Molecular Biology Problem Solver: A Laboratory Guide. Edited by Alan S. Gerstein
Copyright © 2001 by Wiley-Liss, Inc.
ISBNs: 0-471-37972-7 (Paper); 0-471-22390-5 (Electronic)

15

E. coli Expression Systems

Peter A. Bell

Expression Vector Structure 462
What Makes a Plasmid an Expression Vector? 462
Is a Stronger Promoter Always Desirable? 463
Why Do Promoters Leak and What Can You Do about It? 464
What Factors Affect the Level of Translation? 464
What Can Affect the Stability of the Protein in the Cell? 464
Which Protein Expression System Suits Your Needs? 465
Track Record 465
What Do You Know about the Gene to Be Expressed? . . . 465
What Do You Know about Your Protein? 468
Advertisements for Commercial Expression Vectors Are Very Promising. What Levels of Expression Should You Expect? 470
Which *E. coli* Strain Will Provide Maximal Expression for Your Clone? 471
Why Should You Select a Fusion System? 471
When Should You Avoid a Fusion System? 472
Is It Necessary to Cleave the Tag off the Fusion Protein? 474
Will Extra Amino Acid Residues Affect Your Protein of Interest *after* Digestion? 475
Working with Expression Systems 475
What Are the Options for Cloning a Gene for Expression? 475

Is Screening Necessary Prior to Expression? 476
What Aspects of Growth and Induction Are Critical to Success? 478
What Are the Options for Lysing Cells? 479
Troubleshooting 480
No Expression of the Protein 480
The Protein Is Expressed, but It Is Not the Expected Size Based on Electrophoretic Analysis 480
The Protein Is Insoluble, Now What? 481
Solubility Is Essential. What Are Your Options? 482
The Protein Is Made, but Very Little Is Full-Length; Most of It Is Cleaved to Smaller Fragments 483
Your Fusion Protein Won't Bind to Its Affinity Resin 484
Your Fusion Protein Won't Digest 485
Cleavage of the Fusion Protein with a Protease Produced Several Extra Bands 485
Extra Protein Bands Are Observed after Affinity Purification 485
Must the Protease Be Removed after Digestion of the Fusion Protein? 486
Bibliography 486

Over the past decade the variety of hosts and vector systems for recombinant protein expression has increased dramatically. Researchers now select from among mammalian, insect, yeast, and prokaryotic hosts, and the number of vectors available for use in these organisms continues to grow. With the increased availability of cDNAs and protein coding sequencing information, it is certain that these and other, yet to be developed systems will be important in the future. Despite the development of eukaryotic systems, *E. coli* remains the most widely used host for recombinant protein expression. *E. coli* is easy to transform, grows quickly in simple media, and requires inexpensive equipment for growth and storage. And in most cases, *E. coli* can be made to produce adequate amounts of protein suitable for the intended application. The purpose of this chapter is to guide the user in selecting the appropriate host and troubleshooting the process of recombinant protein expression.

EXPRESSION VECTOR STRUCTURE

What Makes a Plasmid an Expression Vector?

Vectors for expression in *E. coli* contain at a minimum, the following elements:

Table 15.1 Characteristics of Popular Prokaryotic Promoters

Promoter		Regulation/Inducer (Concentration)	Strength
LacUV5	Lactose operon	*lacI*/IPTG (0.1–1 mM)	Strong
Trp	Tryptophan operon	*trpR* 3-beta-indoleacrylic acid	Strong
Tac	Hybrid of –35 Trp and –10 lac promoter	*lacI*/IPTG (0.1–1 mM)	Strong
P_L	Phage lambda	Lambda *cI* repressor/heat	Strong
Phage T5	T5 phage	*lacI* (2 operators)/ IPTG (0.1–1 mM)	Strong
Arabinose	Arabinose operon	*AraBAD*/arabinose (1 μm–10 mM)	Variable
T7	T7 phage RNA polymerase	*lacI*/IPTG (0.1–1 mM)	Very strong

- A transcriptional promoter.
- A ribosome binding site.
- A translation initiation site.
- A selective marker (e.g., antibiotic resistance).
- An origin of replication.

In general, things that affect these can affect the level of protein expression. At a minimum, transcription promoters in *E. coli* consist of two DNA hexamers located –35 and –10 relative to the transcriptional start site. Together these elements mediate binding of the about 500 kDa multimeric complex of RNA polymerase.

Suppliers of vectors for expression have selected highly active, and inducible promoter sequences, and there is usually little need to be concerned until a problem is encountered in expression. A list of the commonly used promoters and their regulation is shown in Table 15.1.

Is a Stronger Promoter Always Desirable?

A strong promoter may not be best for all situations. Overproduction of RNA may saturate translation machinery, and maximizing RNA synthesis may not be desirable or necessary. A weaker promoter may actually give higher steady-state levels of soluble, intact protein than one that is rapidly induced.

Why Do Promoters Leak and What Can You Do about It?

Most promoters will have some background activity. Promoters regulated by the lactose operator/repressor will drive a small amount of transcription in the absence of added inducer (e.g., IPTG). To minimize this leakage, 10% glucose can be added to the medium to repress the lactose induction pathway, the growth temperature can be reduced to 15 to 30°C, and a minimal medium that contains no trace amounts of lactose can be used. Promoter leakage is only a problem when the expressed protein is highly toxic to the cells.

The tightly regulated T7 promoter has very low background due to the low levels of T7 RNA polymerase made in the absence of inducer (in specifically engineered host cells such as BL21 (DE3)/pLysS). It has been estimated that the fold induction of transcription in the T7 driven pET vector system is greater than 1000, while the magnitude of induction obtained with lac repressor regulated promoters is generally about 50-fold.

What Factors Affect the Level of Translation?

Translation can be affected by nucleotides adjacent to the ATG initiator codon, the amino acid residue immediately following the initiator, and secondary structures in the vicinity of the start site. Most commercially available vectors for expression use optimal ATG and Shine-Dalgarno sequences. Secondary structures in the mRNA contributed by the gene of interest can prevent ribosome binding (Tessier et al., 1984; Looman et al., 1986; Lee et al., 1987). In addition, the downstream box AAUCACAAAGUG found after the initiator codon in many bacterial genes can also enhance translation initiation. Conversion of the amino terminal sequence of the gene of interest to one that comes close to this consensus may improve the rate of translation of the mRNA (Etchegaray and Inouye, 1999).

What Can Affect the Stability of the Protein in the Cell?

One of the first steps in protein degradation in *E. coli* is the catalyzed removal of the amino terminal methionine residue. This reaction, catalyzed by methionyl aminopeptidase, occurs more slowly when the amino acid in the +2 position has a larger side chain (Hirel et al., 1989; Lathrop et al., 1992). When the methionine residue is intact, the protein will be stable to all but endopeptidase cleavage. Tobias et al. (1991) have determined the relationship between a protein's amino terminal amino acid

and its stability in bacteria, that is, the "N-end rule." They reported protein half-lives of only 2 minutes when the following amino acids were present at the amino terminus: Arg, Lys, Phe, Leu, Trp, and Tyr. In contrast, all other amino acids conferred half-lives of >10 hours when present at the amino terminus of the protein examined. This suggests that one should examine the sequence to be expressed for the residue in the +2 position. If the residue is among those that destabilize the protein, it may be worth the effort to change this residue to one that confers stability.

WHICH PROTEIN EXPRESSION SYSTEM SUITS YOUR NEEDS?

Track Record

What systems are currently used in the laboratory or by others in the field? If the protein coding sequence of interest is well characterized, and the protein or its close relatives have been expressed successfully by others in the field, it is wise to try the same expression system. Go with what has worked in the past. If nothing else, results obtained using the familiar system will serve as a starting point. As an example, most of the recombinant expression of mammalian src homology SH2 protein interaction domains has been done using the pGEX vector series, and similar examples of preferred systems are found in other fields of research. If little is known about the protein to be expressed, it is best to take stock of what information there is before entering the lab. Before beginning any experimentation, it is wise to answer the following question:

What Do You Know about the Gene to Be Expressed?

Source

In general, simple globular proteins from prokaryotic and eukaryotic sources are good candidates for expression in *E. coli.* Monomeric proteins with few cysteines or prosthetic groups (e.g., heme and metals) and of average size (<60 kDa) will likely give good production. Secreted eukaryotic proteins and membrane-bound proteins, especially those with several transmembrane domains, are likely to be problematic in *E. coli*. Solubility of recombinant proteins in *E. coli* can also be estimated by a mathematical analysis of the amino acid sequences (Wilkinson and Harrison, 1991).

Presence of a Start Codon

Some expression vectors provide the start codon for translation initiation, while others rely on the start codon of the gene you're trying to express. Note that in *E. coli*, 5 to 12 base pairs or less separate the ribosome binding site and the start codon. So you would incorporate this requirement into your cloning strategy when the start codon is provided by the gene you plan to express.

GC Content

Coding sequences with high GC (>70%) content may reduce the level of expression of a protein in *E. coli*. Check the sequence using a DNA analysis program.

Codon Usage

Codon usage may also affect the level of protein expression. If the gene of interest contains codons not commonly used in *E. coli*, low expression may result due to the depletion of tRNAs for the rarer codons. When one or more rare codons is encountered, translational pausing may result, slowing the rate of protein synthesis and exposing the mRNA to degradation. This potential problem is of particular concern when the sequence encodes a protein >60 kDa, when rare codons are found at high frequency, or when multiple rare codons are found over a short distance of the coding sequence. For example, rare codons for arginine found in tandem can create a recognition sequence for ribosome binding (e.g., _AGGAGG) that closely approximates a Shine-Dalgarno sequence UAAGGAGG. This may bind ribosomes non-productively and block translation from the bona fide ribosome binding site (RBS) at the initiator codon further upstream. Nonetheless, the appearance of a rare codon does not necessarily lead to poor expression. It is best to try expression of the native gene, and then make changes if these seem warranted later. Strategies include mutating the gene of interest to use optimal codons for the host organism, and co-transforming the host with rare tRNA genes. In one example, introduction into the *E. coli* host of a rare arginine (AGG) tRNA resulted in a several-fold increase in the expression of a protein that uses the AGG codon (Hua et al., 1994). In another case, substitution of the rare arginine codon AGG with the *E. coli*-preferred CGU improved expression (Robinson et al., 1984). Other work has shown that rare codons account for decreased expression of the gene of interest in *E. coli* (Zhang, Zubay, and Goldman, 1991; Sorensen, Kurland, and Pederson, 1989). Rare codons may have an even more dramatic

effect on translation when they occur close to the initiator codon (Chen and Inouye, 1990). While codon usage is not the only or most important factor, be aware that it may influence translation efficiency.

Secondary Structure

Secondary structures that occur near the start codon may block translation initiation (Gold et al., 1981; Buell et al., 1985), or serve as translation pause sites resulting in premature termination and truncated protein. These can be found using DNA or RNA analysis software. Structures with clear stem structures greater than eight bases long may be disrupted by site-specific mutation or by making all or a portion of the coding sequence synthetically.

Depending on the size of the gene, and the importance of obtaining high-expression levels, it may be worth synthesizing the gene. This has been generally done by synthesizing overlapping oligonucleotides that when annealed can be extended using PCR and ligated to form the full-length coding sequence. There are several examples where this approach has been used to optimize codon usage for *E. coli* (Koshiba et al., 1999; Beck von Bodman et al., 1986). In addition, if one takes on the work and expense of synthesizing a gene, secondary structures in the predicted RNA that might stall translation can be removed, and sites for restriction endonucleases can be introduced.

Size of a Gene or Protein

As a rule, very large (>100 kDa) and very small (<5 kDa) proteins are more difficult to express in *E. coli*. Small polypeptides with little secondary structure tend to be rapidly degraded in *E. coli*. Degradation can be minimized by expressing such short oligopeptides as concatemers with proteolytic or chemical cleavage sites in between the monomeric units (Hostomsky, Smrt, and Paces, 1985). Short peptides are also successfully expressed as fusion proteins. Fusion with GST, MalB or other larger, well-folded partners will tend to stabilize a short peptide, making expression possible and purification relatively simple. One publication has shown MBP to be superior to other large fusion proteins at stabilizing short polypeptides (Kapust and Waugh, 1999). At the other extreme, proteins that are above 60 kDa are best made using smaller affinity tags, such as FLAG, his_6, or on their own, without any fusion. While there is no clear upper limit, the larger the protein, the lower the yield is likely to be.

What Do You Know about Your Protein?

Cysteines

There are many things that *E. coli* does not do well, or at all. If the protein of interest is naturally multimeric, or requires post-translational modifications for activity, *E. coli* as an expression host may be a poor choice. Disulfide bonds, formed between two cysteines in an expressed protein, are made inefficiently in the reducing environment of the *E. coli* cytoplasm (Bessette et al., 1999; Derman et al., 1993). If the protein is produced, and can be purified from *E. coli*, in vitro oxidation of the cysteines may be tried (Dodd et al., 1995). Alternatively, the gene of interest can be cloned in a vector that includes a signal sequence (e.g., OmpA, geneIII, and phoA) that will direct the recombinant protein to the relatively oxidizing environment of the periplasm of *E. coli*, where disulfide formation is more efficient. Strains of *E. coli* that are deficient in thioredoxin reductase (trxB) permit proper disulfide formation in the cytoplasm (Derman et al., 1993; Yasukawa et al., 1995). Subsequent work has produced strains that lack both trxB and glutathione oxidoreductase and give better rates of disulfide formation than those seen in native *E. coli* periplasm (Bessette et al., 1999).

Membrane Bound

If the protein to be expressed is naturally associated with membrane and/or has at least one transmembrane domain, addition of a secretion signal to the amino terminus may help to maximize expression of functional protein. Signal sequences, about 20 residues long are derived from proteins that naturally are secreted into the periplasmic space, such as pelB, OmpA, OmpT, MalE, alkaline phosphatase (phoA), or geneIII of filamentous phage (Izard and Kendall, 1994). Protein with an amino terminal signal will be directed to the inner membrane of *E. coli*, and the carboxy terminal portion of the protein will be translocated into the periplasmic space. Depending on the hydrophobicity of the protein of interest, it may not translocate entirely into the periplasm but remain associated with the inner membrane. Secretion may help stabilize proteins from proteolytic attack (Pines and Inouye, 1999), or at least can reduce aggregation of hydrophobic proteins in the cytoplasm, and minimize inclusion body formation. Because of the reducing environment of the periplasmic space, proteins that contain one or more disulfide bonds are best secreted.

The presence of an N-terminal signal sequence appears to

be necessary but not sufficient to direct a target protein to the periplasm. Translocation across the outer membrane and into the growth medium is inefficient. In most cases target proteins found in the growth medium are the result of damage to the cell envelope and do not represent true secretion (Stader and Silhavy, 1990). Translocation across the inner cell membrane of *E. coli* is incompletely understood (reviewed by Wickner, Driessen, and Hartl, 1991), and the efficiency of export will depend on the individual target protein. Currently the export cannot be predicted based on protein sequence, although some generalizations have been made about the sequence immediately following the signal peptide (Boyd and Beckwith, 1990; Yamane and Mizushima, 1988). Therefore it is possible to find target proteins in the cytoplasm (with uncleaved signal sequence) or in the periplasm in partially processed form, in place of or in addition to the expected periplasmic processed species. In some cases the proportion of protein that is exported can be increased by lowering the temperature 15 to 30°C during induction.

Post-translational Modification

E. coli does not glycosylate or phosphorylate proteins or recognize proteolytic processing signals from eukaryotes, so take this into account when designing the cloning strategy. If proteolytic processing is needed, it is best to express only the coding sequences for the fully processed protein. If the protein of interest requires glycosylation for activity, and full activity is important in the final use, consider a eukaryotic host, such as Pichia, insect cells, or mammalian cells.

Is the Protein Potentially Toxic?

Consider whether the protein of interest is likely to have a toxic effect on the host cell. Where the function of the protein is known, this can be guessed at with some accuracy. For example, nonspecific proteases, nucleases, or pore-forming membrane proteins might all be expected to have some toxic effect on *E. coli*. Expression of toxic proteins may be very low, and there will be strong selective pressure on cells to eliminate the gene of interest by point mutation to change the translation frame, insertion of a stop codon, or change in an amino acid residue critical to the protein's function. Larger deletion of parts of the plasmid may also be seen. If there is a suggestion that the gene product will be toxic, use an expression vector with a tightly regulated promoter (e.g., T7, pET

vectors). Minimize propagation of the cells to avoid opportunities for mutation and recombination.

Must Your Protein Be Functional?

Each requirement placed on a recombinant protein will affect the choice of expression system. If a protein is to be used only to prepare antibody, it need not be soluble or active, and the production of inclusion bodies (aggregates of improperly folded protein) in *E. coli* may be all that is needed. Alternatively, if a protein's biological activity will be assayed, or if it is to be used in structural studies (NMR, crystallography, etc.), a properly folded and soluble form will be required.

Will Structural Changes (Additional or Fewer Amino Acids) Affect Your Application?

Depending on the way that a gene is inserted in an expression vector, additional sequences may be added to the clone, and these may lead to extra amino acid residues at the N- or C-termini of the final expressed protein. In many cases these will have no deleterious effect, but if structural studies or precise comparisons to a native protein are to be done, it is wise to eliminate amino acids added by cloning steps. PCR amplification is the most commonly used method to generate inserts for expression, and proper design of PCR primers can eliminate most or all additional residues in the protein.

Is the Sequence of Your Protein Recognized by Specific Proteases?

If you plan to express your gene in a fusion vector that provides an internal protease cleavage site for removal of the affinity tag (discussed below), check that your native protein is not recognized by the protease. Most proteases are highly specific, but thrombin has a variety of secondary cleavage sites (Chang, 1985).

Advertisements for Commercial Expression Vectors Are Very Promising. What Levels of Expression Should You Expect?

There are several systems available for protein expression in mammalian, insect, yeast, and *E. coli*. While it is impossible to predict the yields of protein from these systems for any given protein, some rough guidelines can be given. For any vector it is possible that no expression will be seen! Reported yields in stably transfected mammalian cells are in the range of 1 to $100\,\mu g/10^6$

cells. Insect cell systems will yield between 5 and 200 mg/L of culture (Schmidt et al., 1998), Pichia can produce up to 250 mg/L (Eldin et al., 1997), and reported yields in *E. coli* range from 50 μg to over 100 mg/L. Usually yields of from 1 to 10 mg/L can be expected from *E. coli*. Higher yields, up to a gram or more per liter, can be had using fermentation vessels where oxygen and pH levels can be controlled throughout the cell growth. The above-mentioned values are guidelines; they are entirely dependent on the protein to be expressed. It is always best to test one or more systems in parallel to select the best solution.

Nonbiological synthesis of protein is now possible as an alternative to production in a host organism (Kochendoerfer and Kent, 1999). Oligopeptides are synthesized and then assembled by chemical ligation to give full-length protein. The method has the potential to synthesize gram quantities of >30 kDa proteins, and such preparations would of course be free of host contaminants that might interfere with function or use in diagnostic or therapeutic applications. Unfortunately, chemical synthesis of proteins is not widely available.

Which *E. coli* Strain Will Provide Maximal Expression for Your Clone?

The choice of an expression host depends on the promoter system to be used. Promoters that depend on *E. coli* RNA polymerase can be expressed in most common cloning strains, while T7 promoter vectors must be used in *E. coli* that co-express T7 RNA polymerase (e.g., strains that contain the DE3 lysogen) (Dubendorff and Studier, 1991). Strains that are protease deficient (Bishai, Rappuoli, and Murphy, 1987) or overexpress chaparones have been shown to be useful for some proteins (Georgiou and Valax, 1996; Gilbert, 1994). At a minimum, a recombination deficient strain is advisable. Vendors of the commercially available *E. coli* expression vectors generally will recommend a host for use in expression. As with many questions related to protein expression, the results will depend on the nature of the protein of interest. A given gene may give high yields of intact protein in most strains, while the next would show no product except in a protease deficient host.

Why Should You Select a Fusion System?

Increased Yields

There are several reasons that one would choose to use a fusion system. Translational initiation from the amino terminal fusion

partner may be more efficient than the start contributed by the protein of interest, so larger amounts of protein can be obtained as a fusion. In addition smaller proteins (<20kDa), or subfragments of larger ones often benefit from association with a stable fusion partner, due in part to improved folding or protection from proteolysis. Fusion with GST, MBP, and thioredoxin may be useful for this purpose.

Simplified Purification and Detection

Most of the commonly available fusion partners double as affinity tags, and these make isolation of the protein of interest relatively simple. Protein can often be purified to >90% in a single step. In contrast to conventional chromatographic techniques, little or no information about the sequence, pI, or other physical characteristics of the protein is needed in order to perform the purification. Novice chromatographers or those who have not developed methods for purification of the native protein are advised to begin with an affinity system.

Detection of fusion proteins is a simple matter, since antibodies and colorimetric substrates are available for several of the more common fusion partners. Thus, if there is no established method to detect the protein, detection of the fusion partner can be the most convenient way to assay for the presence of the protein in cells and throughout purification and assay of the protein of interest.

When Should You Avoid a Fusion System?

Since affinity tags make purification relatively simple, and tags can be removed by proteolyic cleavage, use of a tag usually makes sense. If, on the other hand, a nonfusion vector has been used in earlier work, and one wishes to compare results with older data, use the nonfusion system. If there is an established method for purification and a biochemical assay or antibody available to detect the protein of interest, an affinity partner or tag for detection may simply be unnecessary. Ask again what use the protein will be put to. If the end application is likely to be sensitive to the presence of the tag (e.g., NMR, crystallography, therapeutics), and other conditions above are met, there is reason to avoid the tag.

If a fusion affinity tag is desired, several are available. Table 15.2 summarizes some of the characteristics of the most widely used fusion partners.

Table 15.2 Commercially Available Fusion Systems

Tag	Tag Size	Purification	Detection	Cleavage
Calmodulin/CBP	(CBP, 4 kDa)	Calmodulin-agarose EGTA for elution	Biotinylated calmodulin and streptavidin alkaline phosphatase	Thrombin, enterokinase
Chitin binding domain (CBD)	*Bacillus circulans* chitin binding domain (CBD, 52 amino acid residues)	chitin beads	Anti-CBD antibody	Used with intein. On-column cleavage is induced at 4°C by DTT or 2-mercaptoethanol.
E-tag	1.4 kDa	Anti-E sepharose	Anti-E antibodies	NA
FLAG®	1 kDa	Anti-Flag resin	Anti-FLAG antibodies	Enterokinase
Glutathione *S*-transferase GST	26.5 kDa homodimer GST forms a 58 kDa homodimer with two GSH binding sites. The affinity of the enzyme for GSH is approximately 0.1 μM.	Glutathione sepharose/ Glutathione Agarose	Anti-GST antibodies, CDNB substrate	Thrombin Factor Xa PreScission™ protease
HA (hemagglutinin)	~1 kDa YPYDVPDYA	NA	Anti-HA antibodies	
His_6	1 kDa	NTA-agarose, Iminodiacetic acid-sepharose	Anti-His_6 antibodies	Enterokinase, if desired
Maltose binding protein	42.5 kDa K_d of MBP for maltose is 3.5 μM; for maltotriose, 0.16 μM (Miller et al., 1983)	Amylose beads	Anti-MBP	Factor Xa
Myc tag	10 amino acids from human c-Myc EQKLISEEDL	Anti-Myc antibody resin	Anti-Myc antibodies (9E10)	NA
Nus-tag	E. coli NusA protein, 495 amino acids	NA	None	Thrombin
Pinpoint™	12.5 kDa peptide biotinylated *in vivo* (Samols et al., 1988)	Monomeric avidin resin (SoftLink™ soft release avidin resin)	Avidin/strep tavidin conjugates	Factor Xa
S-tag	15 amino acid peptide (S-tag) with strong affinity (K_d = 10^{-9} M) for a 104 amino acid fragment of	S-protein agarose beads	S-protein FITC conjugate	Thrombin, enterokinase

Table 15.2 (*Continued*)

Tag	Tag Size	Purification	Detection	Cleavage
	pancreatic ribonuclease A.			
Strep-tag	A 10 amino acid sequence that binds streptavidin	Streptavidin bead	Streptavidin conjugates	
Z-domain	Two Z domains add a 14 kDa peptide	IgG-sepharose		Factor Xa

Susceptibility To Cleavage Enzymes

As discussed below, some fusion systems allow for the removal of the affinity tag by specific proteolytic or chemical cleavage. Before beginning any experiment, examine the sequence of the protein to be cloned and expressed. The protein of interest may have a binding site for one of the proteases listed in Table 15.3, and if so, this site should be avoided, or a different expression system might be required. Most proteases used for cleavage of fusion protein are quite specific, with theoretical frequencies of 10^{-6}. However, it is best to check as a matter of course.

Is It Necessary to Cleave the Tag off the Fusion Protein?

For many proteins, cleavage is not needed. If the goal of the work is to raise an antibody, the whole fusion protein can be used successfully as antigen—provided that antibodies to the tag do not interfere in the application. If, on the other hand, the protein is to be used in structural studies, or where the function of recombinant protein will be compared with native protein, it may be necessary to remove the fusion tag.

Systems have been developed that use chemical (Nilsson et al., 1985) or specific proteolytic cleavage to separate the protein of interest from the fusion tag. The proteases have the advantage that cleavage is done at near neutral pH and at 4 to 37°C. In addition to proteolytic cleavage, the use of self-splicing inteins has been developed and commercialized by New England Biolabs. In this latter case fusion proteins with chitin-binding domain are bound to high molecular weight chitin chromatography media and incubated in the presence of a reducing agent, generally overnight. Protein splicing takes place, leaving the protein of interest in the flow through, while chitin and the spliced peptide remain bound.

Table 15.3 Characteristics of Popular Fusion Protein Cleavage Enzymes

Protease	Cleavage Site	Comment
Thrombin	?VPR^GS secondary cleavage sites exist; (Chang, 1985)	Widely used, works at 1:1000–1:2000 mass ratio relative to target protein. Purified from bovine sources and may include other proteins.
Factor Xa protease	IEGR^	Leaves defined N-terminus. Works at 1:500–1:1000 mass ratio relative to target protein. Recognition site with proline immediately following Arg residue will not be cleaved.
Enterokinase	DDDDK^	Leaves defined N-terminus. Recombinant.
rTEV	ENLYFQ^G	Recombinant endopeptidase from the Tobacco Etch Virus.
Intein-mediated self-cleavage		No added protease required. Leaves defined N-terminus
PreScission protease	LEVLFQ^GP	Rhinoviral 3C protease expressed as GST fusion protein. Optimal activity at 4°C.

Recognition sites for enzymes commonly used to cleave fusion proteins, and their advantages/disadvanatges are listed in Table 15.3

Will Extra Amino Acid Residues Affect Your Protein of Interest *after* Digestion?

Depending on the protease, and the way in which the protein of interest was cloned in the expression vector, there may be one or more nonnative residues left at the amino terminal of the protein of interest following cleavage. Whether or not this poses a problem depends entirely on the protein and the use to which it will be put. Even the most demanding applications may not be negatively affected by the presence of extra amino terminal residues. Wherever possible, it is best to design a cloning strategy that at least minimizes the number of these residues, and if relatively inoccuous residues (e.g., glycine, serine) can be introduced, all the better.

WORKING WITH EXPRESSION SYSTEMS

What Are the Options for Cloning a Gene for Expression?

In some cases the protein of interest is already cloned in another vector, for example, in a clone isolated from a cDNA

expression library. If the frame of the insertion is known, and compatible restriction sites are found in the expression vector(s) selected, the insert can be cloned directly. In some cases excision from a lambda vector can generate a plasmid vector ready for expression of the insert, without any manipulation at all.

More commonly PCR is used to amplify the target sequence using oligonucleotide primers that have 15 to 20 bases of homology with the 5′ and 3′ ends of the target. These primers will have in addition tails that encode restriction enzyme sites compatible with the expression vector. The PCR products can be digested with the appropriate restriction enzymes, purified, and ligated into an appropriately prepared vector.

The efficiency of cloning can be improved if two different restriction enzyme sites are available. This will allow for directional cloning of inserts into the vector, and all of the clones screened should have the insert in the desired orientation. Please refer to Chapter 9, "Restriction Endonucleases" for a discussion on double digestion strategies. If PCR is used to generate the insert, then primers must be designed appropriately. It is important to leave 4 to 6 random bases at the 5′ end of each PCR primer. These provide a spacer at the ends of the PCR product and allow the restriction enzymes to digest the DNA more efficiently. While *in vitro* ligation is still the most widely used method, ligation independent cloning (LIC) (Li and Evans, 1997) has the advantage that no DNA ligase is required (though an exonuclease activity is), and efficiencies are comparable to those obtained with conventional ligation with T4 DNA ligase.

Is Screening Necessary Prior to Expression?

There are no guarantees that the gene to be expressed will be present in the cell after transformation. As discussed above, most expression vectors are prone to produce small amounts of the protein even in the absence of inducing agent, which can prove toxic to the host. Alternatively, host cells can cause deletions and rearrangements in the expression vector. Either way, it is usually a very good idea to confirm the presence of the inserted gene prior to expression experiments.

Unless a library of clones is to be prepared, the efficiency of ligation and transformation is rarely an issue. Screening of a dozen clones for the presence of an insert should be sufficient to identify one or more positive candidate clones.

The first step is generally to prepare several plasmid DNA minipreps and digest the DNA with the same enzyme(s) used in

cloning to generate the insert. Products should be analyzed by agarose gel electrophoresis to determine if DNA of the predicted size was inserted in the vector. As an alternative, PCR can be done using as template a small scraping from a colony on the plate. Amplification of the plasmid DNA contained in the cells using the same primers used in cloning, or primers that anneal to flanking vector sequences, should show a band of the predicted size. This latter method does not confirm the presence of the restriction sites used in cloning, but has the advantage of being rapid.

Once the presence of an insert of the correct size is confirmed, the DNA sequence at the cloning junctions should be determined. It is not uncommon for a primer sequence to be synthesized with an error—whether by faulty design or at the hands of the oligo supplier. DNA sequencing to confirm the cloning junctions should be done in parallel with a small-scale expression experiment, in which a 1 to 2 ml culture is grown and induced according to a standard protocol. It is important to include a culture that is transformed with the parent expression vector as a negative control in this screening experiment. Following centrifugation, the cell pellet should be suspended in SDS-PAGE loading buffer, and a small amount loaded on an SDS-acrylamide gel. The viscosity of the whole cell lysate (caused by the release of genomic DNA) may make gel loading difficult. However, addition of extra 1× loading buffer, DNaseI (10 μg/ml), extended heating of the sample, or sonication should alleviate the problem.

After electrophoresis, the gel should be stained (e.g., Coomassie Blue) to visualize the proteins in the whole cell lysate. If expression is good, an induced band will clearly be seen at the predicted molecular weight, and this will be absent in the no-insert control culture. If no band is visible and the restriction digestion/DNA sequencing data indicate that all is well, don't despair. Perform an immunoblot of an SDS-acrylamide gel. Screen for the presence of the protein of interest or use an antibody directed against the affinity or epitope tag if one has been used. Use of both N- and C-terminal specific antibodies is ideal in troubleshooting. Be sure to include both positive and negative controls in the immunoblot. Alternatives to immunblotting include ELISA or specific biochemical assays for the protein of interest.

If an antibody is not available for Western blotting, and you have a procedure to purify your protein, attempt the purification. This can visualize a protein that is present in quantities insufficient to stand out on a PAGE gel of a total cell extract.

Once expression of a protein of the predicted molecular weight is found, minimize propagation of the cells. Serial growths under

conditions that permit expression may lead to plasmid loss or rearrangements.

Once analysis is complete, glycerol stocks of positive clones should be prepared. This can be done by streaking culture residue from the DNA miniprep on a plate to get a fresh colony, by reusing the colony that was originally picked, or by re-transforming *E. Coli* with isolated miniprep DNA. In either case a fresh colony should be used to prepare a 2 to 4 ml log-phase culture for the purpose of making a glycerol stock. Be sure to keep protein expression repressed during this step by reduced temperature, use of minimal medium, or adapting it to the vector in use.

What Aspects of Growth and Induction Are Critical to Success?

Aeration, Temperature

The best expression results are had when cultures are grown with sufficient aeration and positive selection for the plasmid. For small-scale experiments, use 2 ml of medium (e.g., LB, SOC or 2XTY) in a 15 ml culture tube. Vigorous shaking (>250 rpm) should be used to maintain aeration. Appropriate antibiotics, such as ampicillin should be added to recommended levels. At larger scales, Ehrlenmeyer flasks should be used. Flasks with baffles improve aeration and $\frac{1}{8}$ to $\frac{1}{2}$ of the total volume of the flask should be occupied by medium. Good results may be obtained using 250 ml to 1 L in a 2 L baffled flask.

Scaling Up

When scaling up growth, monitor the light scattering at 590 or 600 nm. Note that a culture with OD_{600} of one corresponds to about 5×10^8 cells/ml, though this number will vary depending on the strain of *E. coli* used. Two rules of thumb are particularly important: minimize the time in each stage of growth, and monitor both cell density and protein expression at each stage.

From a colony or glycerol stock, begin a small overnight culture (e.g., 2–5 ml) in a selective medium under conditions that repress expression. Don't allow the culture to overgrow. This starter culture is then used to inoculate a larger volume of medium at a volume ratio of about 1:100 (pre-warming the media is a good idea). Monitor the growth by absorbance at 600 nm, and keep the cell density low (OD_{600} below 1). Once the growth has been scaled to give sufficient starter for the final growth vessel, make an inoculum of about 1%. Monitor the OD every 30 minutes or so, and remove aliquots for analysis by SDS-PAGE, immunoblotting, or

functional assay. After an initial lag following the inoculation, the density of the cells should double every 20 to 40 minutes. A graph of the OD coupled with an immunoblot is very useful in selecting optimal conditions for the growth. Once the culture reaches a late log phase (usually about OD_{600} of 0.8–1.2), induction is done by the addition of the appropriate inducing agent. Continue to monitor growth and take aliquots. It is not unusual that cells expressing a foreign protein will either stop growing or show a 10% to 20% decrease in density following induction. While it is common to grow for 1 to 3 hours postinduction prior to harvest, this induction period can vary depending on temperature and other conditions. So it is best determined empirically.

What Are the Options for Lysing Cells?

E. coli are easily broken by several methods including decavitation, shearing, and the action of freeze–thaw cycles. The choice of method depends on the scale of growth, and the type of equipment available (reviewed in Johnson, 1998). For most lab-scale experiments, sonication, or freeze–thaw will be the most practical choices. Ultrasonic distruptors are available from many vendors, but all operate on the conversion of electrical energy through piezoelectric transducers into ultrasonic waves of 18 to 50 kHz. The vibration is transferred to the sample by a titanium tip, and the energy released causes decavitation and shearing of the cells. Several models are available that are microprocessor controlled, programmable, and allow very reproducible cell lysis. It is important to keep the sample on ice and avoid frothing. This latter problem is caused by a probe that is not immersed sufficiently in the sample, or by excessive power. If bubbles begin forming and accumulating on the surface, stop immediately, reposition the probe, and reduce output. Once a sample has been turned to foam, sonication will be ineffective, and there is little to do but start again. Even if frothing is not seen, treatment beyond that needed to cause cell lysis can result in physical damage to the protein of interest. The addition of protease inhibitors to the cell suspension immediately prior to cell lysis is an important precaution, and several commercial cocktails are available for this purpose.

Freeze–thaw, particularly in conjunction with lysozyme and DNase treatment, is one of the mildest procedures to break *E. coli*. Cells are simply resuspended in buffer (PBS, Tris-pH 8.0) containing 10 μg/ml hen egg lysozyme, and the sample is cycled between a dry ice–alcohol bath and a container of tepid water. Generally, 5 to 10 cycles is sufficient to break nearly all of the cells.

As the cells lyse and DNA is liberated, it may be necessary to add DNase to 10 μg/ml to reduce the viscosity of the preparation. Commercial or homemade detergent preparations including *N*-octylamine are also very effective at lysing cells and simple to use.

Whatever method is used, lysis should be monitored. Microscopic examination is the best option. Retain some of the starting suspension, and compare to the lysate. Phase contrast optics will permit direct visualization, though staining can be used as well. Lysis will be evidenced by a slight darkening of the suspension, or clearing, and under the microscope, cells will be broken with membrane fragments or small vesicles present.

Other physical lysis methods include the use of a French Press, Manton-Gaulin, and other devices that place cells under rapid changes of pressure or shear force. These are very effective and reproducible, but generally, they are best used when the original culture volume is >1 L, since most of these cell disruptors have minimum volume requirements.

TROUBLESHOOTING

No Expression of the Protein

If one has checked for small-scale expression as discussed above, there should have been a detected band on a stained gel or immunoblot. If neither are seen, sequencing of the cloning junctions and entire insert should be undertaken to confirm that no frame shifts, stop codons, or rearrangements have occurred. Purification can be tried in parallel to see if even very low levels are made. A slight band on SDS-PAGE of the expected protein will make clear that the cloning went as planned, but the biology of expression is at fault. Varying temperature, time of induction, and the type of plasmid or fusion system can all be tried. In the end some proteins may not express well in *E. coli*, and they should be tried in other organisms.

The Protein Is Expressed, but It Is Not the Expected Size Based on Electrophoretic Analysis

On SDS-PAGE the net charge on the protein of interest will affect mobility. Highly charged proteins will tend to bind less SDS and will have retarded mobility. Proteins rich in proline may also exhibit dramatically slower mobility in SDS-PAGE. If the protein has a calculated pI in the range of 5 to 9, and is not strongly biased in amino acid composition, then a protein that shows multiple

bands or a strong species far from the predicted molecular weight is likely due to something other than an artifact of SDS-PAGE. Probing immunoblots with the appropriate antibodies to N- and/or C-terminal tags of the protein is particularly useful at this stage. Try to identify the halves or pieces of the protein on stained gels and immunoblots to locate likely points in the coding sequence where proteolytic cleavage and/or translation termination may occur. Cleavage at the junction between the protein of interest and the fusion partner (if any) that is used is often seen. Addition of protease inhibitors should be routine in all work, and protease-deficient strains should be tried in parallel or as a next step. If these measures fail, try re-cloning in another vector with a different fusion tag or tags, and promoter.

The Protein Is Insoluble. Now What?

Many heterologous proteins expressed in *E. coli* will be found as insoluble aggregates that are failed folding intermediates (Schein, 1989). Such inclusion bodies are seen as opaque areas in micrographs of *E. coli* that express the protein of interest. Depending on the purpose of expression, the production of inclusion bodies may be a welcome occurrence. If for example, the recombinant protein is to be used solely for the production of antibodies, inclusion bodies may be isolated to high purity by differential centrifugation and used directly as an antigen. If the protein is relatively small, the inclusion bodies may be isolated as above, and refolded with good efficiency. Other (particularly large) proteins will not refold well, and if production of functional protein is required, then an alternative must be found. Before proceeding, it is best to answer the following questions.

Are You Sure Your Protein Is Insoluble?

A first consideration is whether the protein is truly insoluble, or the cells were simply not lysed. Here is where microscopic examination will be of great use. Examine a cell lysate under phase contrast microscopy or after staining. Are intact cells visible? After it sediments, is the pellet large and similar in appearance to the original cell pellet? Is the post-lysis supernatant clear? Any of the above may indicate that cells are not completely disrupted. The protein of interest may be soluble but trapped in intact cells.

If cells are lysed as measured by microscopy, analyze whole cell lysate, clarified lysate, and post-lysis pellet by SDS-PAGE, followed by staining or immunoblotting. If cells are lysed as mea-

sured by microscopy, and the protein of interest is found in the post-lysis pellet, it is likely that it is being made in an insoluble form. While most use a relatively low-speed centrifugation step at around 10,000 × *g*, it is best to do a 100,000 × *g* spin to sediment all aggregates before drawing any conclusion about insolubility. Another indication is microscopic examination of cells under high power (>400×). If inclusion bodies are being made, and expression levels are high, optically dense areas in the *E. coli* cells will be seen. These inclusion bodies may occupy more than half of the cell.

Must Your Protein Be Soluble?

The accumulation of proteins in inclusion bodies is not necessarily undesirable. Insolubility has three important advantages:

1. Inclusion bodies can represent the highest yielding fraction of target protein.
2. Inclusion bodies are easy to isolate as an efficient first step in a purification scheme. Nuclease-treated, washed inclusion bodies are usually 75% to 95% pure target protein.
3. Target proteins in inclusion bodies are generally protected from proteolytic breakdown.

Isolated inclusion bodies can be solubilized by a variety of methods in preparation for further purification and refolding. If the application is to prepare antibodies, inclusion bodies can be used directly for injection after suspension in PBS and emulsification with a suitable adjuvant (e.g., Fischer et al., 1992). If the target protein contains a his_6-tag, purification can be performed under denaturing conditions. The purified protein can be eluted from the resin under denaturing conditions and then refolded.

Solubility Is Essential. What Are Your Options?

Prevent Formation of Insoluble Bodies

A number of approaches have been used to obtain greater solubility, including induction of protein expression at 15 to 30°C (Burton et al., 1991), use of lower concentrations of IPTG (e.g., 0.01–0.1 mM) for longer induction periods, and/or using a minimal defined culture medium (Blackwell and Horgan, 1991).

Solubilize and Refold

Solubilization and refolding methods usually involve the use of chaotropic agents, co-solvents or detergents (Marston and

Hartley, 1990; Frankel, Sohn, and Leinwand, 1991; Zardeneta and Horowitz, 1994). A strategy that has been successful for some proteins is to express as a his_6-tagged fusion, bind under denaturing conditions, and refold while protein is still bound to the resin by running a gradient from 6M to 0M guanidine-HCl in the presence of reduced (GSH) and oxidized (GSSH) glutathione. Once folding has occurred, elution is done with imidazole as usual. Some researchers enhance refolding of enzymes by the addition of substrate or a substrate analogue during gradual removal of denaturant by dialysis (Zhi et al., 1992; Taylor et al., 1992).

The Protein Is Made, but Very Little Is Full-Length; Most of It Is Cleaved to Smaller Fragments

It is important to distinguish among proteolytic breakdown, translation termination, and cryptic translation start sites within the gene of interest. Proteolytic breakdown is most likely to occur at exposed domains of the protein. Examine the pattern of breakdown products by SDS-PAGE, estimate their sizes, and compare the result with the predicted amino acid sequence. Keep an eye out for bends or surface-exposed regions, and any sequences that conform to those for known proteases. While protease inhibitors such as PMSF should be present in the sample prior to cell lysis, expand the group of protease inhibitors and test their effect. Also consider the pattern of expression seen when growth is monitored before and after induction. If there is a switch between intact and fragmented protein after induction, it is likely that proteolysis is the culprit.

Translation Termination

There is little clear-cut evidence for inappropriate translation termination, but in at least one case a stretch of 20 serine residues was suggested to cause premature termination in *E. coli* (Bula and Wilcox, 1996). If a truncated protein is definitely seen, DNA sequencing in the expected termination region should be done to confirm that no cryptic stop codons exist.

Cryptic translation initiation may be seen as well (Preibisch et al., 1988). Cryptic translation initiation can occur within an RNA coding sequence when a sequence resembling the ribosome binding site (AAGGAGG) occurs with the appropriate spacing (typically 5 to 20 nucleotides upstream of an AUG (Met) codon. These smaller products can be problematic when attempting to purify full-length proteins. If some expression of full-length

protein is seen, a useful strategy may be to try dual tag affinity purification, in which the gene of interest is expressed in a vector that encodes two affinity tags, one each at the C- and N-termini. Sequential purification using both affinity tags can give reasonable yields of full-length protein whatever the original cause (Kim and Raines, 1994; Pryor and Leiting, 1997).

Your Fusion Protein Won't Bind to Its Affinity Resin

A lysate is produced, and contacted with the affinity medium. The protein of interest is present in the cell and clarified lysate, as shown by SDS-PAGE, but after purification of the lysate over the medium, all of the protein is found in the flow-through. The presence of a large amount of protein in the eluate after an attempt to bind to the affinity medium does not prove an inability to bind. If there is a very large excess of protein, it may appear that none is binding, when in fact the column has simply been overloaded. Try to wash and elute the protein from the affinity medium before drawing a conclusion. One simple test is to remove 10 to 50 μl of the purification medium after binding and washing, and then boil the sample in an equal volume of 1 × SDS-PAGE loading dye. Gel analysis may show binding of the protein to the resin. Consideration of the amount loaded on the column and the expected capacity of the purification medium will sort out the various causes. If in fact expression is clearly seen in the lysate applied to the purification medium, there are other explanations:

1. The affinity medium was not equilibrated properly, or the protein folded to mask the residues responsible for binding to the affinity medium. Purification in the presence of detergents (e.g., 0–1% Tween-20), or mild chaotropes (e.g., 1–3 M guanidine-HCl or urea) may unmask these residues and enable binding.
2. Your fusion protein won't elute from its affinity resin. Protein may apparently bind to the resin, as measured by the presence of an SDS-PAGE gel band after boiling a sample of the washed resin. Little or no protein of interest may be eluted, however, when the loaded resin is contacted with eluting agent. In this latter case the protein may interact nonspecifically with the base matrix, or the protein precipitated during contact with the resin and is trapped. Addition of detergent, of varying ionic strength and pH, may improve the situation.

Your Fusion Protein Won't Digest

If expression is otherwise good, and the protein is not digested to *any* extent, one should confirm by DNA sequencing that the protease site is intact. Checking the activity of the protease in parallel experiments using a known and well-behaved protein will give some confidence that the protease itself is not to blame. If the site is present, the protease has activity, and buffer conditions are close to those specified for the protease, it may be that the fusion protein folds so that the protease site is inaccessible. Additives that alter the structure slightly, including salts and detergents may unmask the site; see Ellinger et al. (1991). Alternatively, recloning to create a flexible linker flanking the protease site has been shown to increase the efficiency of digestion with Thrombin (Guan and Dixon, 1991) and presumably other proteases.

Cleavage of the Fusion Protein with a Protease Produced Several Extra Bands

Cryptic Sites

The specificity of any protease is inferred from its natural substrates, and there is reason to believe that cryptic sites are also cleaved. (Nagai, Perutz, and Poyart, 1985; Eaton, Rodriguez, and Vehar, 1986; Quinlan, Moir, and Stewart, 1989; Wearne, 1990).

Excess Protease

If multiple bands are seen by SDS-PAGE, a titration of the amount, time and temperature of digestion should be done. Often reducing time or temperature will minimize cleavage at secondary sites, while retaining digestion at the desired site.

Extra Protein Bands Are Observed after Affinity Purification

E. coli host chaperone protein GroEL, with an apparent molecular weight of about 57 to 60 kDa on SDS-PAGE, is often found to co-purify with a protein of interest (Keresztessy et al., 1996) This may be caused by misfolding or by a recombinant protein that is trapped at an intermediate folding stage. High salt concentration (1–2 M), non-ionic detergents, and ligand or co-factors (e.g., ATP or GTP) may be effective in removing chaperones from the protein of interest. Often chaperones and other contaminating proteins are seen following affinity purification; they are best removed by conventional chromatography such as ion exchange.

Their co-purification can be minimized by inducing the culture at a lower density (e.g., OD_{600} = 0.3 vs. 1.0) or by reducing temperature.

Must the Protease Be Removed after Digestion of the Fusion Protein?

The removal of the protease is not necessary for many applications. Generally, protease is added at a ratio of 1:500 or lower relative to the protein of interest, so protease may not interfere with downstream applications. Biochemical assays and antibody production may not require removal, while structural studies, or assays where other proteins are added to the protein of interest in a biochemical assay indicate that a further purification be performed.

The commonly used serine proteases, thrombin and factor Xa, can be removed from a reaction mixture by contacting the digested protein/protease with an immobilized inhibitor such as benzamidine-sepharose (Sundaram and Brandsma, 1996). This purification is not complete due to the equilibrium binding of the inhibitor to the protease, but the majority of the protease can be removed in this way. Better yet, a different purification method like ion-exchange or hydrophobic interaction chromatography can be used to separate the protein of interest from both the protease and other cleavage products including the affinity tag.

Some commercially available proteases (Table 15.3) include affinity tags that can be used effectively to remove the protease from the sample. Biotinylated thrombin can be removed with high efficiency due to the extreme affinity of biotin for avidin or streptavidin-agarose beads. Other proteases containing affinity tags include PreScission protease; a fusion of GST with human rhinoviral 3C protease.

BIBLIOGRAPHY

Beck von Bodman, S., Schuler, M. A., Jollie, D. R., and Sligar, S. G. 1986. Synthesis, bacterial expression, and mutagenesis of the gene coding for mammalian cytochrome b5. *Proc. Nat. Acad. Sci. U.S.A.* 83:9443–9447.

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. Nat. Acad. Sci. U.S.A* 96:13703–13708.

Bishai, W. R., Rappuoli, R., and Murphy, J. R. 1987. High-level expression of a proteolytically sensitive diphtheria toxin fragment in *Escherichia coli*. *J. Bact.* 169:5140–5151.

Blackwell, J. R., and Horgan, R. 1991. A novel strategy for production of a highly expressed recombinant protein in an active form. *FEBS Lett.* 295:10–12.

Boyd, D., and Beckwith, J. 1990. The role of charged amino acids in the localization of secreted and membrane proteins. *Cell* 62:1031–1033.
Buell, G., Schulz, M. F., Selzer, G., Chollet, A., Movva, N. R., Semon, D., Escanez, S., and Kawashima, E. 1985. Optimizing the expression in *E. coli* of a synthetic gene encoding somatomedin-C (ICF-I). *Nucl. Acids Res.* 13:1923–1938.
Bula, C., and Wilcox, K. W. 1996. Negative effect of sequential serine codons on expression of foreign genes in *Escherichia coli*. *Prot. Expr. Purif.* 7:92–103.
Burton, N., Cavallini, B., Kanno, M., Moncollin, V., and Egly, J. M. 1991. Expression in *Escherichia coli*: Purification and properties of the yeast general transcription factor TFIID. *Prot. Expr. Purif.* 2:432–441.
Chang, J.-Y. 1985. Thrombin specificity: Requirement for apolar amino acids adjacent to the cleavage site of polypeptide substrate. *Eur. J. Biochem*. 151:217–224.
Chen, G. F., and Inouye, M. 1990. Suppression of the negative effect of minor arginine codons on gene expression; preferential usage of minor codons within the first 25 codons of the *Escherichia coli* genes. *Nucl. Acids Res.* 18:1465–1473.
Derman, A. I., Prinz, W. A., Belin, D., and Beckwith, J. 1993. Mutations that allow disulfide bond formation in the cytoplasm of *Escherichia coli*. *Science* 262:1744–1747.
Dodd, I., Mossakowska, D. E., Camilleri, P., Haran, M., Hensley, P., Lawlor, E. J., McBay, D. L., Pindar, W., and Smith, R. A. 1995. Overexpression in *Escherichia coli*: Folding, purification, and characterization of the first three short consensus repeat modules of human complement receptor type 1. *Prot. Expr. Purif*. 6:727–736.
Dubendorff, J. W., and Studier, F. W. 1991. Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with lac repressor. *J. Mol. Biol.* 219:45–59.
Eaton, D., Rodriguez, H., and Vehar, G. A. 1986. Proteolytic processing of human factor VIII: Correlation of specific cleavages by thrombin, factor Xa, and activated protein C with activation and inactivation of factor VIII coagulant activity. *Biochem.* 25:505–512.
Eldin, P., Pauza, M. E., Hieda, Y., Lin, G., Murtaugh, M. P., Pentel, P. R., and Pennell, C. A. 1997. High-level secretion of two antibody single chain Fv fragments by *Pichia pastoris*. *J Immunolog. Meth.* 201:67–75.
Ellinger, S., Mach, M., Korn, K., and Jahn, G. 1991. Cleavage and purification of prokaryotically expressed HIV gag and env fusion proteins for detection of HIV antibodies in the ELISA. *Virol.* 180:811–813.
Etchegaray, J. P., and Inouye, M. 1999. Translational enhancement by an element downstream of the initiation codon in *Escherichia coli. J. Biol. Chem.* 274:10079–10085.
Fischer, L., Gerard, M., Chalut, C., Lutz, Y., Humbert, S., Kanno, M., Chambon, P., and Egly, J. M. 1992. Cloning of the 62-kilodalton component of basic transcription factor BTF2. *Science* 257:1392–1395.
Frankel, S., Sohn, R., and Leinwand, L. 1991. The use of sarkosyl in generating soluble protein after bacterial expression. *Proc. Nat. Acad. Sci. U.S.A.* 88:1192–1196.
Georgiou, G., and Valax, P. 1996. Expression of correctly folded proteins in *Escherichia coli*. *Curr. Opin. Biotechnol.* 7:190–197.
Gilbert, H. F. 1994. Protein chaperones and protein folding. *Curr. Opin. Biotechnol.* 5:534–539.
Gold, L., Pribnow, D., Schneider, T., Shinedling, S., Singer, B. S., and Stormo, G. 1981. Translational initiation in prokaryotes. *Ann. Rev. Microbiol*. 35:365–403.
Guan, K. L., and Dixon, J. E., 1991. Eukaryotic proteins expressed in *Escherichia coli*: An improved thrombin cleavage and purification procedure of fusion proteins with glutathione *S*-transferase. *Anal. Biochem.* 192:262–267.

Hirel, P. H., Schmitter, M. J., Dessen, P., Fayat, G., and Blanquet, S. 1989. Extent of N-terminal methionine excision from *Escherichia coli* proteins is governed by the side-chain length of the penultimate amino acid. *Proc. Nat. Acad. Sci. U.S.A.* 86:8247–8251.
Hostomsky, Z., Smrt, J., and Paces, V. 1985. Cloning and expression in *Escherichia coli* of the synthetic proenkephalin analogue gene. *Gene* 39:269–274.
Hua, Z., Wang, H, Chen, D., Chen, Y., and Zhu, D. 1994. Enhancement of expression of human granulocyte-macrophage colony stimulating factor by argU gene product in *Escherichia coli*. *Biochem. Mol. Biol. Int*. 32:537–543.
Izard, J. W., and Kendall, D. A. 1994. Signal peptides: Exquisitely designed transport promoters. *Mol. Microbiol*. 13:765–773.
Johnson, B. 1998. Breaking up isn't hard to do: A cacophony of sonicators, cell bombs, and grinders. *Scientist* 12:23.
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. *Prot. Sci.* 8:1668–1674.
Kelley, R. F., and Winkler, M. E. 1990. Folding of eukaryotic proteins produced in *Escherichia coli*. *Genet. Eng.* (*NY*) 12:1–19.
Keresztessy, Z., Hughes, J., Kiss, L., and Hughes, M. A. 1996. Co-purification from *Escherichia coli* of a plant beta-glucosidase-glutathione *S*-transferase fusion protein and the bacterial chaperonin GroEL. *Biochem. J.* 314:41–47.
Kim, J. S., and Raines, R. T. 1994. Peptide tags for a dual affinity fusion system. *Anal. Biochem.* 219:165–166.
Kochendoerfer, G. G., and Kent, S. B. 1999. Chemical protein synthesis. *Curr. Opin. Chem. Biol.* 3:665–671.
Koshiba, T., Hayashi, T., Miwako, I., Kumagai, I., Ikura, T., Kawano, K., Nitta, K., and Kuwajima, K. 1999. Expression of a synthetic gene encoding canine milk lysozyme in *Escherichia coli* and characterization of the expressed protein. *Prot. Eng*. 12:429–435.
Lathrop, B. K., Burack, W. R., Biltonen, R. L., and Rule, G. S. 1992. Expression of a group II phospholipase A2 from the venom of *Agkistrodon piscivorus piscivorus* in *Escherichia coli*: Recovery and renaturation from bacterial inclusion bodies. *Prot. Expr. Purif.* 3:512–517.
Lee, N., Zhang, S. Q., Cozzitorto, J., Yang, J. S., and Testa, D. 1987. Modification of mRNA secondary structure and alteration of the expression interferon alpha 1 in *Escherichia coli*. *Gene* 58:77–86.
Li, C., and Evans, R. M. 1997. Ligation independent cloning irrespective of restriction site compatibility *Nucl. Acids Res.* 25:4165–4166.
Looman, A. C., Bodlaender, J., de Gruyter, M., Vogelaar, A., and van Knippenberg, P. H. 1986. Secondary structure as primary determinant of the efficiency of ribosomal binding sites in *Escherichia coli*. *Nucl. Acids Res.* 14:5481–5496.
Marston, F. A., and Hartley, D. L. 1990. Solubilization of protein aggregates. *Meth. Enzymol.* 182:264–282.
Miller III, D. M., Olson, J. S., Pflugrath, J. W., and Quiocho, F. A. 1983. Rates of ligand binding to periplasmic proteins involved in bacterial transport and chemotaxis. *J. Biol. Chem.* 258:13665–13672.
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.
Nagai, K., Perutz, M. F., and Poyart, C. 1985. Oxygen binding properties of human mutant hemoglobins synthesized in *Escherichia coli*. *Proc. Nat. Acad. Sci. U.S.A.* 82:7252–7255.
Nilsson, B., Holmgren, E., Josephson, S., Gatenbeck, S., Philipson, L., and Uhlen,

M. 1985. Efficient secretion and purification of human insulin-like growth factor I with a gene fusion vector in *Staphylococci*. *Nucl. Acids Res.* 13: 1151–1162.

Pines, O., and Inouye, M. 1999. Expression and secretion of proteins in *E. coli*. *Mol. Biotechnol*. 12:25–34.

Preibisch, G., Ishihara, H., Tripier D., and Leineweber, M. 1988. Unexpected translation initiation within the coding region of eukaryotic genes expressed in *Escherichia coli*. *Gene* 72:179–186.

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. *Prot. Expr. Purif.* 10:309–319.

Quinlan, R. A., Moir, R. D., and Stewart, M. 1989. Expression in *Escherichia coli* of fragments of glial fibrillary acidic protein: Characterization, assembly properties and paracrystal formation. *J. Cell Sci.* 93:71–83.

Robinson, M., Lilley, R., Little, S., Emtage, J. S., Yarranton, G., Stephens, P., Millican, H., Eaton, M., Humphries, G. 1984. Codon usage can affect efficiency of translation of genes in Escherichia coli. *Nucl. Acids Res.* 12:6663–6671.

Sambrook, J., Fritsch, E. F., and Maniatis, T. 1989. *Molecular Cloning: A Laboratory Manual*, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Samols, D., Thornton, C. G., Murtif, V. L., Kumar, G. K., Haase, F. C., and Wood, H. G. 1988. Evolutionary conservation among biotin enzymes. *J. Biol. Chem.* 263:6461–6464.

Schein C. H. 1989. Production of soluble recombinant proteins in bacteria. *Biotechnol.* 7:1141–1149.

Schmidt, M., Tuominen, N., Johansson, T., Weiss, S. A., Keinänen, K., and Oker-Blom, C. 1998. Baculovirus-mediated large-scale expression and purification of a polyhistidine-tagged rubella virus capsid protein. *Prot. Expr. Purif.* 12:323–330.

Sorensen, M. A., Kurland, C. G., and Pedersen, S. 1989. Codon usage determines translation rate in *Escherichia coli*. *J. Mol. Biol.* 207:365–377.

Stader, J. A., and Silhavy, T. J. 1990. Engineering *Escherichia coli* to secrete heterologous gene products. *Meth. Enzymol.* 185:166–187.

Sundaram, P., and Brandsma, J. L. 1996. Rapid, efficient, large-scale purification of unfused, nondenatured E7 protein of cottontail rabbit papillomavirus. *J. Virol. Meth.* 57:61–70.

Taylor, M. A., Pratt, K. A., Revell, D. F., Baker, K. C., Sumner, I. G., and Goodenough, P. W. 1992. Active papain renatured and processed from insoluble recombinant propapain expressed in *Escherichia coli*. *Prot. Eng.* 5:455–459.

Tessier, L. H., Sondermeyer, P., Faure, T., Dreyer, D., Benavente, A., Villeval, D., Courtney, M., and Lecocq, J. P. 1984. The influence of mRNA primary and secondary structure on human IFN-gamma gene expression in *E. coli*. *Nucl. Acids Res.* 12:7663–7675.

Tobias, J. W., Shrader, T. E., Rocap, G., and Varshavsky, A. 1991. The N-end rule in bacteria. *Science* 254:1374–1377.

Wearne, S. J. 1990. Factor X_a cleavage of fusion proteins: Elimination of nonspecific cleavage by reversible acylation. *FEBS Lett.* 263:23–26.

Wickner, W., Driessen, A. J., and Hartl, F. U. 1991. The enzymology of protein translocation across the *Escherichia coli* plasma membrane. *Ann. Rev. Biochem.* 60:101–124.

Wilkinson, D. L., and Harrison, R. G. 1991. Predicting the solubility of recombinant proteins in *Escherichia coli*. *Biotechnol.* 9:443–448.

Yamane, K., and Mizushima, S. 1988. Introduction of basic amino acid residues after the signal peptide inhibits protein translocation across the cytoplasmic

membrane of *Escherichia coli*: Relation to the orientation of membrane proteins. *J. Biol. Chem.* 263:19690–19696.
Yasukawa, T., Kanei-Ishii, C., Maekawa, T., Fujimoto, J., Yamamoto, T., and Ishii, S. 1995. Increase in solubility of foreign proteins in *E. coli* by coproduction of the bacterial thioredoxin. *J. Biol. Chem.* 270:25328–25331.
Zardeneta, G., and Horowitz, P. M. 1994. Detergent, liposome, and micelle-assisted protein refolding. *Anal. Biochem.* 223:1–6.
Zhang, S. P., Zubay, G., and Goldman, E. 1991. Low-usage codons in *Escherichia coli*, yeast, fruit fly and primates. *Gene* 105:61–72.
Zhi, W., Landry, S. J., Gierasch, L. M., and Srere, P. A. 1992. Renaturation of citrate synthase: Influence of denaturant and folding assistants. *Prot. Sci.* 1:522–529.

Molecular Biology Problem Solver: A Laboratory Guide. Edited by Alan S. Gerstein
Copyright © 2001 by Wiley-Liss, Inc.
ISBNs: 0-471-37972-7 (Paper); 0-471-22390-5 (Electronic)

16

Eukaryotic Expression

John J. Trill, Robert Kirkpatrick, Allan R. Shatzman, and Alice Marcy

Section A: A Practical Guide to Eukaryotic Expression 492
Planning the Eukaryotic Expression Project 493
What Is the Intended Use of the Protein and What Quantity Is Required? 493
What Do You Know about the Gene and the Gene Product? 496
Can You Obtain the cDNA? 497
Expression Vector Design and Subcloning 498
Selecting an Appropriate Expression Host 501
Selecting an Appropriate Expression Vector 506
Implementing the Eukaryotic Expression Experiment 511
Media Requirements, Gene Transfer, and Selection 511
Scale-up and Harvest 514
Gene Expression Analysis 515
Troubleshooting 517
Confirm Sequence and Vector Design 517
Investigate Alternate Hosts 519
A Case Study of an Expressed Protein from cDNA to Harvest 519
Summary 521
Section B: Working with Baculovirus 521
Planning the Baculovirus Experiment 521

Is an Insect Cell System Suitable for the Expression of Your Protein? 521
Should You Express Your Protein in an Insect Cell Line or Recombinant Baculovirus? 522
Procedures for Preparing Recombinant Baculovirus 524
Criteria for Selecting a Transfer Vector 524
Which Insect Cell Host Is Most Appropriate for Your Situation? 525
Implementing the Baculovirus Experiment 527
What's the Best Approach to Scale-Up? 527
What Special Considerations Are There for Expressing Secreted Proteins? 527
What Special Considerations Are There for Expressing Glycosylated Proteins? 528
What Are the Options for Expressing More Than One Protein? 529
How Can You Obtain Maximal Protein Yields? 529
What Is the Best Way to Process Cells for Purification?. . . 530
Troubleshooting 530
Suboptimal Growth Conditions 530
Viral Production Problems 531
Mutation 531
Solubility Problems 532
Summary 532
Bibliography 533

SECTION A: A PRACTICAL GUIDE TO EUKARYOTIC EXPRESSION

Recombinant gene expression in eukaryotic systems is often the only viable route to the large-scale production of authentic, post-translationally modified proteins. It is becoming increasingly easy to find a suitable system to overexpress virtually any gene product, provided that it is properly engineered into an appropriate expression vector. Commercially available systems provide a wide range of possibilities for expression in mammalian, insect, and lower eukaryotic hosts, each claiming the highest possible expression levels with the least amount of effort. Indeed, many of these systems do offer vast improvements in their ease of use and rapid end points over technologies available as recently as 5 to 10 years ago. In addition methods of transferring DNA into cells have advanced in parallel enabling transfection efficiencies approaching 100%. However, one still needs to carefully consider the most

appropriate vector and host system that is compatible with a particular expression need. This will largely depend on the type of protein being expressed (e.g., secreted, membrane-bound, or intracellular) and its intended use. No one system can or should be expected to meet all expression needs.

In this section we will attempt to outline the critical steps involved in the planning and implementation of a successful eukaryotic expression project. Planning the project will begin by answering pertinent questions such as what is known about the protein being expressed, what is its function, what is the intended use of the product, will the protein be tagged, how much protein is needed, and how soon will it be needed. Based on these considerations, an appropriate host or vector system can be chosen that will best meet the anticipated needs.

Considerations during the implementation phase of the project will include choosing the best method of gene transfer and stable selection compared to transient expression and selection methods for stable lines, and clonal compared to polyclonal selection. Finally, we will discuss anticipated outcomes from various methods, commonly encountered problems, and possible solutions to these problems.

PLANNING THE EUKARYOTIC EXPRESSION PROJECT

What Is the Intended Use of the Protein and What Quantity Is Required?

Protein quantity is an important consideration, since substantial time and effort are required to achieve gram quantities while production of 10 to 100 milligrams is often easily obtained from a few liters of cell culture. Therefore we tend to group the expressed proteins into the following three categories: target, reagent, and therapeutic protein. This is helpful both in choosing an appropriate expression system and in determining how much is enough to meet immediate needs (Table 16.1).

Targets

Protein targets represent the majority of expressed proteins used in classical pharmaceutical drug discovery, which involves the configuration of a high-throughput screen (HTS) of a chemical or natural product library in order to find selective antagonists or agonists of the protein's biological activity. Protein targets include enzymes (e.g., kinases or proteases), receptors (e.g., 7

Table 16.1 Categories of Expressed Proteins

Class of Protein	Examples	Expression Amount	Appropriate System
Target	Enzymes and receptors	For screening: 10 mg For structural studies: 100 mg	Stable insect Baculovirus Mammalian Yeast
Reagent	Modifying enzymes Enzyme Substrates	<10 mg	Stable insect Baculovirus Mammalian Yeast
Therapeutic	Therapeutic Monoclonal antibody (mAb) Cytokine Hormone	g/L	Mammalian (CHO, myelomas)

transmembrane, nuclear hormone, integrin), and their ligands and membrane transporters (e.g., ion channels). In basic terms, sufficient quantities of a protein target need to be supplied in order to run the HTS. The actual amounts depend on the size of a given library to be screened and the number of hits that are obtained, which will then need to be further characterized. As a rule of thumb, for purified proteins such as enzymes and receptor ligands, amounts around 10 mg are usually needed to support the screen. For nonpurified proteins such as receptors, one needs to think in terms of cell number and the growth properties of the cell line. For most cell lines, screens are configured by plating between 100,000 to 300,000 cells per milliliter. By way of example, a typical screen of one million compounds in multiwell formats (e.g., 96, 384, or 1536 well) could use between 0.5 to 1.5×10^9 cells. The smaller the volume of the screen, the fewer cells will be required.

Because protein targets require a finite amount of protein, one has the flexibility of choosing from virtually any expression system. Consequently the selection of the system for producing a target protein really depends on considerations other than quantity. The most important goal is to achieve a product with the highest possible biological activity. This will enable a screen to be configured with the least amount of protein and will give the best chance of establishing a screen with the highest possible signal to background ratio. Other considerations include the type of protein being expressed (e.g., intracellular, secreted, and membrane-associated proteins). As discussed below, stable cell systems tend to be more amenable to secreted and membrane-associated proteins, while intracellular proteins are often pro-

duced very efficiently from lytic systems such as baculovirus. Whatever system is used, it should be scaled appropriately to meet the needs of HTS.

A subset of target proteins are those that are used for structural studies. In order to grow crystals that are of sufficient quality to yield high-resolution structures, it is particularly important to begin with properly folded, processed, active protein. Proteins used for structural studies are often supplied at very high concentrations (>5 mg/ml) and must be free of heterogeneity. Glycosylation is often problematic because its addition and trimming tends to be heterogenous (Hsieh and Robbins, 1984; Kornfeld and Kornfeld, 1985). As a result it is often necessary to enzymatically remove some or all of the carbohydrate before crystals can be formed. As a starting point, one often needs approximately 10 mg of absolutely pure protein so that crystallization conditions can be tested and optimized, with the total protein requirement often exceeding 100 mg.

In order to avoid the issue of glycosylation in structural studies altogether, one can express the protein in a glycosylation-deficient host (Stanley, 1989). Alternatively one can remove glycosylation sites by site-directed mutagenesis prior to expression. However, these are very empirical methods that do not often work well for a variety of reasons, including the need in some cases to maintain glycosylation for proper solubility. Thus, for direct expression of a nonglycosylated protein, a first-pass expression approach would likely involve a bacterial system in which high level expression of nonglycosylated protein is more readily attained.

Reagents

A second category of expressed proteins is reagents. These are proteins that are not directly required to configure a screen but are needed to either evaluate compounds in secondary assays or to help produce a target protein itself. Examples of reagent proteins include full-length substrates that are replaced by synthetic peptides for screening. Enzyme substrates themselves are often cleaved to produce biologically active species whose activities can be assessed in vitro. Reagent proteins can also include processing enzymes that are required for the in vitro activation of a purified protein (e.g., cleavage of a zymogen or phosphorylation by an upstream activating kinase). Also included in this category are gene orthologues from species other than the one being used in the screen, whose expression will be used to support animal studies and to determine the cross-species selectivity or activity of selected compounds.

Reagent proteins are usually required in much lower amounts than target proteins. Some can even be purchased commercially in sufficient quantities to meet the required need. Others, because of price or the required quantity, may necessitate recombinant expression. But, since only small quantities are usually required (<10 mg), it is possible to choose an expression system with features that will favor efficient and rapid expression. Furthermore the expression scale can be minimized. The bottom line is that reagent proteins should be the least resource intensive to produce. One should avoid trying to overproduce reagent proteins or scaling them to quantities that will never be used.

Therapeutics

In contrast to reagent proteins, therapeutic protein agents are the most demanding in terms of resource. Therapeutic proteins have intrinsic biological properties like medical drugs. The ultimate objective for expression of a therapeutic protein is the production of clinical-grade protein approaching or exceeding gram per liter quantities. For most expression systems this is not readily achievable. Other than bacterial and yeast expression, the most robust system for producing these levels is the Chinese hamster ovary (CHO) system. Due to the lack of proper post-translational modifications (e.g., glycosylation) in bacteria and yeast, CHO cell expression is often the only choice to achieve sufficient expression. Examples of therapeutic proteins, produced in CHO cells, include humanized monoclonal antibodies (Trill, Shatzman, and Ganguly, 1995), tPA (tissue plasminogen activator; Spellman et al., 1989), and cytokines (Sarmiento et al., 1994). In many cases months are spent selecting and amplifying lines with appropriate growth properties and expression levels to meet production criteria.

What Do You Know about the Gene and the Gene Product?

Information about the gene product or for that matter, its homologues or orthologues, enables one to make an educated guess as to what is the best eukaryotic expression system to use. Is there anything published in the literature about the gene, or is it completely uncharacterized? Do we know in what tissue the gene is expressed, based on either Northern blot analysis or by quantitative or semiquantitative RT-PCR measures? Other factors to determine are whether the protein to be expressed is secreted, cytosolic, or membrane-bound. If it is a receptor, is it a homodimer, heterodimer, multimeric, single, or multispanning

transmembrane receptor or anchored to the surface (e.g., through a glycosyl phosphatidylinositol phosphate (GPI linkage).

Fortunately we usually have the luxury of working with genes that are at least partially characterized by their biological properties. But what about the genes of unknown origin or function? In this new age of genomics, many of the genes we obtain are "like" genes, belonging to large families of related genes that share only a minimal percentage of homology with a known gene. Despite these similarities there is often no way to know whether the same expression and purification methods used for one orthologue or homologue will be effective for another. Thus one is immediately faced with the challenging prospect of having to consider multiple expression strategies in order to get the protein expressed and purified to sufficient levels in an active form, in addition to not knowing what activity to look for.

Can You Obtain the cDNA?

Before embarking on an expression project you will need to locate a cDNA copy of the gene of interest. It is also possible in theory to express genomic DNA containing introns, provided that the expression host will recognize the proper splice junctions. In practice, however, this is not often the most efficient route to expression because it is not usually known how the introns will affect expression levels or whether the desired splice variant will be expressed. Furthermore most mammalian genes are interrupted by multiple intron sequences that can span many kilobases in length. This can make subcloning of genomic DNA considerably more difficult than for the corresponding cDNA.

The three most common ways to obtain a known gene of interest include purchase from a distributor of clones from the Integrated Molecular Analysis of Genomes and their Expression (IMAGE) consortium (*http://image.llnl.gov/*), requests from a published source such as an academic lab, or RT-PCR cloning from RNA derived from a cell or tissue source. IMAGE clones can be found by performing a BLAST search of an electronic database such as GenBank, which can be accessed at the National Library of Medicine PubMed browser (*http://www.ncbi.nlm.nih.gov/PubMed/*). From there you can quickly determine if a sequence is present, if it is full length, publications related to this gene, and possible sources of the gene (tissue sources, personal contacts, etc). Most expressed sequence tags (EST's) matching the gene of interest are available as IMAGE clones. The trick is to find one that is full length. It is

easy to determine if an EST is likely to contain a full-length sequence if it is derived from a directional oligo dT primed library and sequenced from the 5′ end by searching for an ATG and an upstream stop codon. Once you identify a full-length EST, you should then be able to obtain the corresponding IMAGE clone from Incyte Genomics, LifeSeq Public Incyte clones (*http://www.incyte.com/reagents/index.shtml*), Research Genetics (*http://www.resgen.com*), or the American Type Culture Collection (*ATCC, http://www.atcc.org*). If the gene is published, you can also try contacting the author who cloned it in order to obtain a cDNA clone. Most labs, including both academic and pharmaceutical/ biotech companies, will honor a request for a cDNA clone if it is published. Alternatively, you may consider deriving the gene de novo by RT-PCR using the sequence obtained above.

Depending on the size, abundance, and tissue distribution of the mRNA, a PCR approach could be straightforward or complex. One may isolate RNA from tissue, generate cDNA from the RNA using reverse transcriptase, design PCR primers to perform PCR, and fish out the gene of interest. Alternatively, one may simply purchase a cDNA library from which to PCR amplify the gene. Several vendors carry a wide array of high-quality cDNA libraries derived from human and animal tissues. For example, cDNA libraries for virtually every major human or murine tissue/organ can be obtained from Invitrogen (*http://www.invitrogen.com./catalog_project/index.html*) or Clontech (*http://www.clontech.com/products/catalog/Libraries/index.html*). These companies obtain their samples from sources under Federal Guidelines.*

Expression Vector Design and Subcloning

Perhaps the most critical step in the process of expressing a gene is the vector design and subcloning. As much an art as a science, it nevertheless requires complete precision. In many cases you will need to amplify the gene by PCR from RNA. If the gene is in a library, you may also need to trim the 5′ and 3′ UTR (untranslated region) and to add restriction sites and/or a signal sequence if one is not already present. You may also want to add

**Editor's note: In addition to the planning recommended by the authors, it is wise to ask commercial suppliers of expression systems about the existence of patents relating to the components of an expression vector (i.e., promoters) or the use of proteins produced by a patented expression vector/system.*

epitope tags for detection and purification (e.g., His_6 tag). When PCR is involved, the gene will eventually need to be entirely re-sequenced in order to rule out PCR-induced mutations that can occur at a low frequency. If mutations are found, they will need to be repaired, thereby adding to the time required to generate the final expression construct. The best practice is to start with a high-fidelity polymerase with a proofreading (3′–5′ exonuclease activity) function to avoid PCR errors.

Sequence Information

If you are lucky enough to obtain a DNA from a known source, a new litany of questions will need to be answered. Is a sequence and restriction map available? Do you know what vector the gene has been cloned into? Has the gene been sequenced in its entirety? How much do you trust the source from which you have received the gene? It is usually best to have the gene re-sequenced so that you know the junctions and restriction sites and can assure yourself that you are indeed working with the correct gene. What do you do if there are differences between your sequence and the published sequence? You will need to decide if the difference is due to a mutation, an artifact from the PCR reaction, a gene polymorphism, or an error in the published sequence. A search of an EST database coupled with a comparison with genes of other species can help distinguish whether the error is in the database or due to a polymorphism. Alternatively, sequencing multiple, independently derived clones can also help answer these questions.

Control Regions

We now have a gene with a confirmed sequence. But which control regions are present? Does the gene contain a Kozak sequence, 5′-GCCA/GCCAUGG-3′, required to promote efficient translational initiation of the open reading frame (ORF) in a vertebrate host (Kozak, 1987) or an equivalent sequence 5′-CAAAACAUG-3′ for expression in an insect host (Cavener, 1987)? If this sequence is missing, it is essential to add it to your expression vector. It is also advisable to trim the gene to remove any unnecessary sequences upstream of the ATG. The 5′ noncoding regions may contain sequences (e.g., upstream ATG's or secondary structures) that may inhibit translation from the actual start. A noncoding sequence at the 3′ end may destabilize the message.

Epitope Tags and Cleavage Sites

Another sequence you might need to add to your gene is an epitope tag or a fusion partner with or without a protease cleavage site. This will aid in the identification of your protein product (via Western blot, ELISA, or immunofluorescence) and assist in protein purification. Among the various epitope tags available are FLAG® (DYKDDDDK) (Hopp et al., 1988), influenza hemaglutinin or HA (YPYDVPDYA) (Niman et al., 1983), His_6 (HHHHHH) (Lilius et al., 1991), and c-myc (EQKLISEEDL) (Evan et al., 1985). The more popular protease cleavage sites, used to remove the tag from the protein, are thrombin (VPR'GS) (Chang, 1985), factor Xa (IEGR'; Nagai and Thogersen, 1984), PreScission protease (LEVLFQ'GR; Cordingley et al., 1990), and enterokinase (DDDDK'; Matsushima et al., 1994) One may also use larger fusion partners such as the Fc region of human IgG1 or GST. It is crucial to choose a protease that is not predicted to cleave within the protein itself, but this does not preclude spurious cleavages.

The benefits and drawbacks of utilizing epitope tags are discussed in greater detail below in the section, "Gene Expression Analysis."

Subcloning

Your gene is now ready to be cloned into an expression vector of your choice, provided that you have already decided what system to use. This will traditionally involve the use of restriction enzymes to precisely excise the gene on a DNA fragment, which is subsequently ligated into a donor expression vector at the same or compatible sites. If appropriate unique restriction sites are not located in flanking regions they can be added by PCR (incorporating the sequence onto the end of the amplification primer), or by site-directed mutagenesis.

Recent technological advances also offer the possibility of subcloning without restriction enzymes. These new age cloning systems are based on recombinase-mediated gene transfer. Invitrogen offers ECHO™ and Gateway™ cloning technologies, while Clontech markets the Creator™ gene cloning and expression system. Recombinases essentially perform restriction and ligation in a single step, thereby eliminating the time-consuming process of purifying restriction fragments for subcloning and ligating them. These new systems are particularly advantageous when transferring the same gene into multiple expression vectors for expression in different host systems.

Selecting an Appropriate Expression Host

Expressed Protein Issues

The properties of the protein and its intended usage will also have a direct impact on which expression system to choose. Since many eukaryotic proteins undergo post-translational modifications (phosphorylation, signal-sequence cleavage, proteolytic processing, and glycosylation), which can affect function, circulating half-life, antigenicity, and the like, these issues must be addressed when choosing an expression host. These steps have a direct influence on the quality of protein produced. For instance, it has been demonstrated that there is a clear difference in the glycosylation patterns between various mammalian and insect systems. Insect cells lack the pathways necessary to produce glycoproteins containing complex *N*-linked glycans with terminal sialic acids (Ailor and Betenbaugh, 1999; Kornfeld and Kornfeld, 1985), and the absence of sialic acid residues can strongly influence the in vivo pharmacokinetic properties of many glycoproteins (Grossmann et al., 1997). Using tPA as a model system, it has also been shown that glycosylation patterns differ within different mammalian cell types (Parekh et al., 1989).

The expression strategies for both targets and reagents are the same. We desire a purified protein, cell membranes for a binding assay, or attached cell lines for a cell-based assay. The determining factor for selecting a host system depends on the quantity of the protein needed, what signaling components are necessary in the host line, and the degree to which endogenously expressed host proteins generate background responses (e.g., for receptors). For example, insect cell lines often provide a null background for mammalian signaling components, which enable lower basal level activation and high signal to background in cell-based assays.

If the protein is a target and will be used in a cell-based assay, one needs to make a high expressing cell line. In most cases the higher the expression is, the better is the result. But this is not always the case for cytoplasmic or membrane anchored proteins where the expressed protein can be toxic. In these cases it might be better to achieve lower expression or to use some type of regulated promoter vector system as discussed in the following section.

If the desired protein is to be a therapeutic and used to supply clinical trials, the choices are very well documented. There are numerous examples of commercial therapeutic proteins being produced in *E. coli* and yeast. However, if the protein contains numerous disulfide linkages, or requires extensive post-

translational modifications (i.e., folding of antibody heavy and light chains), one needs to consider expression in a mammalian cell line. The gene needs to be cloned into a plasmid system allowing for some type of amplification so that the protein can be expressed at very high levels. In addition one needs to be cognizant of GMP, GLP, and FDA guidelines for the entire expression, selection or amplification process.

The inability to obtain homogeneously pure protein for crystallization is a frequently encountered problem due to the heterogeneous carbohydrate content of many eukaryotic proteins (Grueninger-Leitch et al., 1996). In the past *E. coli* expression systems were exclusively used to produce material for crystallization in order to avoid having glycosylation at all. Recently there have been an increasing number of examples where crystals were generated using baculovirus-expressed protein (Cannan et al., 1999; Sonderman et al., 1999). Another approach has been to use the glycosylation-deficient mutant CHO cell line, Lec3.2.8.1, (Stanley, 1989; Butters et al., 1999; Casasnovas, Larvie, and Stehle, 1999; Kern et al., 1999). In these cases the incomplete or underglycosylation has allowed the formation of high-resolution, diffractable crystals.

Transient Expression Systems

Transient systems are used for the rapid production of small quantities of heterologous gene products and are often suitable to make "reagent" category proteins. The cell lines of choice include the following;

- COS cells (COS-1, ATCC CRL 1650; COS-7 ATCC CRL 1651; see Gluzman, 1981). These are derived from the African green monkey cell line, CV-1, which was infected with an origin-defective SV40 genome. Upon transfection with a plasmid containing a functional SV40 origin of replication, the combination of SV40 replication origin (donor) and SV40 large T-antigen (host cell) results in high copy extrachromosomal replication of the transfected plasmid (Mellon et al., 1981).
- Human embryonic kidney (HEK) 293 cells (ATCC CRL 1573). An immortalized cell line derived from human embryonic kidney cells transformed with human adenovirus type 5 DNA. This cell line contains the adenovirus E1A gene, which transactivates CMV promoter-based plasmids, and this results in increased expression levels. This cell line is widely used to express 7 trans membrane G-protein-coupled receptors (GPCRs) (Ames et al., 1999; Chambers et al., 2000).

In our own experiments involving transient expression systems, we have consistently found that COS cells yield approximately 50% higher expression than HEK 293 cells. (Trill, 2000, unpublished). To take monoclonal antibodies (mAbs) as an example, transient systems such as COS can allow one to examine multiple constructs in two to three days at expression levels ranging from 100 ng/ml to 2 μg/ml. Stable cell lines can yield over 200-fold more protein, but it is often a time-consuming process to achieve those levels, often taking six months to a year to accomplish (Trill, Shatzman, and Ganguly, 1995).

Viral Lytic Systems

Viral lytic systems offer the advantage of rapid expression combined with high-level production. The most popular of the viral lytic systems utilizes baculovirus.

The baculovirus expression system is based on the manipulation of the circular *Autographa californica* virus genome to produce a gene of interest under the control of the highly efficient viral polyhedrin promoter. Engineered viruses are used to infect cell lines derived from pupal ovarian tissue of the fall army worm, *Spodoptera frugiperda* (Vaughn et al., 1977). This lytic system is most useful for the high-level expression of enzymes and other soluble intracellular proteins. Secreted proteins can also be obtained from this system but are more difficult to scale to large volumes due to the rapid onset of the lytic cycle. Cell lines include Sf9, Sf21, and *T. ni* (available as High Five™) cells are from *Trichoplusia ni* egg cell homogenates. Refer to Section B for more detail on baculovirus expression.

Adenovirus expression has also increased in popularity of late. This may be due in part to its use for in vivo gene delivery in animal systems and limited use in experimental gene therapy (Robbins, Tahara, and Ghivizzani, 1999; Ennist, 1999; Grubb et al., 1994). The advantages of this system include a broad host specificity and the ability to use the same expression vector to infect different host cells for contemporaneous animal studies (von Seggern and Nemerow, 1999). Commercial vectors are available for generating recombinant viruses such as the AdEasy™ system sold by Stratagene. This system simplifies the process of generating recombinant viruses since it relies on homologous recombination in *E. coli* rather than in eukaryotic cells (He et al., 1998). The main limitations of this system include moderate to low expression levels and the need to maintain a dedicated tissue culture space in order to avoid crosscontamination with other host

cells. Other animal viruses of interest, including Sindbis, Semliki Forest virus, and the adeno-associated virus (AAV), share many of the same advantages as adenovirus, including broad host specificity (Schlesinger, 1993; Olkkonen et al., 1994; Bueler, 1999). None of these virus expression systems are discussed in detail in this chapter because they do not currently represent mainstream methods for large-scale protein production as is evident from the limitations discussed.

Stable Expression Systems

Stable expression systems are preferred when one desires a continuous source and high levels of expressed heterologous protein. The actual levels of expression largely depend on which host cells are used, what type of plasmids are used, and where the genes are integrated into the host genome (i.e., whether they are influenced by chromosomal position effects).

What are the cell line choices? If it is a mammalian system, the most common choices are as discussed next.

Mouse

Mouse cells such as L-cells (ATCC CCL 1), Ltk$^-$ cells (ATCC CCL 1.3), NIH 3T3 (ATCC CRL 1658), and the myeloma cell lines, Sp2/0 (ATCC CRL 1581), NSO (Bebbington et al., 1992) and P3X63.Ag8.653 (ATCC CRL 1580). These myeloma cell lines have the advantages of suspension growth in serum-free medium and their derivation from secretory cells makes them well-suited hosts for high-level protein production. Because of the presence of the endogenous dihydrofolate reductase (DHFR) gene, none of these cells can be amplified through the use of methotrexate (Schimke, 1988). However, as shown by Bebbington et al. (1992), NSO cells can be amplified using the glutamine synthetase system.

Rat

Rat cell lines, RBL (ATCC CRL 1378), derived from a basophillic leukemia, have been used to express 7TM G-protein-coupled receptors (Fitzgerald et al., 2000; Santini et al., 2000), while the myeloma cell line YB2/0 (ATCC CRL 1662), has been used in the high-level production of monoclonal antibodies (Shitara et al., 1994).

Human

Human cell lines that are frequently used include HEK 293, HeLa (ATCC CCL 2), HL-60 (ATCC CCL 240), and HT-1080 (ATCC CCL 121).

Hamster

Chinese hamster ovary (CHO) cells, such as CHO-K1 (ATCC CCL 61), and two different DHFR$^-$ cell lines DG44 (Urlaub et al., 1983) or DUK-B11 (Urlaub and Chasin, 1983) in which the gene of interest can be amplified via the selection/amplification marker DHFR (Kaufman, 1990). CHO cells have been used to express a large variety of proteins ranging from growth factors (Madisen et al., 1990; Ferrara et al., 1993), receptors (Deen et al., 1988; Newman-Tancredi, Wootton, and Strange, 1992), 7TM G-protein-coupled receptors (Ishii et al., 1997; Juarranz et al., 1999), to monoclonal antibodies (Trill, Shatzman, and Ganguly, 1995).

Also of significance are engineered derivatives of these lines. One example is a CHO cell line containing the adenovirus E1A gene. Cockett, Bebbington, and Yarronton (1991) first established a CHO cell line stably expressing the adenovirus E1A gene, which trans-activates the CMV promoter. Transfection of a human procollagenase gene into this CHO cell line produced a 13-fold increase in stable expression compared with that of CHO-K1. This is significant because an E1A host cell line can be used to rapidly produce sufficient material for early purification and testing without the need for amplification. Stably expressing clones produced from this host can be obtained in as little as two weeks and yield 10 to 20 mg/L of expressed protein.

Baby Hamster Kidney (BHK) Cells (ATCC CCL 10)

BHK cells have also been used to express a variety of genes (Wirth et al., 1988).

Drosophila

Drosophila S2 is a continuous cell line derived from primary cultures of late stage, 20 to 24 hours old, *D. melanogaster* (Oregan-R) embryos (Schneider, 1972). The cell line is particularly useful for the stable transfection of multiple tandem gene arrays without amplification. High copy number genes can be expressed in a tightly regulated fashion under the control of the copper-inducible *Drosophila* metallothionein promoter (Johansen et al., 1989). This cell line is particularly useful for the inducible expression of secreted proteins. S2 cells also grow well in serum-free, conditioned medium, simplifying the purification of expressed proteins.

Yeast Expression Systems (Pichia pastoris and Pichia methanolica)

The main advantages of yeast systems over higher eukaryotic tissue culture systems such as CHO include their rapid growth rate

to high cell densities and a well-defined, inexpensive media. Main disadvantages include significant glycosylation differences of secreted proteins comprised of high mannose, hyperglycosylation consisting of much longer carboydrate chains than those found in higher eukaryotes, and the absence of secretory components for processing certain higher eukaryotic proteins (reviewed in Cregg, 1999). Because of these limitations, yeast systems will not be discussed in full detail in this chapter. More information on *Pichia* expression can be found in the following references: Higgins and Cregg (1998), Cregg, Vedvick, and Raschke (1993), and Sreekrishna et al. (1997).

We all have our preferences for what are the best cell lines to use. Therefore, when setting up an expression laboratory, one should consider obtaining a variety of host cell lines. Listed are a few examples of cell lines that have been routinely used and reasons for their selection: CHO-DG-44 and *Drosophila* S2 (available from Invitrogen), based on consistency in growth, high-level expression, and ability to be easily adapted to serum-free growth in suspension; COS for transient expression; HEK 293, a versatile human cell line which can be used for both transient (but not as good as COS) and stable expression; and Sf9 a host cell for baculovirus infection, a system best suited for internalized proteins rather than secreted proteins. A majority of these cell lines can be grown in serum-free suspension culture, a property that facilitates ease of use and product purification as well as reducing cost.

Selecting an Appropriate Expression Vector

Once an appropriate host system has been chosen, it's time to find a suitable expression vector. For each of the host systems described above, there are a wide variety of vectors to choose from.

A typical expression vector requires the following regulatory elements necessary for expression of your gene: a promoter, translational initiator codon, stop codon, a polyadenylation signal, a selectable marker, and several prokaryotic elements such as a bacterial antibiotic selection marker and an origin of replication for plasmid maintenance. (The presence of prokaryotic elements is for shuttling between mammalian and prokaryotic hosts.) There are numerous choices for each regulatory element, but unfortunately there is no blueprint on which combinations will yield the highest expressing plasmid.

Table 16.2 Promoter Strength Table

Promoter	Source	Strength	Reference
EF-1α	Human elongation factor 1α	40–160	Mizushima and Nagata (1990)
CMV	Human cytomegalovirus immediate-early gene	4	Boshart et al. (1985)
RSV	Rous sarcoma virus LTR	2	Gorman et al. (1982)
SV40 late	Simian virus 40 Late gene	1.1	Wenger, Moreau, and Nielsen (1994)
SV40 early	Simian virus 40 Early gene	1	
Adeno major late	Adenovirus major late promoter	0.4	Mansour, Grodzicker, and Tjian (1986)
Beta-globin	Mouse beta-globin promoter	0.2	Hamer, Kaehler, and Leder (1980)
Beta-actin	Human beta-actin promoter	ND	Ng et al. (1985)

Note: SV40 early promoter strength set as 1 for comparative purposes, and the numbers indicate how much stronger these promoters are.

Promoters

Promoters are DNA sequences that recruit cellular factors and RNA polymerase to activate transcription of a particular gene. They must contain a transcriptional start site, a CAAT box, and TATA box. Examples of various mammalian promoters are given in Table 16.2.

The promoter strength is based on a compilation of comparative experiments where various promoters were compared in transient experiments using the R1610 cell line (Thirion, Banville, and Noel, 1976). The strength of EF-1α and CMV was derived from a comparison to the RSV LTR involving stable expression of various monoclonal antibodies and tPA (Trill, 1998 unpublished). The EF-1α promoter (available from Invitrogen) is by far the strongest promoter and a good choice if you want quick high-level expression.

Polyadenylation Regions

Polyadenylation occurs at a consensus sequence, AAUAAA, and results in increased mRNA stability. Cleavage after the U by poly A polymerase adds a string of adenylate residues (Wahle and Keller, 1992). As with the promoters, there are a number of sources of polyadenylation regions. Several examples are shown in Table 16.3.

Table 16.3 Polyadenylation Regions

Poly A Region	Source	Efficiency
BGH	Bovine growth hormone	3
SV40 late	Simian virus 40	2
TK	Herpes simplex virus thymidine kinase	1.5
SV40 early	Simian virus 40	1
Hep B	Hepatitis B surface antigen	1

Note: SV40 early poly(A) region strength set as 1 for comparative purposes, and the numbers indicate how much more efficient these polyadenylation regions are. The data above and polyadenylation regions are referenced in Pfarr et al. (1985, 1986).

Drug Selection Markers

Choice number three: What drug selection markers should one use? These genes provide resistance to a particular selective drug, and only cells in which the plasmid has been integrated will survive selection. Some effective choices are Blasticidin (Izumi et al., 1991), Histidinol, (Hartman and Mulligan, 1988), Hygromycin B (Gritz and Davies, 1983), Geneticin® (G418) (Colbere-Garapin et al., 1981), Puromycin (de la Luna et al., 1988), mycophenolic acid (Mulligan and Berg, 1981), and Zeocin™ (Mulsant et al., 1988). Whatever marker you decide to use, remember, you will need to determine the effective concentration of drug for each cell line you use. Second, if you are on a tight budget, there is a huge disparity in cost of these drugs. Also there are environmental concerns regarding waste disposal of the conditioned growth medium containing some of these drugs.

Amplification

Finally, if expression is unacceptably low, one solution is to amplify your gene copy number. Two such amplification systems are the use of dihydrofolate reductase (DHFR) as a drug selection marker in the presence of methotrexate, a competitive inhibitor of DHFR (Kaufman, 1990) and inhibition of the enzyme glutamine synthetase (GS) by methionine sulfoxide (MSX) (Bebbington et al., 1992).

Amplification through the DHFR gene is by far the more popular of the two systems. DHFR catalyzes the conversion of folate to tetrahydrofolate, which is necessary in the synthesis of glycine, thymidine monophosphate, and the biosynthesis of purines. If the transfected plasmid contains a DHFR gene, use of the CHO DG-44 and DUK-B11 cell lines allows one to initially select cells in medium devoid of nucleotides and then to amplify gene copy number by selection with increasing concentrations of

methotrexate (Geisse et al., 1996). In the majority of the cases, amplification of the gene copy number results in increased expression.

The glutamine synthetase system can be used as a dominant selectable marker in cell lines that contain GS activity, in glutamine-free growth medium. GS catalyzes the formation of glutamine from glutamate and ammonia. CHO-K1 and NSO are the more widely used cell lines for this method of selection, but myeloma cells offer a distinct advantage over CHO cells because of their low levels of endogenous GS activity. Myeloma cells transfected with a plasmid containing a gene of interest and the GS gene are often selected with low levels of MSX (up to 100 μM), while CHO cells are amplified using higher levels of MSX (up to 1 mM) (Bebbington et al., 1992; Cockett, Bebbington, and Yarronton, 1990).

Regulating Expression

What happens if overexpression of a gene results in a protein which is toxic to the host cell? There are a number of inducible promoters and regulated expression systems available that allow one the ability to control when and how much of the toxic protein is produced. Examples of such promoter-based systems include the Mouse mammary tumor virus (MMTV) promoter which is induced using dexamethasone (James et al., 2000), the *Drosophila* metallothionein promoter which is induced by addition of metal (e.g., cupric sulfate; Johansen et al., 1989), or the mifepristone-dependent plasmid-based gene switch system (Wang et al., 1994). The addition of inducers allows flexible control of expression in these systems. However, inducers such as heavy metals can also interfere with purification efforts, especially if your protein contains an epitope tag. For example, the use of the standard IMAC (immobilized metal affinity chromatography) method for the direct capture of His-tagged proteins from *Drosophila* culture medium is inefficient due to the presence of free copper, which interferes with binding. However, we recently found that when copper-supplemented medium containing an expressed His-tagged protein is loaded directly onto chelating sepharose, the protein binds efficiently to the resin via copper (Lehr et al., 2000). Furthermore this interaction is of greater affinity than that of free copper alone, which can be washed away under low-salt conditions.

Other methods for achieving regulated expression include the Ecdysone-inducible system, based on the heterodimeric ecdysone

receptor of *Drosophila* (Christopherson et al., 1992), and the tetracycline-regulated expression system, based on two regulatory elements derived from the tetracycline-resistance operon of the *E. coli* Tn10 transposon (Gossen and Bujard, 1992).

Single- or Double-Vector Systems?

What type of vector system will we use to house all of these regulatory elements? We can use a two-vector system in which the gene is contained on one plasmid while the selection marker is on the second. *Drosophila S2* cells are an example of a host where a two-vector system is preferable. In this case, varying the proportions of the two plasmids enables one to modulate the number of gene copies inserted onto the chromosome from just a few to more than a thousand (Johansen et al., 1989). Higher gene copies tend to correlate with higher expression levels. Thus the two-vector system can add to the flexibility of the expression outcome. Double-vector systems are also used for mAb expression where the heavy chain and the DHFR gene are located on one vector, and the light chain and the selectable marker are located on the second vector (Trill, Shatzman, and Ganguly, 1995). Alternatively, one could also use a single plasmid where the drug selection cassette and the amplification gene are located on the same plasmid (Aiyar et al., 1994). Again, using mAbs as an example, we can use a single plasmid that contains both the heavy and light chain cDNAs along with the selection and amplifiable drug markers (Trill, Shatzman, and Ganguly, 1995).

Which vector system should you use? This really depends on how much effort you want to expend in your plasmid cloning and transfection and how quickly you need your protein. With two plasmids, it means two separate clonings and two plasmids to sequence. You will also need to co-transfect both plasmids in a ratio that will favor optimal expression. This ratio may need to be empirically determined. A single plasmid, containing two different genes of interest necessitates a unique cloning strategy due to the decrease in unique restriction sites for the cloning process. It also means designing gene-specific bi-directional sequencing primers because of the duplication of regulatory elements.

Summary

There are a large assortment of commercially available mammalian and insect expression vectors to choose from. The majority of the mammalian vectors have common regulatory elements. Most use the CMV promoter to drive expression, contain a

polylinker region to clone in your gene of interest, and use a drug selection marker, most often Neomycin. One of the most popular is pCDNA3.1 sold by Invitrogen. Variations in these vectors include different choices of epitope TAGS for detection using an antibody or through the intrinsic fluorescence of the green fluorescence protein (GFP) and its derivatives. There are also bicistronic vectors that use a single expression cassette containing both the gene of interest and selection marker, separated by an internal ribosome entry site (IRES) from the encephalomyocarditis virus, to promote translation from a bicistronic transcript. In addition there are vectors containing signal sequences designed to aid secretion. Finally, to circumvent the need to develop multiple vectors for each system you use, you can obtain a single expression vector enabling protein expression in bacterial, insect, and mammalian cells from a single plasmid, such as the pTriEx expression vector marketed by Novagen.

It is advisable that one take the time to find a vector that is optimized for a particular host or, if one is not available, to construct a new vector and optimize it for each system that you intend to use. Take the time to create your own polylinker region with convenient, unique restriction sites so that you can easily exchange regulatory elements. CMV is perhaps one of the most versatile promoters available. You will also need to incorporate a resistance marker under the control of its own promoter and including a polyadenylation site. The choice of a selection marker will depend on considerations such as the cost of the drug, the efficiency of its action in a particular host, and environmental concerns for disposal.

IMPLEMENTING THE EUKARYOTIC EXPRESSION EXPERIMENT

Media Requirements, Gene Transfer, and Selection

Stable cell line generation, especially for a therapeutic protein, is a long, labor-intensive process that takes anywhere from six to nine months to complete. Therefore it is essential that one pay close attention to the methods employed to maintain, transfect, and select the cell lines.

Serum

When possible, try to adapt your cells for growth in a chemically defined, serum-free growth medium. Serum contains numerous undefined components, is costly to use, may contain

adventitious agents, and varies from lot to lot. Serum-free medium, available from a number of suppliers, offers several advantages. It allows cell culture to be performed with a defined set of conditions leading to a more consistent performance, possible increases in growth, increased productivity, and easier purification and downstream processing.

If you must use a serum-containing medium, be sure to have the serum lot tested for mycoplasma and other adventitious agents, such as BVDV (bovine viral diarrhea virus). If possible, order gamma-irradiated serum and ask for a certificate of analysis. With the recent concern over bovine spongiform encephalitis (BSE) disease in cattle from the United Kingdom, it is also wise to request serum from regions where BSE is not present (e.g., United States and New Zealand). This extra precaution further adds to the high cost of serum.

Antibiotics

Many researchers supplement their growth media with antibiotics such as penicillin, streptomycin, and antifungals such as amphotericin B. While this is effective in preventing either bacterial or fungal contamination, it does nothing to prevent contamination from mycoplasma or viruses. Furthermore antibiotics can mask poor cell culture sterile technique, lead to drug-resistant bacteria, and increase the risk of mycoplasma contamination. In short, there is no substitute for proper sterile technique, which should eliminate the need to add antibiotics in the first place.

On the subject of sterility, it is also prudent to have your cell lines tested monthly for mycoplasma. Trypsin, for use in removing attached cell lines, should be free of mycoplasma, PPV (porcine parvovirus), and PRRS (porcine respiratory and reproductive syndrome virus). If your medium will be used to support growth of production cell lines expressing therapeutic agents, it is also advisable to consult the FDA guidelines for the use of medium containing animal products.

If one of your cultures should become contaminated with mycoplasma, the best cure is to dispose of the cell lines in question. If this is not an alternative, there are a number of reports indicating that mycoplasma has been eradicated through the use of MRA (mycoplasma removal agent (ICN), a quinolone derivative) (Uphoff, Gignac, and Drexler, 1992; Gignac et al., 1992), either ciprofloxacin (Gignac et al., 1991; Schmitt et al., 1988) or enrofloxacin (Fleckenstein, Uphoff, and Drexler, 1994) both of which are fluoroquinolone antibiotics and BM-cyclin (Roche

Molecular Biochemicals, a combination of tiamulin and minocycline) (Uphoff, Gignac, and Drexler, 1992). However, this is a time-consuming, cost-intensive process that may result in irreversible damage to your cell cultures.

Transfection

The most contemporary methods for transfection of foreign genes into cells employ either cationic lipid reagents or electroporation (Potter, Weir, and Leder, 1984). The former relies on different liposome formulations of cationic or polycationic lipids (as per the manufacturer) that complex with DNA facilitating its uptake into cells. The procedure is simple, very rapid, and can be used for a large variety of cell types. It is the method of choice for transient transfections, especially into COS cells, and is by far the most preferred method for transfecting attached cell lines.

Electroporation, which relies on an electric pulse to reversibly permeabilize the cell's plasma membrane, creates transient pores on the surface of the cell that allow plasmid DNA to enter. This technique is also very rapid, and the protocols are straightforward and can be used in a variety of cell types. Electroporation can be used on suspension cell lines and attached cells, which have been detached from the plate. Electroporation is most efficient when the DNA is linearized prior to transfection (Trill, unpublished). It also offers the unique advantage that a majority of the DNA is integrated in single copies at single sites without any rearrangements (Boggs et al., 1986; Toneguzzo et al., 1988). This is significant when assessing stability and chromosomal location of the gene within the cells and the expressed protein.

Clonal or Polyclonal Selection?

There are advantages and disadvantages to selecting cells as bulk populations over their selection as clones through limit dilution, colony formation, or fluorescence-activated cell sorting (FACS). On the one hand, polyclonal lines can be derived much more quickly than clonal lines, and a reasonable expression level can be achieved in many cases. On the other, there are also many inherent problems with this method. For example, expression levels tend to be diluted by a population of nonproducers within the selected population. These cells contain the transfected plasmid and an intact, fully functioning drug selection gene, but have somehow lost expression from the gene of interest. Within such populations, the risk is great that nonproducers will eventually overgrow the producers, further diluting expression levels.

This problem is compounded by the tendency of overexpressing cells to grow slower than low or nonexpressors.

In general, it is preferable to select clones rather than polyclonal populations in order to achieve the highest reproducible expression. However, the isolation of clonal lines is considerably more time-consuming and labor-intensive. In addition you will need to evaluate expression from tens to hundreds of clonal cell lines rather than a polyclonal population from a single flask.

Whatever selection method you should choose, you will need to do some type of experimentation to assess such cell line characteristics as growth, viability, and protein expression.

Scale-up and Harvest

The final task prior to purification of the recombinant protein is to convert your cell culture into a "factory" for the production of the desired recombinant protein. Again, the type of system that you employ depends largely on the intended use of the protein and how much will be needed. Other deciding factors include cost and complexity of use. Benchtop fermentation systems can be purchased from a number of companies, and each system has its own distinct pro's and con's.

The following systems are restricted to volumes of one liter or less of culture due to limitations in O_2 transfer. These include the following:

- Attachment cell culture using T-flasks, roller bottles, and other carriers such as Cytodex™ (Amersham Pharmacia Biotech), CultiSpher® (HyClone), and Fibra-Cel® disks (New Brunswick Scientific).
- Spinner flasks for use with a stir plate apparatus. One can use suspension cell lines or attached cells grown on carrier surfaces.
- Shake flasks in systems that range from individual platforms placed into incubators to self-contained chambered shakers allowing independent control of temperature and CO_2 gassing. Shake flask systems are mainly used for growth of suspension cell lines.

Medium volumes, more complex than above, include the following:

- CellCube® (marketed by Corning) is a closed loop perfusion system for the culture of attachment-dependent cell lines.
- Wave Bioreactor (*www.wavebiotech.com*) consisting of a fixed rocker base and a disposable plastic Cellbag. This system can

be used in volumes from 100 ml to 10 L for both suspension cells and cells on carriers.

Ideal for larger volumes (ranging from 1 to 10,000 L), although more complex and costly, are the following:

- Stirred tank bioreactors come in all shapes and sizes. They have modular designs, can be upgraded and are versatile, allowing one to control dissolved O_2, airflow, temperature, impeller type and speed, pH, nutrient addition, and vessel size. One can also perform two-compartment fermentation through the use of dialysis membranes separating cells from the medium. You can vary the mode of culture using either a fed-batch or perfusion process to maximize protein expression. These bioreactors are best suited for growing suspension cell cultures. However, a fibrous-bed of polyester disks may be employed as a matrix for high-density growth of cells immobilized on the disks for use in the stirred tank bioreactors.
- Hollow fiber bioreactors are composed of a matrix of hollow fibers that separate the bulk of the culture medium from the cell mass by means of hollow-fiber walls, allowing production of high-density cultures of viable cells in the extracapillary space. Cells are nourished by nutrients circulating in the ICS (intracapillary space) medium that readily diffuse across the hollow-fiber membrane. This system is ideal for production of secreted proteins, specifically monoclonal antibodies, and can be used for both suspension- and attachment-dependent cells.

Gene Expression Analysis

Following gene transfer, the time has come to determine how successful your expression efforts have been. This is done by analysis of either cells or cell lysates in the case of intracellular or membrane proteins, or conditioned medium in the case of secreted proteins. It is presumed at this point that you have specific detection reagents for the expressed protein, that the protein is tagged for detection, or that there is a specific functional assay in place for detecting the protein's biological activity. If the expressed protein is fairly well characterized, there are likely to be commercial antibodies for Western blot analysis and/or enzyme-linked immunosorbant assay (ELISA) detection.

The Pro's and Con's of Tags

If the expressed protein is not well characterized or completely novel, then it is useful to have an epitope tag (e.g., FLAG, HA,

His_6, c-myc as described above) fused to the expressed protein. This will enable detection of protein expression in the absence of specific reagents and will aid in purification. Various tag detection reagents are commercially available through various vendors. In the case of receptors, tagging can be particularly useful when trying to determine if the receptor is expressed onto the cell surface. For example, HA (hemagglutinin) tagging has been used to detect cell surface staining of 7TM receptors (Koller et al., 1997). In our experience we have relied extensively upon the use of immunoglobulin Fc fusions as a reporter to monitor expression. Fc fusions are easy to detect both by ELISA and Western blotting using commercially available reagents. Recently we have employed the Origen technology (IGEN, *www.igen.com*) based electrochemiluminescence detection method (Yang et al., 1994) which we have adapted for the direct detection of Fc expression from individual colonies (Trill, 2001). Following expression of Fc fusions, one can often utilize Fc fusion proteins directly in screens. Alternatively, a protease cleavage site can be engineered for removal of the Fc following purification.

The expression of novel or uncharacterized proteins requires special consideration for detection. On the one hand, there is likely to be very little known about what regions of the protein are important for function. Thus one would ideally like not to have additional residues such as tags, which could potentially interfere with folding (e.g., activity) or expression. However, since there are usually no specific detection reagents or functional assays available, it is often necessary to add a tag anyway in order to detect and purify the protein. Alternatively, one could consider the production of antibodies raised to antigenically pronounced regions. Certain vendors will do both the peptide synthesis and immunization. However, this will take several weeks to months, and there is no guarantee that high titer or neutralizing antibodies will be obtained. Since turnaround time is usually a critical parameter for expression projects, most researchers will take the chance of adding an epitope tag for initial expression. At the same time, if cost and resource are not prohibitive, it is also safest to express both tagged and untagged versions and to prepare peptide antiserum in the process.

Most commercial expression vectors contain modular regions for the optional incorporation of tags. This is a convenient way to fuse tags to an expressed protein. However, the options for tag fusions in commercial vectors are frequently limited to C-terminal tags, which are more prone to clipping through the action of carboxy peptidases in the cell. Furthermore the fusions in most

commercial vectors also include a significant number of extraneous residues derived from the polylinker region that could affect protein folding and activity. Thus, while the one-size-fits-all commercial vectors are generally suitable for initial expression, you will probably want to design more precise fusion constructs with either N-terminal or C-terminal tags.

Tags can provide other benefits for expression. For example, some researchers have found that the use of large soluble tags such as GST can enhance the solubility of certain proteins, which favors the production of active protein (Davies, Jowett, and Jones, 1993; Weiss et al., 1995; Ciccaglione et al., 1998). We have also observed that the additional C-terminal immunoglobulin Fc fusions sometimes result in the enhanced production of certain proteins secreted from CHO cells. Thus the addition of tags for detection and purification remains an empirical process, as does the choice of a system in which to express the protein.

Functional Assays

In many cases the most efficient way to screen for expression is not through direct detection of the protein itself but through some kind of functional assay for the expressed protein's biological activity (e.g., apoptotic, chemotactic, proliferative, or enzymatic). Crude cellular extracts or conditioned medium containing secreted proteins can sometimes be directly screened for biological activity. Functional assays are particularly useful when screening for the expression of a receptor whose ligand is known. In this case, clones can be directly screened for cellular responses to added ligand. Calcium-mobilization and cAMP assays are two of the most commonly used methods of detecting signal transduction through G-coupled-protein receptors.

TROUBLESHOOTING

Finally, after weeks or even months of selection, you have isolated clonal cell lines that should be expressing large quantities of protein. However, Western, ELISA, or functional assays are performed, and they show that little or even no protein is being expressed. What can and should you do now? There are many possible explanations for why you fail to detect a protein.

Confirm Sequence and Vector Design

The first thing to do if you haven't already done so is to double-check the original design of the expression vector and the con-

firmed sequences. In some cases an overlooked point mutation or mistake in the original design is the problem. Ideally such problems are best uncovered before weeks of work have been devoted to selecting lines. It cannot be overstated that one should make every effort to check and double-check sequences and vector designs in order to ensure that this never happens. It is also a good idea to first confirm expression by performing transient assays (e.g., in COS cells).

Once you have ruled out problems with the expression construct, there are a few obvious places where problems may be occurring. First, one could perform Northern blot analysis or RT-PCR to determine if any message is produced. The second possibility is that the protein tag is being proteolytically removed. Many C-terminal tags are prone to clipping as mentioned above. If clipping is the problem and there is no other way of detection, it will be difficult to prove that your protein is being expressed and even more difficult to purify it. However, you might be lucky and find that the expression levels are high enough to enable detection through direct staining of SDS-PAGE gels either with Coomassie Blue or silver stain. If direct detection is ambiguous, then you will either have to wait for specific peptide antibodies to detect the untagged protein or have to modify the expression to limit proteolytic digestion (e.g., removal of arginine-serine rich sequences that may be the target of proteolysis). Baculovirus, being a lytic system, is particularly prone to protease problems. In some cases researchers have even resorted to adding protease inhibitors directly to the infection in order to inhibit proteolysis as it is occurring (Pyle et al., 1995). This is not highly recommended, however, since the protease inhibitors also tend to inhibit cellular and viral functions. On the other hand, the addition of protease inhibitors upon harvest and lysis is imperative in order to prevent such proteolysis during purification.

Secreted proteins present their own particular set of issues related to processing and trafficking the protein out of the cell. If a protein is not naturally secreted to high levels, one may find that the native signal peptide sequence does not guide efficient secretion into the ER (endoplasmic reticulum). In these cases one may consider replacing the native signal sequence for a known efficient signal peptide sequence. For the *Drosophila* S2 system, we have utilized a signal sequence derived from chaperone protein HSC3 (*Drosophila* BIP) (Rubin et al., 1993). This sequence has been adapted into commercial vectors sold by Invitrogen.

Investigate Alternate Hosts

The choice of an expression host is often a critical parameter for efficient expression, but it is not usually possible to predict which system will work for a particular protein. Certain hosts may contain the necessary processing machinery while others do not. Thus it is often worthwhile to switch to a new system if expression is not initially detected. One can learn a great deal by performing transient expression assays in different hosts to narrow the field of compatible host systems. This is best done first, before all the time and effort is expended in the selection of stable cell lines.

Finally, one of the most difficult problems that you can face is expression of an inactive protein. This is particularly troublesome when expression levels are good and the protein appears fully soluble. In many cases the protein requires additional processing that is not supplied by the host cell. Alternatively, the host cell may lack a particular cofactor or signaling component that is necessary to establish activity. For example, G-coupled-protein receptors signal through specific G-proteins, interacting directly through one of several different G alpha subunits. The absence of a specific G-protein subunit could impair receptor function when expressed in certain hosts. Fortunately this specific problem can be ameliorated by co-transfection with one of several promiscuous G-protein subunits that will couple functionally with a broader range of receptors (Offermanns and Simon, 1995). However, not all cofactors are quite so well characterized to enable their supplementation. In most cases, if the cofactor is not endogenous to that host, then expression of active protein will not be directly possible. Again, exploring a number of different cellular hosts will often be the best approach to achieving the desired product.

A Case Study of an Expressed Protein from cDNA to Harvest

It is easy to explain how one goes about expressing a particular gene of interest, but how does this relate to real laboratory situations? The following example of a gene, which we will call ABCD, may help illustrate this.

Information concerning the gene has been published, and its sequence is also contained in the GenBank database. The gene contains 349 amino acids, including the signal peptide. Northern blot analysis indicates that the gene is highly expressed in the vascular endothelium. It is a secreted, cysteine-rich, glycosylated protein that has both chemoattractant and mitogenic

activity. We wish to use this protein as a reagent in screening assays, as a comparison to a homologue we previously expressed. We do not have the gene for this protein, but analysis of in-house cDNA libraries indicates that the gene is present in one of our clones.

The homologue, designated ABCD-Like, was originally expressed in a baculovirus expression system using a N-terminal His_6 epitope tag, separated from the gene of interest with a factor Xa cleavage site. This strategy was based on published reports of similar proteins. Expression levels were very low, which led to purification problems. We were then forced to consider alternative strategies. ABCD-Like was recloned into a mammalian expression vector as a C-terminal Fc fusion protein and expressed in the CHO-DG-44 cell line, which had been adapted for suspension growth in a serum-free medium. We were able to express ABCD-Like at very high levels.

Let's revisit our original three questions, and determine what steps we need to take. We know what the gene is, we know where to find it, and we know a number of facts about ABCD, including its intended use. Finally, we have an idea of what expression system to use based on previous work with the homologue ABCD-Like.

Using the sequence we located in GenBank, a PCR primer is designed to trim the 5′ end of the gene and add a unique restriction site. To the 3′ end of the gene, sequences encoding a Factor Xa cleavage site and a unique restriction site are likewise introduced. The generated PCR fragment could be cloned into our pCDN/Fc vector (Aiyar et al., 1994) as an Fc fusion protein. Upon positive sequencing results, the resulting plasmid, pCDN-ABCD/Fc is linearized and electroporated into our CHO cell line and selected for resistance to maintenance medium without nucleosides, since our plasmid contains the mouse dhfr gene. The colonies that arise are assayed using a Fc sandwich assay with an Origen analyzer, and the high expressors are expanded. A single clone is eventually scaled up into flottles (a cross between a flask and a bottle). A flottle is often referred to as a modified Fernbach Flask, and is available from Corning. The clone is grown for 13 days to produce enough medium for purification and testing. N-terminal sequence analysis of the purified protein revealed the correct mature protein sequence, indicating that processing had occurred. Western blot analysis revealed the presence of two smaller bands. N-terminal analysis of these bands indicated that the protein was cleaved several amino acids before the N-terminus of the Fc region.

The entire process, from inception to purification, took less than three months to complete. There still was the problem of determining how to eliminate the extraneous cleavage products. Analysis of the amino acid sequence revealed a "possible" arginine-rich protease cleavage site. Site directed mutagenesis was performed to eliminate the suspect amino acids. Subsequent re-expression in CHO, using the aforementioned techniques, demonstrated that the correct uncleaved protein was obtained.

SUMMARY

The expression of recombinant proteins in eukaryotic systems represents an important technological advance in the study of the biological function of proteins. This technology enables the isolation of authentic, post-translationally modified proteins in large quantities without having to purify them from a native source. In pharmaceutical research and development, recombinant proteins are used to supply high-throughput drug screens, functional studies, structural biology, and therapeutic agents. In this chapter we have discussed the process by which one goes about finding a gene for expression of a protein, choosing an appropriate expression host, choosing an appropriate vector for that host, cloning the gene into the vector, transfection of the recombinant vector into the host, isolating cells that are expressing the protein, and scaling protein expression for purification. We have also discussed several possible pitfalls commonly encountered and suggestions on how best to fix these problems. The practical considerations on these topics discussed in this chapter are intended to help guide one through the vast array of possible expression systems that one has to choose from including many commercial systems that bring recombinant protein expression technology to virtually anyone who wants to use it.

SECTION B: WORKING WITH BACULOVIRUS

PLANNING THE BACULOVIRUS EXPERIMENT

Is an Insect Cell System Suitable for the Expression of Your Protein?

The first choice for recombinant overexpression of a plain vanilla cytoplasmic protein is nearly always *E. coli*. For many of the remaining proteins that are membrane bound, covalently modified, secreted, or components of multiprotein complexes, expression in eukaryotic cells is the system of choice. Expression

in cells from higher organisms can also be a solution for proteins that, when expressed in bacteria, are insoluble or are expressed as truncated products due to proteolysis, premature translational termination, or the presence of rare codons (Pikaart and Felsenfeld, 1996). In some instances it may be useful to express soluble protein from a eukaryotic source as a "gold standard" to compare with refolded protein from a bacterial source.

The most commonly used eukaryotic cells for recombinant protein expression are derived from mammalian or insect tissues and utilize either viral- or plasmid-based vehicles to transduce your gene of interest. This section will address using baculovirus infection of insect cells as a way to provide modest levels (1–10 mg/L) of proteins in a reasonably quick time frame (7–10 days).

Although recombinant baculoviruses are most often used to infect cultured insect cells and caterpillars, a more recent development has been their use as transfer vectors for mammalian cells (Condreay, 1999; Kost and Condreay, 1999). Several types of mammalian cells are capable of baculovirus uptake and transient expression of recombinant genes, but are incapable of producing progeny virus. This technique has proved particularly valuable for introducing genes into cells that are notoriously difficult to transfect using more traditional methods. It is likely that recombinant baculoviruses incorporating a more specific uptake mechanism by an established receptor-ligand pair will make this approach more common in the future.

Should You Express Your Protein in an Insect Cell Line or Recombinant Baculovirus?

Insect cell expression is relegated to the creation of a cell line or to a lytic infection with recombinant baculovirus infected cells. General descriptions for the creation of stable insect cell lines are given by McCarroll and King (1997), Ivey-Hoyle (1991), and Benting et al. (2000). Invitrogen and Novagen sell reagents to produce such lines and provide detailed manuals available on their Web sites. The most important differences in the two approaches lies in the level of attention needed to maintain the various cell lines, in the elapsed time before it is possible to evaluate expression of a given construct, and in the relative ease of expressing multiprotein complexes (Table 16.4). There are several instances where expression in a baculovirus system makes sense as a first choice. Since insect cells can be infected with multiple different baculoviruses, each expressing an individual protein, this system requires no additional time to analyze the expression of

Table 16.4 Comparison of Protein Expression Systems

	Insect Cell Line	Baculovirus	*E. coli*
Nature of expression	Inducible or constitutive cell line	Lytic viral infection	Inducible
Modifications • Glycosylation • Myristylation • Palmitoylation • Sulfation • Isoprenylation • GPI linker addition	System of choice, since cells are not dying at the time of highest expression	Modifications may not occur efficiently, since cells are dying at the time of highest expression	Not present
Codon preference	Bias against certain codons in *Drosophila*	Very little codon bias	Bias against certain codons
Expression of >1 protein	Time-consuming	Straightforward	Can be done
Ease of cell culture	Easy	Cells need careful attention; plaque assay must be mastered	Easy
How soon after making my expression plasmid will I have 1 L of cells to examine?	3–4 weeks after transfection	7–10 days after transfection	1 day after transformation
Storage	Frozen cells	Virus at 4°C or Frozen stocks	Frozen cells or plasmid DNA

multiple-protein complexes. A comparable insect cell line may require months for the sequential isolation of clonal cell lines that express more than one protein. Baculoviral genes show little evidence of codon bias (Ayres et al., 1994; Levin and Whittome, 2000), and expression in this system may be preferable with genes that contain numerous rare codons for *Drosophila*. If a gene has suspected cellular toxicity, a cell line may be unattainable, making baculovirus a more suitable expression choice. Perhaps the most common reason for using the baculovirus approach is the rapidity with which recombinant protein can be obtained. For the expression of a soluble cytoplasmic protein, it is possible to obtain protein from as much as a few liters of baculovirus infected cells within 7 to 10 days from the initial transfection. The analysis of a comparable amount of cells from a cell line would take 3 to 4 weeks from the initial transfection.

Expression from an insect cell line is preferable for proteins that are secreted or require a modification such as glycosylation or acylation. Most protein expression from baculovirus late or very late promoters occurs just prior to cell lysis, and as a result the

cellular machinery for protein export or modification may be compromised.

Procedures for Preparing Recombinant Baculovirus

This chapter will not discuss the common protocols available for baculovirus expression. References that contain good protocols for cell culture and handling of virus are King and Possee (1992), O'Reilly, Miller, and Luckow (1992), and Murphy et al. (1997). Additionally manuals for cell culture and baculovirus expression can be obtained from the Web sites at Invitrogen, Novagen, Clontech, and Life Technologies. Miller (1997) has details of baculovirus biology.

Criteria for Selecting a Transfer Vector

Epitope Tags

Baculoviruses are most easily formed by homologous recombination between viral DNA containing a lethal deletion and a transfer vector plasmid containing the gene of interest flanked by viral sequences. There are dozens of baculovirus transfer vectors commercially available, and manufacturers are coming up with new ones all the time. Good sources of vectors are Novagen, Pharmingen, Clontech and Invitrogen; check their Web sites for new ones that are not described in the catalogs. Commercial vectors often include sequences for "tags" that are useful for monitoring protein expression by immunoblot analysis. If the protein needs to be purified, the inclusion of an epitope tag that can be bound to an affinity resin (e.g., anti-Flag antibody resin for the Flag® epitope or a metal chelate resin for His_6 tagged proteins) will minimize the processing steps needed to obtain homogeneous recombinant protein.

Choice of Promoter

Most proteins are expressed from transfer vectors containing the very strong p10 or polyhedrin promoters that are most active very late (20–72 hours postinfection). Since expression from these promoters occurs at a time when such modifications as glycosylation are compromised because of the cytopathic effects of the viral infection, modified proteins are best expressed using the moderately strong basic protein or 39K promoters that are active at slightly earlier times (12–24 hours postinfection) (Hill-Perkins and Possee, 1990; Murphy et al., 1990; Jarvis and Summers, 1989; Sridhar et al., 1993; Pajot-Augy et al., 1999).

Cloning Strategy

An alternative approach to obtaining a recombinant baculovirus is available from Life Technologies (Bac-To-Bac™). Instead of recombination occurring in the insect cell, the recombinant viral DNA is recovered from *E. coli* and subsequently transfected into insect cells (Luckow et al., 1993). A disadvantage of this system is that it requires the manufacturer's limited set of transfer vectors. In addition there are less commonly used procedures for making the viral recombinant DNA in vitro (Ernst, Grabherr, and Katinger, 1994; Peakman, Harris, and Gewart, 1992) or in yeast (Patel, Nasmyth, and Jones, 1992). The ability to make recombinants in vitro is essential for creating baculovirus expression libraries, and the in vitro procedure may be required for the expression of proteins that are toxic to insect cells.

Control Elements

Although insect cells have the ability to splice RNA, often just the open reading frame with a minimal amount of untranslated flanking regions is inserted into the transfer vector. It is probably better to utilize baculoviral polyadenylation sequences (often present on the transfer vector) rather than substituting one such as the SV40 terminator (van Oers et al., 1999). Upstream sequences do have an influence on the rates of RNA transcription and/or protein translation, but no pattern has yet emerged (Luckow and Summers, 1988). There is limited evidence for a consensus base context around the initiating ATG (AAAATGA: Ranjan and Hasnain, 1994; Ayers et al., 1994), although experiments with transfected cells suggest a preference for A or T immediately downstream of the initiation codon (Chang, Kuzion, and Blissard, 1999). This apparent lack of a highly preferred initiation sequence ("Kozak" sequence) makes it possible to transplant inserts from bacterial expression vectors directly into a baculovirus transfer vector. As an added benefit, the presence of bacterial sequences upstream of open reading frames may enhance baculovirus expression of the gene of interest (Peakman et al., 1992).

Which Insect Cell Host Is Most Appropriate for Your Situation?

Three cell lines are commonly used for baculovirus expression; Table 16.5 illustrates differences among them. Sf21 cells are ovarian cells derived from *Spodoptera frugiperda* (fall army worm) and Sf9 are a subclone of Sf21. *T. ni* (available as High Five™) cells are from *Trichoplusia ni* egg cell homogenates. For

Table 16.5 Commonly Used Cell Lines for Baculovirus Expression

	Sf9	Sf21	*T. ni* (High Five™)
Initial transfection	✓	✓	
Plaque assay	✓		
Expression of secreted proteins			✓
Expression of cytoplasmic proteins	✓	✓	✓
Adaptation to suspension culture	Easy	Easy	Challenging due to clumping
Media	Serum-containing and serum- or protein-free preparations	Same as Sf9	Same as Sf9 and some made specifically for these cells

initial transfections, Sf9 or Sf21 cells are best because they produce large amounts of virus. Plaque assays are best done with Sf9 or Sf21 cells for the same reason. Sf9 cells are preferred for plaque assays since the plaques on these cells have sharply defined edges with clearer centers compared to plaques on Sf21 monolayers. For expression, *T. ni* often produces more protein than the other two lines, but due to its adherent and clumping habit, it is more difficult to adapt to suspension culture (Saarinen et al., 1999). It is generally best to have two insect cell lines growing—either Sf9 or Sf21—and cells from *T. ni*.

It can't be stressed enough that success with the baculovirus system depends on healthy cells and that careful attention to providing optimal growth conditions will avoid many common expression problems. Insect cells are grown at 27 to 28°C in a non-CO_2 incubator. They can be grown at room temperature on the benchtop, but because of possible unanticipated temperature fluctuations, an incubator is preferred. The best temperature control requires an incubator equipped with cooling capability.

Unfortunately, these cells have a narrow range of densities at which they will grow—between around 1×10^6 and 4×10^6/ml (Sf9 and Sf21 in serum-containing media) or slightly lower densities for *T. ni* cells. Slightly higher densities can be obtained in serum-free media. Cells will cease growing if diluted too much, and they will begin to die if allowed to remain at the higher densities for more than a day or two. With a doubling time of around 24 hours, this means they must be split every two to three days. The cells are generally passaged continuously until there is a noticeable

increase in the doubling time, or a decreased sensitivity to viral infection. This seems to occur when cells are grown in serum-free media after about 30 to 50 passages. At that time a fresh culture is started from frozen cells.

IMPLEMENTING THE BACULOVIRUS EXPERIMENT

What's the Best Approach to Scale-Up?

Cells are initially grown in 75 or 150 cm^2 flasks and transferred to suspension culture in spinner flasks or shake flasks for scale-up. An advantage of suspension culture is that cells are subjected to less handling, so they will attain higher densities than in stationary flasks. The cell volume should not exceed 50% of the flask volume as the oxygen demand increases greatly after infection. There are several types of media that can be used, including serum-free preparations and formulations specifically made for *T. ni* cells. Cells grown in the presence of serum may require a weaning period before being adapted for growth in serum-free media. For the expression of cytoplasmic proteins, all types of media will give adequate expression levels, but for secreted proteins a low-protein or serum-free preparation may be preferred. An important consideration for secreted proteins is that serum-free media often contains Pluronic (a detergent). If a downstream purification step requires a media concentration step, pluronic micelles will be concentrated as well, and this may affect subsequent chromatography efforts.

If large quantities of protein are required, it is worth comparing protein expression with a selection of both serum-supplemented and serum-free media preparations as part of an optimization effort. Unfortunately, no one media preparation seems to be optimal for all proteins. Many manufacturers of serum-free media occasionally have not been able to meet consumer demand at one time or another, so it is worth identifying an alternate commercial source for an acceptable serum-free preparation.

Virus stocks should be prepared in serum containing media or serum-supplemented serum-free media. The presence of pluronic in the growth medium may result in decreased virus production (Palomares, González, and Ramirez, 2000).

What Special Considerations Are There for Expressing Secreted Proteins?

In general, the levels of secreted proteins from baculovirus infected cells are low (less than 10 mg/L), but there are examples

of proteins that are secreted at levels greater than 100 mg/L (Mroczkowski et al., 1994; George et al., 1997). Secreted proteins require a signal sequence for export to the media; commercial vectors (available from Stratagene, Pharmingen, and Novagen) that provide a signal sequence or the native signal sequence can be used. A bacterial signal peptide will also direct secretion of eukaroytic proteins in insect cells (Allet et al., 1997). It may be worth trying several different signal sequences, for no one sequence seems to work best for all proteins (Tessier et al., 1991; Mroczkowski et al., 1994; Golden et al., 1998). Of the commonly used cell lines, *T. ni* cells often produce higher levels of secreted proteins (Hink et al., 1991; Wickham and Nemerow, 1993; Mroczkowski et al., 1994).

Although the baculovirus system can quickly provide recombinant protein, it may not be the optimal approach to obtaining the highest levels of secreted protein possible. It is worth taking the time in parallel with baculovirus efforts to produce an insect cell line that overexpresses the gene of interest (Jarvis et al., 1990; Farrell et al., 1998). That way a backup expression system is in place in case the levels of protein from the baculovirus infection are intolerably low.

What Special Considerations Are There for Expressing Glycosylated Proteins?

Insect cells perform N-linked glycosylation at sites that are similarly targeted in mammalian cells, but in insect cells the modifications are of the high mannose type with inefficient trimming of the core sugar residues or just the trimannosyl core structure (reviewed in Altmann et al., 1999). There are several approaches available to obtain more complex glycosylation patterns typical of mammalian cell expression. Infection of cells from *Estigmene acrea* (available from Novagen) may produce a more mammalian-type of glycosylation pattern (Wagner et al., 1996a; Ogonah, 1996). Co-expression of a mammalian glycosyltransferase may result in a more complex glycosylation pattern (Wagner, 1996b; Jarvis and Finn, 1996; Jarvis, Kawar, and Hollister, 1998). Similarly, use of a Sf9 host cell that has been engineered to constitutively express a glycosyltransferase can be used for the same effect (Hollister, Shaper, and Jarvis, 1997). The addition of mannosamine to infected insect cells can increase the level of terminal N-acetylglucosamine structures in recombinant proteins (Donaldson et al., 1999).

What Are the Options for Expressing More Than One Protein?

A significant advantage to the baculovirus expression system is the ease of expressing multiple proteins. The ability to co-express proteins allows for the expression of heterodimers (Stern and Wiley, 1992; Graber et al., 1992) and even larger multiprotein complexes such as virus particles (Loudon and Roy, 1991). In one notable case, co-expression of seven herpesvirus proteins from seven different baculoviruses allowed replication of a plasmid containing a herpesvirus origin of replication (Stow, 1992). Cells can be simultaneously infected with multiple baculoviruses expresing different proteins, or recombinant baculoviruses can be made that have up to four separate promoters each regulating a different gene (Weyer, Knight, and Possee, 1990; Belyaev, Hails, and Roy, 1995). Vectors that express two or more proteins are available commercially (Pharmingen, Clontech, and Novagen). In contrast to mammalian cells, baculovirus infected insect cells do not make efficient use of an internal ribosomal entry site (IRES) sequence for the expression of multiple proteins (Finkelstein et al., 1999).

Co-expression also enables one to express modifiers of the target protein. Examples of this are co-expression of biotin ligase to obtain biotinylation (Duffy, Tsao, and Waugh, 1998), prohormone convertase to obtain proteolytically processed TGFβ1 (Laprise, Grondin, and Dubois, 1998), and signal peptidase to enhance processing efficiency for a secreted protein (Ailor and Betenbaugh, 1999).

How Can You Obtain Maximal Protein Yields?

Optimizing the host cell selection, cell density at infection, multiplicity of infection, type of media, and the time of harvest will allow maximal recovery of the protein of interest (Licari and Bailey, 1992; Power et al., 1994). All five conditions are interdependent, and it is possible that protein yields may be equal from a relatively low multiplicity of infection (moi) of dilute cells harvested after five to six days compared to high moi infection of a dense culture harvested after two days. If cells are being grown on a larger scale (e.g., in suspension cultures in 1 L spinner flasks), expression optimization should be done under such conditions. Although it may be convenient to examine infection conditions in small culture dishes such as a 24 well cluster dish, optimal parameters for cells growing in a stationary flask are likely to be very different from cells growing in suspension. A reasonable strategy

to start an optimization procedure is to infect 200 ml of 1.5×10^6 cells/ml growing in three 500 ml spinner flasks with moi's of 0.1, 1, and 10, and then to remove 10 ml aliquots of cells every 24 hours for 5 days. For intracellular proteins, the cells should be lysed as they would for downstream purification, and both the soluble and insoluble fractions examined for the presence of the protein of interest.

What Is the Best Way to Process Cells for Purification?

For cytoplasmic proteins, cells are recovered by pelleting and washed with a buffer to remove media components. Infected cell pellets can be further processed or stored frozen until needed. Insect cells can be lysed by hypotonic lysis after incubation in a buffer lacking salt; disruption is completed by using a dounce homogenizer. Cells can also be lysed with a buffer containing a detergent such as Triton, CHAPS, or NP-40. Sonication should not be used as lysis conditions are difficult to control and reproduce from one preparation to another. It is important to keep the preparation on ice and perform cell lysis in the presence of a cocktail of protease inhibitors to avoid proteolysis. The lysate should be cleared by centrifugation at $100{,}000 \times g$ to remove large aggregates and insoluble material. Cleared lysates are then ready for chromatographic purification.

For nuclear proteins, nuclei are obtained following hypotonic lysis or detergent lysis and salt extracted to remove nuclear-associated proteins. Secreted proteins are generally recovered from cell-free clarified supernatants by direct adsorption to a chromatographic resin.

TROUBLESHOOTING

Western blot or a biochemical analysis of transfected cells should indicate expression of the gene of interest three to seven days after the transfection. It is rare that a protein is not expressed at all in baculovirus infected cells, and an observed lack of protein expression may be due to a variety of situations.

Suboptimal Growth Conditions

Many problems with baculovirus expression can be traced to suboptimal cell growth conditions. Healthy cells should show high viability (>98%) and have a doubling time of around 24 hours. If either of these conditions is not met, efforts should be directed toward getting a more robust cell stock. Start with frozen cells

from the American Type Culture Collection (ATCC) or a commercial source, and use heat-inactivated serum that has been certified for insect cell culture (available from Life Technologies) in media without antibiotics. Grow cells initially in stationary flasks as it is easier to monitor their progress. For passaging in flasks, do not scrape or harshly pipette liquid over the cell monolayer. Instead, sharply rap the side of the flask to dislodge as many cells as possible. Remove the cells and media, and distribute these to new flasks containing additional fresh media; add back fresh media to the remaining cells that have adhered to the original flask for further growth. Once cells are growing in flasks in serum-containing media, the cells from several flasks can be pooled for growth in suspension and/or adaptation to serum-free media.

Viral Production Problems

A lack of protein expression may be due to inefficient production of virus in the initial transfection step. The virus may benefit from an amplification step by removing about 100 μl of the media from transfected cells and adding it to freshly plated uninfected cells in a T25 Flask. Cells from this infection should show evidence of viral cytopathic effect and demonstrate protein expression after three to four days. Transfections are generally performed with either a liposome mediated- or a calcium phosphate procedure provided as a "kit" with commercial viral DNA. It is important to follow the manufacturer's instructions carefully. Plasmid DNA should be very pure—preparations made with an anion exchange matrix or cesium chloride banding work well. The DNA should be sterilized by ethanol precipitation and resuspended in a sterile buffer. Viral DNA is very large and susceptible to shearing; use a sterile cut-off blunt pipette tip for transfers and never vortex it. Insect cells must be healthy (>98% viable) and actively growing in log phase when used for a transfection. If possible, transfections should be done in cells growing in serum-containing media to enhance the production of virus. Transfection conditions can be optimized with wild-type baculoviral DNA that produce distinctive polyhedrin in infected cells. Similarly optimization can be done with viral DNA from a baculovirus recombinant that encodes an easily assayed protein (e.g., beta-galactosidase). The presence of insert DNA incorporated into progeny virus can be determined by PCR or Southern blot analysis.

Mutation

A lack of expression may be due to an unwanted mutation or the presence of unintended upstream ATG sequences. The DNA

encoding the open reading frame for the gene of interest in the transfer vector plasmid should be verified by sequencing to rule out this possibility. There is one report of translational initiation occurring at a non-ATG codon, AUU, in a baculovirus expressed protein (Beames et al., 1991). Occasionally larger transfer vectors (>8 kb) suffer deletions and will be unable to give rise to recombinant virus containing an intact gene of interest. Transfer plasmids should be digested with a few restriction enzymes to be sure this has not happened. The use of smaller transfer vectors (<6 kb) often eliminates such genetic instabilities.

Solubility Problems

In general, many recombinant proteins that are insoluble in *E. coli* become soluble when produced in insect cells. There are a few proteins that are completely insoluble in insect cells and will need to be denatured and refolded. More often, a protein will be partially soluble, and for these situations, infection at a low multiplicity (<1 virus/cell) and harvest at an early time (<36 hours) is usually beneficial. Co-expression with protein disulfide isomerase or a chaperonin molecule may enhance the percentage of nonaggregated secreted proteins (Hsu, Eiden, and Betenbaugh, 1994; Hsu et al., 1996; Ailor and Betenbaugh, 1998). Proteins that are susceptible to degradation may benefit from the addition of a signal sequence and export into the media (Mroczkowski et al., 1994). Use of viral DNA from a baculovirus lacking a viral protease (available from Novagen) may also help in the expression of proteins that are degraded in insect cells.

SUMMARY

Baculoviruses represent a versatile, relatively quick, minimal technology approach to recombinant gene expression, especially for proteins that are insoluble in *E. coli* or are covalently modified. All of these features make baculovirus expression an excellent complement to a bacterial expression system, especially for the production of proteins at levels <10 mg. If a cloned gene is on hand, the process of obtaining a recombinant baculovirus and analyzing the expression from approximately 1 L of infected cells can be completed in less than 2 weeks. Recombinant viruses can incorporate large amounts of DNA, making the expression of multiple genes from one virus possible. Additionally, insect cells can be infected with multiple viruses, allowing the expression of entire signaling pathways or protein/modifier combinations. A drawback

to the use of this system is the inability to produce glycoproteins with complex N-linked glycans typical of mammalian cells. There are various approaches to increase this capability in insect cells, but none are truly optimal. Looking to the future, baculoviruses may have a utility as gene-delivery vehicles for protein expression not only in insect cells but also in a wide variety of mammalian cells.

BIBLIOGRAPHY

Ailor, E., and Betenbaugh, M. J. 1998. Overexpression of a cytosolic chaperone to improve solubility and secretion of a recombinant IgG protein in insect cells. *Biotechnol. Bioeng.* 58:196–203.

Ailor, E., and Betenbaugh, M. J. 1999. Modifying secretion and post-translational processing in insect cells. *Curr. Opin. Biotechnol*. 10:142–145.

Ailor, E., Pathmanathan, J., Jongbloed, J. D., and Betenbaugh, M. J. 1999. A bacterial signal peptidase enhances processing of a recombinant single chain antibody fragment in insect cells. *Biochem. Biophys. Res. Commun.* 255:444–450.

Aiyar, N., Baker, E., Wu, H. L., Nambi, P., Edwards, R. M., Trill, J. J., Ellis, C., and Bergsma, D. J. 1994. Human AT1 receptor is a single copy gene: characterization in a stable cell line. *Mol. Cell Biochem*. 131:75–86.

Allet, B., Bernard, A. R., Hochmann, A., Rohrbach, E., Graber, P., Magnenat, E., Mazzei, G. J., and Bernasconi, L. 1997. A bacterial signal peptide directs efficient secretion of eukaryotic proteins in the baculovirus expression system. *Prot. Expr. Purif*. 9:61–68.

Altmann, F., Staudacher, E., Wilson, I. B., and März, L. 1999. Insect cells as hosts for the expression of recombinant glycoproteins. *Glycoconj. J*. 16:109–123.

Ames, R. S., Sarau, H. S., Chambers, J. K., Willette, R. N., Aiyar, N. V., Romanic, A. M., Louden, C. S., Foley, J. J., Sauermelch, C. F., Coatney, R. W., Ao, Z., Disa, J., Holmes, S. D., Stadel, J. M., Martin, J. D., Liu, W. S., Glover, G. I., Wilson, S., McNulty, D. E., Ellis, C. E., Elshourbagy, N. A., Shabon, U., Trill, J. J., Hay, D. W. P., Ohlstein, E. H., Bergsma, D. J., and Douglas, S. A. 1999. Human urotensin-II: A potent vasoconstrictor and agonist for the orphan receptor GPR 14. *Nature* 401:282–286.

Ayres, M. D., Howard, S. C., Kuzio, J., Lopez-Ferber, M., Possee, R. D. 1994. The complete DNA sequence of *Autographa californica* nuclear polyhedrosis virus. *Virol*. 202:586–605.

Beames, B., Braunagel, S., Summers, M. D., and Lanford, R. E. 1991. Polyhedrin initiator codon altered to AUU yields unexpected fusion protein from a baculovirus vector. *Biotechniq.* 11:378–383.

Bebbington, C. R., Renner, G., Thomson, S., King, D., Abrams, D., and Yarronton, G. T. 1992. High level expression of a recombinant antibody from myeloma cells using a glutamine synthetase gene as an amplifiable selectable marker. *Bio/Technol.* 10:169–175.

Belyaev, A. S., Hails, R. S., and Roy, P. 1995. High level expression of five foreign genes by a single recombinant baculovirus. *Gene* 156:229–233.

Benting, J., Lecat, S., Zacchetti, D., and Simons, K. 2000. Protein expression in *Drosophila* Schneider cells. *Anal. Biochem.* 278:59–68.

Boggs, S. S., Gregg, R. G., Borenstein, N., and Smithies, O. 1986. Efficient transformation and frequent single-site, single-copy insertion of DNA can be

obtained in mouse erythroleukemia cells transformed by electroporation. *Exp. Hematol.* 14:988–994.

Boshart, M., Weber, F., Jahn, G., Dorsch-Hasler, K., Fleckenstein, B., and Schaffner, W. 1985. A very strong enhancer is located upstream of an immediate early gene of human cytomegalovirus. *Cell* 41:521–530.

Bueler, H. 1999. Adeno-associated viral vectors for gene transfer and gene therapy. *Biol. Chem.* 380:613–622.

Butters, T. D., Sparks, L. M., Harlos, K., Ikemizu, S., Stuart, D. I., Jones, E. Y., and Davis, S. J. 1999. Effects of *N*-butyldeoxynojirimycin and the Lec3.2.8.1 mutant phenotype on *N*-glycan processing in Chinese hamster ovary cells: Application to glycoprotein crystallization. *Prot. Sci.* 8:1696–1701.

Cannan, S., Roussel, A., Verger, R., and Cambillau, C. 1999. Gastric lipase: Crystal structure and activity. *Biochim. Biophys. Acta* 1441:197–204.

Casasnovas, J. M., Larvie, M., and Stehle, T. 1999. Crystal structure of two CD46 domains reveals an extended measles virus-binding surface. *EMBO J.* 18: 2911–2922.

Cavener, D. R. 1987. Comparison of the consensus sequence flanking translational start sites in *Drosophila* and vertebrates. *Nucl. Acids Res.* 15:1353–1361.

Chambers, J. K., Macdonald, L. E., Sarau, H. M., Ames, R. S., Freeman, K., Foley, J. J., Zhu, Y., McLaughlin, M. M., Murdock, P., McMillan, L., Trill, J., Swift, A., Aiyar, N., Taylor, P., Vawter, L., Naheed, S., Szekeres, P., Hervieu, G., Scott, C., Watson, J. M., Moore, D., Emson, P., Faull, R. L. M., Waldvogel, H. J., Murphy, A. J., Duzic, E., Klein, C., Bergsma, D., Wilson, S., and Livi, G. P. 2000. A G protein-coupled receptor for uridine 5′-diphosphoglucose (UDP-glucose). *J. Biol. Chem.* 275:10767–10771.

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.

Chang, M.-J., Kuzio, J., and Blissard, G. W. 1999. Modulation of translational efficiency by contextual nucleotides flanking a baculovirus initiator AUG codon. *Virol.* 259:369–383.

Christopherson, K. S., Mark, M. R., Bajaj, V., and Godowski, P. J. 1992. Ecdysteroid-dependent regulation of genes in mammalian cells by a *Drosophila* ecdysone receptor and chimeric transactivators. *Proc. Nat. Acad. Sci. U.S.A.* 89:6314–6318.

Ciccaglione, A. R., Marcantonio, C., Equestre, M., Jones, I. M., and Rapicetta, M. 1998. Secretion and purification of HCV E1 protein forms as glutathione-*S*-transferase fusion in the baculovirus insect cell system. *Virus Res.* 55:157–165.

Cockett, M. I., Bebbington, C. R., and Yarranton, G. T. 1991. The use of engineered E1A genes to transactivate the hCMV-MIE promoter in permanent CHO cell lines. *Nucl. Acids Res.* 19:319–325.

Colbere-Garapin, F., Horodniceanu, F., Kourilsky, P., and Garapin, A. C. 1981. A new dominant hybrid selective marker for higher eukaryotic cells. *J. Mol. Biol.* 150:1–14.

Condreay, J. P., Witherspoon, S. M., Clay, W. C., and Kost, T. A. 1999. Transient and stable gene expression in mammalian cells transduced with a recombinant baculovirus vector. *Proc. Nat. Acad. Sci. U.S.A.* 96:127–132.

Cordingley, M. G., Callahan, P. L., Sardana, V. V., Garsky, V. M., and Colonno, R. J. 1990. Substrate requirements of human rhinovirus 3C protease for peptide cleavage in vitro. *J. Biol. Chem.* 265:9062–9065.

Cregg, J. M. 1999. Expression in the Methylotrophic yeast *Pichia pastoris*. In Fernandez, J. M., and Hoeffler, J. P. eds., *Gene Expression Systems: Using Nature for the Art of Expression*. Academic Press, San Diego, CA, pp. 290–325.

Cregg, J. M., Vedvick, T. S., and Raschke, W. C. 1993. Recent advances in the expression of genes in *Pichia pastoris*. *Biotechnol.* 11:905–910.

Davies, A. H., Jowett, J. B., and Jones, I. M. 1993. Recombinant baculovirus vectors expressing glutathione-*S*-transferase fusion proteins. *Bio/Technol.* 11:933–936.

Deen, K. C., McDougal, J. S., Inacker, R., Folena-Wasserman, G., Arthos, J., Rosenberg, J., Maddon, P. J., Axel, R., and Sweet, R. W. 1988. A soluble form of CD4 (T4) protein inhibits AIDS virus infection. *Nature* 331:82–84.

de la Luna, S., Soria, I., Pulido, D., Ortin, J., and Jimenez, A. 1988. Efficient transformation of mammalian cells with constructs containing a puromycin-resistance marker. *Gene* 62:121–126.

Donaldson, M., Wood, H. A., Kulakosky, P. C., and Shuler, M. L. 1999. Use of mannosamine for inducing the addition of outer arm N-acetylglucosamine onto N-linked oligosaccharides of recombinant proteins in insect cells. *Biotechnol. Prog.* 15:168–173.

Duffy, S., Tsao, K.-L., and Waugh, D. S. 1998. Site-specific, enzymatic biotinylation of recombinant proteins in Spodoptera frugiperda cells using biotin acceptor peptides. *Anal. Biochem.* 262:122–128.

Ennist, D. L. 1999. Gene therapy for lung disease. *Trends Pharmacol. Sci.* 20:260–266.

Ernst, W. J., Grabherr, R. M., and Katinger, H. W. 1994. Direct cloning into the *Autographa californica* nuclear polyhedrosis virus for generation of recombinant baculoviruses. *Nucl. Acids Res.* 22:2855–2856.

Evan, G. I., Lewis, G. K., Ramsay, G., and Bishop, J. M. 1985. Isolation of monoclonal antibodies specific for human c-myc proto-oncogene product. *Mol. Cell. Biol.* 5:3610–3616.

Farrell, P. J., Lu, M., Prevost, J., Brown, C., Behie, L., and Iatrou, K. 1998. High-level expression of secreted glycoproteins in transformed lepidopteran insect cells using a novel expression vector. *Biotechnol. Bioeng.* 60:656–663.

Ferrara, N., Winer, J., Burton, T., Rowland, A., Siegel, M., Phillips, H. S., Terrell, T., Keller, G. A., and Levinson, A. D. 1993. Expression of vascular endothelial growth factor does not promote transformation but confers a growth advantage in vivo to Chinese hamster ovary cells. *J. Clin. Invest.* 91:160–170.

Finkelstein, Y., Faktor, O., Elroy-Stein, O., and Levi, B. Z. 1999. The use of bicistronic transfer vectors for the baculovirus expression system. *J. Biotechnol.* 75:33–44.

Fitzgerald, L. R., Dytko, G. M., Sarau, H. M., Mannan, I. J., Ellis, C., Lane, P. A., Tan, K. B., Murdock, P. R., Wilson, S., Bergsma, D. J., Ames, R. S., Foley, J. J., Campbell, D. A., McMillan, L., Evans, N., Elshourbagy, N. A., Minehart, H., and Tsui, P. 2000. Identification of an EDG7 variant, HOFNH30, a G-protein-coupled receptor for lysophosphatidic acid. *Biochem. Biophys. Res. Commun.* 273:805–810.

Fleckenstein, E., Uphoff, C. C., and Drexler, H. G. 1994. Effective treatment of mycoplasma contamination in cell lines with enrofloxacin (Baytril). *Leukemia* 8:1424–1434.

Geisse, S., Gram, H., Kleuser, B., and Kocher, H. P. 1996. Eukaryotic expression systems: A comparison. *Prot. Expr. Purif.* 8:271–282.

George, H. J., Marchand, P., Murphy, K., Wiswall, B. H., Dowling, R., Giannaras, J., Hollis, G., Trzaskos, J. M., and Copeland, R. A. 1997. Recombinant human 92-kDa type IV collagenase/gelatinase from baculovirus-infected insect cells: Expression, purification and characterization. *Prot. Expr. Purif.* 10:154–161.

Gignac, S. M., Brauer, S., Hane, B., Quentmeier, H., and Drexler, H. G. 1991. Elimination of mycoplasma from infected leukemia cell lines. *Leukemia* 5:162–165.

Gignac, S. M., Uphoff, C. C., MacLeod, R. A., Steube, K., Voges, M., and Drexler, H. G. 1992. Treatment of mycoplasma-contaminated continuous cell lines with mycoplasma removal agent (MRA). *Leuk. Res.* 16:815–822.

Golden, A., Austen, D. A., van Schravendijk, M. R., Sullivan, B. J., Kawasaki, E. S., and Osburne, M. S. 1998. Effect of promoters and signal sequences on the production of secreted HIV-1 gp120 protein in the baculovirus system. *Prot. Expr. Purif.* 14:8–12.

Gluzman, Y. 1981. SV40-transformed simian cells support the replication of early SV40 mutants. *Cell* 23:175–182.

Gorman, C. M., Merlino, G. T., Willingham, M. C., Pastan, I., and Howard, B. H. 1982. The Rous sarcoma virus long terminal repeat is a strong promoter when introduced into a variety of eukaryotic cells by DNA-mediated transfection. *Proc. Nat. Acad. Sci. U.S.A.* 79:6777–6781.

Gossen, M., and Bujard, H. 1992. Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proc. Nat. Acad. Sci. U.S.A.* 89:5547–5551.

Graber, S. G., Figler, R. A., Kalman-Maltese, V. K., Robishaw, J. D., and Garrison, J. C. 1992. Expression of functional G protein beta gamma dimers of defined subunit composition using a baculovirus expression system. *J. Biol. Chem.* 267:13123–13126.

Gritz, L., and Davies, J. 1983. Plasmid-encoded hygromycin B resistance: The sequence of hygromycin B phosphotransferase gene and its expression in *Escherichia coli* and *Saccharomyces cerevisiae*. *Gene* 25:179–188.

Grossmann, M., Wong, R., Teh, N. G., Tropea, J. E., East-Palmer, J., Weintrayb, B. D., and Szkudlinski, M. W. 1997. Expression of biologically active human thyrotropin (hTSH) in a baculovirus system: Effect of insect cell glycosylation on hTSH activity in vitro and in vivo. *Endocrinol.* 138:92–100.

Grubb, B. R., Pickles, R. J., Ye, H., Yankaskas, J. R., Vick, R. N., Engelhardt, J. F., Wilson, J. M., Johnson, L. G., and Boucher, R. C. 1994. Inefficient gene transfer by adenovirus vector to cystic fibrosis airway epithelia of mice and humans. *Nature* 371:802–806.

Grueninger-Leitch, F., D'Arcy, A., D'Arcy, B., and Chene, C. 1996. Deglycosylation of proteins for crystallization using recombinant fusion protein glycosidases. *Prot. Sci.* 5:2617–2622.

Hamer, D. H., Kaehler, M., and Leder, P. 1980. A mouse globin gene promoter is functional in SV40. *Cell* 21:697–708.

Hartman, S. C., and Mulligan, R. C. 1988. Two dominant-acting selectable markers for gene transfer studies in mammalian cells. *Proc. Nat. Acad. Sci. U.S.A.* 85:8047–8051.

He, T. C., Zhou, S., da Costa, L. T., Yu, J., Kinzler, K. W., and Vogelstein, B. 1998. A simplified system for generating recombinant adenoviruses. *Proc. Nat. Acad. Sci. U.S.A.* 95:2509–2514.

Higgins, D. R., and Cregg, J. M. 1998. *Pichia Protocols: Methods in Molecular Biology*, vol 103. Humana Press, Totowa NJ.

Hill-Perkins, M. S., and Possee, R. D. 1990. A baculovirus expression vector derived from the basic protein promoter of *Autographa californica* nuclear polyhedrosis virus. *J. Gen. Virol.* 71:971–976.

Hink, W. F., Thomsen, D. R., Davidson, D. J., Meyer, A. L., and Castellino, F. J. 1991. Expression of three recombinant proteins using baculovirus vectors in 23 insect cell lines. *Biotechnol. Prog.* 7:9–14.

Hollister, J. R., Shaper, J. H., and Jarvis, D. L. 1997. Stable expression of mammalian beta 1,4-galactosyltransferase extends the *N*-glycoslylation pathway in insect cells. *Glycobiol.* 8:473–490.

Hopp, T. P., Prickett, K. S., Price, V. L., Libby, R. T., March, C. J., Cerretti, D. P., Urdal, D. L., and Conlon, P. J. 1988. A short polypeptide marker sequence useful for recombinant protein identification and purification. *Bio/Technol.* 6:1204–1210.

Hsieh, P., and Robbins, P. 1984. Regulation of asparagine-linked oligosaccharide processing. *J. Biol. Chem.* 259:2375–2382.

Hsu, T. A., Eiden, J. J., and Betenbaugh, M. J. 1994. Engineering the assembly pathway of the baculovirus-insect cell expression system. *Ann. NY Acad. Sci.* 721:208–217.

Hsu, T. A., Watson, S., Eiden, J. J., and Betenbaugh, M. J. 1996. Rescue of immunoglobulins from insolubility is facilitated by PDI in the baculovirus expression system. *Prot. Expr. Purif.* 7:281–288.

Ishii, I., Izumi, T., Tsukamoto, H., Umeyama, H., Ui, M., and Shimizu, T. 1997. Alanine exchanges of polar amino acids in the transmembrane domains of a platelet-activating factor receptor generate both constitutively active and inactive mutants. *J. Biol. Chem.* 272:7846–7854.

Ivey-Hoyle, M. 1991. Recombinant gene expression in cultured *Drosophila melanogaster* cells. *Curr. Opin. Biotechnol.* 2:704–707.

Izumi, M., Miyazawa, H., Kamakura, T., Yamaguchi, I., Endo, T., and Hanaoka, F. 1991. Blasticidin *S*-resistance gene (bsr): A novel selectable marker for mammalian cells. *Exp. Cell. Res.* 197:229–233.

James, R. I., Elton, J. P., Todd, P., and Kompala, D. S. 2000. Engineering CHO cells to overexpress a secreted reporter protein upon induction from mouse mammary tumor virus promoter. *Biotechnol. Bioeng.* 67:134–140.

Jarvis, D. L., and Summers, M. D. 1989. Glycosylation and secretion of human tissue plasminogen activator in recombinant baculovirus-infected insect cells. *Mol. Cell. Biol.* 9:214–223.

Jarvis, D. L., Gleming, J. G. W., Kovacs, G. R., Summers, M. D., and Guarino, L. A. 1990. Use of early baculovirus promoters for continuous expression and efficient processing of foreign gene products in stably transformed lepidopteran cells. *Bio/Technol.* 8:950–955.

Jarvis, D. L., and Finn, E. E. 1996. Modifying the insect cell *N*-glycosylation pathway with immediate early baculovirus expression vectors. *Nature Biotechnol.* 14:1288–1292.

Jarvis, D. L., Kawar, Z. S., and Hollister, J. R. 1998. Engineering *N*-glycosylation pathways in the baculovirus–insect cell system. *Curr. Opin. Biotechnol.* 9:528–533.

Johansen, H., van der Straten, A., Sweet, R., Otto, E., Maroni, G., and Rosenberg, R. 1989. Regulated expression at high copy number allows production of a growth-inhibitory oncogene product in *Drosophila* Schneider cells. *Genes Dev.* 3:882–889.

Juarranz, M. G., Van Rampelbergh, J., Gourlet, P., De Neef, P., Cnudde, J., Robberecht, P., and Waelbroeck, M. 1999. Different vasoactive intestinal polypeptide receptor domains are involved in the selective recognition of two VPAC(2)-selective ligands. *Mol. Pharmacol.* 6:1280–1287.

Kaufman, R. J. 1990. Selection and coamplification of heterologous genes in mammalian cells. *Meth. Enzymol.* 185:537–566.

Kern, P., Hussey, R. E., Spoerl, R., Reinherz, E. L., and Chang, H. C. 1999. Expression, purification, and functional analysis of murine ectodomain fragments of CD8alphaalpha and CD8alphabeta dimers. *J. Biol. Chem.* 274:27237–27243.

King, L. A., and Possee, R. D. 1992. *The Baculovirus Expression System*. Chapman and Hall, London.

Kitts, P. A., and Possee, R. D. 1993. A method for producing recombinant baculovirus vectors at high frequency. *BioTechniq.* 14:810–817.

Koller, K. J., Whitehorn, E. A., Tate, E., Ries, T., Aguilar, B., Chernov-Rogan, T., Davies, A. M., Dobbs, A., Yen, M., and Barrett, R. W. 1997. A generic method for the production of cell lines expressing high levels of 7-transmembrane receptors. *Anal. Biochem.* 250:51–60.

Kornfeld, R., and Kornfeld, S. 1985. Assembly of asparagine-linked oligosaccharides. *Ann. Rev. Biochem.* 54:631–664.

Kost, T. A., and Condreay, J. P. 1999. Recombinant baculoviruses as expression vectors for insect and mammalian cells. *Curr. Opin. Biotechnol.* 10:428–433.

Kozak, M. 1987. An analysis of 5′-noncoding sequences from 699 vertebrate messenger RNA's. *Nucl. Acids Res.* 15:8125–8148.

Laprise, M. H., Grondin, F., and Dubois, C. M. 1998. Enhanced TGFbeta1 maturation in high five cells coinfected with recombinant baculovirus encoding the convertase furin/pace: Improved technology for the production of recombinant proproteins in insect cells. *Biotechnol. Bioeng.* 58:85–91.

Lehr, R. V., Elefante, L. C., Kikly, K. K., O'Brien, S. P., and Kirkpatrick, R. B. 2000. A modified metal-ion affinity chromatography procedure for the purification of histidine-tagged recombinant proteins expressed in *Drosophila* S2 cells. *Prot. Expr. Purif.* 19:362–368.

Leusch, M. S., Lee, S. C., and Olins, P. O. 1995. A novel host-vector system for direct selection of recombinant baculoviruses (bacmids) in *Escherichia coli*. *Gene* 160:191–194.

Levin, D. B., and Whittome, B. 2000. Codon usage in nucleopolyhedroviruses. *J. Gen. Virol.* 81:2313–2325.

Licari, P., and Bailey, J. E. 1992. Modeling the population dynamics of baculovirus-infected insect cells: Optimizing infection strategies for enhanced recombinant protein yields. *Biotechnol. Bioeng.* 39:432–441.

Lilius, G., Persson, M., Bulow, L., and Mosbach, K. 1991. Metal affinity precipitation of proteins carrying genetically attached polyhistidine affinity tails. *Eur. J. Biochem.* 198:499–504.

Loudon, P. T., and Roy, P. 1991. Assembly of five bluetongue virus proteins expressed by recombinant baculoviruses: Inclusion of the largest protein VP1 in the core and virus-like particles. *Virol.* 180:798–802.

Luckow, V. L., and Summers, M. D. 1988. Signals important for high-level expression of foreign genes in *Autographa californica* nuclear polyhedrosis virus expression vectors. *Virol.* 167:56–71.

Luckow, V. A., Lee, S. C., Barry, G. F., and Olins, P. O. 1993. Efficient generation of infectious recombinant baculoviruses by site-specific transposon-mediated insertion of foreign genes into a baculovirus genome propagated in *Escherichia coli*. *J. Virol.* 67:4566–4579.

Madisen, L., Lioubin, M. N., Marquardt, H., and Purchio, A. F. 1990. High-level expression of TGF-beta 2 and the TGF-beta 2(414) precursor in Chinese hamster ovary cells. *Growth Factors* 3:129–138.

Mansour, S. L., Grodzicker, T., and Tijan, R. 1986. Downstream sequences affect transcription from the adenovirus major late promoter. *Mol. Cell. Biol.* 6:2684–2694.

Matsushima, M., Ichinose, M., Yahagi, N., Kakei, N., Tsukada, S., Miki, K., Kurokawa, K., Tashiro, K., Shiokawa, K., Shinomiya, K., Umeyama, H., Inoues, H., Takahashi, T., and Takahashi, K. 1994. Structural characterization of porcine enteropeptidase. *J. Biol. Chem.* 269:19976–19982.

McCarroll, L., and King, L. A. 1997. Stable insect cell cultures for recombinant protein production. *Curr. Opin. Biotechnol.* 8:590–594.

Mellon, P., Parker, P., Gluzman, Y., and Maniatis, T. 1981. Identification of DNA sequences required for transcription of the human alpha 1-globin gene in a new SV40 host-vector system. *Cell* 27:279–288.

Merrington, C. L., King, L. A., and Possee, R. D. 1999. Baculovirus expression systems. In Higgins, S. J., and Hames, B. D., eds., *Protein Expression—A Practical Approach*. Oxford University Press, Oxford, U.K.

Miller, L. K. 1997. *The Baculoviruses*. Plenum Press, New York.

Mizushima, S., and Nagata, S. 1990. pEF-BOS, a powerful mammalian expression vector. *Nucl. Acids Res.* 18:5322.

Mroczkowski, B. S., Huvar, A., Lernhardt, W., Misono, K., Nielson, K., and Scott, B. 1994. Secretion of thermostable DNA polymerase using a novel baculovirus vector. *J. Biol. Chem.* 269:13522–13528.

Mulligan, R. C., and Berg, P. 1981. Selection for animal cells that express the *Escherichia coli* gene coding for xanthine-guanine phosphoribosyltransferase. *Proc. Nat. Acad. Sci. U.S.A.* 78:2072–2076.

Mulsant, P., Gatignol, A., Dalens, M., and Tiraby, G. 1988. Phleomycin resistance as a dominant selectable marker in CHO cells. *Somat. Cell Mol. Genet.* 14:243–252.

Murphy, C. I., Lennick, M., Lehar, S. M., Beltz, G. A., and Young, E. 1990. Temporal expression of HIV-1 envelope proteins in baculovirus-infected insect cells: Implications for glycoslyation and CD binding. *Genet. Anal. Tech. Appl.* 7:160–171.

Murphy, C. I., Piwnica-Worms, H., Grunwald, S., and Romanow, W. G. 1997. Expression of proteins in insect cells using baculovirus vectors. In Ausubel, F., Brent, R., Kingston, R., Moore, D., Seidman, J. G., Smith, J., and Struhl, K., eds., *Current Protocols in Molecular Biology*. Wiley, New York.

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.

Newman-Tancredi, A., Wootton, R., and Strange, P. G. 1992. High-level stable expression of recombinant 5-HT1A 5-hydroxytryptamine receptors in Chinese hamster ovary cells. *Biochem. J.* 285:933–938.

Niman, H. L., Houghton, R. A., Walker, L. E., Reisfeld, R. A., Wilson, I. A., Hogle, J. M., and Lerner, R. A. 1983. Generation of protein-reactive antibodies by short peptides is an event of high frequency: Implications for the structural basis of immune recognition. *Proc. Nat. Acad. Sci. U.S.A.* 80:4949–4953.

Ng, S. Y., Gunning, P., Eddy, R., Ponte, P., Leaviti, H., and Kedes, L. 1985. Evolution of the functional human beta-actin gene and its multi-pseudogene family: Conservation of noncoding regions of chromosomal dispersion of pseudogenes. *Mol. Cell. Biol.* 5:2720–2730.

O'Reilly, D. R., Miller, L. K., and Luckow, V. A. 1992. *Baculovirus Expression Vectors—A Laboratory Manual*. Freeman, New York.

Offermanns, S., and Simon, M. I. 1995. G alpha 15 and G alpha 16 couple a wide variety of receptors to phospholipase C. *J. Biol. Chem.* 270:15175–15180.

Ogonah, O. W., Freedman, R. B., Jenkins, N., Patel, K., and Rooney, B. C. 1996. Isolation and characterisation of an insect cell line able to perform complex N-linked glycosylation on recombinant proteins. *Nature Biotechnol.* 14:197–202.

Olkkonen, V. M., Dupree, P., Simons, K., Liljestrom, P., and Garoff, H. 1994. Expression of exogenous proteins in mammalian cells with the Semliki forest virus vector. *Methods Cell Biol.* 43: Pt A., 43–53.

Pajot-Augy, E., Bozon, V., Remy, J. J., Couture, L., and Salesse, R. 1999. Critical relationship between glycosylation of recombinant lutropin receptor

ectodomain and its secretion from baculovirus-infected cells. *Eur. J. Biochem.* 260:635–648.

Palomares, L. A., Gouzalez, M., and Ramirez, O. T. 2000. Evidence of Pluronic F-68 direct interaction with insect cells: Impact on shear protection, recombinant protein and baculovirus production, *Enz. Microb. Technol.* 26:324–331.

Parekh, R. B., Dwek, R. A., Rudd, P. M., Thomas, J. R., Rademacher, T. W., Warren, T., Wun, T. C., Hebert, B., Reitz, B., Palmier, M., Ramabhadran, T., and Tiemeier, D. C. 1989. *N*-glycosylation and in vitro enzymatic activity of human recombinant tissue plasminogen activator expressed in Chinese hamster ovary cells and a murine cell line. *Biochem.* 28:7644–7662.

Patel, G., Nasmyth, K., and Jones, N. 1992. A new method for the isolation of recombinant baculovirus. *Nucl. Acids Res.* 20:97–104.

Peakman, T. C., Charles, I. G., Sydenham, M. A., Gewart, D. R., Page, M. J., and Markoff, A. J. 1992. Enhanced expression of recombinant proteins in insect cells using a baculovirus vector containing a bacterial leader sequence. *Nucl. Acids Res.* 20:6111–6112.

Peakman, T. C., Harris, R. A., and Gewart, D. R. 1992. Highly efficient generation of recombinant baculoviruses by enzymatically mediated site-specific in vitro recombination. *Nucl. Acids Res.* 20:495–500.

Pfarr, D. S., Sathe, G., and Reff, M. E. 1985. A highly modular cloning vector for the analysis of eukaryotic genes and gene regulatory elements. *DNA* 4:461–467.

Pfarr, D. S., Rieser, L. A., Woychik, R. P., Rottman, F. M., Rosenberg, M., and Reff, M. E. 1986. Differential effects of polyadenylation regions on gene expression in mammalian cells. *DNA* 5:115–122.

Pikaart, M. J., and Felsenfeld, G. 1996. Expression and codon usage optimization of the erythroid-specific transcription factor cGATA-1 in baculoviral and bacterial systems. *Prot. Expr. Purif.* 8:469–475.

Potter, H., Weir, L., and Leder, P. 1984. Enhancer-dependent expression of human kappa immunoglobulin genes introduced into mouse pre-B lymphocytes by electroporation. *Proc. Nat. Acad. Sci. U.S.A.* 81:7161–7165.

Power, J. F., Reid, S., Radford, K. M., Greenfield, P. F., and Nielsen, L. K. 1994. Modeling and optimization of the baculovirus expression vector system in batch suspension culture. *Biotechnol. Bioeng.* 44:710–719.

Pyle, L. E., Barton, P., Fujiwara, Y., Mitchell, A., and Fiedge, N. 1995. Secretion of biologically active human proapolipoprotein A-1 in a baculovirus-insect cell system: Protection from degradation by protease inhibitors. *J. Lipid Res.* 86:2355–2361.

Ranjan, A., and Hasuain, S. E. 1994. Influence of codon usage and translation initiation codon context in the AcNPV-based expression system: Computer analysis using homologous genes. *Virus Genes* 9:149–153.

Robbins, P. D., Tahara, H., and Ghivizzani, S. C. 1998. Viral vectors for gene therapy. *Trends Biotechnol.* 16:35–40.

Rubin, D. M., Mehta, A. D., Zhu, J., Shoham, S., Chen, X., Wells, Q. R., and Palter, K. B. 1993. Genomic structure and sequence analysis of *Drosophila melanogaster* HSC70 genes. *Gene* 128:155–163.

Saarinen, M. A., Troutner, K. A., Gladden, S. G., Mitchell-Logean, C. M., and Murhammer, D. W. 1999. Recombinant protein synthesis in *Trichoplusia ni* BTI-Tn-5B1–4 insect cell aggregates. *Biotechnol. Bioeng.* 63:612–617.

Santini, F., Penn, R. B., Gagnon, A. W., Benovic, J. L., and Keen, J. H. 2000. Selective recruitment of arrestin-3 to clathrin coated pits upon stimulation of G protein-coupled receptors. *J. Cell. Sci.* 113:2463–2470.

Sarmiento, U. M., Riley, J. H., Knaack, P. A., Lipman, J. M., Becker, J. M., Gately, M. K., Chizzonite, R., and Anderson T. D. 1994. Biologic effects of recombi-

nant human interleukin-12 in squirrel monkeys (*Sciureus saimiri*). *Lab. Invest.* 71:862–873.

Schimke, R. T. 1988. Gene amplification in cultured cells. *J. Biol. Chem.* 263:5989–5992.

Schlesinger, S. 1993. Alphaviruses-vectors for the expression of heterologous genes. *Trends Biotechnol.* 11:18–22.

Schmitt, K., Daubener, W., Bitter-Suermann, D., and Hadding, U. 1988. A safe and efficient method for elimination of cell culture mycoplasmas using ciprofloxacin. *J. Immunol. Meth.* 109:17–25.

Schneider, I. 1972. Cell lines derived from late embryonic stages of *Drosophila melanogaster*. *J. Embryol. Exp. Morph.* 27:353–365.

Shitara, K., Nakamura, K., Tokutake-Tanaka, Y., Fukushima, M., and Hanai, N. 1994. A new vector for the high level expression for chimeric antibodies in myeloma cells. *J. Immunol. Meth.* 167:271–278.

Sondermann, P., Jacob, U., Kutscher, C., and Frey, J. 1999. Characterization and crystallization of soluble human Fc gamma receptor II (CD32) isoforms produced in insect cells. *Biochem.* 38:8469–8477.

Spellman, M. W., Basa, L. J., Leonard, C. K., Chakel, J. A., O'Connor, J. V., Wilson, S., and van Halbeek, H. 1989. Carbohydrate structures of human tissue plasminogen activator expressed in Chinese hamster ovary cells. *J. Biol. Chem.* 264:14100–14111.

Sridhar, P., Panda, A. K., Pal, R., Talwar, G. P., and Hasnain, S. E. 1993. Temporal nature of the promoter and not the relative strength determines the expression of an extensively processed protein in a baculovirus system. *FEBS Lett.* 315:282–286.

Sreekrishna, K., Brankamp, R. G., Kropp, K. E., Blankenship, D. T., Tsay, J. T., Smith, P. L., Wierschke, J. D., Subramaniam, A., and Birkenberger, L. A. 1997. Strategies for optimal synthesis and secretion of heterologous proteins in the methylotrophic yeast *Pichia pastoris*. *Gene* 190:55–62.

Stanley, P. 1989. Chinese hamster ovary cell mutants with multiple glycosylation defects for production of glycoproteins with minimal carbohydrate heterogeneity. *Mol. Cell. Biol.* 9:377–383.

Stern, L. J., and Wiley, D. C. 1992. The human class II MHC protein HLA-DR1 assembles as empty alpha beta heterodimers in the absence of antigenic peptide. *Cell* 68:465–477.

Stow, N. D. 1992. Herpes simplex virus type 1 origin-dependent DNA replication in insect cells using recombinant baculoviruses. *J. Gen. Virol.* 73:313–321.

Tessier, D. C., Thomas, D. Y., Khouri, H. E., Laliberte, F., and Vernet, T. 1991. Enhanced secretion from insect cells of a foreign protein fused to the honeybee melittin signal peptide. *Gene* 98:177–183.

Thirion, J. P., Banville, D., and Noel, H. 1976. Galactokinase mutants of Chinese hamster somatic cells resistant to 2-deoxygalactose. *Genetics* 83:137–147.

Toneguzzo, F., Keating, A., Glynn, S., and McDonald, K. 1988. Electric field-mediated gene transfer: Characterization of DNA transfer and patterns of integration in lymphoid cells. *Nucl. Acids Res.* 16:5515–5532.

Trill, J. J. 2001. Quantification of expression levels from transient and stably transfected cells using an IGEN M-Series analyzer. Manuscript in preparation.

Trill, J. J. 2000. GlaxoSmithkline. Unpublished data.

Trill, J. J. 1998. GlaxoSmithkline. Unpublished data.

Trill, J. J., Shatzman, A. R., and Ganguly, S. 1995. Production of monoclonal antibodies in COS and CHO cells. *Curr. Opin. Biotechnol.* 6:553–560.

Uphoff, C. C., Gignac, S. M., and Drexler, H. G. 1992. Mycoplasma contamination in human leukemia cell lines. II. Elimination with various antibiotics. *J. Immunol. Meth.* 149:55–62.

Urlaub, G., and Chasin, L. A. 1983. Isolation of Chinese hamster ovary cell mutants deficient in dihydrofolate reductase. *Proc. Nat. Acad. Sci. U.S.A.* 77:4216–4220.

Urlaub, G., Kas, E., Carothers, A. M., and Chasin, L. A. 1983. Deletion of the diploid dihydrofolate reductase locus in cultured mammalian cells. *Cell* 33:405–412.

van Oers, M. M., Vlak, J. M., Voorma, H. O., and Thomas, A. A. 1999. Role of the 3′ untranslated region of baculovirus p10 mRNA in high-level expression of foreign genes. *J. Gen. Virol.* 80:2253–2262.

Vaughn, J. L., Goodwin, R. H., Tompkins, G. J., and McCawley, P. 1977. The establishment of two cell lines from the insect *Spodoptera frugiperda* (Noctuidae). *In vitro* 13:213–217.

von Seggern, D. J., and Nemerow, G. R. 1999. Adenoviral–vectors for protein expression. In Fernandez, J. M., and Hoeffler, J. P., eds., *Gene Expression Systems: Using Nature for the Art of Expression*. Academic Press, San Diego, CA, pp. 290–325.

Wagner, R., Geyer, H., Geyer, R., and Klenk, H.-D. 1996a. *N*-acetyl-beta glucosaminidase accounts for differences in glycosylation of influenza virus hemagglutin expressed in insect cells from a baculovirus vector. *J. Virol.* 70:4103–4106.

Wagner, R., Liedtke, S., Kretzschmar, E., Geyer, H., Geyer, R., and Klenk, H. D. 1996b. Elongation of the *N*-glycans of fowl plaque virus hemagglutinin expressed in *Spodoptera frugiperda* (Sf9) cells by coexpression of human beta 1,2-*N*-acetyl glucosaminyltransferase, I. *Glycobiol.* 6:165–175.

Wahle, E., and Keller, W. 1992. The biochemistry of 3′-end cleavage and polyadenylation of messenger RNA precursors. *Ann. Rev. Biochem.* 61:419–440.

Wang, Y., O'Malley, Jr, B. W., Tsai, S. Y., and O'Malley, B. W. 1994. A regulatory system for use in gene transfer. *Proc. Nat. Acad. Sci. U.S.A.* 91:8180–8184.

Weiss, S., Famulok, M., Edenhofer, F., Wang, Y. H., Jones, I. M., Groschup, M., and Winnacker, E. L. 1995. Overexpression of active Syrian golden hamster prion protein PrPc as a glutathione *S*-transferase fusion in heterologous systems. *J. Virol.* 69:4776–4783.

Wenger, R. H., Moreau, H., and Nielsen, P. J. 1994. A comparison of different promoter, enhancer, and cell type combinations in transient transfections. *Anal. Biochem.* 221:416–418.

Weyer, U., Knight, S., and Possee, R. D. 1990. Analysis of very late gene expression by *Autographa californica* nuclear polyhedrosis virus and the further development of multiple expression vectors. *J. Gen. Virol.* 71:1525–1534.

Wickham, T. J., and Nemerow, G. R. 1993. Optimization of growth methods and recombinant protein production in BTI-Tn-5B1–4 insect cells using the baculovirus expression system. *Biotechnol. Progr.* 9:25–30.

Wirth, M., Bode, J., Zettlmeissl, G., and Hauser, H. 1988. Isolation of overproducing recombinant mammalian cell lines by a fast and simple selection procedure. *Gene* 73:419–426.

Yang, H., Leland, J. K., Yost, D., and Massey, R. J. 1994. Electrochemiluminescence: A new diagnostic and research tool. *Bio/Technol.* 12:193–194.