular situations prior to and after induction are on the right (Dubendorff and Studier, 1991; Studier et al., 1990). (B) L-Arabinose induced pBAD expression plasmid (left) and system mechanism on the right (Guzman et al., 1995).

(Khlebnikov et al., 2002; Morgan-Kiss et al., 2002). Regulation of the arabinose operon in *E. coli* is directed by the product of the *araC* gene (Englesberg et al., 1969). The AraC dimer binds three sites in the arabinose operon, $O_1$, $O_2$ and I (Fig. 1B). In the absence of arabinose, the AraC dimer contacts the $O_2$ site located within the *araC* gene, 210 base pairs upstream from the *araBAD* promoter. The other half of the AraC dimer contacts the $O_1$ site in the promoter region and a DNA loop is formed. Transcription from the *araBAD* promoter and the *araC* promoter is repressed by the AraC loop conformation. Upon binding of arabinose the AraC dimer changes its conformation, binding to the $O_2$ site is replaced by binding to the I site at the *araBAD* promoter and transcription by RNA polymerase initiates. Binding of the AraC dimer to the $O_1$ and I sites is stimulated by cAMP receptor protein (CRP) and background expression from *araBAD* can be reduced by glucose mediated catabolite repression (Guzman et al., 1995). AraC regulates transcription of the AraE arabinose transporter from the *araE* promoter in a similar manner resulting in the all-or-nothing response upon induction.

## 4. *E. coli* host strains

The strain or genetic background for recombinant expression is highly important. Expression strains should be deficient in the most harmful natural proteases, maintain the expression plasmid stably and confer the genetic elements relevant to the expression system (e.g., DE3). Advantageous strains for a number of individual applications are available. *E. coli* BL21 is the most common host and has proven outstanding in standard recombinant expression applications. BL21 is a robust *E. coli* B strain, able to grow vigorously in minimal media but however non-pathogenic and unlikely to survive in host tissues and cause disease (Chart et al., 2000). BL21 is deficient in ompT and lon, two proteases that may interfere with isolation of intact recombinant protein. Derivatives of BL21 include *recA* negative strains for the stabilization of target plasmids containing repetitive sequences (Novagen BLR strain), *trxB/gor* negative mutants for the enhancement of cytoplasmic disulfide bond formation (Novagen Origami and AD494 strains), *lacY* mutants enabling adjustable levels of protein expression (Novagen Tuner series) and mutants for the soluble expression of inclusion body prone and membrane proteins (Avidis C41(DE3) and C43(DE3) strains).

## 5. Stability of the messenger RNA

Gene expression levels are mainly determined by the efficiency of transcription, mRNA stability and the frequency of mRNA translation. Transcription and translation has been subject of intense optimization in recombinant expression systems. Stability of the mRNA transcript is however rarely addressed. Gene expression is controlled by the decay of mRNA. The average half-life of mRNA in *E. coli* at 37 °C ranges from seconds to maximally 20 min and the expression rate depends directly on the inherent mRNA stability (Rauhut and Klug, 1999; Regnier and Arraiano, 2000). Messenger RNAs are degraded by RNases, primarily the two exonucleases RNase II and PNPase and the endonuclease RNase E. Protection of mRNAs from RNases depends on RNA folding, protection by ribosomes and modulation of mRNA stability by polyadenylation. Polyadenylation at the three prime end of mRNAs is provided by the PAP I and PAP II polyadenylation polymerases and facilitates degradation by RNase II and PNPase (Cao et al., 1996). Strains containing a mutation in the gene encoding RNaseE (*rne*131 mutation) are available for the enhancement of mRNA stability in recombinant expression systems (Invitrogen BL21 star strain) (Lopez et al., 1999). Control of mRNA stability in recombinant expression systems is desirable. Efficient translation initiation and consequent immediate ribosomal protection from degradation, stabilizes the mRNA and is achieved by selection of ribosomal binding sites lacking inhibitory secondary structure elements. Stable hybrid mRNAs might be constructed by implementation of efficient five prime and three prime stabilizing sequences as a barrier against exonucleases. An mRNA fragment encoding the C-terminal region of *E. coli* $F_o$ ATPase subunit was stabilized by fusion to the sequence encoding green fluorescent protein (GFP). Fusions to *lacZ* however failed to stabilize the fragment and hence the GFP transcript provided mRNA protective structural elements (Arechaga et al., 2003).

Although initiated, universal control of mRNA stabilization still remains to be conveniently incorporated into recombinant expression systems.

## 6. Rare codon interference in recombinant protein biosynthesis

Codon usage in *E. coli* is reflected by the level of cognate amino-acylated tRNAs available in the cytoplasm. Major codons occur in highly expressed genes whereas the minor or rare codons tend to be in genes expressed at low levels. Codons rare in *E. coli* are often abundant in heterologous genes from sources such as eukaryotes, archaeabacteria and other distantly related organisms with different codon frequency preferencies (Kane, 1995). Expression of genes containing rare codons can lead to translational errors, as a result of ribosomal stalling at positions requiring incorporation of amino acids coupled to minor codon tRNAs (McNulty et al., 2003). Codon bias problems become highly prevalent in recombinant expression systems, when transcripts containing rare codons in clusters, such as doublets and triplets accumulate in large quantities. Translational errors arising from rare codon bias include mistranslational amino acid substitutions, frameshifting events or premature translational termination (Kurland and Gallant, 1996; Sørensen et al., 2003c). In-frame two amino acid "hops" have been reported at a single disfavoured AGA codon (Kane et al., 1992). Protein quality is influenced by codon bias by the insertion of lysine for arginine at AGA codons (Calderone et al., 1996; Seetharam et al., 1988). Therefore, expression of full-length protein at high levels is not equivalent with translational integrity. The most problematic codons are decoded by products of the genes *argU* (AGA and AGG), *argX* (CGG), *argW* (CGA and CGG), *ileX* (AUA), *glyT* (GGA), *leuW* (CUA), *proL* (CCC) and *lys* (AAG). AAG is a major *E. coli* codon decoded by $tRNA^{Lys}UUU$, which is enabled to wobble to G by the $xm^5s^2U_{34}$ modification (Yarian et al., 2000). Since UUU reads AAG less efficient there is a problem when a target sequence contain consecutive AAG codons. Most focus has been on the rare arginine codons AGG and AGA, occurring in *E. coli* at frequencies of $\sim$0.14 and $\sim$0.21%, respectively (Kane, 1995).

Two alternative strategies are utilized to remedy codon bias. One approach is site-directed mutagenesis of the target sequence for the generation of codons reflecting the tRNA pool in the host system. This approach is beneficial for increasing expression levels and for alleviation of mistranslation (Calderone et al., 1996; Kane et al., 1992). However, a set of codon-optimized genes was recently shown to suffer from lacking mRNA transcription and stability in a recombinant expression system (Wu et al., 2004). Even though the mutagenesis approach has proven highly effective, it may be too time-consuming in high-throughput biotechnology. A less time consuming method is the co-transformation of the host with a plasmid harbouring a gene encoding the tRNA cognate to the problematic codons (Dieci et al., 2000). By increasing the copy number of the limiting tRNA species, *E. coli* can be controlled to match the codon usage frequency in heterologous genes. Several plasmids are available for rare tRNA co-expression, most of which are based on the p15A replication origin. This enables maintenance in the presence of the ColE1 replication origin present in most expression plasmids (Fig. 2). Numerous reports confirm the concept of plasmid mediated tRNA complementation (Baca and Hol, 2000; Kim et al., 1998; Sørensen et al., 2003c). Commercially Available tRNA complementation plasmids include pR.A.R.E (Novagen) and that implemented in the CodonPlus system from Stratagene.

## 7. Prevention of inclusion body formation

Protein activity demands folding into precise three-dimensional structures. Stress situations such as heat shock impair folding in vivo and folding intermediates tend to associate into amorph protein granules termed inclusion bodies. Rather little is known about the structure of inclusion bodies and the mechanism of their formation (Villaverde and Carrio, 2003). Inclusion bodies are a set of structurally complex aggregates often perceived to occur as a stress response when recombinant protein is expressed at high rates. Macromolecular crowding of proteins at concentrations of 200–300 mg/ml in the cytoplasm of *E. coli,* suggest a highly unfavorable protein-folding environment, especially during recombinant high-level expression (van den Berg et al., 1999). Whether inclusion bodies form through a passive event occurring by hydrophobic interaction between exposed patches on unfolded chains or by specific clustering mechanisms is unknown (Villaverde and Carrio, 2003). The inclusion body aggregates can be observed by optical microscopy as refractile particles of up to 2 $\mu m^3$ and by transmission electron microscopy as electron-dense aggregates

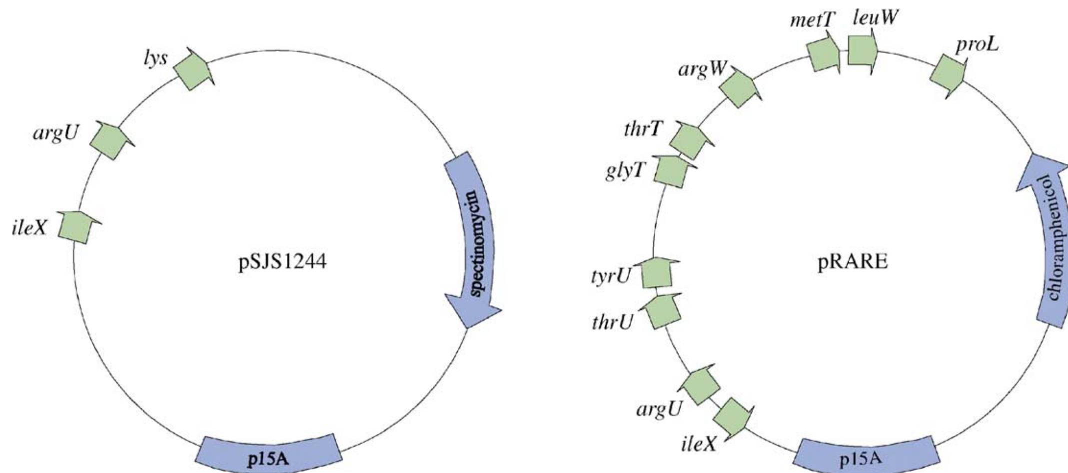


Fig. 2. Two tRNA complementation plasmids. Both plasmids carries the p15A replication origin compatible with the ColE1 origin used in most expression plasmids. Plasmid pSJS1244 is the only tRNA complementation plasmid described including a *lys* tRNA gene for the decoding of AAG (Kim et al., 1998; Sørensen et al., 2003c). Plasmid pRARE harbours ten tRNA genes and is commercialised by Novagen (Baca and Hol, 2000; Novy et al., 2001). The pRARE series of plasmids include versions encoding LacI (pLacIRARE) and T7 lysozyme (pLysSRARE).

lacking defined structure (Carrio et al., 2000). Inclusion bodies are however not inert aggregates but act as a transient reservoir for loosely packaged folding intermediates in vivo (Carrio and Villaverde, 2001). Formation of inclusion bodies in recombinant expression systems is the result of an unbalanced equilibrium between in vivo protein aggregation and solubilization. Aggregation in recombinant systems is minimized through the control of parameters such as temperature, expression rate, host metabolism, target protein engineering including solubility tag-technology and by the co-expression of plasmid-encoded chaperones (Jonasson et al., 2002).

The insoluble recombinant protein normally enriches the inclusion bodies by 50–95% of the proteineous material. Inclusion bodies are easily prepared and their degradation by proteases is limited but present both in vitro and in vivo (Carbonell and Villaverde, 2002). Proteases are directly involved in the in situ degradation of unfolded or misfolded inclusion body associated polypeptides by interaction with exposed hydrophobic patches (Carbonell and Villaverde, 2002). Arrest of recombinant protein synthesis results in the efficient removal and refolding of inclusion bodies but with most protein degraded by proteases and only low fractions reluctant to further processing (Carrio and Villaverde, 2001). The purified aggregates can be solubilized using detergents like urea and guadinium hydrochloride. Native protein can be prepared by in vitro refolding from solubilized inclusion bodies either by dilution, dialysis or on-column refolding methods (Middelberg, 2002; Sørensen et al., 2003a). Refolding strategies might be improved by inclusion of molecular chaperones (Mogk et al., 2002). Optimization of the refolding procedure for a given protein however require time consuming efforts and is not always conducive to high product yields.

A possible strategy for the prevention of inclusion body formation is the co-overexpression of molecular chaperones. This strategy is attractive but there is no guarantee that chaperones improve recombinant protein solubility. *E.coli* encode chaperones some of which drive folding attempts, whereas others prevent protein aggregation (Ehrnsperger et al., 1997; Schwarz et al., 1996; Veinger et al., 1998). As soon as newly synthesized proteins leave the exit tunnel of the *E. coli* ribosome they associate with the trigger factor chaperone (Deuerling et al., 2003). Exposed hydrophobic patches on newly synthesized proteins are protected from unintended interactions by association with trigger factor and folding premature to completion of a protein domain may be prevented (Fig. 3). Proteins can start or continue their folding into the native state after release from trigger factor. Proteins trapped in non-native and aggregation prone conformations are substrate for

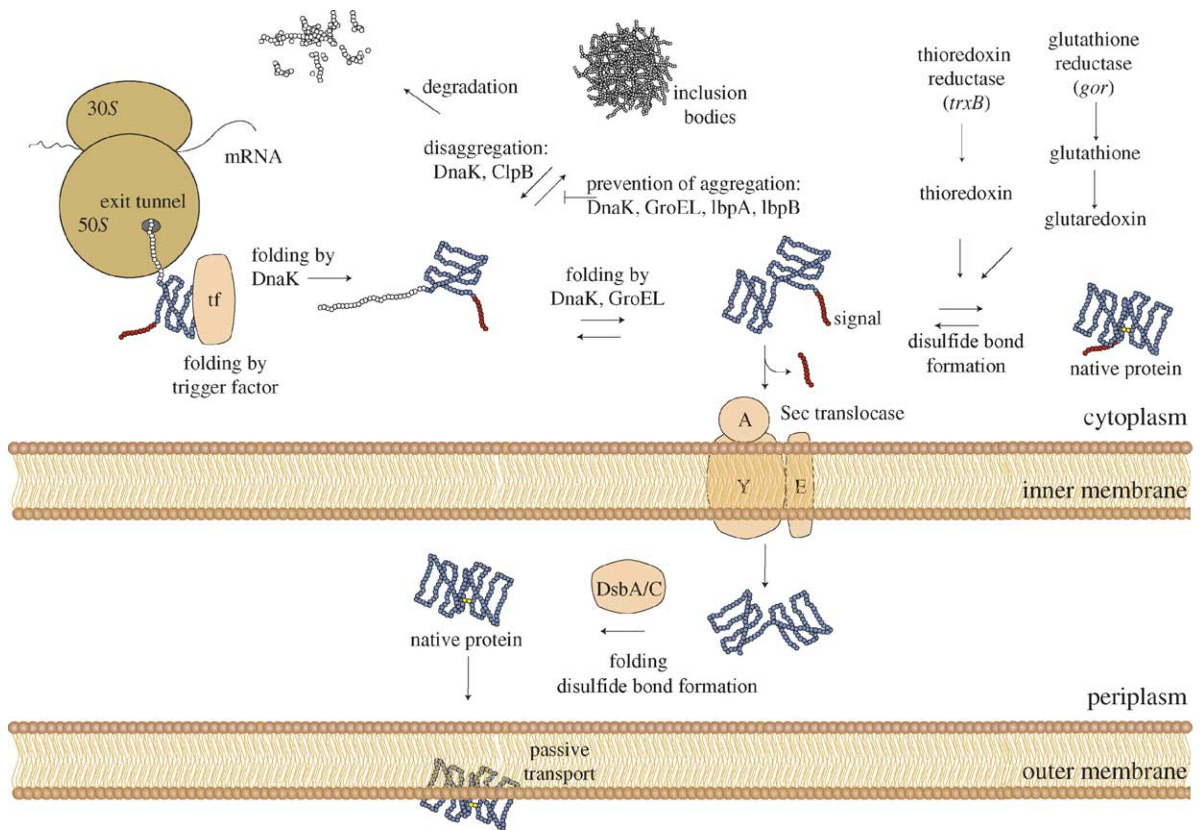


Fig. 3. Protein folding and secretion in *E. coli*. Pathways important for recombinant expression, secretion and disulfide bond formation are shown. See text for details and references.

DnaK and GroEL. DnaK (Hsp70 chaperone family) prevents the formation of inclusion bodies by reducing aggregation and promoting proteolysis of misfolded proteins (Mogk et al., 2002). A bi-chaperone system involving DnaK and ClpB (Hsp100 chaperone family) mediates the solubilization or disaggregation of proteins (Schlieker et al., 2002). GroEL (Hsp60 chaperone family) operates protein transit between soluble and insoluble protein fractions and participates positively in disaggregation and inclusion body formation. Small heat shock proteins lbpA and lbpB protect heat-denatured proteins from irreversible aggregation and have been found associated with inclusion bodies (Kitagawa et al., 2002; Kuczynska-Wisnik et al., 2002).

Simultaneous over-expression of chaperone encoding genes and recombinant target proteins proved effective in several instances. Co-overexpression of trigger factor in recombinants prevented the aggregation of mouse endostatin, human oxygen-regulated protein ORP150, human lysozyme and guinea pig liver transglutaminase (Ikura et al., 2002; Nishihara et al., 2000). Soluble expression was further stimulated by the co-overexpression of the GroEL–GroES and DnaK–DnaJ–GrpE chaperone systems along with trigger factor (Nishihara et al., 2000). The chaperone systems are cooperative and the most favorable strategies involve co-expression of combinations of chaperones belonging to the GroEL, DnaK, ClpB and ribosome associated trigger factor families of chaperones (Amrein et al., 1995; Nishihara et al., 1998).

Two *E. coli* mutant strains have contributed significantly to the soluble expression of difficult recombinant proteins. C41(DE3) and C43(DE3) are mutants that allow over-expression of some globular and membrane proteins unable to express at high-levels in the parent strain BL21(DE3) (Miroux and Walker, 1996).

Expression of the $F_1F_o$ ATP synthase subunit b membrane protein in these strains, in particular C43(DE3) is accompanied by the proliferation of intracellular membranes and inclusion bodies are absent (Arechaga et al., 2000). These strains are now commercialized by Avidis (http://www.avidis.fr) and a high number of reports on their use in expression of difficult proteins have been published (Arechaga et al., 2003; Smith and Walker, 2003; Steinfels et al., 2002; Sørensen et al., 2003c).

## 8. Stress response induced by recombinant *E. coli*

The maintenance of a plasmid often induces a stress response especially when a target protein is highly expressed (Hoffmann and Rinas, 2004). Such stress responses resembles environmental stress situations such as heat shock, amino acid depletion or starvation. Stress induced by plasmid maintenance is often related to plasmid copy number (Bailey, 1993), while the main perturbation can be attributed to genes encoded by the plasmid and even constitutively expressed genes such as antibiotic resistance genes (Hoffmann et al., 2002). Some proteins directly influence host cellular metabolism by their enzymatic properties, but in general expression of recombinant proteins induce a "metabolic burden". The metabolic burden is defined as the amount of resources (raw material and energy) that are withdrawn from the host metabolism for maintenance and expression of the foreign DNA (Bentley and Kompala, 1990). In general the specific growth rate of cells expressing a product correlates inversely with the rate of recombinant protein synthesis (Dong et al., 1995; Hoffmann and Rinas, 2004). The expression of recombinant proteins therefore, usually results in impaired growth rates and lowered increase in biomass. This is a direct response to the high-energy requirements induced by recombinant protein synthesis, the synthesis of stress proteins and elevated respiration rates (Hoffmann and Rinas, 2004).

The response triggered by the cells under energy limiting conditions is extremely complex and includes the activation of alternative pathways for energy generation and adjustment of the level of energy generating enzymes. Recombinant expression results in high rates of protein synthesis. However, while the recombinant protein is highly expressed, housekeeping genes including components of the protein synthesis machinery are down regulated (Hoffmann et al., 2002). Amino acid starvation tends to occur during recombinant expression if the product deviates considerably from the average *E. coli* protein. The response includes an extensive reprogramming of gene expression patterns and down regulation of the majority of genes involved in transcription, translation and amino acid biosynthesis (Chang et al., 2002). Addition of the appropriate amino acid(s) can alleviate this phenomenon known as the stringent response.

Another response to stress induced by recombinant expression is an increase of the in vivo proteolysis of the target protein. This response has been circumvented by the use of protease deficient host strains, heat shock deficient strains, chaperone co-expression and protease inhibitor co-expression (Dong et al., 1995). These strategies rely on engineering of the host. Other strategies target the specific product, which can be stabilized by fusion protein technology and site directed mutagenesis at protease specific sites or directed differently by a signal peptide (e.g., to inclusion bodies or the periplasm).

Stress can be reduced in recombinant systems by slow adaptation of cells to a specific production task. This can be accomplished by gradually increasing the level of inducer or by slowly increasing the plasmid copy number during cultivation (Trepod and Mott, 2002). Stress situations can clearly be avoided and should be circumvented if the desired quality and quantity of recombinant protein is impaired.

## 9. Fusion protein technology and cleavage by site specific proteolysis

A wide range of protein fusion partners has been developed in order to simplify the purification and expression of recombinant proteins (Stevens, 2000). Fusion proteins or chimeric proteins usually include a partner or "tag" linked to the passenger or target protein by a recognition site for a specific protease. Most fusion partners are exploited for specific affinity purification strategies. Fusion partners are also advantageous in vivo, where they might protect passengers from intracellular proteolysis (Jacquet et al., 1999; Martinez et al., 1995), enhance solubility (Davis et al., 1999; Kapust and Waugh, 1999; Sørensen et al., 2003b) or

be used as specific expression reporters (Waldo et al., 1999). High expression levels can often be transferred from a N-terminal fusion partner, to a poorly expressing passenger, most probably as a result of mRNA stabilization (Arechaga et al., 2003). Common affinity tags are the polyhistidine tag (His-tag), which is compatible with immobilized metal affinity chromatography (IMAC) and the glutathione *S*-transferase (GST) tag for purification on glutathione based resins. Several other affinity tags exist and have been extensively reviewed (Terpe, 2003).

Fusion partners of particular interest with regard to optimization of recombinant expression, include the *E. coli* maltose binding protein (MBP) and *E. coli* N-utilizing substance A (NusA). MBP (40 kDa) and NusA (54.8 kDa) act as solubility enhancing partners and are especially suited for the expression of inclusion body prone proteins. Although many proteins are highly soluble, they are not all effective as solubility enhancers. *E. coli* MBP proved to be a much more effective solubility partner than the highly soluble GST and thioredoxin proteins in a comparison of solubility enhancing properties (Kapust and Waugh, 1999). Solubility enhancement is a common trait of maltodextrin-binding proteins (MBPs) from a number of organisms and some of them are even more effective than *E. coli* MBP (Fox et al., 2003). A precise mechanism for the solubility enhancement of MBP has not been found. However, MBP might act as a chaperone by interactions through a solvent exposed "hot spot" on its surface, which stabilizes the otherwise insoluble passenger protein (Bach et al., 2001; Fox et al., 2001).

Wilkinson and Harrison proposed a model for the theoretical calculation of solubility percentages of recombinant proteins expressed in the *E. coli* cytoplasm (Wilkinson and Harrison, 1991). A web-server for the calculation of this index is found at http://www.biotech.ou.edu. The Wilkinson–Harrison model along with experimental data identified NusA as a highly favorable solubility partner (Davis et al., 1999). The major advantage of NusA, in addition to the good solubility characteristics, is its high expressivity. Both MBP and NusA have been used for the solubilization of highly insoluble ScFv antibodies in the cytoplasm of *E. coli* (Bach et al., 2001; Zheng et al., 2003).

MBP and NusA are relatively large fusion partners. We recently suggested the use of a highly soluble N-terminal fragment of translation initiation factor IF2 (17.4 kDa) as a solubility partner (Sørensen et al., 2003b). The use of a small partner reduces the amount of energy required to obtain a certain number of molecules, diminishes steric hindrance and simplify applications such as NMR. The outcome of fusion to a solubility partner is protein specific and is not a universal method for the prevention of inclusion-body formation.

A newly introduced strategy is to screen for soluble proteins using a folding reporter. Fluorescence of *E. coli* cells expressing target genes fused to GFP is related to the solubility of the target gene expressed alone (Waldo et al., 1999). Hence, protein folding in *E. coli* can be improved by directed evolution approaches for a certain target protein by screening for fluorescing mutants. This approach evolved three insoluble proteins including *Pyrobaculum aerophilum* methyl transferase, tartrate dehydratase β-subunit and nucleoside diphosphate kinase to be 50, 95 and 90% soluble, respectively (Pedelacq et al., 2002). The GFP reporter system was further used to screen for solubilizing interaction partners to insoluble targets. Fusion of integration host factor β upstream to GFP resulted in aggregation, whereas co-expression of the binding partner, integration host factor α, increased fluorescence dramatically (Wang and Chong, 2003).

Typically, it is desirable to separate the recombinant protein from exploited fusion partners such as affinity tags, solubility enhancers or expression reporters. This is achieved by site-specific proteolysis of the isolated fusion protein, in vitro. Two serine proteases belonging to the eukaryotic blood-clotting cascade, namely factor Xa and thrombin are extensively employed (Jenny et al., 2003). Factor Xa cleaves at the amino acid sequence IEGR/X, where X can be any amino acid except arginine or proline. Thrombin recognizes the sequence LVPR/G.

While these enzymes are highly efficient for cleavage at the inserted recognition sequence, proteolysis is frequently occurring at other sites in target proteins. More specific proteases in use include enterokinase (recognizes DDDDK/X, where X can be any amino acid except proline) and the highly specific precision protease (Amersham Biosciences), which cleaves at LEVLFQ/GP (Walker et al., 1994). The latter is a picornavirus 3C protease, a class of proteases that have

not been reported to cleave fusion proteins at unintended locations. Another 3C protease has come in use, namely the tobacco etch virus protease (TEV). This protease cleaves the sequence ENLYFQ/G efficiently (Kapust et al., 2002). TEV is extensively used for in vitro cleavage of fusion proteins but it has also found use for controlled intracellular fusion protein processing (Kapust and Waugh, 2000). Co-expression of TEV from a recombinant plasmid can be used as a tool to study cleavage of a recombinant target protein in vivo and is relevant for the detection of cleavage problems before cost-effective purification procedures are initiated (Ehrmann et al., 1997; Herskovits et al., 2001; Kapust and Waugh, 2000).

Selection of optimal reaction conditions and a specific protease depends on the recombinant target protein. Hence, the use of proteases for fusion protein cleavage is not a trivial procedure.

## 10. Secretion of recombinant proteins and disulfide bond formation

Recombinantly expressed proteins can in principle be directed to three different locations namely the cytoplasm, the periplasm or the cultivation medium. Various advantages and disadvantages are related to the direction of a recombinant protein to a specific cellular compartment. Expression in the cytoplasm is normally preferable since production yields are high. Disulfide bond formation is segregated in *E. coli* and is actively catalyzed in the periplasm by the Dsb system (Rietsch and Beckwith, 1998). Reduction of cysteines in the cytoplasm is achieved by thioredoxin and glutaredoxin. Thioredoxin is kept reduced by thioredoxin reductase and glutaredoxin by glutathione. The low molecular weight glutathione molecule is reduced by glutathione reductase (Fig. 3). Disruption of the *trxB* and *gor* genes encoding the two reductases, allow the formation of disulfide bonds in the *E. coli* cytoplasm. The *trxB* (Novagen AD494) and *trxB/gor* (Novagen Origami) negative strains of *E. coli* have been selected in several expression situations (Bessette et al., 1999; Lehmann et al., 2003; Premkumar et al., 2003). Folding and disulfide bond formation in the target protein is enhanced by fusion to thioredoxin in strains lacking thioredoxin reductase (*trxB*) (Stewart et al., 1998). Overexpression of the periplasmic foldase DsbC in the cytoplasm stimulates disulfide bond formation further (Bessette et al., 1999).

Transmembrane transport is normally mediated by N-terminal signal peptides by direction of the protein to a specific transporter complex in the membrane (Fig. 3). Most proteins are exported across the inner membrane to the periplasm by the well-known Sec translocase apparatus (Manting and Driessen, 2000). Frequently used periplasmic leader sequences for potential export are derived from ompT, ompA, pelB, phoA, malE, lamB and β-lactamase (Blight et al., 1994). Systems are available for the potential export and enhanced disulfide bond formation via fusion to DsbA or DsbC, the enzymes catalyzing disulfide bond formation and isomerization in the periplasm (Collins-Racie et al., 1995). A direct consequence of periplasmic production is a considerable reduction in the amount of contaminating proteins in the starting material for purification. Other benefits include the much higher probability of obtaining an authentic N-terminus in the target protein, decreased proteolysis and simplified protein release by osmotic shock procedures (Jonasson et al., 2002).

Efficient pathways for translocation through the outer membrane are absent, albeit some proteins exported to the periplasm diffuse or leaks into the extra cellular medium. Passive transport across the outer membrane can be stimulated by external or internal destabilization of the *E. coli* structural components. Destabilization is achieved either by lysis proteins working from the interior of the cell, by using strains lacking structural membrane components or by permeabilization directed from the cell exterior either mechanically, enzymatic or chemically (Shokri et al., 2003). Another strategy involves the engineering of secretion mechanisms into *E. coli* either from pathogenic *E. coli* or other species. Direction of recombinant proteins to the periplasm often results in protein leak to the extra cellular cultivation medium. This uncontrolled strategy enabled the purification of potato carboxypeptidase inhibitor and cholera toxin B subunit (Molina et al., 1992; Slos et al., 1994). The ompA signal sequence was recently used to translocate a recombinant peptide to the periplasm for probable secretion to the cultivation medium. Translocation was enhanced by co-expression of two secretion factors (secE and secY) and the level of recombinant peptide in the cultivation medium increased (Ray et al., 2002). Recombinant proteins prob-

ably leave the periplasm passively through destabilized membrane structures, either when cells age or when culture conditions change. However, detailed knowledge and standardized methods for directed secretion are missing.

## 11. Systems for co-overexpression of multiple targets

Elucidation of macromolecular structure as well as functional investigation migrates towards even more complicated entities. These studies often require preparation of large quantity multi-component protein complex. Complex production in vivo has therefore gained increasing interest. In vivo preparation of protein complexes is achieved by plasmid-mediated co-expression of the cognate interaction partners. The in vivo co-expression approach has multiple advantages as compared to in vitro complex reconstitution from isolated components. Several reports indicate the importance of an interaction partner for the proper in vivo folding of a recombinant protein. Co-expression often results in increased amounts of properly folded target protein, in some instances protected from proteolytic degradation by another component of the complex (Li et al., 1997; Stebbins et al., 1999; Tan, 2001). Two general strategies are available, namely co-expression from two separate plasmids maintained in the cell simultaneously, or expression of multiple recombinant proteins from a plasmid polycistron. More than two plasmids are difficult two maintain in *E. coli*, since each plasmid must replicate from unique and compatible replicons. Different selectable markers are obviously necessary as well. A polycistronic plasmid allows for the co-overexpression of more than two genes (Tan, 2001). Such a system successfully expressed binary and ternary complexes including the VHL-elonginC-elonginB complex (Stebbins et al., 1999). Another study used double cistronic vectors to gain a dramatic increase in soluble expression of both interaction partners in a heterodimeric receptor complex (Li et al., 1997).

A new system for double cistronic co-expression of maximally eight recombinant proteins from four different plasmids have recently been commercialized by Novagen. Each plasmid carries different replication origins namely ColE1, p15A, RSF and CDF (Held et al., 2003). Similarly four different selectable markers are used (spectinomycin, kanamycin, chloramphenicol and ampicillin). Future challenges in recombinant co-expression will elucidate the amenability of this and similar systems.

## 12. Conclusions

We have reviewed the most recent improvements in recombinant expression of proteins in *E. coli* as well as the difficulties arising from this unnatural stress situation. Improvement of recombinant expression relies on the modulation and circumvention of many issues such as mRNA stability, codon bias and inclusion body formation. Genetic strategies are the primary source of innovation for recombinant expression in *E. coli* and the limits are constantly pushed as we learn. We conclude that the primary key to successful preparation of recombinant proteins in *E. coli* is the skillfull combination of the utensils from the vast genetic toolbox.

## Acknowledgements

The authors thank Brian Søgaard Laursen, Janni Egebjerg Kristensen and Max Vejen, Department of Molecular Biology, Aarhus University, Denmark for critical reading of the manuscript. K.K.M. is funded by grants from the Danish Natural Science Research Council and Carlsberg (Grants no. 21-03-0465 and ANS-0987/40).

## References

Amrein, K.E., Takacs, B., Stieger, M., Molnos, J., Flint, N.A., Burn, P., 1995. Purification and characterization of recombinant human p50csk protein–tyrosine kinase from an *Escherichia coli* expression system overproducing the bacterial chaperones GroES and GroEL. Proc. Natl. Acad. Sci. U. S. A. 92, 1048–1052.

Arechaga, I., Miroux, B., Karrasch, S., Huijbregts, R., de Kruijff, B., Runswick, M.J., Walker, J.E., 2000. Characterisation of new intracellular membranes in *Escherichia coli* accompanying large scale over-production of the b subunit of F(1)F(o) ATP synthase. FEBS Lett. 482, 215–219.

Arechaga, I., Miroux, B., Runswick, M.J., Walker, J.E., 2003. Over-expression of *Escherichia coli* F1F(o)-ATPase subunit a is inhibited by instability of the *uncB* gene transcript. FEBS Lett. 547, 97–100.

Baca, A.M., Hol, W.G., 2000. Overcoming codon bias: a method for high-level overexpression of *Plasmodium* and other AT-rich parasite genes in *Escherichia coli*. Int. J. Parasitol. 30, 113–118.

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

Bailey, J.E., 1993. Host-vector interactions in *Escherichia coli*. Adv. Biochem. Eng. Biotechnol. 48, 29–52.

Baneyx, F., 1999. Recombinant protein expression in *Escherichia coli*. Curr. Opin. Biotechnol. 10, 411–421.

Bentley, W.E., Kompala, D.S., 1990. Optimal induction of protein synthesis in recombinant bacterial cultures. Ann. N. Y. Acad. Sci. 589, 121–138.

Bessette, P.H., Aslund, F., Beckwith, J., Georgiou, G., 1999. Efficient folding of proteins with multiple disulfide bonds in the *Escherichia coli* cytoplasm. Proc. Natl. Acad. Sci. U. S. A. 96, 13703–13708.

Blight, M.A., Chervaux, C., Holland, I.B., 1994. Protein secretion pathway in *Escherichia coli*. Curr. Opin. Biotechnol. 5, 468–474.

Calderone, T.L., Stevens, R.D., Oas, T.G., 1996. High-level misincorporation of lysine for arginine at AGA codons in a fusion protein expressed in *Escherichia coli*. J. Mol. Biol. 262, 407–412.

Calos, M.P., 1978. DNA sequence for a low-level promoter of the *lac* repressor gene and an 'up' promoter mutation. Nature 274, 762–765.

Cao, G.J., Pogliano, J., Sarkar, N., 1996. Identification of the coding region for a second poly(A) polymerase in *Escherichia coli*. Proc. Natl. Acad. Sci. U. S. A. 93, 11580–11585.

Carbonell, X., Villaverde, A., 2002. Protein aggregated into bacterial inclusion bodies does not result in protection from proteolytic digestion. Biotechnol. Lett. 24, 1939–1944.

Carrio, M.M., Cubarsi, R., Villaverde, A., 2000. Fine architecture of bacterial inclusion bodies. FEBS Lett. 471, 7–11.

Carrio, M.M., Villaverde, A., 2001. Protein aggregation as bacterial inclusion bodies is reversible. FEBS Lett. 489, 29–33.

Chang, D.E., Smalley, D.J., Conway, T., 2002. Gene expression profiling of *Escherichia coli* growth transitions: an expanded stringent response model. Mol. Microbiol. 45, 289–306.

Chart, H., Smith, H.R., La Ragione, R.M., Woodward MJ, 2000. An investigation into the pathogenic properties of *Escherichia coli* strains BLR, BL21, DH5alpha and EQ1. J. Appl. Microbiol. 89, 1048–1058.

Collins-Racie, L.A., McColgan, J.M., Grant, K.L., DiBlasio-Smith, E.A., McCoy, J.M., LaVallie, E.R., 1995. Production of recombinant bovine enterokinase catalytic subunit in *Escherichia coli* using the novel secretory fusion partner DsbA. Biotechnology (N. Y.) 13, 982–987.

Connell, S.R., Tracz, D.M., Nierhaus, K.H., Taylor, D.E., 2003. Ribosomal protection proteins and their mechanism of tetracycline resistance. Antimicrob. Agents Chemother. 47, 3675–3681.

Davis, G.D., Elisee, C., Newham, D.M., Harrison RG, 1999. New fusion protein systems designed to give soluble expression in *Escherichia coli*. Biotechnol. Bioeng. 65, 382–388.

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

Deuerling, E., Patzelt, H., Vorderwulbecke, S., Rauch, T., Kramer, G., Schaffitzel, E., Mogk, A., Schulze-Specking, A., Langen, H., Bukau B, 2003. Trigger Factor and DnaK possess overlapping substrate pools and binding specificities. Mol. Microbiol. 47, 1317–1328.

Dieci, G., Bottarelli, L., Ballabeni, A., Ottonello, S., 2000. tRNA-assisted overproduction of eukaryotic ribosomal proteins. Protein Expr. Purif. 18, 346–354.

Dong, H., Nilsson, L., Kurland, C.G., 1995. Gratuitous overexpression of genes in *Escherichia coli* leads to growth inhibition and ribosome destruction. J. Bacteriol. 177, 1497–1504.

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

Ehrmann, M., Bolek, P., Mondigler, M., Boyd, D., Lange, R., 1997. TnTIN and TnTAP: mini-transposons for site-specific proteolysis in vivo. Proc. Natl. Acad. Sci. U. S. A. 94, 13111–13115.

Ehrnsperger, M., Graber, S., Gaestel, M., Buchner, J., 1997. Binding of non-native protein to Hsp25 during heat shock creates a reservoir of folding intermediates for reactivation. EMBO J. 16, 221–229.

Englesberg, E., Squires, C., Meronk Jr., F., 1969. The L-arabinose operon in *Escherichia coli* B-r: a genetic demonstration of two functional states of the product of a regulator gene. Proc. Natl. Acad. Sci. U. S. A. 62, 1100–1107.

Fox, J.D., Kapust, R.B., Waugh, D.S., 2001. Single amino acid substitutions on the surface of *Escherichia coli* maltose-binding protein can have a profound impact on the solubility of fusion proteins. Protein Sci. 10, 622–630.

Fox, J.D., Routzahn, K.M., Bucher, M.H., Waugh, D.S., 2003. Maltodextrin-binding proteins from diverse bacteria and archaea are potent solubility enhancers. FEBS Lett. 537, 53–57.

Guzman, L.M., Belin, D., Carson, M.J., Beckwith, J., 1995. Tight regulation, modulation and high-level expression by vectors containing the arabinose PBAD promoter. J. Bacteriol. 177, 4121–4130.

Hannig, G., Makrides, S.C., 1998. Strategies for optimizing heterologous protein expression in *Escherichia coli*. Trends Biotechnol. 16, 54–60.

Hardy, K.G., 1987. Plasmids—A Practical Approach. IRL Press, Oxford.

Held, D., Yaeger, K., Novy, R., 2003. New coexpression vectors for expanded compatibilities in *E. coli*. InNovations 18, 4–6.

Herskovits, A.A., Seluanov, A., Rajsbaum, R., ten Hagen-Jongman, C.M., Henrichs, T., Bochkareva, E.S., Phillips, G.J., Probst, F.J., Nakae, T., Ehrmann, M., Luirink, J., Bibi, E., 2001. Evidence for coupling of membrane targeting and function of the signal recognition particle (SRP) receptor FtsY. EMBO Rep. 2, 1040–1046.

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

Hoffmann, F., Weber, J., Rinas, U., 2002. Metabolic adaptation of *Escherichia coli* during temperature-induced recombinant protein production. Part 1. readjustment of metabolic enzyme synthesis. Biotechnol. Bioeng. 80, 313–319.

Ikura, K., Kokubu, T., Natsuka, S., Ichikawa, A., Adachi, M., Nishihara, K., Yanagi, H., Utsumi, S., 2002. Co-overexpression of folding modulators improves the solubility of the recombinant

guinea pig liver transglutaminase expressed in *Escherichia coli*. Prep. Biochem. Biotechnol. 32, 189–205.

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

Jenny, R.J., Mann, K.G., Lundblad, R.L., 2003. A critical review of the methods for cleavage of fusion proteins with thrombin and factor Xa. Protein Expr. Purif. 31, 1–11.

Jonasson, P., Liljeqvist, S., Nygren, P.A., Stahl, S., 2002. Genetic design for facilitated production and recovery of recombinant proteins in *Escherichia coli*. Biotechnol. Appl. Biochem. 35, 91–105.

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.

Kane, J.F., Violand, B.N., Curran, D.F., Staten, N.R., Duffin, K.L., Bogosian, G., 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, 6707–6712.

Kapust, R.B., Tozser, J., Copeland, T.D., Waugh, D.S., 2002. The P1' specificity of tobacco etch virus protease. Biochem. Biophys. Res. Commun. 294, 949–955.

Kapust, R.B., Waugh, D.S., 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.

Kapust, R.B., Waugh, D.S., 2000. Controlled intracellular processing of fusion proteins by TEV protease. Protein Expr. Purif. 19, 312–318.

Khlebnikov, A., Keasling, J.D., 2002. Effect of *lac*Y expression on homogeneity of induction from the P(tac) and P(trc) promoters by natural and synthetic inducers. Biotechnol. Prog. 18, 672–674.

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

Kim, R., Sandler, S.J., Goldman, S., Yokota, H., Clark, A.J., Kim, S.H., 1998. Overexpression of archaeal proteins in *Escherichia coli*. Biotechnol. Lett. 20, 207–210.

Kitagawa, M., Miyakawa, M., Matsumura, Y., Tsuchido, T., 2002. *Escherichia coli* small heat shock proteins, IbpA and IbpB, protect enzymes from inactivation by heat and oxidants. Eur. J. Biochem. 269, 2907–2917.

Kuczynska-Wisnik, D., Kedzierska, S., Matuszewska, E., Lund, P., Taylor, A., Lipinska, B., Laskowska, E., 2002. The *Escherichia coli* small heat-shock proteins IbpA and IbpB prevent the aggregation of endogenous proteins denatured in vivo during extreme heat shock. Microbiology 148, 1757–1765.

Kurland, C., Gallant, J., 1996. Errors of heterologous protein expression. Curr. Opin. Biotechnol. 7, 489–493.

Laursen, B.S., Steffensen, S., Hedegaard, J., Moreno, J.M., Mortensen, K.K., Sperling-Petersen, H.U., 2002. Structural requirements of the mRNA for intracistronic translation initiation of the enterobacterial *infB* gene. Genes Cells. 7, 901–910.

Lehmann, K., Hoffmann, S., Neudecker, P., Suhr, M., Becker, W.M., Rosch, P., 2003. High-yield expression in *Escherichia coli*, purification and characterization of properly folded major peanut allergen Ara h 2. Protein Expr. Purif. 31, 250–259.

Li, C., Schwabe, J.W., Banayo, E., Evans, R.M., 1997. Coexpression of nuclear receptor partners increases their solubility and biological activities. Proc. Natl. Acad. Sci. U. S. A. 94, 2278–2283.

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

Manting, E.H., Driessen, A.J., 2000. *Escherichia coli* translocase: the unravelling of a molecular machine. Mol. Microbiol. 37, 226–238.

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

Mayer, M.P., 1995. A new set of useful cloning and expression vectors derived from pBlueScript. Gene 163, 41–46.

McNulty, D.E., Claffee, B.A., Huddleston, M.J., Kane, J.F., 2003. Mistranslational errors associated with the rare arginine codon CGG in *Escherichia coli*. Protein Expr. Purif. 27, 365–374.

Middelberg, A., 2002. Preparative protein refolding. Trends Biotechnol. 20, 437.

Miroux, B., Walker, J.E., 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, 289–298.

Mogk, A., Mayer, M.P., Deuerling, E., 2002. Mechanisms of protein folding: molecular chaperones and their application in biotechnology. Chembiochem 3, 807–814.

Molina, M.A., Aviles, F.X., Querol, E., 1992. Expression of a synthetic gene encoding potato carboxypeptidase inhibitor using a bacterial secretion vector. Gene 116, 129–138.

Morgan-Kiss, R.M., Wadler, C., Cronan Jr., J.E., 2002. Long-term and homogeneous regulation of the *Escherichia coli araBAD* promoter by use of a lactose transporter of relaxed specificity. Proc. Natl. Acad. Sci. U. S. A. 99, 7373–7377.

Newbury, S.F., Smith, N.H., Robinson, E.C., Hiles, I.D., Higgins, C.F., 1987. Stabilization of translationally active mRNA by prokaryotic REP sequences. Cell 48, 297–310.

Nishihara, K., Kanemori, M., Kitagawa, M., Yanagi, H., Yura, T., 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. Environ. Microbiol. 64, 1694–1699.

Nishihara, K., Kanemori, M., Yanagi, H., Yura, T., 2000. Overexpression of trigger factor prevents aggregation of recombinant proteins in *Escherichia coli*. Appl. Environ. Microbiol. 66, 884–889.

Novy, R., Yaeger, K., Mierendorf, R., 2001. Overcoming the codon bias of *E. coli* for enhanced protein expression. InNovations 12, 1–3.

Pedelacq, J.D., Piltch, E., Liong, E.C., Berendzen, J., Kim, C.Y., Rho, B.S., Park, M.S., Terwilliger, T.C., Waldo, G.S., 2002. Engineering soluble proteins for structural genomics. Nat. Biotechnol. 20, 927–932.

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

Premkumar, L., Bageshwar, U.K., Gokhman, I., Zamir, A., Sussman, J.L., 2003. An unusual halotolerant alpha-type carbonic anhydrase from the alga *Dunaliella salina* functionally expressed in *Escherichia coli*. Protein Expr. Purif. 28, 151–157.

Rauhut, R., Klug, G., 1999. mRNA degradation in bacteria. FEMS Microbiol. Rev. 23, 353–370.

Ray, M.V., Meenan, C.P., Consalvo, A.P., Smith, C.A., Parton, D.P., Sturmer, A.M., Shields, P.P., Mehta, N.M., 2002. Production of salmon calcitonin by direct expression of a glycine-extended precursor in *Escherichia coli*. Protein Expr. Purif. 26, 249–259.

Regnier, P., Arraiano, C.M., 2000. Degradation of mRNA in bacteria: emergence of ubiquitous features. Bioessays 22, 235–244.

Rietsch, A., Beckwith, J., 1998. The genetics of disulfide bond metabolism. Annu. Rev. Genet. 32, 163–184.

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

Schlieker, C., Bukau, B., Mogk, A., 2002. Prevention and reversion of protein aggregation by molecular chaperones in the *E. coli* cytosol: implications for their applicability in biotechnology. J. Biotechnol. 96, 13–21.

Schwarz, E., Lilie, H., Rudolph, R., 1996. The effect of molecular chaperones on in vivo and in vitro folding processes. Biol. Chem. 377, 411–416.

Seetharam, R., Heeren, R.A., Wong, E.Y., Braford, S.R., Klein, B.K., Aykent, S., Kotts, C.E., Mathis, K.J., Bishop, B.F., Jennings, M.J., et al., 1988. Mistranslation in IGF-1 during over-expression of the protein in *Escherichia coli* using a synthetic gene containing low frequency codons. Biochem. Biophys. Res. Commun. 155, 518–523.

Shokri, A., Sanden, A.M., Larsson, G., 2003. Cell and process design for targeting of recombinant protein into the culture medium of *Escherichia coli*. Appl. Microbiol. Biotechnol. 60, 654–664.

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

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

Smith, V.R., Walker, J.E., 2003. Purification and folding of recombinant bovine oxoglutarate/malate carrier by immobilized metal-ion affinity chromatography. Protein Expr. Purif. 29, 209–216.

Stebbins, C.E., Kaelin Jr., W.G., Pavletich, N.P., 1999. Structure of the VHL-ElonginC-ElonginB complex: implications for VHL tumor suppressor function. Science 284, 455–461.

Steinfels, E., Orelle, C., Dalmas, O., Penin, F., Miroux, B., Di Pietro, A., Jault, J.M., 2002. Highly efficient over-production in *E. coli* of YvcC, a multidrug-like ATP-binding cassette transporter from *Bacillus subtilis*. Biochim. Biophys. Acta 1565, 1–5.

Stenstrom, C.M., Jin, H., Major, L.L., Tate, W.P., Isaksson, L.A., 2001. Codon bias at the 3′-side of the initiation codon is correlated with translation initiation efficiency in *Escherichia coli*. Gene 263, 273–284.

Stevens, R.C., 2000. Design of high-throughput methods of protein production for structural biology. Struct. Fold Des. 8, R177–R185.

Stewart, E.J., Aslund, F., Beckwith, J., 1998. Disulfide bond formation in the *Escherichia coli* cytoplasm: an in vivo role reversal for the thioredoxins. EMBO J. 17, 5543–5550.

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

Studier, F.W., Rosenberg, A.H., Dunn, J.J., Dubendorff, J.W., 1990. Use of T7 RNA polymerase to direct expression of cloned genes. Methods Enzymol. 185, 60–89.

Summers, D., 1998. Timing, self-control and a sense of direction are the secrets of multicopy plasmid stability. Mol. Microbiol. 29, 1137–1145.

Sørensen, H.P., Laursen, B.S., Mortensen, K.K., Sperling-Petersen, H.U., 2002. Bacterial translation initiation–mechanism and regulation. Recent Res. Dev. Biophys. Biochem. 2, 243–270.

Sørensen, H.P., Sperling-Petersen, H.U., Mortensen, K.K., 2003a. Dialysis strategies for protein refolding. Preparative streptavidin production. Protein Expr. Purif. 32, 252–259.

Sørensen, H.P., Sperling-Petersen, H.U., Mortensen, K.K., 2003b. A favorable solubility partner for the recombinant expression of streptavidin. Protein Expr. Purif. 32, 252–259.

Sørensen, H.P., Sperling-Petersen, H.U., Mortensen KK, 2003c. Production of recombinant thermostable proteins expressed in *Escherichia coli*: completion of protein synthesis is the the bottleneck. J. Chromatogr. B 786, 207–214.

Tan, S., 2001. A modular polycistronic expression system for overexpressing protein complexes in *Escherichia coli*. Protein Expr. Purif. 21, 224–234.

Terpe, K., 2003. Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. Appl. Microbiol. Biotechnol. 60, 523–533.

Trepod, C.M., Mott, J.E., 2002. A spontaneous runaway vector for production-scale expression of bovine somatotropin from *Escherichia coli*. Appl. Microbiol. Biotechnol. 58, 84–88.

van den Berg, B., Ellis, R.J., Dobson, C.M., 1999. Effects of macromolecular crowding on protein folding and aggregation. EMBO J. 18, 6927–6933.

Veinger, L., Diamant, S., Buchner, J., Goloubinoff, P., 1998. The small heat-shock protein IbpB from *Escherichia coli* stabilizes stress-denatured proteins for subsequent refolding by a multichaperone network. J. Biol. Chem. 273, 11032–11037.

Villaverde, A., Carrio, M.M., 2003. Protein aggregation in recombinant bacteria: biological role of inclusion bodies. Biotechnol. Lett. 25, 1385–1395.

Waldo, G.S., Standish, B.M., Berendzen, J., Terwilliger, T.C., 1999. Rapid protein-folding assay using green fluorescent protein. Nat. Biotechnol. 17, 691–695.

Walker, P.A., Leong, L.E., Ng, P.W., Tan, S.H., Waller, S., Murphy, D., Porter, A.G., 1994. Efficient and rapid affinity purification of proteins using recombinant fusion proteases. Biotechnology (N. Y.) 12, 601–605.

Wang, H., Chong, S., 2003. Visualization of coupled protein folding and binding in bacteria and purification of the heterodimeric complex. Proc. Natl. Acad. Sci. U. S. A. 100, 478–483.

Wilkinson, D.L., Harrison, R.G., 1991. Predicting the solubility of recombinant proteins in *Escherichia coli*. Biotechnology (N. Y.) 9, 443–448.

Wu, X., Jornvall, H., Berndt, K.D., Oppermann, U., 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.

Yarian, C., Marszalek, M., Sochacka, E., Malkiewicz, A., Guenther, R., Miskiewicz, A., Agris PF, 2000. Modified nucleoside dependent Watson-Crick and wobble codon binding by tRNALysUUU species. Biochemistry 39, 13390–13395.

Zheng, L., Baumann, U., Reymond, J.L., 2003. Production of a functional catalytic antibody ScFv-NusA fusion protein in bacterial cytoplasm. J. Biochem. (Tokyo) 133, 577–581.