Recombinant protein expression in *Escherichia coli*

François Baneyx

Escherichia coli is one of the most widely used hosts for the production of heterologous proteins and its genetics are far better characterized than those of any other microorganism. Recent progress in the fundamental understanding of transcription, translation, and protein folding in *E. coli*, together with serendipitous discoveries and the availability of improved genetic tools are making this bacterium more valuable than ever for the expression of complex eukaryotic proteins.

Addresses
Department of Chemical Engineering, University of Washington, Box 351750, Seattle, WA 98195, USA;
e-mail: baneyx@cheme.washington.edu

Current Opinion in Biotechnology 1999, **10**:411–421

0958-1669/99/$ – see front matter

Abbreviations

bp	basepair
Csp	cold-shock protein
DB	downstream box
Hsp	heat-shock protein
IPTG	isopropyl-β-D-1-thiogalactopyranoside
MBP	maltose-binding protein
nt	nucleotide
PNPase	polynucleotide phosphorylase
PPIase	peptidyl prolyl *cis/trans* isomerase
SD	Shine-Dalgarno
SRP	signal recognition particle
TF	trigger factor
UP	upstream
UTR	untranslated region

Introduction

Among the many systems available for heterologous protein production, the Gram-negative bacterium *Escherichia coli* remains one of the most attractive because of its ability to grow rapidly and at high density on inexpensive substrates, its well-characterized genetics and the availability of an increasingly large number of cloning vectors and mutant host strains. Although there is no guarantee that a recombinant gene product will accumulate in *E. coli* at high levels in a full-length and biologically active form, a considerable amount of effort has been directed at improving the performance and versatility of this workhorse microorganism. This review will examine the salient features of *E. coli*-based expression systems with an emphasis on how a number of limitations have recently been addressed. Many additional details and references can be found in Makrides' excellent review [1].

Plasmid copy number and maintenance

To achieve high gene dosage, heterologous cDNAs are typically cloned into plasmids that replicate in a relaxed fashion and are present at 15–60 (e.g. pMB1/ColE1-derived plasmids) or a few hundred copies per cell (e.g. the pUC series of pMB1 derivatives). When co-overexpression of additional gene products is desired, ColE1 derivatives are usually combined with compatible plasmids bearing a p15A replicon and maintained at about 10–12 copies per cell. Under laboratory conditions, such multicopy plasmids are randomly distributed during cell division and, in the absence of selective pressure, are lost at low frequency (10^{-5}–10^{-6} per generation) primarily as a result of multimerization [2•]. Nevertheless, plasmid loss can increase tremendously in the case of very high copy number plasmids, when plasmid-borne genes are toxic to the host or otherwise significantly reduce its growth rate, or when cells are cultivated at high density or in continuous processes.

The simplest way to address this problem is to take advantage of plasmid-encoded antibiotic-resistance markers and supplement the growth medium with antibiotics to kill plasmid-free cells. The drawbacks of this approach are loss of selective pressure as a result of antibiotics degradation, inactivation, or leakage of periplasmic detoxifying enzymes (e.g. β-lactamase) into the growth medium, and the contamination of the product or biomass by antibiotics, which may be unacceptable from a medical or regulatory standpoint.

A number of alternative strategies have therefore been developed to ensure that plasmid-free cells will not overtake a culture (Table 1). In most cases, cloning vectors are engineered to carry gene(s) or repressors that cause cell death upon plasmid loss. Although all these approaches have proved valuable, they may place restrictions on the growth medium composition in the case of complementation, and introduce a metabolic burden on the cell by requiring transcription, and often translation, of additional plasmid-encoded genes. To circumvent these problems, Williams *et al.* [3•] created a host strain containing a conditionally essential gene under control of the *lac* operator/promoter region and a companion multicopy plasmid bearing the *lac* operator. Titration of the LacI repressor protein by plasmid-encoded *lac* operators resulted in the expression of the chromosomal gene (in this case kanamycin resistance) and the selective growth of plasmid-bearing cells in medium supplemented with the antibiotic. Replacement of the kanamycin resistance cartridge by an essential host gene will improve the value of this system by removing the disadvantages associated with the use of antibiotic resistance.

A radically different solution to the problem of plasmid instability is the direct insertion of heterologous genes within the chromosome of *E. coli*. Although simple delivery vehicles (e.g. bacteriophage λ) are available for this purpose, little emphasis has been placed on this strategy owing to the perceived notion that gene dosage will necessarily be

Table 1

Programmed cell death in *E. coli*: selected approaches to enhance plasmid stability.

Genetic tool	Principle of action
hoh/sok (*parB*) locus of plasmid R1	Hok is a 52 amino acid-long membrane-damaging protein encoded on a very stable but translationally inactive transcript. Sok is a highly unstable antisense RNA that binds to the *hok* mRNA leader region. Rapid decay of the Sok pool in plasmid-free cells leads to the processing of the 3′ end of *hok* to yield an active transcript. Related system: *pndAB* of plasmid R483.
ccdAB locus of plasmid F	CcdB is a proteolytically stable 11 kDa protein that inhibits DNA gyrase. CcdA is a 9 kDa protein that binds to CcdB and blocks its action. Because the half-life of CcdA is much shorter than that of CcdB, plasmid-free segregants are killed upon degradation of the 'antidote'. Related systems: *parDE* of plasmids RP4/RK2, *phd/doc* of plasmid P1, *parD/pem* of plasmids R1/R100.
Complementation	An essential chromosomal gene is deleted or mutated and an intact copy or a suppressor is supplied in *trans* on a plasmid. Plasmid loss leads to cell death under non-permissive growth conditions. Examples of chromosomal alterations include deletions of genes necessary for the synthesis of essential amino acids, and thermosensitive and nonsense mutations in essential chromosomal genes.

low. However, once chromosomal insertion of a single DNA fragment containing a drug-resistance marker and flanked by two short direct repeats has been achieved, the entire fragment can be amplified to 15–40 copies through *recA*-mediated duplications by increasing the antibiotic concentration [4••]. Although such amplified structures have been reported to be unstable in the absence of selective pressure [5], Olson *et al.* [4••] recently reported that tandem repeats of an IGF-I fusion located at the *attλ* site of the *E. coli* chromosome remained stable in high-density fermentations conducted without antibiotic. An elegant but more time-consuming approach for the insertion of multiple DNA fragments at different locations of the chromosome was developed by Peredelchuk and Bennett [6]. This scheme uses elements of the Tn*1545* site-specific recombination module to randomly integrate a target gene and a drug-resistance marker into the chromosome of a host strain provided with transposon integrase, thereby yielding a collection of clones containing single insertions at different locations of the chromosome. Inserts from the resulting population can be accumulated within a single strain by successive cycles of DNA transfer through bacteriophage P1-mediated transduction, selection for drug resistance and removal of the marker using the excision system of phage λ. As a result, instability problems are eliminated although strain performance may be compromised if important chromosomal loci are disrupted.

Promoters

For many years the *E. coli* lactose utilization (*lac*) operon has served as one of the paradigms of prokaryotic regulation. It is therefore not surprising that many of the promoters used to drive the transcription of heterologous genes have been constructed from *lac*-derived regulatory elements. Although the *lac* promoter and its close relative, *lac*UV5 (which is theoretically not subject to cAMP-dependent regulation, but see [7•]), are rather weak and rarely used for the high-level production of recombinant polypeptides, they are extremely valuable tools to achieve graded expression of helper or toxic proteins provided that *lacY* mutant hosts are used and that induction is performed with the non-hydrolyzable lactose analog isopropyl-β-D-1-thiogalactopyranoside (IPTG) (see [8•,9] for a discussion). The synthetic *tac* and *trc* promoters, which consist of the –35 region of the *trp* promoter and the –10 region of the *lac* promoter, only differ by 1 bp in the length of the spacer domain separating the two hexamers. Both promoters are quite strong and routinely allow the accumulation of polypeptides to about 15–30% of the total cell protein. Although it is often argued that the cost of IPTG limits the usefulness of these promoters, this is rarely a problem for high-added-value products. Furthermore, as little as 50–100 μM IPTG is usually sufficient to achieve full induction. The more serious issue of IPTG toxicity can be circumvented by utilizing lactose as an inducer or by making use of thermosensitive variants of the LacI repressor protein that allow thermal induction of recombinant protein synthesis.

The leakiness of *lac*-derived promoters may be a concern for the production of membrane proteins or other gene products that are toxic to the cell. For medium copy number plasmids (e.g. pBR322), repression can be efficiently achieved by using host strains carrying the *lacI*^{Q} allele. This single nucleotide mutation in the –35 hexamer of the chromosomal *lacI* promoter leads to an increase in the number of LacI repressor molecules from 10–20 to over 100 per cell. For higher copy number plasmids (e.g. pUC derivatives or pMB1 derivatives containing a *rom/rop* mutation), the *lacI* or *lacI*^{Q} genes are typically cloned onto the expression plasmid or provided in *trans* on a compatible plasmid. It was recently shown, however, that a 15 bp deletion in the *lacI* promoter that fortuitously replaces the native –35 hexamer by the consensus sequence for σ^{70}-dependent promoters increases the strength of the *lacI* promoter 170-fold [10•]. Strains bearing the resulting *lacI*^{Q1} allele efficiently repress *lacI*-regulated genes on high copy number plasmids and full activation of plasmid-borne *tac* promoters can be achieved with as little at 3–10 μM IPTG [10•].

In recent years, the pET vectors (commercialized by Novagen, Madison, WI) have gained increasing popularity. In this system, target genes are positioned downstream of the bacteriophage T7 late promoter on medium copy number plasmids. The highly processive T7 RNA polymerase

is supplied in *trans*. Typically, production hosts contain a prophage (λDE3) encoding the enzyme under control of the IPTG-inducible *lac*UV5 promoter. While this system leads to the synthesis of large amounts of mRNA, and, in most cases, the concomitant accumulation of the desired protein at very high concentrations (40–50% of the total cell protein), it is not without drawbacks. For example, high level of mRNA can cause ribosome destruction and cell death, and leaky expression of T7 RNA polymerase may result in plasmid or expression instability. Furthermore, even 'empty' pET plasmids are toxic to *E. coli* in the presence of IPTG [11]. Some of the strategies that have been developed to address these issues are co-overexpression of phage T7 lysozyme (which degrades T7 RNA polymerase) from the compatible pLysS and pLysE plasmids (Novagen) and the insertion of a *lac* operator sequence downstream of plasmid-encoded T7 promoters, in order to reduce leaky transcription. In addition, empirical selection has yielded strains that are superior to the traditional BL21(DE3) host by overcoming toxic effects associated with the overexpression of membrane and globular proteins under T7 transcriptional control [11]. Finally, because it has been reported that the *lac*UV5 promoter becomes activated in stationary phase cultures in a process requiring cAMP, cAMP-deficient (*cya*) mutants of BL21(DE3) should be used for clone selection and fermentation to avoid counter-selection of plasmids carrying toxic genes under T7-control [7•].

An additional limitation of the T7 and other strong promoter systems is that the target protein is often unable to reach a native conformation and either partially or completely segregates within inclusion bodies. Although this problem may be addressed by co-overexpressing folding modulators or through fusion protein technology (see below), an alternative approach is to use promoters that are activated by temperature downshift, as proper protein folding is often favored under low temperature cultivation conditions (see [12•] and references therein). The best characterized cold-shock promoter is that of the major *E. coli* cold-shock protein CspA [13•]. Although the *cspA* core promoter is only weakly induced upon temperature downshift, a 159 nucleotide (nt) long untranslated region (UTR) at the 5′ end of *cspA*-driven transcripts makes them highly unstable at 37°C but significantly increases their stability at low temperatures while simultaneously favoring their preferential engagement by a cold-modified translational machinery containing fewer polysomes and a larger number of monosomes, 30S and 50S ribosomal particles. The *cspA* promoter is rather well repressed at and above 37°C, compares favorably to the *tac* promoter for the expression of an aggregation-prone fusion protein at reduced temperatures and remains functional at 10°C [14]. The major disadvantage of the *cspA* system is that it becomes repressed 1–2 hours after temperature downshift, a time period that is too short to allow high-level accumulation of recombinant proteins. However, the use of a host strain carrying a null mutation in *rbfA*, a gene encoding a 15 kDa protein associating with free 30S ribosomal subunits, allowed continuous *cspA*-driven production of β-galactosidase in high density cell fermentations for several hours following transfer to low temperatures [15••]. The recent demonstration that *cspA*-driven transcription is beneficial for the expression of toxic and proteolytically sensitive gene products, together with the availability of cloning vectors designed for rapidly positioning cDNAs under *cspA* transcriptional control (M Mujacic, K Cooper, F Baneyx, unpublished data) should stimulate interest in this system. Interestingly, the strong bacteriophage λP_L promoter, which is typically used to drive the synthesis of recombinant proteins by transferring strains containing a thermosensitive version of the λcI repressor protein (cI857) from 30 to 42°C, is also cold-inducible [16]. In this case, the main drawback is a high basal level of expression as low-temperature induction must be performed in strains lacking λcI.

Among the various nutritionally inducible promoters (e.g. *phoA* and *trp*, which are induced by phosphate and tryptophan limitation, respectively), the arabinose promoter (*ara*BAD or P_{BAD}) has recently become commercialized by Invitrogen Corp (Carlsbad, CA). This system uses the inexpensive sugar L-arabinose as an inducer and is somewhat weaker than the *tac* promoter. Although it is commonly believed that *ara*BAD can be used to achieve graded levels of protein expression by varying the arabinose concentration, there is extensive heterogeneity in cell populations treated with subsaturating concentrations of the inducer, with some bacteria fully induced and others not at all [9]. Thus, *ara*BAD will not be useful for precisely controlling the levels of protein accumulation until a host strain that efficiently uptakes arabinose by constitutively synthesizing the arabinose transporter(s), or a gratuitous inducer that does not employ them is identified [8•,9]. Additional promoters regulated by a variety of signals (pH, dissolved oxygen concentration, osmolarity, etc.) are available and have been reviewed in detail elsewhere [17].

Upstream elements

The DNA regions that flank core promoters play an important role in determining transcription efficiency. Upstream (UP) elements located 5′ of the –35 hexamer in certain bacterial promoters are A+T rich sequences that increase transcription by interacting with the α subunit of RNA polymerase [18•]. Because few UP elements have been isolated, Gourse and co-workers [19••] used *in vitro* selection to identify upstream sequences conferring increased activity to the *rrnB* P1 core promoter. The best UP sequence was portable and increased *in vivo* transcription from the *rrnB* P1 and *lac* core promoters 326- and 108-fold, respectively.

The degree of homology with the deduced consensus sequence (–59 NNAAA[A/T][A/T]T[A/T]TTTTNNAANNN –38; where N is any nucleotide) was also shown to correlate with the strength of natural UP elements fused upstream of the *lac* core promoter [20•]. These results suggest that the

positioning of highly active UP sequences upstream of well repressed promoters may increase their strength to a level only achieved thus far with phage promoters, but without the drawbacks associated with phage polymerase expression (e.g. leakiness, toxicity and counter-selection).

mRNA stability

E. coli mRNAs are rather unstable, with half-lives ranging between 30 s and 20 min. The major enzymes involved in mRNA degradation are two 3′→5′ exonucleases (RNase II and polynucleotide phosphorylase [PNPase]) and the endonuclease RNase E [21•,22•]. The catalytic activity of RNase E is located at the protein amino terminus, whereas the carboxy-terminus serves as a scaffold for the assembly of a highly efficient RNA 'degradasome' involving PNPase, the DEAD-box RNA helicase RhlB and the glycolytic enzyme enolase. There is considerable controversy over whether RNase E-dependent mRNA decay proceeds in the 5′→3′ or in the opposite direction. In either case, stable secondary structures present in the 5′ UTR of certain transcripts as well as in 3′ rho-independent terminators can both increase mRNA stability; however, their efficiency is modulated by fine features. For example, addition of poly(A) tails to the 3′ end of mRNAs by the seemingly redundant poly(A) polymerases PAP I (the *pcnB* gene product) and PAP II [23] provides a single stranded 'toehold' for RNase II and PNPase that facilitates transcript degradation. In general, polyadenylation is not a problem for recombinant protein expression as only a small fraction of mRNAs contain poly(A) tails in wild-type *E. coli* strains.

The stabilizing effect conferred by untranslated 5′ hairpins was first demonstrated in the case of the long-lived *ompA* mRNA. Fusion of the *ompA* 5′ UTR to a variety of heterologous mRNAs significantly increased transcript half-life, presumably by interfering with RNase E binding [24]. This protective effect is abrogated, however, when the hairpin is preceded by 5′ unpaired nucleotides [25•]. Because RNase E is much more efficient at cleaving substrates with 5′ monophosphate ends than 5′ triphosphate ends [26••], the stabilizing function of 5′ hairpins may be related to their ability to sequester the end of transcripts. The 5′ UTR of the *ompA* mRNA appears particularly well suited for this task as among 10 synthetic hairpins, only one was slightly more effective than the *ompA* UTR in stabilizing *lacZ* transcripts [25•].

Translational issues

Initiation of translation of *E. coli* mRNAs requires a Shine-Dalgarno (SD) sequence complementary to the 3′ end of the 16S rRNA and of consensus 5′-UAAGGAGG-3′, followed by an initiation codon, which is most commonly AUG. About 8% of start sites use GUG, whereas UUG and AUU are rare initiators that are only present in autogenously regulated genes (e.g. those encoding ribosomal protein S20 and initiation factor 3). Although the optimal spacing between these two features is 8 nt, translation initiation is only severely affected if this distance is reduced below 4 nt or increased above 14 nt [27]. Because of the close coupling between transcription and translation in prokaryotes, engineering of the translation initiation region is a powerful tool for modulating gene expression in a promoter-independent fashion [27]. This also means that stable mRNA secondary structures encompassing the SD sequence and/or the initiation codon can dramatically reduce gene expression by interfering with ribosome binding. This problem can be circumvented by increasing the homology of SD regions to the consensus, and by raising the number of A residues in the initiation region through site-directed mutagenesis. An additional mRNA feature affecting translation initiation is the downstream box (DB), which is located after the initiation codon and complementary to bases 1469–1483 of the 16S rRNA. DBs have a 5′-AUGAAUCACAAAGUG-3′ consensus sequence and recent evidence suggests that they play a major role as translational enhancers [28••]. Although introduction of a consensus DB at the 5′ end of genes encoding recombinant proteins would change their amino acid sequence, increasing the homology of this region to that of a DB by using synonymous codons may improve translation initiation of certain transcripts.

Differences in codon usage between prokaryotes and eukaryotes can have a significant impact on heterologous protein production. The arginine codons AGA and AGG are rarely found in *E. coli* genes, whereas they are common in *Saccharomyces cerevisiae* and eukaryotes. The presence of such codons in cloned genes affects protein accumulation levels, mRNA and plasmid stability and, in extreme cases, inhibits protein synthesis and cell growth [29]. An important, but much less obvious effect of AGA codons, is primary structure changes due to the misincorporation of lysine for arginine, particularly when cells are grown in minimal medium [30•]. Fortunately, these problems can usually be addressed by using site-directed mutagenesis to replace rare arginine codons by the *E. coli*-preferred CGC codon or by co-overexpressing the *argU*(*dnaY*) gene which encodes $tRNA^{Arg(AGG/AGA)}$.

Folding in the cytoplasm

Overproduction of heterologous proteins in the cytoplasm of *E. coli* is often accompanied by their misfolding and segregation into insoluble aggregates known as inclusion bodies. Although inclusion body formation can greatly simplify protein purification, there is no guarantee that the *in vitro* refolding will yield large amounts of biologically active product (unsuccessful refolding attempts are seldom reported in the literature). A traditional approach to reduce protein aggregation is through fermentation engineering, most commonly by reducing the cultivation temperature (see [12•] and references therein). The more recent realization that *in vivo* protein folding is assisted by molecular chaperones, which promote the proper isomerization and cellular targeting of other polypeptides by transiently interacting with folding intermediates, and by foldases, which accelerate rate-limiting steps along the folding

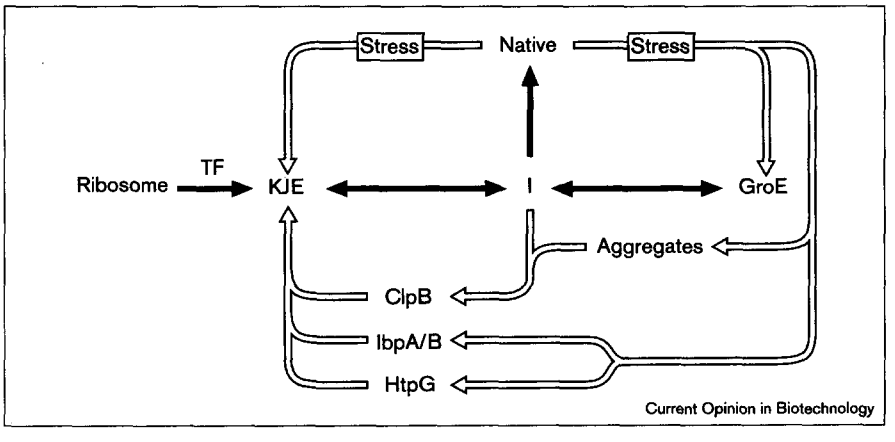
pathway, has provided powerful new tools to combat the problem of inclusion body formation [12•,31].

The best characterized molecular chaperones in the cytoplasm of *E. coli* are the ATP-dependent DnaK-DnaJ-GrpE and GroEL-GroES systems [32•,33•]. DnaK binds to hydrophobic regions exposed to the solvent by nascent or stress-unfolded polypeptides, thereby preventing off-pathway reactions leading to aggregation. The promiscuity of DnaK binding is well explained by the fact that it recognizes heptameric stretches of amino acids consisting of a 4–5 residues-long hydrophobic core flanked by basic residues. This motif occurs every 36 residues on the average protein [34]. DnaJ, which independently binds folding intermediates, activates DnaK for tight substrate binding and might direct it to high-affinity sites. The nucleotide exchange factor GrpE mediates complex resolution: released proteins may either fold into a proper conformation, be recaptured by DnaK-DnaJ for additional cycles of interaction or be reversibly transferred to the 'downstream' GroEL-GroES chaperonins. GroEL is an ~800 kDa hollow toroid consisting of two stacked homoheptameric rings. It binds both substrate proteins and GroES (a 70 kDa dome-shaped homoheptamer) via a ring of hydrophobic residues located in its apical domain. Although no clear consensus sequence has been identified, GroEL, like DnaK, appears to favor hydrophobic and basic residues in its substrates [35]. Upon GroES binding, partially structured folding intermediates are released into the inner cavity of GroEL where they can fold in a capped and hydrophilic environment. There is extensive evidence that co-overproduction of the DnaK-DnaJ or GroEL-GroES chaperones can greatly increase the soluble yields of aggregation-prone proteins (see [31] and references therein) and a number of plasmids compatible with pMB1-derived cloning vectors are available for this purpose [36•,37,38]. The process does not involve dissolution of preformed recombinant inclusion bodies but is related to improved folding of newly synthesized protein chains [38]. It is important to point out, however, that the beneficial effect associated with an increase in the intracellular concentration of DnaK-DnaJ and GroEL-GroES is highly dependent on the nature of the overproduced protein, and that success is by no means guaranteed (and highly unlikely if the protein is inherently incapable of folding).

Based on *in vitro* studies and homology considerations, a number of additional cytoplasmic proteins have been proposed to function as molecular chaperones. They include ClpB, HtpG and IbpA/B, which, like DnaK-DnaJ-GrpE and GroEL-GroES, are heat-shock proteins (Hsps) belonging to the σ^{32} stress regulon. Although inactivation of these Hsps has a modest effect on the ability of *E. coli* to handle thermal stress [39•], they appear to have a supporting role in cellular protein folding by acting as minor chaperones that bind folding intermediates or misfolded proteins and transfer them to the DnaK-DnaJ-GrpE team for subsequent reactivation (Figure 1). Although overproduction of IbpA/B, HtpG or ClpB did not suppress the misfolding of an aggregation-prone fusion protein in the *E. coli* cytoplasm (JG Thomas, F Baneyx, unpublished data), increased intracellular levels of these Hsps might improve the solubility of other substrates, particularly if coordinated with DnaK-DnaJ overexpression.

The *trans* conformation of X–Pro bonds is energetically favored in nascent protein chains; however, ~5% of all prolyl peptide bonds are found in a *cis* conformation in native proteins. The *trans* to *cis* isomerization of X–Pro bonds is rate limiting in the folding of many polypeptides and is catalyzed *in vivo* by peptidyl prolyl *cis/trans* isomerases (PPIases). Three cytoplasmic PPIases, SlyD, SlpA and

Figure 1



A model for chaperone-mediated protein folding in the cytoplasm of *E. coli*. The peptidyl-prolyl *cis/trans* isomerase trigger factor (TF) might be the first folding factor to interact with proteins as they emerge from the ribosome. Under non-stressful conditions (filled arrows), nascent or newly synthesized proteins enter the chaperone cycle through interactions with DnaK-DnaJ-GrpE (KJE). Folding intermediates (I) are released into a native conformation following cycles of binding and release by KJE and/or GroEL-GroES (GroE). Upon heat shock or other stressful conditions (e.g. recombinant protein expression), a number of alternate pathways become prevalent (open arrows). Misfolded proteins re-enter the KJE-GroE pathway directly or are bound by minor chaperones. HtpG and IbpA/B capture proteins as they denature. ClpB is likely to have a more general function including dissolution of preformed aggregates and/or disaggregation of improperly associated intermediates. The minor chaperones interact with KJE to allow protein substrates to re-enter the major chaperone protein folding pathways. This model is based on *in vitro* studies from a number of laboratories and *in vivo* data (JG Thomas, F Baneyx, unpublished data).

trigger factor (TF), have been identified to date [40,41]. The most potent is TF, a 48 kDa protein associated with 50S ribosomal subunits that has been postulated to cooperate with chaperones to guarantee proper folding of newly synthesized proteins. Whether TF overproduction will improve the folding of recombinant proteins synthesized in the *E. coli* cytoplasm remains to be determined. It should be noted, however, that this may not be without physiological consequences as TF and SlyD overproduction lead to cell filamentation. Interestingly, co-overproduction of a leader-less version of PpiA (thus confining to the cytoplasm a PPIase that normally resides in the periplasm) has been shown to increase the yields of a cytoplasmic fusion protein [42•].

Structural disulfide bonds do not form in the cytoplasm of wild-type *E. coli* strains, as this environment is reducing and at least five proteins (thioredoxins 1 and 2, and glutaredoxins 1, 2 and 3, the products of the *trxA*, *trxC*, *grxA*, *grxB* and *grxC* genes, respectively) are involved in the reduction of disulfide bridges that transiently arise in cytoplasmic enzymes [43••]. Nevertheless, disulfide-bonded recombinant proteins can accumulate in the cytoplasm of surprisingly healthy *trxB* mutants that lack thioredoxin reductase, a protein responsible for the reduction of oxidized thioredoxins. Oxidation occurs post-translationally and is favored at low temperatures (see [44] and references therein). While mutants lacking both *trxB* and genes involved in the reduction of glutaredoxins (e.g. *gshA* and *gor*) are even more efficient at accumulating oxidized recombinant proteins in dithiothreitol-free medium, they exhibit severe growth deficiencies in the absence of the reducing agent [45]. As the majority of a cysteine-rich eukaryotic protein was found to accumulate in an almost completely oxidized, but inter-molecular disulfide bonded form in *trxB* mutants held on ice [44], the main challenge will be to engineer protein disulfide isomerases capable of reshuffling disulfide bridges into their native pattern in this environment.

Cytoplasmic degradation

Protein folding and proteolytic degradation are intimately linked as catabolism is an efficient way to conserve cellular resources by recycling improperly folded or irremediably damaged proteins into their constituent amino acids. In the cytoplasm of *E. coli*, most — if not all — early degradation steps are carried out by five ATP-dependent Hsps: Lon/La FtsH/HflB, ClpAP, ClpXP, and ClpYQ/HslUV [46]. ClpAP and ClpXP are two-component proteases that share the same degradation subunit (ClpP) but have different ATPase regulatory subunits (ClpA or ClpX). The latter appear to bind substrates in a chaperone-like manner and use ATP hydrolysis to feed them to the proteolytic center of mini-proteasome structures. Along with FtsH (an inner membrane-associated protease the active site of which faces the cytoplasm), ClpAP and ClpXP are responsible for the degradation of proteins modified at their carboxyl termini by addition of the non-polar destabilizing tail AANDENYALAA (using amino acid single letter code) [47•,48•]. The tagging mechanism involves the 10Sa (SsrA) stable RNA and is designed to prevent ribosome stalling at the 3′ end of damaged mRNAs [49]. Proteases Lon and ClpYQ appear to be more generic as they efficiently degrade puromycin-truncated proteins; however, there is some evidence that they also exhibit tail specificity. An obvious consequence of the existence of the SsrA tagging system is that any heterologous proteins rich in non-polar residues at its carboxyl terminus will be an appetizing substrate for cellular proteases.

A possible strategy to avoid degradation is to make use of host strains bearing mutations in protease genes; however, there are drawbacks to this approach. For example, inactivation of Lon leads to filamentation and FtsH is an essential protein for which only thermosensitive mutants are available. In addition, several proteases are usually involved in the degradation of a given protein substrate but multiple mutations in genes encoding proteases reduce cell growth rates and compromise strain fitness. An alternative is to target the polypeptide of interest to the insoluble fraction of the cell, as inclusion-body proteins are generally protected from degradation. For a normally soluble protein, this can be achieved by using strains bearing thermosensitive mutations in the major molecular chaperone systems [50]. It is important, however, to bear in mind that certain proteases (e.g. OmpT) adsorb to the surface of inclusion bodies during the recovery process and may degrade the desired protein while it is being refolded. The inner membrane protease FtsH is also active under denaturing conditions and can process recombinant proteins associated with the inner-membrane during their refolding (KW Cooper, F Baneyx, unpublished data).

Fusion proteins

Although fusion proteins were originally constructed to facilitate protein purification and immobilization and to couple the activity of enzymes acting in a single metabolic pathway, it soon became apparent that certain fusion partners could greatly improve the solubility of passenger proteins that would otherwise accumulate within inclusion bodies in the cell cytoplasm. Systems suitable for the construction of fusions to maltose-binding protein (MBP), thioredoxin and glutathione S-transferase are commercially available and additional 'solubilizing' fusion partners (e.g. variants of DsbA and gpHD) have recently been described [42•,51•]. The most probable reason for improved folding (and/or reduced degradation) of passenger proteins is that the fusion partner efficiently and rapidly reaches a native conformation as it emerges from the ribosome (or soon after its release), and promotes the acquisition of correct structure in downstream folding units by favoring on-pathway isomerization reactions. In the case of unfused cytoplasmic MBP, proper folding requires both DnaK-DnaJ-GrpE and GroEL-GroES, which may recruit chaperones in the vicinity of the passenger protein (JG Thomas, F Baneyx, unpublished data). It has also been proposed that MBP

may directly interact with passenger proteins [52••], thereby acting as an 'intramolecular' chaperone, much like protease propeptides do [53]. These mechanisms require the MBP domain to be synthesized first and are in agreement with a study showing that, whereas mammalian asparatic proteinases are soluble when fused to the carboxyl-terminus of MBP, they become insoluble when the order of the fusion proteins is reversed [54]. It should finally be noted that, despite outlandish claims, all fusion partners are not equally proficient at alleviating inclusion body formation. In a systematic comparison of the effectiveness of various fusion partners in increasing the solubility of six aggregation-prone passenger polypeptides, Kapust and Waugh [52••] found that MBP was far superior to either thioredoxin or glutathione-*S*-transferase as a 'solubilizing' partner.

The affinity of certain fusion partners for immobilized ligands can facilitate the purification of the desired fusion protein; however, binding usually occurs with low affinity (which precludes the use of stringent wash conditions) and can be disrupted by passenger proteins. The use of polyhistidine tags at the amino-terminus or at the junction region of the fusion partner can solve this problem by allowing efficient purification via immobilized metal affinity chromatography [51•,55]. An additional advantage of fusion proteins is that they appear to permit the synthesis of otherwise poorly translated polypeptides. A probable explanation for this result is that the translation of passenger proteins containing rare codons occurs with higher efficiency; however, this may also lead to Lys→Arg misincorporation at rare codons.

Currently, the main disadvantages of fusion-protein technologies are that: firstly, liberation of the passenger proteins requires expensive proteases (e.g. Factor Xa and enterokinase); secondly, cleavage is rarely complete leading to reduction in yields; thirdly, additional steps may be required to obtain an active product (e.g. formation and isomerization of disulfide bonds); and finally, solubility is never guaranteed.

Secretion

Polypeptides destined for export are synthesized as preproteins containing an amino-terminal signal sequence (leader peptide) that is cleaved during the translocation process by inner-membrane-associated leader peptidases, the active sites of which face the periplasm. Typical signal sequences are 18–30 amino acids in length and consist of two or more basic residues at the amino terminus, a central hydrophobic core of seven or more amino acids, and a hydrophilic carboxyl-terminus motif recognized by leader peptidases (usually small residues at positions –1 and –3 [A, G or S] preceeded by a helix-breaking residue at position –6 [P or G]; where +1 denotes the first amino acid of the mature protein). Many signal sequences derived from naturally occuring secretory proteins (e.g. OmpA, OmpT, PelB, β-lactamase and alkaline phosphatase) support the efficient translocation of heterologous polypeptides across the inner membrane when fused to their amino termini. In some cases, however, preproteins are not readily exported and either become 'jammed' in the inner membrane, accumulate in precursor inclusion bodies, or are rapidly degraded within the cytoplasm. While membrane jamming is an indication that translocation may be physically impossible (e.g. in the case of large cytoplasmic proteins, unatural fusion proteins, and mutant proteins evolved by combinatorial approaches), an improved understanding of secretory mechanisms in *E. coli* has provided clues to circumvent other problems.

Efficient translocation requires that secretory proteins be brought into the vicinity of the inner membrane in a loosely folded form. This is guaranteed by molecular chaperones, which can be either generic (e.g. DnaK and GroEL) or specific for secretory proteins. SecB, a tetrameric polypeptide present at low levels in the cytoplasm binds to the mature domain of a subset of preproteins destined for the outer membrane and transfers them to peripheral membrane protein SecA. The latter uses energy derived from ATP hydrolysis and the proton motive force to mediate preprotein export by cycles of insertion and de-insertion into the SecYEG translocon [56•,57•]. The signal recognition particle (SRP), which consists of a 4.5S RNA and a 48 kDa GTPase termed Ffh/P48 binds highly hydrophobic signal sequences in certain preproteins (e.g. integral inner membrane proteins) and delivers them to the peripheral membrane protein FtsY in the vicinity of SecA and SecYEG [58••]. It is therefore probable that the majority of secretory proteins are delivered to the SecA motor via a variety of targeting mechanisms for export through SecYEG; however, some inner-membrane proteins also appear to directly integrate into the lipid bilayer [59].

In view of the above mechanistic information, it is tempting to hypothesize that the misfolding and degradation of a number of heterologous proteins targeted for the periplasm results from their inefficient chaperoning to the translocase, either because they fold (or misfold) too rapidly in the cytoplasm, or because the necessary chaperone(s) become limiting. Attempts to co-overproduce SecB, DnaK-DnaJ and GroEL-GroES have met with variable success and improved secretion depends heavily on the signal-sequence–mature protein combination [60]. This suggests that the signal sequence influences secondary and tertiary structure formation in the mature region of secretory proteins, which in turn affects chaperone recognition. It may therefore be necessary to try several signal sequences and/or overproduce different chaperones to optimize the translocation of a given heterologous protein. At present, there are no reports on how overproduction of components of the SRP (and in particular FfH) affects protein secretion. This route may be particularly valuable for improving the assembly of inner membrane proteins. It should finally be noted that strains selected for their ability to restore the export of preproteins with defective signal sequences are

useful hosts for facilitating export. The most potent mutation (*prlA4* in SecY) was recently shown to function by stabilizing SecA at the SecYEG translocon [61••].

Folding and degradation in the periplasm

The periplasm is an oxidizing environment that contains enzymes catalyzing the formation and rearangement of disulfide bonds [43••,62]. As a result, it is a particularly attractive destination for the production of secreted eukaryotic proteins. Recent studies have shown that co-overproduction of protein disulfide isomerases (and in particular DsbC) can greatly improve proper disulfide bond formation in cysteine-rich recombinant proteins, such as human tissue plasminogen activator [63••].

In addition to inefficient secretion, one of the drawbacks of periplasmic expression is that recombinant proteins may misfold or form inclusion bodies in this cellular compartment. A systematic search for periplasmic factors improving phage display [64••] led to the identification of Skp/OmpH, a protein previously implicated in the folding of outer membrane proteins [65]. In contrast to specialized periplasmic chaperones (e.g. PapCD which is involved in pilus biogenesis), Skp appears to be a broad substrate range chaperone, as its overexpression improves the folding of a number of aggregation-prone single-chain antibody fragments [64••,66•]. The periplasm also contains a number of PPIases, including SurA, FkpA, RotA/PpiA and the recently discovered PpiD [67•]. The primary function of SurA and PpiD is to catalyze prolyl peptide bond isomerization in outer membrane proteins. Both SurA and FkpA might, however, have a more general role as their overproduction facilitates the folding of recombinant proteins that aggregate or are degraded in the periplasm [65].

Because there is no periplasmic ATP pool, misfolded proteins are degraded by energy-independent proteases, the most active of which are DegP/HtrA and Tsp, a protease that recognizes secreted proteins tagged by the SsrA system [49]. A number of additional proteolytic enzymes are present in the periplasm and cell enveloppe and participate in the degradation of secreted proteins [46]. Strains lacking individual or combinations of cell envelope proteases have been constructed and, although their use can help alleviate accute degradation problems, they have the disadvantage of exhibiting slower growth rates. It should finally be noted that periplasmic proteins can be further translocated to the growth medium of the cell by deliberate permeabilization of the outer membrane using a variety of systems (see [68•,69•] and references therein). This process greatly simplifies the purification of target polypeptides although it increases their dilution.

Conclusions

Recent advances in the understanding of the function, regulation and interactions of cellular gene products, together with the availability of new genetic tools, are making *E. coli* a more attractive host than ever for the production of heterologous proteins. The facts that only a small amount of information has been exploited for practical purposes and that many fundamental aspects of *E. coli* physiology remain to be uncovered will continue to fuel progress in optimizing this microorganism for protein expression. Many improvements have resulted from serendipitous discoveries (e.g. the usefulness of fusion proteins and the fact that disulfide bridges can form in the cytoplasm of *trxB* strains) and this trend is likely to continue. Although certain post-translational modifications (e.g. glycosylation) will probably remain beyond the reach of *E. coli*, robust engineered strains suitable for the cost-effective production of a wide variety of complex eukaryotic proteins should become available in the near future.

Acknowledgements

This work was supported by awards from the Bioengineering and Environmental Systems division of the US National Science Foundation. I thank Jeff Thomas for drawing Figure 1.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Makrides SC: **Strategies for achieving high-level expression of genes in *Escherichia coli*.** *Microbiol Rev* 1996, **60**:512-538.

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

An excellent review of the mechanisms through which multicopy plasmids (primarily ColE1) are stably maintained in *E. coli*.

3. • Williams SG, Cranenburgh RM, Weiss AME, Wrighton CJ, Sherratt DJ, Hanak JAJ: **Repressor titration: a novel system for selection and stable maintenance of recombinant plasmids.** *Nucleic Acids Res* 1998, **26**:2120-2124.

This paper describes a new approach for stable maintenance of multicopy plasmids. A plasmid-borne *lac* operator titrates chromosomal LacI, leading to the constitutive expression of a chromosomally encoded kanamycin gene preceded by a *lacO/P* control region. Only plasmid-containing cells survive in medium supplemented with kanamycin. The authors highlight the fact that the system may be particularly useful for producing plasmid vectors for gene therapy.

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

An entertaining article documenting approaches that can be used to circumvent broad patent claims protecting *E. coli* expression systems. A strain containing T7 RNA polymerase under *lac* transcriptional control is constructed by first introducing a defective version of T7 gene 1 within the chromosomal *lac* operon and rescuing it through homologous recombination with the missing fragment from the same gene. A strategy for multiple integration of heterologous genes in the chromosome which does not require that the gene be plasmid-borne, or that gene and promoter be 'operably linked' is also described. This paper should significantly reduce claim language plagiarism and force patent lawyers to become more creative.

5. Chopin M-C, Chopin A, Ronault A, Galleron M: **Insertion and amplification of foreign genes in the *Lactococcus lactis* subsp. *lactis* chromosome.** *Appl Environ Microbiol* 1989, **55**:1769-1774.

6. Peredelchuk MY, Bennett GN: **A method for construction of *E. coli* strains with multiple DNA insertions in the chromosome.** *Gene* 1997, **187**:231-238.

7. • Grossman TH, Kawasaki ES, Punreddy SR, Osburne MS: **Spontaneous cAMP-dependent derepression of gene expression in stationary phase plays a role in recombinant expression instability.** *Gene* 1998, **209**:95-103.

An interesting study showing that the *lac* and *lacUV5* promoters become activated in a cAMP-dependent fashion as uninduced cells approach stationary phase. Leaky induction is significantly reduced in cAMP-deficient (*cya*) mutants. It is pointed out that although early stationary phase cells are rarely used for protein production, this stage is reached on solid medium

during clone selection and by a fraction of the population in liquid cultures. The use of *cya* mutants that exhibit enhanced expression and plasmid stability is recommended for the production of toxic proteins placed under transcriptional control of *lac*-derived promoters or T7 promoters relying on a *lacUV5*-transcribed T7 RNA polymerase.

8. Hashemzadeh-Bonehi L, Mehraein-Ghomi F, Mitsopoulos C, Jacob JP, • Hennessy ES, Broome-Smith JK: **Importance of using *lac* rather than *ara* promoter vectors for modulating the levels of toxic gene products in *Escherichia coli*.** *Mol Microbiol* 1998, **30**:676-678.
This article re-emphasizes the importance of not using *ara*-driven transcription for exerting fine control over protein synthesis levels. *lacUV5*-based cloning vectors suitable for the (low yield) production of toxic bacterial and eukaryotic membrane proteins in uninduced cells are described.

9. Siegele DA, Hu JC: **Gene expression from plasmids containing the *araBAD* promoter at subsaturating inducer concentrations represents mixed populations.** *Proc Natl Acad Sci USA* 1997, **94**:8168-8172.

10. Glascock CB, Weickert MJ: **Using chromosomal *lacI*^{Q1} to control • expression of genes on high-copy number plasmids in *Escherichia coli*.** *Gene* 1998, **223**:221-231.
The *lacI*^{Q1} allele arises from a mutation that replaces the −35 region of the *lacI* promoter by the consensus hexamer for Eσ^{70}-dependent promoters. P*lacI*^{Q1} is found to be 170-fold stronger than P*lacI*. Strains carrying *lacI*^{Q1} on their chromosome are as efficient as plasmid-encoded *lacI* at repressing *tac*-driven transcription of genes carried by high copy number plasmids and high-magnitude induction is achieved with very low concentrations of IPTG. The authors speculate that certain strains reported to be *lacI*^{Q} are actually *lacI*^{Q1} and describe methods to test this possibility.

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

12. Baneyx F: ***In vivo* folding of recombinant proteins in *Escherichia* • *coli*.** In *Manual of Industrial Microbiology and Biotechnology*, edn 2. Edited by Davies JE, Demain AL, Cohen G, Hershberger CL, Forney LJ, Holland IB, Hu W-S, Wu J-HD, Sherman DH, Wilson RC. Washington, DC: American Society for Microbiology; 1999:551-565.
A review of protein folding in *E. coli* with an emphasis on folding modulators and the usefulness of pragmatic approaches to prevent heterologous protein aggregation.

13. Phadtare S, Alsina J, Inouye M: **Cold-shock response and cold- • shock proteins.** *Curr Opin Microbiol* 1999, **2**:175-180.
A review of cold-shock responses and cold-shock proteins with a focus on CspA and its homologs.

14. Vasina JA, Baneyx F: **Expression of aggregation-prone recombinant proteins at low temperatures: a comparative study of the *Escherichia coli cspA* and *tac* promoter systems.** *Protein Expr Purif* 1997, **9**:211-218.

15. Vasina JA, Peterson MS, Baneyx F: **Scale-up and optimization of the •• low-temperature inducible *cspA* promoter system.** *Biotechnol Prog* 1998, **14**:714-721.
This paper examines the usefulness of fermentation and strain engineering approaches in circumventing the problem of *cspA* promoter repression following prolonged incubation at low temperatures. Temperature cycling and successive temperature downshifts in fermentors are found to only marginally improve yields. Host strains bearing a *rbfA* null mutation are shown to allow constitutive protein expression for up to seven hours after temperature downshift and their performance is validated in high density fermentations.

16. Giladi H, Goldenberg G, Koby S, Oppenheim AB: **Enhanced activity of the bacteriophage lambda P_L promoter at low temperatures.** *Proc Natl Acad Sci USA* 1995, **92**:2184-2188.

17. Goldstein MA, Doi RH: **Prokaryotic promoters in biotechnology.** *Biotechnol Annu Rev* 1995, **1**:105-128.

18. Aiyar SE, Gourse RL, Ross W: **Upstream A-tracts increase • bacterial promoter activity through interactions with the RNA polymerase α subunit.** *Proc Natl Acad Sci USA* 1998, **95**:14652-14657.
Phased A-tracts of sequence $([A]_6TAGGC[A]_5TAGGC)_n$ are shown to enhance transcription from the *lac* and *rrnB* P1 core promoters 15–20-fold when their 3′ end is located at position −40. The beneficial effect is less, however, than that conferred by the native *rrnB* P1 UP element and decreases when A-tracts are positioned one helical turn further upstream from the −35 hexamer. This paper also demonstrates that the α subunit of RNA polymerase is required for transcription enhancement both *in vivo* and *in vitro*.

19. Estrem ST, Gaal T, Ross W, Gourse RL: **Identification of an UP •• element consensus sequence for bacterial promoters.** *Proc Natl Acad Sci USA* 1998, **95**:9761-9766.
A modified SELEX procedure is used to select nucleotide sequences that enhance transcription when positioned upstream of the *rrnB* P1 core promoter. A group of 31 such sequences increasing transcription 136–326-fold (compared to 69-fold for the native *rrnB* P1 UP element) is identified and used to deduce a consensus sequence for UP elements. The best sequence increases transcription 108-fold when placed upstream of the *lac* core promoter (compared to 39-fold for the native *rrnB* P1 UP element) suggesting that engineering of UP regions will be a powerful tool to improve transcription of heterologous genes.

20. Ross W, Aiyar SE, Salomon J, Gourse RL: ***Escherichia coli* • promoters with UP elements of different strengths: modular structure of bacterial promoters.** *J Bacteriol* 1998, **180**:5375-5383.
This paper analyzes the effect of UP elements from *rrnD* P1, *rrnB* P2, λ P_R, *lac*, *merT* and RNA II on the activity of native and *lac* core promoters. Transcription enhancement in hybrid *lac* promoters spans a wide range of values (between 1.5 and 33-fold) and correlates well with the degree of homology to the UP consensus sequence. The authors conclude that *E. coli* promoters are modular composites of three RNA polymerase recognition elements – the UP sequence and the −35 and −10 hexamers – whose proximity to consensus determines promoter strength.

21. Carpousis AJ, Vanzo NF, Raynal LC: **mRNA degradation: a tail of • poly(A) and multiprotein machines.** *Trends Genet* 1999, **15**:24-28.
A review of recent advances in the understanding of mRNA degradation that highlights differences and similarities between eubacteria and eukaryotes.

22. Coburn GA, Mackie GA: **Degradation of mRNA in *Escherichia coli*: • an old problem with some new twists.** *Prog Nucleic Acid Res Mol Biol* 1999, **62**:55-108.
An in-depth review of the mRNA degradation process in *E. coli*.

23. Cao GJ, Pogliano J, Sarkar N: **Identification of the coding region for a second poly(A) polymerase in *Escherichia coli*.** *Proc Natl Acad Sci USA* 1996, **93**:11580-11585.

24. Hansen M, Chen L, Fejzo M, Belasco J: **The *ompA* 5′ untranslated region impedes a major pathway for mRNA degradation in *E. coli*.** *Mol Microbiol* 1994, **12**:707-716.

25. Carrier TA, Keasling JD: **Library of synthetic 5′ secondary structures • to manipulate mRNA stability in *Escherichia coli*.** *Biotechnol Prog* 1999, **15**:58-64.
This article examines how 10 rationally designed UTRs forming 5′ hairpins of different strengths influence the stability of *lacZ* transcripts to which they are fused. The mRNA half-life is influenced over an order of magnitude with wide data spread. Although there is a good correlation between increased hairpin strength and message stability for certain UTRs, others do not follow this trend. The library of 5′ UTR constructed in this study should be a useful tool to modulate gene expression in a promoter-independent fashion.

26. Mackie GA: **Ribonuclease E is a 5′-end dependent endonucleases.** •• *Nature* 1998, **395**:720-723.
A very elegant set of experiments that establishes several key properties of RNase E: it favors linear substrates; requires unpaired 5′ nucleotides for degradation; and cleaves RNAs with 5′ monophosphates much more efficiently than those with 5′ triphosphates. These results explain how 5′ hairpins may provide their stabilizing effect.

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

28. Etchegaray J-P, Inouye M: **Translational enhancement by an •• element downstream of the initiation codon in *Escherichia coli*.** *J Biol Chem* 1999, **274**:10079-10085.
Using the DB of CspA as a starting point, the authors created a downstream element that perfectly matches the consensus sequence and demonstrated translational enhancement both at low temperatures and at 37°C. This study further shows that a Shine-Dalgarno sequence is required for proper DB function, and that the location of this element is flexible, as it functions equally well when positioned at either codon 1 or 6.

29. Zahn K: **Overexpression of an mRNA dependent on rare codons inhibits protein synthesis and cell growth.** *J Bacteriol* 1996, **178**:2926-2933.

30. Forman MD, Stack RF, Masters PS, Hauer CR, Baxter SM: **High • level, context-dependent misincorporation of lysine for arginine in *Saccharomyces cerevisiae* a1 homeodomain expressed in *Escherichia coli*.** *Protein Sci* 1998, **7**:500-503.
In cultures grown in minimal medium and in the presence or absence of glucose, 20–25% of the yeast a1 homeodomain is found to contain misincorporated lysine for arginine at the 115 AGA codon. Misincorporation is reduced to ~5% when cells are grown in LB broth. Mutagenesis of codon 115 for CGC reduces misincorporation to 5% in minimal medium, while overexpression of *argU* yields a more than 99% pure product under the same growth conditions.

31. Thomas JG, Ayling A, Baneyx F: **Molecular chaperones, folding catalysts and the recovery of active recombinant proteins from *E. coli*: to fold or to refold.** *Appl Biochem Biotechnol* 1997, **66**:197-238.

32. • Richardson A, Landry SJ, Georgopoulos C: **The ins and outs of a molecular chaperone machine.** *Trends Biochem Sci* 1998, **23**:138-143.
A review of the mechanism of action of the GroEL-GroES system with an emphasis on structural aspects.

33. • Bukau B, Horwich AL: **The Hsp70 and Hsp60 chaperone machines.** *Cell* 1998, **92**:351-366.
An excellent review of the mechanism of action of the major cytoplasmic chaperone systems.

34. Rüdiger S, Germeroth L, Schneider-Mergener J, Bukau B: **Substrate specificity of the DnaK chaperone determined by screening cellulose-bound peptide libraries.** *EMBO J* 1997, **16**:1501-1507.

35. Coyle JE, Jaeger J, Gross M, Robinson CV, Radford SE: **Structural and mechanistic consequences of polypetide binding by GroEL.** *Fold Des* 1997, **2**:93-104.

36. • Nishihara K, Kanemori M, Kitagawa M, Yanaga H, Yura T: **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* 1998, **64**:1694-1699.
Folding of Cryj2 is shown to depend on the synergistic action of the DnaK-DnaJ-GrpE and GroEL-GroES systems. A ColE1-compatible plasmid encoding the *dnaKJ* and *groESL* operons under the control of differentially inducible promoters is constructed and used to characterize the effect of chaperone coexpression on the folding, degradation and aggregation of Cryj2.

37. Castanié MP, Bergès H, Oreglia J, Prère MF, Fayet O: **A set of pBR322-compatible plasmids allowing the testing of chaperone-assisted folding of proteins overexpressed in *Escherichia coli*.** *Anal Biochem* 1997, **254**:150-152.

38. Thomas JG, Baneyx F: **Protein misfolding and inclusion body formation in recombinant *Escherichia coli* cells overproducing heat-shock proteins.** *J Biol Chem* 1996, **271**:11141-11147.

39. • Thomas JG, Baneyx F: **Roles of the *Escherichia coli* small heat shock proteins IbpA and IbpB in thermal stress management: comparison with ClpA, ClpB, and HtpG *in vivo*.** *J Bacteriol* 1998, **180**:5165-5172.
This study uses a set of isogenic strains to examine how the inactivation of several putative molecular chaperones (IbpA/B, ClpB, HtpG and the non-Hsp ClpA) affects the ability of *E. coli* to grow at high temperatures and recover from thermal stress. Deletion of the minor chaperones causes growth defects (but no obvious increase in host protein misfolding) at 46°C. Double mutant studies indicate that IbpA/B, ClpB and HtpG play a supporting role in protein folding by cooperating with the DnaK-DnaJ-GrpE and GroEL-GroES systems.

40. Hottenrott S, Schumann T, Plückthun A, Fischer G, Rahfeld J-U: **The *Escherichia coli* SlyD is a metal ion-regulated peptidyl-prolyl *cis/trans* isomerase.** *J Biol Chem* 1997, **272**:15697-15701.

41. Stoller G, Rücknagel KP, Nierhaus KH, Schmid FX, Fischer G, Rahfeld J-U: **A ribosome-associated peptidyl-prolyl *cis/trans* isomerase associated as the trigger factor.** *EMBO J* 1995, **14**:4939-4948.

42. • Zhang Y, Olsen DR, Nguyen KB, Olson PS, Rhodes ET, Mascarenhas D: **Expression of eukaryotic proteins in soluble form in *Escherichia coli*.** *Protein Expr Purif* 1998, **12**:159-165.
E. coli DsbA, a normally periplasmic thiol/disulfide oxidoreductase is modified by removal of the signal sequence and conversion of the two active-site cysteine residues into serines. The resulting protein ($DsbA^{mut}$) is shown to act as a powerful solubilizing partner when eukaryotic proteins are fused to its carboxyl terminus. The same study demonstrates that the accumulation levels of a fusion protein between a signal-sequence-less DsbA and the rhinoviral 3C proteinase are enhanced when the expression plasmid encodes a leaderless version of PpiA (a PPIase that normally resides in the periplasm). It is instructive to note that the solubility of a $DsbA^{mut}$–GFP fusion is increased about twofold relative to a Trx–GFP fusion, whereas a MBP–GFP fusion is about ninefold more soluble than the same control (see [52••]). Although this suggests that MBP is over fourfold more efficient than $DsbA^{mut}$ as a solubilization partner, differences in junction domains, growth conditions and plasmid design may partially contribute to the difference.

43. •• Åslund F, Beckwith J: **The thioredoxin superfamily: redundancy, specificity and gray area genomics.** *J Bacteriol* 1999, **181**:1375-1379.
This short review of the thioredoxin superfamily of proteins (which includes thioredoxins, glutaredoxins, their reductases and the Dsb cell-envelope proteins) raises the important point that functional genomics will not be sufficient to unambiguously determine why specific proteins evolved and under what conditions apparently redundant proteins become absolutely necessary.

44. Schneider EL, Thomas JG, Bassuk JA, Sage EH, Baneyx F: **Manipulating the aggregation and oxidation of human SPARC in the cytoplasm of *Escherichia coli*.** *Nat Biotechnol* 1997, **15**:581-585.

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

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

47. • Gottesman S, Roche E, Zhou YN, Sauer RT: **The ClpXP and ClpAP proteases degrade proteins with carboxyl-terminal peptide tails added by the SsrA-tagging system.** *Genes Dev* 1998, **12**:1338-1347.
This study implicates ClpXP and ClpAP as major players in the degradation of proteins that have been tagged at their carboxyl termini by the SsrA system at 32°C. The authors point out that in addition to this general housekeeping function, the Clp proteases recognize additional determinants in some of their other targets (at the amino terminus in the case of ClpAP and at the carboxyl terminus in the case of ClpXP).

48. • Herman C, Thévenet D, Bouloc P, Walker GC, D'Ari R: **Degradation of carboxy-terminal-tagged cytoplasmic proteins by the *Escherichia coli* protease HflB (FtsH).** *Genes Dev* 1998, **12**:1348-1355.
This paper shows that, in addition to the Clp proteases, the inner membrane protein HflB (FtsH) participates in the degradation of SsrA-tagged proteins. HflB-mediated degradation is postulated to dominate at high temperatures and in the case of abnormal membrane proteins.

49. Keiler KC, Waller PRH, Sauer RT: **Role of a peptide tagging system in the degradation of proteins synthesized from damaged messenger RNA.** *Science* 1996, **271**:990-993.

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

51. • Forrer P, Jaussi R: **High-level expression of soluble heterologous proteins in the cytoplasm of *Escherichia coli* by fusion to the bacteriophage lambda head protein D.** *Gene* 1998, **224**:45-52.
The bacteriophage λ head protein modified by addition of a His-tag at its amino terminus (gpHD) is shown to promote the folding of some – but not all – passenger proteins fused to its carboxyl terminus. It is shown that, unlike GST fusions, certain gpHD fusions are thermostable and remain soluble upon incubation at 80°C. The authors emphasize the fact that, because gpD is an established partner for phage display, the system will be useful to couple the selection of heterologous proteins from libraries and their high-level expression.

52. •• Kapust RB, Waugh DS: ***Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused.** *Protein Sci* 1999, **8**:1668-1674.
A careful comparison of the ability of three fusion partners (MBP, GST and Trx) to inhibit the misfolding of six aggregation-prone polypeptides. MBP is found to be dramatically superior to either Trx or GST as a 'solubilizing' partner, and to exert its beneficial effect during, rather than after folding. The study also reveals that Trx fusions are marginally, but consistently more soluble than their GST counterparts. Based on the observation that MBP binds maltose within a deep hydrophobic cleft, the authors suggest that MBP may be particularly effective at promoting folding by physically sequestering hydrophobic stretches in passenger protein folding intermediates.

53. Shinde U, Inouye M: **Propeptide-mediated folding in subtilisin: the intramolecular chaperone concept.** *Adv Exp Med Biol* 1996, **379**:147-154.

54. Sachdev D, Chirgwin JM: **Order of fusions between bacterial and mammalian proteins can determine solubility in *Escherichia coli*.** *Biochem Biophys Res Commun* 1998, **244**:933-937.

55. Pryor KD, Leiting B: **High level expression of soluble proteins in *Escherichia coli* using a His6-tag and the maltose-binding protein double-affinity fusion system.** *Protein Expr Purif* 1997, **10**:309-319.

56. • Economou A: **Bacterial preprotein translocase: mechanism and conformational dynamics of a processive enzyme.** *Mol Microbiol* 1998, **27**:511-518.
A review of Sec-dependent secretion with an emphasis on the mechanism of action of SecA.

57. • Driessen AJ, Fekkes P, van der Wolk JP: **The Sec system.** *Curr Opin Microbiol* 1998, **1**:216-222.
A review of the Sec secretion pathway.

58. Valent QA, Scotti PA, de Gier J-WL, von Heijne G, Lentzen G,
•• Wintermeyer W, Oudega B, Luirink J: **The *Escherichia coli* SRP and SecB targeting pathways converge at the translocon.** *EMBO J* 1998, **17**:2504-2512.
Crosslinking experiments show that upon release from the SRP–FtsY complex, model secretory proteins are found exclusively in the vicinity of SecA, SecY and SecG. This important result suggests that the multiple pathways used by the cell to bring preproteins into contact with the inner membrane (e.g. SecB-, SRP-, and generic-molecular-chaperone-dependent pathways) all employ SecA and the SecYEG translocon to mediate protein translocation across the cytoplasmic membrane.

59. de Gier JWL, Scotti PA, Sääf A, Valent QA, Kuhn A, Luirink J, von Heijne G: **Differential use of the signal recognition particle translocase targeting pathway for inner membrane assembly in *Escherichia coli*.** *Proc Natl Acad Sci USA* 1998, **95**:14646-14651.

60. Bergès H, Joseph-Liauzun E, Fayet O: **Combined effects of the signal sequence and the major chaperone proteins on the export of human cytokines in *Escherichia coli*.** *Appl Environ Microbiol* 1996, **62**:55-60.

61. van der Wolk JPW, Fekkes P, Boorsma A, Huie JL, Silhavy TJ, Driessen
•• AJM: ***PrlA4* prevents the rejection of signal sequence defective preproteins by stabilizing the SecA-SecY interaction during the initiation of translocation.** *EMBO J* 1998, **17**:3631-3639.
A long overdue explanation of why *prlA4* mutations (which map in the *secY* gene) are such excellent suppressors of signal sequence mutations. Studies with inverted membrane vesicles (IMVs) show that SecA has a much higher affinity for the mutant form of SecY than for the wild-type version and that SecA–precursor complexes remain bound at the translocon for longer periods of time, which permits the secretion of proteins with defective (or completely absent) signal sequences. It is also observed that translocation of authentic secretory proteins increases in *prlA4* IMVs due to the loss of SecA 'proofreading' function. As a result, *prlA4* mutants should be powerful hosts not only to facilitate the secretion of hard to translocate proteins but also to further improve that of proteins which efficiently reach the periplasm in wild-type cells.

62. Missiakas D, Raina S: **Protein folding in the bacterial periplasm.** *J Bacteriol* 1997, **179**:2465-2471.

63. Qiu J, Swartz JR, Georgiou G: **Expression of active human tissue
•• type plasminogen activator in *Escherichia coli*.** *Appl Environ Microbiol* 1998, **64**:4891-4896.
This paper examines the effect of co-overexpressing various thiol/disulfide oxidoreductases in the *E. coli* periplasm on the recovery of active human tissue plaminogen activator (tPA; a protein that contains 17 disulfide bridges in its native form). The folding helpers tested include *E. coli* DsbA (a soluble periplasmic protein that primarily functions as a strong oxidizer), *E. coli* DsbC (a soluble periplasmic protein that primarily functions as a protein disulfide isomerase [PDI]), and two eukaryotic PDIs from rat and yeast. Co-overexpression of DsbC is found to greatly increase tPA recovery in both shake flasks and fermentors and the purified protein has the same specific activity as native tPA. Although the final yields are low, this study demonstrates that engineering of disulfide-bond-formation pathways holds great promise for the production of complex eukaryotic proteins in *E. coli*.

64. Bothmann H, Plückthun A: **Selection for a periplasmic factor
•• improving phage display and functional periplasmic expression.** *Nat Biotechnol* 1998, **16**:376-380.
A systematic search for factors improving the phage display of a fusion between a poorly folding single-chain antibody fragment (ScFv) and the gene-3 protein is conducted by subcloning a genomic library on the phagemid. The only insert that favors display encodes Skp, a protein involved in outer membrane protein biogenesis. Skp co-overexpression is also shown to improve the folding of numerous ScFvs secreted in the periplasm. These important results suggests that Skp acts as a general periplasmic molecular chaperone. However, the authors caution that the effect of Skp may be indirect and related to the ability of this protein to prevent deleterious interactions between recombinant proteins and lipopolysaccharides by improving the transport of the latter species.

65. Missiakas D, Betton J-M, Raina S: **New components of protein folding in extracytoplasmic compartments of *Escherichia coli* SurA, FkpA and Skp/OmpH.** *Mol Microbiol* 1996, **21**:871-884.

66. Hayhurst A, Harris WJ: ***Escherichia coli* Skp chaperone
• coexpression improves solubility and phage display of single chain antibody fragments.** *Protein Expr Purif* 1999, **15**:336-343.
This study shows that Skp coexpression can greatly improve the solubility of a highly toxic ScFv secreted in the periplasm and allow its phage display. Skp accumulation is higher when the protein is synthesized from a dicistronic operon than when provided on compatible plasmids under *ara* transcriptional control. High levels of Skp correlate with increase ScFv solubility.

67. Dartigalongue C, Raina S: **A new heat-shock gene, *ppiD*, encodes a
• peptidyl-prolyl isomerase required for the folding of outer membrane proteins in *Escherichia coli*.** *EMBO J* 1998, **17**:3968-3980.
PpiD, an inner-membrane anchored periplasmic PPIase belonging to the parvulin-like family, is isolated as a multicopy suppressor of *surA* mutations. PpiD is shown to be involved in the maturation of outer membrane proteins and double mutations in *ppiD* and *surA* are found to be lethal. By contrast, combinations of *ppiD* mutations with deletions in other periplasmic PPIases (*ppiA* or *fkpA*) or *skp* do not confer a lethal phenotype.

68. van der Wal FJ, Koningstein G, ten Hagen CM, Oudega B, Luirink J:
• **Optimization of bacteriocin release protein (BRP)-mediated protein release by *Escherichia coli*: random mutagenesis of the pCloDF13-derived gene to uncouple lethality and quasi-lysis from protein release.** *Appl Environ Microbiol* 1998, **64**:392-398.
Bacteriocin-release protein (BRP) is a small lipoprotein that causes the release of periplasmic proteins into the growth medium when exported to the cell envelope. Its usefulness is dampened by the fact that, even when secreted from the Lpp leader peptide, accumulation of mature BRP causes turbidity decrease in liquid cultures. In this paper, BRP mutants that retain their ability to release protein into the growth medium but do not have a detrimental effect on cell growth are selected and characterized.

69. Wan EW, Baneyx F: **TolAIII co-overexpression facilitates the
• recovery of periplasmic recombinant proteins into the growth medium of *Escherichia coli*.** *Protein Expr Purif* 1998, **14**:13-22.
E. coli strains bearing mutations in the *tolQRAB* gene cluster or certain other cell envelope proteins spontaneously release periplasmic proteins in the extracellular medium. Unfortunately, these bacteria are fragile and not amenable to high density fermentation. In this study, a ColE1-compatible plasmid encoding the third topological domain of TolA (TolAIII) is used to mimic the 'leaky' phenotype of *tolA* mutants. The usefulness of TolAIII coexpression in promoting the efficient release of resident and recombinant periplasmic proteins is demonstrated. It is noted that although this strategy reduces the total levels of recombinant β-lactamase secreted from the *ompA* leader, formation of periplasmic inclusion bodies is completely abolished.