Commentary

Pharmaceutical BIOPROCESSING

Pharm. Bioprocess.
(2013) 1(3), 243–247

Valuing the future: recent advances and future directions in cell line development

Keywords: biopharmaceuticals, cell engineering, CHO, genome, NS0, 'omics, productivity, selection, stability

Mohamed Al-Rubeai[1]* & Darrin Kuystermans[2]

[1]School of Chemical & Bioprocess Engineering & Conway Institute of Biomolecular & Biomedical Research, University College Dublin, Dublin 4, Ireland

[2]Sanford-Burnham Medical Research Institute, Medical City, Lake Nona, Orlando, FL 32827, USA

*Author for correspondence:
E-mail: m.al-rubeai@ucd.ie

Mammalian cell lines have been valuable research tools for the past few decades and have now begun to be utilized extensively in industry for biopharmaceutical production. It has become evident that the implications for using mammalian cell lines is wide but the ability to grow these cells to high cell densities and at a high sustained viability, and more importantly produce high titre of recombinant product is vital for their use as commercial cell banks. When recombinant proteins were brought into the spotlight in the 1970s they gave rise to a variety of production systems and cultivation techniques [1]. The objectives of most of these systems were to overcome the low cell densities and increase product titres being secreted from the cells either in suspension or anchored to a solid support. Over the years, the advancements made in cell cultivation and engineering has led to a 100-fold increase in volumetric productivity, mainly due to developments of fed batch and perfusion culture systems, media and process optimizations in conjunction with improved expression technologies [2]. However, despite these advances it is still a well-known fact that in comparison to bacteria the scale-up and optimization of mammalian cell cultures to meet production demands has been problematic due to the low productivity and instability of the cell lines used.

Biopharmaceutical mammalian host cell lines

One predominately used cell line for the production of biologics is the Chinese hamster ovary (CHO) cell, followed by mouse myeloma (NS0), baby hamster kidney (BHK), human embryonic kidney (HEK-293) and more recently, the human retina derived PerC6. As much as improvements in volumetric productivity are important in cell lines, specific productivity and growth characteristics also have a significant impact on the process.

The mammalian cell lines selected for manufacturing of recombinant protein should possess several characteristics, including a well-documented origin, confirming identity, purity and suitability. It must have appropriate vector utilization for optimum transcription efficiency [3]. This also links to a high capability of translating protein products' mRNAs with efficient post-processing of proteins under high rates with minimal incorrect processed polypeptides and inappropriate glycosylation.

A good cell line's productive and proliferative capacity can result in high volumetric production rates. Thus the mammalian biopharmaceutical industry has research interest directed towards the development of cell lines with high proliferation rates that can be grown to high viable densities and have high production capabilities. However, achieving cell numbers and productivities has been shown to enhance environmental stress factors through suboptimal conditions, which can trigger drastic changes in cellular physiology, resulting in an alteration in cell viability, proliferation, product quality and metabolic activity. Applying 'omic technologies (transcriptomics,

FUTURE SCIENCE

10.4155/PBP.13.27 ISSN 2048-9145

proteomics, metabolomics, glycomics and cytomics), a global investigation of gene expression, protein regulation, glycosylation profiling, flow/mass cytometry and metabolic profile data of cells under special bioprocess conditions, can provide detailed views of the physiological state of the culture during the bioreaction process. The generation of expression profiles can provide further information about genetic markers responsible for critical metabolic changes in industrial cell lines.

The use of these host cell lines has given rise to common expression technologies that have become the standard in the industry. A good example of this is the CHO DHFR, where a mutant CHO DHFR$^-$ cell line is transfected with a vector containing the target gene along with the DHFR marker gene. The selection procedure for this system occurs in the absence of glycine, thymidine and hypoxanthine while methotrexate concentration is used to amplify the expression of the protein product due to methotrexate being able to inhibit the DHFR enzyme. Another standardized expression/selection system is the CHO GS and NS0 GS, where cells are transfected with a vector containing the gene of interest in addition to the GS gene. Cells possessing the GS enzyme synthesize their own intracellular glutamine from glutamate and the ammonia group provided by asparagines. NS0 cells are preferred for this system because they have no endogenous GS activity unlike its CHO counterpart. Using glutamine-free media the GS cells are selected in the presence of methionine sulfoximine. By inhibiting the GS activity the gene of interest is amplified with gradual increase in methionine sulfoximine concentration. The GS cells require lower gene copy numbers, when compared with the DHFR system, for a similar expression level and thus give the added advantage of a shorter selection period. An additional advantage of the GS selection system is that the cell cultures produce less ammonia [4], which can negatively affect the glycosylation of the recombinant protein.

Expression vectors

The generation of a cell line producing a high level of antibodies is a labour intensive process. The expression system utilized during this process can enhance the cloning and screening procedures apart from providing optimum mRNA transcription and translation, where for example, the heavy and light chain transcription ratio of antibodies may impede secretion due to translation kinetics. The majority of expression systems rely on random transgene integration events to insert the foreign nucleic acid material into the cellular chromosome. This process has the drawback of a positional effect that can affect overall transcription efficiency, for example, if integration occurs in the heterochromatin it can result in minimal transgene expression and it is also not necessary that integration into the euchromatin will guarantee maximal transgene expression. The position effect is further supported by studies where CHO clones with amplified genes located on specific chromosomal regions have been found to be stable and more productive than other types of clones [5]. What can be concluded from these studies is that stable expression cannot be inferred by the copy number of an integrated gene since deterioration over time may occur due to repeat induced silencing. Targeting transgene integration has been made possible by several systems and one of the first used the Cre/loxP recombination system and Flp/FRT (Flp-recognition target) along with recombinase-mediated cassette exchange means it is feasible to produce cell lines which can be selected for high expression and stability of recombinant proteins. The system can be used to select, stable clones of GFP gene integration by being assessed for stability of expression. This is then followed by transfection with the gene of interest so that recombinase-mediated cassette exchange can take place and a high-producing cell line with the gene of interest can be obtained.

Cell engineering

The direct approach to cell engineering has been applied for years to enhance cell lines through the manipulation of single genes that play important roles in key metabolic and regulatory pathways but recent consideration has been given to the use of an indirect cell engineering approach, utilizing genomic and proteomic techniques and tools, to aid the discovery process of novel targets for metabolic manipulation within the host cell line [6]. Since the relationship between gene and protein abundance is not always linear, and each technique has its drawbacks, combining both techniques is particularly useful in analyzing, interpreting and predicting genotype–phenotype relationships in cell lines for more complete conclusions. Over the last few years, several 'omic studies have been established in order to determine global and specific changes in the gene and protein expression patterns of hybridoma, myeloma and CHO cell lines in response to environmental perturbation and to unravel the molecular basis of clonal variation in productivity. The research in this area has been boosted by the publication of a genomic sequence of the CHO-K1 ancestral cell line from which many CHO cell lines have been derived [7]. It is expected that the availability of the genomic sequence to provide a valuable resource for genomic research on CHO cells and promises to accelerate the

discovery and development of new biotherapies. It will also improve the procedures for metabolic and cell engineering strategies in a significant way to approach those currently available for bacteria and yeast. Today, many individual genes have been introduced to – or omitted from – industrial cell lines to endow a particular phenotype in order to improve the cellular processes. Some of these strategies of cell engineering attempted to control apoptosis, proliferation and productivity within cell lines. With a well-established genomic and proteomic technology and availability of genomic sequence, the next few years will most likely see the further integration of genomic and proteomic technological platforms to describe complex biological systems and provide us with a better understanding of the regulatory pathways that interact within the cell. At the same time, pathway engineering will enable us to redesign CHO and other cell lines to increase yield, improve stability and enhance production efficiency.

Selection of high producers

The selection of high-producing cell lines can be a costly process. It is the main bottleneck of progressing a cell line from its developmental stage into the manufacturing environment and each method used to accomplish this has its advantages and disadvantages. Some of the most popular methods have been reviewed recently by Browne and Al-Rubeai [8]. Traditional protocols can be time consuming and labor intensive by nature, relying on cell selection using limited dilution single cell cloning techniques which involves multiple screens of hundreds of wells. The reason selection can be a laborious exercise is partly due to the nature of high-producing clones in a heterogeneous population. These high-producing clones tend to be overgrown by the low or non-producers – this is mainly due to energy being diverted towards productivity instead of cell proliferation – and can still require proper screening for stability of integration of the transgene. As a result, hundreds of wells undergo multiple screens to locate the high producers that also have good growth characteristics.

Other methods include the use of semisolid substrates such as agarose. The secretion of soluble recombinant protein precipitates because antibodies to these proteins form a halo-like structure around the colony formations. This halo size can be correlated to the amount of secreted recombinant protein [9]. This method encountered many associated problems when first introduced, such as poor mammalian cell growth and difficulty in viewing the halo formation, but modifications have resolved most of these issues starting with the use of methylcellulose-based semisolid media (such as ClonaCell™-TCS) and using lower capture antibody concentrations to increase sensitivity of detection under phase microscopy conditions.

Selection throughput has greatly increased over the years due to improved technology and the development of flow cytometric protocols, however, when it comes to distinct individual cell lines, optimization may be required including the need to address sensitivity for cell lines lacking robustness [8]. By introducing automated systems, a significant process efficiency increase can be obtained for the number of clones screened, alongside a reduction in the time needed to select for suitable cellular attributes. However, these systems are still based on relatively simple screening tools and the increase in the efficiency and capacity of single cell screening does not improve upon the traditional format of cloning, outgrowth and screening where improved technology has just allowed for additional automation in the process. In a paper reviewing the works on developing high-producing cell lines, Seth *et al.* concluded that a cell's high productivity is not a process regulated by 'master controllers', instead the differences in many pathways, and is due to the accumulation of subtle complementary changes in the protein secretory pathways, including others, such as those pathways involved in the control of cell death, proliferation, energy metabolism, and redox balance [10]. These complementary changes would indicate that it is unlikely that we will discover a basic 'biomarker' for productivity.

Stability

As a continuous cell line increases its generation number, process consistency may be affected due to instabilities of growth productivity or product characteristics. Across a cell line's manufacturing window, stability is defined where no change of the recombinant product's quantity or quality is seen produced from a genetic and physiological heterogeneic cell line [101]. In order to assess quality, extensive batteries of tests are carried out on the recombinant product. The regulatory authorities consider the quality control of the end product a key assessment criterion for a bioprocess approval. The quantity of end product, or the cell line productivity, has also been studied and research indicates that stability can be determined by the location of the amplified genes themselves, with higher stability seen when these genes are located in telomere regions of the chromosome. Work has demonstrated that overexpression of human telomerase reverse transcriptase can enhance chromosomal stability in CHO cells and possibly production stability by adding hexameric repeats to the chromosome ends preventing telomeric loss.

Research has looked into finding viable predictive indicator of cell line stability and productivity. So far, the mRNA level, rather than the amplified gene copy number, has been suggested to be a good indicator of cell line stability [11]. This warrants further research into the use of mRNA levels as an indicator in different cell lines and the effect of long-term cryopreservation, which may or may not effect differing cell lines and the stability.

Integration of 'omic analysis tools

The age of 'omic tools has given rise to new avenues of research and the possibility of merging the omic disciplines can give a deeper understanding of the biological system being addressed. Before a merging of disciplines can give valid information, we must address certain technological and methodological aspects, including the variation in biological samples and the lack of standard proceedures. The reason this integration would give a more global view is because the flow of intracellular expression information does not flow in a unidirectional path from transcriptome to proteome, thus looking at the system as a whole might give the information required to decipher more of the biological networks involved. If one were to look at the mRNA level, the observations in gene expression only provides information on some of the effectors of biological function. Proteins are effectors of biological function, thus merging these two should give a better identification of most of the effectors of biological function in a pathway. The advancment of the proteomic field means that it will be possible to start merging it with transcriptomic technologies, and we are likely to see this integration gaining a lot of attention in the next few years. The possibility of new tools emerging, with a more integrated approach, that allow the study of the transcriptome and proteome from the same biological system, will help advance this technology even further. One must remember that although specific time point snapshots of both transcriptomic and proteome information provide a wealth of information, it still might not be enough to fully understand the system, since temporal control may also play a part.

It is known that proteins and mRNA transcripts have a varied life. The half-life of these components can give discrepancies in the data, which should be accounted for when analyzing profiles and integrating transcriptomic and proteomic data. To obtain a more in-depth profile, temporal integrative profiling could give us a better understanding of the regulatory pathways involved, while the bioinformatics analysis of all the data sets on one platform will help better integrate the data for interpretation. Another approach may involve using chemostat cultures whenever possible in order to further control temporal effects of batch processes, although these effects will always be present.

Dawn of designer cell lines

With the availability of whole cell genomes and the combination of 'omic technology, we can set our sights on designer cells as biopharmaceutical platforms. Metabolomic data can also reinforce media formulations for these designer cell lines, engineered for stable expression using high-throughput selection techniques. These designer cells can be tweaked for stability and optimal volumetric production aided by the discovery of RNAi pathways utilizing siRNA and miRNA and recently, the piwi-interacting RNA (piRNA) complexes (distinct from miRNA) [12] linked both to epigenetic and post-transcriptional gene silencing. This opens up the possibility to engineer cellular pathways at desired gene expression levels, to fine tune cellular functions such as recombinant protein expression, cellular growth, apoptosis, and metabolism via RNA regulation networks. These RNA regulation network tools together with the possibility to use specific genome engineering tools for targeted genome editing such as zinc finger nucleases, the transcription activator-like effector nuclease system and RNA-guided endonucleases, such as clustered regularly interspaced short palindromic repeats technology which allows for targeted gene insertion into specified locations of the genome, are a powerful toolkit to produce these designer cells that can meet the production requirements at a faster cell line development rate then current cell line development strategies utilized for large scale biopharmaceutical recombinant protein production systems.

Ultimately, the combination of 'omics RNA regulation networks, genome editing tools, media optimization of specific cell lines, and high-through put selection methods will all contribute to advancing cell line development. The future is looking bright for the next generation of engineered recombinant protein production platforms as our understanding of these complex biological systems increases, thus decreasing development time and overall costs.

Financial & competing interests disclosure

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

References

1 Kuystermans D, Al-Rubeai M. Bioreactor systems for producing antibody from mammalian cells. *Cell Engineering* 7, 25–52 (2011).

2 Wurm FM. Production of recombinant protein therapeutics in cultivated mammalian cells. *Nat. Biotechnol.* 22(11), 1393–1398 (2004).

3 Kaufman RJ. Overview of vector design for mammalian gene expression. *Mol. Biotechnol.* 16(2), 151–160 (2000).

4 Yang M and Butler M. Effects of ammonia on CHO cell growth, erythropoietin production and glycosylation. *Biotechnol. Bioeng.* 68(4), 370–380 (2000).

5 Yoshikawa T, Nakanishi F, Ogura Y *et al.* Amplified gene location in chromosomal DNA affected recombinant protein production and stability of amplified genes. *Biotechnol. Prog.* 16, 710–715 (2006).

6 Kuystermans D, Krampe B, Swiderek H, Al-Rubeai M. Using cell engineering and omic tools for the improvement of cell culture processes. *Cytotechnology* 53(1), 3–22 (2007).

7 Xu X, Nagarajan H, Lewis NE *et al.* The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. *Nat. Biotechnol.* 29(8), 735–741 (2011).

8 Browne SM, Al-Rubeai M. Selection methods for high-producing mammalian cell lines. *Trends Biotechnol.* 25, 425–432 (2009).

9 Gibson MG, Ashey L, Newell AH, Bradley C. Soft agar cloning with antibody overlay to identify Chinese hamster ovary clones secreting recombinant products. *Biotechniques* 15(4), 594–597 (1993).

10 Seth G, Charaniya S, Wlaschin KF, Hu WS. In pursuit of a super producer-alternative paths to high producing recombinant mammalian cells. *Curr. Opin. Biotechnol.* 18(6), 557–564 (2007).

11 Barnes LM, Bentley CM, Dickson AJ. Molecular definition of predictive indicators of stable protein expression in recombinant NS0 myeloma cells. *Biotechnol. Bioeng.* 85(2), 115–121 (2004).

12 Gerstl MP, Hackl M, Graf AB, Borth N, Grillari J. rediction of transcribed PIWI-interacting RNAs from CHO RNAseq data. *J. Biotechnol.* 166, 51–57 (2013).

» Website

101 Racher AJ. GS gene expression system: GS, biochemical comparability, and cell line stability. Waterside Workshop Presentation Slide (2004). http://bio.lonza.com/uploads/tx_mwaxmarketingmaterial/Lonza_PowerpointSlidesCollections_050304GS_Waterside_Workshop.pdf