# Protein Production in Mammalian Cells

Advanced article

Article Contents



**David L Hacker,** *École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland*

**Florian M Wurn,** *École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland*

doi: 10.1002/9780470015902.a0020209

Mammalian cells are an important host for the production of clinically relevant recombinant proteins. The most widely used approach for this purpose is to establish a cell line with an actively expressed recombinant gene stably integrated in its genome. Alternatively, protein can be transiently produced in cells for a few days immediately after deoxyribonucleic acid (DNA) transfer without the necessity of recombinant gene integration.

## Introduction

Recombinant genes were first stably transfected into cultured mammalian cells in the 1970s (Pellicer *et al*., 1978). Today, this technology is widely used not only to generate recombinant proteins for fundamental studies of their structure and function but also to produce biopharmaceuticals and vaccines for the prevention and treatment of human disease. In 1986, human tissue plasminogen activator (tPA or Activase®) for thrombolysis therapy became the first biopharmaceutical from mammalian cells to obtain regulatory approval. Now, the majority of therapeutic proteins are manufactured in mammalian cells. From 2001 to 2005 alone, 33 new biopharmaceuticals were approved with the majority of these being synthesized in Chinese hamster ovary (CHO) cells (Reichert, 2006). A significant percentage of biopharmaceuticals derived from mammalian cells are monoclonal antibodies. As of 2005, 18 therapeutic monoclonal antibodies had obtained regulatory approval and over 150 others were in clinical trials (Reichert, 2006). Several therapeutic proteins synthesized in mammalian cells, beginning with erythropoietin (Epogen®), have achieved 'blockbuster' status, a category of drugs with annual sales over $1 billion (US). **See also**: History of Biotechnology

Of the two general methods of protein production in mammalian cells, the most important one involves transfer of a recombinant gene(s) into cells followed by the covalent integration of the gene into the host's genome. Individual cells carrying the integrated recombinant gene(s) can be recovered by genetic selection and separately grown as clonal populations that are termed recombinant or stable cell lines. Since these genetically modified cells maintain the recombinant gene and its expression as protein, this process is termed stable protein production. In contrast, transient protein production does not depend on the selection of cells that have received the recombinant gene by deoxyribonucleic acid (DNA) transfer. Instead, the cells synthesize the protein for one or more days from unintegrated (episomal) plasmid DNA. Currently, the manufacture of clinically relevant recombinant proteins in mammalian cells based on stable production, and this method is also used for making small amounts of protein for research and for preclinical studies as is true for the transient production. This trend appears to be changing with the recent establishment of methods for the large-scale transient transfection of single-cell suspension cultures of 100 L or more that yield gram quantities of protein. As a result, there is increased interest in developing the transient approach for the production of proteins for clinical applications in order to shorten the timeline necessary for getting new biopharmaceuticals to patients.

There are several reasons for choosing mammalian cells rather than a microbial host for protein production, but the major one is the complexity of most biopharmaceutical proteins. For example, antibodies are glycosylated protein complexes composed of four individual proteins (two copies each of the light and heavy chains) covalently linked through disulfide bonds. Correct protein folding, complex formation and glycosylation can only be achieved with high efficiency in mammalian cells. The same is true for other protein modifications such as proteolytic processing and phosphorylation. In almost all cases, the biological activity and pharmacokinetic properties of the protein depend on correct folding and posttranslational modifications. Secondly, recombinant proteins produced in mammalian cells are relatively easy to purify since in the best cases for secreted proteins they represent up to 90% of the total protein in the culture medium. By comparison, recombinant proteins produced in bacteria are usually not correctly folded or modified and represent only a minor component in bacterial lysates prior to purification. **See also**: Protein Production for Biotechnology

## Stable Protein Production

### Cells

CHO cells were chosen as the host for the production of Activase tPA because of the availability of a genetic selection method for recovering recombinant CHO cell lines (see below). Other protein manufacturers followed suit because the approval barriers for a second product

from the same host were considerably easier to overcome. CHO cells have several advantages in that they produce no infectious endogenous retrovirus, they are not a permissive host for most human viruses, they grow well in single-cell suspension cultures, and they can be easily transfected with DNA. Although CHO cells are usually the preferred host, other cell lines such as those derived from baby hamster kidney (BHK-21), human embryo kidney (HEK-293) and mouse myeloma (NS0) have gained regulatory approval for recombinant protein production.

The original CHO cell line was isolated from primary ovarian cells following a spontaneous immortalization event. A derivative of the original cell line (strain CHO-K1) was mutagenized to generate cells lacking dihydrofolate reductase (DHFR) activity. One of the resulting cell lines named CHO-DXB11 (also referred to as CHO-DUKX or CHO-DUK-XB11) has one deleted *dhfr* allele and a missense mutation in the second allele. Subsequently, the CHO-pro3$^-$ strain, another derivative of the original CHO cell line, was mutagenized to yield CHO-DG44, a cell line with deletions of both *dhfr* alleles. These DHFR-minus cell lines require glycine, hypoxanthine and thymidine (GHT) for growth. Although not initially intended for the purpose of recombinant protein production, these cell lines were used for a number of pioneering experiments in which cells were stably transfected with an exogenous *dhfr* gene and then selected in medium lacking GHT (Ringold *et al*., 1981). As described in more detail below, this genetic selection scheme remains the standard method to establish stably transfected CHO cell lines for the production of recombinant proteins (Wurm, 2004).

## Expression vectors

In most cases, recombinant genes are transferred into mammalian cells with nonviral vectors (plasmids), but recombinant cell lines have also been established by infection with retrovirus vectors. The major advantages for using nonviral vectors include the ease of their construction, the ease of delivery into cells using inexpensive chemical agents and their biosafety. However, they also have a size limitation making it difficult to insert more than two genes into a single plasmid. This problem is easily overcome by the simultaneous transfection of multiple vectors. For both nonviral and viral vectors, the recombinant gene is typically expressed from a strong viral or cellular promoter. Examples include the immediate early promoters of human (hCMV) and mouse cytomegalovirus (mCMV), the simian virus 40 (SV40) virus early promoter, the *Rous sarcoma virus* long-terminal repeat promoter, and the promoters of constitutively expressed housekeeping genes like the human elongation factor-1 $\alpha$ (EF-1$\alpha$) and the chicken $\alpha$-actin genes (Makrides, 1999). Hybrid promoters generated by combining transcriptional control elements from two of these sources are also used.

More recently, inducible promoters have been developed for controlled gene expression in mammalian cells. This allows the protein production process to be divided into two phases; one for the exponential growth of the cells to a high density and the other for induced expression of the recombinant gene at a time when the cells are no longer dividing. This approach is especially valuable for the production of cytotoxic proteins. The major disadvantage of inducible promoters is that they do not support a comparable level of recombinant gene expression as observed for the strong constitutive promoters mentioned above.

The recombinant gene is almost always inserted into the expression vector as a complementary DNA (cDNA) rather than as a full-length gene. Since splicing of the messenger ribonucleic acid (mRNA) and its subsequent transport to the cytoplasm are functionally linked, mRNAs transcribed from intronless cDNAs are only found at low levels in the cytoplasm. Most expression vectors therefore include an intron sequence that is usually located between the promoter and the cDNA. Finally, processing and polyadenylation of the 3′ end of the recombinant transcript are controlled by a cloned polyadenylation signal at the 3′ end of the cDNA. Those acquired from the bovine growth hormone gene, the mouse β-globin gene and the SV40 virus early transcription unit are frequently used for this purpose since they are well-characterized (Makrides, 1999).

Recombinant gene expression is not only regulated at the level of transcription and mRNA processing, but the eventual yield of protein is also determined by several other factors including mRNA stability and the efficiency of translation. The latter is controlled in large part by the nucleotide sequence around the translation start site and by the secondary structure of the 5′ untranslated region (UTR) of the mRNA. Protein expression is also influenced by the sequence of the coding region of the gene. More specifically, a high percentage of G + C at the third (wobble) position of codons results in enhanced levels of both the recombinant mRNA and protein.

Protein production from recombinant cell lines may diminish over time to the point where it is no longer detectable, a phenomenon termed gene silencing. For integrated transgenes this may be due to the influence of endogenous condensed chromatin (heterochromatin) near the integration site. Heterochromatin is associated with transcriptionally inactive regions of the genome and is marked by the absence of histone acetylation. Transcriptionally active euchromatin, in contrast, is less condensed and is characterized by the presence of acetylated histones. Gene silencing not only correlates with histone deacetylation but also with other histone modifications and the methylation of cytosines within CpG dinucleotides located in the promoter and the coding region of the recombinant gene. DNA elements like scaffold or matrix attachment regions (S/MARs) and ubiquitous chromatin opening elements (UCOEs) that may function to prevent condensation of euchromatin have been inserted into expression vectors or into separate plasmids that are then cotransfected with the plasmid carrying the recombinant gene. In either case, integration of these DNA elements at the same site as the recombinant gene often increases the productivity and stability of cell lines (Wurm, 2004).

## Nonviral gene delivery

As mentioned above, recombinant mammalian cell lines are typically generated from nonviral vectors that are transferred into cells by chemical or physical methods. The three main chemical reagents for this purpose are calcium phosphate (CaPi), cationic polymers especially the polyamine polyethylenimine (PEI) and cationic liposomes. All of these reagents form complexes with negatively charged DNA, and they all face the same physical barriers to the efficient delivery of DNA into the nucleus. After binding to the cell surface, the complexes are endocytosed. A fraction of the DNA eventually escapes the endosome before digestion by nucleases. Within the cytoplasm, the DNA, unbound or within complexes, is trafficked to the nuclear membrane by an unknown mechanism(s). For all these transfection methods, it appears that very little plasmid DNA passes through nuclear pore complexes. Instead, entry of the transfected DNA is dependent on the breakdown of the nuclear membrane during mitosis. Once in the nucleus, the plasmid DNA must be accessible to the cellular components involved in DNA integration. In contrast to chemical methods, physical gene transfer methods such as microinjection and nucleofection allow transfer of naked plasmid DNA directly to the nucleus. Therefore, these methods do not face the physical barriers mentioned above. **See also**: Transfection of DNA into Mammalian Cells in Culture

## Selection of recombinant cell lines

Following DNA transfer, the cells harbouring an integrated recombinant gene must be selected from among the entire cell population. As mentioned above, the standard method is to transfer the recombinant gene along with the *dhfr* gene into a DHFR-deficient CHO cell line (Ringold *et al*., 1981). The *dhfr* gene is usually expressed from a weak promoter, either on the same plasmid as the recombinant gene or on a separate plasmid. After DNA transfer, the cells are grown in medium lacking GHT. Only cells with one or more integrated and actively expressed copies of the exogenous *dhfr* will survive under these conditions. Normally, less than 5% of the transfected cells survive selection, with the highest rates observed when the plasmid DNA is linearized prior to transfection. The surviving cells always have a single site of plasmid integration, but the plasmid copy number varies widely at each integration site (Wurm, 2004). If the *dhfr* gene and the recombinant gene are transfected on separate plasmids, both are found integrated at the same site. These observations suggest that the covalently linked multimeric complexes of transfected plasmid DNA (concatamers) are assembled prior to integration, most likely by cellular ligases.

Selected cells divide and form clonal colonies that are individually transferred to separate cultivation vessels. The clonality of each cell line is often assured by several rounds of limiting dilution, a technique whereby the cells in a presumptive clonal population are diluted and then transferred to multiwell plates such that only one cell is plated per well. The cell population arising from a single cell can be considered clonal. From each transfection, numerous clonal cell lines are recovered, but there is usually extensive heterogeneity in recombinant protein productivity, sometimes with a range exceeding two orders of magnitude, among the clones generated from a single transfection. Thus, to obtain cell lines producing the highest levels of protein, many recombinant clones must be evaluated.

Other selectable metabolic markers besides the *dhfr* gene are available for establishing stable cell lines. The glutamate synthetase (*gs*) gene, for example, is used for selection of recombinant NS0 cell lines since their endogenous GS activity is low (Barnes *et al*., 2000). Following transfection with the *gs* gene and the recombinant gene, the cells are maintained in medium lacking glutamate. Only those cells with one or more integrated copies of the functional *gs* gene survive. Although CHO cells have a higher endogenous level of GS activity than NS0 cells, it is also possible to use this selection system in these cells. Resistance genes for the antibiotics neomycin, hygromycin B, zeocin, blasticidin and puromycin represent another class of selectable markers. Following DNA transfer, the cells are maintained in medium containing the antibiotic, and those cells lacking the resistance gene are eliminated. Lastly, recombinant cell lines may be selected on the basis of the expression of a fluorescent marker such as the green fluorescent protein (GFP) whose gene is delivered along with the recombinant gene.

With both DHFR and GS selection, expression of the recombinant gene can be significantly increased by exposing the cells to drugs that block the enzymatic activity of the selection marker. DHFR and GS are inhibited by methotrexate (MTX) and methionine sulphoximine (MSX), respectively (Wurm, 2004). For CHO-derived cell lines expressing the *dhfr* gene, a majority of the cells die after 2–3 weeks of exposure to MTX while a few survive due to a sufficiently high level of DHFR. The survivors usually have a higher integrated plasmid copy number than they did prior to exposure to MTX, and their stepwise treatment with increasing concentrations of MTX may eventually result in the isolation of cells that contain dramatically increased copies of the integrated genes. CHO cell lines containing several hundred to a few thousand copies of transfected plasmid DNA have been established through this method of gene amplification. The higher gene copy number accounts for the increased level of recombinant protein productivity observed in most 'amplified' cell lines. However, there is a risk that the amplified DNA will be unstable if the MTX is withdrawn. Similar observations have been made following the exposure of recombinant NS0 cell lines to MSX (Barnes *et al*., 2000).

Following the selection, cloning and amplification of recombinant cell lines, those with the highest levels of growth and protein productivity are further characterized for the stability. As mentioned above, recombinant gene expression is not necessarily maintained at the same level over time. Gene silencing in the entire population or in a

subpopulation is a common observation. Also, cytogenetic (chromosomal) instabilities may contribute to the emergence of heterogeneity in clonal cell populations over time. Thus, the candidate cell lines must be cultivated for 2–6 months to assess the long-term stability of protein production. It is recommended that all the steps described above be performed in serum-free suspension culture if these conditions are used for the manufacture of the protein. The productivity of recombinant cell lines often decreases when the cells are adapted to serum-free suspension growth after being maintained as adherent cells in the presence of serum.

### Cell cultivation systems

Most recombinant protein production processes for clinical manufacture utilize single-cell suspension cultivation in stirred-tank bioreactors of various sizes up to 20 000 L (Wurm, 2004). The cells are maintained in media that are specifically designed for suspension growth to a high cell density, preferably in the absence of serum or other animal-derived components. The cells may be cultivated during the entire protein production run without medium additives (batch culture) or medium components may be periodically added to the culture to prolong cell viability and protein production (extended- or fed-batch culture) (Wurm, 2004). The latter strategy is used for most high-yielding processes today. **See also**: Animal Cell Culture Scale-up

The other major type of production process with single-cell suspension cultures is continuous perfusion (Wurm, 2004). In this case, the cells are maintained in a stirred-tank bioreactor, and several bioreactor volumes of fresh medium are fed into the culture each day while the same volume is withdrawn from the vessel. Perfused cultures can achieve even higher cell densities than batch or extended-batch cultures and can be maintained for many weeks, with product harvests occurring repeatedly throughout the cultivation period. The antihaemophilic factor VIII (Kogenate®) is manufactured from recombinant BHK-21 cells using continuous perfusion. This is the largest therapeutic recombinant protein (2332 amino acids) available, and its size and sensitivity to proteases contributed to the decision to produce it in a perfused system rather than in a batch process since it is rapidly degraded if it remains in the culture medium.

Recombinant proteins are also produced from cells attached to a surface rather than in suspension (Wurm, 2004). For example, Epogen is synthesized from a recombinant CHO cell line grown in roller bottles filled with medium to 10–30% of their capacity. The cells attach to the surface of the bottle whose slow rotation assures the regular wetting of the cells. A few days after the cells reach confluency on the bottle surface, the medium is removed and the protein is recovered. Cells are also grown attached to polymeric spheres (microcarriers) that are made of dextran, polyacrylamide or polystyrene. Each bead is seeded with a few cells, and the beads are maintained in suspension in stirred tank bioreactors. Eventually, after several rounds of cell division, each bead may hold several hundred cells. The disadvantage of this method is that volumetric scale-up requires the proteolytic removal of cells from the beads and then reseeding of the cells on a greater quantity of beads in a larger volume of medium.

## Transient Protein Production

Transient protein production begins with DNA transfer into suspension cultures of cells using a nonviral or a viral vector. Following delivery of the recombinant gene, the culture is maintained in the absence of genetic selection for 1–14 days for protein production. This method has been performed at a range of scales up to 100 L.

### Cells

HEK-293, the most commonly used cell line for transient protein production, was established in the 1970s by transformation of primary human embryonic kidney cells with the E1A and E1B genes from human adenovirus. In the attempt to specifically improve protein production through the promotion of episomal replication of the transfected plasmid DNA, two other HEK-293 derivatives were generated. HEK-293 T and HEK-293EBNA cells constitutively express the SV40 large T antigen and the *Epstein-Barr virus* (EBV) EBNA1 protein, respectively, to support the replication of plasmid DNA having the appropriate viral origin of replication. Although HEK-293 cells are the preferred host for transient protein production and yield the highest levels of protein, CHO cells have also been used for this purpose (Derouazi *et al*., 2004). It may be advantageous to use CHO cells when it is anticipated that the same protein will eventually be produced in a recombinant CHO cell line to assure the maintenance of the same characteristics, especially with regard to glycosylation.

### Expression vectors

Although the same nonviral vectors may be used for both transient and stable protein production, there are some significant changes that can be made to improve vectors for transient production. For example, it is not necessary to maintain a selectable gene on the plasmid since selection is not required. The elimination of this gene allows the size of the vector to be reduced. Alternatively, a second recombinant gene may be inserted in the vector. Secondly, a viral origin of replication that supports the replication of the plasmid in cells expressing the appropriate viral protein may be added.

A few viral vectors have been designed for applications in transient protein production, and they can efficiently deliver genes to a broad range of cell lines, a significant factor if the host cell for protein production is not HEK-293 or CHO. Viral vectors pose a biosafety risk, and they are more difficult to generate than plasmids. Although many viruses have been engineered as expression vectors,

members of the alphavirus (*Semliki Forest virus* and *Sindbis virus*) and baculovirus (*Autographa californica* nuclear polyhedrosis virus) families are the only ones to a significant extent for transient protein production.

## DNA transfection

CaPi remains one of the most efficient DNA delivery agents, facilitating the transfer of plasmid DNA to approximately 80–95% of transfected HEK-293 cells under some conditions (Meissner *et al*., 2001). Although this method is simple and cost-effective, efficient DNA transfer into cells is dependent on the presence of serum in the medium (Jordan and Wurm, 2004), and this may complicate the purification of secreted proteins. Another disadvantage of CaPi-mediated transfection is the time-sensitive formation of the precipitate, making the transfection of large volumes technically challenging (Jordan and Wurm, 2004).

The most frequently used delivery agent for transient protein production is PEI (Boussif *et al*., 1995). This polymer ranges in size from 800 Da to over 750 kilodaltons (kDa) in branched and linear forms. The gene transfer efficiency of PEI increases with its molecular weight up to 25 kDa and then declines with further increases in size. The cytotoxicity of PEI also increases with molecular weight. Since they combine high gene delivery efficiency with low cytotoxicity, the linear and branched forms of 25 kDa PEI are the ones most often used for transfection. Generally, PEI-mediated transfection is performed in the absence of serum, giving this method a major advantage over CaPi. However, PEI is not biodegradable, meaning any PEI remaining in the culture medium after transfection may interfere with protein purification.

Cationic lipids are not extensively used for transient protein production due to their high cost in comparison to PEI and CaPi. In addition, physical DNA transfer methods have not been successfully adapted to the large culture volumes necessary for transient protein production.

## Cell cultivation systems

Unlike the stable production processes described above, transient protein production is only performed in single-cell suspension culture. The cells are either stirred as in spinner flasks and stirred-tank bioreactors or agitated (Meissner *et al*., 2001). A variety of cultivation vessels have been developed for agitated systems including shake flasks, wave-type bioreactors and plastic or glass bottles. Since recombinant proteins produced transiently are for research applications, it is possible to add animal-derived components such as serum to the medium. Nevertheless, methods for transient protein production in serum-free medium have been developed for culture volumes up to 100 L, demonstrating the feasibility of these processes for the production of recombinant proteins for clinical applications (Derouazi *et al*., 2004; Makrides, 1999.).

## References


Barnes LM, Bentley CM and Dickson AJ (2000) Advances in animal cell recombinant protein production: GS-NS0 expression system. *Cytotechnology* **32**: 109–123.

Boussif O, Lezoualc'h F, Zanta MA *et al*. (1995) A versatile vector for gene and oligonucleotide transfer into cells in culture and *in vivo*: polyethylenimine. *Proceedings of the National Academy of Sciences of the USA* **92**: 7297–7301.

Derouazi M, Girard P, Van Tilborgh F *et al*. (2004) Serum-free large-scale transient transfection of CHO cells. *Biotechnology and Bioengineering* **87**: 537–545.

Jordan M and Wurm FM (2004) Transfection of adherent and suspended cells by calcium phosphate. *Methods* **33**: 136–143.

Makrides SC (1999) Components of vectors for gene transfer and expression in mammalian cells. *Protein Expression and Purification* **17**: 183–202.

Meissner P, Pick H, Kulangara A *et al*. (2001) Transient gene expression: recombinant protein production with suspension-adapted HEK293-EBNA cells. *Biotechnology and Bioengineering* **75**: 197–203.

Pellicer A, Wigler M and Axel R (1978) The transfer and stable integration of the HSV thymidine kinase gene into mouse cells. *Cell* **14**: 133–141.

Reichert JM (2006) Trends in US approvals: new biopharmaceuticals and vaccines. *Trends in Biotechnology* **24**: 293–298.

Ringold G, Dieckmann B and Lee F (1981) Co-expression and amplification of dihydrofolate reductase cDNA and the *Escherichia coli* XGPRT gene in Chinese hamster ovary cells. *Journal of Molecular and Applied Genetics* **1**: 165–175.

Wurm FM (2004) Production of recombinant protein therapeutics in cultivated mammalian cells. *Nature Biotechnology* **22**: 1393–1398.


## Further Reading


Adams GP and Weiner LM (2005) Monoclonal antibody therapy of cancer. *Nature Biotechnology* **23**: 1147–1157.

Anderson DC and Krummen L (2002) Recombinant protein expression for therapeutic applications. *Current Opinion in Biotechnology* **13**: 117–123.

Dinnis DM and James DC (2005) Engineering mammalian cells factories for improved recombinant monoclonal antibody production: lessons from nature? *Biotechnology and Bioengineering* **91**: 180–189.

Kwaks THJ and Otte A (2006) Employing epigenetics to augment the expression of therapeutic proteins in mammalian cells. *Trends in Biotechnology* **24**: 137–142.

Lechardeur D, Verkman AS and Lukacs GL (2005) Intracellular routing of plasmid DNA during non-viral gene transfer. *Advanced Drug Delivery Reviews* **57**: 755–767.

Smales CM and James DC (eds) (2005) *Therapeutic Proteins: Methods and Protocols*. Totowa, NJ: Humana Press.

Wurm F and Bernard A (1999) Large-scale transient expression in mammalian cells for recombinant protein production. *Current Opinion in Biotechnology* **10**: 156–159.