Pharmaceutical BIOPROCESSING

Industry perspective on Chinese hamster ovary cell 'omics'

"Applying a systems biology approach, which employs analysis via interpretation of whole omics datasets, will be crucial to attain future achievements and the next step."

Keywords: cell line engineering • CRISPR • genomics • integration site • metabolomics • NGS • proteomics • TALEN • transcriptomics • ZFN

Holger Laux
Novartis Pharma AG, Werk Klybeck, CHBS, WKL-681.1.08A, Klybeckstrasse 141, CH-4057 Basel, Switzerland
holger.laux@novartis.com

The biopharmaceutical sector has become a significant segment of pharmaceutical industry. More than 200 biopharmaceutical products have reached market approval and more than 350 additional products are currently in clinical trials [1]. Chinese hamster ovary (CHO) cells are the main host for production of recombinant therapeutic proteins. The 2011 CHO-produced biopharmaceutical sales were nearly US$60 billion [2]. Approximately 45% of new products approved in the period from 2006 to 2010 produced in CHO cells and seven out of ten of the world's top-selling biopharmaceutical drugs are expressed in CHO cells [1]. Despite some progress in improving therapeutic protein productivity, the cellular machinery of CHO cells is still not well understood and consequently good knowledge of recombinant protein production is missing. Therefore, functional genomics to better understand and improve productivity of CHO cell lines in pharmaceutical and biotech industry is getting more and more relevant. This includes applied genomic screening as well as cell line engineering tools. Thanks to recent advances in decoding the CHO cell genome and transcriptome [3–6] novel cell line engineering technologies as gene knockout (e.g., zinc finger nucleases [ZFNs], transcription activator-like effector nucleases [TALENs], clustered regularly interspaced short palindromic repeats [CRISPRs]), RNA interference (e.g., si/shRNA) as well as overexpression or introduction of new genes can now be applied to modify CHO cell lines. The expectation is that the now available Chinese hamster genomic information will lead to better understanding of the expression of recombinant proteins in CHO cells. This includes the utilization of genomics/transcriptomic tools (e.g., next-generation sequencing [NGS], expression/comparative genomics microarray techniques), proteomic tools (e.g., mass spectrometry [MS]), as well as metabolomic techniques (e.g., nuclear magnetic resonance spectroscopy).

Production stability

An important aspect in biopharmaceutical industry is the generation of stable production cell lines/clones. These must maintain their production characteristics as, for example, productivity, cell growth and product quality over a specific time frame (usually several months). Depending of upscaling and cell-growth characteristic the cell line/clone might have to double up to 50 times without losing their characteristics, especially their high productivity. To ensure that production cell lines can be upscaled to high volume, bioreactors stability studies of several month have to be performed which are labor intensive (usually several cell lines/clones are evaluated at the

FUTURE SCIENCE part of fsg

same time). Therefore, it would be of high value for pharmaceutical industry to determine predictive gene expression and/or miRNA expression pattern, epigenetic pattern and plasmid integration site of stable versus unstable producing cells. Developing such omics-based prediction tools for productivity stability would enable the selection of mainly stable cell lines/clones and therefore significantly reduce resources and shorten development time lines of biopharmaceuticals. The ultimate goal of cell line development in industry is to obtain clonal cell lines that secrete the recombinant protein with high specific productivity (Qp) and at consistently high levels over an extended number of cell generations, allowing for scale-up and cost-efficient manufacturing. Therefore, identification of genes involved in productivity stability and subsequently performing cell line engineering (e.g., knockout/down or respectively overexpress identified genes) to generate more stable production cell lines would be even more important for biopharmaceutical industry. Moreover, productivity stability is also associated with transgene integration sites. Currently, plasmids encoding the protein of interest are mainly inserted into the cellular chromosome by random transgene-integration events leading to a variety of chromosomal integration sites and copy numbers. Expression level as well as stability of the transgene is dependent on the integration site as well as the copy number [7]. With the now available genome sequences, researchers can investigate which regions in the genome are preferable regarding high-level and stable gene expression and can target the insertion of the expression system to specific regions using, for example, meganucleases, recombinases or TALEN technologies.

Productivity

Another relevant aspect in biopharmaceutical industry is the improvement of production processes and increase productivity. Hereby different omics approaches can support a better understanding of CHO cell lines and can ultimately enable a better production process as well as generation of an improved host cell line. Due to insufficient Chinese hamster sequence information available, no Chinese hamster microarrays for expression screening were available till recently. Thanks to the now available Chinese hamster genome [3,5,6] and massive improvements of deep sequencing technologies [8], comparison of transcriptomes or even genomes (e.g., between high vs low producing clones/cell lines) is feasible at low costs and without substantial resources. In 2011 the CHO-K1 genome was published from Beijing Genomics Institute [5]. Two years later the genome of the Chinese Hamster as well as genomes of several CHO cell lines were publicly available as well [3,6]. These genomes have already resulted in commercial availability of gene expression microarrays and enabled industry and academia to apply NGS methods to detect relevant biomarkers for their processes. RNA-seq is feasible nowadays due to the available genomes to which the sequence reads can be aligned. Overall identification of genes involved in higher productivity and yields will help to improve cell lines using cell line engineering approaches, adapt bioprocess or apply high-throughput screening methods to detect clones with the desired transcriptomics profile (marker genes). Additionally, identification of genes and their promoters with desired expression dynamics during fed-batch culture is of high interest for pharmaceutical industry. Having a list of endogenous CHO promoters available with specific intrinsic dynamic activities will enable the control of the expression of transgenes. An example would be the control of apoptosis for process improvements. Most apoptotic genes are expressed at high levels only when cells are under stress conditions or at the late stage of a culture. Overexpression of antiapoptotic genes delays the decline of culture viability, but overexpression of antiapoptotic genes during the exponential phase may impact rapid proliferation. Antiapoptosis strategies such as overexpressing antiapoptotic genes or shRNAs against pro-apoptotic genes may be best carried out using dynamic promoters. Implementation of such toolboxes in biopharmaceutical industry as, for example, the use of dynamically promoters to overexpress desired genes or use cell line engineering techniques to knock down/out undesired genes will improve specific productivity aiming to have processes of more than 10 g/l available. These processes would enable batches of more than 100 kg of antibodies in 10 kl bioreactors. Especially production of therapeutic proteins requiring very large production scales (e.g., blockbusters) will benefit from increased titer due to reduction of cost of goods with savings of up to 25% of raw material costs.

NGS also enables the measurement of expression levels of miRNA. miRNA can regulate global gene expression at post-transcriptional level via mRNA cleavage or translational repression. Especially in the last 2 years numerous miRNA data were published [9,10]. Nevertheless, the knowledge about the functionality of specific miRNAs is still very limited. As soon as there is a better understanding of miRNAs, they can be used to influence, for example, proliferation, apoptosis or metabolism. This can be achieved via overexpression of beneficial miRNAs or suppression respective knockout of disadvantageous miRNAs. First attempts in this direction were already performed by Loh *et al.* [11], who detected differentially expressed miRNA between high- and low-monoclonal antibody-producing CHO cells and subsequently stable trans-

fected these miRNA resulting in higher productivity. It is likely that miRNA cell line engineering will become more relevant for industry in the next few years with increasing knowledge about their functions.

Medium optimization

Omics can also be applied for medium optimization. Over the past two decades, more than 100-fold yield improvement of titers in CHO cell culture has been observed. This improved yield is to a large extend due to the development of serum-free medium as well as the optimization of feeding strategies [12]. Nevertheless, the demands of an increasing competitive market still require CHO cells to be more productive. Especially metabolomics can support media optimization. Metabolomics analysis of the culture medium presents a suitable platform to profile extracellular metabolites. The metabolic profiles provide a read-out of cellular phenotype and can help improve bioprocessing by suggesting alterations in nutrient utilization and highlighting accumulated metabolic byproducts [13]. In general, the extent of changes in metabolite is significantly higher compared with the extent of changes in gene expression profile. To improve measurements of intracellular metabolites, special quenching and extraction methods have been developed [14]. During bioprocess optimization metabolite profiling can be used to design feeding that improves recombinant protein production and to identify metabolites correlating with growth rates [15]. Identification of crucial extracellular metabolites as well as intracellular metabolite flux for media development and apoptosis reduction were recently described [16,17]. Methods such as liquid chromatography–MS-based or nuclear magnetic resonance-spectroscopy-based metabolomics were applied. It was, for example, detected that malate accumulates in the medium of fed-batch culture [16]. Subsequent cell engineering to overexpress MDH IIresulted in almost twofold higher viable cell numbers. Similar approaches are conceivable also for other targets.

Proteomics

Also proteomics benefits from the now available Chinese hamster genome. In the past, CHO proteomic studies have relied on finding homologous peptides in orthologous organisms or using expressed sequence tags, which have limited the proteome sizes to only a few hundred proteins [18]. With the recently published Chinese hamster genome, complementary proteomic and transcriptomic studies can be applied to provide insights into more than 25,000 genes. Proteomics can be used to directly monitor thousands of CHO proteins that can play key roles in, for example, cell growth, cell death, protein processing, glycosylation and metabolism. Transcriptomic and proteomic data can together clarify key properties about, for example, differentially expressed proteins and depleted or enriched pathways. Since the correlation between the transcriptome and proteome is not always one-to-one, a combined approach is useful to better understand the cellular physiology [19]. Especially 2D gel electrophoresis combined with MS analysis is a suitable tool for identifying protein changes under specific conditions in CHO cells. Recently, differentially expressed proteins of CHO cells cultivated in medium supplemented with improved hydrolysates mixtures, yielding in higher productivity, were identified by comparative proteomics [18]. Similar approach would be possible comparing protein expression between, for example, high- and low-producing cell lines or clones. These proteomic approaches also enable the analysis of pathways associated with proteins that were enriched or depleted in the proteome such as protein processing and apoptosis. Accordingly identified proteins can be used for future cell engineering efforts aiming to alter CHO cell growth, metabolism, protein expression, glycosylation or other protein modifications for improving production of biopharmaceuticals of interest.

> "Better understanding of post-translational modification processing will enable glycoengineering and therefore the generation of Chinese hamster ovary cell lines with enhanced glycosylation capabilities."

Generica

During the last years, biosimilars have become relevant parts of pharmaceutics industry. Biosimilars are designed to be highly similar to existing branded biologics. Due to the fact that biologics cannot be exactly copied, the goal of biosimilar development is to demonstrate that the product is highly similar to the reference biologic product. Differences in any characteristics, for example, post-translational modifications (PTM), may potentially impact the efficacy of biosimilars as this increases the potential for adverse events (e.g., immunogenicity) [20]. The most common PTM in CHO-produced recombinant protein products involves the incorporation of oligosaccharides. Comparison of different CHO cell lines and/or clones using genomics/transcriptomics approaches will facilitate the development of biosimilars due to better understanding and prediction of glycopattern on recombinant proteins with the aim to be as close as possible to the glycopattern of the originator product. Better understanding of PTM processing will enable glycoengineering and therefore the generation of CHO cell lines

with enhanced glycosylation capabilities [21]. These approaches will significantly reduce development time and resources of biosimilars.

Cell line sequencing

Using the public Chinese hamster or CHO-K1 genome as a reference is beneficial for biopharmaceutical companies for whole-genome sequencing of their own proprietary CHO cell lines. This allows comparative genomic experiments to detect chromosomal rearrangements that could have a profound effect on protein expression productivity and cell line stability and enables the identification of beneficial/disadvantageous mutations/rearrangements. It also facilitates cell line engineering approaches using, for example, TALEN, ZFNs or CRISPR technology, having the target sequence available. In the same context it is beneficial to know how many copies of a specific gene exist in the genome as knockout approaches are becoming significantly more complex with higher number of gene copies. In long term the CHO genome gives also insight into viral susceptibility if, for example, key genes associated with viral entry are expressed in CHO production cell lines/clones and therefore if viral infections can contaminate cell culture processes. Also analyzing the CHO genome or transcriptome for retrovirus-like sequences that are actively transcribed is now feasible.

Conclusion & future perspective

Omics technologies, including techniques as genomics, proteomics, metabolomics and transcriptomics, can be applied to globally understand CHO metabolic as well as proteomic profiling, genome transcription, and to identify the key metabolic activity and regulation genes or elements. The application of multiple cell engineering steps in improving CHO cells to achieve superior phenotypes can now be achieved with limited efforts. Integrated multiomics approaches enable the improvement of biopharmaceutical production by combining metabolite nutrient development, high-producing cell line construction and bioreactor process optimization. Although all these approaches require extensive phenotyping and omic technologies, costs are rapidly decreasing for these techniques. These omics techniques will shed further light on the subtleties of the cellular machine. Especially NGS is a fast-moving field. For example, 'single-molecule real-time sequencing' will become more relevant in *de novo* and resequencing activities of CHO cell lines. Additionally, this technique can distinguish methylated bases from the normal A,C,G,T and therefore will make it possible to sequence full methylomes relevant for prediction and improvement of clone productivity stability. Novel cell line engineering techniques such as CRISPR will improve cell line engineering approaches such as gene disruption, gene repair or DNA insertion as promoter, gene tags of single or multiple genes. Using CRISPR technology in genome-scale screening has also the potential to identify genes for cell line engineering to improve production process. Lentiviral delivery of a genome-scale CRISPR-Cas9 knockout library can target more than 20,000 genes and can be used for screening of desired phenotypes. A similar approach can also be performed using shRNA libraries. It can be expected that such tools will be available for Chinese hamster genome within the next few years. So far systems biology approaches have also been restricted due to missing CHO genome information. Applying system biology approach, which employs analysis via interpretation of whole omics datasets, will be crucial to attain future achievements and the next step. Having an integrated view of complex omics datasets and better understand genotype-phenotype relationship, pathways, correlation of proteomics/genomics/transcriptomics/metabolomics/epigenomics datasets will enable a significant improved understanding of production cell lines, processes and product quality. This will in future result in optimized media, feeds and processes in combination with superior engineered cell lines and customized biopharmaceutical production.

Financial & competing interests disclosure

H Laux is a permanent employee of Novartis Pharma. The author has no other 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 apart from those disclosed.

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

References

1 Walsh G. Biopharmaceutical benchmarks 2010. *Nat. Biotechnol.* 28(9), 917–924 (2010).

2 Hammond S, Kaplarevic M, Borth N, Betenbaugh MJ, Lee KH. Chinese Hamster Genome Database: an online resource for the CHO community at www.CHOgenome.org. *Biotechnol. Bioeng.* 109, 1353–1356 (2012).

3 Brinkrolf K, Rupp O, Laux H *et al.* Chinese hamster genome sequenced from sorted chromosomes. *Nat. Biotechnol.* 31(8), 694–695 (2013).

4 Becker J, Hackl M, Rupp O *et al.* Unraveling the Chinese hamster ovary cell line transcriptome by next-generation sequencing. *J. Biotechnol.* 156(3), 227–235 (2011).

5 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).

6 Lewis NE, Liu X, Li Y *et al.* Genomic landscapes of Chinese hamster ovary cell lines as revealed by the *Cricetulus griseus* draft genome. *Nat. Biotechnol.* 31(8), 759–765 (2013).

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

8 Metzker ML. Sequencing technologies – the next generation. *Nat. Rev. Genet.* 11(1), 31–46 (2010).

9 Baik JY, Lee KH. miRNA expression in CHO: nature knows best. *Biotechnol. J.* 9(4), 459–460 (2014).

10 Hackl M, Jadhav V, Klanert G *et al.* Analysis of microRNA transcription and post-transcriptional processing by Dicer in the context of CHO cell proliferation. *J. Biotechnol.* doi:10.1016/j.jbiotec.2013.12.018 (2014) (Epub ahead of print).

11 Loh WP, Loo B, Zhou L *et al.* Overexpression of microRNAs enhances recombinant protein production in Chinese hamster ovary cells. *Biotechnol. J.* doi:10.1002/biot.201400050 (2014) (Epub ahead of print).

12 Hacker DL, De Jesus M, Wurm FM. 25 years of recombinant proteins from reactor-grown cells – where do we go from here? *Biotechnol. Adv.* 27(6), 1023–1027 (2009).

13 Kildegaard HF, Baycin-Hizal D, Lewis NE, Betenbaugh MJ. The emerging CHO systems biology era: harnessing the 'omics revolution for biotechnology. *Curr. Opin. Biotechnol.* 24(6), 1102–1107 (2013).

14 Sellick CA, Croxford AS, Maqsood AR *et al.* Metabolite profiling of recombinant CHO cells: designing tailored feeding regimes that enhance recombinant antibody production. *Biotechnol. Bioeng.* 108(12), 3025–3031 (2011).

15 Dietmair S, Hodson MP, Quek LE *et al.* Metabolite profiling of CHO cells with different growth characteristics. *Biotechnol. Bioeng.* 109(6), 1404–1414 (2012).

16 Chong WP, Goh LT, Reddy SG *et al.* Metabolomics profiling of extracellular metabolites in recombinant Chinese Hamster Ovary fed-batch culture. *Rapid Commun. Mass Spectrom.* 23(23), 3763–3771 (2009).

17 Goudar C, Biener R, Boisart C *et al.* Metabolic flux analysis of CHO cells in perfusion culture by metabolite balancing and 2D [^{13}C, ^{1}H] COSY NMR spectroscopy. *Metabol. Eng.* 12(2), 138–149 (2010).

18 Kim JY, Kim YG, Han YK, Choi HS, Kim YH, Lee GM. Proteomic understanding of intracellular responses of recombinant Chinese hamster ovary cells cultivated in serum-free medium supplemented with hydrolysates. *Appl. Microbiol. Biotechnol.* 89(6), 1917–1928 (2011).

19 Doolan P, Meleady P, Barron N *et al.* Microarray and proteomics expression profiling identifies several candidates, including the valosin-containing protein (VCP), involved in regulating high cellular growth rate in production CHO cell lines. *Biotechnol. Bioeng.* 106(1), 42–56 (2010).

20 Cai XY, Thomas J, Cullen C, Gouty D. Challenges of developing and validating immunogenicity assays to support comparability studies for biosimilar drug development. *Bioanalysis* 4(17), 2169–2177 (2012).

21 Zhang P, Chan KF, Haryadi R, Bardor M, Song Z. CH^2O glycosylation mutants as potential host cells to produce therapeutic proteins with enhanced efficacy. *Adv. Biochem. Eng. Biotechnol.* 131, 63–87 (2013).