Mammalian Expression Cassette Engineering for High-Level Protein Production

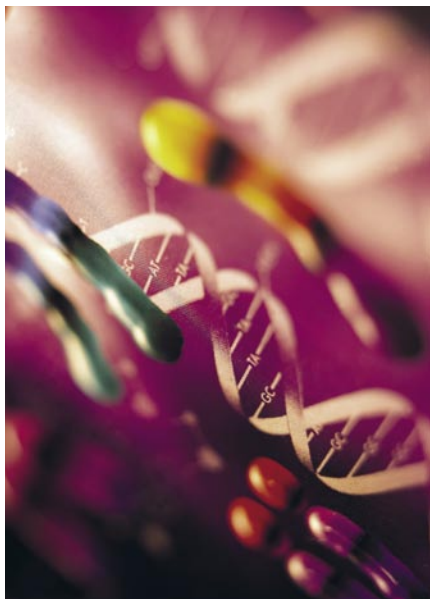
Components, Strategies, and Options

Dale L. Ludwig

Pharmaceutical and biotechnology company interest in developing biologic therapies has increased dramatically in recent years. The prospect of a blockbuster biologic has been realized with the clinical success of several products including Rituxan, Epogen, and Avastin. Some 16 monoclonal antibodies (MAbs) have been approved for pharmaceutical use, and more than 100 are currently in clinical trials.

Until relatively recently, the major impediment to widespread development of these biologics was their cost of goods: Typical expression levels in mammalian cells, the preferred host, were relatively poor by comparison with other production systems. However, the productive potential of engineered mammalian cell culture expression systems has greatly increased in recent years (**1**). Recombinant-based production cell lines now boast routine productivity levels in the 1–2 g/L range, and yields of greater than 5 g/L were recently reported by Lonza scientists (www.lonzabiologics.com). Productivity levels in engineered cell lines even 10 years ago was routinely in the range of just a few hundred milligrams per liter.

The industry has clearly found means to improve on animal cell culture and make its cost of goods for manufacture competitive with alternative production methods such as transgenic plants or animals as well as microbial fermentation. Advances seen in mammalian production systems have come from changes to all facets of mammalian cell culture production, including improvements in the components and strategies used to generate and select expression vectors.



WWW.PHOTOS.COM

Like vectors for expression in microbes and plants, a mammalian expression vector consists of a regulated cassette to drive efficient and high-level expression of a heterologous transcript, a gene-selectable marker (often a drug or metabolic selectable marker), and a vector backbone containing elements for engineering and propagating the vector in bacteria (**2**). Systems for generating stable, high-producing cell lines are typically nonreplicating and require integration into the genome of a host cell to be passed into daughter cells upon division. At least one functional copy of the vector must be integrated so both selectable marker and gene of interest are expressed.

Unlike with lower eukaryotes such as yeasts, integration of DNA into mammalian chromosomes is primarily a nontargeted event mediated by nonhomologous end-joining into random sites of DNA breakage within the genome (**3**). To facilitate this process and promote ligation of plasmids into genomic DNA, entry DNA is typically linearized before transfection. That generates free DNA ends located within a plasmid distal from the essential elements of the expression cassette and selection marker. In MAb production, cassettes for both heavy- and light-chain genes on a single vector facilitate cointegration (and hopefully coexpression) of both genes.

Selection of stably transfected cells is usually performed using media containing a selective pressure such as the absence of an essential amino acid or inclusion of a cytotoxic antibiotic. In many cases, both a selectable resistance marker and a gene of interest are present on an integrating

plasmid in separate cassettes under the control of two independent regulatory units. Thus, selection of the resistance marker does not necessarily ensure expression of the gene of interest. In many cases, a large number of selected colonies must be screened for positive expression of the gene of interest, and only a small fraction are high-producing lines.

Depending on the methods of transfection and selective pressure, copies of integrated plasmids can be increased in number, which may or may not effect an increase in output, depending on the sites of integration. Integration into tightly packed and heavily methylated heterochromatin (as opposed to euchromatin, the gene-rich open regions) can lead to poor or no expression. Ultimately, the expression strength of the cassette driving a gene of interest is crucial to output and depends on its context within a chromosomal integration site. Molecular biologists have designed a number of methods to improve expression and selection of productive clones by manipulating expression vector components.

The Mammalian Expression Cassette

Expression of a heterologous gene depends on the inherent strength and efficiency of regulatory sequences directing its transcription and processing into messenger RNA (mRNA) and the efficiency of that gene's translation into a protein product. At minimum, an expression cassette (Figure 1A) consists of a promoter that is 5′ (upstream) of the multiple cloning site (MCS), into which a gene to be expressed is inserted, and a 3′ untranslated region (3′ UTR) containing one or more polyadenylation (polyA) sequences (AATAAA), which are necessary for transcription termination and polyadenylation of the mRNA 3′ end. Enhancer elements from selected genes may also be placed upstream of the promoter to boost expression. An expression cassette may further include additional flanking sequences to facilitate or enhance position-independent expression from randomly integrated cassettes (Figure 1B): e.g., locus control regions (LCRs), matrix attachment sequences (MARs), and ubiquitous chromatin opening elements (UCOEs).

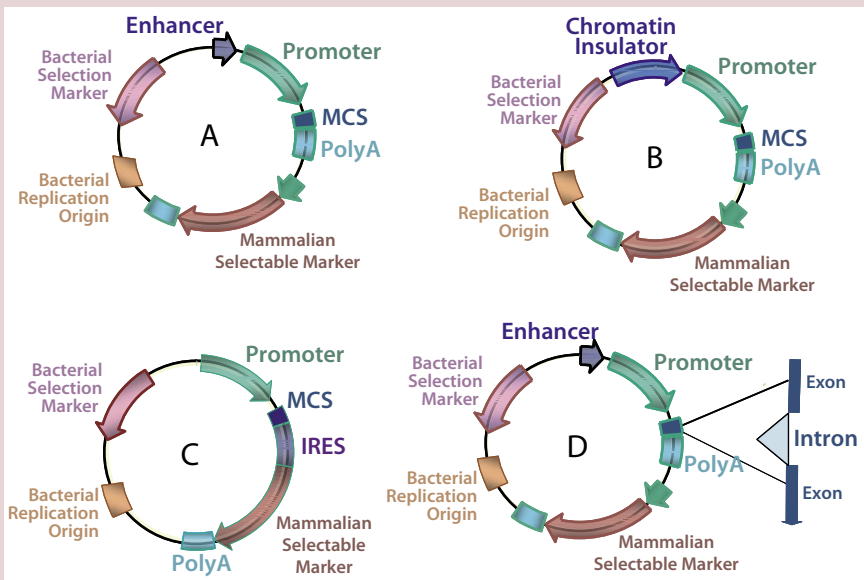
An additional element that has been included in some cassettes is the internal ribosome entry site (IRES) (Figure 1C). Typically from a viral source such as the encephalomyocarditis virus (ECMV), this element enables expression of a single bicistronic mRNA encoding both the gene of interest and a selectable marker. So a single expression cassette drives expression of both gene products. Expression cassettes may also be selectively engineered to include one or more intergenic sequences (Figure 1D) that promote more efficient transcription processing.

Promoters: A promoter constitutes the upstream or 5′ flanking sequences that drive transcription initiation. It is in essence a varying arrangement of short regulatory sequences (e.g., TATA box, downstream promoter element, Sp1) for transcription factor and RNA polymerase transcription complex binding that are coordinately arranged by evolutionary processes to initiate and promote constitutive or inducible, low-to-high levels of transcription.

By far, the most common promoters used for driving heterologous gene expression in mammalian cells are the human and mouse cytomegalovirus (CMV) major immediate early promoters. They effect high-level expression in a variety of cell hosts in both single- and dual- (e.g. immunoglobulin) gene systems (4). For preferred hosts such as CHO and NS0, CMV continues to be the promoter of choice (Table 1). However, there have been published reports that such promoters are cell cycle dependent, with their greatest transcriptional activity observed in the S phase. They can actually be silenced in certain stable cell lines, leading to considerable heterogeneity between transfectants.

Breuning et al. demonstrated that CMV promoters are upregulated in response to stress (5). This is a positive attribute that probably contributes to elevated heterologous gene expression (particularly later in culture growth) because of stress induced by

Figure 1: Mammalian expression vector configurations with plasmids depicted as single gene constructs



[A] In a basic expression vector, gene coding sequences are inserted into a multiple cloning site (MCS) under control of a 5′ promoter (with or without an enhancer element) and a 3′ polyadenylation sequence. The selectable marker is under control of a separate set of regulatory elements. Sequences for propagation of the plasmid in bacteria are present on the vector backbone.

[B] In vectors containing chromatin insulators (e.g., MARs) or chromatin opening elements (e.g., UCOEs), the element is typically placed upstream — and possibly also downstream — of the promoter.

[C] Bistronic vectors contain a single cassette for expression of a gene of interest inserted into the MCS and a selectable marker, separated by the IRES and under control of an upstream promoter and 3′ polyA.

[D] To facilitate expression, one or more introns are frequently inserted into the coding sequence for a gene of interest.

Table 1: Selected mammalian expression systems for antibody production

System	Promoter	PolyA	Selection	Host Cell	Enhancer	Insulator	IRES	Citation
GS	hCMV-MIE	SV40	Glutamine synthetase	CHO, NS0	CMV	None	None	Lonza Biologics
PER.C6	hCMV-MIE	BGH	*neo* (G418)	PER.C6	CMV	None	None	(**48**)
CHEF-1	CHEF-1 5′	CHEF-1 3′	DHFR/*neo*	CHO	None[1]	None[1]	None	(**6**)
EASE	hCMV	ND[2]	DHFR/*neo*	CHO	None[3]	EASE	ECMV	(**30**)
UCOE	hCMV	ND[2]	*neo*/hygro	CHO	None	UCOE	None	(**31**)
Chick Lysozyme MAR	SV40	SV40	DHFR/*neo*	CHO	SV40	MAR	None	(**29**)
Ig Heavy-Chain Enhancer	MT1 Ig κ	Mouse Ig κ, Ig γ	DHFR	CHO	Mouse Ig heavy chain	None	None	(**13**)

[1] Sequences contained within the CHEF-1 DNA fragments may contain enhancer or chromosomal insulator activity.

[2] Element used could not be determined from available literature. [3] Sequences contained within the EASE DNA fragments may contain enhancer activity.

accumulated metabolic byproducts and a reduction in available nutrients. It may also lessen the potential for silencing the CMV promoter during passage and cultivation of transfected cell lines. Indeed, many successful systems involve this promoter, including Lonza's Glutamine synthetase and Crucell's PER.C6 vectors (Table 1).

The SV40 immediate early promoter and the Rous Sarcoma Virus (RSV) long-terminal-repeat (LTR) promoter are also used frequently in expression cassette development. Often, the SV40 promoter drives expression of a selectable marker (e.g., G418 or DHFR) instead of a gene of interest. Recently, the CHO-derived elongation factor-1 promoter (CHEF-1) has been described as a novel system for driving high-level expression in mammalian cells (**6**). For a single gene product, expression levels were significantly higher than with comparable CMV-based vectors. The authors of that study noted that the level of expression achieved with CHEF-1 sequences are probably attributable to more than the promoter alone. They suggested that other regulatory elements such as insulators, enhancers, or matrix attachment regions (MARs) found within the cloned sequence may participate in driving expression. Even so, the CHEF-1 system appears to possess a number of positive attributes that make it directly comparable with, and potentially better than, CMV promoters.

Inducible rather than constitutive promoters have also been tested for expression in CHO cells (**7–9**): e.g., stress-inducible GADD153, zinc-inducible metallothionine, and dexamethasone-inducible mouse mammary tumor virus (MMTV). These all have the advantage of controlled expression, so they may have value in production of proteins whose expression are inhibitory to host cell growth.

Using myeloid cells (e.g., Sp2/0), we were interested in screening for strong endogenous promoters to use in vector development. As Figures 2A–2C illustrate, differential expression array analysis determined that the vitamin-D3 up-regulated protein 1 (VDUP1) gene was dramatically induced in response to cell density (**10**). We subsequently determined that a cloned mouse VDUP1 promoter was considerably stronger than the SV40 promoter in driving expression of secreted alkaline phosphatase (SEAP). That imparted inducible expression to a secreted heterologous gene product in stable mammalian cells (Figure 2C). We are continuing to explore the potential of this promoter element for driving high-level protein production in myelomas and other host cell lines. This promoter has multiple VDRE and heat-shock factor element sites for potential upregulation in response to exogenous agents or stress, and it has recently been confirmed that endogenous VDUP1 is inducible in response to heat shock (**11**).

One caveat with inducible rather than constitutive expression of such proteins as MAbs is that the short-term lag in expression (before induction) may compromise production yields relative to those obtained using a constitutive promoter (e.g., CMV). Depending on the method of induction, this problem might be overcome by inducing expression earlier in culture through media supplementation — if the inducer itself is not inhibitory to cell growth.

Although relatively few established promoters have been used for driving high-level heterologous gene expression in large-scale mammalian cell culture, many viable alternatives do exist. It was recently proposed that a slight bias toward enhanced expression of the light chain for immunoglobulins may improve secretion of mature gene products (**12**). It is therefore plausible to use two different promoters for separately driving the heavy- and light-chain genes, with the light chain under control of a modestly stronger regulatory element.

Enhancers: An evolutionarily conserved component of some mammalian genes is the enhancer element. Similar to upstream activation sequences (UAS) in microorganisms such as yeasts, enhancers can further elevate gene expression often independent of their orientation. Viral promoter regions (e.g., CMV and SV40) contain proximal enhancer elements, which are frequently included when such promoters are used in cassette engineering.

In many mammalian genes, enhancers ensure precise and coordinated control of gene expression. Interestingly, they are often located at a considerable distance from the regulated gene, so the mechanism by which they interact with promoter regions is not altogether clear. Some enhancers have been identified through their ability to confer tissue-specific gene expression.

One was identified within the human immunoglobulin heavy-chain locus and incorporated into vectors for improving antibody expression in mammalian myeloid cells such as Sp2/0 (**13**).

Polyadenylation Sequences: Just as the promoter region is essential for mediating initiation of transcription, the 3′ UTR is critical to ensuring its proper termination. The efficiency with which that occurs can dramatically affect stability of an RNA product as well as facilitate transport of processed mRNA from nucleus to cytoplasm — and it may also play a role in promoting translation (**14**). At minimum, the 3′-UTR requires one or more polyA sequences with consensus AATAAA and a GT-rich downstream sequence element (DSE) for transcription termination and polyadenylation of RNA transcripts (**15, 16**).

A prominent feature in certain polyA regions is the presence, not of transcript enhancers or stabilizers, but of instability elements (e.g., AREs). Such elements can dramatically reduce the half-life of a nascent transcript and are thus deleterious to the performance of an optimal expression cassette (**17, 18**). Similar to the frequent use of CMV in mammalian expression cassette development, the SV40 virus polyA fragment is a common 3′ UTR found in many expression systems.

The 3′ UTR from the rabbit ß-globin gene, without its AREs, has also been used for driving heterologous expression in mammalian cells (**18**). It demonstrates comparable activity to the SV40 polyA as determined by RNA levels and protein production. Furthermore, polyA from the bovine growth hormone (BGH) gene has been shown to mediate efficient transcript termination and polyadenylation to heterologous genes (**19, 20**). Thus it provides improved expression over the SV40 polyA in comparison studies.

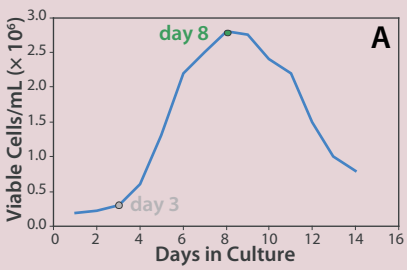
Optimization of the 3′ end in expression cassettes doesn't seem to have been well studied. Because of the relative scarcity of available alternatives, SV40 and BGH polyAs predominate.

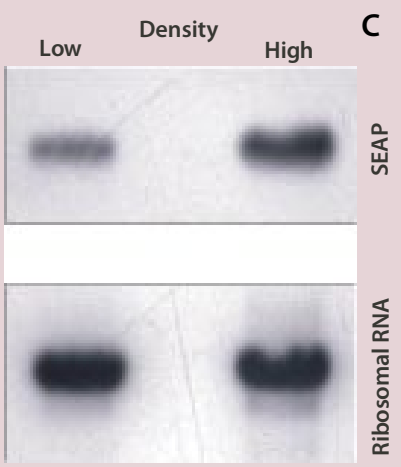
Introns: Mammalian genes typically possess multiple intergenic sequences of varying lengths spaced between the exonic or coding sequences of each given gene. These have considerable evolutionary importance in maintaining the integrity of coding regions, protecting them from mutation, rearrangement, or deletion (**21**). Introns also appear to influence transcription efficencies (**22**).

Cloning and manipulation of cDNAs for heterologous gene expression may be considerably more facile, but studies have increasingly supported inclusion of native or even heterologous intron sequences within selected areas of coding regions to improve on transcription efficiency (**21, 23**). For antibody expression, the genomic version of immunoglobulin constant regions (containing multiple native introns) is frequently used in place of a constant region cDNA sequence (**13**). Many constructs using the CMV IE promoter include the first intron of the gene, which is likely to improve the promoter's performance in heterologous expression (**24, 25**). Some postulate that processes involved in pre-mRNA splicing of introns facilitate RNA transport from the nucleus or enhance RNA stability and/or half-life, thus providing greater abundance of transcript templates for translation (**26**).

Intron inclusion within coding sequences has been more heavily investigated in development of gene therapy vectors, but the effects are also relevant to production vector development. Our own investigations of MAb expression in rodent cells suggests a consistently positive benefit for improving expression with intron-containing cassettes. That effect can be observed in transient expression systems, and it can improve

Figures 2A–C: Identification and characterization of VDUP1 expression in high-density Sp2/0 cell culture





[A] Growth curve of Sp2/0 myeloma cells in culture: An array of 5000 mouse genes was screened for differential expression in RNA samples taken early (day 3) and late (day 8) when cell densities were minimal and maximal, respectively.

[B] Schematic representation of the relative expression level for the 10 most abundantly expressed genes in late Sp2/0 culture and their differential relative to the early time point.

[C] Northern blot analysis of mVDUP1–promoter–driven, secreted alkaline phosphatase (SEAP) gene expression from a stably transfected mouse cell line: The blot was probed first for SEAP, then rehybridized with a ribosomal RNA-specific probe. Expression of the heterologous SEAP reporter gene was determined to be induced about threefold by the 1.5kb VDUP1 promoter during high-density cell culture.

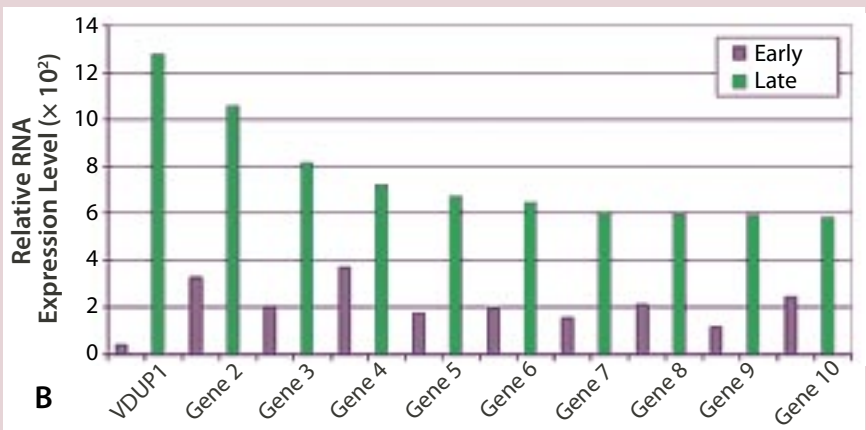
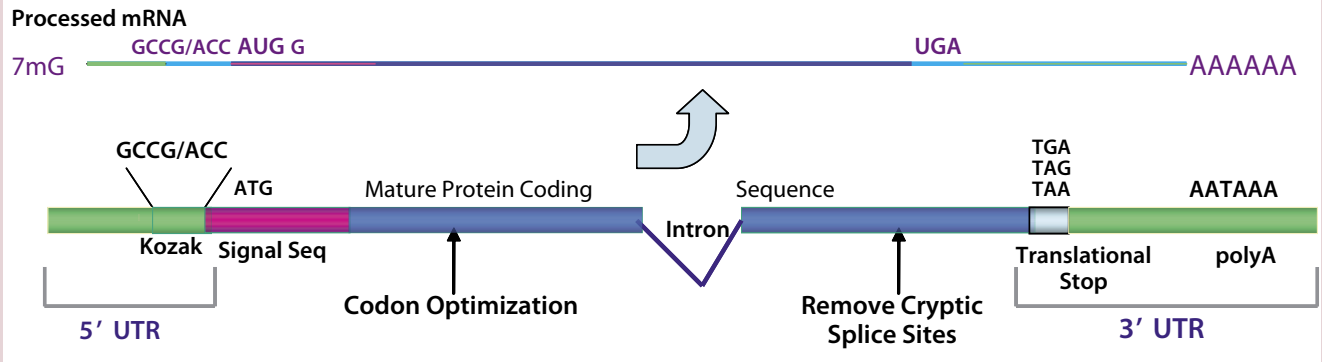


Figure 3: Inserted gene coding region and immediate upstream and downstream flanking sequences transcribed and processed into mRNA



production in stable cell lines as well. It is important, however, to evaluate the probability of splice donor and acceptor sequences for efficient splicing and to optimize sequences if the coding region permits. A number of websites (e.g., www.fruitfly.org/seq_tools/splice-instrucs.html) enable free sequence splice analysis, which can be particularly helpful in identifying unwanted cryptic splice sites within a gene coding region.

Chromatin Modifiers: A major impediment in selecting high-production cell lines for recombinant proteins is the considerable variability inherent to stable transfection of mammalian cells through random chromosomal integration. For any exogenously transfected gene, output from independent clones typically ranges from poor to very high (the former occurring much more frequently than the latter), and no expression at all may be detected in as many as 50% or more of clones screened. With the gene of interest and the selectable marker under control of separate regulatory elements, growth in selective media does not necessarily confer expression of a desired gene product, even if the related vector is inserted completely intact. These so called *position-effects* are attributed to the context and structure of a chromosomal integration site (**27**).

A number of mechanisms contribute to position effects. For example, heterologous gene sequences can be effectively silenced or compromised by endogenous upstream transcription units that affect transcription read-though of the promoter contained within an expression cassette. In addition, methylation of exogenous DNA sequences, a cellular protective mechanism, is a frequent event that can dramatically suppress promoter activity in transgenic systems.

Molecular biologists have found several approaches to overcome this obstacle and make high-producing clones a more frequent result. Where chromosomal elements influence the expression of integrated foreign sequences, scientists have successfully isolated chromosomal insulator or buffer elements that block or suppress chromosomal position effects, imparting position-effect independent expression. Such elements — e.g., matrix or scaffold attachment regions (MARs or SARs) or locus control regions (LCRs) — flank an expression cassette, typically on both 5′ and 3′ ends. As a result, expression of an integrated transgene is minimally affected by chromosomal position. It potentially brings about a chromatin structure that is permissive for transcription factor and polymerase binding. That has the effect of increasing the frequency of productive expression in individual clones. It thus enables greater control of expression by regulatory elements of a cassette (e.g., promoters) as well as a more proportional affect on expression of integrant copy number (**28, 29**).

Novel *cis*-acting chromosomal elements that exert a positive effect on recombinant protein expression have recently been described and tested for facilitating gene expression in stable cell lines. These include expression augmenting sequence elements (EASEs) and ubiquitous chromatin opening elements (UCOEs) (**30, 31**). They require only a single upstream element to impart their positive effect. The overall benefit is an ability to enhance the frequency of productive expression in stable clones, thereby improving the probability of isolating a high-productivity line. Screening and analysis are still required, but there are more positive clones to start with. In this case, as with CHEF-1 vectors, data suggest that such elements in fact increase cassette performance to provide not only larger numbers of positive clones but also greater numbers of more productive clones.

IRES Elements: A different approach with a similar endpoint to the use of chromatin insulators (that of increasing the frequency of stably expressing transfectants) is the use of internal ribosome entry sites (IRESs). Typically of viral origin, these sequences enable creation of functional bicistronic mRNAs that encode both a gene of interest and a stable transfection selection marker (**32**). With clones selected based on their selection marker expression, expression of the gene of interest is forced when both genes are transcribed by a single RNA.

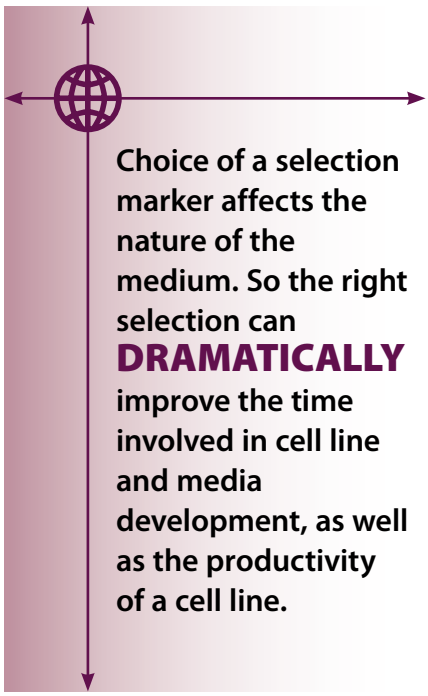
A series of tricistronic vectors, pTRIDENT, uses two IRES elements to link three genes (**33, 34**). This type of construct is therefore functionally capable of expressing both heavy- and light-chain antibody genes with a selectable marker from a single cassette. Ribosomes load onto and initiate translation of sequences significantly downstream from the capped 5′ end of the mRNA, which is where they traditionally initiate binding. Although this process

appears to be less efficient than cap-dependent translation, it is preferable to place the selection marker (rather than the gene of interest) downstream of the IRES. Indeed, the resulting modest impairment of selection marker expression can actually improve selection of better-expressing transfectants because higher transcription levels may be necessary to confer growth in the presence of a selection medium. Although greater stringency of selection may increase the frequency of higher producing clones, the overall number of selected clones is undoubtedly reduced.

Selection Markers

A considerably large number of selective markers are available for enabling stable selection of plasmid integrants in mammalian cells. Markers for selection of mammalian cells tend to be either antibiotic resistance genes or metabolic enzymes. By far, the most common antibiotic resistance marker used is the bacterial neomycin–kanamycin phosphotransferase type II enzyme encoded by the *neo* gene. It confers selection by the drug G418 in eukaryotic cells. *Neo* is found frequently in commercially available expression vectors and high-level production systems (e.g., PER.C6, UCOE, and CHEF-1).

Although a number of alternative antibiotic selection markers are available (e.g., puromycin, zeocin, and blastocidin), little information is available on their use in mammalian expression systems for large-scale production. The above-mentioned tricistronic vector system, pTRIDENT, involves zeocin selection. More commonly used in high-level expression systems are metabolic selection markers such as the hamster gene encoding glutamine synthetase (GS) (**35, 36**) and the mouse dihydrofolate reductase (DHFR) gene (**37–39**). In both cases, cell lines defective in each enzyme are available: NS0 for GS (**35**) and CHO for DHFR (**40**) allow direct selection in media lacking either glutamine or hypoxanthine/thymidine (HT), respectively. Of considerable importance is the existence of metabolic inhibitors for those enzymes — methionine sulfoximine (MSX) and methotrexate (MTX), respectively — which enable greater selection stringency to transfectants. This creates the potential for selection of clones that have undergone gene amplification.



Because posttransfection cell line amplification with DHFR selection in MTX can delay product development by several months, an alternative approach was taken by Lo and Gillies (**41**). They used protoplast fusion to effect high–copy-number integration into a genome for initial transfectants. They subsequently increased production in resulting clones by adaptation (rather than gene amplification) to higher methotrexate levels, essentially eliminating the amplification step.

With GS, similar selection of high-level production lines can be frequently achieved without amplification. The enzyme catalyzes production of glutamine from glutamate and ammonia, a metabolic byproduct that negatively affects growth if it accumulates in culture. So such selection of transfectants may actually promote positive cell growth and effect higher productivities. Indeed, gram-per-liter levels of recombinant antibody have been achieved with both GS and DHFR selection.

Other metabolic markers including histidinol dehydrogenase (HisD) (**42**) and asparagine synthetase (**43**) have also been identified for use in mammalian expression systems. HisD catalyzes the two-step, NAD+ dependent oxidation of L-histidinol to L-histidine, so stable transfection is performed in medium lacking histidine and containing histidinol. Increasing amounts of histidinol in the medium could potentially be used to impart increased stringency of selection to transfectants. This amino acid is not quite as labile as glutamine and therefore is present in plant and yeast hydrolysates, which are common supplements for adding amino acids and dipeptides to media and thus could compromise selective pressure for this marker. Asparagine synthetase may possess similar host and media problems. Both it and HisD may represent interesting alternatives for future initiatives in vector and cell engineering toward high-level expression.

Choice of a selection marker obviously affects the nature of a basal cell culture medium, supplements, and feeds. Therefore, an appropriate selection can dramatically affect the time involved in cell line and media development, as well as the productivity of a cell line. The strength of selection has also been investigated considerably for its influence on the stringency of selection. In theory, the weaker a selection marker is expressed, the better the stringency of site selection will enable sufficient production of the marker. That should in turn lead to higher production of the gene of interest present on a vector. Strategies include alternate splicing, weakened transcription or translation initiation, and even segmenting selection markers for transcomplementation.

Translation and Secretion Optimization

Vector components discussed thus far have dealt with options that enable proficient transcription of a gene of

interest. Any discussion of mammalian expression cassettes also should include some mention of translation optimization (Figure 3). When the start of the 5´ untranslated leader is conferred by a promoter, that stretch of upstream sequence preceding the gene-coding region can influence both transcription and translation. In particular, sequences prone to forming secondary structures could influence the speed of translation, delaying movement of ribosomes to the coding region (**44**). In this regard, including upstream stretches of MCS that contain multiple palindromic restriction endonuclease sequences could negatively affect translation (**45**).

In addition, sequences immediately upstream of an initiation start codon play a significant role in the initiation of translation. The 6–9 bases proximal to an initation AUG, referred to as the *Kozak sequence,* represent the eukaryotic equivalent of the Shine-Delgarno sequence, which in prokaryotic genes facilitates start-codon recognition by ribosomes to begin translating sequences into polypeptides. An optimal Kozak sequence (GCCG/ACCAUGG) has been determined with two bases of critical significance. A purine at –3 and a G at +4 within the coding sequence can clearly influence translation efficiency (**44**).

It is advisable to include such sequences in a final expression construct, if not the entire consensus sequence. Also important is the choice of a signal sequence to promote efficient intracellular transport of a recombinant gene product into the secretory apparatus. From personal experience, I can say that trial and error is the best approach. A number of sequences from secreted proteins have been tested for secretion of heterologous proteins, including native immunoglobulin signal sequences. They can serve as universal signals to efficiently direct secretion of a desired gene product.

Finally, codon optimization of a gene of interest should be considered (**46**), especially when a host cell line is sufficiently distinct from the gene's host of origin. Whenever possible, during codon optimization of the gene sequence, it is further advisable to remove cryptic splice sites and polyAs as well because they can remotely influence transcript processing. Such manipulation of a gene-coding sequence may or may not ultimately influence the productive outcome of transfected production cells, but it is more likely to improve rather than hinder expression of the gene.

Acronym Glossary

Acronyms and other alphanumerics that are not spelled out in the text can be found here.

AREs: all-rich sequence elements

AUG: mRNA version of the "ATG" DNA sequence

DHFR: dihydrofolate reductase

GADD153: growth-arrest and DNA damage inducible gene 153

NAD+: nicotinamide adenine dinucleotide

SV40: simian virus 40

TATA: DNA sequence for a specific transcription element

VDRE: Vitamin D3 response element

Giant Steps Forward

Many options are afforded the molecular biologist in designing a suitable mammalian expression cassette. Although CMV-promoter–driven cassettes with GS, DHFR, or *neo* selection seem to predominate, the framework of that basic unit is under constant investigation. New and promising strategies have recently been described including the use of artificial chromosomes to facilitate stable high expression of recombinant protein (**47**). Like many others, we have evaluated innovative licensable vector technologies and performed our own internal investigations, finding good results with some and poor performance with others.

Ultimately, companies focused on making a specific class of biologics (e.g., MAbs) may choose only a single platform and thus maintain continuity with downstream process development among products. That can place a restraint on a host and selection system, but it does maintain flexibility to improve each contruct itself. Now that 1–2 g/L is common, the near-term expectation is of yields approaching 10 g/L. Successful approaches to achieve this milestone will no doubt involve further manipulation of expression vectors as well as optimization of host cells, media, and reactor processes.

It would be of considerable interest to mix and match the most promising elements of each within one system. However, it is important to note that many of the regulatory elements described here require licensing for commercial use, which may restrict investigation of novel combinations. Vector engineering remains a critical component of every recombinant protein production process, certainly as important as host selection and engineering as well as media formulations development.

Acknowledgments

I would like to thank Dr. Venkat Mangalampalli and Ms. Pinki Singh for their assistance in the preparation of this manuscript and the accompanying figures.

References

1 Wurm F. Production of Recombinant Protein Therapeutics in Cultivated Mammalian Cells. *Nat. Biotech.* 22(11) 2004: 1393–1398.

2 Kaufman RJ. Overview of Vector Design for Mammalian Gene Expression. *Mol. Biotechnol.* 16, 2000: 151–160.

3 Manivasakam P, et al. Restriction Enzymes Increase Efficiencies of Illegitimate DNA Integration but Decrease Homologous Integration in Mammalian Cells. *Nucl. Acids Res.* 29(23) 2001: 4826–4833.

4 Foecking MK, Hofstetter H. Powerful and Versatile Enhancer–Promoter Unit for Mammalian Expression Vectors. *Gene* 45(1) 1986: 101–105.

5 Breuning W, et al. Activation of Stress-Activated MAP Protein Kinases Up-Regulates Expression of Transgenes Driven By the Cytomegalovirus Immediate/Early Promoter. *Nucl. Acids Res.* 26(2) 1998: 486–489.

6 Running Deer J, Allison DS. High-Level Expression of Proteins in Mammalian Cells Using Transcription Regulatory Sequences from the Chinese Hamster EF-1a Gene. *Biotechnol. Prog.* 29, 2004: 880–889.

7 de Boer L, Gray PP, Sunstrom N-A. Enhanced Productivity of G1 Phase Chinese Hamster Ovary Cells Using the GADD153 Promoter. *Biotechnol. Lett.* 26, 2004: 61–65.

8 Mangalampalli VRM, et al. Increased Production of a Secreted Glycoprotein in Engineered CHO Cells through Amplification of a Transfcription Factor. *Cytotechnol.* 38, 2002: 23–35.

9 James RI, et al. Engineering CHO Cells to Overexpress a Secreted Reporter Protein Upon Induction of Mouse Mammary Tumor Virus Promoter. *Biotech. Bioeng.* 67, 2000: 134–140.

10 Ludwig DL, et al. Cloning, Genetic Characterization, and Chromosomal Mapping of the Mouse VDUP1 Gene. *Gene* 269, 2001: 103–112.

11 Kim KY, et al. Heat Shock Factor Regulates VDUP1 Gene Expression. *Biochem. Biophys. Res. Commun.* 315(2) 2004: 369–75.

12 Schlatter S, et al. On the Optimal Ratio of Heavy to Light Chain Genes for Efficient Recombinant Antibody Production By CHO Cells. *Biotechnol. Prog.* 21(1) 2005: 122–133.

13 Gillies SD, Lo KM, Wesolowski J. High-Level Expression of Chimeric Antibodies Using Adapted cDNA Variable Region Cassettes. *J. Immunol. Meth.* 125, 1989: 191–202.

14 Colgan DF, Manley JL. Mechanism and Regulation of mRNA Polyadenylation. *Genes Dev.* 11 (21) 1997: 2755–2766.

15 Yonaha M, Proudfoot NJ. Transcriptional Termination and Coupled Polyadenylation In Vitro. *EMBO J.* 19(14) 2000: 3770–3777.

16 Proudfoot N. New Prespectives on Connecting Messenger RNA 3´ End Formation to Transcription. *Cur. Opin. Cell Biol.* 16, 2004: 272–278.

17 Chen C-YA, Shyu A-B. AU-Rich Elements: Characterization and Importance in mRNA Degradation. *Trends Biochem. Sci.* 20, 1995: 465–470.

18 Rotondaro L, et al. High Level Expression of a cDNA from Human Granulocyte Colony-Stimulating Factor in Chinese Hamster Ovary Cells: Effect of 3´-Noncoding Sequences. *Mol. Biotechnol.* 7, 1997: 231–240.

19 Pfarr DS, et al. Differential Effects of Polyadenylation Regions on Gene Expression in Mammalian Cells. *DNA* 5(2) 1986: 115–122.

20 Goodwin EC, Rottman FM. The 3´-Flanking Sequence of the Bovine Growth Hormone Gene Contains Novel Elements Required for Efficient and Accurate Polyadenylation. *J. Biol. Chem.* 267(23) 1992: 16330–16334.

21 Lacy-Hulbert A, et al. Interruption of Coding Sequences By Heterologous Intron Can Enhance the Functional Expression of Recombinant Genes. *Gene Ther.* 8, 2001: 649–653.

22 Buchman AR, Berg P. Comparison of Intron-Dependent and Intron-Independent Gene Expression. *Mol. Cell. Biol.* 8(10) 1988: 4395–4405.

23 Melecher R, Grosch HW, Hasilik A. Plasmid Vectors with a 5´-Hybrid Intron Facilitate High-Level Glycoprotein Expression in CHO Cells. *Biochim. Biophys. Acta.* 1575(1–3) 2002: 49–53.

24 Simari RD, et al. Requirements for Enhanced Transgene Expression By Untranslated Sequences from the Human Cytomegalovirus Immediate-Early Gene. *Mol. Med.* 4(11) 1998: 700–706.

25 Xu ZL, et al. Strength Evaluation of Transcriptional Regulatory Elements for Transgene Expression By Adenovirus Vector. *J. Contr. Rel.* 81(1–2) 2002: 155–163.

26 Neuberger MS, Williams GT. The Intron Requirement for Immunoglobulin Gene Expression Is Dependent Upon the Promoter. *Nucl. Acids Res.* 16(14) 1988: 6713–6724.

27 Bell AC, West Ag, Felsenfeld G. Insulators and Boundaries: Versatile Regulatory Elements in the Eukaryotic Genome. *Science* 291, 2001: 447–450.

28 Kim J-M, et al. Improved Recombinant Gene Expression in CHO Cells Using Matrix Attachment Regions. *J. Biotechnol.* 107, 2004: 95–105.

29 Girod P-A, Zahn-Zabal M, Mermod N. Use of Chicken Lysozyme 5´ Matrix Attachment Region to Generate High Producer CHO Cell Lines. *Biotech. Bioeng.* 91(1) 2005: 1–10.

30 Aldrich TL, Vaije A, Morris AE. EASE Vectors for Rapid Stable Expression of Recombinant Antibodies. *Biotechnol. Prog.* 19, 2003: 1433–1438.

31 Benton T, et al. The Use of UCOE Vectors in Combination with a Preadapted Serum Free, Suspension Cell Line Allows for Rapid Production of Large Quantities of Protein. *Cytotechnol.* 38, 2002: 43–46.

32 Martinez-Salas E. Internal Ribosome Entry Site Biology and Its Use in Expression Vectors. *Curr. Opin. Biotech.* 10, 1999: 458–464.

33 Fussenegger M, Mazur X, Bailey JE. pTrident, a Novel Vector Family for Tricistronic Gene Expression in Mammalian Cells. *Biotechnol. Bioeng.* 57(1) 2000: 1–10.

34 Moser S, et al. An Update of pTRIDENT Multicistronic Expression Vectors: pTRIDENTs Containing Novel Streptogramin-Responsive Promoters. *Biotechnol. Prog.* 16(5) 2000: 724–735.

35 Bebbington CR, et al. High-Level Expression of a Recombinant Antibody from Myeloma Cells Using a Glutamine Synthetase Gene As an Amplifiable Selectable Marker. *Biotechnol.* 10(2) 1992: 169–175.

36 Pu H, et al. Rapid Establishment of High-Producing Cell Lines Using Dicistronic Vectors with Glutamine Synthetase As the Selection Marker. *Mol. Biotechnol.* 10(1) 1998: 17–25.

37 Lo KM, Lynch CA, Gillies SD. The Use of a Wild-Type Dihydrofolate Reductase–Encoding cDNA As a Dominant Selectable Marker and Induction of Expression By Methotrexate. *Gene* 121, 1992: 365–369.

38 Kim SJ, et al. Characterization of Chimeric Antibody Producing CHO Cells in the Course of Dihydrofolate Reductase-Mediated Gene Amplification and Their Stability in the Absence of Selective Pressure. *Biotechnol. Bioeng.* 58(1) 1998: 73–84.

39 Bianchi AA, McGrew JT. High-Level Expression of Full-Length Antibodies Using Trans-Complementing Expression Vectors. *Biotechnol. Bioeng.* 84(4) 2003: 439–444.

40 Urlaub G, Chasin LA. Isolation of Chinese Hamster Cell Mutants Deficient in Dihydrofolate Reductase Activity. *Proc. Natl. Acad. Sci. USA* 77(7) 1980: 4216–4220.

41 Lo KM, Gillies SD. High-Level Expression of Human Proteins in Murine Hybridoma Cells: Induction By Methotrexate in the Absence of Gene Amplification. *Biochim. Biophys. Acta.* 1088(2) 1991: 217–224.

42 Hartman SC, Mulligan RC. Two Dominant-Acting Selectable Markers for Gene Transfer Studies in Mammalian Cells. *Proc. Natl. Acad. Sci. USA* 85(21) 1988: 8047–8051.

43 Cartier M, Chang MWM, Stanners CP. Use of the *Escherichia coli* Gene for Aspargine Synthetase as a Selective Marker in a Shuttle Vector Capable of Dominant Transfection and Amplification in Animal Cells. *Mol. Cell. Biol.* 7(5) 1987: 1623–1628.

44 Kozak M. Regulation of Translation via mRNA Structure in Prokaryotes and Eukaryotes. *Gene* 361, 2005: 13–37.

45 Beliveau A, et al. Multiple Cloning Sites from Mammalian Expression Vectors Interfere with Gene Promoter Studies In Vitro. *Eur. J. Biochem.* 261, 1999: 585–590.

46 Kim CH, Oh Y, Lee TH. Codon Optimization for High-Level Expression of Human Erythropoietin (EPO) in Mammalian Cells. *Gene* 199(1–2) 1997: 293–301.

47 Lindenbaum M, et al. A Mammalian Artificial Chromosome Engineering System (ACE System) Applicable to Biopharmaceutical Protein Production, Transgenesis, and Gene-Based Cell Therapy. *Nucl. Acids Res.* 32(21) 2004: e172.

48 Jones D, et al. High-Level Expression of Recombinant IgG in the Human Cell Line PER.C6. *Biotechnol Prog.* 19(1) 2003: 163–168.

Dale L. Ludwig, *PhD, is senior director for the Department of Cell Engineering and Expression at ImClone Systems Inc., New York, NY 10014; 1-646-638-5187, fax 1-212-645-2054; dale@imclone.com.*