Toward genome-scale models of the Chinese hamster ovary cells: incentives, status and perspectives

Bioprocessing of the important Chinese hamster ovary (CHO) cell lines used for the production of biopharmaceuticals stands at the brink of several redefining events. In 2011, the field entered the genomics era, which has accelerated omics-based phenotyping of the cell lines. In this review we describe one possible application of this data: the generation of computational models for predictive and descriptive analysis of CHO cellular metabolism. We describe relevant advances in other organisms and how they can be applied to CHO cells. The immediate implications of the implementation of these methods will be accelerated development of the next generation of CHO cell lines and derived biopharmaceuticals.

Christian S Kaas[1,2], Yuzhou Fan[1,3], Dietmar Weilguny[3], Claus Kristensen[2], Helene F Kildegaard[4] & Mikael R Andersen*[,1]

[1]Department of Systems Biology, Technical University of Denmark, Denmark
[2]Mammalian Cell Technology, Biopharmaceutical Research Unit, Novo Nordisk A/S, Maaloev, Denmark
[3]Symphogen A/S, Ballerup, Denmark
[4]The Novo Nordisk Foundation Center for Biosustainability, HØrsholm, Denmark
*Author for correspondence: mr@bio.dtu.dk

It is generally appreciated that cell culture based on Chinese hamster ovary (CHO) cells holds substantial economical and medical importance. The global market for biologics was US\$99 billion in 2009, where 60–70% of the products were produced in CHO cells [1]. Over 40 biopharmaceuticals have been produced in CHO cells so far, including monoclonal antibodies, hormones, cytokines and blood-coagulation factors. It is furthermore evident that the impact of CHO cell culture will only increase in the immediate future: the US market for biologics alone has been climbing from US\$51.3 billion in 2010 to US\$63.6 billion in 2012, and expected to increase at higher rates with the US Affordable Care Act [2]. The global market for biologics is expected to rise to US\$190 billion in 2015 [3], and the percentage of CHO-derived products in approved new biologics are climbing. In 2010 and 2011 combined, 14 out of 19 approved biopharmaceuticals were derived from cell culture, the majority of these using CHO cells as hosts [4].

Despite this impact, the development of CHO cell processes – although highly successful – has been mainly driven by medium development and process engineering and to a lesser extent genomic technologies such as enhanced expression technologies for heter-ologous proteins [5]. Metabolic engineering, such as seen in microbial cell factories [6,7], has been very limited, although with some notable exceptions (for example, see [8,9]). We will argue that this has been due to the relatively late arrival of genome sequences for CHO cell lines; even though the first CHO expressed sequence tags (EST) sequences were published in 2005 [10], the first CHO genome sequences were published in 2011 [11,12], an entire decade after the first draft publication of the human genome [13], and two decades after the genome of the first eukaryote, *Saccharomyces cerevisiae* [14]. As a result, most early genome-based studies of CHO cells were performed by using the genome sequences from other mammals, for example, human, mouse or rat [15,16], which generally limits the possible experiments and interpretation of the results.

However, there is now ample genomic information available for the CHO cell lines. The CHO-K1 genome sequence [11] has been supplemented by the 2013 release of two draft genomes for the Chinese hamster (*Cricetulus griseus*) [17,18] from which the CHO cell line was originally isolated in 1957 [19]. Additionally, draft sequences for a number of CHO cell lines including the industrially relevant CHO-S and CHO DG44 have

---

Key terms

Computational framework: Modeling metabolism is typically done using linear programming, which allows optimization of fluxes to a single criterion; for example, maximum possible growth rate, which is a typical approach for microbial cultures. Alternative methods include quadratic programming, which allows optimization for two criteria. Quadratic programming methods are often used for modeling effects of gene deletions.

Cellular compartments: In a genome-scale metabolic model, cellular compartments are modeled by assigning reactions to a given compartment, and adding known and required transport reactions in and out of the compartment to the model. Predicting in which compartment a specific reaction takes place is challenging for enzymes with a low degree of characterization.

---

been published [17]. While this still leaves a few widely used cell lines, for example, the CHO DXB11 cell line [20,21], unsequenced, and a general need for improved genome quality, it is clear that CHO cell research has reached the genomic era.

One highly promising application of genomics-based research is the generation of genome-scale models of CHO cells (Figure 1).

Potential applications of metabolic models to CHO cell cultures

A genome-scale metabolic model (GSM) is a systematic correlation of the genomic information of an organism to a metabolic network, effectively reconstructing the metabolic network of the cell type in question. Such a network is most often built from available generic pathway databases (e.g., Kyoto Encyclopedia of Genes and Genomes [KEGG] [22]) and specific literature for the organism being modeled, combined with an annotated genome [23]. This underlying network is often called a genome-scale reconstruction or genome-scale metabolic network reconstruction (GENRE). Integration of the GENRE with a linear programming-based mathematical framework allows modeling of the metabolic fluxes of the cell, which is often predictive and nearly always helpful in data interpretation. The actual model and **computational framework** apply the laws of mass conservation and balances of metabolic fluxes around single metabolites to compute enzymatic rates for every single enzyme present in the model. These rates are seen as averages for the culture and are most often given as specific rates relative to a certain number of cells. Additional algorithms may be applied to predict the effect of, for example, gene deletions/insertions, perturbations of feeding rates/nutrient uptake or increased production rates [24]. Pioneering work and additional application such as integration of the protein secretion network and regulatory infor-

mation has been driven forward in *Escherichia coli* [25–27]. As CHO cells are arguably more complex in terms of gene numbers and **cellular compartments** than *E. coli*, the work associated with building a CHO GSM is more laborious and complicated, in particular in terms of assigning correct genes to enzymatic functions, and assigning enzymatic reactions to the correct compartments. However, the algorithms and uses of these models are general, and examples of potential applications from *E. coli* are equally relevant for CHO cells. In addition to this, implementation has been performed in a wide span of eukaryotic organisms as well, several of which with a complexity and quality of annotation resembling CHO cells. Examples of eukaryotic models include eukaryotic microbes, for example, industrially relevant yeasts such as *S. cerevisiae*, *Kluyveromyces lactis* and *Pichia pastoris* [28–31], filamentous fungi applied for enzyme production, for example, *Aspergillus niger* [32–34], and also higher eukaryotes such as *Arabidopsis* [35], mouse hybridoma cells [36], and human cells [37,38]. In these examples, cells, arguably as complex as CHO cells, have had their metabolism reconstructed. Cells from mouse, *Arabidopsis* and human are evidently of similar or higher complexity than the CHO cell. Even eukaryotic microbes such as filamentous fungi have a more complex growth physiology, with multicellular growth compared with the relatively homogeneous CHO cells with a more uniform growth. While the current annotation of the CHO genome is far from the quality of annotation and gene characterization found for human cells or even mouse cells [39], models can to a large part be generated by inferring function by homology to organisms with better annotation, for example, mouse or human in the case of CHO.

The primary applications of these models can be divided into at least five major categories: metabolic engineering, model-directed discovery, interpretations of phenotypes, analysis of network properties and studies of evolutionary processes [40,41]. All of these applications are highly relevant and interesting for CHO cell culture in their omics-driven approach to cellular physiology (Figure 1E).

Metabolic engineering holds considerable promise for CHO cell culture, as GSMs have the possibility of predicting the effect of gene deletions, additions and over-/under-expression. Several phenotypic traits of the CHO cells are sub-optimal for prolonged culture and protein productivity. Some examples of this are the conversion of high **glycolytic flux** to lactate, or the formation of ammonium by conversion of amino acids in the medium. Both are detrimental to cell growth and product quality [42,43]. Accordingly, these processes have been subjected to metabolic

engineering with varying success [8,9], but definitive solutions have not been found at the cell engineering level, and have to a large extent been addressed and alleviated with process and medium design (see, [44] for an example of lactate production). Model-driven engineering presents an interesting angle on these problems, by generating platform cell lines incapable of producing such by-products or producing them in highly reduced amounts. Other possibilities are found in areas addressed in microbial cell factories, for example, increasing the number of sugars available for carbon catabolism to decrease the problems with high glycolytic flux [45]. GSMs also provide attractive possibilities to model decreased by-product formation [46], generally applicable to any biotechnological production process. Another tantalizing possibility is the extension of the significant advances in CHO culture medium development [5]. One could imagine that it would be interesting to perform model-guided systemic cell line engineering to tailor the cell lines to a specific medium, or a certain feeding profile. The capabilities of GSMs to model cellular metabolism on a systemic scale allow the combination of, for example, consumption rates of medium components with model predictions. The result of this would be further improvement of platform media and processes with a tailored cell line.

Model-directed discovery has been useful in many microbial systems, in particular for improving gene annotation and functional assignment [32,33,41]. With the current state of **CHO genomics** being in its infancy, the annotation of the identified genes is very preliminary. The state of the CHO genome annotation on one hand complicates accurate model reconstructions, due to the challenges of linking genes with function, but also presents opportunities. In particular, the reconstruction of CHO metabolism (GENRE) will suggest tentative assignment for a high percentage of the metabolic genes, as is seen in other eukaryotic microbes organisms [28,32,34], where the genes involved in metabolism have not been studied as extensively as in leading model organisms, for example, yeast or *E. coli* [25,29]. However, even in *E. coli* the metabolic network has been applied to find and characterize candidate genes for a specific function [47]. This type of application of the metabolic network is particularly powerful when combined with other omics-data types such as metabolomics and proteomics (see below). Here one can, for example, integrate orphan metabolites into the metabolic network and thus improve our understanding of CHO metabolism, or identify active isoenzymes for specific pathways from protein expression data coupled with tentative metabolic enzyme networks. As this application of the model is fairly

Key terms

Glycolytic flux: In this article, glycolytic flux is defined as being the conversion rate of glucose through glycolysis.

CHO genomics: Sequencing of multiple CHO cell lines and the progenitor hamster has revealed that while individual cell line genomes have a similar number of genes as the hamster, there is a large number of structural variations, in particular insertions, deletions and single-nucleotide polymorphisms. Such variation suggests that models should be tailorized to individual cell lines.

independent of the computational predictive power, it is relevant to CHO cells no matter how accurate the models might become. The same holds true for the next application.

Interpretation of phenotypes is perhaps the most universally applicable use of the GSMs and GENREs. The network – irrespective of predictive power – provides a framework for interpretation of experimental data. For CHO cells, the calculation of metabolite consumption/production rates, growth rates and specific product formation rates has long been standard for cell culture medium and process design. However, a theoretical and computational framework for holistic interpretation of the data has not previously been available. These reconstructed networks can aid the interpretation of experimental data related to growth, as well as multiple types of omics data [41]. Models have proven important in interpretation of metabolic/flux data [48], transcriptomics [49] and proteomics [50]. The approaches are listed in an excellent recent review [51], which also covers mammalian cell types. A related example from medical research is the interpretation of proteomic and DNA microarray data from human macrophages through a reconstructed metabolic network of the cell type [38]. In this study, a holistic view of the process of activation of macrophages was achieved, and systemic activation and inactivation of parts of metabolism was identified.

In many cases, models are also capable of quite persuasive prediction of phenotypes. Especially interesting, considering the costs and time required to produce stable genetic changes in CHO cells, is prediction of phenotypes of genetic mutants [52,53].

Analysis of network properties is generally speaking a computational exercise, in which one analyzes the network structure of the GENRE to discover inherent features or emergent properties of the metabolic network. In some cases, this analysis has become mainly an arithmetic exercise, applying standard network topology algorithms, and has generated limited biological insight. However, in some cases, deep insights are found. The most easily applicable example in the context of CHO cell and process engineering is the application of elementary flux modes [54] to identify
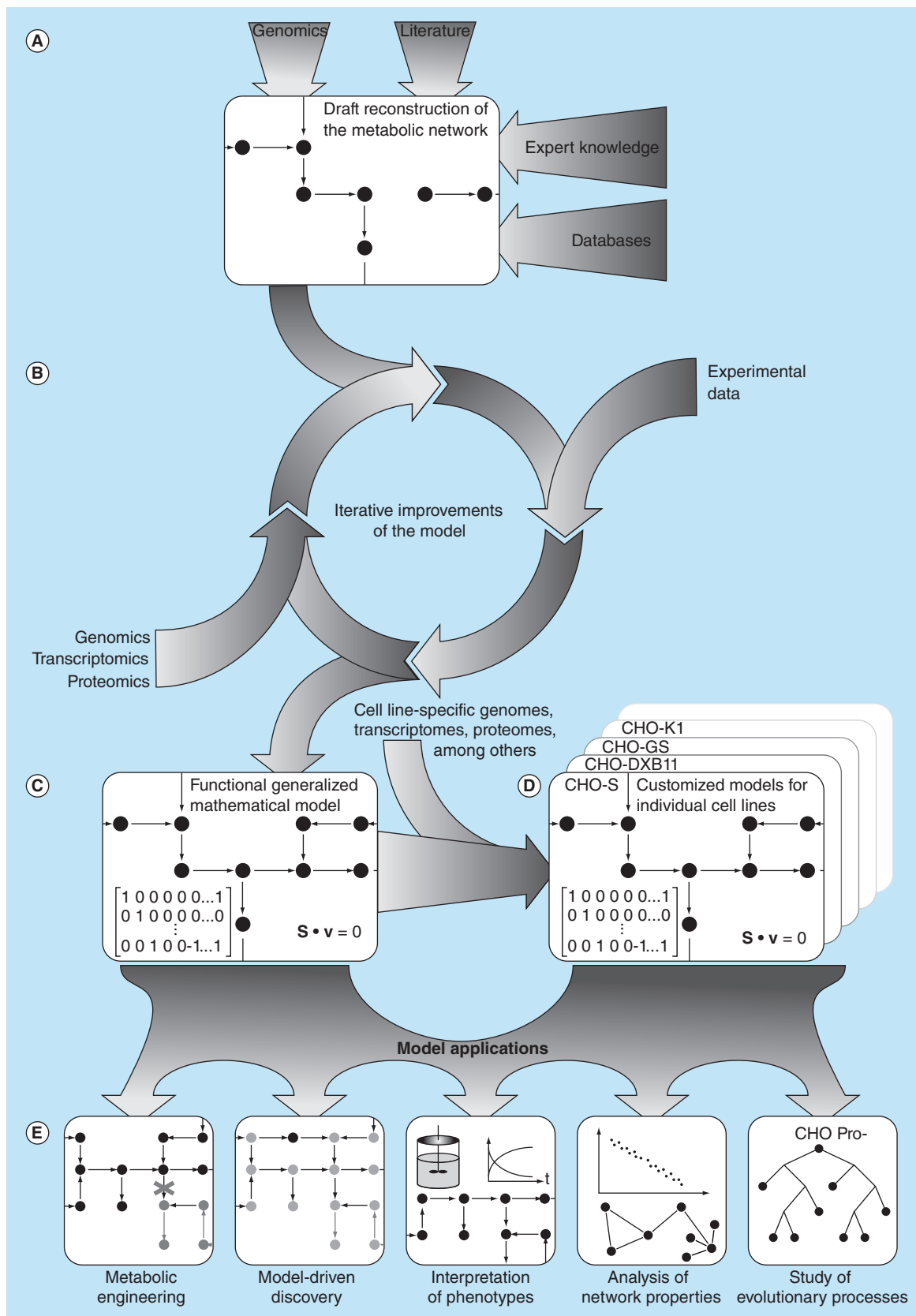
Figure 1. Model building and application in Chinese hamster ovary cells (see facing page). (A) Initial drafting of a GSM requires genomic annotation supplemented with available literature and knowledge on the metabolism of choice. This generates a draft genome-scale metabolic network reconstruction. **(B)** Multiple iterations of model improvement and testing, supplemented with available omics and phenotypic data, generate a GSM capable of computation. **(C)** A generalized GSM for CHO cells can be tailored to specific cell lines **(D)** by the inclusion of additional omics data specific for the individual cell lines. **(E)** Both generalized and specialized CHO GSMs may be applied to engineer cell lines, interpret data and increase functional understanding of these important cell lines. CHO: Chinese hamster ovary; GSM: Genome-scale metabolic model; t: Time.

the smallest possible set of essential metabolic genes in CHO cells, such as that achieved for individual pathways in *E. coli* [55] or *S. cerevisiae* [56]. Such identification could be applied for generating cell lines with a trimmed metabolism, thus decreasing the variability of the system. This would be feasible for CHO, despite the presence of extra isoenzymes or alternate pathways in many enzymatic steps. For several enzymatic functions, only one of the isoenzymes is detected at the protein level [57]. Alternatively, one can delete steps in the enzymatic pathways/elementary flux modes, where only one enzyme exists.

Studies of evolutionary processes have been a reoccurring theme in several applications of bacterial network, in particular the *E. coli* GENRE [40,41]. In such studies, specialized models have, for example, been developed to describe specific strains of *E. coli*, and compared these to identify the genetic origin of specific phenotypes [58]. Given the availability of the genomic sequence for multiple CHO cell lines with varying properties [17] and surely more to come, such an exercise would hold exciting perspectives for interpreting these genomes. One possible application would be the genetic basis for certain metabolic features in cell lines generated by mutagenesis, and the possibility of *de novo* engineering the features into a 'clean' background. This would reduce possible complications from decreased genetic stability in mutants subjected to mutagenesis [59].

With the above-mentioned being only a small percentage of the possible applications of such models for CHO cell lines, the potential is clearly large for the generation and application of GENREs and GSMs for CHO cells.

Additional available data sources for increased applicability of CHO models

The availability of an annotated genome for CHO cells is the bare minimum of information required to generate a draft model for CHO cells. Models of microbial systems have been published based mainly on literature on characterized genes (e.g., for *Corynebacterium glutamicum* [60] or *Aspergillus niger* [32,61]), but this requires detailed legacy data for a wide selection of metabolic pathways. In general, most recent generation of GSMs is based on variations of a standardized protocol for metabolic network reconstruction and model validation

published in 2010 [23], using basic genome annotation as a starting point for the organism of choice.

However, other types of omics data have proven highly valuable for model generation, validation and application. Here, the CHO field is maturing at an impressive pace, considering the quite recent publication of the first public CHO genome [11], followed by a wealth of other omics types being published in these years [62]. Here we will briefly emphasize selected studies which provide data highly applicable to CHO modeling, either due to the experimental setup of the study, the type of the data or the perspectives these offer for CHO modeling.

Genomics

A well-annotated genome with a high coverage is a crucial component in building a GSM with predictive power or a GENRE with a potential for informative data integration and interpretation. It is essential to be able to identify the genes of all major metabolic pathways in order to generate an accurate GENRE and following that a GSM. This requires a genome coverage of ideally >99% of the genes. Early EST sequencing efforts [10] identified only less than 20% of the genes, shown to be present in the CHO-K1 cell line draft genome [11]. Convenient for model construction, the genome sequence has been made accessible at the online database CHOgenome.org [63,64] as well as at the NCBI genbank. The coverage appears to be at least 99%, at least it was demonstrated that homologs existed for 99% of the genes in the human genome associated with glycosylation.

Due to the variability of the cell lines, it can be argued that it would be most appropriate to use the progenitor Chinese hamster (*C. griseus*) both as the source of a reference genome and as a scaffold for a master CHO GSM, from which specialized models can be generated for individual cell lines. This is now possible due to

---

Key term
Reference genome: Typically a genome sequence with a high (e.g., >99.5%) percentage of coverage, and a gene annotation of high quality. It is typically used to provide context and mapping of sequence from closely related genomes with a low sequence depth. This can be highly advantageous for especially large genomes, where sequence depth is expensive, or to be able to use the same set of reference genes in comparisons and data analysis.

---

the publication of draft genomes for *C. griseus* [17,18]. This publication showed that there are more than 3.7 million point mutations between the progenitor hamster and the CHO cell lines [17] and extensive chromosomal rearrangements that have occurred between CHO-K1 and CHO-DG44 [65] emphasizing the effects of the mutagenesis that occurred in the process of creating the various cell lines. Currently, in order to fully exploit these sequences, these genomes must be mapped against reference genomes with gene annotation. Improving the gene annotation for the *C. griseus* genome and CHO-K1, which has become the *de facto* reference genome for cell lines, would be beneficial to model building. Current and new genomes could thus be aligned to these references for detection of mutations.

Furthermore, the state of genome assembly should be improved. Currently the genomes for both hamster and cell lines are divided into at least 4000 contigs per genome, which means that genes for important metabolic functions may be lost in the sequencing gaps. Such gaps can to some extent be detected and fixed in the network reconstruction process [23,32]. Even so, an appropriate solution would be to apply third generation sequencing to yield longer sequenced reads that can assemble the contigs to improve the coverage of the genomes. Such efforts are in progress in the community [Borth N, Pers. Comm.], and should have a substantial positive impact on the models, which can be constructed for CHO cells.

Transcriptomics

In general, it is only a low percentage of the CHO genes which are actually expressed under normal condition, for example, only approximately 50% of the genes involved in protein glycosylation are transcriptionally active [11]. Consequently, integration of dynamic omics data such as transcriptomics and proteomics are important for accurate prediction of gene deletion/silencing effects. Several studies and tools are now available for this, including both sequencing and DNA microarray based methods. Naturally, some of the first transcriptome data were generated prior to the genome sequence based on EST sequences from CHO and mouse used for design of microarrays [66,67].

Worth particular mention is a large-scale comparison of microarray data from more than 120 individual CHO cultures [68]. The data can be accessed through the web-based CHO gene coexpression database allowing easy access to the list of genes found to coexpress with, for example, cell specific productivity and growth rate. Such data could be used for model improvement and validation. For easy and relatively inexpensive assessment of the CHO transcriptome in future experiments, a new generation of the Affymetrix® CHO DNA microarray (Affymetrix, CA, USA) has been launched with up to 26 unique sequences of each transcript with a total of more than 644,000 probes [69].

RNA-sequencing is expanding for CHO culture as in many other fields [70]. Recently, a transcriptome database for CHO RNA sequencing data has been developed and is available at GenDBE [71,72].

In summary, transcriptomics data are abundantly available, and will only increase in the coming years.

Proteomics

The CHO proteome is interesting in the context of CHO metabolic modeling as it can provide additional functional information, in some cases expanding on transcriptomic evidence. The proteome of CHO-K1 was thoroughly characterized by Baycin-Hizal in 2012 [57]. Here, 6164 proteins were detected. Of these, only 60% were also detected at the mRNA level by Xu in 2011 [11]. The functional application of the data and the need for having models specialized to individual cell lines become apparent from this study. Statistical analysis indicated that some pathways such as fatty acid metabolism, amino sugar and nucleotide sugar metabolism, which provide important precursors for recombinant protein synthesis, as well as protein processing and apoptosis, were enriched in CHO-K1 [57].

Given the principal application of CHO cells for production of secreted proteins, in this content, secretome data, such as characterized from the CHO DG44 and CHO-S cell lines by Slade [73], are interesting to incorporate. Such data can help identify secretory bottlenecks or extracellular proteases as seen for microbial cell factories [74,75].

CHO-specific protein databases have been constructed based on data from the CHO-K1 genome [11] and the CHO transcriptome [76], and have been shown to increase the number of identified proteins by 40–50% from proteomics studies compared with only using protein databases based on, for example, the murine proteome [77].

It is generally accepted that the generation of proteomics data is more technically challenging than transcriptome data, but the pilot studies within CHO cells, such as those mentioned above, show that there is clearly additional value to be gained from interrogating this data set.

Metabolomics

As mentioned in the text above, metabolomics have considerable value to add in the model building pro-

cess, as this type of data can help identify metabolic pathways, which are experimentally shown to be occurring in the cells due to the presence of metabolic intermediates or products, but the genetic basis is not necessarily clear. Being able to identify and include these pathways can increase the predictive power of the model.

For such inclusion, several studies of high-quality [78,79] and standardized protocols [79–82] have been available for several years, as metabolomics are not as such dependent on the availability of a genome.

Two studies are of particular interest in terms of getting data of a sufficient quality for model integration. The first was published by Dietmair et al. [83] correlating intracellular and extracellular metabolite concentrations with growth. The second is the work of Chong et al. [84] where intracellular metabolite profiles were obtained for eight single-cell clones with high and low production rates of monoclonal antibodies at the mid-exponential phase during shake flask batch cultures. Such studies can give insight in metabolic responses, and help validate CHO metabolic models, in that one can examine and adapt the ability of the model to predict these responses.

The application of the metabolic networks to interpret metabolomics data can also be exemplified in a 2012 study by Selvarasu et al. [85], where a generalized metabolic network of mammalian cells was adapted to CHO cells to aid in metabolomics data interpretation (see further details below). The coupling of the network, genome-scale-modeling and metabolomics data allowed the identification of growth-limiting factors.

Overall, the CHO field is at this point uniquely poised to utilize the substantial amounts of available omics data in building high-quality models for CHO cells. An overview is presented in Figure 1. During any future model-building efforts, one should draw upon the current availability of computational models for CHO and similar systems, and incorporate this where appropriate.

Overview of cellular modeling efforts in CHO cells & beyond

So far, no dedicated effort to building a CHO GSM de novo has been published. The closest example is the adaptation of a model of mouse hybridoma cells [36] to CHO cells by the addition of 35 CHO-specific metabolic reactions and subsequent model curation resulting in a model comprising 1540 reactions and 1302 metabolites [85]. This model has been further developed by other groups, although not published through a journal at this time, but is available for download from CHO.sf.net [86]. A similar approach

of adapting a mouse GSM was employed by Martínez et al. [87] for examining the energy consumption and metabolism surrounding lactate formation and consumption in CHO cells.

Dedicated models have been developed for related cell lines in other systems, as mentioned a generic model for mouse cells, applied to mouse hybridoma cell lines [36], and a model for the HEK-293 cell line has been developed as well [88]. This study is particularly promising for CHO cell modeling, as the HEK-293 model was developed by reducing the generic model for human cellular metabolism [37] to a model specific for HEK-293 metabolism. Furthermore, this model was employed to interpret both transcriptomic, metabolomic and flux data to gain functional understanding of glucose and glutamine metabolism; both key features for CHO metabolism [88]. A similar study has been seen for baby hamster kidney (BHK) cells for interpretation of metabolomics data [89].

These models listed above represent the full list of available metabolic genome-scale models with relevance to CHO cells. However, to the best of our knowledge, a model specific for CHO cells or any specific cell line has still not been generated.

One area, where modeling in CHO cells is more developed, is the kinetic modeling of protein N-glycosylation, in particular integrated with mass spectrometry on glycans. Here, very accurate predictions and substantial networks have been generated and improved over the last two decades. The first mathematical model for protein N-glycosylation process was built in 1997 by the complementary studies of a single-compartmental model [90] and a multi-compartmental model [91]. Later work expanded upon the previous work to involve glycosylation processes as galactosylation, fucoslation, sialylation and addition of N-acetyllactosamine residues [92]. This model had up to 7565 N-glycans and 22,871 reactions included. Furthermore, two glycosylation models based on different views of protein transport across the Golgi, namely Golgi maturation mechanism and vesicular transport mechanism, were studied and compared. This model was highly expanded and sophisticated by the same group to include interpretative power of N-glycan mass spectrometry data [93]. More recently, an optimized model considering 77 N-glycans, 8 enzymes, 4 nucleotide transporters and 95 reactions with individual rate expressions were built on the basis of Golgi maturation mechanism with an improvement of taking Golgi protein recycling into account [94]. On top of that, a more comprehensive glycosylation model that links a model that described the metabolism of nucleotides and nucleotide sugars

to the previous N-glycosylation model was developed by the same group [95]. These networks have been shown to have both high predictive and interpretative power, and would be unique key features to have integrated in CHO GSMs, to the extent that it is possible. Such additions could predict effects of glycosylation engineering and/or the effect of different substrate uptake rates.

Conclusion & future perspective

With the potential of GSMs tailored to CHO cells as demonstrated above, it is not surprising that several groups in the CHO community are working on building whole and partial reconstructions of CHO metabolism, including some of the authors of this review. These groups have this year formed a consortium compiling their work, and are working toward generating a community consensus model for CHO cells [Lewis ne, Pers. Comm.]. Such models and network reconstructions are known from several other research communities, including Salmonella Typhimurium [96], yeast [29,97] and human [37] metabolism.

The future arrival of the CHO GSM will probably raise the same discussion that followed the release of the first CHO genome: How well does this model describe each of the different CHO cell lines? Each of the cell lines has undergone rearrangements and has diverse transcriptomes and for this reason several parameters will need to be investigated. Future and current sequencing projects for individual cell lines should be combined with bioreactor characterization of the cell lines and their corresponding models to gain a functional understanding of the differences (Figure 1C–D). The next years will tell whether these models will be able to model the complex behavior of the CHO cell and open up new design targets such as it has been the case in microbes. Making such specialized models will be a substantial amount of work, but this task will be made easier, if a generic CHO model of high quality based on an assembled and annotated reference genome is generated first. From that, specialized models can be made in semi-automated fashion through comparative genomics. This would have the additional advantage that annotation of the genomes of the individual cell lines would not be required (as this is currently not available [17]), but could be achieved by alignment to the reference genome.

Should the models be able to deliver on the promise and potential seen in other cells, it is bound to trigger a second wave of CHO cell line engineering. Notably CRISPR-Cas9-based genome-editing systems being made available at non-cost prohibitive prices [98] and efficient high-throughput mammalian vector design systems [99] support the development of faster and cheaper genome engineering tools to accelerate future cell line engineering efforts in CHO.

In summary, the potential of genome-scale models stands to be unleashed in CHO cells within a very short time span. Combined with the genomes ushering in the genomics era for CHO, substantial amounts of omics data are being generated, and the development of efficient genetic engineering tools of CHO cell culture will soon move into the next generation of cell line development. Such advances promise better and cheaper development of biopharmaceuticals for this important group of cell factories.

Financial & competing interests disclosure

Executive summary

- Genome-scale metabolic models have been applied with success in many other prokaryotic and eukaryotic cell factories.
- The Chinese hamster ovary (CHO) field now has all of the relevant information and methods needed to construct and apply such models.
- A CHO metabolic model will have applications both in design and engineering of cells, but equally important also in interpretation of omics data. The potential is large.
- Initial CHO models have adapted from models of mouse metabolism, but no *de novo* CHO models have been published at this time.
- The community is currently constructing a consensus model for CHO metabolism.
- Added value will come from generating specialized models for individual cell lines.

References

Papers of special note have been highlighted as:
• of interest; •• of considerable interest

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

2   Aggarwal RS. What's fueling the biotech engine – 2012 to 2013. *Nat. Biotechnol.* 32(1), 32–39 (2014).

3   IMS Institute for Healthcare Informatics. The global use of medicines: outlook through 2015. www.imshealth.com/ims/Global/Content/Insights

4   Walsh G. New biopharmaceuticals. *Biopharm. Int.* 25(6), 34–36 (2012).

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

6   Bailey JE. Toward a science of metabolic engineering. *Science* 252(5013), 1668–1675 (1991).

7   Nielsen J. Metabolic engineering: techniques for analysis of targets for genetic manipulations. *Biotechnol. Bioeng.* 58(2–3), 125–132 (1998).

8   Kim SH, Lee GM. Down-regulation of lactate dehydrogenase-A by siRNAs for reduced lactic acid formation of Chinese hamster ovary cells producing thrombopoietin. *Appl. Microbiol. Biotechnol.* 74(1), 152–159 (2007).

•   **Article is a solid example of metabolic engineering of Chinese hamster ovary (CHO) cells.**

9   Zhang F, Sun X, Yi X, Zhang Y. Metabolic characteristics of recombinant Chinese hamster ovary cells expressing glutamine synthetase in presence and absence of glutamine. *Cytotechnology* 51(1), 21–28 (2006).

•• **Expression of glutamine synthetase in CHO cells is an important technology as well as an example of a type of engineering which can be predicted using genome-scale metabolic models.**

10 Wlaschin KF, Nissom PM, de Leon Gatti M *et al.* EST sequencing for gene discovery in Chinese hamster ovary cells. *Biotechnol. Bioeng.* 91(5), 592–606 (2005).

11 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).

12 Hammond S, Swanberg JC, Kaplarevic M, Lee KH. Genomic sequencing and analysis of a Chinese hamster ovary cell line using Illumina sequencing technology. *BMC Genomics* 12(1), 67 (2011).

13 Lander ES, Linton LM, Birren B *et al.* Initial sequencing and analysis of the human genome. *Nature* 409(6822), 860–921 (2001).

14 Goffeau A, Barrell BG, Bussey H *et al.* Life with 6000 genes. *Science* 274(5287), 546–567 (1996).

15 Gibbs RA, Weinstock GM, Metzker ML *et al.* Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* 428(6982), 493–521 (2004).

16 Waterston RH, Lindblad-Toh K, Birney E *et al.* Initial sequencing and comparative analysis of the mouse genome. *Nature* 420(6915), 520–562 (2002).

17 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).

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

19 Puck TT, Cieciura SJ, Robinson A. Genetics of somatic mammalian cells. III. Long-term cultivation of euploid cells from human and animal subjects. *J. Exp. Med.* 108(6), 945–956 (1958).

20 Graf LH, Chasin LA. Direct demonstration of genetic alterations at the dihydrofolate reductase locus after gamma irradiation. *Mol. Cell. Biol.* 2(1), 93–96 (1982).

21 Urlaub G, Chasin LA. Isolation of Chinese hamster cell mutants deficient in dihydrofolate reductase activity. *Proc. Natl Acad. Sci. USA* 77(7), 4216–4220 (1980).

22 Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res.* 42, D199–D205 (2014).

23 Thiele I, Palsson BØ. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat. Protoc.* 5(1), 93–121 (2010).

24 Burgard AP, Pharkya P, Maranas CD. Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol. Bioeng.* 84(6), 647–657 (2003).

25 Orth JD, Conrad TM, Na J *et al.* A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism. *Mol. Syst. Biol.* 7(535), 535 (2011).

26 Monk JM, Charusanti P, Aziz RK *et al.* Genome-scale metabolic reconstructions of multiple *Escherichia coli* strains highlight strain-specific adaptations to nutritional environments. *Proc. Natl Acad. Sci. USA* 110(50), 20338–20343 (2013).

27 Edwards JS, Palsson BO. The Escherichia coli MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities. *Proc. Natl Acad. Sci. USA* 97(10), 5528–5533 (2000).

28 Chung BKS. Genome-scale metabolic reconstruction and in silico analysis of methylotrophic yeast *Pichia pastoris* for strain improvement. *Microb. Cell Fact.* 9, 50 (2010).

29 Herrgård MJ, Swainston N, Dobson P *et al.* A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. *Nat. Biotechnol.* 26(10), 1155–1160 (2008).

30 Zomorrodi AR, Maranas CD. Improving the iMM904 *S. cerevisiae* metabolic model using essentiality and synthetic lethality data. *BMC Syst. Biol.* 4, 178 (2010).

31 Dias O, Pereira R, Gombert AK, Ferreira EC, Rocha I. iOD907, the first genome-scale metabolic model for the milk yeast *Kluyveromyces lactis*. *Biotechnol. J.* 9(6), 776–790 (2014).

32 Andersen MR, Nielsen ML, Nielsen J. Metabolic model integration of the bibliome, genome, metabolome and reactome of *Aspergillus niger*. *Mol. Syst. Biol.* 4, 178 (2008).

33 Vongsangnak W, Olsen P, Hansen K, Krogsgaard S, Nielsen J. Improved annotation through genome-scale

metabolic modeling of *Aspergillus oryzae*. *BMC Genomics* 9, 245 (2008).

34 Dreyfuss JM, Zucker JD, Hood HM, Ocasio LR, Sachs MS, Galagan JE. Reconstruction and validation of a genome-scale metabolic model for the filamentous fungus *Neurospora crassa* using FARM. *PLoS Comput. Biol.* 9(7), e1003126 (2013).

35 De Oliveira Dal'Molin CG, Quek L-E, Palfreyman RW, Brumbley SM, Nielsen LK. AraGEM, a genome-scale reconstruction of the primary metabolic network in Arabidopsis. *Plant Physiol.* 152(2), 579–589 (2010).

36 Selvarasu S, Karimi IA, Ghim G-H, Lee D-Y. Genome-scale modeling and *in silico* analysis of mouse cell metabolic network. *Mol. Biosyst.* 6(1), 152–161 (2010).

37 Thiele I, Swainston N, Fleming RMT *et al.* A community-driven global reconstruction of human metabolism. *Nat. Biotechnol.* 31(5), 419–425 (2013).

38 Bordbar A, Mo ML, Nakayasu ES *et al.* Model-driven multi-omic data analysis elucidates metabolic immunomodulators of macrophage activation. *Mol. Syst. Biol.* 8, 558 (2012).

39 Karp PD, Keseler IM, Shearer A *et al.* Multidimensional annotation of the *Escherichia coli* K-12 genome. *Nucleic Acids Res.* 35(22), 7577–7590 (2007).

40 Feist AM, Palsson BØ. The growing scope of applications of genome-scale metabolic reconstructions using *Escherichia coli*. *Nat. Biotechnol.* 26(6), 659–667 (2008).

41 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).

42 Yang M, Butler M. Effects of ammonia and glucosamine on the heterogeneity of erythropoietin glycoforms. *Biotechnol. Prog.* 18(1), 129–138 (2002).

43 Lao MS, Toth D. Effects of ammonium and lactate on growth and metabolism of a recombinant Chinese hamster ovary cell culture. *Biotechnol. Prog.* 13(5), 688–691 (1997).

44 Gagnon M, Hiller G, Luan Y-T, Kittredge A, DeFelice J, Drapeau D. High-end pH-controlled delivery of glucose effectively suppresses lactate accumulation in CHO fed-batch cultures. *Biotechnol. Bioeng.* 108(6), 1328–1337 (2011).

•   **Gagnon *et al.* demonstrates how process design in many cases can improve fed-batch performance.**

45 Richard P, Verho R, Putkonen M, Londesborough J, Penttila M. Production of ethanol from L-arabinose by containing a fungal L-arabinose pathway. *FEMS Yeast Res.* 3(2), 185–189 (2003).

46 Vemuri GN, Eiteman MA, McEwen JE, Olsson L, Nielsen J. Increasing NADH oxidation reduces overflow metabolism in *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA* 104(7), 2402–2407 (2007).

47 Fuhrer T, Chen L, Sauer U, Vitkup D. Computational prediction and experimental verification of the gene encoding the $NAD^+/NADP^+$-dependent succinate semialdehyde dehydrogenase in *Escherichia coli*. *J. Bacteriol.* 189(22), 8073–8078 (2007).

48 Choi HS, Kim TY, Lee D-Y, Lee SY. Incorporating metabolic flux ratios into constraint-based flux analysis by using

artificial metabolites and converging ratio determinants. *J. Biotechnol.* 129(4), 696–705 (2007).

49 Lee KH, Park JH, Kim TY, Kim HU, Lee SY. Systems metabolic engineering of *Escherichia coli* for L-threonine production. *Mol. Syst. Biol.* 3, 149 (2007).

50 Lewis NE, Hixson KK, Conrad TM *et al.* Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models. *Mol. Syst. Biol.* 6, 390 (2010).

51 Hyduke DR, Lewis NE, Palsson BØ. Analysis of omics data with genome-scale models of metabolism. *Mol. Biosyst.* 9(2), 167–174 (2013).

•• **Provides a solid overview of demonstrated data integration approaches for metabolic networks.**

52 Famili I, Forster J, Nielsen J, Palsson BO. *Saccharomyces cerevisiae* phenotypes can be predicted by using constraint-based analysis of a genome-scale reconstructed metabolic network. *Proc. Natl Acad. Sci. USA* 100(23), 13134–13139 (2003).

53 Fong SS, Palsson BØ. Metabolic gene-deletion strains of *Escherichia coli* evolve to computationally predicted growth phenotypes. *Nat. Genet.* 36(10), 1056–1058 (2004).

54 Schuster S. Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering. *Trends Biotechnol.* 17(2), 53–60 (1999).

55 De Figueiredo LF, Podhorski A, Rubio A *et al.* Computing the shortest elementary flux modes in genome-scale metabolic networks. *Bioinformatics* 25(23), 3158–3165 (2009).

56 Jonnalagadda S, Srinivasan R. An efficient graph theory based method to identify every minimal reaction set in a metabolic network. *BMC Syst. Biol.* 8, 28 (2014).

57 Baycin-Hizal D, Tabb DL, Chaerkady R *et al.* Proteomic analysis of Chinese hamster ovary cells. *J. Proteome Res.* 11(11), 5265–5276 (2012).

58 Baumler DJ, Peplinski RG, Reed JL, Glasner JD, Perna NT. The evolution of metabolic networks of *E. coli*. *BMC Syst. Biol.* 5, 182 (2011).

59 Morgan WF, Hartmann A, Limoli CL, Nagar S, Ponnaiya B. Bystander effects in radiation-induced genomic instability. *Mutat. Res.* 504(1–2), 91–100 (2002).

60 Kjeldsen KR, Nielsen J. In silico genome-scale reconstruction and validation of the *Corynebacterium glutamicum* metabolic network. *Biotechnol. Bioeng.* 102(2), 583–597 (2009).

61 David H, Akesson M, Nielsen J. Reconstruction of the central carbon metabolism of *Aspergillus niger*. *Eur. J. Biochem.* 270(21), 4243–4253 (2003).

62 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).

63 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(6), 1353–1356 (2012).

64   CHOgenome.org.
      www.chogenome.org/

65   Cao Y, Kimura S, Itoi T, Honda K, Ohtake H, Omasa T.
      Construction of BAC-based physical map and analysis of
      chromosome rearrangement in Chinese hamster ovary cell
      lines. *Biotechnol. Bioeng.* 109(6), 1357–1367 (2012).

66   Kantardjieff A, Nissom PM, Chuah SH *et al.* Developing
      genomic platforms for Chinese hamster ovary cells.
      Biotechnol. Adv. 27(6), 1028–1035 (2009).

67   Yee JC, de Leon Gatti M, Philp RJ, Yap M, Hu W-S.
      Genomic and proteomic exploration of CHO and hybridoma
      cells under sodium butyrate treatment. *Biotechnol.
      Bioeng.* 99(5), 1186–1204 (2008).

68   Clarke C, Henry M, Doolan P *et al.* Integrated miRNA,
      mRNA and protein expression analysis reveals the role of
      post-transcriptional regulation in controlling CHO cell
      growth rate. *BMC Genomics* 13(1), 656 (2012).

69   Affymetrix. Data Sheet GeneChip® CHO Gene 2.0 ST
      Array.
      http://media.affymetrix.com/support/technical/datasheets

70   McGettigan PA. Transcriptomics in the RNA-seq era. *Curr.
      Opin. Chem. Biol.* 17(1), 4–11 (2013).

71   Rupp O, Becker J, Brinkrolf K *et al.* Construction of a
      public CHO cell line transcript database using versatile
      bioinformatics analysis pipelines. *PLoS ONE* 9(1), e85568
      (2014).

72   GenDBE – ProCell. A eukaryotic genome browser and
      annotation system.
      https://gendbe.cebitec.uni-bielefeld.de/cho.html

73   Slade PG, Hajivandi M, Bartel CM, Gorfien SF.
      Identifying the CHO secretome using mucin-type O-linked
      glycosylation and click-chemistry. *J. Proteome Res.* 11(12),
      6175–6186 (2012).

74   Meerman HJ, Georgiou G. Construction and characterization
      of a set of *E. coli* strains deficient in all known loci affecting
      the proteolytic stability of secreted recombinant proteins.
      Biotechnology 12(11), 1107–1110 (1994).

75   Jin FJ, Watanabe T, Juvvadi PR, Maruyama J, Arioka M,
      Kitamoto K. Double disruption of the proteinase genes, tppA
      and pepE, increases the production level of human lysozyme
      by *Aspergillus oryzae*. *Appl. Microbiol. Biotechnol.* 76(5),
      1059–1068 (2007).

76   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).

77   Meleady P, Hoffrogge R, Henry M *et al.* Utilization and
      evaluation of CHO-specific sequence databases for mass
      spectrometry based proteomics. *Biotechnol. Bioeng.* 109(6),
      1386–1394 (2012).

78   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).

79   Sellick CA, Hansen R, Maqsood AR *et al.* Effective
      quenching processes for physiologically valid metabolite
      profiling of suspension cultured mammalian cells. *Anal.
      Chem.* 81(1), 174–183 (2009).

80   Sellick CA, Hansen R, Stephens GM, Goodacre R,
      Dickson AJ. Metabolite extraction from suspension-cultured
      mammalian cells for global metabolite profiling. *Nat.
      Protoc.* 6(8), 1241–1249 (2011).

81   Kronthaler J, Gstraunthaler G, Heel C. Optimizing high-
      throughput metabolomic biomarker screening: a study of
      quenching solutions to freeze intracellular metabolism in
      CHO cells. *OMICS* 16(3), 90–97 (2012).

82   Dietmair S, Timmins NE, Gray PP, Nielsen LK, Krömer JO.
      Towards quantitative metabolomics of mammalian cells:
      development of a metabolite extraction protocol. *Anal.
      Biochem.* 404(2), 155–164 (2010).

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

84   Chong WPK, 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).

85   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).

••   **Possibly the best example of the application of
      genome-scale metabolic modeling to CHO cells.**

86   CHO model download.
      http://cho.sourceforge.net/

87   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).

•   **Martínez *et al*. have a solid analysis of carbon fluxes
      through the CHO cell.**

88   Dietmair S, Hodson MP, Quek L-E, Timmins NE, Gray P,
      Nielsen LK. A multi-omics analysis of recombinant protein
      production in Hek293 cells. *PLoS ONE* 7(8), e43394 (2012).

89   Vernardis SI, Goudar CT, Klapa MI. Metabolic profiling
      reveals that time related physiological changes in mammalian
      cell perfusion cultures are bioreactor scale independent.
      Metab. Eng. 19, 1–9 (2013).

90   Monica TJ, Andersen DC, Goochee CF. A mathematical
      model of sialylation of N-linked oligosaccharides in the
      trans-Golgi network. *Glycobiology* 7(4), 515–521 (1997).

91   Umaña P, Bailey JE. A mathematical model of N-linked
      glycoform biosynthesis. *Biotechnol. Bioeng.* 55(6), 890–908
      (1997).

92   Krambeck FJ, Betenbaugh MJ. A mathematical model of
      N-linked glycosylation. *Biotechnol. Bioeng.* 92(6), 711–728
      (2005).

93   Krambeck FJ, Bennun S V, Narang S, Choi S, Yarema KJ,
      Betenbaugh MJ. A mathematical model to derive
      N-glycan structures and cellular enzyme activities from
      mass spectrometric data. *Glycobiology* 19(11), 1163–1175
      (2009).

94   Jimenez Del Val I, Nagy JM, Kontoravdi C. A dynamic
      mathematical model for monoclonal antibody N-linked

glycosylation and nucleotide sugar donor transport within a maturing Golgi apparatus. *Biotechnol. Prog.* 44(0), 1–44 (2011).

95 Jedrzejewski PM, Del Val IJ, Constantinou A *et al.* Towards controlling the glycoform: a model framework linking extracellular metabolites to antibody glycosylation. *Int. J. Mol. Sci.* 15(3), 4492–4522 (2014).

96 Thiele I, Hyduke DR, Steeb B *et al.* A community effort towards a knowledge-base and mathematical model of the human pathogen *Salmonella typhimurium* LT2. *BMC Syst. Biol.* 5, 8 (2011).

97 Heavner BD, Smallbone K, Price ND, Walker LP. Version 6 of the consensus yeast metabolic network refines biochemical coverage and improves model performance. *Database (Oxford)* 2013, bat059 (2013).

98 Ronda C, Pedersen LE, Hansen HG *et al.* Accelerating genome editing in CHO cells using CRISPR Cas9 and CRISPy, a web-based target finding tool. *Biotechnol. Bioeng.* 111(8), 1604–1616 (2014).

99 Lund AM, Kildegaard HF, Petersen MBK *et al.* A versatile system for USER cloning-based assembly of expression vectors for mammalian cell engineering. *PLoS ONE* 9(5), e96693 (2014).