

Introduction

MAMMALIAN CELL LINES are used to produce therapeutic and diagnostic proteins as well as viruses for gene therapy and vaccines. In all applications, the cell line used for the production has a large effect on productivity and product quality. Viruses have diverse tissue tropism and host species range. The infectivity of a virus is often affected by the tissue and species that the host cell is derived from. Virus infection elicits various responses from host cells that suppress its replication (Panel 6.1).

Panel 6.1. Factors Affecting Host Cell Selection for Industrial Applications

Viruses (Vaccines, Gene Therapy)

- Tropism of virus for tissues and species
- Virus productivity
 - Host cell response
- Host cell response to virus infection
- Genetic stability (tumorigenicity)
- Processability
- Regulatory acceptability

Heterologous Gene Products

- Ease of genetic manipulation
- Genetic stability
- Secretory plasticity
- Processability
- Regulatory acceptability
(endogenous retrovirus, cell history)

The selection of a host cell with temperate responses to virus infection is thus important for high virus productivity. Most virus products do not undergo the extensive purification seen in the production of protein therapeutics. Their recovery often involves only the capture and concentration of the virus, the removal of major contaminants, and buffer exchange. Extensive purification causes virus titer loss. Selecting a host cell with a minimum of contaminants can ease the demand on downstream processing. Because of the possibility of incorporation of genetic elements from the host cell into the viral genome, the use of tumorigenic host cells for virus production has traditionally been shunned.

Selecting the right host cell line or cell strain for virus production is thus of paramount importance.

For the manufacturing of most viruses, once a host cell line is selected the development effort focuses on the process rather than on cell line development. In some cases, the host cell line is engineered to endow it with some “helper” functions for virus replication. However, with the increasingly easy access to genome engineering, the use of cellular engineering to alter host cell response or even a synthetic biology approach to engineer virus production systems will likely be exploited.

For the biomanufacturing of therapeutic proteins, subunit vaccines, and virus-like particles (VLPs), the production cell line is among the most important factors affecting productivity and product quality. Having a cell line with the ability to synthesize a high-quality product at a sufficiently high specific productivity is the prerequisite to a high productivity process. For the production of recombinant proteins, the factors affecting the selection of the production cell line are different from those for virus production (Panel 6.1). This chapter will focus on the development of cell lines for therapeutic proteins.

Transient Production vs. Stable Expression Cell Line

Transient Expression

RECOMBINANT PROTEINS are produced in mammalian cells through a transient expression of the gene of interest (GOI), or by establishing a cell line that produces the GOI product permanently (Panels 6.2–6.4). Transient expression is frequently used for the generation of small quantities of proteins (up to a few grams) for protein characterization or toxicity testing. First, cells are transfected with a high concentration of plasmids that include the GOI. A large number of plasmids are taken up by cells into the cytoplasm. A fraction of those plasmids then enter the nucleus, where the GOI is transcribed into mRNA. The mRNA molecules are exported back to the cytosol, translated into proteins, and secreted into the medium. Due to the large number of plasmids transfected and the strong promoter used, a large amount of product

Panel 6.2.

Protein Production in Mammalian Cells

Transient Expression

- The transgene is encoded in a plasmid or non-integrating virus vector
- The DNA-based vector must enter the nucleus for transfection, after which the mRNA of the transgene is translated in the cytoplasm
- The vector is not integrated into the genome; product synthesis stops after the vector is degraded or diluted due to cell division
- Useful for rapid production of research quantities of proteins

Stable Cell Line Expression

- The heterologous gene is integrated in the chromosome and expressed
- The heterologous gene can stay in the genome and be expressed stably for a large number of generations

Panel 6.3. Host Cells Frequently Used for Transient Expression

- HEK293 (human embryonic kidney fibroblasts)
- COS (from green monkey kidney cells)
- BHK cells
- CHO cells

Panel 6.4. Host Cells for Stable Expression of Heterologous Proteins

- CHO, NS0, SP2/0, HEK293, BHK, PER.C6
- Different host cell lines have different post-translational capabilities
- Different mammalian host cells produce somewhat different glycan structures
 - CHO and NS0 have different sialylated glycans (NANA and NGNA, respectively)

is produced. The plasmids in the cell are diluted as the cells divide and degraded over time. Thus, the transient production of the protein is usually limited to a few days. Nevertheless, the process is easily carried out (requiring only plasmid preparation) and is frequently used to produce a small quantity of proteins for characterization and testing.

Stable Expression

Stable expression involves the integration of the GOI into the genome of the host cell and the isolation of a cell line that has the capability of expressing and producing the GOI over a long period of time. Establishing a stable cell line for a given product has been the core of cell culture-based biomanufacturing. The cell line must be capable of producing product of the same quality in different batches, and in different locations, over the years. Once a production line is selected, cells are expanded to establish master and working cell banks. The stockpiled cells, typically stored in liquid nitrogen, are used for manufacturing purposes throughout the life cycle of the product.

Host Cells

AN IDEAL PRODUCTION CELL LINE has a high productivity, is able to synthesize the product to the desired quality, and is genetically stable for consistent product synthesis over the product's life cycle. Many factors affect the ease or difficulty of establishing a production cell line, including the choice of host cell, the vector carrying the GOI, and the method of inserting the GOI into the genome of the host cell. Of all the factors, the most important is the selection of the host cell.

Factors Affecting Host Cell Choice

In the early days of recombinant therapeutic protein technology, many studies explored the possibility of using various cell lines of rodent, monkey, and human origins for production, including mouse C127 and myeloma (NS0), Baby Hamster Kidney (BHK), and Human Embryonic

Kidney (HEK293) (Panel 6.4). It did not take long for CHO cells to dominate the scene.

When developing a production cell line, the host cell line is transformed from cells producing low levels of secreted proteins to highly secretory cells. Many different cell lines can be transfected with the GOI and made to produce the protein. However, only a few are able to have their genetic circuitry reprogrammed to expand their protein folding and other processing and secretory capabilities to the level necessary to become a hyperproducer. Different host cells have different degrees of “plasticity” that enable them to undergo such a transformation. The extraordinary capability of CHO cell lines to undergo the still poorly characterized transformation process is a major factor that made them the host cell line of choice for stable expression of recombinant proteins.

Different host cell lines have different capabilities and specialized post-translational modifications, such as gamma-carboxylation, phosphorylation, and lipidation. Cell lines derived from different species or different tissues can have different compositions of glycosyltransferases that catalyze different glycosylation reactions, or have different isoforms with different substrate specificity. Glycans on proteins produced in CHO, NS0, and other cell lines are somewhat different from each other because of their different sialylation.¹ Murine cells like NS0 and SP2/0 express galactose- α 1,3-galactosyl transferase and can potentially add α -galactose to the glycan. As discussed in Chapter 3, non-human mammalian cells produce N-glycolylneuraminic acid (Neu5Gc/NGNA) that is absent in human glycan. Human cells express only Neu5Ac/NANA (N-acetylneuraminic acid). CHO cells express only α 2,3-sialyltransferase thus do not synthesize the α 2,6-sialic acid that is seen in human glycans.

Nowadays, most new therapeutic protein products are produced in CHO cells. The CHO cell line was derived from an inbred strain of Chinese hamster.^{2,3} Many sublines and mutants were subsequently derived.⁴ The early passing-on of those sublines was not well documented. Among the CHO cell lines commonly used in industry, DXB11 and CHO-K1 were derived from the same ancestor line, while CHO-S and DG44 share the same lineage.⁴ It is common practice for a company to select a CHO cell line as the host cell platform for virtually all future products.

The choice of host cell also determines the possible host cell components that may contaminate the product. Non-human host cell proteins that are released into the medium upon cell lysis may become residual contaminants in the final product and elicit an immunogenic response upon administration to the patient. In the manufacturing process, the contaminating host cell proteins are removed and reduced to an acceptable level in downstream processing.

Most mammalian genomes harbor sequences of endogenous retroviruses. Sequences of endogenous retroviruses constitute about 1% of the

human genome. Many of these sequences are heavily mutated and are inactive. However, some viral proteins are expressed, and viral particles or even retroviruses can be produced from some host cells. The release of these viral particles by a host cell may contaminate the product and pose a safety concern. The level of viral particles released varies among host cell species and cell lines. Murine cell lines tend to produce high levels of viral particles. Both host cell proteins and viral particle contamination issues are mitigated through the manufacturing process. Nevertheless, the issues are factors that must be considered in host cell selection.

A human host cell line has the advantage of synthesizing “human form” glycans and eliminating the problem of non-human host cell protein contamination. The issue of human retrovirus particle inactivation and removal can be mitigated with the unit operation steps already implemented in contemporary product purification. However, in the past three decades the host cell line employed for biologic manufacturing has converged on CHO cells, rather than on one of the human cell lines established as a producer. CHO cells have a high level of genetic plasticity and can be readily reprogrammed into high protein secretory cells. They are also relatively easy to “adapt” to different culture conditions.

Host Cell Adaptation and Engineering

Many cell lines commonly used for recombinant protein production, including CHO, BHK, and HEK293 cells, are derived from adherent cells that proliferate by attaching to a compatible surface. They can be adapted to acquire the ability to grow in suspension (Panel 6.5). The adaptation to suspension growth is often achieved by eliminating the surface for attachment and waiting for adapted cells to grow. This is sometimes aided by adding a dispersing agent like heparan sulfate or dextran sulfate to the culture medium to reduce cell aggregation. After a couple of weeks, cell growth gradually resumes and cell concentration increases to allow for subculturing.

In addition to suspension growth, host cells can also be adapted to other culture conditions favorable for manufacturing. For example, the requirement of complex lipid additives and growth factors may be reduced or even eliminated by gradually forcing cells to grow under decreasing concentrations of the additives. The molecular mechanism of cell adaptation is not well understood. One case that is known is the adaption of NS0 cells to become cholesterol-independent. In the auxotrophic NS0 cells, a gene in the cholesterol biosynthetic pathway is silenced due to extensive methylation in the CpG island region of its promoter.⁵ The adapted cells have a reduced methylation in the promoter region to allow for the expression of the gene, thus alleviating the cholesterol dependency. Whether the adaptation to other culture conditions is similarly affected by epigenetic changes is not known.

Many traits are highly desirable in a producing cell line, including favorable metabolic characteristics, enhanced protein secretion capacity, and the capability to synthesize specific glycan features. However, many of these traits are highly complex and are most likely under the influence of an array of genes instead of a single gene. For example, the capability of growth in suspension likely involves alterations in the expression of a plethora of genes in the signaling pathway of growth control and in the organization of the cytoskeleton. Nevertheless, many desirable traits, such as the cholesterol independence described above, are under the control of a single gene or a small number of genes that are amenable to genetic manipulation.

Increasingly, the host cell line is being genetically altered to harbor favorable traits (Panel 6.5).⁶ Host cells can be engineered to overexpress anti-apoptotic genes that suppress apoptosis in the late stage of fed-batch culture and prolong the production phase. Cellular metabolism may be engineered to reduce waste metabolite formation and to facilitate the switch to a lactate-consumption type of metabolism. Host cells can be engineered to alter their glycosylation enzyme expression profile.⁷ Expression of heterologous 2,6-sialyl transferase allows for the addition of sialic acid through the 2,6-glycosidic linkage that is absent in the glycan synthesized by CHO cells.⁸ Knock-out of endogenous fucosyl transferase eliminates the fucose residue in the glycan-attached Asn₂₉₇ of IgG heavy chain.⁹ The afucosylated form of IgG has been reported to enhance antibody-dependent cellular cytotoxicity.⁹

Panel 6.5. Alteration of Host Cell Lines for Protein Production

Host Cell Adaptation to Acquire the Capability of:

- Suspension growth
- Serum-free, defined medium growth
- Altered nutritional requirements
 - e.g., glutamine- or cholesterol-free growth
- The genetic/epigenetic changes incurred in adaptation are likely complex and are not well understood

Host Engineering

- Effective when the gene-trait relationship is known
- Used to achieve:
 - Favorable metabolic characteristics
 - Apoptosis resistance
 - Improved glycosylation or other post-translational modifications

Generating a Recombinant Protein High-Producing Cell Line

THE FIRST STEP TO CONSTRUCTING a producing cell line from a host cell line is establishing a vector that contains the expression unit of the GOI coding for the product (Panel 6.6). The sequence encoding the GOI, often along with the vector sequence, is then inserted into the genome of the host cells so that a production cell line can be established (Figure 6.1). Some methods of gene insertion insert a defined

Panel 6.6. Steps in Cell Line Generation

- Vector/GOI transfection and integration into host cell genome
- Selection
- Amplification (may be omitted)
- Single-cell cloning
- Screening/productivity assessment/isolation of clones
- Cell stability evaluation
- Cell/product characterization

DNA sequence into either a predefined locus or random locations in the host cell genome. Most commonly, multiple copies of the vector, often in somewhat modified vector structure, are inserted into random locations in the genome. However, targeting a defined sequence into a prescribed site is increasingly being explored for creating a production cell line.

Typically, the vector is propagated in *E. coli* to generate sufficient quantities for transfection into the host cell. The vector typically carries a selectable marker gene so that the cells that have acquired the vector can be enriched after transfection using a selection agent (Figure 6.1).

In vitro screening of post-selection cells is then performed to screen for clones in this population that produce the product at high levels.

Until recently, it was common practice to amplify the copy number of the integrated vector in the isolated clones by subjecting them to a high concentration of selection agent. This kills the vast majority of cells, sparing only those with multiple copies of the vector; some of the amplified copies will have both the selectable marker gene as well as the GOI. Screening of the amplified cell population is then carried out to identify the candidate clones that produce the highest levels of product.

After vector integration and amplification, the cell population is rather heterogeneous. Different cells have different genomic loci of vector integration. They may also have different regions of their genome that have been rearranged, amplified, or mutated. Their growth behavior may also be different. Such a genetically heterogeneous cell population

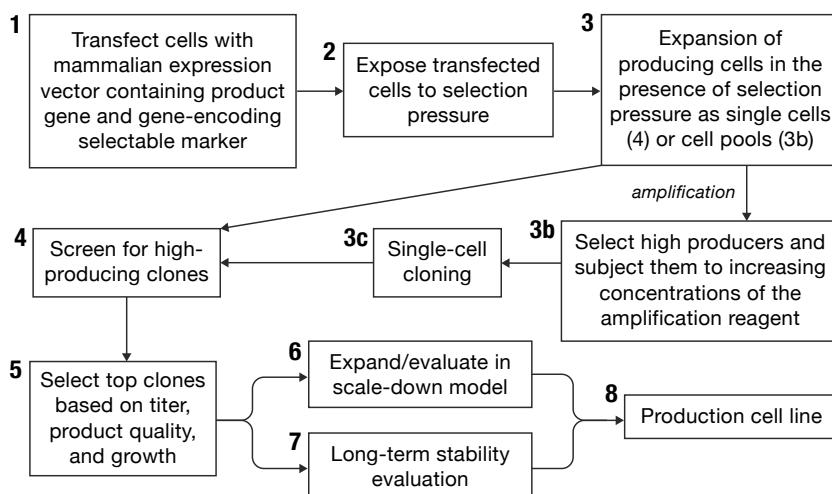


Figure 6.1. Typical steps in developing a high-producing cell line.

is called a “cell pool.” Subsequently, single-cell cloning is performed by sorting single cells into culture wells using flow cytometry or by culturing each individual cell into separated colonies in a plate. After single-cell cloning, the productivity of each clone is assessed. Those with high productivity are isolated and further expanded for growth characterization and product quality assessment for final production cell line selection. In some cases, the producing cells are then adapted to growth conditions more amenable to manufacturing conditions.

Vectors

A number of different types of vectors are used to insert a GOI into the host cell genome. Most commonly used is a plasmid vector that contains a complete GOI expression unit (Figure 6.2, Panel 6.7). The vector typically includes two sets of sequences: one for expressing the GOI and for selection of the host cell, and one consisting of bacterial sequences that enable the plasmid to replicate in *E. coli* so that a large quantity of plasmids can be obtained. The bacterial elements include a bacterial selectable marker for plasmid-carrying *E. coli* and the bacterial origin of replication for propagation in *E. coli*. These sequences, while essential in plasmid propagation, are not needed for GOI expression in the host cell.

Often, the plasmid is cut with a restriction enzyme to “linearize” the plasmid at a site that is not essential for the expression of the GOI before its delivery to the host cell. One may also use a specialized plasmid replication system that makes “minicircles,” thus eliminating most of the bacterial vector sequences from the final minicircle. The minicircle plasmid has a smaller size than the complete plasmid and gives a better efficiency in transfection.

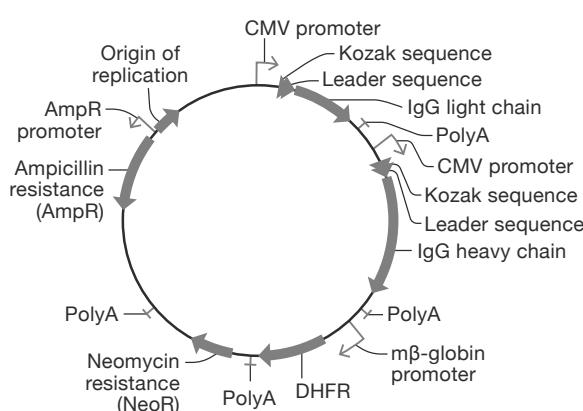


Figure 6.2. A typical vector for introducing a transgene into a host cell for recombinant protein production.

Panel 6.7. Elements in a Plasmid Vector

- Promoter
- Coding sequence of the gene of interest (GOI)
- PolyA signal
- May have integration signal/homology arms to facilitate targeted integration
- Selectable marker/amplification marker
- *E. coli* plasmid element (origin of replication, selectable marker)

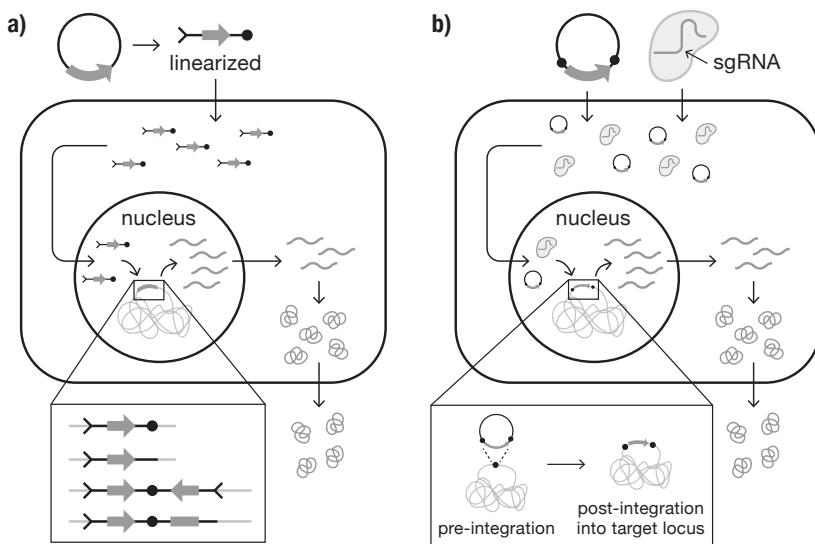


Figure 6.3. Integration of the GOI into the host cell genome. (a) Random integration using plasmids. (b) Targeted integration. In random integration, the GOI may be structurally modified before integration.

The plasmid is then transfected into the host cell. The plasmid is often subjected to intracellular degradation, and the truncated or intact plasmid may be ligated in various orientations. Through random events during host DNA replication and repair, the vector is integrated into the host genome without additional elements provided exogenously. The integrated sequence is thus highly variable in terms of its length and vector-host genome junction, and so is the locus on the host cell genome that the vector is integrated into (Figure 6.3a).

Increasingly, the GOI in a vector is directed to a specific locus and integrated as a defined sequence (Figure 6.3b). This is usually done by using DNA integration methods like CRISPR, transposon, and recombinase-mediated cassette exchange (RMCE) (such as LoxP or Flp systems). These methods all involve guiding the GOI to the locus using matching sequences that are homologous, or complementary (inset in Figure 6.3b). For CRISPR, the destination is a locus in the genome as defined by the DNA sequence (Figure 6.4a). A guide RNA directs Cas9 nuclease to the locus to cut the DNA. Through the cell's DNA repair system, the DNA segment containing the GOI is then inserted into the cut site. Next, the GOI is integrated into the genome at the target site. For LoxP and Flp, the destination is a pre-inserted LoxP or Flp sequence in the genome (Figure 6.4b). In the case of transposon-mediated gene insertion, the sequence marking the insertion loci is shorter and less defined. The integration is often numerous in the genome (Figure 6.4c). In all these methods, an enzyme (Cas9 for CRISPR, recombinase for LoxP/Flp, and transposase for transposon) is provided in the form of a plasmid component,

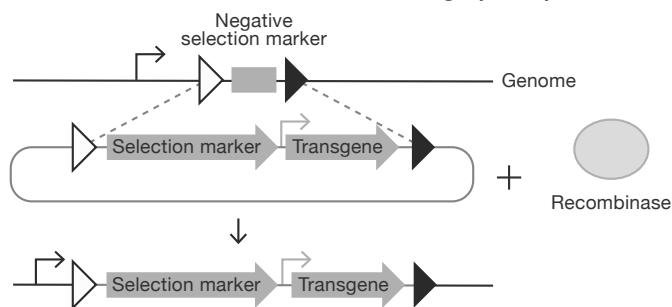
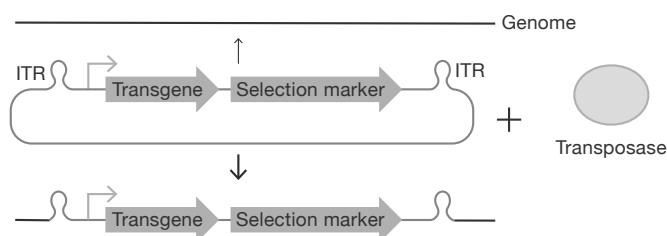
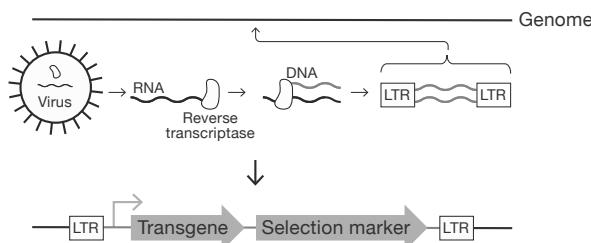
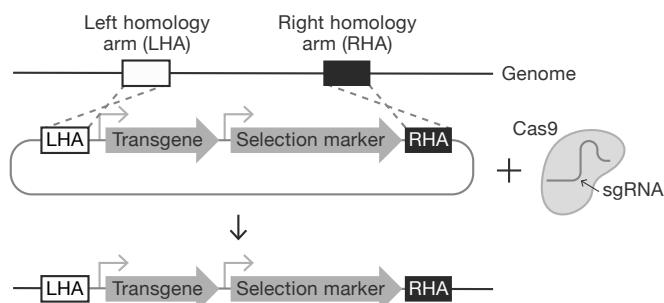
a) Recombinase-Mediated Cassette Exchange (RMCE)**b) Transposon****c) Lentivirus****d) CRISPR**

Figure 6.4. Schematic of GOI integration by (a) recombinase-mediated cassette exchange (RMCE), (b) transposase, (c) lentiviral, and (d) CRISPR systems. In CRISPR and RMCE, the GOI integrates into a defined locus. In transposon and lentiviral vectors, the integration is random.

mRNA, or protein, and co-transfected with the GOI and other components to facilitate the cutting of the genomic DNA and the integration of the insertion sequence.

Viral vectors are also used to deliver the GOI to host cells. Viruses enter the cell by binding to specific cell surface receptors, and then integrate into a chromosome to generate cells expressing the GOI. Using a lentivirus, the number of vectors needed for an infection is rather small compared to that for a plasmid vector. Because of the traditional safety concerns, viral vectors have not been used extensively for introducing GOIs for recombinant protein production until recent years. Lentivirus has been used as an *ex vivo* gene delivery system to cells for gene therapy. The viruses are modified to be devoid of viral protein sequences in the genome and are replication deficient. In place of viral genes, the lentiviral vector carries a payload that includes the GOI (Figure 6.4d). Upon integration into the host cell genome, the GOI is inserted along with the viral vector. Lentiviral vectors have a propensity for integrating into actively transcribed regions of the host cell's genome. They insert a defined DNA segment consisting of the virus genome into the host cell genome. Viral vectors have a very high efficiency of infection and integration. Using low multiplicity of infection (MOI), it is possible to isolate cells that have a very small number or even just one copy of the virus integrated into the host cell genome, and which have a defined viral genome sequence. Retroviral vectors are also frequently used in gene therapy; however, they do not transduce CHO cells as efficiently as lentivirals do. Both lentiviruses and retroviruses are RNA viruses, and are reverse transcribed to double-stranded DNA for replication upon infecting host cells. Reverse transcription has a higher error frequency than DNA replication. Thus, sequence verification after cell lines are isolated is necessary, just as in the case of other vectors.

Promoters and other regulatory elements

Many commonly used promoters that drive the transcription of the GOI are derived from mouse or human viruses, for example the CMV (cytomegalovirus), SV40, and HSV (herpes simplex virus) thymidine kinase (TK) promoters (Panel 6.8). Promoters for highly expressed genes from the host cell have also been used. Additionally, synthetic promoters have been isolated with the aim of increasing transcript levels.¹⁰ The promoters used for driving GOI expression are mostly constitutive and are expressed all the time.

The promoter used for product protein expression is also subjected to epigenetic modification of the host cells. Silencing of the promoter will affect the expression level of the GOI and therefore productivity. For example, the methylation of CpG dinucleotides in the CMV promoter/enhancer in a recombinant CHO cell line can cause a decrease in productivity.¹¹

In addition to the strength of the promoter, other elements affect the transcript level of the GOI. The human CMV, in the presence of intron A, has been shown to enhance the protein expression levels of the GOI.¹² By including the 5' and 3' flanking control regions of EF-1 α , the expression of the GOI driven by CMV or the human EF-1 α promoter in CHO cells can be increased.¹³ Enhancers are DNA sequences that increase gene transcription and may not necessarily be located close to the promoter sequence. In some cases, a promoter such as CMV is used in conjunction with its associated enhancer sequences in the vector construct.

The expression of the inserted GOI is affected not only by the promoter and enhancer used, but also by the regional context of the genome near the integration locus. Some regions on the genome may be unfavorable to transcription. For example, the presence of CpG islands in close vicinity to a heterochromatin region may negatively affect gene expression. To minimize the probability that a randomly integrated GOI falls under the influence of negative position effect, a cis-acting barrier element called an insulator can be placed in front of the promoter of the GOI.

GOI-coding sequence and leader sequence

The cDNA of a GOI isolated from a source organism may not be the most efficient for expression in a host cell from a different species. The coding sequence of the GOI is often modified to enhance the efficiency and fidelity of its translation (Panel 6.9). The codon usage can be changed according to the abundance profile of tRNAs in the host cell. Codons may also be changed to remove cryptic splice sites, optimize GC content, and remove direct repeats.

Secreted proteins have a leader sequence for translocation into the endoplasmic reticulum during translation. The leader sequence differs in different proteins and species. A strong signal sequence, such as one derived from human albumin, may give a higher productivity. However, the effectiveness of a signal sequence may also be influenced by the downstream protein sequence. Furthermore, the leader sequence needs to be cleaved upon the entry of the nascent protein molecule into the ER. Incomplete or nonspecific post-translational cleavage of the leader

Panel 6.8. Driving Transcription of a GOI

Promoters for Transgene Expression

- Mostly constitutive promoters, virus origin
 - SV40, CMV, thymidine kinase (TK)
- Strong promoters from host cells
- Synthetic promoters
- Sometimes used with an enhancer, flanking control regions

Epigenetic Effects on Promoter Expression

- Some promoters are subject to CpG silencing
- Transcription affected by local context of integration site
- The position effect of heterochromatin may be minimized by using insulators

Panel 6.9. Expression Unit of a GOI

- Coding sequence is often codon-optimized
- A leader sequence for efficient secretion (e.g., human albumin leader)
 - The leader sequence selection affects secretion and cleavage
- 5' UTR and 3' UTR (untranslated region) influence transcript stability, transcription termination

sequence can result in heterogeneity in the mature protein sequence at the N terminus. The efficiency of cleavage of the leader signal is thus an additional consideration in its selection.

The 3' UTR, i.e., the region between the stop codon and the polyA tail on the mRNA, is important for proper termination of transcription and also plays a role in mRNA stability. It contains the signal for the addition of polyA to the transcript that allows for its transport from the nucleus to the cytoplasm. Commonly used polyA signals for recombinant protein production include SV40 and BGH (bovine growth hormone) polyA.

Selectable marker and reporter

Upon the introduction of the plasmid vector into the host cells, some cells will receive the vector. A portion of those vectors will enter the nucleus and express the encoded genes. Among the genes encoded in the plasmid is a selectable marker such as antibiotic resistance. The expression of the selectable marker allows the cell to survive in the presence of the selection reagent. Eventually, some plasmids that have entered the nucleus will be integrated into the genome and become a permanent part of the genome. Plasmids are not capable of self-replication, and free plasmids in the cell are gradually degraded. Only those cells that express the selectable marker will survive under selective conditions. Those that survive and grow under selection over time therefore have the vector, including the selectable marker gene, integrated into the genome. Many of the integrated vectors will also contain the GOI. The selectable marker gene and the selective condition, typically the presence or absence of the selective agent, are thus used in pairs. Some commonly used selectable markers and their mode of action in mammalian cells are listed in Table 6.1.

Selectable markers fall into two categories: dominant and recessive (Panel 6.10). A recessive marker resides in cells with a particular genetic background and causes a growth deficiency. The introduction of a compensatory gene leads to overcoming the deficiency. For example, cells without a functional dihydrofolate reductase (DHFR) gene will require the additional supplementation of thymidine and glycine in the culture medium for growth. The introduction of a functional DHFR enables them to grow without the supplements. Therefore, after the transfection of a plasmid containing DHFR, only the transfected cells that express the newly acquired DHFR will grow in the absence of thymidine/glycine. Similarly, thymidine kinase (TK)-defective mutants require thymidine to be included to the culture medium. The introduction of a functional TK

Table 6.1. Commonly Used Antibiotics for a Selection of Stably Transfected Mammalian Cells

Antibiotic	Mode of selection	Resistance gene	Mode of resistance
Geneticin (G418)	Blocks protein synthesis by inhibiting elongation step	Neomycin phosphotransferase (npt)	Phosphorylation of Geneticin
Hygromycin B	Inhibits protein synthesis by disrupting translocation and promoting mistranslation	Hygromycin phosphotransferase (hpt)	Phosphorylation of Hygromycin B
Puromycin	Blocks protein synthesis by causing premature chain termination	Puromycin N-acetyltransferase (pac)	Acetylation of Puromycin
Blasticidin S	Inhibits protein synthesis by interfering with peptide bond formation	Blasticidin S deaminase (bsr)	Deamination of Blasticidin S
Zeocin (Bleomycin)	Intercalates into and cleaves DNA	Bleomycin resistance protein (ble)	Binds stoichiometrically and prevents Zeocin from binding to DNA

gene allows for cell growth in the absence of thymidine. Conversely, the presence of a dominant selectable marker is protective to the cell under selective conditions. By introducing the selectable gene to the cell, the cell is endowed with a resistance to a given lethal condition.

A resistance gene typically encodes an enzyme that modifies the selective chemical agent to destroy its activity. The phosphorylation and acetylation reactions employed for inactivating antibiotics require intracellular reactants (ATP, acetyl group donor). Those enzymes are therefore only effective in destroying the selective agents intracellularly. The hydrolysis enzyme, on the other hand, may be active even when released into the medium after cell lysis. In the selection process, the concentration of the selective chemical agent decreases with time and the rate of decrease is dependent on the type of resistance and the concentration of transfected cells. Thus, the optimal concentration for the selection for each agent is not only dependent on cell line but also on cell concentration. Clonal selection and population selection may have rather different optimal concentrations of the selective agent.

Another class of selective agents interferes with the uptake of a toxic selective agent. The multidrug resistance gene (MDR) confers cells

Panel 6.10. Selective Markers for Transgene Expression

Recessive Selection

- DHFR (dihydrofolate reductase)
 - Effective on DHFR-deficient background
- TK (thymidine kinase)
 - Effective on a TK-deficient background

Dominant Selection

- Antibiotic resistance
 - Neomycin, hygromycin, blasticidin
- Use a mutated weaker selectable marker and a weak promoter to select the integrant into a transcriptionally active site

with resistance by increasing their ability to pump toxic substances, such as colchicine, out of cells. Its overexpression allows for selection from a background of cells not expressing MDR.

The expression of the selectable marker is usually driven by a weaker promoter than the promoter driving the expression of the protein of interest. Sometimes a mutated selectable marker with a lower activity to rescue the missing enzyme function is used. With such a “weakened” expression of the selectable marker, the integration site must be transcriptionally active and allow for a higher transcript level of the selectable marker in order to counter the selection pressure. This may lead to increased expression of the protein of interest.

Not all cells that have received transfected plasmids and express the selectable marker protein will also express the GOI. The frequency of the false positives (those expressing the selectable marker but not the protein of interest) varies with the selectable marker used and the vector design. Some false positives may arise through failure to co-integrate the GOI, or through mutations that develop alternate mechanisms to neutralize the effect of the selectable marker. Thus, choice of the selectable marker can be important for reducing false positive clones.

The selection of transfectants that harbor the GOI is performed at single-cell cloning stage to isolate clones, as well as at a population level to generate the transfected cell pool. The isolation of GOI-transfected cells may also be performed without using a selectable marker and selective pressure. Instead of a selectable marker gene, a reporter gene may be used in the vector. The expression of the reporter protein gives the transfected cell fluorescence upon excitation with light of the appropriate wavelength. The transfected cells can be sorted using fluorescence-activated cell sorting (FACS) into single cells for single-cell cloning or into a pool. Commonly used reporter proteins include GFP (green fluorescent protein), mCherry, tdTomato, and RFP (red fluorescent protein). Such reporter-gene-based transfectant isolation is frequently used in research, but less often for industrial cell lines.

Delivery of Foreign DNA and Its Integration into the Host Cell Genome

Successful transfection of a plasmid vector requires that the host cell membrane first become permeable or else that a mechanism is in place for the plasmid uptake. Calcium phosphate precipitation, cationic polymers, and liposomes of the DNA molecules are commonly used to deliver plasmids to the cytoplasm of the host cell (Panel 6.11). The plasmids are first prepared as particles or encapsulated into lipid vesicles. Then, through particle interactions with the cellular membrane and an endocytosis-like mechanism, plasmids are taken up by cells. Electroporation

is another frequently used method. It delivers the DNA molecules to the cytoplasm through the openings in the membrane generated by an electric field. All of these methods use a very high plasmid-to-cell ratio, but only a moderate DNA concentration. While thousands of the plasmid molecules may enter each cell, only a small portion translocate to the nucleus where they can be transcribed to express the gene(s) encoded in the plasmid, including the selectable marker gene.

Typically, the transfection efficiency is evaluated by the expression of the resistance marker gene or another reporter gene. This is in turn dependent on the plasmids' entry into the nucleus and subsequent expression. Because of this, the efficiency rarely approaches 100%.

The plasmid, in its circular and supercoil form, may be used for transfection directly. Often, it is first linearized using a restriction enzyme whose recognition sequence is outside of the GOI and selectable marker. Linearized plasmid DNA is thought to result in a higher efficiency of genome integration. While some transfected cells may have a single copy of the integrated vector carrying the GOI, others may have multiple copies that integrated into different loci upon transfection. Furthermore, the vector integrated into the host cell genome may not be intact, but rather be truncated or fused by multiple copies (Figure 6.1).

Gene Amplification and Integration of the GOI

For three decades, an effective and commonly practiced step in cell line development has been to subject the vector-transfected cells to very high levels of selective pressure to increase the number of copies of the GOI in the host cell genome. The rationale is that more copies of the GOI give a high transcript level and a high productivity. Upon exposure to a very high level of selective pressure, most cells will die, leaving only those with many copies of the selectable marker gene. Since the GOI is nearby the selectable marker gene, it is highly probable that it will be co-amplified during the amplification and produce the product at high levels. Structurally, the vector sequence may be amplified alone, or it may

Panel 6.11. Methods of Gene Transfer

Lipofection/Lipid-Mediated Gene Transfer

- A mixture of DNA with an amphipathic compound (DOTMA, DOPE, etc.) that simultaneously interacts with DNA and hydrophobic portions of the membrane, allowing passage of DNA into the cell

Electroporation

- Exposes cells to a high-voltage electro-pulse in the presence of the DNA solution
- Introduces pores in the plasma membrane, allowing entry of DNA
- Duration of pulse and strength of the electric field varies with cell type

DNA-Calcium Phosphate Co-Precipitation

- DNA and calcium chloride are added dropwise into a HEPES buffer with sodium phosphate (1 mM)
- A fine precipitate forms in 5–30 min, and is added directly to the cells
- Cells take up DNA-containing particles via an endocytosis-type mechanism

be co-amplified with adjacent segments of the genome. The amplicons may be in tandem or combined in different orientations. Some amplified segments may be translocated to different loci nearby or far away in the same or different chromosome.

DHFR amplification system

Two gene amplification markers are commonly used in mammalian cells: DHFR and glutamine synthetase (GS). The DHFR gene as a selectable marker for gene amplification is used in conjunction with its chemical antagonist, methotrexate (MTX) (Figure 6.2, Panel 6.12). DHFR is an enzyme that catalyzes the conversion of folate to tetrahydrofolate, a compound required for the biosynthesis of glycine, thymidine monophosphate, and purine. MTX is a folate analogue that inhibits DHFR, thereby leading to cell death in the absence of thymidine and purine in the medium.

After exposure to a high concentration of MTX for a period of time, the surviving cells have multiple copies of DHFR. The size of amplicons surrounding DHFR varies widely, from 10 kbp to 10 Mbp. The GOI and DHFR genes are co-amplified. All copies of the GOI may not be transcriptionally active. Some may be structurally compromised due to DNA truncation, fusion, or other modifications. It is difficult to assess which of the intact copies of the GOI are transcriptionally more active. The amplification process can be performed as a single step of MTX exposure over one to two weeks, or in multiple step-wise increases in MTX concentration. Highly MTX-resistant cells may contain several thousand copies of DHFR.

DHFR-based gene amplification is more efficient in a DHFR-defective genetic background to minimize or eliminate the possibility that the resistance arises from the amplification of the endogenous DHFR without concurrent amplification of the GOI. CHO mutants (DSB-11, DG44) that are deficient in DHFR are commonly used for amplification-based cell line development. These DHFR-deficient cells require the addition of thymidine, glycine, and hypoxanthine to the medium. These cells do not grow in the absence of added nucleosides unless they acquire a functional DHFR gene.

GS amplification system

The GS selection system is based on the reaction that synthesizes glutamine from the substrates glutamate and ammonia. Most mammalian cell lines require glutamine to grow, since their endogenous GS activity is low (Panel 6.13). The GS expression vector contains a GS gene, along with the GOI, to allow for the selection by growth in glutamine-free cell culture media. GS is usually driven by a weaker promoter than the GOI.

Panel 6.12. DHFR Amplification System

- Use DHFR^{-/-} (DG44) or DHFR^{+/-} (DXB11) background, vector contains coding sequences of DHFR and the linked GOI
- DHFR catalyzes conversion of dihydrofolate to tetrahydrofolate for purine biosynthesis, and the reaction is inhibited by methotrexate (MTX)
- Treated with a high concentration of MTX, the surviving cells attain multiple copies of the DHFR gene and the GOI

Panel 6.13. GS Amplification System

- Cultured cells have very low glutamine synthetase activity and require glutamine for growth
- GS catalyzes conversion of glutamate/NH₃ to glutamine, can alleviate glutamine requirement
- The reaction is inhibited by methionine sulfoximine (MSX)
- After exposure to a high concentration of MSX, the surviving cells have multiple copies of the GS gene and GOI

By supplementing a high concentration of GS inhibitor methionine sulfoximine (MSX) in the medium, it is possible to select for transfectants that have multiple copies of GS.

Low multiplicity of GOI integration

In the development of B lymphocytes, only one copy of the IgG heavy gene is actively transcribed, while the copy in the other allele is inactivated so as to render it untranscribable, a phenomenon called allele exclusion. Yet, B cells continue to develop into high antibody-producing plasma cells. The phenomenon suggests that, with the right cellular machinery, a single copy of the GOI is sufficient for high protein expression and secretion. This was realized in a high-producing NS0 cell line that was engineered to have only one copy of IgG heavy chain and light chain genes.¹⁴ Having multiple copies of the GOI certainly facilitates high transcription, but a high level of expression is achievable even with a single copy of the GOI.

A high degree of gene amplification through the application of a high concentration of selective pressure often gives rise to phenotypically and genotypically unstable producing cells. It is advantageous to accomplish the same high level of transcription using a lower degree of amplification. A number of methods aim to attain high expression levels with a low multiplicity of the GOI. DHFR-selectable markers are used in conjunction with an impaired neomycin resistance gene. After plasmid transfection and antibiotic G418 selection, only cells to whose genome the vector has integrated in a transcriptionally active region and that give a high level of neomycin resistant transcripts can survive. The method thus allows for the selection of cells that the GOI has integrated into in a very transcriptionally active region known as a “hot spot” region. The amplification is then performed at a lower level of the DHFR

inhibitor MTX. Since the locus of integration is transcriptionally active, the high-expressing clones isolated have only a few copies of the GOI.

A low multiplicity of the GOI can also be obtained by omitting the amplification step. After transfection, a large number of plasmids enter the cell and only a small fraction of those integrate into the genome. The number of copies of the vector (and the GOI) integrated into a cell distributes over a range. By titrating the selective agent concentration, transfecants above different thresholds of multiplicity of the GOI are selectively enriched.

Integration site of the GOI

The locus of GOI integration influences how actively it is transcribed and how stable it is over time (Panel 6.14). The genome region of GOI integration may also affect the stability of the GOI structurally. The genome of aneuploid cells constantly acquires structural variants due to

the reduced DNA replication fidelity. Some regions in the genome may be prone to breakage or gain or loss of copies. The data on genome structural variants for industrial cell lines, and even for research cell lines, is scarce, but can be readily acquired through genome sequencing.

Some genome segments are transcriptionally more active than others, or are considered to be 'hot spots' because of their high accessibility to transcription factors. The genome regions that have an abundance of such highly accessible segments are considered to be super-enhancer regions. Although data is still lacking, it is possible that the insertion of the GOI into such a highly accessible region will facilitate high transcription

Panel 6.14. GOI Integration Site

- Genome regions of integration may have varying structural stabilities
 - Unstable if near a telomere
 - Some regions may be prone to segmental deletion, insertion, or translocation
- Local context of genome may affect transcription
 - Vicinity to heterochromatin may encounter heterochromatin invasion and silencing
 - Integration in transcriptionally active, highly accessible, or super-enhancer regions may favor transcription

of the GOI. Local context may negatively affect transcription. The genome consists of euchromatin regions and heterochromatin regions. In heterochromatin regions, DNA is densely packed, making it less accessible to transcription factors. Super-enhancer and heterochromatin regions are both lineage-specific and are inherited in cell division. However, variation does occur and the boundary of a heterochromatin region may extend, causing a decrease in transcription of nearby genes.

Hence, in targeted integration of GOI by CRISPR or RMCE, the ideal target site on the genome should reside in a structurally stable region where the transcription activity is very high. In the case of random integration, the integration locus of the GOI can be identified by genome

sequencing or focused integration site sequencing, and a candidate cell line that has a favorable integration can be selected for production.

Isolating a Production Cell Line

Hyperproductivity traits

Classical cell line development largely revolves around means that increase the transcript level of the GOI. For example, GOI amplification and its hot spot integration primarily affect the transcript levels of the GOI. Although highly abundant transcription is necessary for high productivity, by itself it is hardly sufficient for converting a host cell into a hyperproducer.

In our body, B cells differentiate to antibody secretory plasma cells upon antigen stimulation over the course of a few days. The changes in cell properties in such a transition may give hints to the traits that a hyperproducing cell must acquire in addition to achieving highly active transcription of the GOI (Figure 6.5).

After antigen stimulation, B cells show the characteristics of unfolded protein response (UPR). UPR is typically induced in cells overloaded with secretory protein molecules and is marked by an expanded ER. Over the course of about four days, B cells increase their size significantly and expand their ER by at least 15-fold. The expansion appears to involve the whole content of the ER, not merely the selective increase of some ER proteins. The expansion starts before mass antibody production. The enzyme disulfide isomerase (PDI), which catalyzes disulfide bond formation, increases in the ER as the B cells differentiate. Many of the redox balance enzymes in the cytosol and mitochondria are also upregulated. There is evidence that the Golgi increases along with the ER. Along with these changes, the metabolic capacity also increases significantly. The transcription factor XBP-1, which activates transcription of many ER proteins, is upregulated, coinciding with the increase in antibody production around a day post-stimulation.

Taken together, the traits that give rise to the hyperproductivity of plasma cells include high secretory capacity, elevated energy generation,

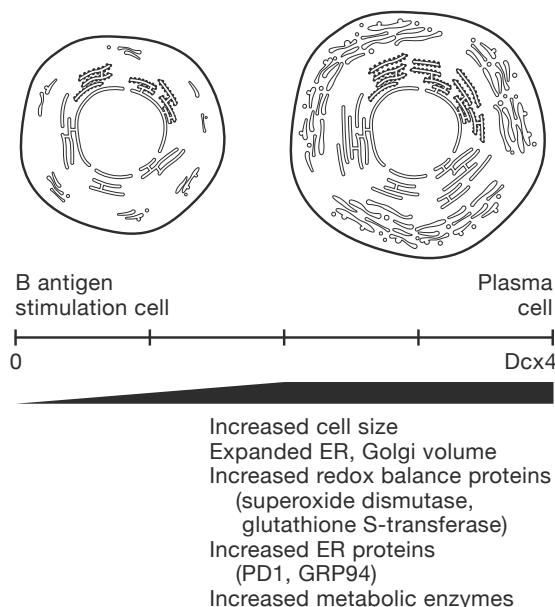


Figure 6.5. Cellular changes occur in the differentiation of a B cell to a plasma cell.

Panel 6.15. Cellular Properties Contributing to Hyperproductivity

- High transcript level of the GOI
- High translation capacity
- High secretory and post-translational processing capacity
- High activity for energy metabolism
- High redox balance capacity
- Ability to grow to high cell concentration and sustain a high viability in stationary phase

and redox balance (Panel 6.15). For glycoprotein products, elevated post-translational processing capacity, including high mannose glycan trimming in the ER, glycan extension in the Golgi, and membrane vesicle trafficking for membrane homeostasis, also contribute to a high productivity.

In addition to an increased product synthesis capacity, a hyperproducer should have superior growth kinetics. A higher tolerance to adverse growth conditions in the late culture will enhance product accumulation by

sustaining a high cell viability over a longer duration in the stationary growth stage. Hyperproductivity is thus a complex combination of many cellular properties that each by itself is a complex trait, including high secretory and post-translational modification capacity, redox balance, and increased antiapoptotic activities. Each complex property is likely to be subjected to the control of multiple regulatory networks rather than that of a single master regulation gene. The path that leads a transfected host cell to becoming a hyperproducer is not unique. It is likely that many different combinations of cellular changes can lead to similar levels of high productivity. In other words, different hyperproducing cell lines are likely to have different changes in their gene expression that make each of them a hyperproducer. A more complex and comprehensive combination of important traits may lead to higher productivity. The possible involvement of many alternative combinations of traits that can give rise to hyperproductivity may make the identification of “marker” genes for screening or engineering a hyperproducer elusive.

Single-cell cloning and cell characterization

After the selection of the transfectants, the resulting cell population is heterogeneous in their integration sites, copy number of the GOI, growth behavior, product secretion kinetics, etc. In a heterogeneous population, the fastest growing subpopulation will become predominant over time. The population shift may cause the product quality to shift as well. Single-cell cloning is performed on the cell pool by limiting dilution or by FACS to ensure that the production cell line starts with a homogeneous genetic and epigenetic background (Panel 6.16). As will be discussed next, it is inevitable that the population will become heterogeneous again. Nevertheless, barring further genetic change to the GOI, the ensuing producing cell line will have identical genetic information and local gene expression context.

It takes roughly 10–14 days for a single cell to grow into a colony of about 500–1000 cells. The duration is constrained by the cell growth

rate. The low cell concentration used in single-cell cloning (~1 cell/100 µL in a culture well, or ~10 cell/mL) likely incurs a long lag phase and extends the duration of single-cell cloning.

When using well plates, the number of clones that can be screened is often limited. The adoption of high-throughput methods using colony-in-gel systems allows thousands of colonies to be screened, drastically increasing the probability of isolating hyperproducers. Typically, an in-gel product concentration assay is integrated into an automated colony-in-gel system to identify and isolate cell colonies that produce more product. Fluorescence-labeled antibodies against the product are added to the cell-gel, where they bind to the secreted product molecules that accumulate around the colony. The radius and the total fluorescence around the colony is a quantitative indicator of the amount of product produced, while the size of the colony is used to estimate the growth rate. The increased screening capability has allowed very high secretory cell clones to be isolated.

After single-cell cloning, clones with a high productivity are isolated. Those which also exhibit robust growth and production characteristics are further characterized for product quality and production stability before the final production line is selected.

Single-cell cloning and population heterogeneity

Aneuploid cells have abnormal chromosomal counts, and many of their chromosomes have macroscopic structural aberrations. Such chromosomal alterations accumulate over time, generating a heterogeneous cell population created by different cells carrying different sets of chromosomes with different structural abnormalities. Some cell lines are highly heterogenous, with a wide range of chromosome abnormality and numbers. For example, some CHO cell lines have a wide range of structural variation.^{4, 15} Others, such as Vero cells,¹⁶ have a relatively low level of heterogeneity, with a narrow distribution of chromosome numbers and variations of karyotypes. At a detailed molecular level, base-pair mutations, varying lengths of segmental DNA duplications, deletions, inversions, and translocations frequently occur in aneuploid cells. These structural variations occur during DNA replication at the single-cell level. Some genomic structural variants may carry phenotypic or physiological variations as well. Cells with some types of genome or karyotype variations may grow to become a noticeable

Panel 6.16.

Isolating Hyperproducing Cell Clones

- Single-cell cloning
 - To ensure producing line starts with a homogeneous population
- Improve efficiency by
 - Using conditioned medium
 - Clone-picking instrumentation
 - Cell colony-gel system, entrapment, or 3-D printing
 - Robotic microfluidic nano-well system (increases cell concentration by using nano-wells)
- The increased capacity of screening enables the isolation of rare hyperproducers

subpopulation. The population of an aneuploid cell line is thus heterogeneous in genomic variants.

In single-cell cloning, the progeny population starts with a uniform integration locus and genotype. However, the karyotype in the cell population inevitably becomes heterogeneous again. In CHO cell lines, the heterogeneous distribution in the karyotype is re-established within a couple dozen generations after single-cell cloning.¹⁵ What single-cell cloning from a cell pool achieves is thus the homogenization of the GOI integration loci and other genetic and epigenetic alterations derived during cell line development. The population heterogeneity at the genome level is re-established just like in a typical aneuploid cell line.

Assessing Production Performance of Candidate Cell Lines

After the single-cell cloned candidate cells are expanded, several cell characteristics are evaluated for the selection of the production cell line, including productivity, product gene integrity, the GOI transcription level, and product quality. The productivity of the candidate clones determined in cell-gel or well plates is under conditions that are rather different from those during manufacturing. Before selecting the

finalist candidates, small-scale cultures ought to be carried out under simulated process-scale conditions. The selection is based on productivity and protein quality parameters such as structural variants and the glycosylation profile (Panel 6.17). A variety of culture devices are used for high-throughput small-scale culture, from culture tubes and shaker flasks to automated miniaturized systems that mimic stirred tank bioreactors.

The integrity of all of the expressed copies of the product gene(s) integrated in the genome can be verified by genome sequencing or by PCR amplification of the integrated GOI followed by amplicon sequencing. Mutations or structural variants in one or more copies of the GOI must be absent in the cell line. RNAseq may also be used to detect mutations and other structural variants.

The transcript of the GOI is usually expressed at a very high level, and is covered at a very great depth in RNAseq. The

Panel 6.17. Characterization of Candidate Cell Lines for Production

Process Performance Evaluation

- Use high-throughput systems, parallel evaluation of multiple candidates, and replicated experiments for statistical confidence

Cell Characterization

- Integrity of integrated GOI copies
 - DNA sequencing of the GOI
 - RNAseq, analyze GOI transcripts
 - Mutation may be present only in one GOI copy and only in some cells if the mutation occurs after GOI integration or single-cell cloning
- Integration site determined by sequencing
 - Structural stability of the GOI integration region
 - Distance to heterochromatins

Product Characterization

- Protein integrity
- Glycan profile

depth of coverage allows for the identification of rare base changes and insertions or deletions (indels). Therefore, a sequence change is detectable even if it occurs only in a small fraction of mutant cells.

Ultimately, the structural integrity is evaluated by structural analysis at protein levels. Some glycan structure features affect the biological functions of some classes of therapeutic proteins with the same mode of action. If the preferred glycan feature is known, the glycan profile of the product is an important parameter in cell line selection. For example, for an IgG product that mediates ADCC, the preferred cell line is one that produces the afucosylated glycan as the dominant form.

Evaluation of Productivity Stability

From a single cell at the beginning of single-cell cloning until the progeny cells reach the final stage of the manufacturing culture, a production cell line is likely to undergo more than 60 population doublings (Panel 6.18). The number of cell divisions exceeds that of a fertilized egg growing into a hamster, a mouse, or even a human adult. DNA replication, both *in vivo* and in cultured diploid cells (before senescence), has very high fidelity, maintaining sequence identity and structural integrity at a chromosomal level. However, aberrant changes occur constantly in aneuploid cells. Some of these changes may lead to a decrease in productivity. Hence, the finalist cell lines are serially cultured for 60 doublings, or more than the number of doublings required to reach the manufacturing scale, and their productivity examined over time to verify their stability over the product life cycle (Panel 6.19).

Besides the structural instability and transcriptional silencing of the GOI, changes unrelated to the cell's capability to synthesize the product may impact the product titer accumulated in the culture (Panel 6.20). For example, the transition

Panel 6.18. Stability of Producing Cell Lines

- Diploid cells used for vaccine production are stable within their allowable passage number
- Recombinant producing lines are aneuploid, have an unstable karyotype
- From cell cloning to product, the life span may take 60 cell doublings
- From cell thawing and expansion to manufacturing, may take 40 cell doublings
- Production cell line must remain productive throughout its life cycle

Panel 6.19. Evaluation of Productivity Stability

- Serial expansion of cells for a large number of doublings (e.g., 40–60), evaluate titer at the end
 - Labor intensive, long time duration
 - Can detect only significant decrease in product titer
- Detect change in clonal distribution of productivity over 10–20 cell doublings
 - Productivity will decrease only if the low-producing subpopulation increases, or if (unlikely) the entire population becomes less productive
 - Examine the proportion of low-producing clones over time
 - By single-cell cloning and determining the percentage of clones whose titer is below the threshold
 - By single-cell sorting followed by single-cell quantitative PCR to quantify the transcript of the GOI to check transcriptional stability

Panel 6.20. Productivity Instability

- Transgene alterations cause instability
 - Loss of GOI copy
 - Transcriptional silencing: promoter silencing, heterochromatin invasion
- To enhance stability
 - Robust promoter, native promoter
 - Low transgene copy number
 - Integrate to a structurally and transcriptionally stable genomic region
- Cell property alterations may cause secondary effects on productivity
 - Metabolic behavior change
 - Secretory pathway change
 - Change in growth kinetics

from the production of lactate to its consumption in the stationary phase of growth has an impact on growth and production kinetics. A change in this metabolic shift behavior of the cell line will cause the productivity to decrease. It is also plausible that changes that broadly affect cell physiology, such as protein processing capacity or altered redox balance, may cause productivity to decrease. In those cases, the drop in productivity is real, but its cause is most likely unrelated to the instability of the productivity. In the following section, we will focus on that instability.

*Possible causes of lost productivity***Loss of functional copy of the GOI**

In cell division, a cell may lose a copy of the GOI that makes a large contribution to transcription and become a low- or non-producer. In spite of their propensity to develop genomic structural variations, the vast majority of genes in aneuploid cells are stable and not easily lost. GOIs, if integrated into the genome like an endogenous gene, should also be stable. However, GOI copies that are amplified in tandem, integrated near a telomere, or integrated in a region that is vulnerable to structural change may have a higher frequency of loss.¹⁵

Transcriptional silencing

Mutation or epigenetic changes in the regulatory region of the GOI may diminish its transcription. The methylation of the CMV promoter at its CpG residues causes a reduced transcription of the GOI.¹¹ The invasion or spread of the neighboring heterochromatin region into the locus of the GOI insertion may cause transcription silencing of the GOI.

Population heterogeneity and low producer dominance

Genomic structural variations and epigenetic silencing occur at a low frequency in every cell division. At a low frequency, some low- or non-producing cells may arise through such genomic or epigenetic changes. The low-frequency occurrence of low- or non-producing cells is not deleterious by itself. A decrease in the product titer is detectable only when the mutant population becomes large. Two mechanisms can lead to an increase in the representation of the mutant: (1) an accumulation of mutants through continued mutations of the parent population,

or (2) a faster growth rate of the mutants than of the parent. As shown in Figure 6.6, if the mutant does not grow faster, continued mutations alone will not lead to a high enough non-producer abundance level to cause a detectable drop (20%) in productivity over 60 generations unless the mutation rate is at an exceedingly high level of 1/1000 in a generation. On the other hand, if the non-/low-producing cell grows 20% faster than the parent, even at a moderate mutation rate of 1 in 10^5 per generation, there is a significant drop in the percentage of parent cells (and in productivity) over time. In a continuous culture study where non-producing hybridoma cells became a significant fraction of the population over time, the non-producing cells were found to grow nearly 20% faster than the parent producing cells.¹⁷

It is highly probable that a mutant cell with a low productivity grows faster. The energetic cost of product synthesis in a hyperproducing cell line is high. Free of that energetic load, a mutant cell may grow faster, although this may be dependent on culture conditions.

One way to evaluate cell line stability is thus to perform single-cell cloning on two populations that are 10–20 generations apart to quantify the change in the percentage of low- and non-producing clones (Panel 6.19). Whether the percentage increases or not over generations will give a measure of the stability of the productivity. Alternatively, one can perform single-cell quantitative PCR on the product transcript level after cell sorting. An increase in a subpopulation with a reduced transcript level will suggest productivity instability.

With targeted integration, one can guide the integration of the GOI to a stable genomic locus that has a very low

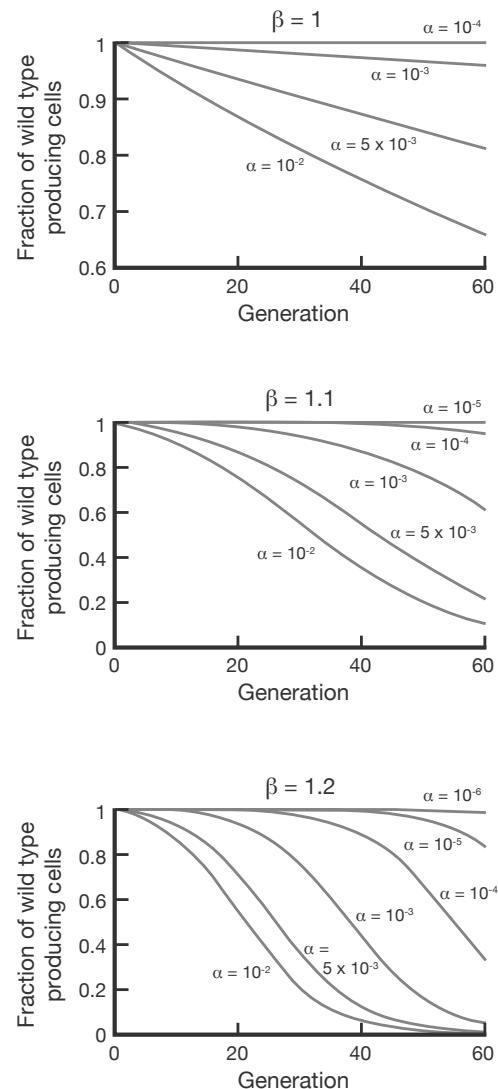


Figure 6.6. An increasing mutant sub-population through continued mutation or a faster growth rate. α is the frequency that a low-/non-producing mutant cell arises in each generation. β is the ratio of the growth rate of the mutant to the parent. At $\beta = 1$, mutants arise only from mutation. At $\beta = 1.1$ and 1.2 , mutant cells grow 10% and 20% faster, respectively.

frequency of structural variation in the host cell and which is distal to heterochromatin regions. Hence, an advantage of targeted integration is a high degree of confidence that the resulting production cell line will be stable.

Streamlined Cell Line Development

Cell line development is a long process that is repeated for each new drug candidate. Shortening the timeline of cell line development will facilitate new drug development (Figure 6.7, Panel 6.21). A bottleneck of cell line development is the single-cell cloning step. Tens of thousands of producing cells can be generated after the transfection of the GOI. The single-cell cloning step reverts the cell number back to 1, adding at least 25 cell doublings, or more than three weeks to the timeline.

Single-cell cloning may increase the risk of generating a population that has a skewed distribution of genomic content and cellular properties. When applied to a normal and genetically nearly uniform population, single-cell cloning serves to “purify” the population. However, cell lines used in bioprocessing are highly heterogeneous in many aspects. A cloned single cell may not give back the same cell distribution as the original cell population. For example, a mutant cell lacking both alleles of gene a ($a^{-/-}$) isolated from a mixed population of $a^{-/+}$, $a^{-/-}$, and $a^{+/+}$ will not give back genotype $a^{+/+}$ in its progeny population. In order to streamline cell line development and preserve the properties of the population, single-cell cloning should be practiced to the extent of ensuring the initial homogeneity of the production cell line, but not unnecessarily repeated afterwards.

Directed integration of the GOI into a stable and accessible locus is being increasingly adopted in cell line development. After integration, the resulting producing cells consist of three types of GOI insertion status

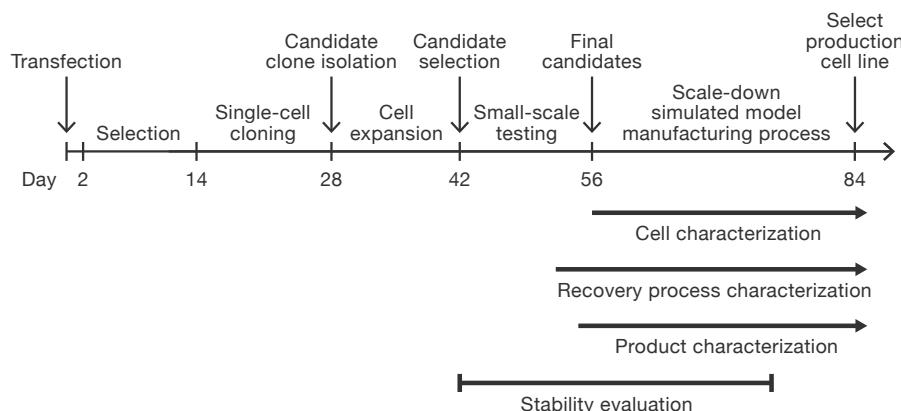


Figure 6.7. A timeline for generating an industrial producing cell line.

at the locus: $^{+/-}$, $^{-/+}$, and $^{+/+}$ (assuming the cell line has two sets of chromosomes). Note that a pair of sister chromosomes may not be identical in aneuploid cells, as one or both may be abnormal, fragmented, or translocated, and given different regional contexts. Insertion of the GOI into two alleles may give different phenotypes. Thus, directed GOI integration will likely still be followed by single-cell cloning. However, directed GOI integration may increase the frequency of obtaining stable hyperproducers.

To avoid single-cell cloning, one may attempt to integrate the GOI in all alleles in the cell. By eliminating the occurrence of off-target integration and then collecting cells of $^{+/+}$ with respect to GOI integration, one can generate a population of cells that have the same distribution of genomic contents as the host cell population except for the homogenous insertion of the GOI.

Panel 6.21. Streamlining Cell Line Development

- Single-cell cloning is a bottleneck in making cell line development more efficient
- Single-cell cloning homogenizes GOI integration sites and genomic structural variations
 - However, karyotype heterogeneity and structural variability quickly return to the single-cell cloning “purified” production cell line
 - Repeated single-cell cloning may give rise to a skewed cell genomic distribution that is different from the original early cell line
- Targeted integration may generate (GOI-integration locus) $^{+/+}$, $^{+/-}$, $^{-/+}$ cells
 - ($^{+/-}$) and ($^{-/+}$) may not be identical
 - A pool of ($^{+/+}$) with no off-site integration would have retained the distribution of the genomic structure of the host cell population. Perhaps this can accomplish what single-cell cloning is intended to achieve?

Concluding Remarks: Genomic Technology & Cell Line Development

THE PAST TWO DECADES have seen a tremendous increase in the productivity of cell culture process. Much of the productivity increase can be attributed to the generation of better-producing cell lines. Although the cell line development process remains largely empirical, nowadays industrial producing cell lines can secrete proteins at levels that rival the professional secretors in our bodies. The increased ease and versatility in performing genetic manipulations has played a major role in better cell line development. Perhaps even more important is the enhanced capability to identify and isolate high-producing clones that perform well at manufacturing scales. Automation of single-cell cloning and high-producing clone identification and isolation allows more clones to be evaluated and greatly increases the probability of isolating a high producer that is at the far high end of the distribution curve. Gene amplification, a main step in cell line development for three decades, is giving way to controlled low- or single-copy GOI for improved cell stability. The