Pharmaceutical BIOPROCESSING

gy

From random mutagenesis to systems biology in metabolic engineering of mammalian cells

"Ultimately, we envision these models as tools to supplement established techniques for strain development by interpreting existing and newly generated -omic data..."

Keywords: bioprocessing • Chinese hamster ovary cells • metabolism • metabolic engineering • recombinant protein • systems biology

Metabolic engineering is rapidly developing, with a continuous stream of technological developments being employed to expand the portfolio of molecules produced in cell factories. For chemical production (e.g., amino acids, biofuels, among others), metabolic engineering has progressed through three phases [1]. Initially, biological products were obtained through random mutagenesis of production strains and large screening efforts. Improved microbial strains could be isolated, but mechanisms underlying the desired phenotype were often poorly understood [2]. Diverse molecular biology techniques facilitated the second phase, in which simple, intuitive modifications were made. The third phase now employs systems biology techniques to understand the effect of modifications on all other metabolic pathways and on cell physiology. Thus, we have entered an era in which metabolic engineering aims to improve microbial strains in a reproducible fashion, using complex designs based on detailed biochemical knowledge and computational model simulations. Here, we highlight the historical progression toward using systems biology in microbial metabolic engineering and compare this to the current status of mammalian production cell line development. Finally, we discuss the unique challenges in engineering mammalian cell lines for biotherapeutic production and outline how systems biology can facilitate metabolic engineering efforts for these platforms.

The systems biology approach to metabolic engineering has been enabled by three primary advancements: whole-genome sequencing, gene editing tools and genome-scale models of cellular metabolism. The completion of the *Escherichia coli* K-12 genome sequencing effort in 1997 [3] provided a comprehensive parts list for targeted metabolic engineering and expanded the scope of our understanding of the machinery within this microbe. The further development of efficient genetic modification systems, such as the lambda Red recombination system [4], enabled the deployment of targeted metabolic engineering designs, such as the removal of competing pathways that divert flux away from the formation of a desired product. Predictions of the systemic effects of genetic modifications were enabled when the information in the sequenced genome was harnessed for the development of genome-scale models of metabolism [5]. These models contain all known biochemical reactions in a cell, thus allowing one to predict the overall impact of modifications on phenotypic traits such as growth rate and small molecule secretion.

Systems biology approaches are now important tools in microbial metabolic engineering. Yim *et al.* genetically modified *E. coli* to produce 1,4-butanediol (BDO) by introducing heterologous genes to allow

Hooman Hefzi
Department of Bioengineering, University of California, San Diego, La Jolla, CA 92093, USA

Nathan E Lewis
Author for correspondence: Department of Pediatrics, University of California, San Diego, La Jolla, CA 92093, USA
n4lewis@ucsd.edu

FUTURE SCIENCE part of fsg

for biosynthesis of the non-native molecule [6]. Further strain optimization was carried out by using a genome-scale model to identify genetic modifications that couple production of BDO to cell growth. A quadruple knockout strategy was identified by this *in silico* approach, and was implemented in conjunction with additional enzyme engineering, resulting in a host cell line that could produce high yields of BDO. Following the same principles, Curran *et al.* produced muconic acid in *Saccharomyces cerevisiae* by introducing heterologous genes, deleting *G6pd* and overexpressing *Tkt* [7]. These modifications shunted metabolic flux toward the desired product, resulting in a twofold increase in final titer. These are just two of many examples in which modeling and genome editing are being employed together in metabolic engineering in academic and industrial settings.

> "...it is important to note that the nature of the desired product (almost exclusively recombinant proteins) and the use of a mammalian production host raise their own unique set of challenges."

While similar approaches can be applied to biotherapeutic production, additional considerations must be made, and engineering efforts beyond metabolism can prove valuable. For example, a chief consideration in biotherapeutic development is that of ensuring biocompatibility and efficacy of the recombinant protein, factors that are almost inexorably tied to protein modifications, such as glycosylation. Simple therapeutic proteins such as insulin can be produced in microbial strains such as *E. coli*. However, more complex proteins require a production system that can produce 'human-like' molecules. Herculean efforts have successfully introduced humanized glycosylation into nonmammalian hosts, including the approach taken by GlycoFi, who engineered the yeast *Pichia pastoris* with a quadruple gene knockout and the introduction of 14 additional genes [8]. However, mammalian cell lines remain the dominant hosts for biotherapeutic production, in part because their existing glycosylation machinery is more similar to that found in humans. Chinese hamster ovary (CHO) cells have become the *de facto* workhorse of the industry following early successes in recombinant protein production during the late 1980s (e.g., Activase® developed by Genentech, CA, USA)), and remains in a strong position bolstered by regulatory hurdles facing other potential hosts and a strong history of bioprocess development that has been optimized for CHO.

The existing norm in generation of production CHO cell lines for biotherapeutics often relies on random integration of a construct consisting of a selection marker (e.g., *dhfr* in a *dhfr*-deficient line) along with the gene of interest, followed by amplification of the construct using a drug (e.g., methotrexate for *dhfr*). Such approaches are effective in generating high-producing cell lines, but random integration and the accumulation of uncharacterized mutations make reproducibility difficult and the elucidation of precise mechanisms behind high productivity nearly impossible. Current metabolic engineering efforts in mammalian cell lines are focused on supplementing the drug-selected strain's properties with intuitive genetic modifications to gain desirable traits. For example, buildup of malate in the medium of CHO culture was alleviated by overexpression of malate dehydrogenase [9]. Similar logic was used in decreasing lactate production by decreasing lactate dehydrogenase and increasing pyruvate dehydrogenase activities [10], but little has been done in understanding this cell line from a systemic perspective.

Fortunately, the same drivers that have allowed systems approaches in microbial metabolic engineering have recently been developed for CHO. The genomic sequence of CHO and the Chinese hamster have been published [11,12] and the accompanying genome annotations are providing a comprehensive catalog of enzymes in CHO. These resources are enabling the construction of a genome-scale metabolic network model, furthering previous efforts built upon murine models [13,14]. Novel genome editing technologies such as transcription activator-like effector nucleases (TALENs), zinc-finger nucleases (ZFNs), and the clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 system now allow targeted genetic modification of mammalian cells lines [15]. Taken together these tools should enable a more systemic approach toward engineering mammalian cell lines.

The leap toward systems biology-based engineering of CHO will be an important one and this effort comes with several opportunities. However, it is important to note that the nature of the desired product (almost exclusively recombinant proteins) and the use of a mammalian production host raise their own unique set of challenges. First, the trade-off between growth and recombinant protein production is much more significant than the trade-off for production of small molecules. This is because recombinant proteins directly drain protein synthesis (the largest constituent of cell mass) and make growth-coupled design incredibly difficult. However, systems biology models can potentially elucidate nonintuitive interventions that enhance resource allocation to protein synthesis. Second, microbial engineering often addresses stoichiometric limitations on product formation after ensuring that necessary enzymes are expressed. Biotherapeutic protein produc-

tion, however, further requires that the cell is capable of properly folding, post-translationally modifying and secreting the non-native product; each step can be a bottleneck [16]. These additional requirements can be met in large part by the development of models of the protein secretion pathway, and the use of such models to analyze -omics measurements from individual host cell lines that exhibit differences in secreted protein yield. Third, it can be difficult to produce a homogeneous product since multiple glycoforms are often produced for a given recombinant protein. Analyses of glycomic data in the context of models of glycosylation may remedy this as they provide a more mechanistic understanding of the factors leading to diversity among glycoforms. Fourth, metabolic enzymes in mammals are far less defined and studied than their microbial counterparts and cell line properties (e.g., gene expression) that vary from tissue to tissue must be accounted for. However, detailed studies of metabolic enzymes and phenotypic assessment of mutations will help remedy these challenges over time. Finally, while techniques exist to construct and analyze metabolic models for mammalian cell lines – and even account for cell line specific knowledge – the unique properties of recombinant protein production suggest that some of the exact approaches used in microbial metabolic engineering (e.g., coupling production to growth or elimination of competing pathways) may be difficult to implement. Fortunately, novel approaches complementing existing techniques targeted at mammalian systems are being developed to specifically address these challenges.

There will be certain areas in which established methods in genome-scale modeling will be particularly informative. Indeed, constraint-based modeling with genome-scale models is powerful at predicting growth and production capabilities. A plethora of *in silico* methods have been developed for use with these models and will prove useful for bioprocess optimization [17]. Additionally, methods for expanding these models to account for processes such as transcription/translation [18] or protein secretion [19] have been developed and can guide the expansion of CHO models to allow for analysis of specific areas of importance, including those discussed above. The biochemical knowledge contained within genome-scale models is also informative for understanding the mechanistic basis behind relationships observed in various -omics datasets and should shed light on potential areas for cell line improvement [20]. Ultimately, we envision these models as tools to supplement established techniques for strain development by interpreting existing and newly generated -omic data in order to pinpoint bottlenecks, metabolic or otherwise and identify actionable solutions.

Financial & competing interests disclosure

This work was completed with generous support from the Novo Nordisk Foundation that had been provided to the Center for Biosustainability at the Technical University of Denmark. The authors have 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 Park JH, Lee SY. Towards systems metabolic engineering of microorganisms for amino acid production. *Curr. Opin. Biotechnol.* 19(5), 454–460 (2008).

2 Backus MP, Stauffer JF. The production and selection of a family of strains in *Penicillium chrysogenum*. *Mycologia* 47(4), 429–463 (1955).

3 Blattner FR, Plunkett G, Bloch CA *et al.* The complete genome sequence of *Escherichia coli* K-12. *Science* 277(5331), 1453–1462 (1997).

4 Datsenko KA, Wanner BL. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl Acad. Sci. USA* 97(12), 6640–6645 (2000).

5 McCloskey D, Palsson BØ, Feist AM. Basic and applied uses of genome-scale metabolic network reconstructions of Escherichia coli. *Mol. Syst. Biol.* 9, 661 (2013).

6 Yim H, Haselbeck R, Niu W *et al.* Metabolic engineering of *Escherichia coli* for direct production of 1,4-butanediol. *Nat. Chem. Biol.* 7(7), 445–452 (2011).

7 Curran KA, Leavitt JM, Karim AS, Alper HS. Metabolic engineering of muconic acid production in *Saccharomyces cerevisiae*. *Metab. Eng.* 15, 55–66 (2013).

8 Hamilton SR, Davidson RC, Sethuraman N *et al.* Humanization of yeast to produce complex terminally sialylated glycoproteins. *Science* 313(5792), 1441–1443 (2006).

9 Chong WPK, Reddy SG, Yusufi FNK *et al.* Metabolomics-driven approach for the improvement of Chinese hamster ovary cell growth: overexpression of malate dehydrogenase II. *J. Biotechnol.* 147(2), 116–121 (2010).

10 Zhou M, Crawford Y, Ng D *et al.* Decreasing lactate level and increasing antibody production in Chinese hamster ovary cells (CHO) by reducing the expression of lactate dehydrogenase and pyruvate dehydrogenase kinases. *J. Biotechnol.* 153(1–2), 27–34 (2011).

11 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).

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

13 Selvarasu S, Ho YS, Chong WPK *et al.* Combined in silico modeling and metabolomics analysis to characterize fed-batch CHO cell culture. *Biotechnol. Bioeng.* 109(6), 1415–1429 (2012).

14 Martínez VS, Dietmair S, Quek L-E, Hodson MP, Gray P, Nielsen LK. Flux balance analysis of CHO cells before and after a metabolic switch from lactate production to consumption. *Biotechnol. Bioeng.* 110(2), 660–666 (2013).

15 Kim H, Kim J-S. A guide to genome engineering with programmable nucleases. *Nat. Rev. Genet.* 15(5), 321–334 (2014).

16 Peng R-W, Fussenegger M. Molecular engineering of exocytic vesicle traffic enhances the productivity of Chinese hamster ovary cells. *Biotechnol. Bioeng.* 102(4), 1170–1181 (2009).

17 Lewis NE, Nagarajan H, Palsson BO. Constraining the metabolic genotype–phenotype relationship using a phylogeny of *in silico* methods. *Nat. Rev. Microbiol.* 10(4), 291–305 (2012).

18 O'Brien EJ, Lerman JA, Chang RL, Hyduke DR, Palsson BØ. Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Mol. Syst. Biol.* 9, 693 (2013).

19 Feizi A, Österlund T, Petranovic D, Bordel S, Nielsen J. Genome-scale modeling of the protein secretory machinery in yeast. *PLoS ONE* 8(5), e63284 (2013).

20 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).