Strategies to Optimize Protein Expression in *E. coli*

Dana M. Francis[1] and Rebecca Page[1]

[1]Brown University, Providence, Rhode Island

ABSTRACT

Recombinant protein expression in *Escherichia coli* (*E. coli*) is simple, fast, inexpensive, and robust, with the expressed protein comprising up to 50 percent of the total cellular protein. However, it also has disadvantages. For example, the rapidity of bacterial protein expression often results in unfolded/misfolded proteins, especially for heterologous proteins that require longer times and/or molecular chaperones to fold correctly. In addition, the highly reductive environment of the bacterial cytosol and the inability of *E. coli* to perform several eukaryotic post-translational modifications results in the insoluble expression of proteins that require these modifications for folding and activity. Fortunately, multiple, novel reagents and techniques have been developed that allow for the efficient, soluble production of a diverse range of heterologous proteins in *E. coli*. This overview describes variables at each stage of a protein expression experiment that can influence solubility and offers a summary of strategies used to optimize soluble expression in *E. coli*. *Curr. Protoc. Protein Sci.* 61:5.24.1-5.24.29. © 2010 by John Wiley & Sons, Inc.

Keywords: protein expression • *E. coli* • fusion proteins • proteases • heterologous protein • purification tags • expression tags • expression strains and vectors • folded protein • active protein

INTRODUCTION

Recombinant protein expression has revolutionized all aspects of the biological sciences. Most significantly, it has dramatically expanded the number of proteins that can be investigated both biochemically and structurally. Previously, protein production was the domain of experts, as purification from a natural source (i.e., plants, rabbits, bovine) was often difficult and time consuming. However, the availability of new commercial systems for recombinant protein expression, combined with advanced protein purification techniques, has made protein production prevalent throughout the biological and biomedical sciences. This has enabled the research community to study thousands of low abundance and novel proteins from a large variety of organisms. Notably, 31 recombinant proteins were approved for therapeutic use between 2003 and 2006, highlighting the importance of heterologous protein expression in biopharmaceutical research (Walsh, 2006). As the number of recombinantly produced proteins increases, so too does an appreciation for the difficulties and limitations inherent to this process.

In spite of the development of multiple nonbacterial recombinant expression systems over the last three decades (yeast, baculovirus, mammalian cell, cell free systems; see Table 5.24.1), *Escherichia coli* is still the preferred host for recombinant protein expression (Yin et al., 2007). The rationale is clear: *E. coli* is easy to genetically manipulate, it is inexpensive to culture, and expression is fast, with proteins routinely produced in one day. Moreover, protocols for isotope-labeling for NMR spectroscopy and selenomethionine incorporation for X-ray crystallography are well established, making it highly suitable for structural studies. Thus, *E. coli* has multiple, significant benefits over other expression systems including cost, ease-of-use, and scale.

Despite its many advantages and widespread use, there are also disadvantages to using *E. coli* as an expression host. In contrast to eukaryotic systems, transcription and translation are fast and tightly coupled. Since many eukaryotic proteins require longer times and/or the assistance of folding chaperones to fold into their native state, this rate enhancement often leads to a pool

Production of Recombinant Proteins

Current Protocols in Protein Science 5.24.1-5.24.29, August 2010
Published online August 2010 in Wiley Interscience (www.interscience.wiley.com).
DOI: 10.1002/0471140864.ps0524s61
Copyright © 2010 John Wiley & Sons, Inc.

5.24.1

Supplement 61

Optimizing
Protein
Expression in
E. coli

5.24.2

Supplement 61

Table 5.24.1 Comparison of Recombinant Protein Expression Systems

	E. coli	Yeast	Insect cells	Mammalian cells	Cell-free[a] (*E. coli*)	Cell-free[a] (wheat germ)
Average time of cell division	30 min	90 min	18 hr	24 hr	N/A	N/A
Cost of expression	Low	Low	High	High	High	High
Expression level	High	Low-High	Low-High	Low-Moderate	Low-High	Low-High
Success rate (% soluble)	40-60	50-70	50-70	80-95	Variable	Variable
Advantages	Simple, low cost, rapid, robust, high yield, easy labeling for structural studies	Simple, low cost	Post-translational modifications	Natural protein configuration, post-translational modifications	High yield, fast, flexible, disulfide-bonded and membrane proteins, easy labeling for structural studies	Fast, flexible, disulfide-bonded and membrane proteins, post-translational modifications
Disadvantages	No post-translational modifications, insoluble protein, production of disulfide-bonded and membrane proteins is difficult	Less post-translational modifications, production of membrane proteins is difficult	Slow, higher cost, production of membrane proteins is difficult	Slow, high cost, lower yield	High cost, less post-translational modifications, efficient production requires highly-specialized setup	High cost, lower yield than *E. coli* cell-free system, efficient production requires highly specialized setup

[a]Spirin, 2004; Langlais et al., 2007; for other references, see text.

of partially folded, unfolded, or misfolded, insoluble proteins (Oberg et al., 1994). Thus, some targets, especially larger multidomain and membrane proteins, either fail to express in *E. coli* or express insolubly as inclusion bodies. Moreover, insolubility is not just restricted to heterologous proteins, as many bacterial proteins also cannot be produced in soluble form when overexpressed in *E. coli* (Vincentelli et al., 2003). In addition, the reducing environment of the bacterial cytoplasm makes the efficient production of disulfide-containing proteins challenging (Stewart et al., 1998; Ritz and Beckwith, 2001). Finally, *E. coli* lacks the machinery required to perform certain eukaryotic post-translational modifications, such as glycosylation, which can be critical for the formation of folded, active protein (Zhang et al., 2004).

Considerable efforts have been made in recent years to maximize the efficient production of soluble recombinant proteins in bacteria. A remarkable number of novel reagents (new vectors, new host strains) and strategies (chaperone co-expression, low temperature induction) have been developed that allow many of these disadvantages to be readily and successfully overcome. It is this topic, how to optimize soluble protein expression in *E. coli*, that is the focus of this review. Section I outlines a typical expression protocol for the production of heterologous proteins in *E. coli*. While this protocol has proven highly successful for a broad range of targets, every protein has its own unique set of biophysical characteristics that often requires protocol changes in order to successfully express the target. Thus, in the second part of this review, critical parameters that are essential to consider when designing a protein expression strategy are highlighted (Fig. 5.24.1). These parameters have the common goal of maximizing the yield of soluble, active protein. Section II describes optimization of the target DNA, section III discusses modifications for the optimization of expression vectors, section IV details bacterial host strains that aid heterologous protein expression, section V outlines optimization of protein expression conditions, and section VI describes how to enhance soluble expression by coexpression with other proteins.

I. REPRESENTATIVE PROTOCOL FOR EXPRESSING PROTEINS IN BACTERIA

Before initiating a protein expression project, it is essential to determine the definitive use of the recombinantly produced protein. Must the protein be soluble? Does the final product need to be active? Is the native protein conformation important? In some instances, such as antibody production, the production of soluble protein is not necessary, as the protein sequence, rather than the correct 3-dimensional fold, is required for successful antibody production (Yan et al., 2007). For these cases, expression that causes the protein to be incorporated into inclusion bodies is suggested, as the recombinant protein, while misfolded or unfolded, is highly enriched and protected from proteases (Valax and Georgiou, 1993). However, most often, the objective of recombinant protein expression in *E. coli* is to produce a protein product that is soluble, folded, and active.

Expression in *E. coli* requires four elements: (1) the protein of interest, (2) a bacterial expression vector, (3) an expression cell line, and (4) the equipment/materials for bacterial cell culture (i.e., shaker/media). There are multiple parameters that can be varied when optimizing an expression protocol, from selecting a vector with the appropriate promoter, to choosing an appropriate induction temperature. With each selection affecting the solubility and activity of the protein product, this task can appear daunting. However, decades of work by individual protein chemists, coupled with the recent experiences of high-throughput structural genomics efforts, have resulted in the identification of a consensus protocol that allows a diverse set of proteins to be successfully expressed in *E. coli*. A flowchart of the protocol typically used by the authors to express a diverse set of proteins is shown in Figure 5.24.2 (Mustelin et al., 2005; Brown et al., 2008; Critton et al., 2008). More extensive protocols are also available (Peti and Page, 2007; Gräslund et al., 2008a).

The protocol outlined in Figure 5.24.2 consists of nine steps. First, the optimal residue boundaries of the desired protein product, the "target protein," are determined. The target gene is then subcloned into a bacterial expression vector that utilizes the T7 *lac*O promoter system (i.e., that used in the pET system) and contains both an N-terminal hexahistadine (his$_6$)-tag and a tobacco etch virus (TEV) protease site (Peti and Page, 2007). The T7 promoter system provides strong, robust expression, the his$_6$-tag facilitates purification, while the protease site allows the (his$_6$)-tag to be proteolytically removed from the purified target protein. In the third step, the expression vector is transformed into a derivative of the BL21 (DE3) strain, such as BL21
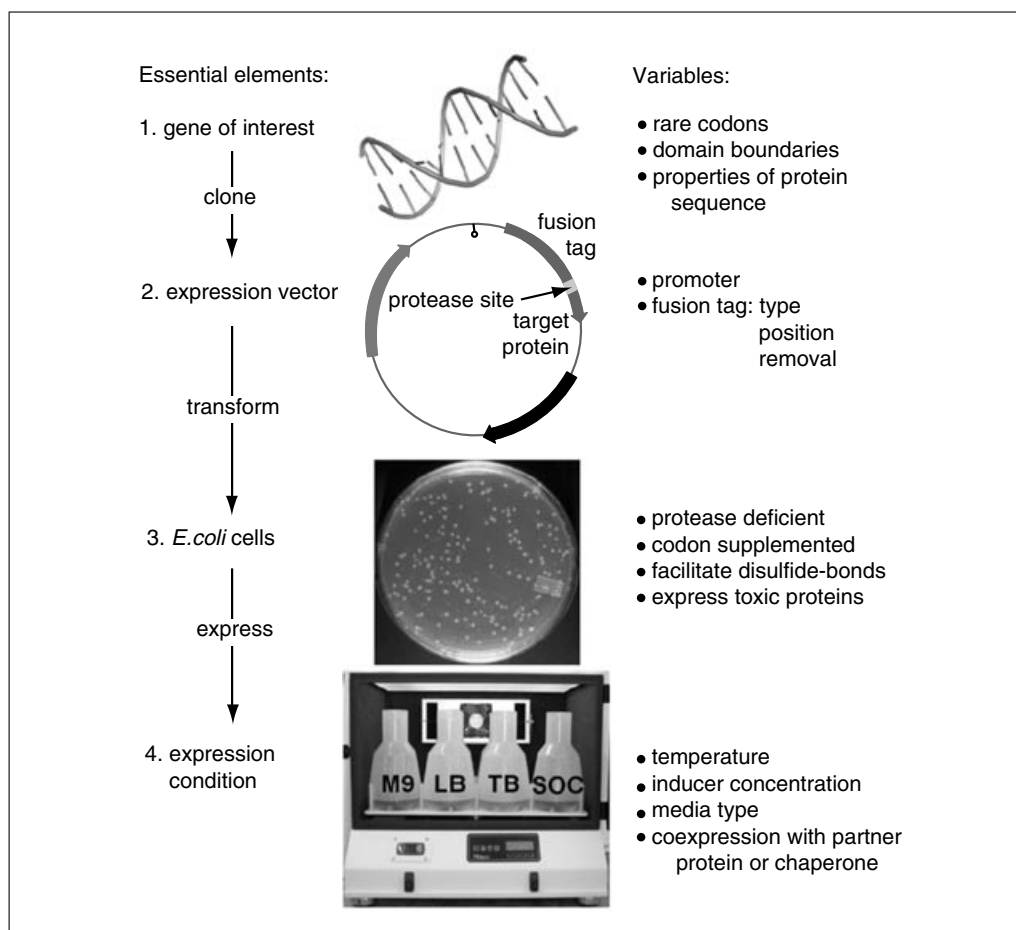
Figure 5.24.1 Schematic overview of the topics covered in this review, highlighting the multiple parameters (listed on the right) that can greatly impact the success of soluble expression.

(DE3)-RIL cells. These cells, which are compatible with the T7 promoter system, contain plasmids that encode arginine, isoleucine, and leucine tRNAs that are rare in *E. coli*, and are deficient in both *lon* and *ompT* proteases, which minimizes in vivo degradation of the target protein. After the overnight starter culture is used to inoculate the large-scale culture (step 5), the cells are grown to mid-log phase (OD_{600} of ~0.6 to 0.9) in Luria broth (LB) in baffled shaker flasks (which increases aeration and thus yield) at 37°C with constant shaking. The cultures are then transferred to a lower temperature (18°C), and, once cooled, protein expression is induced using isopropyl-β-thio-galactoside (IPTG). Expression is continued overnight with vigorous shaking (200 to 250 rpm) at 18°C. The lower expression temperature facilitates the production of folded, soluble protein. Finally, the cells are pelleted by centrifugation and stored at −80°C until needed.

This protocol is meant to serve as a starting point for designing an expression strategy in *E. coli*. However, because multiple char-

acteristics of the protein, vector, host strain, and/or expression conditions may need to be modified in order for folded, active protein to be produced, it is not uncommon to see more complicated protocols that test these variables in parallel (Peti and Page, 2007). Thus, the remainder of this review describes which factors have the largest effect on soluble protein expression and how to change them in order to express folded, active protein in *E. coli*.

II. PROPERTIES OF THE GENE AND PROTEIN THAT INFLUENCE EXPRESSION AND SOLUBILITY

Here, we discuss critical characteristics of the gene and/or protein sequence that influence its soluble expression in *E. coli*.

Rare codons

One of the most common reasons that heterologous proteins fail to express in *E. coli* is the presence of "rare" codons in the target mRNA. Many proteins, especially human proteins, have mRNA sequences that include codons that are infrequently used in
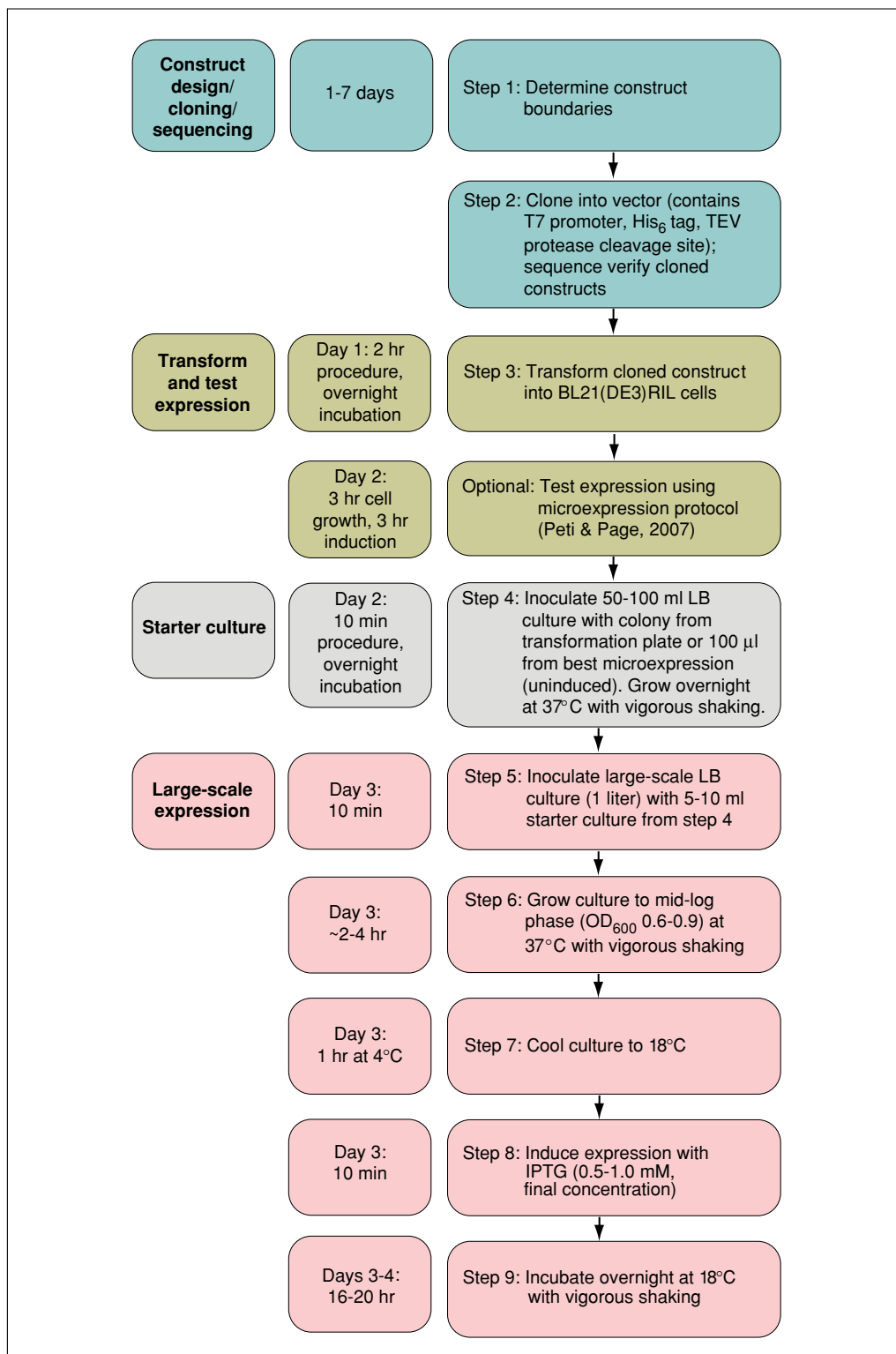
Figure 5.24.2 Flowchart of a general expression protocol used by the authors to express a broad range of targets, from phosphatases, to neuronal scaffolding proteins, to bacterial signaling proteins. The approximate time required to complete each segment of the protocol is listed to the left of the corresponding step.

E. coli (rare codons) (Sharp and Li, 1987). This includes codons for arginine (AGA, AGG), isoleucine (AUA), leucine (CUA), and proline (CCC). Target genes that contain significant numbers of individual rare codons, or smaller numbers of tandem rare codons, are more likely to experience translational stalling in *E. coli*, and thus often either completely fail to express, express at very low levels, or are expressed as truncated proteins (Kane, 1995, Cruz-Vera et al., 2004). Moreover, when they do express, these rare codon–rich genes can also be incorrectly translated, as high level misincorporation of lysine for arginine at AGA codons has been observed for protein targets expressed in *E. coli* (Calderone et al., 1996).

Fortunately, the codon bias of *E. coli* is straightforward to overcome. Multiple Web sites are now available that quantify the number and the location of rare codons in a gene (e.g., the rare codon calculator, RaCC; *http://nihserver.mbi.ucla.edu/RACC/*). These programs also often highlight the number of consecutive rare codons. If the target protein contains a significant number of rare codons, especially tandem rare codons, two approaches can be taken. In the first, changes are made to the gene. Specifically, the gene is "codon optimized," i.e., the rare codons are replaced with those that are common to the host. This can be achieved using site-directed mutagenesis. However, it is often faster and cheaper to simply have the codon-optimized gene synthesized (this service is offered by multiple companies). Gene synthesis has the added benefit that most gene optimization algorithms optimize not only rare codons, but also mRNA secondary structure, which has also been shown to affect translation efficiency (Hatfield and Roth, 2007). In the second approach, changes are made to the expression host. Namely, genes encoding the rare tRNAs are co-expressed with the wild-type (non-optimized) target gene (Wakagi et al., 1998). *E. coli* strains are now available that contain plasmids that encode rare tRNAs [i.e., BL21 (DE3)-RIL/RP/RILP cells from *Stratagene* or Rosetta cells from *Novagen*; Schenk et al., 1995; Tegel et al., 2009].

Both approaches effectively overcome codon bias. For example, codon optimization of the human β-defensin 2 (hBD2) gene led to a nine-fold enhancement in the expression level (Peng et al., 2004) while co-expression of inorganic phosphatase from *Sulfolobus* sp. with tRNA^Arg (AGA codon) more than doubled its already high expression level (Wakagi et al., 1998). Critically, recent large-scale, comparative studies have shown that, for most targets, either approach is equally effective (Burgess-Brown et al., 2008). Thus, many research groups, including those of structural genomics consortia, protein production facilities, and the authors (see the standard protocol in Fig. 5.24.2), use codon-supplemented cells for all initial expression trials.

Protein size and domain boundaries

The evolution of eukaryotes has been characterized by a significant increase in the size and complexity of proteins; e.g., the average protein length in *E. coli* is 317 residues while in humans it is 510 residues (Netzer and Hartl, 1997, Sakharkar et al., 2006). This increase in size is due to an increase in the number of complex, multidomain proteins, in which individual domains have distinct and independent functions. Comparative studies have shown that the probability of soluble expression in *E. coli* decreases with increasing molecular weight, especially for proteins >60 kD (Canaves et al., 2004, Goh et al., 2004, Gråslund et al., 2008a). For example, a large-scale study examining the protein properties of 95 recombinantly expressed mammalian proteins found that smaller proteins (average molecular weight of 22.8 kD) were often expressed solubly alone, while larger proteins (average molecular weight of 40.4 kD) were only solubly expressed when fused to solubility-enhancing tags (Dyson et al., 2004). In a separate study, small proteins (100 amino acids or less) were expressed solubly in *E. coli* at levels suitable for purification 47% of the time, while large proteins (600 to 800 amino acids) were expressed solubly only 33% of the time (Gråslund et al., 2008a). Thus, when using *E. coli* as an expression host, it is typically advantageous to express individual protein domains, as opposed to the full-length protein, whenever possible.

The starting and ending residues of the target domain can also greatly affect expression yield and solubility. For example, Klock et al. (2008) showed that deletion of just four residues at either the N- or C-terminus can convert a solubly expressing protein into one that expresses insolubly. In a separate study, Gråslund et al. (2008b) generated 10 constructs of a single target domain of interest: full-length and 9 deletion constructs that differed in length from one another by 2 to 10 residues at either the N- or C-terminus. Thus, all available functional and structural data should be used to determine optimal boundaries for a protein domain

construct. For a protein of unknown domain structure, threading the target protein sequence onto a homologous protein structure (i.e., SWISS-MODEL; Arnold et al., 2006) or using structure-based/fold recognition sequence alignment programs (i.e., FFAS; Jaroszewski et al., 2005) can aid in determining the optimal domain boundaries. When a homologous protein structure is not available, the prediction of secondary structural elements (i.e., PSIPRED; Jones, 1999) should be utilized. The disruption of predicted secondary structural elements must be avoided. Many of the bioinformatics tools needed to carry out these types of analyses are freely available at *http://www.expasy.org*. In addition, a program has been developed that integrates these bioinformatics tools to aid in protein boundary identification (the SGC Domain Boundary Analyzer; described in more detail in Gråslund et al., 2008b). Finally, it is typically advantageous to subclone four or more constructs with different N- and C-terminal boundaries and test, in parallel, which construct results in the highest level of soluble expression (Brown et al., 2008).

Protein sequence

Hydrophobic residues, low complexity regions. In addition to molecular weight, other biophysical properties of the protein, such as hydrophobicity and sequence complexity, can influence expression yields. In the study by Dyson et al. (2004), 95 mammalian proteins were fused to a variety of N- and C-terminal expression and purification tags in order to elucidate the properties of the proteins and fusion tags that facilitate soluble expression. They determined that contiguous hydrophobic residues (AILFWV) and low complexity regions (LCRs) negatively correlate with soluble expression. LCRs are regions of biased sequence composition, such as homopolymeric runs, short-period repeats, and overrepresentations of one or a few residues that typically adopt disordered coil conformations (DePristo et al., 2006). This has also been reported by other groups (Canaves et al., 2004; Gråslund et al., 2008a). Thus, it is common to design protein expression constructs to avoid hydrophobic residues and low complexity segments in the extreme N- and C-termini (Peti and Page, 2007). However, LCRs do not always inhibit soluble expression. Many intrinsically unstructured proteins (IUPs; proteins that do not adopt a single folded conformation yet are still biologically

active) contain LCRs, yet have been robustly and solubly expressed in *E. coli* (Dyson and Wright, 2005; Dancheck et al., 2008). Because LCRs may play an active role in mediating protein function or protein-protein interactions (Karlin et al., 2002), their inclusion in an expression protein construct must be determined on a protein-by-protein basis.

Disulfide bonds. The presence of disulfide bonds in a protein also negatively correlates with soluble expression in *E. coli*. The reducing environment of the bacterial cytoplasm makes the efficient production of disulfide-containing proteins, such as growth factors and antibody FAB fragments, challenging (Stewart et al., 1998). Thus, the expression of disulfide-bond containing proteins in *E. coli* commonly results in the production of insoluble protein (due to misfolding) sequestered into inclusion bodies (Veldkamp et al., 2007; Chen and Leong, 2009); for a review of refolding strategies, see Tsumoto et al., (2003). When refolding conditions cannot be successfully identified, the target protein must be produced solubly in vivo. The three most common strategies to express disulfide-containing proteins, all of which are discussed later in this review, are to try the following: (1) target the expressed protein to the *E. coli* periplasm, which is highly oxidative (Leichert and Jakob, 2004), (2) fuse the protein to thioredoxin (Lefebvre et al., 2009a), and/or (3) express the protein in bacterial strains containing thioredoxin reductase and glutathione reductase mutants (Xu et al., 2008).

Transmembrane segments. Finally, the solubility of the recombinantly expressed protein will typically be compromised if the construct includes transmembrane-spanning regions. Thus, the soluble expression of transmembrane-containing proteins, especially integral membrane proteins, in *E. coli* is exceptionally challenging, requiring specialized materials and strategies (Mohanty and Wiener, 2004; Gordon et al., 2008; Dvir and Choe, 2009).

III. PROPERTIES OF THE VECTOR THAT INFLUENCE EXPRESSION AND SOLUBILITY

Once the protein target and corresponding construct(s) are determined, it must be subcloned into a vector that contains all DNA sequence elements that direct the transcription and translation of the target gene (Studier and Moffatt, 1986a). These elements include promoters, regulatory sequences, the

Production of Recombinant Proteins

5.24.7

Shine-Dalgarno box, transcriptional terminators, and origins of replication, among others. In addition, expression vectors contain a selection element, typically an antibiotic-resistance gene, to aid in plasmid selection within the host cell. Another critical feature of *E. coli* expression vectors is the presence of a *fusion tag*. In contrast to the elements described above, the fusion tags are transcribed in-frame with the construct of interest. When translated, a single fusion protein, which includes the protein of interest and the fusion tag, is obtained. Today, nearly all proteins are expressed with some kind of fusion tag, and the number and diversity of tags is continually increasing.

Origin of replication

The origin of replication of a vector is the site where replication is initiated. It also determines copy number of the vector in the host. The copy number for common *E. coli* expression plasmids ranges from low (2 to 20) to high (20 to 40). Typically, high-copy-number plasmids are desired for protein expression in *E. coli*, as they result in the maximum protein yield for a given culture volume (Jing et al., 1993; Huang et al., 1994). The origin of replication is also important to consider when carrying out protein co-expression experiments in which two different plasmids, each of which contains a different protein/biomolecule, are simultaneously transformed into the same expression cell (Johnston and Marmorstein, 2003). For these experiments, the origins of the two plasmids should be different to allow the cell to support both expression vectors.

Promoter systems

Promoters are another element of the vector that can have a profound effect on the strength and duration of transcription and, in turn, protein yield. Synthesis of mRNA is initiated when RNA polymerase binds to a specific DNA sequence, the promoter, adjacent to the target gene. This sequence contains the transcription start site, as well as two hexanucleotide sequences approximately 10 and 35 bases upstream of the initiation site that direct the binding of essential elements of the polymerase machinery (Rosenberg and Court, 1979; Hawley and McClure, 1983; Harley and Reynolds, 1987). An effective promoter for expressing heterologous proteins in *E. coli* has three characteristic features. First, it is strong, resulting in robust expression of the target gene (typically 10% to 50% of the total cellular

protein). Second, it exhibits low basal transcriptional activity to prevent unwanted transcription prior to induction. Third, induction is simple and cost-effective.

When selecting a promoter system, the nature of the protein target and its desired downstream use must be considered. If the protein target is a toxic protein (like a ribonuclease), one should consider using promoter systems that have extremely low basal expression, such as the *araBAD* promoter (Lee et al., 1987). Alternatively, for maximal protein yields, a strong promoter should be selected, such as T7 or *tac*. Finally, for aggregation-prone proteins, a cold-shock promoter, in which expression is carried out at low temperatures, may be tested. Multiple promoters have been developed for expression in *E. coli* and are summarized in Table 5.24.2. The four most widely used promoters are the T7 RNA promoter, the *araBAD* promoter, hybrid promoters, and the *cspA* promoter.

T7 promoter (T7 RNA polymerase system). The T7 RNA polymerase system is the most commonly used promoter system in *E. coli*. Gene expression is driven by the T7 RNA polymerase (from the T7 bacteriophage), which transcribes DNA five times faster than the bacterial RNA polymerase (Studier et al., 1990). Because *E. coli* lacks this enzyme, the polymerase must be delivered to the cell, via an inducible plasmid or, more often, by using an *E. coli* strain that contains a chromosomal copy of the T7 polymerase gene (Studier and Moffatt, 1986b). In the absence of an inducer, the polymerase, which itself is under the control of the *lacUV5* promoter, is not produced, and correspondingly, the gene of interest is not transcribed (Studier et al., 1990). Upon addition of the nonhydrolyzable lactose analog, IPTG, the T7 RNA polymerase is transcribed and synthesized. The polymerase then initiates transcription of the target gene by binding to a T7 polymerase–specific promoter.

Once induced, most of the cellular machinery is devoted to the production of the recombinant protein, comprising up to 50% of the total cellular protein (Studier and Moffatt, 1986b; Studier et al., 1990). However, such robust transcription can have undesirable effects. First, even minimal basal production of T7 RNA polymerase results in "leaky" expression (expression prior to induction) of the target protein (Moffatt and Studier, 1987). This can be detrimental if the protein is toxic to the host, resulting in cell death or growth arrest. To minimize leaky expression, several host strains

Table 5.24.2 Promoter Systems used to Direct Recombinant Protein Expression in *E. coli*

E. coli promoter system	Description	Induction	Advantages	Disadvantages
Most common promoter systems				
T7 RNA polymerase	T7 RNA polymerase gene under the control of L8-UV5 *lac* promoter	IPTG	High level of expression: accumulate up to 50% total cell protein Well characterized, used most often Titrate expression using Tuner strains	Leaky expression: use pLysS strains for expression of proteins toxic to host
araBAD promoter (P_{BAD})	Promoter is controlled by AraC regulator	L-arabinose (repressed by glucose)	Tight regulation Titrate expression levels from low to high Low basal expression: suitable for production of proteins toxic to host	Repressed expression state is not always zero, gene-dependent
trc and *tac* promoter	−35 sequence from *trp* promoter and −10 sequence from *lacUV5* promoter (*trc*-17 bp spacing, *tac*-16 bp spacing between sequences)	IPTG	High level of expression: accumulate 15%-30% of total cell protein	Very leaky expression: not optimal for expression of proteins toxic to host Newer, more efficient systems are available
cspA promoter	Promoter from the major cold-shock protein in *E.coli*	Temperature downshift from 37°C (expression optimal between 10°C-25°C)	Efficient expression at low temperatures Can improve folding, lower inclusion body formation Advantageous for expression of aggregation-prone and proteolytically-sensitive proteins Induction is cost efficient	Leaky expression: not optimal for expression of proteins toxic to host Translational efficiency slows at lower temperatures Not titratable
Less common promoter systems				
Phage promoter *pL*[a]	Phage promoter that is regulated by the temperature-sensitive cI repressor	Temperature shift from 30°C to 42°C	Moderately high expression Induction is cost efficient	High basal level of expression at temperatures below 30°C Induction cannot be performed at low temperatures
PhoA promoter[b]	Promoter for the gene of the periplasmic alkaline phosphatase	Lower phosphate concentration in the growth medium	Promotes secretion to the periplasm Inexpensive induction	Phosphate limitation can have negative effects on metabolism of host cell Not titratable Limited media options

continued

Production of Recombinant Proteins

5.24.9

Table 5.24.2 Promoter Systems used to Direct Recombinant Protein Expression in *E. coli, continued*

E. coli promoter system	Description	Induction	Advantages	Disadvantages
RecA promoter[c]	Promoter for *recA* gene Regulated by *lexA*	Nalidixic acid	Tight regulation No growth media or temperature restrictions	Not titratable
TetA promoter[d]	Regulated by the tetR repressor	Anhydrotetracycline	Moderately high expression Low basal expression Independent of *E. coli* strain	Not titratable

[a]See Elvin et al., 1990.

[b]See Kikuchi et al., 1981; Miyake et al., 1985; Chang et al., 1986.

[c]See Shirakawa et al., 1984; Olins and Rangwala, 1990.

[d]See Delatorre et al., 1984, Skerra, 1994.

[i.e., BL21(DE3)pLysS, Rosetta(DE3)pLysS, Rosetta-gami(DE3)pLysS] have been developed that contain a plasmid encoding T7 lysozyme, an enzyme that binds and inhibits T7 polymerase (Moffatt and Studier, 1987). The concentrations of T7 lysozyme produced in these strains are sufficient to inhibit basal transcription of target genes. However, it continues to inhibit T7 RNA polymerase after expression has been induced, and thus, it is typical for the level of recombinant protein expression to be significantly reduced in pLysS versus non-pLysS strains. For nontoxic proteins, this results in lower yields (Studier, 1991). However, for toxic proteins, this often results in higher yields because the cell growth is not arrested prior to induction by the premature expression of the toxic protein (Yeo et al., 2009). In addition, because T7 is such a strong promoter, some translated proteins aggregate and form inclusion bodies because they fail to fold before encountering another unfolded protein. In these cases, expression parameters can be changed (see section IV) to maximize the yield of soluble, folded protein. Alternatively, a weaker promoter can be used.

araBAD promoter. The arabinose promoter system (*araBAD* promoter) is a strong, titratable promoter which, unlike the T7 promoter, has almost no basal transcriptional activity (Lee et al., 1987). Thus, it is advantageous for the expression of highly toxic proteins. The induction agent for this promoter is L-arabinose (Lee et al., 1987). In the absence of L-arabinose, transcription is exceptionally low and, if needed, can be even further suppressed by the addition of glucose (Miyada et al., 1984). As reported by Guzman et al., protein expression levels increase linearly with increasing concentrations of L-arabinose over two logarithms (Guzman et al., 1995). This allows the expression level to be titrated over a wide range of inducer concentrations, which can be important when trying to either maximize expression yields (higher L-arabinose concentrations) or to increase the yield of soluble protein (lower L-arabinose concentrations). It should be noted that although this system can efficiently repress gene expression, the repression level is not always zero and the efficiency of repression is gene dependent (Guzman et al., 1995). Finally, studies that have directly compared protein yields from the *araBAD* and the T7 promoters have found that T7 promoters generally result in higher expression yields (Goulding and Perry, 2003). For example, in our hands, we have seen protein expression yields increase by 2- to as much as 10-fold by switching from the *araBAD* to the T7 promoter system.

Hybrid promoters: trc and tac promoters. The *trc* and *tac* promoters are hybrids of naturally occurring promoters, consisting of the −35 region of the *trp* promoter and the −10 region of the *lacUV5* promoter (Amann et al., 1983; Deboer et al., 1983). The only difference between these two systems is the spacing between the −35 and −10 consensus sequences, with 16 bp and 17 bp separations in the *tac* and *trc* promoters, respectively (Brosius et al., 1985). Expression is induced with IPTG (Brosius et al., 1985). *Trc* and *tac* are both considered to be strong promoters, with the *trc* promoter ∼90% as active as the *tac* promoter, and both can result in the accumulation of up to 15% to 30% of the total cellular protein (Brosius et al., 1985). Because these promoters are leaky, they can be problematic when expressing proteins that are toxic to the cell (Brosius et al., 1985, Otto et al., 1995).

Optimizing Protein Expression in *E. coli*

5.24.10

cspA promoter. Following a downshift in growth temperature from 37°C to 10°C, the bacterial cellular machinery is largely devoted to the production of 13 "cold-shock" proteins (Jones et al., 1987). After 90 min, the major cold-shock protein, CspA, accounts for 13% of total cellular protein (Goldstein et al., 1990). Accordingly, the *cspA* promoter has been exploited to direct the expression of recombinant proteins at low temperatures. For this system, induction is achieved by simply changing the growth temperature of the bacterial culture from 37°C to between 10°C and 25°C (Vasina and Baneyx, 1996); no chemical inductant is required. Moreover, because expression is carried out at low temperatures, this system has the added benefit that it promotes the soluble expression of aggregation-prone proteins (Vasina and Baneyx, 1996, 1997). One drawback is that the cold shock promoter is not completely repressed at higher temperatures, which can result in basal expression of the target protein (Qing et al., 2004). Finally, expression levels using the *cspA* promoter are typically lower than that seen with the T7 promoter.

Fusion tags

Fusion tags are proteins or peptides that are genetically fused to the target protein. They are useful because they can improve protein expression, promote folding, increase protein solubility, and facilitate downstream processes such as purification and detection. However, the "perfect" tag, i.e., one that can perform all of these tasks for every protein, still does not exist. Thus, it is often necessary to test multiple fusion tags to determine which tag results in the highest yields of soluble protein (Peti and Page, 2007; Brown et al., 2008) and to also use a combination of tags in order to facilitate both expression and purification (Nilsson et al., 1996; Pryor and Leiting, 1997; Routzahn and Waugh, 2002). For a comprehensive review of fusion tags, see Terpe (2003). Many of the most commonly used fusion tags, their biophysical characteristics, and their uses in expression and/or purification are listed in Table 5.24.3. The placement of the tag, either N-terminal or C-terminal, is also important as it can have a profound effect on soluble protein expression levels. Additionally, the presence of a fusion tag may interfere with the biological activity of the recombinantly expressed protein, and thus, in these cases, it may be important to enzymatically remove the tag after the fusion protein has been purified.

Common tags: hexahistidine (his6). The his6 tag does not enhance soluble expression, but it does facilitate purification and because of its widespread use, it is described here. His6-tagged fusion proteins are purified using immobilized metal affinity chromatography (IMAC; Porath et al., 1975). In IMAC, metal ions, typically nickel or cobalt, are immobilized to resin via a metal chelator, such as nitrilotriacetic acid, with only three or four of the six metal coordination sites occupied. The unprotonated histidines of the his6-tagged fusion protein coordinate the metal, allowing the expressed protein to be readily purified from *E. coli* lysate. The bound proteins are then eluted using either imidazole or by lowering the pH. Because this tag is small (<1 kD), it is frequently used in conjunction with solubility-enhancing fusion tags to provide a bi-functional tag that facilitates both solubility and purification. Increasing the length of the tag to eight or even ten histidine residues typically increases the purity of the purified fusion protein (Tanaka et al., 1999; Mobley et al., 2007).

Common tags: Thioredoxin (Trx). Trx is a small protein that catalyzes dithiol-disulfide exchange reactions, an activity that is important for diverse cellular processes. When overexpressed in *E. coli*, it can also accumulate to up to 40% of the total cellular protein, yet still remain soluble (LaVallie et al., 1993, Lauber et al., 2001). Thus, trx facilitates both expression and solubility. Its utility for enhancing soluble expression was recently demonstrated in two studies. In the first, 30 full-length mammalian proteins were expressed with a series of fusion tags in order to identify those tags that promote soluble expression for the most diverse set of targets. It was found that two tags, Trx and maltose-binding protein (MBP, discussed below), consistently promoted soluble expression when fused to the N-terminus of the target protein (Dyson et al., 2004). In a second, similar study, in which the expression of 27 eukaryotic proteins of suitable size for NMR studies were examined, Trx was ranked highest for its ability to promote soluble protein expression (Hammarstrom et al., 2002).

Trx is also the fusion tag of choice if expressing a protein that contains disulfide bonds. To overcome the highly reductive environment of the bacterial cytosol, new bacterial strains, which contain mutations in two proteins that play key roles in maintaining the reducing environment of the cytosol, thioredoxin reductase (*trxB*) and glutathione

Production of Recombinant Proteins

5.24.11

Table 5.24.3 Characteristics of Commonly used Fusion Tags

Tag	Protein	Amino acids	Size (kDa)	Source organism	Purification aid	Affinity matrix	Comments
Tags with solubility-enhancing properties							
GST	Glutathione S-transferase	243	28.1	*Schistosoma japonicum*	Yes	Glutathione agarose	Forms dimer in solution
MBP	Maltose binding protein	390	43.0	*Escherichia coli*	Yes	Amylose resin	Strong solubility-enhancer
DsbA	Disulfide oxidoreductase	228	25.4	*Escherichia coli*	No		Aids periplasmic disulfide-bond formation
NusA	N-utilizing substance A protein	535	59.3	*Escherichia coli*	No		Strong solubility-enhancer
Trx	Thioredoxin	135	14.7	*Escherichia coli*	No		Aids cytosolic disulfide-bond formation
Z-domain[a]	Protein A IgG ZZ repeat domain	91	10.6	*Staphylococcus aureus*	Yes	Protein A-sepharose	
GB1[b]	Protein G β1 domain	85	9.7	*Streptococcus* sp.	Yes	IgG-resins	Used often with proteins for NMR
SUMO	Small ubiquitin-like modifier	99	11.1	*Homo sapiens*	No		
SET	Solubility-enhancing tags	<40		T7 phage gene 10B; synthetic	No		Small, highly acidic peptide tags that limit protein aggregation
HaloTag-7[c]	Catalytically-inactive derivative of DhaA	296	34.0	*Rhodococcus* sp.	Yes	HaloLink resin	Strong solubility-enhancer
Tags without solubility-enhancing properties							
His$_6$	Hexahistadine	6	0.8	Synthetic	Yes	Immobilized metal resin	Often combined with solubility-enhancing tags
Intein[d]	Protein splicing element	128-1650		Variable	Yes	Chitin resin	Remove from resin by induced self-cleavage

[a]See Nilsson et al., 1987; Zhao et al., 2005.
[b]See Bao et al., 2006.
[c]See Ohana et al., 2009.
[d]See Chong et al., 1998; Singleton et al., 2002.

5.24.12

reductase (*gor*), have been developed (Origami and Rosetta-gami cells; discussed further in section IV). Because the formation of disulfide bonds appears to be dependent on the presence of thioredoxin in *trxB* mutant strains (Stewart et al., 1998), increased yields of soluble, active protein may be obtained by fusing the target protein to Trx and expressing the Trx-fusion protein in *trxB/gor* mutant strains.

Common tags: Maltose-binding protein (MBP). MBP is a soluble periplasmic protein that binds maltose and delivers it to the MalFGK2 transporter (Fox et al., 2001). As a fusion tag, MBP facilitates expression, solubility, and purification. It can also be used to target proteins to either the cytosol or periplasm (fusion proteins targeted to the periplasm include the endogenous *malE* signal sequence in the MBP gene), although it is most frequently used for cytosolic expression. Like Trx, MBP constantly ranks as one of the most effective fusion tags for promoting protein solubility (Dyson et al., 2004; Niiranen et al., 2007). In addition, MBP has been shown to passively promote the folding of its fused partner (Kapust and Waugh, 1999; Nallamsetty and Waugh, 2006, 2007). The mechanism by which MBP promotes folding is not well understood, but it is has been suggested that MBP possesses chaperone-like qualities and may function by either interacting directly with its fusion protein, stabilizing folding intermediates, or inhibiting protein aggregation (Kapust and Waugh, 1999; Fox et al., 2001).

MBP is most commonly used as an N-terminal tag. However, unlike Trx, which is only effective as an N-terminal fusion tag, MBP has been shown to effectively enhance protein solubility as both an N- and a C-terminal fusion tag (Dyson et al., 2004). Thus, MBP can be used effectively in both positions. Unlike Trx, MBP also functions as an affinity tag as it binds tightly to sugars, such as amylose or dextrin, which, when coupled to agarose or sepharose beads, can be used to purify MBP fusion proteins from the *E. coli* lysate (Diguan et al., 1988).

While the effective increase in solubility and affinity properties are highly favorable, MBP also has less desirable traits, which is why it is not the "perfect" tag. The protein is very large (42 kD) and its presence can interfere with biological activity of the recombinant protein if not removed. Moreover, MBP enhances solubility so strongly that it can even solubilize unfolded/misfolded proteins. In these cases, the target proteins often precipitate after enzymatic cleavage of MBP, resulting in wasted time and resources (Lee et al., 1993; Saavedraalanis et al., 1994; Kishore et al., 1998). In order to verify that MBP facilitates the production of folded, active protein, and does not simply solubilize misfolded proteins, it is advantageous to co-express the MBP fusion protein with its site-specific protease. In this system, the MBP fusion protein is proteolytically cleaved in vivo by the co-expressed protease (Nallamsetty and Waugh, 2006). Proteins/protein constructs for which folding and solubility is enhanced by MBP will remain soluble following cleavage while those that do not will precipitate. Small-scale expression tests can be used to rapidly identify which constructs benefit from fusion to MBP. Because the MBP tag is proteolytically removed in vivo, it is advantageous to add a purification tag, typically a his-tag, *after* the protease site (i.e., MBP–TEV site–his_6tag–protein target).

Common tags: Glutathione-S-transferase (GST). GST, a protein that catalyzes the nucleophilic attack of glutathione on electrophilic substrates in order to reduce their reactivity with other biomolecules (Armstrong, 1997), is another commonly used fusion tag. GST facilitates both expression and purification, but is not generally considered a solubility-enhancing tag as many studies have shown that it is a poor solubility enhancer (Esposito and Chatterjee, 2006; Brown et al., 2008). While GST (∼28 kDa) is smaller than MBP, it dimerizes. Thus, if the target protein also forms oligomers, the use of GST can lead to aggregation and precipitation of the expressed fusion protein, even though the target protein is folded and active (Kaplan et al., 1997). In spite of these disadvantages, GST is still widely used because it binds tightly and specifically to glutathione agarose, allowing the target protein to be purified in a single step (Smith and Johnson, 1988).

Recently developed tags: NusA. Recently, novel solubility tags have been developed with the common aim to enhance the solubility of a diverse set of targets with minimal drawbacks (Chong et al., 1998; Zhang et al., 2004; Butt et al., 2005; Ohana et al., 2009). N-utilization substance A (NusA), a transcription termination/antitermination factor in *E. coli*, is one of these new tags. Its utility as a solubility-enhancing fusion protein was identified through a systematic study of more than 4000 *E. coli* proteins. In this study, NusA was predicted to have a high solubility probability when expressed in

Production of Recombinant Proteins

5.24.13

E. coli (>90%), and subsequent experiments demonstrated that it was more effective at promoting solubility than both GrpE and baterioferritin (BFR; Davis et al., 1999). Since then, multiple studies have shown that NusA effectively enhances protein solubility for a diverse set of protein targets (De Marco et al., 2004; Douette et al., 2005; Niiranen et al., 2007); in fact, in some cases it has been shown to be even more effective than MBP (Kohl et al., 2008). However, NusA is not an affinity tag. Thus, it is typically coupled to the his$_6$-tag to facilitate purification (de Marco, 2006). Finally, it is also large (55 kDa), which can sometimes make proteolytic removal of the NusA fusion protein difficult.

Recently developed tags: SET. It was also recently shown that small highly acidic peptide tags, which are based on the C-terminal portion of the T7 phage gene 10B, promote the soluble expression of folded, active protein for highly aggregation prone proteins. These tags are known as solubility enhancing tags (SET tags). Namely, Zhang and colleagues showed that a series of T7 phage gene 10B-based peptide tags, with net negative charges of greater or equal than −6, fully solubilized the highly aggregation-prone Ig variable-type domain, CAR D1 (Zhang et al., 2004). Moreover, it was folded. Significantly, in the same study, a peptide tag with a net charge of −12 partially solubilized the myelin P-zero protein, P$_0$ex, a protein that is highly recalcitrant to solubilization (Zhang, Y.B. et al., 2004). An advantage of the SET tags is that they are small, typically <40 amino acids. They are also highly acidic, with net charges between −6 and −18. Although the mechanism by which SET tags improve solubility is still under investigation, it has been suggested that the negatively charged fusion peptide increases electrostatic repulsion between folding-intermediates (Zhang, Y.B. et al., 2004). This effectively delays polypeptide aggregation, increasing the time available for the target protein to properly fold.

Recently developed tags: SUMO. Another novel and increasingly routinely used tag is the small ubiquitin-like modifier (SUMO) tag. Its development as a prokaryotic protein expression system was based on the observation that the addition of ubiquitin to recombinant proteins increases their solubility (Butt et al., 1989; Wittliff et al., 1990). The utility of SUMO as both an expression and solubility-enhancing tag has been demonstrated in multiple studies. One of the first showed that fusion of SUMO to the N-terminus of GFP and MMP13, a protein highly recalcitrant to soluble expression, dramatically enhanced their expression and solubility (Malakhov et al., 2004). Subsequent studies have shown SUMO also effectively promoted the soluble expression of SARS coronavirus proteins (Zuo et al., 2005b) and membrane proteins (Zuo et al., 2005a). One of the advantages of SUMO is that it is small (11.2 kD). It also has its own, highly specific protease (Ulp), which functions by recognizing the tertiary structure of the SUMO protein, rather than a short, specific amino acid sequence. It cleaves the peptide chain immediately after the C-terminal residue of the SUMO protein. This is an advantage because, unlike when using other proteases, no non-native residues are left on the target protein following cleavage (Butt et al., 2005). The only restriction of the SUMO protease is that the N-terminal amino acid of the target protein cannot be a proline, as this residue restricts access to the SUMO protease active site.

Facilitating the formation of disulfide bonds: export signal sequences. An additional method to produce disulfide bond–containing proteins in *E. coli* is to export them to the periplasm, the space between the inner cytoplasmic membrane and the external outer membrane of Gram-negative bacteria like *E. coli*. The bacterial periplasm is highly oxidative and thus promotes the formation of disulfide bonds (for a review of methods used to produce disulfide containing proteins in *E. coli* see de Marco, 2009). The *E. coli* maltose-binding protein is targeted to the periplasm in vivo via an N-terminal export signal sequence. Thus, when a construct of MBP that includes the export sequence is used as a fusion tag, the fusion protein is exported to the periplasm and, in turn, the oxidative environment promotes disulfide bond formation (Riggs, 2000). Because MBP is used to target proteins both to the cytoplasm and to the periplasm, it is important to verify that the expression vector used localizes the target protein to the desired cellular compartment. In addition to MBP, DsbA, a Trx homolog, has also been used as a fusion partner to target and promote the proper folding of proteins in the periplasm (Collinsracie et al., 1995; Couprie et al., 2000; Winter et al., 2000). Finally, it is important to point out that proteins targeted to the periplasm have specialized protocols for purification, which differ from those used for proteins targeted to the bacterial cytosol (Malik et al., 2007).

Position of fusion tag. All described solubility and affinity tags can be fused either to the N- or C-terminus of the target protein. However, N-terminal fusions are the most common and have the added benefit that they often enhance soluble protein expression more successfully than C-terminal fusions (Sati et al., 2002; Dyson et al., 2004; Busso et al., 2005). One of the reasons N- versus C-terminally fused tags effectively improve protein expression and solubility is because many of the fusion tags are native *E. coli* proteins. Thus, they provide an efficient 5′ sequence for the initiation of transcription (Esposito and Chatterjee, 2006). Correspondingly, the N-terminal mRNA sequence and structure is efficiently recognized by the ribosome and compatible with robust translation (Cebe and Geiser, 2006). In contrast, the transcription and translation initiation sites of target genes and proteins are variable and can impede these processes when tags are fused to the target protein C-terminus. Notably, there are exceptions, as MBP and SET tags have been shown to be equally effective in enhancing solubility regardless of which terminus they are fused to (Dyson et al., 2004; Zhang, Y.B. et al., 2004), although this is still protein dependent. One tag that is often fused to the protein C-terminus is the his$_6$-tag. Typically, it is equally effective at either location, but has the advantage that when fused to the C-terminus, only fully translated proteins are purified as incompletely translated target proteins do not include the purification tag.

Removal of the fusion tag. All fusion tags have the potential to influence the behavior of the expressed protein, and thus it is typically desirable to remove the tag following purification. This is achieved by engineering a protease-specific cleavage site (a sequence of ∼7 amino acids that is specifically recognized by the protease) between the tag and target protein. Following expression and purification, the corresponding protease is used to cleave the tag from the protein in vitro.

Several proteases are widely used for fusion tag removal (see Table 5.24.4). However, many enzymes, such as thrombin, factor Xa, and enterokinase, can promiscuously cleave the target protein at nonspecific sites (Chang, 1985; Choi et al., 2001). Today, one of the most widely used proteases is tobacco etch virus (TEV) protease. It has a number of advantages. First, it is more stringent and thus efficiently cleaves fusion tags without nonspecific secondary cleavage (Phan et al., 2002). Second, it can accommodate a number of different residues on the C-terminal side of the sessile bond, eliminating nonnative residues on the N-terminus of the target protein (Kapust et al., 2002). TEV protease is also active over a diverse set of experimental conditions, such as pH, buffers, and temperatures (Kapust et al., 2001, 2002). A disadvantage is that TEV protease is inhibited by a large number of detergents, making the removal of fusion tags from integral membrane proteins, which require protein-detergent complexes for solubilization, challenging (Mohanty et al., 2003). If detergents are an essential component of the protein buffer, thrombin protease, which retains its activity in a wide range of detergents, can be used to cleave the fusion tag.

Sometimes a protease will fail to cleave the fusion protein. Most often, this is due to steric hindrance in which the protease site is not accessible to the enzyme (Kapust and Waugh, 2000; Lee et al., 2008). Including a short linker of glycine, asparagine, or alanine residues between the protease site and the fusion protein often alleviates this problem (Esposito and Chatterjee, 2006). Inefficient cleavage may also be overcome by altering the cleavage conditions, such as increasing protease concentrations, extending the cleavage time, and altering temperature, among other parameters.

IV. *E. COLI* HOST STRAINS THAT AID EXPRESSION OF HETEROLOGOUS PROTEINS

Several elements of the *E. coli* strain used for recombinant protein production have a large impact on the success of soluble expression. As mentioned in previous sections of this review, bacterial host strains have been specifically developed to aid the expression of heterologous proteins (e.g., codon-supplemented cells to aid the expression of proteins with rare codons). Brief descriptions of commercially available *E. coli* strains designed for the specific expression of proteins that are susceptible to proteolysis, contain rare codons, or require disulfide-bonds are provided below.

Protease-deficient strains: BL21, BL21 (DE3), and BL21Star (DE3)

E. coli BL21 and its derivatives are most frequently used for routine protein expression. These strains are deficient of *ompT* and *lon* proteases. OmpT is a bacterial endoprotease that readily cleaves T7 RNA polymerase (Grodberg and Dunn, 1988), while the Lon (La) protease, encoded by the *lon* gene, is an ATP-dependent enzyme that rapidly degrades

Table 5.24.4 Commonly used Proteases to Remove Fusion Tags from Recombinant Proteins

Protease	Description	Cleavage site[a]	Protease inhibitors	Activity	Comments
TEV	Catalytic domain of the Nuclear Inclusion a (NIa) protein, a cysteine protease, found in the tobacco etch virus (TEV)	EXXYXQ-(G/S) Most common: ENLYFQ-X	PMSF, AEBSF, TLCK, pepstatin A, bestatin, E-64, zinc (>5 mM), EDTA (1 mM), reagents that react with cysteine; various detergents	pH: 4-9 Temperature: 4°C-37°C (max. at 34°C) Can accommodate most buffers	Sequence specificity more stringent than factor Xa, thrombin, and enterokinase Can be produced in *E. coli*
3C[b]	Recombinant form of the 3C protease from human rhinovirus type 14	EVLFQ-GP	PMSF, TLCK, leupeptin, zinc (100 mM), urea (1 M), guanidine (1 M)	pH: 3-10 Temperature: 4°C-37°C (max. at 4°C) Test activity in buffer of choice	High sequence specificity Can be produced in *E. coli*
Xa[c]	Factor Xa is a serine protease that converts prothrombin to thrombin	I(E/N)GR- Will not cleave at site followed by P or R	PMSF, AEBSF, DFP, aprotinin, antithrombin III, antipain, α1-antitypsin, chymostatin, hirudin, leupeptin, urea (100 mM), guanidine (10 mM), NaCl (100 mM), imidazole (100 mM)	pH: 6.5-9 Temperature: 4°C-37°C Activity is highly reduced in phosphate buffers compared to Tris or HEPES, $CaCl_2$ should be included in cleavage reaction	Nonspecific proteolysis may occur at secondary sites Cleavage usually performed near physiological conditions
Thr	Thrombin is a serine protease that converts fibrinogen into fibrin	LVPR-GS	PMSF, AEBSF, aprotinin, antithrombin III, antipain, α1-antitypsin, chymostatin, hirudin, leupeptin, reducing agents	pH: 5-10 Temperature: 4°C-37°C (max. at 37°C) Can accommodate most buffers	Nonspecific proteolysis may occur at secondary sites Cleavage usually performed near physiological conditions
EntK	Catalytic subunit of bovine enterokinase	DDDDK-	Serine protease inhibitors, PMSF, imidazole (250 mM), NaCl (250 mM), urea (2 M), SDS	pH: 6.0-8.5 Temperature: 4°C-37°C Activity is highly reduced in phosphate buffers compared to Tris or MES	Nonspecific proteolysis may occur at secondary sites

[a]The "–" indicates the site of cleavage within the single letter amino acid code.
[b]See Miyashita et al., 1992.
[c]See Nagai and Thogersen, 1984; Jenny et al., 2003.

misfolded and recombinant proteins (Phillips et al., 1984). Correspondingly, deletion of *lon* and *ompT* is correlated with increased expression and stability of recombinant proteins (Gottesman, 1990). Since, the T7 RNA polymerase system is the most widely used promoter system for protein expression, most BL21 strains, designated (DE3) strains, contain a chromosomal copy of the T7 RNA polymerase gene, allowing for simple and efficient expression of genes from T7-based expression vectors.

BL21Star (DE3) (Invitrogen), a derivative of the BL21 (DE3) strain, contains an additional mutation in the *rne* gene (Lopez et al., 1999). This gene encodes RNase E, an enzyme that functions as an essential part of the "degradosome" to actively degrade mRNA within the cell (Grunberg-Manago, 1999; Lopez et al., 1999; Carpousis, 2007). Consequently, the use of this strain results in an increase in mRNA stability, and, in turn, protein expression. The strain developers have routinely noted a 2- to 10-fold increase in heterologous protein expression when compared to non-RNase E defective BL21 strains (Invitrogen). However, the basal expression level of the target gene is also increased when using these strains, and therefore, a variant that exhibits repressed basal expression, BL21Star (DE3)pLysS (see description of pLysS strains below), should be used when expressing toxic proteins.

Codon-supplemented strains: BL21-RP, BL21-RIL, BL21-RPIL, Rosetta, and Rosetta-gami

As described in section II, differences in codon frequency between the target gene and the expression host can lead to translational stalling, premature translation termination, and amino acid mis-incorporation (Kane, 1995). Instead of generating a codon-optimized gene, this disparity may be overcome by supplying the rare tRNAs during expression (Burgess-Brown et al., 2008). Numerous bacterial strains (BL21-RP, BL21-RIL, BL21-RPIL, Rosetta, and Rosetta-gami) are available that contain plasmids that encode rare tRNAs to promote the efficient expression of genes that contain high frequencies of rare codons. The tRNAs that are supplemented in each strain differ, and therefore, the appropriate host strain should be determined on a protein-to-protein basis. Additionally, the Rosetta-gami cell line, which contains *trxB/gor* mutants, has the added benefit of facilitating cytoplasmic disulfide bond formation (see next section) (Milisavljevic et al., 2009).

Strains to express disulfide-bonded proteins: BL21 trxB, Origami, Rosetta-gami

As discussed in section III, some host strains (BL21 trxB, Origami, Rosetta-gami) have mutations in thioredoxin reductase (*trxB*) and/or glutathione reductase (*gor*) genes, two proteins that maintain the reducing environment of the bacterial cytosol (Prinz et al., 1997). Consequently, these strains aid the formation of cytosolic disulfide bonds, greatly enhancing the solubility of folded, active disulfide bond–containing proteins produced in the cytosol. Moreover, these strains can also be more effective than exporting the target protein to the periplasm. For example, in one study, a *trxB/gor* mutant strain (FA113; a suppressor strain that overcomes the normally slow growth of *trxB/gor* mutant strains in the absence of reductant) increased the active yield of multiple proteins [*E. coli* alkaline phosphatase, a truncated form of the human tissue plasminogen activator (vtPA) and full-length human tPA, which contained two, nine, and seventeen disulfide bonds, respectively] when compared to yields obtained by secretion of the proteins to the periplasm (Bessette et al., 1999). Because inter-protein disulfide bonds can also form, leading to high-molecular-weight oligomers, these strains should only be used to express proteins that require disulfide bonds for proper folding.

Strains to express toxic proteins: BL21-AI and pLysS

The BL21-AI host strain, in which the T7 RNA polymerase gene is under the control of the *araBAD* promoter, was developed to allow the expression of toxic genes from any T7-based expression vector. As described earlier, the low basal activity of the *araBAD* promoter system is optimal for the expression of proteins that are toxic to the host. In BL21-AI cells, basal expression of the T7 RNA polymerase, and the subsequent target gene, is highly repressed in the absence of arabinose and the presence of glucose. This cell line has been used for the robust expression of several heterologous enzymes, with soluble yields exceeding 30 mg/liter (Chen et al., 2009; Yao et al., 2009).

As an alternative to the BL21-AI strain, several BL21 derivatives, known as pLysS strains (also described in section III), express T7 phage lysozyme, an enzyme that effectively

inhibits T7 RNA polymerase activity. Inhibition of T7 RNA polymerase, in turn, decreases basal expression of the target protein. Like BL21-AI strains, pLysS strains have been used to successfully express toxic proteins under the control of T7 RNA polymerase (Jensen et al., 1999; Cinquin et al., 2001).

Typically, protein expression is accomplished using host strains that have a combination of the elements described above. For example, Origami (DE3) pLysS cells have *trxB/gor* mutations, promoting the production of disulfide-bonds, contain a chromosomal copy of T7 RNA polymerase for the efficient expression of genes under the control of this enzyme, and express pLysS to decrease basal expression and aid production of toxic target proteins. This strain was essential for the soluble production of BSPH1, a novel human sperm-binding protein that contains four disulfide-bonds and is completely insoluble when expressed in BL21 (DE3) cells (Lefebvre et al., 2009b). Therefore, all properties of the protein should be considered when choosing the optimal host strain to express the protein of interest.

V. THE SOLUBILITY OF PROTEINS CAN BE IMPROVED BY CHANGING EXPRESSION CONDITIONS

In *E. coli*, transcription and translation are tightly coupled, with ~60,000 polypeptide chains being synthesized per minute (Lorimer, 1996). The use of strong expression promoters and high inducer concentrations can result in nascent protein concentrations that are so high that the proteins aggregate before folding. Thus, reducing the rates of transcription and/or translation facilitates folding by allowing the newly synthesized protein to fold before it aggregates and forms inclusion bodies. Below we describe the common expression condition parameters that can be manipulated to enhance protein solubility.

Temperature

Lowering the expression temperature routinely improves the solubility of recombinantly expressed proteins (Shirano and Shibata, 1990; Kataeva et al., 2005; Volonte et al., 2008; Piserchio et al., 2009). At lower temperatures, cell processes slow down, leading to reduced rates of transcription, translation, and cell division (Chou, 2007), while also leading to reduced protein aggregation (Vasina and Baneyx, 1997; Sahdev et al., 2008). Moreover, most proteases are less active at lower temperatures, and thus lowering the expression temperature also results in a reduction in the degradation of proteolytically sensitive proteins (Spiess et al., 1999; Hunke and Betton, 2003; Pinsach et al., 2008).

Because of the profound increase in the yield of soluble protein at low temperatures, it is strongly suggested to use a low induction temperature as the default (see typical protocol in Fig. 5.24.2). The bacterial culture should be cultivated at 37°C until mid-to-late log phase (Chou, 2007). The culture is then induced and the recombinant protein synthesized between 15°C to 25°C (Peti and Page, 2007; Gråslund et al., 2008a). Due to the reduced protein synthesis rate, longer induction times are necessary to obtain a sufficient protein yield (typical induction time: 4 hr at 37°C, 16 to 20 hr at 18°C).

Concentration of the inducer

In addition to lowering the growth temperature, a reduction in transcription rate can also be achieved by lowering the concentration of the induction agent. For example, the *araBAD* promoter system is titratable, and thus lower concentrations of the inducer, L-arabinose, result in the production of less protein (Guzman et al., 1995). Additionally, decreasing the concentration of IPTG can also enhance the production of soluble protein. For example, the solubility of recombinant cyclomaltodextrinase (CDase) was shown to be highly sensitive to the concentration of the inducer. When the protein was induced using 0.05 mM IPTG, the protein was soluble and active; however, when the inducer concentration was doubled to 0.1 mM the expressed protein was insoluble and inactive (Turner et al., 2005). Thus, although the most common IPTG concentrations for protein induction range from 0.1 to 1.0 mM, a decrease to even lower levels can effect solubility. Finally, a derivative of the BL21 (DE3) expression strain, containing a *lacY1* deletion mutation, has been developed that allows the T7 promoter system to be titratable over a range of IPTG concentrations (Turner et al., 2005). This cell line, called Tuner (DE3), permits accurate tuning of the induction level and can, in some cases, promote the production of soluble protein (Turner et al., 2005).

Media

Although fermentation has its clear advantages, batch culture is the most common method to cultivate cells for recombinant protein expression. Since there is limited control over the growth parameters using

this approach, all nutrients that are required for growth must be supplied from the beginning by inclusion in the growth medium. Luria broth (LB) is the standard for the expression of proteins. This broth is composed of bactotryptone, which provides peptides, peptones and essential amino acids, yeast extract, which provides vitamins and trace elements, and sodium chloride, which provides sodium ions to maintain osmotic balance (Sahdev et al., 2008). Terrific broth (TB), the second most widely used expression medium, is formulated to increase protein solubility and yield (Sahdev et al., 2008). M9, or minimal medium, contains the minimum nutrients essential for bacterial species to grow. Minimal medium usually lacks amino acids, and thus is most widely used for selective labeling, i.e., isotope labeling for proteins studied using NMR spectroscopy (Reilly and Fairbrother, 1994; Pryor and Leiting, 1997; Chen et al., 2006).

VI. ENHANCING SOLUBILITY BY COEXPRESSION WITH OTHER PROTEINS

Some proteins only express solubly when they are coexpressed with additional biomolecules, typically other proteins, like a binding partner, or molecular chaperones. In these cases, the target protein is coexpressed with a second protein that is encoded on either the same plasmid or a separate plasmid. Multiple vectors have been designed for coexpression of two or more proteins, such as the Duet vectors (Novagen; coexpression of two or more target proteins from the same vector using the T7 promoter system) or separate vectors (Expression Technologies; coexpression of two proteins using two different vectors). Moreover, multiple Duet vectors can be used in the same cell, allowing the expression of up to eight proteins simultaneously. Finally, multiple vectors have been developed that contain *E. coli* chaperones (GroEL/ES, dnaJKE, trigger factor; Takara Bio), allowing the protein of interest to be coexpressed with one or a combination of chaperones to facilitate proper folding (Kyratsous et al., 2009). Below are a few examples of proteins that require coexpression with either a partner protein or molecular chaperone for successful expression.

Coexpression with a partner protein

Expressing proteins whose activities disrupt *E. coli* growth and/or signaling are often toxic to the cell. While host cells can be cultivated to a high density under repressive conditions, induction of the toxic protein may result in rapid growth arrest and/or cell death. For some proteins, this toxicity can be mitigated when the toxin is coexpressed with a second protein that either binds and inhibits the toxin or does not bind yet counteracts its activity. Both strategies allow the toxic protein to be successfully solubly expressed to high levels. Two families of proteins that are routinely associated with cell toxicity are endogenous *E. coli* toxins, such as RelE, which inhibits translation through mRNA cleavage (Pedersen et al., 2003), and protein tyrosine kinases (PTK), which can promiscuously phosphorylate nonnative *E. coli* targets (Wang, Y.H. et al., 2006). However, multiple groups have successfully expressed these proteins in *E. coli* by coexpressing them with an appropriate protein partner. Thus, if the target protein is toxic to the host cell, coexpression with a protein that antagonizes the toxic effect should be considered.

Toxin-antitoxin pairs. Toxin:antitoxin pairs (Gerdes et al., 2005) are composed of an unstable antitoxin and a stable toxin. Under normal conditions, the toxin and antitoxin associate to form a tight, nontoxic complex. However, under conditions of stress, the antitoxins are degraded, leading to cell growth arrest due to the cellular effects of the toxin (inhibiting replication by blocking DNA gyrase or, more commonly, inhibiting translation via mRNA cleavage) (Gerdes et al., 2005). One such toxin:antitoxin system is *mqsRA*, which encodes the MqsR toxin, a ribonuclease, and its cognate MqsA antitoxin (Brown et al., 2009). We showed that overexpression of MqsR alone results in growth arrest, as translation is inhibited via mRNA cleavage by MqsR. However, coexpression of MqsR with MqsA, which binds and mitigates MqsR-mediated toxicity, resulted in the robust expression of both proteins. In this case, MqsR and MqsA were induced simultaneously using IPTG from separate T7 polymerase-based vectors (MqsR: pet28A, Kan resistance, pBR322 promoter; MqsA: pAC21a, Amp resistance, pACYC promoter) (Brown et al., 2009). There are also other examples in which other bacterial toxins have been solubly produced in *E. coli* by coexpression with their cognate antitoxins (Gerdes et al., 2005).

More recently, these toxin-antitoxin systems have been used as plasmid-stabilization systems that can effectively increase recombinant protein production levels. While the antitoxin protein is constitutively expressed from a plasmid that encodes the protein of interest, the toxin gene is chromosomally incorporated

Production of Recombinant Proteins

and highly repressed in the presence of the antitoxin (StabyExpress, Delphi Genetics). If the plasmid is lost from the host cell, the highly labile antitoxin is readily degraded and the activity of the toxin is no longer inhibited, leading to cell growth arrest. This plasmid-stabilization strategy, using the CcdA/CcdB toxin-antitoxin pair, often increases the final recombinant protein production levels by three- to five-fold (Stieber et al., 2008).

Kinase-phosphatase partners. Several, but not all, tyrosine kinases have proven difficult to express in *E. coli*. The expression of an active kinase is often accompanied by low yield, extensive degradation, insolubility, and a heterogeneously autophosphorylated sample (Y.H. Wang et al., 2006). In contrast, catalytically inactive kinases can be expressed to high yields with little difficulty. Thus, the activity of the kinase is not well tolerated by the host cell. To counteract the negative effects of the kinase, multiple groups have coexpressed tyrosine kinases with an opposing tyrosine phosphatase. For example, Src kinase, a tyrosine kinase, either expresses insolubly or spontaneously mutates to generate an inactive kinase when expressed alone. However, when coexpressed with PTP1B, a tyrosine phosphatase, milligram amounts of soluble, active kinase were produced (Y.H. Wang et al., 2006). In this case, rather than using two different plasmids, an MBP-PTP1B-thrombin site-Src fusion protein was expressed from a single plasmid. Following purification, the protein fusion tag (MBP-PTP1B) was removed by thrombin cleavage (Y.H. Wang et al., 2006). This kinase-phosphatase coexpression method has also proven effective for the soluble expression of Met kinase, which was solubly coexpressed with PTP1B in which both proteins were expressed using two compatible vectors (Seeliger et al., 2005), and Abl kinase, which was solubly coexpressed with YopH using a single bicistronic vector (W.R. Wang et al., 2006).

Coexpression with molecular chaperones

Small bacterial proteins typically fold rapidly, due to fast folding kinetics. However, larger bacterial and heterologous proteins fold more slowly, and thus require protein chaperones and folding catalysts to prevent aggregation and facilitate folding in *E. coli* (Baneyx and Mujacic, 2004). While molecular chaperones are ubiquitous in the cell, they are rapidly titrated during overexpression of recombinant proteins. Thus, supplementation

of these folding-partners using coexpression methods has proven to effectively aid solubilization of target proteins.

DnaK-DnaJ-GrpE and GroEL-GroES. The DnaK-DnaJ-GrpE and GroEL-GroES systems are the most extensively characterized molecular chaperones in *E. coli*. Both groups are folding chaperones that bind to solvent-exposed hydrophobic domains, preventing nascent peptide aggregation, and through iterative rounds of ATP-driven conformational changes, mediate the folding/unfolding of their substrates (for an extensive review of cytosolic molecular chaperones see Hartl and Hayer-Hartl, 2002). While the soluble expression of certain proteins is improved by coexpression with DnaK-DnaJ-GrpE (Chen et al., 2003), other proteins result in higher yields of soluble protein when coexpressed with GroEL-GroES (Ayling and Baneyx, 1996; Park et al., 2004; Sahu et al., 2009). Unfortunately, it is not possible to predict which, if any, molecular chaperone system will improve protein solubility. Due to the reduced metabolic burden, it is recommended to test protein expression with the two chaperone systems separately (for a thorough coexpression protocol see Baneyx and Palumbo, 2003).

In our laboratory, we have found that coexpression of protein phosphatase 1 (PP1α), a metal-dependent serine/threonine phosphatase, with GroEL and GroES results in the maximum yield of soluble protein (Kelker et al., 2009). To ensure ample chaperones are available to aid protein folding at the start of recombinant protein production, GroEL and GroES expression, which is under the control of the *araBAD* promoter (induced with L-arabinose), is induced at a culture optical density of $\lambda_{600} = 0.5$, while PP1α, which is under the control of the T7 promoter (induced with IPTG), is induced later, at a culture optical density of $\lambda_{600} = 1.0$. After 20 hr (10°C), the cultivated cells are pelleted and then resuspended in fresh medium that contains 200 µg/ml chloroamphenical to block all ribosome activity and the cultures are allowed to shake for an additional 2 hr. This last step allows for in vivo refolding and results in a substantial increase in the amount of folded, active PP1α (Kelker et al., 2009).

cpn10-cpn60. Unfortunately, most molecular chaperone systems display reduced activity at reduced temperatures. For example, while GroEL/GroES is most efficient at 30°C, this system is only 30% active at 12°C. As discussed throughout the review, there are extensive advantages to expressing recombinant
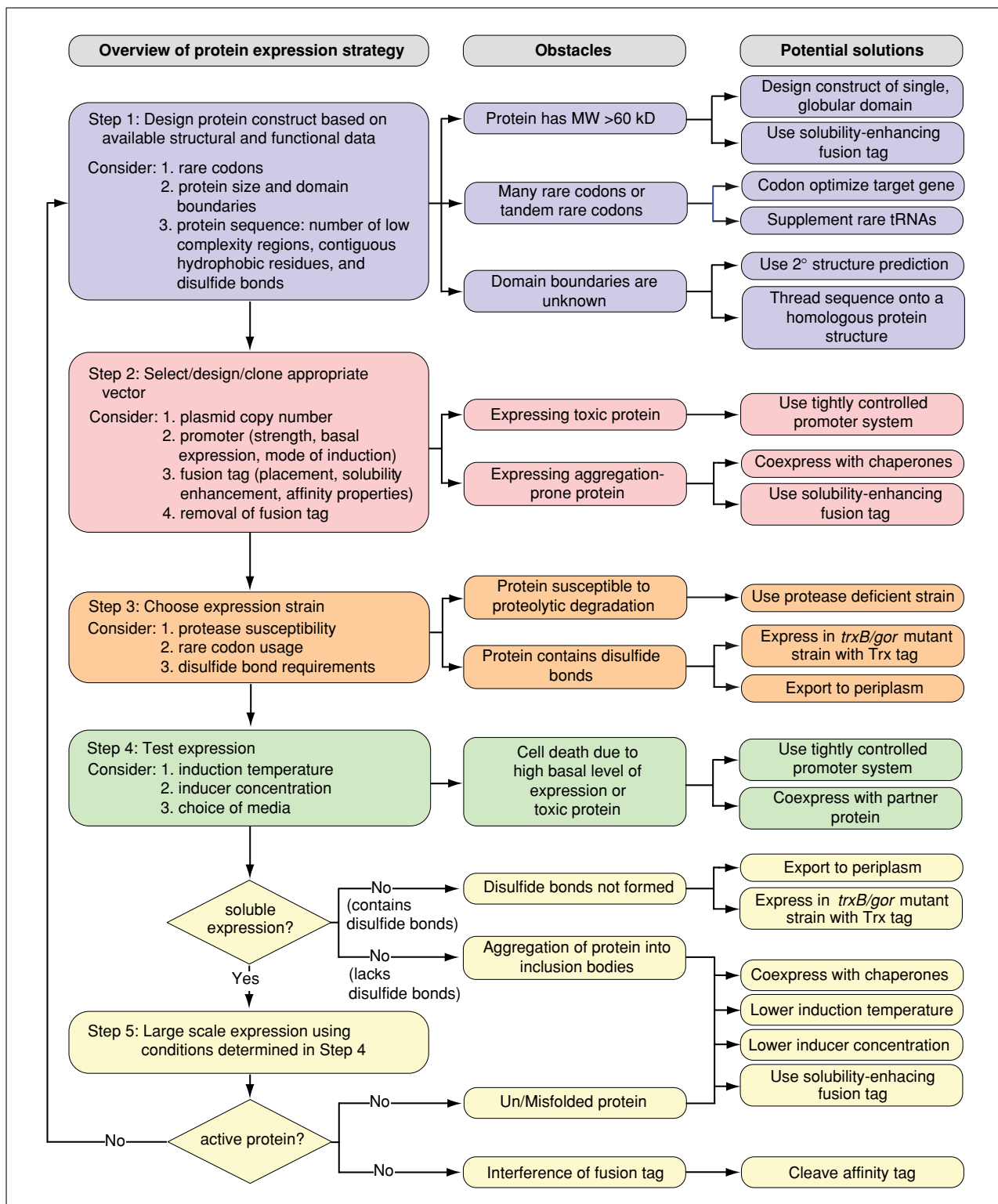
Figure 5.24.3 Flowchart depicting the critical factors to consider, common obstacles, and potential solutions for each stage of protein expression in *E. coli*. The left column lists the major steps of recombinant protein expression with key variables to consider. The middle column includes common obstacles encountered at each step, while possible solutions for each obstacle are presented in the right column.

proteins at low temperatures. A second system is available that is maximally functional at low temperatures: the cold-adapted chaperonins Cpn10 and Cpn60 from psychrophilic bacterium, *Oleispira antarctica*. These chaperonins, which have 74% and 54% amino acid identity with GroEL and GroES, respectively, are effective folding-modulators at low temperatures (4°C to 12°C) (Amada et al., 1995). Using this system, a temperature-sensitive esterase expressed at 10°C with Cpn10 and Cpn60 exhibited a 180-fold increase in activity over expression at 37°C (Ferrer et al., 2004). A derivative of the BL21 host strain, ArcticExpress (Stratagene), that coexpresses these two chaperonins, has been developed and used to successfully express several proteins, including interleukin-2 tyrosine kinase (Joseph and Andreotti, 2008).

ANTICIPATED RESULTS AND TIME CONSIDERATIONS

All of the parameters listed in this unit, from construct length to inducer concentration, can affect the solubility of recombinant proteins produced in *E. coli*. Accordingly, it is often necessary to vary one or many elements in the expression protocol to successfully express soluble protein (Fig. 5.24.3). This can mean testing scores of expression protocols, which is time consuming and costly. Thus, it is advantageous to use small-scale expression tests to select an optimal construct and expression conditions prior to scaling up. The microexpression protocol used in our laboratory to determine expression, as well as protein solubility, has been explicitly described (Peti and Page, 2007). It is recommended to use this strategy to determine an optimal protocol for large-scale expression. Be aware that results from small-scale growths do not always translate to large-scale systems. While positive small-scale results can usually be reproduced in large-scale studies, proteins that appear to have low or insoluble expression on a small-scale may be expressed in soluble form when grown on a larger scale (Gräslund et al., 2008a).

The importance of folded, active recombinant protein to the proposed research project defines the amount of time and effort that is devoted to creating the optimal expression protocol. For example, five years were devoted to identifying the optimal expression and purification protocol for protein phosphatase 1 (PP1α) (Kelker et al., 2009). This optimized protocol includes coexpressing PP1α

with molecular chaperones, supplementing the cultivation medium with metals, growing the culture and inducing protein expression at low temperatures, using low concentrations of inductants, and cleaving the fusion tag after the protein has been stabilized by a PP1α ligand (either an inhibitor or a PP1α binding protein). While time-consuming, the high yields of soluble, active PP1α have been essential for the successful completion of multiple projects (Dancheck et al., 2008, Kelker et al., 2009); thus, the effort devoted to determining the optimal protocol was invaluable.

ACKNOWLEDGMENTS

LITERATURE CITED

Amada, K., Yohda, M., Odaka, M., Endo, I., Ishii, N., Taguchi, H., and Yoshida, M. 1995. Molecular-cloning, expression, and characterization of Chaperonin-60 and Chaperonin-10 from a thermophilic bacterium, Thermus-Thermophilus Hb8. *J. Biochem.* 118:347–354.

Amann, E., Brosius, J., and Ptashne, M. 1983. Vectors bearing a hybrid Trp-Lac promoter useful for regulated expression of cloned genes in *Escherichia coli*. *Gene* 25:167-178.

Armstrong, R.N. 1997. Structure, catalytic mechanism, and evolution of the glutathione transferases. *Chem. Res. Toxicol.* 10:2-18.

Arnold, K., Bordoli, L., Kopp, J., and Schwede, T. 2006. The SWISS-MODEL workspace: A web-based environment for protein structure homology modelling. *Bioinformatics* 22:195-201.

Ayling, A. and Baneyx, F. 1996. Influence of the GroE molecular chaperone machine on the in vitro refolding of *Escherichia coli* beta-galactosidase. *Protein Sci.* 5:478-487.

Baneyx, F. and Mujacic, M. 2004. Recombinant protein folding and misfolding in *Escherichia coli*. *Nat. Biotechnol.* 22:1399-1408.

Baneyx, F. and Palumbo, J.L. 2003. Improving heterologous protein folding via molecular chaperone and foldase co-expression. *Methods Mol. Biol.* 205:171-197.

Bao, W.J., Gao, Y.G., Chang, Y.G., Zhang, T.Y., Lin, X.J., Yan, X.Z., and Hu, H.Y. 2006. Highly efficient expression and purification system of small-size protein domains in *Escherichia coli* for biochemical characterization. *Protein Expr. Purif.* 47:599-606.

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

Brosius, J., Erfle, M., and Storella, J. 1985. Spacing of the −10 and −35 regions in the tac promoter.

Effect on its in vivo activity. *J. Biol. Chem.* 260:3539-3541.

Brown, B.L., Hadley, M., and Page, R. 2008. Heterologous high-level *E. coli* expression, purification and biophysical characterization of the spine-associated RapGAP (SPAR) PDZ domain. *Protein Expr. Purif.* 62:9-14.

Brown, B.L., Grigoriu, S., Kim, Y., Arruda, J.M., Davenport, A., Wood, T.K., Peti, W., and Page, R. 2009. Three dimensional structure of the MqsR:MqsA complex: A novel TA pair comprised of a toxin homologous to RelE and an antitoxin with unique properties. *PLoS Pathog.* 5:e1000706.

Burgess-Brown, N.A., Sharma, S., Sobott, F., Loenarz, C., Oppermann, U., and Gileadi, O. 2008. Codon optimization can improve expression of human genes in *Escherichia coli*: A multi-gene study. *Protein Expr. Purif.* 59:94-102.

Busso, D., Delagoutte-Busso, B., and Moras, D. 2005. Construction of a set Gateway-based destination vectors for high-throughput cloning and expression screening in *Escherichia coli*. *Anal. Biochem.* 343:313-321.

Butt, T.R., Jonnalagadda, S., Monia, B.P., Sternberg, E.J., Marsh, J.A., Stadel, J.M., Ecker, D.J., and Crooke, S.T. 1989. Ubiquitin fusion augments the yield of cloned gene-products in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 86:2540-2544.

Butt, T.R., Edavettal, S.C., Hall, J.P., and Mattern, M.R. 2005. SUMO fusion technology for difficult-to-express proteins. *Protein Expr. Purif.* 43: 1-9.

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

Canaves, J.M., Page, R., Wilson, I.A., and Stevens, R.C. 2004. Protein biophysical properties that correlate with crystallization success in Thermotoga maritima: Maximum clustering strategy for structural genomics. *J. Mol. Biol.* 344:977-991.

Carpousis, A.J. 2007. The RNA degradosome of *Escherichia coli*: An mRNA-degrading machine assembled on RNase E. *Annu. Rev. Microbiol.* 61:71-87.

Cebe, R. and Geiser, M. 2006. Rapid and easy thermodynamic optimization of the 5′-end of mRNA dramatically increases the level of wild type protein expression in *Escherichia coli*. *Protein Expr. Purif.* 45:374-380.

Chang, C.N., Kuang, W.J., and Chen, E.Y. 1986. Nucleotide-sequence of the alkaline-phosphatase gene of *Escherichia coli*. *Gene* 44:121-125.

Chang, J.Y. 1985. Thrombin specificity - requirement for apolar amino-acids adjacent to the thrombin cleavage site of polypeptide substrate. *Eur. J. Biochem.* 151:217-224.

Chen, L.F., Maloney, K., Krol, E., Zhu, B., and Yang, J. 2009. Cloning, overexpression, purification, and characterization of the maleylacetate reductase from sphingobium chlorophenolicum strain ATCC 53874. *Curr. Microbiol.* 58:599-603.

Chen, X., Tong, X.T., Xie, Y.H., Wang, Y., Ma, J.B., Gao, D.M., Wu, H.M., and Chen, H.B. 2006. Over-expression and purification of isotopically labeled recombinant ligand-binding domain of orphan nuclear receptor human B1-binding factor/human liver receptor homologue 1 for NMR studies. *Protein Expr. Purif.* 45:99-106.

Chen, Y. and Leong, S.S.J. 2009. Adsorptive refolding of a highly disulfide-bonded inclusion body protein using anion-exchange chromatography. *J. Chromatogr. A* 1216:4877-4886.

Chen, Y., Song, J.M., Sui, S.F., and Wang, D.N. 2003. DnaK and DnaJ facilitated the folding process and reduced inclusion body formation of magnesium transporter CorA overexpressed in *Escherichia coli*. *Protein Expr. Purif.* 32:221-231.

Choi, S.I., Song, H.W., Moon, J.W., and Seong, B.L. 2001. Recombinant enterokinase light chain with affinity tag: Expression from *Saccharomyces cerevisiae* and its utilities in fusion protein technology. *Biotechnol. Bioeng.* 75:718-724.

Chong, S.R., Montello, G.E., Zhang, A.H., Cantor, E.J., Liao, W., Xu, M.Q., and Benner, J. 1998. Utilizing the C-terminal cleavage activity of a protein splicing element to purify recombinant proteins in a single chromatographic step. *Nucleic Acids Res.* 26:5109-5115.

Chou, C.P. 2007. Engineering cell physiology to enhance recombinant protein production in *Escherichia coli*. *Appl. Microbiol. Biotechnol.* 76:521-532.

Cinquin, O., Christopherson, R.I., and Menz, R.I. 2001. A hybrid plasmid for expression of toxic malarial proteins in *Escherichia coli*. *Mol. Biochem. Parasitol.* 117:245-247.

Collinsracie, L.A., McColgan, J.M., Grant, K.L., Diblasio-Smith, E.A., McCoy, J.M., and Lavallie, E.R. 1995. Production of recombinant bovine enterokinase catalytic subunit in *Escherichia coli* using the novel secretory fusion partner Dsba. *Biotechnology* 13:982-987.

Couprie, J., Vinci, F., Dugave, C., Quemeneur, E., and Moutiez, M. 2000. Investigation of the DsbA mechanism through the synthesis and analysis of an irreversible enzyme-ligand complex. *Biochemistry* 39:6732-6742.

Critton, D.A., Tortajada, A., Stetson, G., Peti, W., and Page, R. 2008. Structural basis of substrate recognition by hematopoietic tyrosine phosphatase. *Biochemistry* 47:13336-13345.

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

Production of Recombinant Proteins

5.24.23

Dancheck, B., Nairn, A.C., and Peti, W. 2008. Detailed structural characterization of unbound protein phosphatase 1 inhibitors. *Biochemistry* 47:12346-12356.

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

de Marco, A. 2006. Two-step metal affinity purification of double-tagged (NusA-His(6)) fusion proteins. *Nat. Protoc.* 1:1538-1543.

de Marco, A. 2009. Strategies for successful recombinant expression of disulfide bond-dependent proteins in *Escherichia coli. Microb. Cell Fact.* 8:26.

De Marco, V., Stier, G., Blandin, S., and de Marco, A. 2004. The solubility and stability of recombinant proteins are increased by their fusion to NusA. *Biochem. Biophys. Res. Commun.* 322:766-771.

Deboer, H.A., Comstock, L.J., and Vasser, M. 1983. The Tac promoter - a functional hybrid derived from the Trp and Lac promoters. *Proc. Natl. Acad. Sci. U.S.A.* 80:21-25.

Delatorre, J.C., Ortin, J., Domingo, E., Delamarter, J., Allet, B., Davies, J., Bertrand, K.P., Wray, L.V., and Reznikoff, W.S. 1984. Plasmid vectors based on Tn10 DNA - gene-expression regulated by tetracycline. *Plasmid* 12:103-110.

DePristo, M.A., Zilversmit, M.M., and Hartl, D.L. 2006. On the abundance, amino acid composition, and evolutionary dynamics of low-complexity regions in proteins. *Gene* 378:19-30.

Diguan, C., Li, P., Riggs, P.D., and Inouye, H. 1988. Vectors that facilitate the expression and purification of foreign peptides in *Escherichia coli* by fusion to maltose-binding protein. *Gene* 67:21-30.

Douette, P., Navet, R., Gerkens, P., Galleni, M., Levy, D., and Sluse, F.E. 2005. *Escherichia coli* fusion carrier proteins act as solubilizing agents for recombinant uncoupling protein 1 through interactions with GroEL. *Biochem. Biophys. Res. Commun.* 333:686-693.

Dvir, H. and Choe, S. 2009. Bacterial expression of a eukaryotic membrane protein in fusion to various Mistic orthologs. *Protein Expr. Purif.* 68:28-33.

Dyson, H.J. and Wright, P.E. 2005. Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* 6:197-208.

Dyson, M.R., Shadbolt, S.P., Vincent, K.J., Perera, R.L., and McCafferty, J. 2004. Production of soluble mammalian proteins in *Escherichia coli*: Identification of protein features that correlate with successful expression. *BMC Biotechnol.* 4:32.

Elvin, C.M., Thompson, P.R., Argall, M.E., Hendry, P., Stamford, N.P.J., Lilley, P.E., and Dixon, N.E. 1990. Modified bacteriophage-lambda promoter vectors for overproduction of proteins in *Escherichia coli. Gene* 87:123-126.

Esposito, D. and Chatterjee, D.K. 2006. Enhancement of soluble protein expression through the use of fusion tags. *Curr. Opin. Biotechnol.* 17:353-358.

Ferrer, M., Chernikova, T.N., Timmis, K.N., and Golyshin, P.N. 2004. Expression of a temperature-sensitive esterase in a novel chaperone-based *Escherichia coli* strain. *Appl. Environ. Microbiol.* 70:4499-4504.

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

Gerdes, K., Christensen, S.K., and Lobner-Olesen, A. 2005. Prokaryotic toxin-antitoxin stress response loci. *Nat. Rev. Microbiol.* 3:371-382.

Goh, C.S., Lan, N., Douglas, S.M., Wu, B., Echols, N., Smith, A., Milburn, D., Montelione, G.T., Zhao, H., and Gerstein, M. 2004. Mining the structural genomics pipeline: Identification of protein properties that affect high-throughput experimental analysis. *J. Mol. Biol.* 336:115-130.

Goldstein, J., Pollitt, N.S., and Inouye, M. 1990. Major cold shock protein of *Escherichia coli. Proc. Natl. Acad. Sci. U.S.A.* 87:283-287.

Gordon, E., Horsefield, R., Swarts, H.G., de Pont, J.J., Neutze, R., and Snijder, A. 2008. Effective high-throughput overproduction of membrane proteins in *Escherichia coli. Protein Expr. Purif.* 62:1-8.

Gottesman, S. 1990. Minimizing proteolysis in *Escherichia coli* - Genetic solutions. *Methods in Enzymol.* 185:119-129.

Goulding, C.W. and Perry, L.J. 2003. Protein production in *Escherichia coli* for structural studies by X-ray crystallography. *J. Struct. Biol.* 142:133-143.

Gråslund, S., Nordlund, P., Weigelt, J., Hallberg, B.M., Bray, J., Gileadi, O., Knapp, S., Oppermann, U., Arrowsmith, C., Hui, R., Ming, J., dhe-Paganon, S., Park, H.W., Savchenko, A., Yee, A., Edwards, A., Vincentelli, R., Cambillau, C., Kim, R., Kim, S.H., Rao, Z., Shi, Y., Terwilliger, T.C., Kim, C.Y., Hung, L.W., Waldo, G.S., Peleg, Y., Albeck, S., Unger, T., Dym, O., Prilusky, J., Sussman, J.L., Stevens, R.C., Lesley, S.A., Wilson, I.A., Joachimiak, A., Collart, F., Dementieva, I., Donnelly, M.I., Eschenfeldt, W.H., Kim, Y., Stols, L., Wu, R., Zhou, M., Burley, S.K., Emtage, J.S., Sauder, J.M., Thompson, D., Bain, K., Luz, J., Gheyi, T., Zhang, F., Atwell, S., Almo, S.C., Bonanno, J.B., Fiser, A., Swaminathan, S., Studier, F.W., Chance, M.R., Sali, A., Acton, T.B., Xiao, R., Zhao, L., Ma, L.C., Hunt, J.F., Tong, L., Cunningham, K., Inouye, M., Anderson, S., Janjua, H., Shastry, R., Ho, C.K., Wang, D., Wang, H., Jiang, M., Montelione, G.T., Stuart, D.I., Owens, R.J., Daenke, S., Schutz, A., Heinemann, U., Yokoyama, S., Bussow, K., and Gunsalus, K.C. 2008a. Protein production and purification. *Nat. Methods* 5:135-146.

Gråslund, S., Sagemark, J., Berglund, H., Dahlgren, L.G., Flores, A., Hammarstroem, M., Johansson, I., Kotenyova, T., Nilsson, M., Nordlund, P., and Weigelt, J. 2008b. The use of systematic N- and C-terminal deletions to promote production and structural studies of recombinant proteins. *Protein Expr. Purif* 58:210-221.

Grodberg, J. and Dunn, J.J. 1988. Ompt encodes the *Escherichia coli* outer-membrane protease that cleaves T7-Rna polymerase during purification. *J. Bacteriol.* 170:1245-1253.

Grunberg-Manago, M. 1999. Messenger RNA stability and its role in control of gene expression in bacteria and phages. *Ann. Rev. Genet.* 33:193-227.

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

Hammarstrom, M., Hellgren, N., van Den Berg, S., Berglund, H., and Hard, T. 2002. Rapid screening for improved solubility of small human proteins produced as fusion proteins in *Escherichia coli. Protein Sci.* 11:313-321.

Harley, C.B. and Reynolds, R.P. 1987. Analysis of *Escherichia coli* promoter sequences. *Nucleic Acids Res.* 15:2343-2361.

Hartl, F.U. and Hayer-Hartl, M. 2002. Protein folding - Molecular chaperones in the cytosol: From nascent chain to folded protein. *Science* 295:1852-1858.

Hatfield, G.W. and Roth, D.A. 2007. Optimizing scaleup yield for protein production: Computationally Optimized DNA Assembly (CODA) and translation engineering. *Biotechnol. Annu. Rev.* 13:27-42.

Hawley, D.K. and McClure, W.R. 1983. Compilation and analysis of *Escherichia coli* promoter DNA-sequences. *Nucleic Acids Res.* 11:2237-2255.

Huang, B.H., Shi, Z.T., and Tsai, M.D. 1994. A small, high-copy-number vector suitable for both in-vitro and in-vivo gene-expression. *Gene* 151:143-145.

Hunke, S. and Betton, J.M. 2003. Temperature effect on inclusion body formation and stress response in the periplasm of *Escherichia coli. Mol. Microbiol.* 50:1579-1589.

Jaroszewski, L., Rychlewski, L., Li, Z., Li, W., and Godzik, A. 2005. FFAS03: A server for profile–profile sequence alignments. *Nucleic Acids Res.* 33:W284-W288.

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

Jensen, P.Y., Bonander, N., Horn, N., Tumer, Z., and Farver, O. 1999. Expression, purification and copper-binding studies of the first metal-binding domain of Menkes protein. *Eur. J. Biochem.* 264:890-896.

Jing, G.Z., Huang, Z., Liu, Z.G., and Zou, Q. 1993. Plasmid Pkkh - an improved vector with higher copy number for expression of foreign genes in *Escherichia coli. Biotechnol. Lett.* 15:439-442.

Johnston, K. and Marmorstein, R. 2003. Co-expression of proteins in *E. coli* using dual expression vectors. *Methods Mol. Biol.* 205:205-213.

Jones, D.T. 1999. Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* 292:195-202.

Jones, P.G., Vanbogelen, R.A., and Neidhardt, F.C. 1987. Induction of proteins in response to low-temperature in *Escherichia coli. J. Bacteriol.* 169:2092-2095.

Joseph, R.E. and Andreotti, A.H. 2008. Bacterial expression and purification of Interleukin-2 Tyrosine kinase: Single step separation of the chaperonin impurity. *Protein Expr. Purif.* 60:194-197.

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

Kaplan, W., Husler, P., Klump, H., Erhardt, J., Sluis Cremer, N., and Dirr, H. 1997. Conformational stability of pGEX-expressed Schistosoma japonicum glutathione S-transferase: A detoxification enzyme and fusion-protein affinity tag. *Protein Sci.* 6:399-406.

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

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

Kapust, R.B., Tozser, J., Fox, J.D., Anderson, D.E., Cherry, S., Copeland, T.D., and Waugh, D.S. 2001. Tobacco etch virus protease: Mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. *Protein Eng.* 14:993-1000.

Kapust, R.B., Tozser, J., Copeland, T.D., and Waugh, D.S. 2002. The P1′ specificity of tobacco etch virus protease. *Biochem. Biophys. Res. Commun.* 294:949-955.

Karlin, S., Brocchieri, L., Bergman, A., Mrazek, J., and Gentles, A.J. 2002. Amino acid runs in eukaryotic proteomes and disease associations. *Proc. Natl. Acad. Sci. U.S.A.* 99:333-338.

Kataeva, I., Chang, J., Xu, H., Luan, C.H., Zhou, J., Uversky, V.N., Lin, D., Horanyi, P., Liu, Z.J., Ljungdahl, L.G., Rose, J., Luo, M., and Wang, B.C. 2005. Improving solubility of Shewanella oneidensis MR-1 and Clostridium thermocellum JW-20 proteins expressed into *Esherichia coli. J. Proteome Res.* 4:1942-1951.

Kelker, M.S., Page, R., and Peti, W. 2009. Crystal structures of protein phosphatase-1 bound to nodularin-R and tautomycin: A novel

scaffold for structure-based drug design of serine/threonine phosphatase inhibitors. *J. Mol. Biol.* 385:11-21.

Kikuchi, Y., Yoda, K., Yamasaki, M., and Tamura, G. 1981. The nucleotide-sequence of the promoter and the amino-terminal region of alkaline-phosphatase structural gene (Phoa) of *Escherichia coli. Nucleic Acids Res.* 9:5671-5678.

Kishore, U., Leigh, L.E.A., Eggleton, P., Strong, P., Perdikoulis, M.V., Willis, A.C., and Reid, K.B.M. 1998. Functional characterization of a recombinant form of the C-terminal, globular head region of the B-chain of human serum complement protein, C1q. *Biochem. J.* 333:27-32.

Klock, H.E., Koesema, E.J., Knuth, M.W., and Lesley, S.A. 2008. Combining the polymerase incomplete primer extension method for cloning and mutagenesis with microscreening to accelerate structural genomics efforts. *Proteins* 71:982-994.

Kohl, T., Schmidt, C., Wiemann, S., Poustka, A., and Korf, U. 2008. Automated production of recombinant human proteins as resource for proteome research. *Proteome Sci.* 6:10.

Kyratsous, C.A., Silverstein, S.J., DeLong, C.R., and Panagiotidis, C.A. 2009. Chaperone-fusion expression plasmid vectors for improved solubility of recombinant proteins in *Escherichia coli. Gene* 440:9-15.

Langlais, C., Guilleaume, B., Wermke, N., Scheuermann, T., Ebert, L., LaBaer, J., and Korn, B. 2007. A systematic approach for testing expression of human full-length proteins in cell-free expression systems. *BMC Biotechnol.* 7:64.

Lauber, T., Marx, U.C., Schulz, A., Kreutzmann, P., Rosch, P., and Hoffmann, S. 2001. Accurate disulfide formation in *Escherichia coli*: Overexpression and characterization of the first domain (HF6478) of the multiple Kazal-type inhibitor LEKTI. *Protein Expr. Purif.* 22:108-112.

LaVallie, E.R., DiBlasio, E.A., Kovacic, S., Grant, K.L., Schendel, P.F., and McCoy, J.M. 1993. A thioredoxin gene fusion expression system that circumvents inclusion body formation in the *E. coli* cytoplasm. *Biotechnology* 11:187-193.

Lee, C.D., Sun, H.C., Hu, S.M., Chiu, C.F., Homhuan, A., Liang, S.M., Leng, C.H., and Wang, T.F. 2008. An improved SUMO fusion protein system for effective production of native proteins. *Protein Sci.* 17:1241-1248.

Lee, H.S., Berger, D.K., and Kustu, S. 1993. Activity of purified nifa, a transcriptional activator of nitrogen-fixation genes. *Proc. Natl. Acad. Sci. U.S.A.* 90:2266-2270.

Lee, N., Francklyn, C., and Hamilton, E.P. 1987. Arabinose-induced binding of arac protein to aral2 activates the *arabad* operon promoter. *Proc. Natl. Acad. Sci. U.S.A.* 84:8814-8818.

Lefebvre, J., Boileau, G., and Manjunath, P. 2009a. Recombinant expression and affinity purification of a novel epididymal human sperm-binding protein, BSPH1. *Mol. Hum. Reprod.* 15:105-114.

Lefebvre, J., Boileau, G., and Manjunath, P. 2009b. Recombinant expression and affinity purification of a novel epididymal human sperm-binding protein, BSPH1. *Mol. Hum. Reprod.* 15:105-114.

Leichert, L.I. and Jakob, U. 2004. Protein thiol modifications visualized in vivo. *PLoS Biol.* 2:e333.

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

Lorimer, G.H. 1996. A quantitative assessment of the role of the chaperonin proteins in protein folding in vivo. *FASEB J.* 10:5-9.

Malakhov, M.P., Mattern, M.R., Malakhova, O.A., Drinker, M., Weeks, S.D., and Butt, T.R. 2004. SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins. *J. Struct. Funct. Genomics* 5:75-86.

Malik, A., Jenzsch, M., Lubbert, A., Rudolph, R., and Sohling, B. 2007. Periplasmic production of native human proinsulin as a fusion to *E. coli* ecotin. *Protein Expr. Purif.* 55:100-111.

Milisavljevic, M.D., Papic, D.R., Timotijevic, G.S., and Maksimovic, V.R. 2009. Successful production of recombinant buckwheat cysteine-rich aspartic protease in *Escherichia coli. J. Serb. Chem. Soc.* 74:607-618.

Miyada, C.G., Stoltzfus, L., and Wilcox, G. 1984. Regulation of the arac gene of *Escherichia coli* - Catabolite repression, auto-regulation, and effect on arabad expression. *Proc. Natl. Acad. Sci. U.S.A.* 81:4120-4124.

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

Miyashita, K., Kusumi, M., Utsumi, R., Komano, T., and Satoh, N. 1992. Expression and purification of recombinant 3c-proteinase of coxsackievirus-B3. *Biosci. Biotechnol. Biochem.* 56:746-750.

Mobley, C.K., Myers, J.K., Hadziselimovic, A., Ellis, C.D., and Sanders, C.R. 2007. Purification and initiation of structural characterization of human peripheral myelin protein 22, an integral membrane protein linked to peripheral neuropathies. *Biochemistry* 46:11185-11195.

Moffatt, B.A. and Studier, F.W. 1987. T7 lysozyme inhibits transcription by T7 rna-polymerase. *Cell* 49:221-227.

Mohanty, A.K. and Wiener, M.C. 2004. Membrane protein expression and production: Effects of polyhistidine tag length and position. *Protein Expr. Purif.* 33:311-325.

Optimizing Protein Expression in *E. coli*

5.24.26

Mohanty, A.K., Simmons, C.R., and Wiener, M.C. 2003. Inhibition of tobacco etch virus protease activity by detergents. *Protein Expr. Purif.* 27:109-114.

Mustelin, T., Tautz, L., and Page, R. 2005. Structure of the hematopoietic tyrosine phosphatase (HePTP) catalytic domain: Structure of a KIM phosphatase with phosphate bound at the active site. *J. Mol. Biol.* 354:150-163.

Nagai, K. and Thogersen, H.C. 1984. Generation of beta-globin by sequence-specific proteolysis of a hybrid protein produced in *Escherichia coli*. *Nature* 309:810-812.

Nallamsetty, S. and Waugh, D.S. 2006. Solubility-enhancing proteins MBP and NusA play a passive role in the folding of their fusion partners. *Protein Expr. Purif.* 45:175-182.

Nallamsetty, S. and Waugh, D.S. 2007. Mutations that alter the equilibrium between open and closed conformations of *Escherichia coli* maltose-binding protein impede its ability to enhance the solubility of passenger proteins. *Biochem. Biophys. Res. Commun.* 364:639-644.

Netzer, W.J. and Hartl, F.U. 1997. Recombination of protein domains facilitated by co-translational folding in eukaryotes. *Nature* 388:343-349.

Niiranen, L., Espelid, S., Karlsen, C.R., Mustonen, M., Paulsen, S.M., Heikinheimo, P., and Willassen, N.P. 2007. Comparative expression study to increase the solubility of cold adapted Vibrio proteins in *Escherichia coli*. *Protein Expr. Purif.* 52:210-218.

Nilsson, B., Moks, T., Jansson, B., Abrahmsen, L., Elmblad, A., Holmgren, E., Henrichson, C., Jones, T.A., and Uhlen, M. 1987. A synthetic igg-binding domain based on staphylococcal protein-A. *Protein Eng.* 1:107-113.

Nilsson, J., Larsson, M., Stahl, S., Nygren, P.A., and Uhlen, M. 1996. Multiple affinity domains for the detection, purification and immobilization of recombinant proteins. *J. Mol. Recognit.* 9:585-594.

Oberg, K., Chrunyk, B.A., Wetzel, R., and Fink, A.L. 1994. Native-like secondary structure in interleukin-1-beta inclusion-bodies by attenuated total reflectance Ftir. *Biochemistry* 33:2628-2634.

Ohana, R.F., Enccell, L.P., Zhao, K., Simpson, D., Slater, M.R., Urh, M., and Wood, K.V. 2009. HaloTag7: A genetically engineered tag that enhances bacterial expression of soluble proteins and improves protein purification. *Protein Expr. Purif.* 68:110-120.

Olins, P.O. and Rangwala, S.H. 1990. Vector for enhanced translation of foreign genes in *Escherichia coli*. *Methods Enzymol.* 185:115-119.

Otto, C.M., Niagro, F., Su, X.Z., and Rawlings, C.A. 1995. Expression of recombinant feline tumor-necrosis-factor is toxic to *Escherichia coli*. *Clin. Diagn. Lab. Immunol.* 2:740-746.

Park, S.L., Kwon, M.J., Kim, S.K., and Nam, S.W. 2004. GroEL/ES chaperone and low culture temperature synergistically enhanced the soluble expression of CGTase in *E-coli*. *J. Microbiol. Biotechnol.* 14:216-219.

Pedersen, K., Zavialov, A.V., Pavlov, M.Y., Elf, J., Gerdes, K., and Ehrenberg, M. 2003. The bacterial toxin RelE displays codon-specific cleavage of mRNAs in the ribosomal A site. *Cell* 112:131-140.

Peng, L., Xu, Z.N., Fang, X.M., Wang, F., Yang, S., and Cen, P.L. 2004. Preferential codons enhancing the expression level of human beta-defensin-2 in recombinant *Escherichia coli*. *Protein Pept. Lett.* 11:339-344.

Peti, W. and Page, R. 2007. Strategies to maximize heterologous protein expression in *Escherichia coli* with minimal cost. *Protein Expr. Purif.* 51:1-10.

Phan, J., Zdanov, A., Evdokimov, A.G., Tropea, J.E., Peters, H.K., Kapust, R.B., Li, M., Wlodawer, A., and Waugh, D.S. 2002. Structural basis for the substrate specificity of tobacco etch virus protease. *J. Biol. Chem.* 277:50564-50572.

Phillips, T.A., Vanbogelen, R.A., and Neidhardt, F.C. 1984. Ion gene-product of *Escherichia coli* is a heat-shock protein. *J. Bacteriol.* 159:283-287.

Pinsach, J., de Mas, C., Lopez-Santin, J., Striedner, G., and Bayer, K. 2008. Influence of process temperature on recombinant enzyme activity in *Escherichia coli* fed-batch Cultures. *Enzyme Microb. Technol.* 43:507-512.

Piserchio, A., Ghose, R., and Cowburn, D. 2009. Optimized bacterial expression and purification of the c-Src catalytic domain for solution NMR studies. *J. Biomol. NMR* 44:87-93.

Porath, J., Carlsson, J., Olsson, I., and Belfrage, G. 1975. Metal chelate affinity chromatography, a new approach to protein fractionation. *Nature* 258:598-599.

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

Pryor, K.D. and Leiting, B. 1997. High-level expression of soluble protein in *Escherichia coli* using a His(6)-tag and maltose-binding-protein double-affinity fusion system. *Protein Expr. Purif.* 10:309-319.

Qing, G., Ma, L.C., Khorchid, A., Swapna, G.V., Mal, T.K., Takayama, M.M., Xia, B., Phadtare, S., Ke, H., Acton, T., Montelione, G.T., Ikura, M., and Inouye, M. 2004. Cold-shock induced high-yield protein production in *Escherichia coli*. *Nat. Biotechnol.* 22:877-882.

Reilly, D. and Fairbrother, W.J. 1994. A novel isotope labeling protocol for bacterially expressed Proteins. *J. Biomol. NMR* 4:459-462.

Production of Recombinant Proteins

5.24.27

Riggs, P. 2000. Expression and purification of recombinant proteins by fusion to maltose-binding protein. *Mol. Biotechnol.* 15:51-63.

Ritz, D. and Beckwith, J. 2001. Roles of thiol-redox pathways in bacteria. *Annu. Rev. Microbiol.* 55:21-48.

Rosenberg, M. and Court, D. 1979. Regulatory sequences involved in the promotion and termination of rna-transcription. *Annu. Rev. Genet.* 13:319-353.

Routzahn, K. and Waugh, D. 2002. Differential effects of supplementary affinity tags on the solubility of MBP fusion proteins. *J. Struct. Funct. Genomics* 2:83-92.

Saavedraalanis, V.M., Rysavy, P., Rosenberg, L.E., and Kalousek, F. 1994. Rat-liver mitochondrial processing peptidase - both alpha-subunit and beta-subunit are required for activity. *J. Biol. Chem.* 269:9284-9288.

Sahdev, S., Khattar, S.K., and Saini, K.S. 2008. Production of active eukaryotic proteins through bacterial expression systems: A review of the existing biotechnology strategies. *Mol. Cell Biochem.* 307:249-264.

Sahu, S.K., Rajasekharan, A., and Gummadi, S.N. 2009. GroES and GroEL are essential chaperones for refolding of recombinant human phospholipid scramblase 1 in *E. coli*. *Biotechnol Lett.* 31:1745-1752.

Sakharkar, M.K., Kangueane, P., Sakharkar, K.R., and Zhong, Z. 2006. Huge proteins in the human proteome and their participation in hereditary diseases. *In Silico Biol.* 6:275-279.

Sati, S.P., Singh, S.K., Kumar, N., and Sharma, A. 2002. Extra terminal residues have a profound effect on the folding and solubility of a *Plasmodium falciparum* sexual stage-specific protein over-expressed in *Escherichia coli*. *Eur. J. Biochem.* 269:5259-5263.

Schenk, P.M., Baumann, S., Mattes, R., and Steinbiss, H.H. 1995. Improved high-level expression system for eukaryotic genes in *Escherichia coli* using T7 RNA polymerase and rare ArgtRNAs. *Biotechniques* 19:196-200.

Seeliger, M.A., Young, M., Henderson, M.N., Pellicena, P., King, D.S., Falick, A.M., and Kuriyan, J. 2005. High yield bacterial expression of active c-Abl and c-Src tyrosine kinases. *Protein Sci.* 14:3135-3139.

Sharp, P.M. and Li, W.H. 1987. The codon Adaptation Index–a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.* 15:1281-1295.

Shirakawa, M., Tsurimoto, T., and Matsubara, K. 1984. Plasmid vectors designed for high-efficiency expression controlled by the portable *reca* promoter-operator of *Escherichia coli*. *Gene* 28:127-132.

Shirano, Y. and Shibata, D. 1990. Low temperature cultivation of *Escherichia coli* carrying a rice lipoxygenase L-2 cDNA produces a soluble and active enzyme at a high level. *FEBS Lett.* 271:128-130.

Singleton, S.F., Simonette, R.A., Sharma, N.C., and Roca, A.I. 2002. Intein-mediated affinity-fusion purification of the *Escherichia coli* RecA protein. *Protein Expr. Purif.* 26:476-488.

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.

Smith, D.B. and Johnson, K.S. 1988. Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione s-transferase. *Gene* 67:31-40.

Spiess, C., Beil, A., and Ehrmann, M. 1999. A temperature-dependent switch from chaperone to protease in a widely conserved heat shock protein. *Cell* 97:339-347.

Spirin, A.S. 2004. High-throughput cell-free systems for synthesis of functionally active proteins. *Trends Biotechnol.* 22:538-545.

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

Stieber, D., Gabant, P., and Szpirer, C. 2008. The art of selective killing: Plasmid toxin/antitoxin systems and their technological applications. *Biotechniques* 45:344-346.

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

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

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

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

Tanaka, T., Kubota, M., Samizo, K., Nakajima, Y., Hoshino, M., Kohno, T., and Wakamatsu, E. 1999. One-step affinity purification of the G protein beta gamma subunits from bovine brain using a histidine-tagged G protein alpha subunit. *Protein Expr. Purif.* 15:207-212.

Tegel, H., Steen, J., Konrad, A., Nikdin, H., Pettersson, K., Stenvall, M., Tourle, S., Wrethagen, U., Xu, L., Yderland, L., Uhlen, M., Hober, S., and Ottosson, J. 2009. High-throughput protein production–lessons from scaling up from 10 to 288 recombinant proteins per week. *Biotechnol. J.* 4:51-57.

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

Tsumoto, K., Ejima, D., Kumagai, I., and Arakawa, T. 2003. Practical considerations in refolding proteins from inclusion bodies. *Protein Expr. Purif.* 28:1-8.

Optimizing Protein Expression in *E. coli*

5.24.28

Turner, P., Holst, O., and Karlsson, E.N. 2005. Optimized expression of soluble cyclomaltodextrinase of thermophilic origin in *Escherichia coli* by using a soluble fusion-tag and by tuning of inducer concentration. *Protein Expr. Purif.* 39:54-60.

Valax, P. and Georgiou, G. 1993. Molecular characterization of beta-lactamase inclusion-bodies produced in *Escherichia coli*. 1. Composition. *Biotechnol. Prog.* 9:539-547.

Vasina, J.A. and Baneyx, F. 1996. Recombinant protein expression at low temperatures under the transcriptional control of the major *Escherichia coli* cold shock promoter cspA. *Appl. Environ. Microbiol.* 62:1444-1447.

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

Veldkamp, C.T., Peterson, F.C., Hayes, P.L., Mattmiller, J.E., Haugner, J.C., de la Cruz, N., and Volkman, B.F. 2007. On-column refolding of recombinant chemokines for NMR studies and biological assays. *Protein Expr. Purif.* 52:202-209.

Vincentelli, R., Bignon, C., Gruez, A., Canaan, S., Sulzenbacher, G., Tegoni, M., Campanacci, V., and Cambillau, C. 2003. Medium-scale structural genomics: Strategies for protein expression and crystallization. *Accounts Chem. Res.* 36:165-172.

Volonte, F., Marinelli, F., Gastaldo, L., Sacchi, S., Pilone, M.S., Pollegioni, L., and Molla, G. 2008. Optimization of glutaryl-7-aminocephalosporanic acid acylase expression in *E-coli*. *Protein Expr. Purif.* 61:131-137.

Wakagi, T., Oshima, T., Imamura, H., and Matsuzawa, H. 1998. Cloning of the gene for inorganic pyrophosphatase from a thermoacidophilic archaeon, *Sulfolobus* sp. strain 7, and overproduction of the enzyme by coexpression of tRNA for arginine rare codon. *Biosci. Biotechnol. Biochem.* 62:2408-2414.

Walsh, G. 2006. Biopharmaceutical benchmarks 2006. *Nature Biotechnol.* 24:769-765.

Wang, W.R., Marimuthu, A., Tsai, J., Kumar, A., Krupka, H.I., Zhang, C., Powell, B., Suzuki, Y., Nguyen, H., Tabrizizad, M., Luu, C., and West, B.L. 2006. Structural characterization of autoinhibited c-Met kinase produced by coexpression in bacteria with phosphatase. *Proc. Natl. Acad. Sci. U.S.A.* 103:3563-3568.

Wang, Y.H., Ayrapetov, M.K., Lin, X.F., and Sun, G.Q. 2006. A new strategy to produce active human Src from bacteria for biochemical study of its regulation. *Biochem. Biophys. Res. Commun.* 346:606-611.

Winter, J., Neubauer, P., Glockshuber, R., and Rudolph, R. 2000. Increased production of human proinsulin in the periplasmic space of *Escherichia coli* by fusion to DsbA. *J. Biotechnol.* 84:175-185.

Wittliff, J.L., Wenz, L.L., Dong, J., Nawaz, Z., and Butt, T.R. 1990. Expression and characterization of an active human estrogen-receptor as a ubiquitin fusion protein from *Escherichia coli*. *J. Biol. Chem.* 265:22016-22022.

Xu, Y., Yasin, A., Tang, R., Scharer, J.M., Moo-Young, M., and Chou, C.P. 2008. Heterologous expression of lipase in *Escherichia coli* is limited by folding and disulfide bond formation. *Appl. Microbiol. Biotechnol.* 81:79-87.

Yan, F., Qian, M.L., Yang, F., Cai, F., Yuan, Z., Lai, S.T., Zhao, X.Y., Gou, L.T., Hu, Z.G., and Deng, H.X. 2007. A novel pro-apoptosis protein PNAS-4 from *Xenopus laevis*: Cloning, expression, purification, and polyclonal antibody production. *Biochemistry (Moscow)* 72:664-671.

Yao, J.W., Patrone, J.D., and Dotson, G.D. 2009. Characterization and kinetics of phosphopantothenoylcysteine synthetase from *Enterococcus faecalis*. *Biochemistry* 48:2799-2806.

Yeo, Y.J., Shin, S., Lee, S.G., Park, S., and Jeong, Y.J. 2009. Production, purification, and characterization of soluble NADH-flavin Oxidoreductase (StyB) from *Pseudomonas putida* SN1. *J. Microbiol. Biotechnol.* 19:362-367.

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

Zhang, Y.B., Howitt, J., McCorkle, S., Lawrence, P., Springer, K., and Freimuth, P. 2004. Protein aggregation during overexpression limited by peptide extensions with large net negative charge. *Protein Expr. Purif.* 36:207-216.

Zhang, Z.W., Gildersleeve, J., Yang, Y.Y., Xu, R., Loo, J.A., Uryu, S., Wong, C.H., and Schultz, P.G. 2004. A new strategy for the synthesis of glycoproteins. *Science* 303:371-373.

Zhao, Y.X., Benita, Y., Lok, M., Kuipers, B., van der Ley, P., Jiskoot, W., Hennink, W.E., Crommelin, D.J.A., and Oosting, R.S. 2005. Multi-antigen immunization using IgG binding domain ZZ as carrier. *Vaccine* 23:5082-5090.

Zuo, X., Li, S., Hall, J., Mattern, M.R., Tran, H., Shoo, J., Tan, R., Weiss, S.R., and Butt, T.R. 2005a. Enhanced expression and purification of membrane proteins by SUMO fusion in *Escherichia coli*. *J. Struct. Funct. Genomics* 6:103-111.

Zuo, X., Mattern, M.R., Tan, R., Li, S., Hall, J., Sterner, D.E., Shoo, J., Tran, H., Lim, P., Sarafianos, S.G., Kazi, L., Navas-Martin, S., Weiss, S.R., and Butt, T.R. 2005b. Expression and purification of SARS coronavirus proteins using SUMO-fusions. *Protein Expr. Purif.* 42:100-110.

Production of Recombinant Proteins

5.24.29