Strain Improvement of *Escherichia coli* To Enhance Recombinant Protein Production

MICHAEL E. PYNE, KARAN S. SUKHIJA, AND C. PERRY CHOU

19

Among many host systems developed for recombinant protein production, the gram-negative bacterium *Escherichia coli* still retains high popularity due to several technical advantages. First, its fast growth rate and superior environmental adaptability allow *E. coli* to be cultured easily, inexpensively, and if necessary, to a high cell density. Second, its genome-encoded proteins and metabolic pathways are well characterized. Third, numerous technologies and protocols for engineering *E. coli* strains are well developed. Fourth, heterologous proteins of interest can be expressed to high levels (up to 50% of total cellular protein). However, *E. coli* also has technical limitations in recombinant protein production. For example, many human therapeutic proteins require specific glycosylation for full bioactivity. These proteins cannot be produced in *E. coli*, which does not perform such glycosylation, although in some cases it has been possible to engineer a glycosylation pathway in *E. coli* (28). Also, many eukaryotic proteins have multiple domains or disulfide bonds and cannot be functionally expressed in *E. coli* primarily because the organelles and mechanisms in eukaryotic cells for assembling such highly complex protein structures are missing in *E. coli*.

The wild-type *E. coli* cell is not optimally designed for industrial applications. Hence, developing a bioprocess often starts with the construction of the host/vector system so that the *E. coli* strain can be genetically transformed and phenotypically suitable for biomanufacturing. Since the gene of interest is heterologously expressed along with cell growth during the cultivation stage for recombinant protein production, high-level gene expression and high-cell-density cultivation have to be performed simultaneously in order to optimize the culture performance. After cultivation for recombinant protein purification, either the cell pellet (for recombinant proteins retained intracellularly) or the extracellular medium (for recombinant proteins released extracellularly) is harvested and a series of downstream bioprocessing steps are carried out. Experience over the past few decades has shown that most biomanufacturing issues arising at various bioprocessing stages can be addressed through genetic manipulation of the host/vector system. It is the most effective and economic approach to enhance recombinant protein production in terms of mediating functional expression, increasing protein yield, and facilitating downstream purification. However, factors associated with the genetic mechanisms (i.e., replication and gene expression steps) and bioprocessing conditions are often tangled in an unpredictable manner, resulting in complications in strategy development.

Various technical issues that potentially limit recombinant protein production in *E. coli* and the corresponding strategies to address them are summarized in Table 1. Many of these strategies involve alleviating (or even eliminating) the obstacle or limitation associated with a particular bioprocessing stage (i.e., gene expression, cell cultivation, or downstream processing) or with the recombinant protein itself. Typically, the biotechnological basis for strain improvement to enhance recombinant protein production relies on the permanent implementation of desirable traits into the production strain to stimulate both cell growth and functional expression of the target gene during the cultivation (Fig. 1). This is done by genetic restructuring of the production strain, involving both expression vector construction and genetic modifications to the host genome using either recombinant DNA technology, classical mutagenesis and screening technology, or both. Since genetic manipulation of *E. coli* host/vector systems is extensively described in the literature, this chapter reviews the major technical issues associated with *E. coli* strain engineering to enhance recombinant protein production and directs the reader to protocols appropriate for specific applications.

19.1. CONSTRUCTION OF EXPRESSION VECTORS

19.1.1. Genetic Elements of Expression Vectors

Heterologous gene expression in *E. coli* typically involves two major intracellular molecular mechanisms, namely, (i) replication of the expression vector and (ii) expression of the target gene, which includes transcription, translation, and posttranslational processing steps (Fig. 2). To improve a strain for recombinant protein production, effective replication should be ensured such that a high dosage of the target gene is maintained in all the cells of the population. Various genetic elements found in a typical expression vector used for recombinant protein production in *E. coli* are shown schematically in Fig. 3. The functions of each element and a list of common features are summarized in Table 2. Basically, these genetic elements are in place to increase the gene dosage and the efficiency of transcription, translation,

TABLE 1 Factors limiting recombinant protein production in *E. coli* and corresponding strategies for overcoming these limitations

Stage	Limiting events	Strategies for strain improvement	Reference(s)
Gene expression			
Replication	Plasmid instability	Genetically stabilize the plasmid, apply proper culture conditions	7, 39
	Gene dosage	Use a high-copy-number plasmid	35, 99
Transcription	Transcriptional efficiency	Use a strong inducible promoter	11, 95
	Promoter leakage	Use a tightly regulated promoter or other genetic configurations	37, 91
Posttranscriptional stage	mRNA secondary structure near the 5′ terminus	Optimize gene sequence	53, 108
	mRNA stability	Use an RNase-deficient mutant	48, 64
Translation	Translational efficiency	Optimize the Shine-Dalgarno sequence for optimal binding of ribosome, optimize coding sequence	26, 53, 108
	Ribosome availability	Enhance the production of functional ribosome	104, 109
	tRNA availability (codon bias)	Coexpress the rare tRNA gene, optimize the codon	53, 108
Posttranslational stage	Protein translocation	Use a proper signal peptide	18
	Protein folding	Coexpress chaperones, use a low culture temperature, use protein fusion technology, secrete the protein into the extracellular medium	43, 51, 56, 82
	Protein disulfide bond formation	Secrete the protein into the periplasm or extracellular medium, coexpress factors involved in disulfide bond formation, use a mutant that can form disulfide bonds in the cytoplasm	24, 51
	Proteolysis	Use a protease-deficient mutant, use protein fusion technology, perform extracellular protein secretion	4, 32, 97
	Glycosylation	Use a genetically engineered strain	57, 84
Cultivation	Physiological deterioration	Eliminate various deterioration factors	8, 30
	Heat shock responses	Manipulate heat shock responses	15, 110
	Nutrient uptake	Avoid nutrient overfeeding	61
	Toxic metabolite accumulation	Reduce the accumulation	17, 23
	Growth inhibition	Eliminate inhibitory factors	54, 76
Downstream processing	Extracellular protein secretion	Use a proper secretion strategy	32
	Protein recovery	Use protein fusion technology	97
	Protein solubility	Use protein fusion technology, modify the protein sequence	66, 97
Protein feature	Toxicity	Use a tightly regulated promoter to prevent expression leakage	37, 91
	High complexity and difficult expression	Use a special strain for functional expression	71, 105
	Proteolysis	Modify protein sequence	75

and posttranslational processing so that the target gene can be functionally expressed at a high rate. The copy number of the expression vector is intrinsically determined by the replication origin. A strong inducible promoter is required to increase the transcriptional efficiency and boost the mRNA level. Because of the coupling of transcription and translation, recombinant protein production is often improved simply by using a strong promoter. It may be necessary that the promoter be tightly regulated to avoid leaky expression under noninduced conditions, which is particularly important for proteins with a high toxicity. The molecular details of the transcriptional regulation mechanism and whether it involves an activator or repressor can be very important for achieving maximum production. The terminator used is also important to ensure proper termination of transcription of the target gene, since transcription running past the target gene can interfere with other functions of the expression vector. The ribosome binding site and sequences near

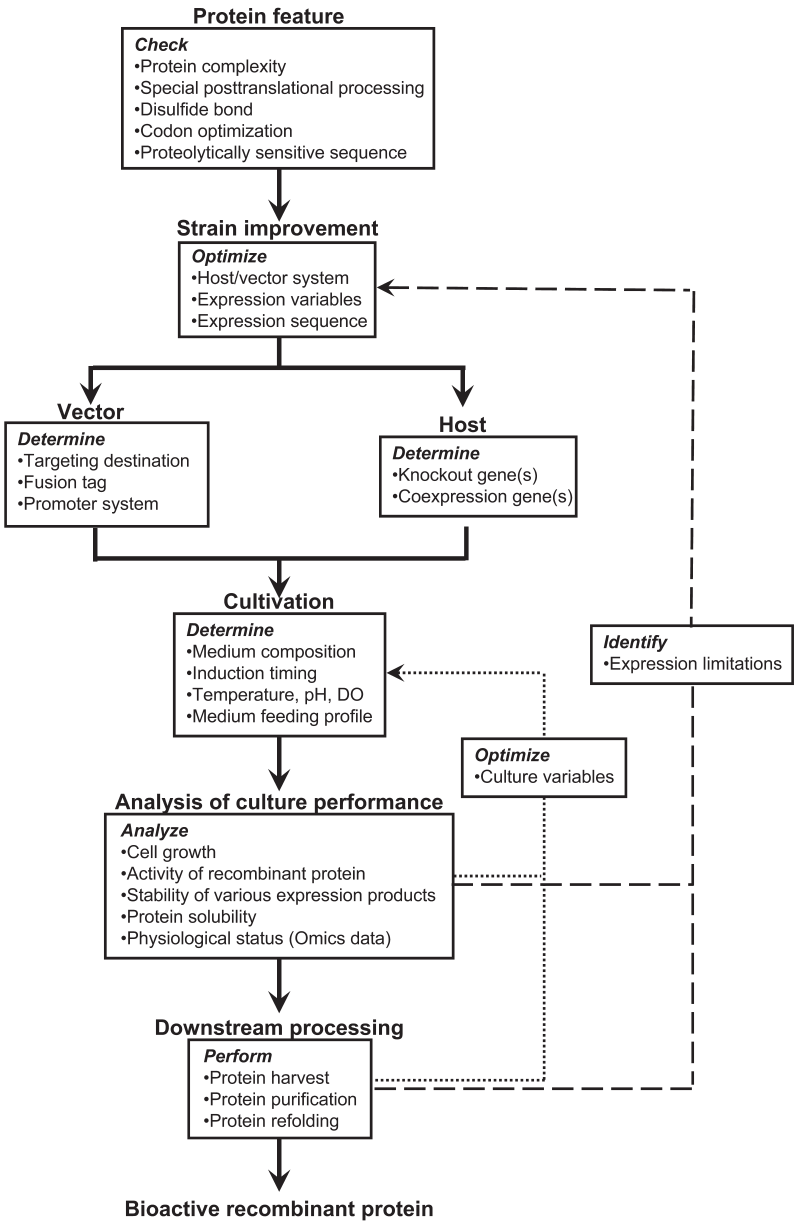


FIGURE 1 Flowchart for *E. coli* strain improvement to enhance recombinant protein production. The strategies depend on identification of the specific factors limiting the overall recombinant protein yield and include optimization of the host/vector system, expression variables, and expression sequence. Refer to Table 1 for a list of technical limitations and how to deal with them. DO, dissolved oxygen.

the start codon affect the efficiency of ribosome binding to the target mRNA and the rate of translation initiation. The antibiotic resistance marker enables selection for cells harboring the expression vector. The fusion tag, situated either at the 5′ or 3′ end of the target gene (so that the expressed tag is fused with the N or C terminus of the target protein), offers several expression advantages, particularly for soluble expression and protein purification. Most expression vectors have multiple cloning sites to enable flexible cloning of various target genes with transcriptional or translational fusions for different expression scenarios (Fig. 3). Many expression vectors with a selection of these genetic elements are commercially available and significantly facilitate the construction work. While the construction of the expression vector is considered a key step for recombinant protein production, developing an effective bioprocess also relies on a proper design of the expression host, cultivation, and downstream processing that are compatible with the expression vector.

19.1.2. Limiting Steps in Gene Expression

Recombinant protein production involves a series of molecular mechanisms that must be optimized, regulated, and coordinated. A malfunction at any of the steps (replication, transcription, translation, posttranslational processing, etc.) or the presence of various undesired events (such

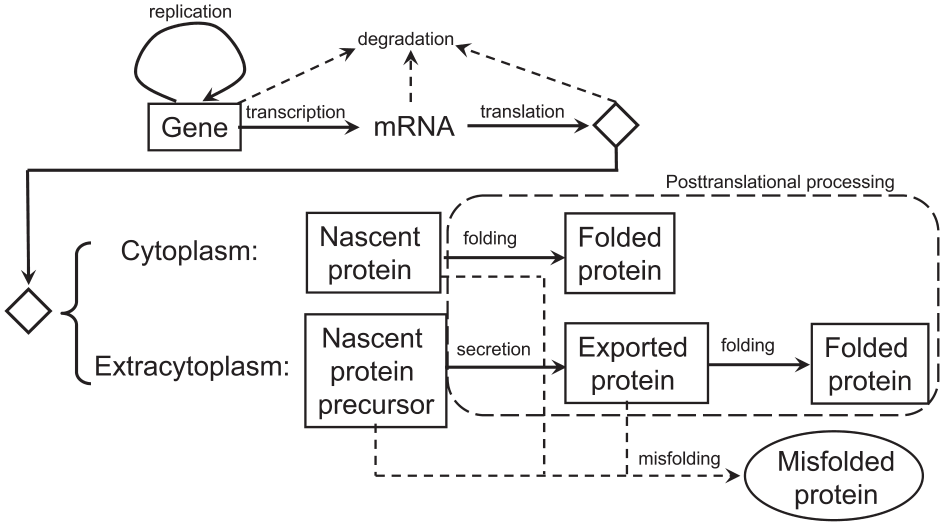
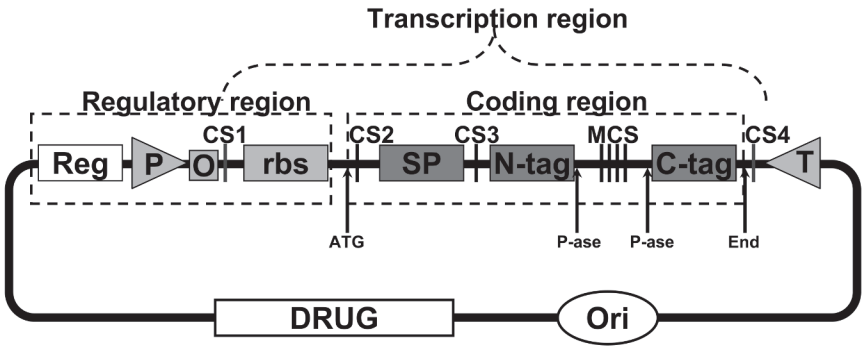


FIGURE 2 Molecular events associated with recombinant protein production in *E. coli*. Recombinant protein production involves a series of complex molecular mechanisms, such as replication of the expression vector, transcription and translation of the gene of interest, and various posttranslational processing steps (including protein secretion, folding, and disulfide bond formation). Production can be limited by low efficiency at any one of these steps or by an abnormal event that diverts protein synthesis into a nonproductive pathway (e.g., protein misfolding or degradation of DNA, mRNA, or protein).



Cloning sites	Expression product	Comments
CS2(or CS1)/CS4	ORF	1. Another rbs is needed when CS1 is used. 2. ORF might contain a SP for secretion.
CS2(or CS1)/MCS	ORF::C-tag	
CS3/CS4	SP::ORF	1. SP might not be present for cytoplasmic expression. 2. P-ase (protease cleavage site) might be available to release the target protein moiety.
CS3/MCS	SP::ORF::C-tag	
MCS/CS4	SP::N-tag::ORF	
MCS/MCS	SP::N-tag::ORF::C-tag	

FIGURE 3 A typical *E. coli* expression vector (i.e., plasmid) for recombinant protein production. Several expression and cloning features are shown, including Reg (gene encoding the regulator, either transcriptional activator or repressor), P (promoter), O (operator), rbs (ribosome binding site), SP (signal peptide), N-tag (N-terminal fusion tag), C-tag (C-terminal fusion tag), T (terminator), CS (cloning site), MCS (multiple cloning sites), Ori (replication origin), DRUG (drug resistance gene), ATG (initiation codon encoding methionine), P-ase (protease cleavage site), and End (stop codon). Note that, depending on the cloning site(s) for insertion of a target gene (i.e., open reading frame [ORF]), a transcriptional or translational fusion vector can be constructed to express a gene product (either ORF or ORF-fusion) containing various feature domains and targeting in the cytoplasm or extracytoplasmic compartment.

TABLE 2 Genetic elements found in a typical expression vector used for recombinant protein production

Element	Common features	Function(s)
Origin	pUC, pMB1, ColE1, p15A, pSC101	Regulate gene dosage
Promoter	T7, *araB*, *lacUV5*, *tac*, *trc*, *tet*A, λ_L, λ_R	Enhance transcriptional efficiency
Terminator	*rrnB* T1, T2	Ensure transcriptional termination
Regulator	Transcriptional activators or repressors	Regulate transcriptional efficiency
Ribosome binding site	Shine-Dalgarno sequence, translational enhancer	Improve translational initiation
Drug marker	Ampicillin, chloramphenicol, tetracycline, kanamycin	Select the transformed cell
Signal sequence	Tat, TorA, OmpA, PhoA, PelB, SpA	Translocate the protein precursor across the cytoplasmic membrane
Tag	Streptavidin-binding peptide (Strep), poly-histidine, FLAG peptide, DsbA, DsbC, thioredoxin (TRX), glutathione *S*-transferase (GST), maltose-binding protein (MBP), N-utilization substance protein A (NusA), calmodulin binding peptide (CBP), small ubiquitin-like modifier (SUMO)	Facilitate protein purification, improve protein solubility

as degradation or misfolding of the target protein) can jeopardize the production process (Fig. 2). A high yield of bioactive protein implies not only a high rate for all the expression steps (i.e., fluxes), but also coordination of various intracellular events to balance these expression fluxes. Typically, for a particular host/vector system under a specific growth condition, overall performance can be limited by a single expression step, which can result in the accumulation of an inactive intermediate prior to the limiting step or a reduced level of the intermediate after that step. Theoretically, strategies based on enhancing the limiting step can lead to an overall improvement in recombinant protein production. For example, transcription is recognized as a common step limiting the abundance of mRNA for subsequent translation. A stronger promoter can boost the mRNA level through increased transcriptional efficiency, and the recombinant protein yield is often increased as a result. In addition to the high expression rate, all the expression intermediates (i.e., mRNA and protein precursors) and the final gene product should remain stable in order to achieve a high protein yield during the production. There are cases where one or more of these intermediates are subject to intracellular degradation and recombinant protein production is reduced as a result. Strategies to stabilize these labile biomolecules can be developed for strain improvement.

While the limiting step is determined by the intracellular genetic makeup of the host/vector system, it is prone to change depending on the transient physiological state. This implies that the strategies for genetic construction of the production strain and for its biological cultivation should be simultaneously configured to optimize the culture performance for recombinant protein production. Technically, it is important that the limiting step be properly identified since improving the efficiency of that step should enhance the overall expression performance. However, improving a single expression step can simply shift the limitation to a different step, giving little to no improvement in the overall recombinant protein yield. For example, though the transcriptional efficiency can be enhanced by using a strong inducible promoter, the boosted mRNA level does not necessarily lead to the increased production of the target protein due to the limitation from another expression step, such as translation or protein folding. Even more troublesome, the genetic strategy used to resolve one limiting step might actually cause the development of another limitation. For example, the presence of high-copy-number plasmids to increase the gene dosage or the induction of strong promoters to enhance the transcriptional efficiency could negatively affect the physiological condition of the producing cells, resulting in growth inhibition and a reduced recombinant protein yield. Such complications often limit the applicability of these genetic strategies.

19.1.3. Localization of Recombinant Protein

Another key issue for recombinant protein production in *E. coli* is the location to which the final protein product should be targeted, including the cytoplasm, periplasm, cell surface, or extracellular medium. This is determined by various factors associated with the recombinant protein, such as its susceptibility to intracellular proteolysis, its tendency to misfold, the requirement for disulfide bond formation, and the process used for downstream protein purification. While the cytoplasm is the intracellular compartment where all the nascent proteins are first synthesized and where most of them reside, it might not be a suitable destination for a recombinant protein. Therefore, the target gene product might require the secretion to either the periplasm or extracellular medium for functional expression (18, 32). The periplasm provides an oxidative environment that allows disulfide bond formation, as well as a separate compartment from the cytoplasm for facilitating protein purification. The expression of a recombinant protein on the *E. coli* cell surface is performed only for special applications other than mass production (62). Secretion to the extracellular medium is an alternative strategy for functional expression of recombinant proteins that are otherwise difficult to express intracellularly due to technical limitations such as protein misfolding or disulfide bond formation. With the host/vector system acting as a whole-cell biocatalyst, extracellular secretion enables the continuous production of recombinant protein and facilitates downstream purification. However, recombinant protein yields are often low when secretion is involved.

There are two major structural barriers to protein secretion, the cytoplasmic membrane and the outer membrane. Protein translocation across the cytoplasmic membrane is usually driven by the Sec-dependent type II secretion

pathway under the direction of a cleavable signal peptide at the N terminus of the secreted polypeptide. Technically, a DNA sequence encoding a signal peptide can be attached at the 5′ end of the open reading frame encoding the target protein to form a translational fusion. The expressed polypeptide precursor is expected to be translocated across the cytoplasmic membrane under the direction of the signal peptide. Several signal peptides (see Table 2) have been adopted successfully for exporting recombinant proteins into the periplasm or onto the outer membrane of *E. coli*. Very few *E. coli* proteins pass the outer membrane and are secreted extracellularly. Two other secretion pathways (i.e., type I and type III) are known to be responsible for the extracellular secretion of certain *E. coli* proteins without involving a periplasmic intermediate (10, 98). Although biochemical and genetic strategies (36, 90, 102, 106) based on permeabilizing the outer membrane have been developed for the extracellular release of periplasmic proteins (which are typically exported from the cytoplasm via the type II secretion system), they often impair cell physiology and reduce productivity. The type I and type III secretion systems have been successfully used to secrete target proteins extracellularly, though the host/vector system often requires a specific design (9, 22, 29, 101).

19.1.4. Protein Fusion Technology

Protein fusion technology serves as a powerful tool with multiple applications for recombinant protein production (97). Basically, the target gene product is expressed as a fusion protein, where the fusion tag does not affect the bioactivity of the target protein. The typical fusion tag is one that can be captured by affinity chromatography, greatly facilitating downstream purification. In addition, a specific cleavage site can be engineered at the junction between the two fusion partners so that the target protein can be released as the native protein. While such tags theoretically can be fused with the target protein at either the N or C terminus, the N terminus is more often used because many N-terminal tags contain local DNA sequences optimized for translational initiation. With an appropriate design, the production of fusion protein can be enhanced. Furthermore, several fusion tags have been demonstrated to promote protein folding because they can rapidly attain their native conformation. As a result, the solubility and stability of a fusion protein are often superior to the heterologous protein alone (107). In some cases, particularly with a small tag, the target protein retains its bioactivity and can be used without removing the tag.

19.1.5. Protein Misfolding

Another common technical issue for recombinant protein production in *E. coli* is protein misfolding, which results in loss of bioactivity. The problem can be viewed as the expression flux imbalance between polypeptide formation and polypeptide processing at any posttranslational stage (including folding). Such imbalance diverts protein synthesis into a nonproductive pathway and results in the accumulation of misfolded protein species, which typically aggregate into insoluble inclusion bodies that are inactive. Inclusion bodies can be formed in either the cytoplasm or periplasm. Factors that can cause this expression flux imbalance and protein misfolding include heat shock, high protein concentration, certain amino acid sequences within the expressed protein, and physiological stress on the cells. Some of the strategies that can be used to alleviate this problem include lowering the cultivation temperature or coexpressing particular chaperones. However, in some expression scenarios, the formation of inclusion bodies is considered advantageous, particularly for toxic proteins, since inclusion bodies are often easy to harvest and are less toxic to the production cell. The use of inclusion bodies as an expression strategy hinges primarily on the ease of recovering bioactivity through in vitro protein refolding.

19.2. DEVELOPMENT OF THE EXPRESSION HOST

19.2.1. Genetic Modification

Enhancing recombinant protein production relies on the construction of a compatible host/vector system for effective expression of the target gene. Wild-type *E. coli* cells have several mechanisms to protect them from environmental stress, and these can be activated during high-level recombinant protein production. These natural safeguards can potentially limit performance in industrial applications. For example, several nucleases and proteases can be induced to digest foreign DNA and protein, respectively, in response to the physiological stress caused by recombinant protein production. Hence, the host strain requires genetic modification to improve its suitability for biomanufacturing through the implementation of desirable traits on either the host chromosome or expression vector. Thus, obstacles to recombinant protein production due to the natural safeguard mechanisms mentioned above can be eliminated or at least mitigated through appropriate genetic engineering.

19.2.2. Identifying the Targets that Require Genetic Modification

Essentially, genetic modifications involve changing the expression level of certain key genes. At one extreme native *E. coli* genes can be knocked out, and at the other extreme heterologous genes encoding a useful activity can be expressed. Expression vectors provide a flexible way to coexpress multiple genes and can be used for such genetic modification (100). If necessary, the expression level can be fine-tuned with tunable promoters (25) or gene-silencing mechanisms (48, 49). Identifying the key genes that can directly or indirectly affect recombinant protein production is one of the most critical and difficult aspects in developing the genetic engineering strategy. While individual genes of potential value are constantly being identified in many research programs and large-scale random overexpression projects (65), recent progress in transcriptomic and proteomic analyses offers a systematic approach for rapid identification of hundreds of such candidates (1, 31, 38, 83, 111). However, proper selection of the key genes among the candidates that might facilitate optimal engineering of the production strain remains challenging.

General methods for engineering microbial strains to derive desired phenotypes have been extensively reviewed (78, 88). Whole-organism mutagenesis followed by phenotypic screening and selection has been the classical strategy for obtaining desirable traits for biomanufacturing, such as increased product titer (yield or productivity) and greater physiological robustness under the overproduction conditions. Random mutations can be introduced by exposing the production strain to UV light or a chemical mutagen such as *N*-methyl-*N*′-nitro-*N*-nitrosoguanidine, or by random genome-wide transposon insertions (2). Multiple rounds of mutagenesis and screening are often required to optimize the strain. This approach can potentially identify

multiple mutation targets leading to the desired phenotype, and the targets can be related to the production pathway in obscure ways. Given the increasing accessibility of genomic sequencing services, such mutations can be identified and novel mechanisms relevant to recombinant protein production can be discovered. A major drawback with multiple rounds of mutagenesis and screening is that the process can be time-consuming and resource intensive. Furthermore, the result of identifying the manipulation targets might not be reproducible. While mutagenesis is easy to perform, establishing an effective screening protocol can be challenging. Recent development in high-throughput screening technologies with advanced analytical tools such as flow cytometry, chromatography, and mass spectrometry significantly improves the screening efficiency, particularly in the area of metabolic engineering applications. However, these approaches are often product specific.

19.2.3. Physiological Improvement

The metabolic burden arising from recombinant protein production can induce physiological stress and in turn reduce recombinant protein yield. One strategy to alleviate the impact of this burden is to mimic or boost the natural stress response, thus maintaining healthy cell physiology during recombinant protein overproduction (20). Several types of genes can be useful for this strategy, one example being those encoding chaperones, which are often induced during the heat shock response and assist with folding of damaged or misfolded proteins in the cytoplasm (5) or periplasm (72). These chaperones can enhance the solubility, structural stability, translocation/secretion efficacy, or even disulfide bond formation of the expressed gene products (41, 55, 87). Various proteases (cytoplasmic, periplasmic, or membrane-bound) are also induced during stress responses that degrade damaged or abnormal proteins. Such proteases can also degrade the desired recombinant protein (42) and therefore represent major targets for gene knockout. While there are reports of using protease-deficient mutants to improve recombinant protein production (69), knocking out a large number of protease genes for this purpose will result in a sick strain because *E. coli* requires these proteases for regulated protein turnover and physiological adaptation. Some have reported improvement in both cell physiology and recombinant protein yield via protease coexpression, which is presumably mediated by selective proteolysis of misfolded proteins that may be toxic (47, 77). Various stress-sensing pathways have been identified (27, 45, 63), and genes involved in this process can be logical targets for genetic modification. Stationary-phase genes encode proteins that may lead to a reduction in cellular and metabolic activity, which can negatively affect recombinant protein production, and as such, these genes are targets for strain improvement (19, 46). Other metabolic engineering strategies that have been used to improve a strain's adaptation to harsh production conditions include reducing the secretion of toxic metabolites (34, 103), boosting the nutrient assimilation rate (67), or increasing the supply of limiting cofactors (86).

19.2.4. Genotypes Suitable for Recombinant Protein Production

E. coli strains engineered for the application of recombinant DNA technology have been derived from both K-12 and B strains. The K-12-derived strains (e.g., DH5α and JM109) often carry several endogenous mutations (e.g., *endA*, *recA*, and *lacZ*) specifically designed for molecular cloning purposes, though they can also be used as a host for recombinant protein production. On the other hand, *E. coli* B strains, e.g., BL21 F^- *ompT dcm lon gal hsd*$S_B(r_B^-, m_B^-)$ (95) and its derivatives, are probably the most common production strains primarily owing to the two mutations in protease genes, i.e., *lon*, encoding intracellular ATP-dependent protease, and *ompT*, encoding outer membrane protease, which can potentially alleviate the degradation of heterologous protein products. In addition, BL21 is less sensitive to glucose than K-12 strains (80) and can tolerate higher levels of metabolic stress associated with recombinant protein production (44), making it suitable for high-cell-density cultivation. These technical advantages may be related to its more integrated outer membrane structure, since it is observed that BL21 cannot be used as a suitable expression host for cell-surface display or extracellular secretion of recombinant protein (N. Narayanan and C. P. Chou, unpublished data).

Several genetic traits have been introduced into BL21 to improve its utility as a protein overproducer. The most popular is the genomic integration of a DE3 lysogen that contains the T7 RNA polymerase gene, expression of which is regulated by the *lacUV5* promoter. The resulting strain, BL21(DE3), can be used with expression vectors containing the T7 promoter system for overexpression of recombinant protein (96). Other genetic features have also been included in BL21 to enhance its tolerance against physiological impacts associated with recombinant protein production (particularly for toxic proteins). For example, the basal level arising from leaky expression during noninducing conditions is minimized by replacing the isopropyl-β-D-thiogalactopyranoside (IPTG)-inducible *lacUV5* promoter in DE3 with the *araB* promoter for tighter transcriptional regulation (14) (e.g., BL21-AI™, available from Invitrogen [Carlsbad, CA]). Alternatively, a plasmid (pLysS or pLysE with a p15A replication origin) or a mini-F plasmid containing the gene encoding T7 lysozyme, which is a natural inhibitor of T7 RNA polymerase (94), can be included to minimize leaky expression [e.g., BL21(DE3) (pLysS) and BL21(DE3) (pLysE) from Novagen (Madison, WI) or NEB Express and T7 Express from NEB (Ipswich, MA)].

Another interesting approach is to challenge the expression host with a plasmid that overexpresses a toxic protein and then screen for mutants that gain the ability to produce the protein. C41(DE3) and C43(DE3) (available from Lucigen Corp. [Middleton, WI]) contain mutations that facilitate overexpression of certain membrane proteins and were derived from BL21(DE3) by this strategy (71). The mutations resulting in the improved expression performance of these strains were recently identified to be located in the *lacUV5* promoter, and the discovery led to the construction of a BL21(DE3) derivative, i.e., Lemo21(DE3), with a precise T7 RNA polymerase activity for tunable protein expression (105). Precise control of expression has also been achieved with a *lacY* mutation (e.g., Tuner™ and Origami™ B from Novagen). It appears that precise control of the expression level of the heterologous gene of interest is critical for achieving maximum protein production.

Since T7 RNA polymerase synthesizes mRNA rapidly, resulting in the uncoupling of transcription and translation, transcripts are subject to degradation by endogenous RNases. Mutation of the *rne* gene, encoding RNase E, the major enzyme for RNA degradation (64), stabilizes such highly expressed mRNA transcripts, and BL21 strains with this mutation are available (e.g., Star™ from Invitrogen).

The *endA* gene encodes a nonspecific endonuclease, endonuclease I, which can degrade plasmid DNA prepared by

the typical mini-prep protocols. In addition, *recA* encodes an ATPase involved in homologous DNA recombination in vivo (6). K-12 strains often carry mutations in these two genes to maintain the structural stability and product quality of plasmid DNA for molecular cloning. These two mutations have been introduced into BL21 strains for large-scale production of plasmid DNA (79). Though the functions associated with the *recA* and *endA* gene products might seem to be required for healthy cell physiology, protein overexpression using such double-mutant BL21 strains (e.g., Acella™ from EdgeBio [Gaithersburg, MD]) does not seem to encounter any problems associated with recombination or DNA degradation. Furthermore, such strains allow both protein expression and molecular cloning to be conducted in a single host.

Overexpression of heterologous proteins can potentially result in transient shortage of amino acids and the corresponding charged tRNAs for translation, which in turn can lead to the problems discussed above due to uncoupling of transcription and translation. This physiological regulation, known as the stringent response (12, 40), can be relaxed by the *relA* mutation such that RNA is still synthesized in the absence of protein synthesis. On the other hand, protein overexpression can be limited by depletion of certain rare tRNAs, particularly for expression of genes having a codon bias significantly different from that of *E. coli* (92). This limitation can be alleviated by coexpressing genes for those rare tRNAs, such as *argU*, *proL*, *ileY*, and *leuW*. Several BL21-CodonPlus® (Stratagene [La Jolla, CA]) and Rosetta™ (Novagen) strains expressing these tRNA genes are available, and their use for recombinant protein production appears to be justified (50). Alternatively, the heterologous gene can be modified or completely synthesized to have the optimal codon usage of *E. coli* (53, 74, 108).

For certain eukaryotic or therapeutic proteins, heterologous overexpression is limited by posttranslational processing, such as disulfide bond formation, which can be crucial for a protein's bioactivity. Though *E. coli* contains the oxidative milieu of the periplasm, where disulfide bonds can be formed in vivo, targeting of proteins to this compartment can be limited by the efficiency of translocation or periplasmic folding. Hence, there is a desire for strains that can produce proteins containing disulfide bonds in the cytoplasm. However, the *E. coli* cytoplasm contains two thioredoxins and three glutaredoxins, all of which are maintained in a reduced state by thioredoxin reductase (TrxB) and glutathione. Consequently, disulfide bond formation hardly occurs in this reduced compartment. Introducing mutations in the genes for both thioredoxin reductase (*trxB*) and glutathione reductase (*gor*) (or glutathione synthetase [*gshB*]) can transform the reduced cytoplasm into an oxidized one (93). This double mutation in the BL21 background is found in the Origami™ B strains (Novagen), which offer an alternative for the overexpression of proteins containing disulfide bonds in the oxidative cytoplasm. However, cell growth appears to be impaired by this double mutation.

Another host feature that can be useful for downstream processing is genomic inclusion of the λ R gene, encoding the λ lysozyme (i.e., endolysin). The resulting strain (XJb Autolysis™ from Zymo Research Corp. [Orange, CA]) is efficiently lysed by arabinose induction.

Note that most of the above mutations are implemented with an antibiotic resistance marker, which limits the choice of markers for the expression vector. In practice, it will be more convenient to construct mutants without drug markers using chromosomal engineering strategies developed recently (21).

19.2.5. Chromosomal Engineering of *E. coli*

Chromosomal engineering refers to the techniques for making site-specific insertions, replacements, or deletions in the chromosome. With an extensive ability to knock out an existing gene or express a foreign gene, the technique becomes a powerful tool for strain improvement. The most common one, known as recombineering, is based on homologous recombination (Fig. 4). In general, exogenous DNA containing an allele of interest is first delivered into the recipient cell directly or via a shuttle vector (either plasmid or virus based) or conjugation. Because *E. coli* is artificially transformable, conjugation is seldom used for delivering the exogenous DNA. The efficiency for homologous recombination in vivo can be enhanced (e.g., by expressing exonuclease and recombinase) so that the key allele in the delivered DNA can be site-specifically exchanged with the corresponding genomic allele or can be inserted into the chromosome. A cotransduced drug marker is often used for selection of the mutant derivative, whose genotype can be verified by colony PCR or genomic sequencing. Several methods based on transposon mutagenesis and homologous recombination (21, 33, 73, 112–114) have been developed for genomic modification to construct *E. coli* strains. Very large fragments carried on bacterial artificial chromosomes can be inserted into the *E. coli* genome (58, 85). This is particularly useful for metabolic engineering of *E. coli*, since novel strains can be constructed that express entire pathways for nonnative metabolites. A comprehensive *E. coli* mutant library was made using the chromosomal engineering protocol involving the λ Red-mediated recombination system (21). Each gene in the chromosome was replaced with a kanamycin resistance cassette flanked by homologous regions to the target allele. This library of single-gene knockout strains is called the Keio collection (3) (www.shigen.nig.ac.jp/ecoli/strain/top/top.jsp). Each mutant gene in the Keio collection can be transferred via P1-phage transduction (70) from its original BW25113 background to other genetic backgrounds using kanamycin for selection. Furthermore, the kanamycin resistance cassette between two flanking FLP recombination target (FRT) sites can be removed via FLP recombination using a helper plasmid (16). This will allow the construction of drug-resistance-free strains with multiple gene knockouts. The establishment of the Keio collection and its availability to researchers worldwide have marked an important biotechnological milestone by significantly facilitating *E. coli* mutant construction. The recently commercialized recombineering protocol from Gene Bridges (Heidelberg, Germany; www.genebridges.com), also based on λ Red-mediated recombination, allows versatile chromosomal engineering, including gene disruption, deletion, insertion, point mutation, modification, and even promoter fine-tuning, and can serve as a versatile manipulation tool for strain improvement and even optimization.

Protocol 1. Gene Knockouts in *E. coli* Using λ Red-Mediated Recombination

This protocol is based on Datsenko and Wanner (21) using the *E. coli* K-12 strain. There are other similar recombineering methods that use plasmids containing different promoters and markers, but these protocols are also based on the recombinases from either the Rac prophage or bacteriophage λ or both (58). Genetic modification using recombineering is not limited to chromosomal genes, but can

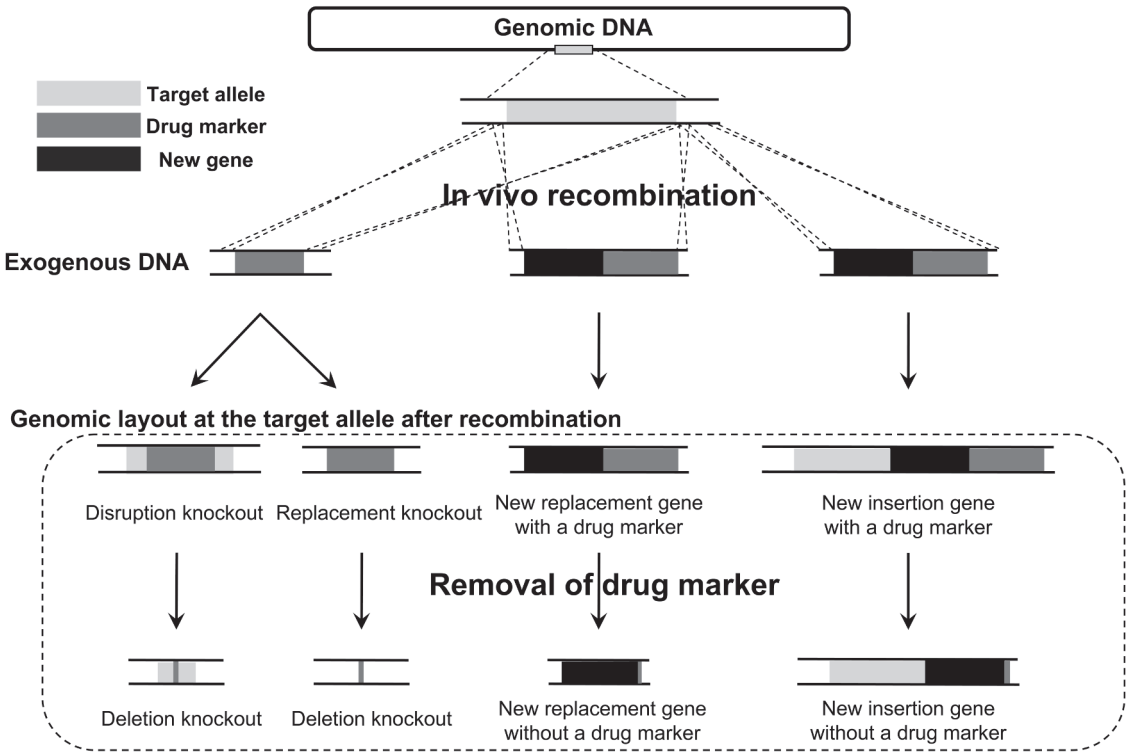


FIGURE 4 Chromosomal engineering of *E. coli* based on homologous recombination for either site-specific gene knockout or gene insertion. The target allele/site is first selected and the exogenous segment is prepared in vitro (e.g., by PCR). Because *E. coli* is artificially transformable, the exogenous DNA can be delivered into the recipient cell through electroporation. The efficiency of in vivo recombination can be enhanced by expressing key enzyme(s) associated with the recombination. A drug resistance marker is often introduced as the major replacing cassette or cotransduced with a new gene for selection of transformed cells, and can be subsequently deleted in vivo (e.g., by FLP recombination).

also be applied to extrachromosomal DNA molecules, such as plasmids and bacterial artificial chromosomes.

1. Obtain the set of recombineering vectors denoted pKD46, pKD3 (or pKD4), and pCP20.
2. Transform pKD46 into the *E. coli* strain containing the gene you wish to modify. The plasmid has a low copy number and a temperature-sensitive origin of replication, repA101. Thus, the cells should be incubated at 30°C during the transformation.
3. Design the primers for PCR amplification of the insertion cassette on a template plasmid (pKD3 or pKD4). The insertion cassette is flanked by approximately 35- to 50-bp homology extensions. These extensions are homologous to the sequences surrounding the chromosomal region to be replaced. The sense and antisense primers for the PCR should be designed such that the last 20 bp of each primer anneal to the insertion cassette to be amplified. The PCR product is gel purified, DpnI digested, and then repurified to ensure that all template plasmid is removed. The insertion cassette will contain the chloramphenicol resistance gene (for pKD3) or kanamycin resistance gene (for pKD4) between two FRT sites, which are used to remove the antibiotic resistance genes after recombination.
4. Make an electrocompetent cell suspension by growing a colony obtained from step 2 in LB medium containing 10% arabinose at 30°C to induce the expression of the three main λ Red genes on pKD46, *exo, bet*, and *gam*.
5. Electroporate approximately 0.5 to 1.0 μg of the insertion cassette prepared in step 3 into the electrocompetent cells by selection of the transformants on agar plates containing 34 μg/ml chloramphenicol or 50 μg/ml kanamycin at 37°C.
6. Verify the absence of pKD46 in the transformants by streaking and growing a few colonies on agar plates containing 100 μg/ml ampicillin as well as agar plates containing 34 μg/ml chloramphenicol (or 50 μg/ml kanamycin). The cells that have lost pKD46 will only grow on the chloramphenicol (or kanamycin) plates, and not on the ampicillin plates. Insertion of the antibiotic resistance genes can be further verified using colony PCR.
7. Remove the antibiotic resistance gene via FLP-FRT recombination. Transform the recombineered cells with pCP20, which contains a temperature-sensitive origin of replication, repA101. Plate the cells on LB agar plates containing 100 μg/ml ampicillin at 30°C.
8. Prepare overnight cultures of a few of the colonies obtained from step 7 in LB medium without antibiotic at 37°C to induce the expression the flippase (*flp*) gene for excising the antibiotic resistance gene region between the FRT sites.
9. Verify the loss of all antibiotic resistance genes by streaking the overnight cultures on LB agar plates containing appropriate antibiotics and incubating overnight at 37°C. The loss of the antibiotic resistance genes can be further verified using colony PCR.

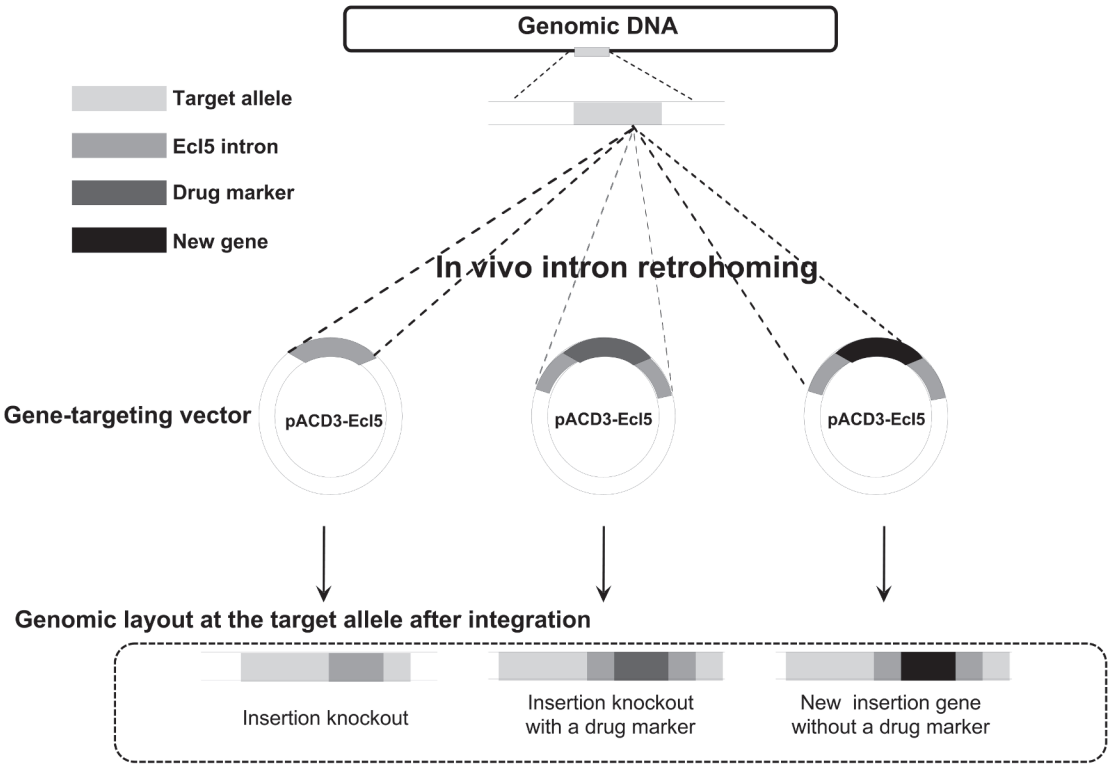


FIGURE 5 Chromosomal engineering of *E. coli* based on the intron gene-targeting system for either site-specific gene knockout or gene insertion. The target allele/site is first selected and the gene sequence is entered into the EcI5 computer algorithm to obtain putative insertion sites and corresponding mutagenesis primers. The intron is then retargeted, ligated into a targetron vector, and expressed within the appropriate host strain. The pACD3-EcI5 vectors contain a convenient MluI restriction site for inserting cargo genes such as a drug marker for knockout selection or a foreign gene for chromosomal expression in *E. coli*.

A chromosomal gene knockout strategy not involving homologous recombination is also available (Fig. 5). The technique utilizes a bacterial group II intron, such as L1.LtrB from *Lactococcus lactis*, and an intron-encoded protein (IEP), for site-specific integration via an in vivo retrohoming process (68). The mechanism begins with the IEP locating certain fixed nucleotide positions within the target DNA sequence, thereby facilitating specific base-pairing interactions between the intron RNA and target DNA. Due to both endonuclease and reverse transcriptase activity of the IEP, the intron RNA is able to hybridize with one strand of the target DNA and serve as a template for cDNA synthesis. Once the integration and reverse transcription events are complete, the nicked insertion site is sealed by host repair mechanisms to generate a stable and permanent 0.9-kb L1.LtrB intron insertion, which contains a stop codon in all six reading frames. Since specific base pairing between the intron RNA and target DNA occurs at only a few key regions (denoted IBS, EBS1d, and EBS2), it is possible to mutate and retarget the L1.LtrB intron for insertion into virtually any chromosomal site. Recently, a new bacterial group II intron isolated from *E. coli* O157:H7, called EcI5, has been shown to possess higher site-specific integration frequencies than the L1.LtrB intron (115), eliminating the need for antibiotic selection markers.

Protocol 2. Gene Knockouts in *E. coli* Using the EcI5 Intron

This protocol, based on the recent work by Zhuang et al. (115), describes gene knockout starting from EcI5 intron retargeting through induction of intron transcription and mutant screening.

1. Obtain the set of EcI5 gene-targeting vectors denoted pACD3-EcI5A, pACD3-EcI5C, pACD3-EcI5G, and pACD3-EcI5T. The vectors differ in just one nucleotide position at the EBS3 site. The appropriate vector must be chosen such that the EBS3 nucleotide is complementary to the +1 position (IBS3) within the predicted EcI5 intron insertion site.
2. Select a gene to be knocked out and input the coding sequence into the computer algorithm developed by Zhuang et al. (115). The algorithm ranks putative intron insertion sites and provides primer sequences corresponding to each insertion site.
3. Select at least two insertion sites and obtain the IBS1/IBS2, EBS1, and EBS2 primers for each site, in addition to the universal EcI5 primer.
4. The EcI5 intron is retargeted by a two-step PCR procedure in which the IBS1/IBS2 and EBS2 primers are used to generate a 336-bp PCR product, while the EBS1 and universal EcI5 primer are used to generate a 230-bp PCR product. These small PCR products are then gel purified and used in a second PCR to generate the fully mutated and retargeted 546-bp product. The internal EBS2 and EBS1 primers contain 20 bp of overlap and are fused together in the second PCR using the flanking IBS1/IBS2 and universal primers. The final product is also gel purified.
5. The retargeted PCR product is flanked by XbaI and AvaII sites, which are present in the IBS1/IBS2 and

universal EcI5 primers, respectively. The retargeted PCR product is swapped for the corresponding region in a previously targeted EcI5 targetron vector. Prepare restriction digests using AvaII and XbaI to digest both the intron PCR product and pACD3-EcI5A/G/C/T vector.
6. Gel purify the digested EcI5 targetron vector.
7. Prepare a ligation reaction containing 15 to 30 ng digested PCR product and 50 ng digested and purified targetron vector. This corresponds to a molar insert-to-vector ratio of approximately 3:1 to 6:1.
8. Inactivate the ligase using heat or phenol extraction.
9. Transform 1 μl of the ligation into an *E. coli* cloning strain using the heat shock or electroporation method. Plate the cells onto LB agar plates containing 25 μg/ml chloramphenicol. All pACD3-EcI5 targetron vectors contain a chloramphenicol antibiotic marker for plasmid selection and propagation.
10. Screen approximately five colonies by restriction analysis or colony PCR to verify the presence of the 546-bp insert.
11. Prepare overnight cultures of two or three colonies containing the 546-bp insert at 37°C for approximately 16 h.
12. Verify the IBS1, IBS2, EBS1, and EBS2 mutations by DNA sequencing using the universal T7 promoter sequencing primer.
13. Transform the retargeted plasmid into the desired knockout strain using the appropriate strain-specific transformation protocol. Since the intron is expressed under the control of the T7 promoter system, the knockout strain should be a DE3 strain of *E. coli*. If the knockout is to be conducted in a non-DE3 strain, the cells must be supplied with a source of T7 RNA polymerase. Alternatively, gene knockouts can be conducted in DE3 strains and then transduced to non-DE3 strains using P1-phage transduction.
14. Revive electroporated cells in 1 ml warm SOC medium (a nutrient-rich broth used to increase transformation efficiency) for 1 to 2 h at 37°C.
15. Dilute the transformation reaction into 4 ml warm LB medium containing 25 μg/ml chloramphenicol and grow for 16 h at 37°C.
16. Inoculate 5 ml of fresh LB medium containing 25 μg/ml chloramphenicol with 50 μl of the overnight culture and grow to an optical density at 600 nm of 0.2 to 0.3.
17. To induce intron transcription, inoculate 5 ml of fresh LB medium containing 100 μM IPTG and grow for 3 h at 37°C.
18. To remove excess IPTG, pellet the culture by centrifugation and resuspend in 1 ml fresh LB medium without antibiotic or IPTG.
19. Dilute a portion of the culture 1,000-fold in LB medium and spread 50 to 100 μl onto LB agar without antibiotic.
20. Perform colony PCR on colonies using primers flanking the intron insertion site or one flanking primer and one intron-specific primer. The intron insertion is ~0.9 kb.

There are also efforts to reduce the *E. coli* genome, thereby removing nonessential regions without affecting the growth behavior (52, 81). With a cleaner and more efficient background, such reduced-genome *E. coli* strains (e.g., Clean Genome® *E. coli* from Scarab Genomics, LLC [Madison, WI; www.scarabgenomics.com]) have been demonstrated to be effective in the production of recombinant proteins, DNAs, and metabolites (13, 60, 89).

19.3. CHALLENGES FOR STRAIN IMPROVEMENT

Proteins having a complex structure or multiple domains or requiring complex folding, multiple disulfide bonds, or other special posttranslational processing steps are difficult to express functionally in *E. coli*, and these technical difficulties still present a great challenge for strain improvement. Once the target protein is determined to be suitable for expression in *E. coli*, strain improvement is aimed at the host/vector system, target protein sequence, and various expression variables (Fig. 1). The general guidelines for strain improvement are (i) to ensure the genetic stability of the host/vector system, (ii) to maximize the synthesis fluxes for all the gene expression steps (i.e., transcription, translation, and posttranslational processing steps), (iii) to ensure the flux balance of these protein synthesis steps, (iv) to stabilize all the expression intermediates and final products, and (v) to minimize the physiological impact associated with high-level gene expression and high-cell-density cultivation. Successful strain improvement often relies on identification of the key steps limiting the overall culture performance. However, these limiting steps tend to be specific to the product of interest, difficult to identify, and subject to change. While certain types of limitations are now relatively easy to overcome, achieving a precise balance of the protein synthesis fluxes through optimization of the host/vector system remains challenging. In other words, we still do not have sufficient knowledge to undertake complete genome restructuring to achieve optimum strain performance in a given situation. Modern approaches based on systems biology and "-omics" studies continue to improve our knowledge, however. The genetic modifications currently adopted in most practical applications are primarily limited to gene knockouts and extrachromosomal overexpression, which cannot be fine-tuned. Modern chromosomal engineering techniques will potentially allow the inclusion of multiple exogenous genes as well as more precise regulation of the expression level of the modified genes in the chromosome. As a result, innovative strains can be more flexibly, effectively, and optimally tailored to enhance recombinant protein production. Despite the power of *E. coli* as an expression system, much work remains to be done to expand the types of protein products that can be functionally expressed and to improve its performance relative to the numerous alternative expression systems.

We greatly appreciate critical comments from George N. Bennett, Chris Reeves, and Richard Baltz.

REFERENCES

1. **Aldor, I. S., D. C. Krawitz, W. Forrest, C. Chen, J. C. Nishihara, J. C. Joly, and K. M. Champion.** 2005. Proteomic profiling of recombinant *Escherichia coli* in high-cell-density fermentations for improved production of an antibody fragment biopharmaceutical. *Appl. Environ. Microbiol.* **71:**1717–1728.
2. **Alexeyev, M. F., and I. N. Shokolenko.** 1995. Mini-Tn*10* transposon derivatives for insertion mutagenesis and gene delivery into the chromosome of gram-negative bacteria. *Gene* **160:**59–62.
3. **Baba, T., T. Ara, M. Hasegawa, Y. Takai, Y. Okumura, M. Baba, K. A. Datsenko, M. Tomita, B. L. Wanner, and H. Mori.** 2006. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* **2:**2006.0008.
4. **Baneyx, F., and G. Georgiou.** 1991. Construction and characterization of *Escherichia coli* strains deficient in multiple secreted proteases: protease III degrades high-molecular-weight substrates in vivo. *J. Bacteriol.* **173:**2696–2703.
5. **Baneyx, F., and M. Mujacic.** 2004. Recombinant protein folding and misfolding in *Escherichia coli*. *Nat. Biotechnol.* **22:**1399–1408.

6. **Bell, C. E.** 2005. Structure and mechanism of *Escherichia coli* RecA ATPase. *Mol. Microbiol.* **58:**358–366.
7. **Benito, A., M. Vidal, and A. Villaverde.** 1993. Enhanced production of P_L-controlled recombinant proteins and plasmid stability in *Escherichia coli* $RecA^+$ strains. *J. Biotechnol.* **29:**299–306.
8. **Bentley, W. E., N. Mirjalili, D. C. Andersen, R. H. Davis, and D. S. Kompala.** 2009. Plasmid-encoded protein: the principal factor in the "metabolic burden" associated with recombinant bacteria. *Biotechnol. Bioeng.* **102:**1284–1297.
9. **Bergmann, S., D. Wild, O. Diekmann, R. Frank, D. Bracht, G. S. Chhatwal, and S. Hammerschmidt.** 2003. Identification of a novel plasmin(ogen)-binding motif in surface displayed α-enolase of *Streptococcus pneumoniae*. *Mol. Microbiol.* **49:**411–423.
10. **Binet, R., S. Letoffe, J. M. Ghigo, P. Delepelaire, and C. Wandersman.** 1997. Protein secretion by Gram-negative bacterial ABC exporters: a review. *Gene* **192:**7–11.
11. **Boer, H. A., L. J. Comstock, and M. Vasser.** 1983. The *tac* promoter: a functional hybrid derived from the *trp* and *lac* promoters. *Proc. Natl. Acad. Sci. USA* **80:**21–25.
12. **Cashel, M., and K. E. Rudd.** 1987. The stringent response, p. 1410–1438. *In* F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), Escherichia coli *and* Salmonella typhimurium: *Cellular and Molecular Biology*, vol. 2. American Society for Microbiology, Washington, DC.
13. **Chakiath, C. S., and D. Esposito.** 2007. Improved recombinational stability of lentiviral expression vectors using reduced-genome *Escherichia coli*. *BioTechniques* **43:**466, 468, 470.
14. **Chao, Y. P., C. J. Chiang, and W. B. Hung.** 2002. Stringent regulation and high-level expression of heterologous genes in *Escherichia coli* using T7 system controllable by the araBAD promoter. *Biotechnol. Prog.* **18:**394–400.
15. **Chen, J. Q., T. B. Acton, S. K. Basu, G. T. Montelione, and M. Inouye.** 2002. Enhancement of the solubility of proteins overexpressed in *Escherichia coli* by heat shock. *J. Mol. Microbiol. Biotechnol.* **4:**519–524.
16. **Cherepanov, P. P., and W. Wackernagel.** 1995. Gene disruption in *Escherichia coli*: Tc^R and Km^R cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant. *Gene* **158:**9–14.
17. **Cho, S. H., D. Shin, G. E. Ji, S. Heu, and S. Ryu.** 2005. High-level recombinant protein production by overexpression of Mlc in *Escherichia coli*. *J. Biotechnol.* **119:**197–203.
18. **Choi, J. H., and S. Y. Lee.** 2004. Secretory and extracellular production of recombinant proteins using *Escherichia coli*. *Appl. Microbiol. Biotechnol.* **64:**625–635.
19. **Chou, C.-H., G. N. Bennett, and K.-Y. San.** 1996. Genetic manipulation of stationary-phase genes to enhance recombinant protein production in *Escherichia coli*. *Biotechnol. Bioeng.* **50:**636–642.
20. **Chou, C. P.** 2007. Engineering cell physiology to enhance recombinant protein production in *Escherichia coli*. *Appl. Microbiol. Biotechnol.* **76:**521–532.
21. **Datsenko, K. A., and B. L. Wanner.** 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. USA* **97:**6640–6645.
22. **Davis, S. J., and R. D. Vierstra.** 1998. Soluble, highly fluorescent variants of green fluorescent protein (GFP) for use in higher plants. *Plant Mol. Biol.* **36:**521–528.
23. **De Anda, R., A. R. Lara, V. Hernandez, V. Hernandez-Montalvo, G. Gosset, F. Bolivar, and O. T. Ramirez.** 2006. Replacement of the glucose phosphotransferase transport system by galactose permease reduces acetate accumulation and improves process performance of *Escherichia coli* for recombinant protein production without impairment of growth rate. *Metab. Eng.* **8:**281–290.
24. **de Marco, A.** 2009. Strategies for successful recombinant expression of disulfide bond-dependent proteins in *Escherichia coli*. *Microb. Cell Fact.* **8:**26.
25. **De Mey, M., J. Maertens, G. J. Lequeux, W. K. Soetaert, and E. J. Vandamme.** 2007. Construction and model-based analysis of a promoter library for *E. coli*: an indispensable tool for metabolic engineering. *BMC Biotechnol.* **7:**34.
26. **de Smit, M. H., and J. van Duin.** 1990. Secondary structure of the ribosome binding site determines translational efficiency: a quantitative analysis. *Proc. Natl. Acad. Sci. USA* **87:**7668–7672.
27. **Diaz-Acosta, A., M. L. Sandoval, L. Delgado-Olivares, and J. Membrillo-Hernandez.** 2006. Effect of anaerobic and stationary phase growth conditions on the heat shock and oxidative stress responses in *Escherichia coli* K-12. *Arch. Microbiol.* **185:**429–438.
28. **Feldman, M. F., M. Wacker, M. Hernandez, P. G. Hitchen, C. L. Marolda, M. Kowarik, H. R. Morris, A. Dell, M. A. Valvano, and M. Aebi.** 2005. Engineering N-linked protein glycosylation with diverse O antigen lipopolysaccharide structures in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **102:**3016–3021.
29. **Fernandez, L. A., I. Sola, L. Enjuanes, and V. de Lorenzo.** 2000. Specific secretion of active single-chain Fv antibodies into the supernatants of *Escherichia coli* cultures by use of the hemolysin system. *Appl. Environ. Microbiol.* **66:**5024–5029.
30. **Flores, S., R. de Anda-Herrera, G. Gosset, and F. G. Bolivar.** 2004. Growth-rate recovery of *Escherichia coli* cultures carrying a multicopy plasmid, by engineering of the pentose-phosphate pathway. *Biotechnol. Bioeng.* **87:**485–494.
31. **Franchini, A. G., and T. Egli.** 2006. Global gene expression in *Escherichia coli* K-12 during short-term and long-term adaptation to glucose-limited continuous culture conditions. *Microbiology* **152**(Pt. 7)**:**2111–2127.
32. **Georgiou, G., and L. Segatori.** 2005. Preparative expression of secreted proteins in bacteria: status report and future prospects. *Curr. Opin. Biotechnol.* **16:**538–545.
33. **Goryshin, I. Y., J. Jendrisak, L. M. Hoffman, R. Meis, and W. S. Reznikoff.** 2000. Insertional transposon mutagenesis by electroporation of released Tn5 transposition complexes. *Nat. Biotechnol.* **18:**97–100.
34. **Gosset, G.** 2005. Improvement of *Escherichia coli* production strains by modification of the phosphoenolpyruvate: sugar phosphotransferase system. *Microb. Cell Fact.* **4:**14.
35. **Grabherr, R., and K. Bayer.** 2002. Impact of targeted vector design on ColE1 plasmid replication. *Trends Biotechnol.* **20:**257–260.
36. **Gumpert, J., and C. Hoischen.** 1998. Use of cell wall-less bacteria (L-forms) for efficient expression and secretion of heterologous gene products. *Curr. Opin. Biotechnol.* **9:**506–509.
37. **Guzman, L.-M., D. Belin, M. J. Carson, and J. Beckwith.** 1995. Tight regulation, modulation, and high-level expression by vectors containing the arabinose P_{BAD} promoter. *J. Bacteriol.* **177:**4121–4130.
38. **Haddadin, F. T., and S. W. Harcum.** 2005. Transcriptome profiles for high-cell-density recombinant and wild-type *Escherichia coli*. *Biotechnol. Bioeng.* **90:**127–153.
39. **Hagg, P., J. W. de Pohl, F. Abdulkarim, and L. A. Isaksson.** 2004. A host/plasmid system that is not dependent on antibiotics and antibiotic resistance genes for stable plasmid maintenance in *Escherichia coli*. *J. Biotechnol.* **111:**17–30.
40. **Harcum, S. W., and W. E. Bentley.** 1993. Response dynamics of 26-, 34-, 39-, 54-, and 80-kDa proteases in induced cultures of recombinant *Escherichia coli*. *Biotechnol. Bioeng.* **42:**675–685.
41. **Heo, M. A., S. H. Kim, S. Y. Kim, Y. J. Kim, J. H. Chung, M. K. Oh, and S. G. Lee.** 2006. Functional expression of single-chain variable fragment antibody against

c-Met in the cytoplasm of *Escherichia coli*. *Protein Expr. Purif.* **47:**203–209.
42. **Hoffmann, F., and U. Rinas.** 2004. Roles of heat-shock chaperones in the production of recombinant proteins in *Escherichia coli*. *Adv. Biochem. Eng. Biotechnol.* **89:**143–161.
43. **Hoffmann, F., J. van den Heuvel, N. Zidek, and U. Rinas.** 2004. Minimizing inclusion body formation during recombinant protein production in *Escherichia coli* at bench and pilot plant scale. *Enzyme Microb. Technol.* **34:**235–241.
44. **Hoffmann, F., J. Weber, and U. Rinas.** 2002. Metabolic adaptation of *Escherichia coli* during temperature-induced recombinant protein production: 1. Readjustment of metabolic enzyme synthesis. *Biotechnol. Bioeng.* **80:**313–319.
45. **Ihssen, J., and T. Egli.** 2004. Specific growth rate and not cell density controls the general stress response in *Escherichia coli*. *Microbiology* **150**(Pt. 6)**:**1637–1648.
46. **Jeong, K. J., J. H. Choi, W. M. Yoo, K. C. Keum, N. C. Yoo, S. Y. Lee, and M. H. Sung.** 2004. Constitutive production of human leptin by fed-batch culture of recombinant *rpoS*$^-$ *Escherichia coli*. *Protein Expr. Purif.* **36:**150–156.
47. **Kadokura, H., H. Kawasaki, K. Yoda, M. Yamasaki, and K. Kitamoto.** 2001. Efficient export of alkaline phosphatase overexpressed from a multicopy plasmid requires *degP*, a gene encoding a periplasmic protease of *Escherichia coli*. *J. Gen. Appl. Microbiol.* **47:**133–141.
48. **Kemmer, C., and P. Neubauer.** 2006. Antisense RNA based down-regulation of RNaseE in *E. coli*. *Microb. Cell Fact.* **5:**38.
49. **Kim, J. Y. H., and H. J. Cha.** 2003. Down-regulation of acetate pathway through antisense strategy in *Escherichia coli*: improved foreign protein production. *Biotechnol. Bioeng.* **83:**841–853.
50. **Kleber-Janke, T., and W.-M. Becker.** 2000. Use of modified BL21(DE3) *Escherichia coli* cells for high-level expression of recombinant peanut allergens affected by poor codon usage. *Protein Expr. Purif.* **19:**419–424.
51. **Kolaj, O., S. Spada, S. Robin, and J. G. Wall.** 2009. Use of folding modulators to improve heterologous protein production in *Escherichia coli*. *Microb. Cell Fact.* **8:**9.
52. **Kolisnychenko, V., G. Plunkett, C. D. Herring, T. Feher, J. Posfai, F. R. Blattner, and G. Posfai.** 2002. Engineering a reduced *Escherichia coli* genome. *Genome Res.* **12:**640–647.
53. **Kudla, G., A. W. Murray, D. Tollervey, and J. B. Plotkin.** 2009. Coding-sequence determinants of gene expression in *Escherichia coli*. *Science* **324:**255–258.
54. **Kurland, C. G., and H. J. Dong.** 1996. Bacterial growth inhibition by overproduction of protein. *Mol. Microbiol.* **21:**1–4.
55. **Kurokawa, Y., H. Yanagi, and T. Yura.** 2000. Overexpression of protein disulfide isomerase DsbC stabilizes multiple-disulfide-bonded recombinant protein produced and transported to the periplasm in *Escherichia coli*. *Appl. Environ. Microbiol.* **66:**3960–3965.
56. **Kyratsous, C. A., S. J. Silverstein, C. R. DeLong, and C. A. Panagiotidis.** 2009. Chaperone-fusion expression plasmid vectors for improved solubility of recombinant proteins in *Escherichia coli*. *Gene* **440:**9–15.
57. **Langdon, R. H., J. Cuccui, and B. W. Wren.** 2009. *N*-linked glycosylation in bacteria: an unexpected application. *Future Microbiol.* **4:**401–412.
58. **Lee, E.-C., D. Yu, J. Martinez de Velasco, L. Tessarollo, D. A. Swing, D. L. Court, N. A. Jenkins, and N. G. Copeland.** 2001. A highly efficient *Escherichia coli*-based chromosome engineering system adapted for recombinogenic targeting and subcloning of BAC DNA. *Genomics* **73:**56–65.
59. Reference deleted.
60. **Lee, J. H., B. H. Sung, M. S. Kim, F. R. Blattner, B. H. Yoon, J. H. Kim, and S. C. Kim.** 2009. Metabolic engineering of a reduced-genome strain of *Escherichia coli* for L-threonine production. *Microb. Cell Fact.* **8:**2.
61. **Lee, S. Y.** 1996. High cell density culture of *Escherichia coli*. *Trends Biotechnol.* **14:**98–105.
62. **Lee, S. Y., J. H. Choi, and Z. Xu.** 2003. Microbial cell-surface display. *Trends Biotechnol.* **21:**45–52.
63. **LeThanh, H., P. Neubauer, and F. Hoffmann.** 2005. The small heat-shock proteins IbpA and IbpB reduce the stress load of recombinant *Escherichia coli* and delay degradation of inclusion bodies. *Microb. Cell Fact.* **4:**6.
64. **Lopez, P. J., I. Marchand, S. A. Joyce, and M. Dreyfus.** 1999. The C-terminal half of RNase E, which organizes the *Escherichia coli* degradosome, participates in mRNA degradation but not rRNA processing in vivo. *Mol. Microbiol.* **33:**188–199.
65. **Lynch, M. D., T. Warnecke, and R. T. Gill.** 2007. SCALEs: multiscale analysis of library enrichment. *Nat. Methods* **4:**87–93.
66. **Malissard, M., and E. G. Berger.** 2001. Improving the solubility of the catalytic domain of human β-1,4-galactosyltransferase 1 through rationally designed amino-acid replacements. *Eur. J. Biochem.* **268:**4352–4358.
67. **March, J. C., M. A. Eiteman, and E. Altman.** 2002. Expression of an anaplerotic enzyme, pyruvate carboxylase, improves recombinant protein production in *Escherichia coli*. *Appl. Environ. Microbiol.* **68:**5620–5624.
68. **Matsuura, M., R. Saldanha, H. W. Ma, H. Wank, J. Yang, G. Mohr, S. Cavanagh, G. M. Dunny, M. Belfort, and A. M. Lambowitz.** 1997. A bacterial group II intron encoding reverse transcriptase, maturase, and DNA endonuclease activities: biochemical demonstration of maturase activity and insertion of new genetic information within the intron. *Genes Dev.* **11:**2910–2924.
69. **Meerman, H. J., and G. Georgoiu.** 1994. Construction and characterization of a set of *E. coli* strains deficient in all known loci affecting the proteolytic stability of secreted recombinant proteins. *Bio/Technology* **12:**1107–1110.
70. **Miller, J. H.** 1992. *A Short Course in Bacterial Genetics*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
71. **Miroux, B., and J. E. Walker.** 1996. Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J. Mol. Biol.* **260:**289–298.
72. **Mogensen, J. E., and D. E. Otzen.** 2005. Interactions between folding factors and bacterial outer membrane proteins. *Mol. Microbiol.* **57:**326–346.
73. **Mori, H., K. Isono, T. Horiuchi, and T. Miki.** 2000. Functional genomics of *Escherichia coli* in Japan. *Res. Microbiol.* **151:**121–128.
74. **Nakamura, Y., T. Gojobori, and T. Ikemura.** 2000. Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic Acids Res.* **28:**292.
75. **Narayanan, N., and C. P. Chou.** 2009. Alleviation of proteolytic sensitivity to enhance recombinant lipase production in *Escherichia coli*. *Appl. Environ. Microbiol.* **75:**5424–5427.
76. **Neubauer, P., H. Y. Lin, and B. Mathiszik.** 2003. Metabolic load of recombinant protein production: inhibition of cellular capacities for glucose uptake and respiration after induction of a heterologous gene in *Escherichia coli*. *Biotechnol. Bioeng.* **83:**53–64.
77. **Pan, K.-L., H.-C. Hsiao, C.-L. Weng, M.-S. Wu, and C. P. Chou.** 2003. Roles of DegP in prevention of protein misfolding in the periplasm upon overexpression of penicillin acylase in *Escherichia coli*. *J. Bacteriol.* **185:**3020–3030.
78. **Patnaik, R.** 2008. Engineering complex phenotypes in industrial strains. *Biotechnol. Prog.* **24:**38–47.
79. **Phue, J.-N., S. J. Lee, L. Trinh, and J. Shiloach.** 2008. Modified *Escherichia coli* B (BL21), a superior producer of plasmid DNA compared with *Escherichia coli* K (DH5). *Biotechnol. Bioeng.* **101:**831–836.

80. **Phue, J. N., S. B. Noronha, R. Hattacharyya, A. J. Wolfe, and J. Shiloach.** 2005. Glucose metabolism at high density growth of *E. coli* B and *E. coli* K: differences in metabolic pathways are responsible for efficient glucose utilization in *E. coli* B as determined by microarrays and Northern blot analyses. *Biotechnol. Bioeng.* **90:**805–820.
81. **Posfai, G., G. Plunkett, T. Feher, D. Frisch, G. M. Keil, K. Umenhoffer, V. Kolisnychenko, B. Stahl, S. S. Sharma, M. de Arruda, V. Burland, S. W. Harcum, and F. R. Blattner.** 2006. Emergent properties of reduced-genome *Escherichia coli*. *Science* **312:**1044–1046.
82. **Rabhi-Essafi, I., A. Sadok, N. Khalaf, and D. M. Fathallah.** 2007. A strategy for high-level expression of soluble and functional human interferon α as a GST-fusion protein in *E. coli*. *Protein Eng. Des. Sel.* **20:**201–209.
83. **Raman, B., M. P. Nandakumar, V. Muthuvijayan, and M. R. Marten.** 2005. Proteome analysis to assess physiological changes in *Escherichia coli* grown under glucose-limited fed-batch conditions. *Biotechnol. Bioeng.* **92:**384–392.
84. **Rich, J. R., and S. G. Withers.** 2009. Emerging methods for the production of homogeneous human glycoproteins. *Nat. Chem. Biol.* **5:**206–215.
85. **Rong, R., M. M. Slupska, J.-H. Chiang, and J. H. Miller.** 2004. Engineering large fragment insertions into the chromosome of *Escherichia coli*. *Gene* **336:**73–80.
86. **San, K. Y., G. N. Bennett, S. J. Berrios-Rivera, R. V. Vadali, Y. T. Yang, E. Horton, F. B. Rudolph, B. Sariyar, and K. Blackwood.** 2002. Metabolic engineering through cofactor manipulation and its effects on metabolic flux redistribution in *Escherichia coli*. *Metab. Eng.* **4:**182–192.
87. **Sandee, D., S. Tungpradabkul, Y. Kurokawa, K. Fukui, and M. Takagi.** 2005. Combination of Dsb coexpression and an addition of sorbitol markedly enhanced soluble expression of single-chain Fv in *Escherichia coli*. *Biotechnol. Bioeng.* **91:**418–424.
88. **Santos, C. N., and G. Stephanopoulos.** 2008. Combinatorial engineering of microbes for optimizing cellular phenotype. *Curr. Opin. Chem. Biol.* **12:**168–176.
89. **Sharma, S. S., F. R. Blattner, and S. W. Harcum.** 2007. Recombinant protein production in an *Escherichia coli* reduced genome strain. *Metab. Eng.* **9:**133–141.
90. **Shokri, A., A. M. Sande'n, and G. Larsson.** 2003. Cell and process design for targeting of recombinant protein into the culture medium of *Escherichia coli*. *Appl. Microbiol. Biotechnol.* **60:**654–664.
91. **Skerra, A.** 1994. Use of the tetracycline promoter for the tightly regulated production of a murine antibody fragment in *Escherichia coli*. *Gene* **151:**131–135.
92. **Sorensen, H. P., and K. K. Mortensen.** 2005. Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. *J. Biotechnol.* **115:**113–128.
93. **Stewart, E. J., F. Aslund, and J. Beckwith.** 1998. Disulfide bond formation in the *Escherichia coli* cytoplasm: an in vivo role reversal for the thioredoxins. *EMBO J.* **17:**5543–5550.
94. **Studier, F. W.** 1991. Use of bacteriophage-T7 lysozyme to improve an inducible T7 expression system. *J. Mol. Biol.* **219:**37–44.
95. **Studier, F. W., and B. A. Moffatt.** 1986. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* **189:**113–130.
96. **Studier, F. W., A. H. Rosenberg, J. J. Dunn, and J. W. Dubendorff.** 1990. Use of T7 RNA polymerase to direct the expression of cloned genes. *Methods Enzymol.* **185:**60–89.
97. **Terpe, K.** 2003. Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Appl. Microbiol. Biotechnol.* **60:**523–533.
98. **Thomas, N. A., and B. B. Finlay.** 2003. Establishing order for type III secretion substrates—a hierarchical process. *Trends Microbiol.* **11:**398–403.
99. **Togna, A. P., M. L. Shuler, and D. B. Wilson.** 1993. Effects of plasmid copy number and runaway plasmid replication on overproduction and excretion of β-lactamase from *Escherichia coli*. *Biotechnol. Prog.* **9:**31–39.
100. **Tolia, N. H., and L. Joshua-Tor.** 2006. Strategies for protein coexpression in *Escherichia coli*. *Nat. Methods* **3:**55–64.
101. **Tzschaschel, B. D., C. A. Guzman, K. N. Timmis, and V. de Lorenzo.** 1996. An *Escherichia coli* hemolysin transport system-based vector for the export of polypeptides: export of Shiga-like toxin IIeB subunit by *Salmonella typhimurium aroA*. *Nat. Biotechnol.* **14:**765–769.
102. **van der Wal, F. J., C. M. ten Hagen-Jongman, B. Oudega, and J. Luirink.** 1995. Optimization of bacteriocin-release-protein-induced protein release by *Escherichia coli*: extracellular production of the periplasmic molecular chaperone FaeE. *Appl. Microbiol. Biotechnol.* **44:**459–465.
103. **Vemuri, G. N., M. A. Eiteman, and E. Altman.** 2006. Increased recombinant protein production in *Escherichia coli* strains with overexpressed water-forming NADH oxidase and a deleted ArcA regulatory protein. *Biotechnol. Bioeng.* **94:**538–542.
104. **Vind, J., M. A. Sorensen, M. D. Rasmussen, and S. Pedersen.** 1993. Synthesis of proteins in *Escherichia coli* is limited by the concentration of free ribosomes: expression from reporter genes does not always reflect functional mRNA levels. *J. Mol. Biol.* **231:**678–688.
105. **Wagner, S., M. M. Klepsch, S. Schlegel, A. Appel, R. Draheim, M. Tarry, M. Hagbom, K. J. van Wijk, D. J. Slotboom, J. O. Persson, and J.-W. de Gier.** 2008. Tuning *Escherichia coli* for membrane protein overexpression. *Proc. Natl. Acad. Sci. USA* **105:**14371–14376.
106. **Wan, E. W., and F. Baneyx.** 1998. TolAIII co-overexpression facilitates the recovery of periplasmic recombinant proteins into the growth medium of *Escherichia coli*. *Protein Expr. Purif.* **14:**13–22.
107. **Waugh, D. S.** 2005. Making the most of affinity tags. *Trends Biotechnol.* **23:**316–320.
108. **Welch, M., A. Villalobos, C. Gustafsson, and J. Minshull.** 2009. You're one in a googol: optimizing genes for protein expression. *J. R. Soc. Interface* **6**(Suppl. 4)**:** S467–S476.
109. **Wood, T. K., and S. W. Peretti.** 1991. Construction of a specialized-ribosome vector for cloned-gene expression in *E. coli*. *Biotechnol. Bioeng.* **38:**891–906.
110. **Wu, M. S., K. L. Pan, and C. P. Chou.** 2007. Effect of heat-shock proteins for relieving physiological stress and enhancing the production of penicillin acylase in *Escherichia coli*. *Biotechnol. Bioeng.* **96:**956–966.
111. **Yoon, S. H., M. J. Han, S. Y. Lee, K. J. Jeong, and J. S. Yoo.** 2003. Combined transcriptome and proteome analysis of *Escherichia coli* during high cell density culture. *Biotechnol. Bioeng.* **81:**753–767.
112. **Yu, B. J., K. H. Kang, J. H. Lee, B. H. Sung, M. S. Kim, and S. C. Kim.** 2008. Rapid and efficient construction of markerless deletions in the *Escherichia coli* genome. *Nucleic Acids Res.* **36:**e84.
113. **Yu, D., H. M. Ellis, E.-C. Lee, N. A. Jenkins, N. G. Copeland, and D. L. Court.** 2000. An efficient recombination system for chromosome engineering in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **97:**5978–5983.
114. **Zhang, Y. M., F. Buchholz, J. P. Muyrers, and A. F. Stewart.** 1998. A new logic for DNA engineering using recombination in *Escherichia coli*. *Nat. Genet.* **20:**123–128.
115. **Zhuang, F. L., M. Karberg, J. Perutka, and A. M. Lambowitz.** 2009. EcI5, a group IIB intron with high retrohoming frequency: DNA target site recognition and use in gene targeting. *RNA* **15:**432–449.