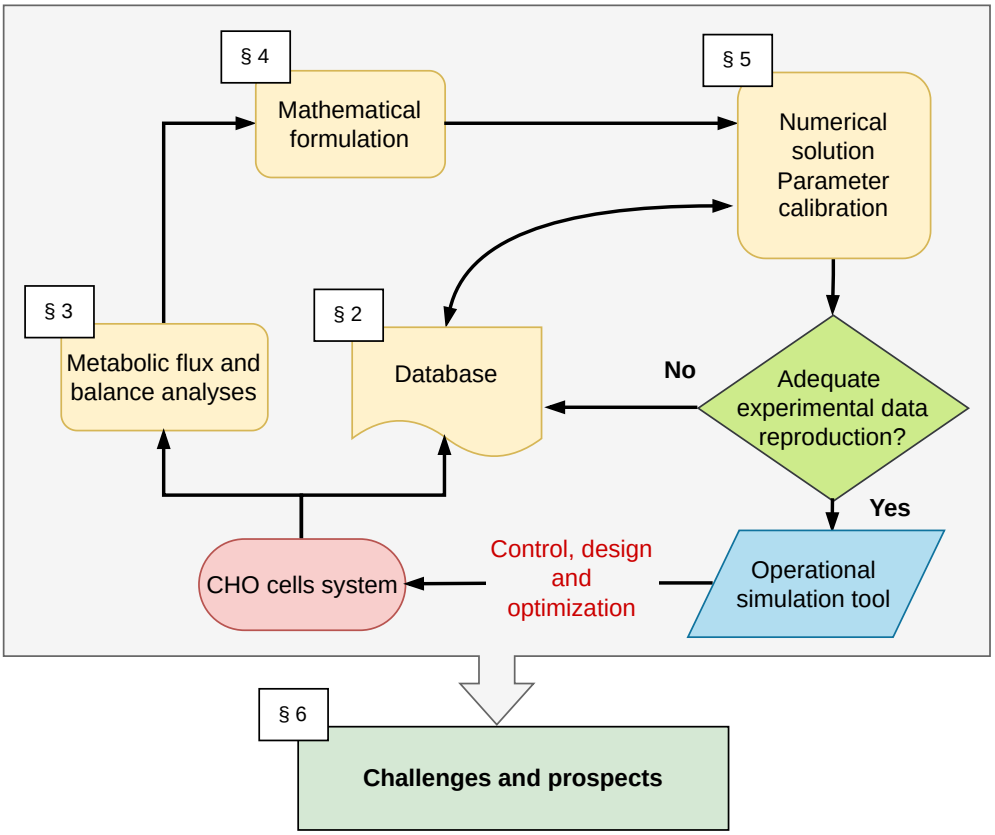
Building blocks needed for mechanistic modeling of bioprocesses: A critical review based on protein production by CHO cells

Yusmel González-Hernández, Patrick Perré

Graphical Abstract

Building Blocks Needed for Mechanistic Modeling of Bioprocesses: a Critical Review Based on Protein Production by CHO Cells

Yusmel González-Hernández, Patrick Perré



Building Blocks Needed for Mechanistic Modeling of Bioprocesses: a Critical Review Based on Protein Production by CHO Cells

Yusmel González-Hernández[a], Patrick Perré[a,]

[a] *Université Paris-Saclay, CentraleSupélec, Laboratoire de Génie des Procédés et Matériaux, Centre Européen de Biotechnologie et de Bioéconomie (CEBB), 3 rue des Rouges Terres 51110 Pomacle, France*

Abstract

This paper reviews the key building blocks needed to develop a mechanistic model for use as an operational production tool. The Chinese Hamster Ovary (CHO) cell, one of the most widely used hosts for antibody production in the pharmaceutical industry, is considered as a case study. CHO cell metabolism is characterized by two main phases, exponential growth followed by a stationary phase with strong protein production. This process presents an appropriate degree of complexity to outline the modeling strategy. The paper is organized into four main steps: (1) CHO systems and data collection; (2) metabolic analysis; (3) formulation of the mathematical model; and finally, (4) numerical solution, calibration, and validation. The overall approach can build a predictive model of target variables. According to the literature, one of the main current modeling challenges lies in understanding and predicting the spontaneous metabolic shift. Possible candidates for the trigger of the metabolic shift include the concentration of lactate and carbon dioxide. In our opinion, ammonium, which is also an inhibiting product, should be further investigated. Finally, the expected progress in the emerging field of hybrid modeling, which combines the best of mechanistic modeling and machine learning, is presented as a fascinating breakthrough. Note that the modeling strategy discussed here is a general framework that can be applied to any bioprocess.

Email address: patrick.perre@centralesupelec.fr (Patrick Perré)

Highlights

- An overview of the main metabolic pathways in CHO cells is provided,
- A simple and intuitive strategy for modeling CHO cell metabolism is outlined,
- Inverse analysis using evolutionary algorithms is a powerful tool for model calibration,
- The main open questions and challenges are summarized,
- Recommendations and medium-term prospects are proposed for further work.

Keywords: CHO cells, model calibration, metabolic shift, activation-inhibition, digital twin

---

Contents

1 Introduction **4**

2 CHO cells system **7**

2.1 CHO culture conditions 7

2.2 Gene amplification systems 10

2.2.1 DHFR-based MTX selection 10

2.2.2 GS-based MSX selection 10

2.3 Measured data 11

3 Metabolic pathway identification **13**

3.1 Metabolic flux analyses 13

3.2 Metabolic pathways description 17

3.3 Metabolic shift 18

3.4 Main acid contributors to pH 21

4 Mathematical formulation **22**

4.1 Kinetics of biological processes 23

4.2 Kinetics of glutamine metabolism 28

4.3 Kinetics of physical and chemical processes 29

4.4 Critical analysis of the published CHO mechanistic models 31

5 Numerical solution and calibration **35**

5.1 Computational solutions of ODEs 35

5.2 Calibration and validation 38

5.2.1 Parameter estimation by inverse analysis method 39

5.2.2 Calibration challenges 41

5.3 Optimization methods for minimizing the objective function 42

6 Challenges and prospects of mechanistic modeling of CHO cells **47**

6.1 Needs, concerns and improvement gaps 47

6.2 The digital twin at the crossroads of mechanistic modeling and data science 50

7 Conclusions **53**

8 Acknowledgments **54**

1. Introduction

The operation of bioreactors in bioproduction still needs to be exploited empirically, with a small amount of information collected online by sensors. However, the potential of mechanistic models as operational production tools is likely to increase efficiency, for example, through the early detection and correction of issues. In this review paper, bioproduction by Chinese Hamster Ovary (CHO) cells is used as an example to detail the key bricks needed to build a mechanistic model and how they work together. We chose this because CHO cells are the most commonly mammalian host used for therapeutic protein production in the pharmaceutical industry (Yang

et al., 2022). The metabolism of CHO cells includes an initial growth phase, with a high level of lactate production, and a subsequent stationary phase, where cell growth has slowed or stopped, and recombinant protein production is high (Dean and Reddy, 2013). This succession of phases is efficient as it separates the growth phase, where resources are used to increase the population, from the stationary phase, where the cells use resources to produce the protein of interest (Sengupta et al., 2011). Nevertheless, the factors and mechanisms that trigger this metabolic shift between the exponential and stationary phases observed in CHO cultures remain poorly understood (Hartley et al., 2018; Yahia et al., 2021). Metabolic flux analysis (MFA) and flux balance analysis (FBA) are essential in mechanistic understanding of cell metabolism for optimal production process planning and design (Huang et al., 2017; Sha et al., 2018). Despite tremendous progress, much remains to be done to understand CHO cell metabolism (Marx et al., 2022).

This production process is complex, and combining productivity, product quality, efficiency, and consistency remains challenging (Luo et al., 2021). This complexity explains why many industrial strategies for protein production by CHO cells remain mainly based on empirical results (Calmels et al., 2019). Mechanistic modeling is a powerful tool that combines knowledge from different sciences (biology, chemistry, physics, maths, etc.) that can be used for assumption testing, experimental design, and control/optimization of bioprocesses. However, before mechanistic modeling can be used as an operational tool, developing such models requires a demanding experimental effort for calibration and validation. Over-parameterized models could affect this process, which could turn them into models lacking robustness and universality (Tsopanoglou and del Val, 2021). Several models have already been developed for simulating the CHO cell metabolism (Xing et al., 2010; Nolan and Lee, 2011; López-Meza et al., 2016; Jimenez del Val et al., 2016; Kotidis et al., 2019; Yahia et al., 2021). But most models are poorly identified, and the description of

the metabolic pathways and indicators used for the prediction of metabolic shift is still inadequate. Sometimes, models are either too complex or straightforward for realistic industry use.

To further advance the field, this paper reviews the progress in the mathematical modeling of CHO cells. For clarity, the comprehensive work to be produced to obtain an efficient mechanistic model in bioprocesses is divided into a set of building blocks (Fig. 1).

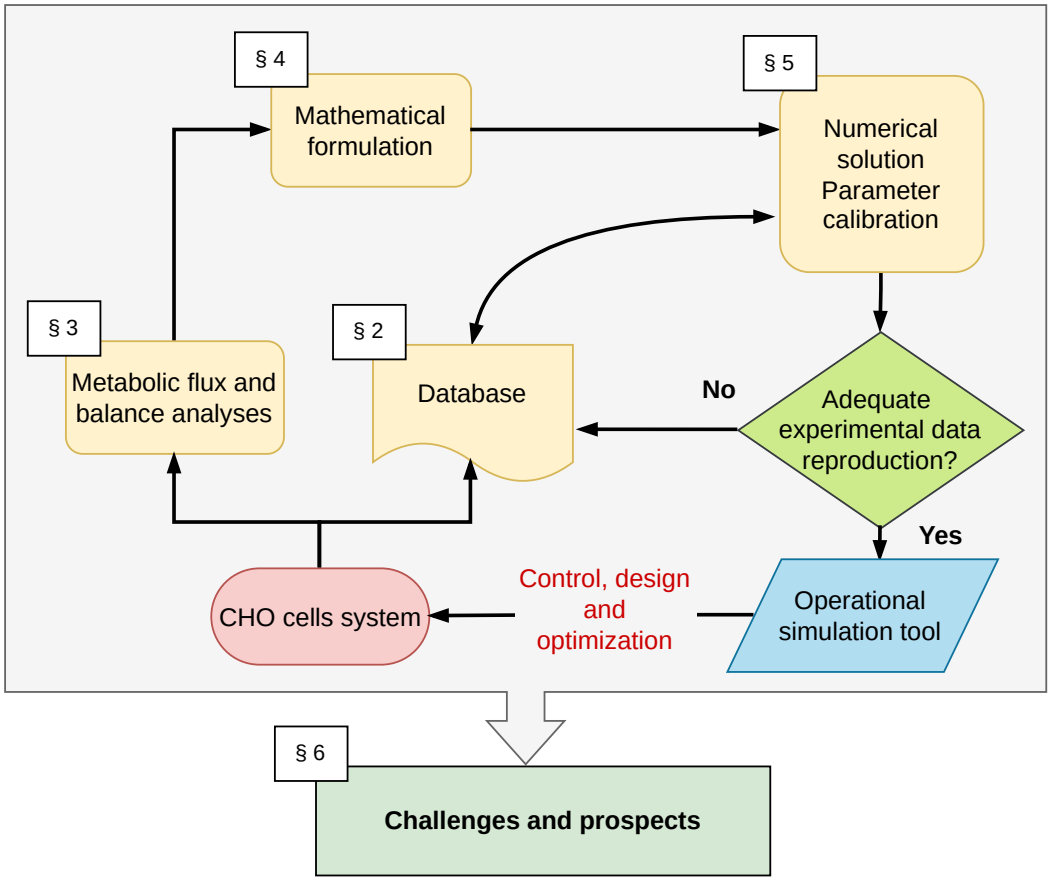


Fig. 1: Strategy for modeling CHO cell metabolism.

The paper is organized as follows. It is first necessary to understand the biological system (substrate, culture media, operating parameters, metabolic pathways, etc.). This should be done at the bioreactor level, at which cultures should be performed over a wide range of growing conditions with comprehensive instrumentation (online

and offline) to build a database (section 2). At the fundamental level, metabolic pathways need to be described, using metabolic flux analysis and flux balance analysis (section 3). The metabolic pathways are then formulated to build a system of ordinary differential equations (ODEs) (section 4). With relevant applied mathematical tools, the set of equations can finally be solved efficiently for predictive simulation (section 5). At this stage, the database of section 2 is crucial to define the model parameters by direct determination or inverse analysis. Once validated, the mechanistic model can be used as an operational control command, design, and optimization tool.

A final section 6 is devoted to the remaining challenges and prospects. After some application examples of mechanistic modeling to CHO cell systems, the main open questions and remaining challenges will be exposed. Then, as part of the answers to these challenges, we explain how the combination of mechanistic modeling and machine learning, together with database and online information, is about to change the vision of the sequential loop depicted in Fig. 1. Finally, recommendations are provided to the reader as a guide for future studies.

2. CHO cells system

2.1. CHO culture conditions

The production of recombinant proteins using CHO cells requires optimal culture media that support cell growth and productive yield, providing the necessary resources for high viable cell densities, stimulating synthesis processes, and extracellular transport of biological products (Ritacco et al., 2018). Currently, several media that provide good yields for CHO cells (high density of viable cells, cell longevity, and increased product titers) in therapeutic protein production are commercially available (Pan et al., 2017b). The optimal composition of a basal medium for protein production depends mainly on the CHO cells strain, specific characteristics of

the subclones, and the protein of interest (Rodrigues et al., 2012; Reinhart et al., 2015; Pan et al., 2017b).

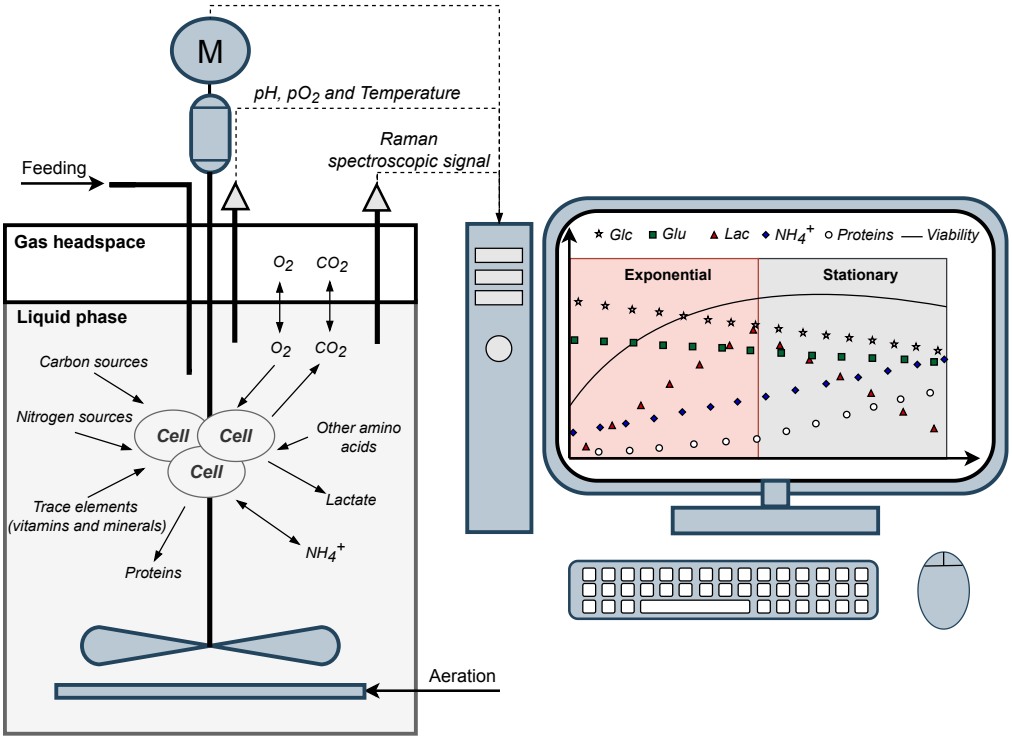


Fig. 2: Schematic representation of the bioreactor functioning during CHO cell culture (Glc: Glucose, Glu: Glutamate, and Lac: Lactate).

Whatever the CHO strain, the culture medium must be rich in carbon sources (sugar, mostly glucose), nitrogen sources (mainly glutamate or glutamine, and other amino acids, especially those that cannot be synthesized de novo by the CHO cells and known as essentials: histidine, phenylalanine, leucine, isoleucine, lysine, methionine, threonine, tryptophan, arginine, cysteine, proline, and valine (Carrillo-Cocom et al., 2015; Hefzi et al., 2016; Pan et al., 2017a,b)), and trace elements (vitamins and minerals), under aerobic conditions (Huang et al., 2017). This system is usually operated in fed-batch mode to maintain optimal nutrient concentration values and ensure maximum protein production (Fig. 2). In recent years, perfusion reactors have attracted a great deal of interest due to the advantages they offer over batch-feed systems: higher viable cell densities for extended periods of time, resulting

in increased volumetric titers and a smaller footprint (MacDonald et al., 2022). CHO cell metabolism is highly complex and divided into two fundamental phases. The first stage is exponential growth and low protein production, followed by a stationary phase, characterized by low cell growth and strong protein production using the lactate produced during the exponential phase as substrate. A final phase, the decline phase with lower production and higher mortality, could occur, but the industrial process usually stops before this decline phase. Lactate and ammonium are the most significant inhibitory by-products of CHO cell metabolism, but an excess of specific metabolites, including amino acids (e.g., phenylalanine, leucine, threonine, tryptophan, tyrosine, serine, methionine, etc.) and by-products (e.g., formate, indolelactate, homocysteine, phenylacetate, etc.), can also impact CHO cell metabolism negatively (Pereira et al., 2018).

The stationary phase can be manually induced by a temperature shift or triggered spontaneously due to the accumulation of inhibiting secondary metabolites for cell metabolism, primarily lactate and ammonium. The temperature strategy is timed to increase longevity and improve final antibody yield (Torres et al., 2018; McHugh et al., 2020).

Due to the great complexity of these systems, many studies are currently being carried out to optimize their operation by reducing the production of inhibiting secondary metabolites for cellular metabolism and increasing the yield in protein production. Note that experiments in these systems are costly, time-consuming, and prone to bacterial contamination. With recent developments in computing capabilities, mathematical modeling has become an undeniable ally in bioprocess research, enabling the validation of various hypotheses with considerable savings in resources and time.

2.2. Gene amplification systems

Stable CHO cell lines for recombinant protein production are obtained using two main gene amplification systems: dihydrofolate reductase (DHFR)-based methotrexate (MTX) selection or glutamine synthetase (GS)-based methionine sulfoximine (MSX) selection (Matasci et al., 2008; Costa et al., 2010; Fan et al., 2012; Budge et al., 2021; Yang et al., 2022).

2.2.1. DHFR-based MTX selection

The primary function of the DHFR enzyme is to catalyze the production of tetrahydrofolate from folic acid (Costa et al., 2010), a process involved in the biosynthesis of glycine, purines, and thymidylic acid (Cacciatore et al., 2010; Budge et al., 2021; Yang et al., 2022). The DHFR gene is transfected into host cells in the same gene expression vector as the protein of interest, serving as a marker to select cells transfected with the protein gene of interest in a medium deficient in glycine, purines, and thymidylic acid (Noh et al., 2013). MTX, a DHFR inhibitor, is also used to create more pressure in the selection process until only cells with an elevated gene copy number prevail (Noh et al., 2013). Although this selection system has been the most widespread, as it allows for greater efficiency in gene amplification, it requires many rounds of selection involving considerable time consumption (Noh et al., 2018).

In modeling, this amplification method should be considered when working with systems with deficient glycine, purines, and thymidylic acid. However, these nutrients are usually fulfilled in the production process, and their consumption is less extensive than other nutrients like glucose, glutamate, and glutamine.

2.2.2. GS-based MSX selection

The primary function of the enzyme glutamine synthetase is to catalyze the synthesis of glutamine from glutamate and ammonium (Cacciatore et al., 2010; Yang et al., 2022). This amplification method is adapted to cells that do not survive in glutamine-poor media. Then, the GS gene is transfected into host cells in the same

expression vector as the protein gene of interest, serving as a marker for selecting cells transfected with the protein gene of interest in the glutamine-deficient medium (Noh et al., 2013). MSX, an enzyme inhibitor, is then used to increase the selection pressure until only cells with elevated gene copy numbers prevail (Zhang et al., 2022). The GS system can achieve adequate expression levels through a single round of selection and amplification, thus significantly reducing the time required for cell line generation (Kingston et al., 2002; Noh et al., 2013). The shorter amplification time needed and its contribution to ammonium reduction by converting glutamate and ammonium to glutamine have increased its application in the pharmaceutical industry (Fan et al., 2013).

Unlike the previous amplification method, the GS-based MSX selection method has a significant impact during the production process. This method endows the cell with the ability to synthesize glutamine from glutamate, ammonium, and ATP, helping to counteract the inhibition of ammonium in CHO cell metabolism (Fan et al., 2013). In the case of glutamate limitation, the use of glutamine has a more pronounced negative effect on the cell than glutamate utilization, as it results in an increased release of ammonium (Dang, 2010). Therefore, when modeling a system in which this selection method has been applied, it is crucial to consider both the synthesis and utilization of glutamine. These processes involve extensively consumed substrates (glutamate and glutamine) and an inhibiting metabolite (ammonium), significantly impacting cell metabolism during protein production.

2.3. Measured data

In CHO systems, the experimental determination of variables such as total and viable cell number (or viability), offline pH, partial pressures of oxygen (pO_2) and carbon dioxide (pCO_2), osmolality, glucose, lactate, amino acids, ammonia, and mAbs concentrations are common (Yahia et al., 2021; Huang et al., 2017; Xing et al., 2010; Nolan and Lee, 2011, 2012). Determining these parameters during the

process enables its performance to be assessed and the operating conditions to be adjusted by manipulating parameters such as the feeding of substrates or media. In most cases, the sampling and subsequent analytical techniques used to determine the variables require considerable resources and time. This is a lock that prevents the operator from reacting in time to ensure optimum performance.

Recently, probes have been developed that allow online monitoring of most variables, as is the case with probes designed using spectroscopy. Raman spectroscopy is one of the most widely used techniques for online monitoring and control of cell cultures. It stands out for its spectra's sharpness and compatibility with aqueous systems (Li et al., 2018). Several studies have reported the use of Raman spectroscopy for online monitoring in CHO systems in the last few years (Li et al., 2018; Santos et al., 2018; Feidl et al., 2019; Yilmaz et al., 2020; W Eyster et al., 2021; Chen et al., 2021; A Gibbons et al., 2022; Domján et al., 2022; Schwarz et al., 2022; Romann et al., 2022; Yousefi-Darani et al., 2022; Yang et al., 2024). The online monitoring of key process variables facilitates the development of mathematical feedback models where operating conditions can be adjusted to obtain the maximum yield in protein production. Mechanistic modeling and Raman spectroscopy have been combined for monitoring antibody chromatographic purification by Feidl et al. (2019). This monitoring was achieved by combining data from a kinetic model and a Raman analyzer employing an extended Kalman filter. This technique proved robust, allowing accurate estimation of antibody concentrations with reduced noise. The richness of the database that will be used later for model calibration is of crucial importance, namely regarding the robustness and prediction potential. To optimize the experimental work needed to build this database, it is strongly recommended to use a design of experiment (DOE) that considers the variability of the main system variables, within and outside the operational ranges used in industry. Both fed-batch and batch tests are recommended. Notably, one part of the tests should not

be used in the learning step but kept for validation.

3. Metabolic pathway identification

In bioprocess modeling, identifying metabolic pathways is crucial as it represents the core of mechanistic modeling and its formulation. To that purpose, metabolic flux analysis (MFA) is the most widely used technique. MFA uses *in vivo* isotopic markers of metabolites (^{13}C) and modeling to quantify fluxes through the major metabolic pathways at steady state (Sengupta et al., 2011). In recent years, MFA has been extended to assess metabolic transients during fermentation, considering cellular dynamics, referred to as dynamic MFA (DMFA) (Antoniewicz, 2013; Martínez et al., 2015). Conventional and dynamic MFA approaches can be complemented by flux balance analysis (FBA), a mathematical modeling method frequently employed by metabolic engineers to quantitatively simulate steady-state genome-scale metabolic reconstructions (Kauffman et al., 2003; Gianchandani et al., 2010; Martínez et al., 2013; Ivarsson et al., 2015; Hefzi et al., 2016; Huang et al., 2017; Gutierrez et al., 2020; Schinn et al., 2021), in which, as for DMFA, dynamic conditions can also be applied (DFBA) (Antoniewicz, 2013). FBA utilizes linear programming to optimize a flux distribution towards a defined objective, considering physicochemical and thermodynamic constraints (Orth et al., 2010). While FBA demands significantly fewer experimental resources than the extensive requirements of ^{13}C-MFA, the latter is the preferred method for identifying metabolic pathways within biological systems. It offers higher precision and the ability to generate a more comprehensive flux map (Antoniewicz, 2021). Consequently, this study will emphasize the ^{13}C-MFA approach.

3.1. Metabolic flux analyses

Several recent MFA approaches have been developed to understand CHO cell metabolism. In most cases, two distinct metabolic phases can be observed. The first is the ex-

ponential phase, characterized by rapid growth and the secretion of inhibitory by-products such as lactate and ammonium. During this phase, protein production is limited. It is followed by the stationary or non-growth phase, during which protein production increases significantly and cell growth decreases. During the exponential phase, a high flux of glycolysis with considerable lactate production and a strong association of anaplerotic processes with the TCA cycle is observed. In contrast, the stationary or non-growth phase exhibits a reduced glycolysis flux, a net lactate consumption, an oxidative flux of the pentose phosphate pathway, and a reduced rate of anaplerosis (Ahn and Antoniewicz, 2011; Templeton et al., 2013).

Fig. 3 summarizes the metabolic pathways involving carbon-containing compounds that play a crucial role in catabolic and anabolic processes in CHO cells (known as central metabolism carbon) determined by using ^{13}C as reported in the literature.

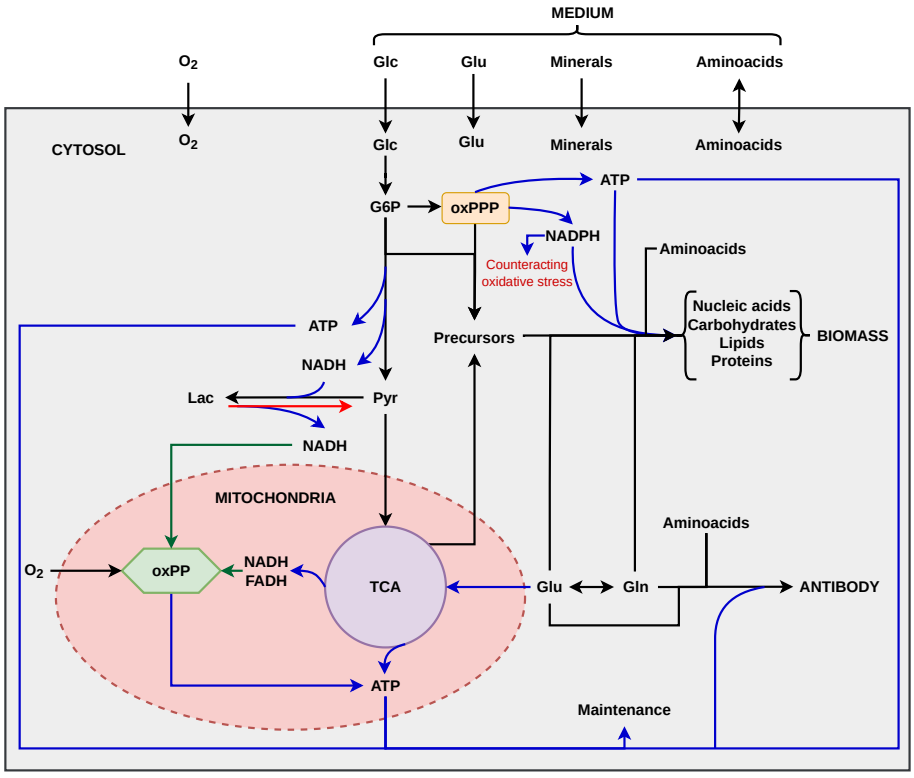


Fig. 3: Simplified schema of central metabolism carbon in CHO cells based on literature reports (Glc: Glucose, Glu: Glutamate, G6P: Glucose-6-phosphate, Pyr: Pyruvate, Lac: Lactate, Gln: Glutamine, oxPPP: Oxidative pentose phosphate pathway, TCA: Tricarboxylic acid cycle, and oxPP: Oxidative phosphorylation).

Anaplerosis is mainly observed through converting pyruvate to oxaloacetate and glutamate to α-ketoglutarate. In both phases, pyruvate dehydrogenase, and TCA cycle fluxes are similar (Ahn and Antoniewicz, 2011). During the stationary phase, an elevated glucose flux diverted to oxPPP, providing high NADPH production, is observed (Sengupta et al., 2011). This elevated NADPH production via oxPPP could be associated with macromolecule biosynthesis or a defensive process of the cell to counteract oxidative stress (Fig. 3). Sengupta et al. (2011) also found that, during the stationary phase, most of the pyruvate produced in glycolysis was metabolized in the TCA cycle with little or no lactate production. This finding aligns with Templeton et al. (2013), who developed a more detailed MFA, observing a correlation between the peak of antibody production and the highest oxidative activity in the Krebs cycle. The authors confirmed the absence of lactate production during the stationary phase until the metabolite reached low levels, restarting lactate production. Interestingly, during the stationary phase, the energy efficiency of the cells using lactate (total ATP produced per total C-mol substrate consumed) six times higher than in lactate-producing cells was reported by Martínez et al. (2013) by flux balance analysis. These results underline the importance of the exponential phase, during which rapid growth ensures a large number of cells or "protein micro-factories," simultaneously providing lactate, an efficient energy source for the stationary phase.

Martínez et al. (2015) conducted an interesting DMFA work for studying the dynamics of metabolic shifts caused by temperature changes. Their study demonstrated that inducing mild hypothermia in the system significantly reduces growth and overall metabolic rates, potentially improving the stability of recombinant protein production. These results suggest that protein production is antagonistic to the growth process. As the cell reduces its metabolic flux for cell growth due to temperature decrease, its biological activity is redirected towards protein production. This

behavior is especially relevant considering these cells have been engineered, selected, and adapted to produce recombinant proteins, behaving like malignant cells to grow indefinitely. Cell decay is also an important process that should not be neglected in these systems. Being present throughout the process is more significant during the decline phase. The specific mortality averaged over the whole process rate ranges between 0.013 and 0.107 d^{-1} (Templeton et al., 2013). In brief, most MFA studies report strong cell growth and weak protein production during the exponential phase mainly based on glucose and glutamate consumption, which is followed by a transition of metabolism to the stationary phase, where strong protein production and weak cell growth are based mainly on the simultaneous utilization of glutamate, lactate, and glucose, with low or no lactate production. In general, peak antibody production is associated with increased oxidative metabolism activity. Cell decay has been reported during the entire process. Fig. 4 shows a compilation of the phenomena occurring during CHO cell metabolism identified from the analysis of metabolic fluxes reported in the literature.

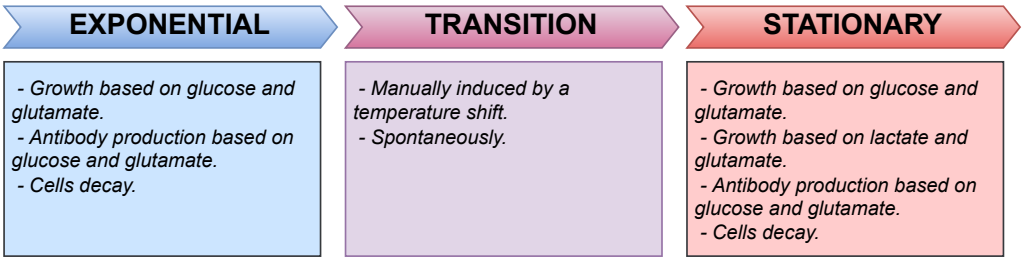


Fig. 4: Metabolic pathways involved in the different stages of CHO cell metabolism.

The exponential phase is accompanied by the production of by-products, such as lactate and ammonia, which inhibit cell metabolism. However, the lactate production during the exponential phase ensures the redox balance in the cytosol. It becomes an essential carbon and energy source during the stationary phase (Fig. 3).

3.2. Metabolic pathways description

The metabolic pathways identified from the metabolic flux analysis works reported in the previous section are summarized in figures 5 and 6, respectively, for the exponential and stationary phases. The extent of each metabolic pathway identified may depend on the CHO strain.

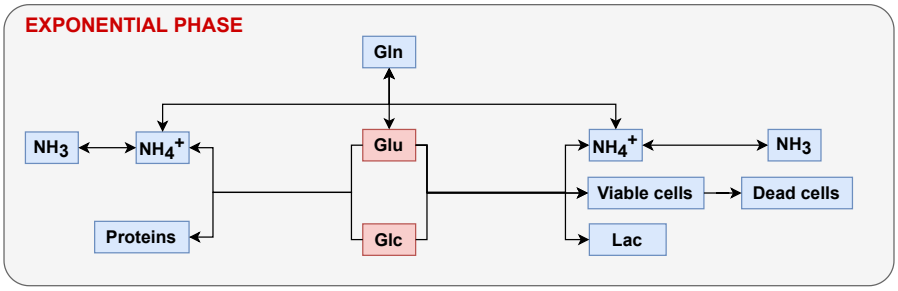


Fig. 5: Simplified schema of metabolic pathways taking place during the exponential phase (Glc: Glucose, Glu: Glutamate, Gln: Glutamine, and Lac: Lactate, with pink and blue colors denoting substrates and products, respectively).

The modeling of the stationary phase is more complex than the exponential phase, as it involves a more significant number of metabolic pathways co-occurring (e.g., protein production based on lactate, cell growth based on lactate, protein production based on glucose, and cell growth based on glucose). In this phase, lactate is an energy source primarily for cell growth. While lactate plays a role in protein production, its contribution is minimal compared to glucose. Consequently, some authors have dismissed the significance of lactate in protein production (Jimenez del Val et al., 2016; Kotidis et al., 2019). However, its minor contribution is crucial when the model distinguishes anabolism from catabolism by accounting for the ATP/ADP and NADH/NAD^+ electron transport chains (Nolan and Lee, 2011).

This remark on the role of lactate illustrates how the choice of metabolic pathways to be considered is a crucial question for modeling. The answer depends on the costs and benefits between their influence on the main output variables and the complexity of the parameter determination. Additionally, the number of stoichiometric and

dynamic parameters to be defined depends on the model complexity, which could be problematic for direct experimental determination. As it will be explained below, this question of parameter determination can be partly addressed by inverse analysis. Note that a more comprehensive model does not necessarily mean greater precision. A complexity that is too high and unbalanced with sufficient data can hinder its practical application (numerical solution problems, over-parametrized model, error propagation, large CPU time, etc.). Therefore, the challenge for modelers is to obtain a suitable trade-off.

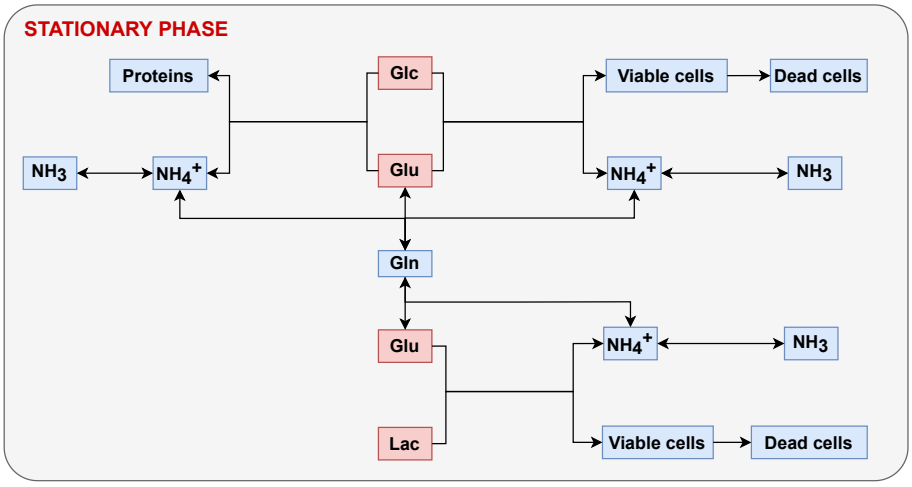


Fig. 6: Simplified schema of metabolic pathways occurring during the stationary phase (Glc: Glucose, Glu: Glutamate, Gln: Glutamine, and Lac: Lactate, with pink and blue colors denoting substrates and products, respectively).

Moreover, although both phases have common metabolic pathways, they differ considerably in stoichiometry and kinetics. Consequently, both phases must be separately formulated using submodels. These two submodels can coexist, but the pathways of the stationary phase submodel (Fig. 6) should be activated once the metabolic pathways of the exponential phase submodel (Fig. 5) have been deactivated.

3.3. Metabolic shift

The metabolic transition of CHO from the exponential to the stationary phase is usually intentionally induced by a temperature shift (temperature drop) (López-

Meza et al., 2016; Jimenez del Val et al., 2016). Nevertheless, factors that trigger a spontaneous metabolic shift remain a challenge for modelers, as it is crucial to obtain a predictive model that can address unexpected situations or propose innovative protocols. The metabolic transition from lactate production to lactate consumption in CHO cells is an essential phenomenon strongly linked to cell culture longevity and protein production yield (Brunner et al., 2018). However, this metabolic shift is difficult to control due to the unknown mechanisms involved, which are still under investigation (Zagari et al., 2013; Hartley et al., 2018; Hong et al., 2018).

Several studies have been conducted to identify the factors or the combination of factors that trigger this phenomenon. The literature has highlighted various major triggering factors (Fig. 7): limiting substrates such as glucose (Altamirano et al., 2004; Tsao et al., 2005; Altamirano et al., 2006; Martínez et al., 2013; Zagari et al., 2013; Brunner et al., 2021), glutamate, or glutamine (Zagari et al., 2013; Ghorbaniaghdam et al., 2014; Wahrheit et al., 2014), the effect of pH (Zalai et al., 2015; Liste-Calleja et al., 2015; Ivarsson et al., 2015), and inhibitory by-products such as lactate (Mulukutla et al., 2012; Kyriakopoulos and Kontoravdi, 2014; Pereira et al., 2018), and carbon dioxide (Brunner et al., 2018; Xu et al., 2018).

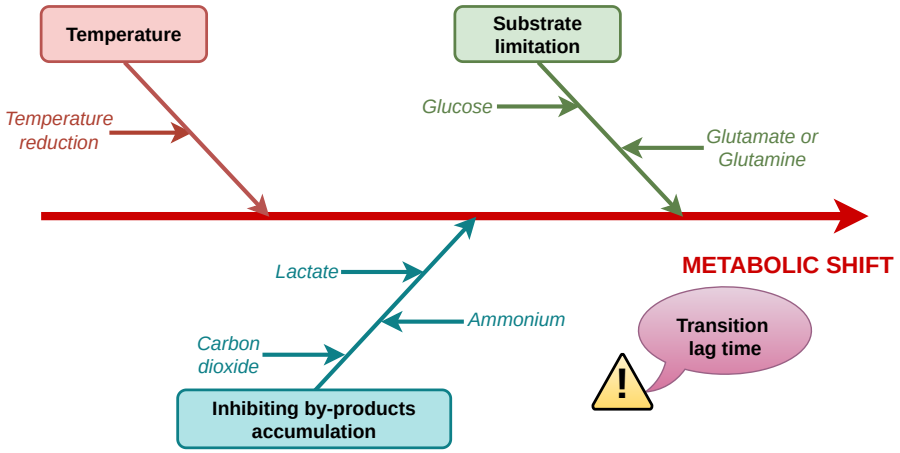


Fig. 7: Variables proposed in the literature as possible triggers for metabolic shift in the CHO cell system.

Substrate limitation (glucose, glutamate, or glutamine) can activate alternative metabolic pathways in which the carbon source is alternated from glucose to lactate. Many studies have reported that spontaneous metabolic shift is associated with pH (Zalai et al., 2015; Liste-Calleja et al., 2015; Ivarsson et al., 2015), which contradicts other studies in which metabolic shift has occurred spontaneously at constant pH (Altamirano et al., 2006; Mulukutla et al., 2012; Martínez et al., 2013; Ghorbaniaghdam et al., 2014). Martínez-Monge et al. (2019) tested pH-controlled and uncontrolled CHO cell metabolism. When pH was controlled in the bioreactor, only lactate consumption was observed once glucose was completely depleted. However, since these experiments were performed in batch mode, it can be assumed that the metabolic shift was triggered by substrate limitation, in this case, glucose limitation. It would be interesting to perform this experiment with sufficient glucose (batch feeding). In contrast, when pH was not controlled, it decreased due to the accumulation of secreted lactic acid, triggering its consumption when pH was below 6.80. According to these authors, the metabolic shift could be induced by adding lactate to the initial medium and setting the pH below 6.80.

Finally, the antagonism between cell growth and protein production reported in section 3.1 could explain the spontaneous metabolic shift. For example, inhibiting cell growth due to the accumulation of by-products could lead the cell to redirect their production capacity towards protein synthesis (Mulukutla et al., 2012; Kyriakopoulos and Kontoravdi, 2014; Pereira et al., 2018). This redirection could be a way to maintain maximum biological activity of the cell, consistent with its malignant behavior, while conveniently consuming an essential cell growth inhibitor by-product such as lactate.

In conclusion, according to the literature analysis, a combination of temperature and limiting substrates triggers the metabolic shift. On the other hand, the spontaneous metabolic shift occurs even with controlled and uncontrolled pH, implying that pH

could be an indirect macroscopic indicator of the metabolic shift. This parameter is directly related to inhibiting by-products already identified as triggering factors, like lactate and carbon dioxide. Another factor to consider is ammonium, which is an inhibitory by-product of cell metabolism that, at the same time, contributes to lowering the pH of the system. To our knowledge, ammonium has not yet been reported in the literature as a triggering factor. For this reason, it would be interesting to intentionally try to induce metabolic shifts by adding these inhibiting by-products to discern their contribution to metabolic shifts.

3.4. Main acid contributors to pH

To quantify the pH variation in the CHO system, it is necessary to consider the main metabolites, the substrate feed, and the acid-base solutions added during metabolism to regulate pH, where the main acid-base contributors must be identified. The reduction of pH during the degradation process is a direct consequence of CHO cell metabolism. All metabolic pathways are accompanied by carbon dioxide production and other inhibiting by-products such as lactate and ammonium that shift equilibrium towards hydronium formation, thus reducing the pH value. The chemical equilibria corresponding to the substrates and metabolites involved in CHO cell metabolism are shown in Table 1.

Table 1: Main acid contributors to pH variation

No.	**pH contributor**	**Chemical equilibrium**	K_a **at 25 °C**
1	Ammonium	$NH_4^+ + H_2O \rightleftharpoons NH_3 + H_3O^+$	$5.60\text{x}10^{-10}$
2	Lactate	$HLac + H_2O \rightleftharpoons Lac^- + H_3O^+$	$1.38\text{x}10^{-4}$
3	Carbon dioxide	$H_2CO_3 = CO_2 + H_2O$	
		$H_2CO_3 + H_2O \rightleftharpoons HCO_3^- + H_3O^+$	$4.20\text{x}10^{-7}$
		$HCO_3^- + H_2O \rightleftharpoons CO_3^{2-} + H_3O^+$	$4.80\text{x}10^{-11}$

Considering the equilibrium constant values at 25 °C and the same species concentration, the order of contribution to the system acidity is HLac $>$ NH_4^+ $>$ CO_2. However, the actual acid contribution depends on the extent to which each species

it is available in the system. One must remember that lactate is produced during the exponential phase and consumed during the stationary phase. Meanwhile, ammonium and CO_2 are produced throughout the process. Nevertheless, as pH is generally regulated, the influence of pH is discarded in most metabolic models. In contrast, it is important to note hyperosmolality, a non-physiological increase in osmolality that affects cell physiology, morphology, and proteome, resulting from the addition of concentrated feed and base solutions for pH regulation. While it enhances the specific antibody production rate, it inhibits cell growth, thereby affecting the final antibody titers (Kim et al., 2002; Min Lee and Koo, 2009; Zhang et al., 2010; Romanova et al., 2022). Incorporating hyperosmolality into mechanistic metabolic models could improve the accuracy of the impact of feeding on CHO cell metabolism.

4. Mathematical formulation

The formulation is a key building block to derive a mechanistic model. Even though they are not mandatory, the following assumptions are considered in most published works:

1. Perfectly stirred bioreactor,
2. Constant stoichiometric and kinetics parameter values,
3. Metabolic shift can be induced by temperature change and the combination of ammonium and lactate inhibiting effect,
4. Different stoichiometry and kinetics for exponential and stationary phases, even when they can change over the metabolic process, as can be corroborated in Templeton et al. (2013),
5. Constant chemical composition of CHO cells, regardless of the metabolic process and the nature of the substrate. It is to be expected that the chemical

composition of CHO cells will be different when glucose or lactate is used as the primary carbon source,

6. No oxygen limitation is considered, as aeration is regulated to ensure adequate oxygen levels during the process to avoid the reactive oxygen species formation that affects the cell's metabolism, reducing the proteins' productivity (Handlogten et al., 2018) considerably,

7. pH regulation.

Assumption 1 allows the 0D model to be derived. They are based on ODEs (Ordinary Differential Equations), which allow the efforts to be concentrated on the metabolism. Constant stoichiometric and kinetics parameter values are quasi-mandatory for tackling complexity. Yet, due to the existence of two phases and co-current pathways, the global stoichiometry over an entire process depends on culture conditions. Assumption 3 is essential as it connects the submodel describing the exponential phase with the submodel describing the stationary phase. Consequently, successful identification and mathematical description of the combination of effects of the factors that trigger this metabolic shift represents a significant challenge for modelers. Although not mandatory, the two last assumptions are usually fulfilled in production and allow the model to be simplified without restriction. Finally, it is essential to note that all the above assumptions are based on the condition that the CHO cell system is fully acclimatized to the culture medium and ready for production.

4.1. Kinetics of biological processes

Models based on physical, chemical, and biological principles offer a robust option in process engineering. However, the rates at which these principles occur is mandatory to obtain a predictive model. This must include the dynamics of all metabolic pathways, but also side phenomena such as the viable/non-viable cell densities and balances in the compartments considered by the formulation: at least the cells

and the bioreactor (nutrient/metabolite concentrations) (Tsopanoglou and del Val, 2021) and, depending on the model, the concentration obtained from balances in subcompartments of the cells, such as mitochondria. For a bioprocess operated in fed-batch mode, the following matter balances reads as:

$$V\frac{dS_{(j)}}{dt} = Q_{in}(S_{in(j)} - S_{(j)}) + V\sum_{i=1}^{i=n} r_{S_{(i)}}, \tag{1}$$

$$V\frac{dX}{dt} = -Q_{in}X + V\sum_{i=1}^{i=n} r_{X_{(i)}}, \tag{2}$$

$$V\frac{dP_{(j)}}{dt} = -Q_{in}P_{(j)} + V\sum_{i=1}^{i=n} r_{P_{(i)}}, \tag{3}$$

$$\frac{dV}{dt} = Q_{in}, \tag{4}$$

where Q_{in} is the volumetric flow rate of feed (l/h), $S_{(j)}$ is the limiting substrate $-j$ (g/l), $S_{in(j)}$ is the $S_{(j)}$ concentration in the feed (g/l), $P_{(j)}$ is the product concentration $-j$ (g/l), V is the bioreactor volume (l), X is the biomass concentration (cells/l), $r_{X(i)}$ is the volumetric growth rate for process $-i$ (cells/l/h), $r_{P(i)}$ is the volumetric metabolites production rate for process $-i$ (cells/l/h), and $r_{S(i)}$ is the volumetric substrates consumption rate for process $-i$ (cells/l/h) .

In general, the most widely accepted mathematical expressions for considering the limiting effect and the inhibitory effect on the specific growth rates are Monod's law and its variant, respectively, which are represented as a single expression as the product of all limiting and inhibitory substrates (Bree et al., 1988; Xing et al., 2010; Jimenez del Val et al., 2016; López-Meza et al., 2016; Kotidis et al., 2019; Yahia et al., 2021; Tsopanoglou and del Val, 2021). The Monod variant mathematical expression is also commonly used as a turnover rate function for changing metabolic pathways. The first term of Eq. 6 represents the substrate's limiting effect, whereas

the by-products' inhibitory effect is represented in the second term of Eq. 6:

$$r_{X_{(i)}} = \mu_{(i)} X, \tag{5}$$

$$\mu_{(i)} = \mu_{max(i)} \prod_{j=1}^{j=m} \frac{S_{(j)}}{K_{S(j)} + S_{(j)}} \prod_{j=1}^{j=k} \frac{K^{inh}_{S(j)}}{K^{inh}_{S(j)} + S^{inh}_{(j)}}, \tag{6}$$

where $\mu_{(i)}$ is the specific growth rate for process $-i$ (h^{-1}), $\mu_{max(i)}$ is the maximum growth rate for process $-i$ (h^{-1}), $S_{(j)}$ is the limiting substrate $-j$ (g/l), $K_{S(j)}$ is the half-saturation coefficient for the limiting substrate $S_{(j)}$ (g/l), $S^{inh}_{(j)}$ is the inhibiting by-product $-j$ (g/l) and $K^{inh}_{S(j)}$ is the half-saturation coefficient for the inhibiting by-product $S^{inh}_{(j)}$ (g/l).

The expression rates of the remaining model variables are expressed as a function of cell growth rate considering the stoichiometric relationships between the different chemical species and the cells formed during the biological process in a manner analogous to chemical reactions. Negative and positive sign values are added to the expression rate for substrate consumption (Eq. 7) and metabolite production (Eq. 8), respectively:

$$r_{S_{(i)}} = -\frac{1}{Y_{X/S_{(j)}}} r_{X_{(i)}}, \tag{7}$$

$$r_{P_{(i)}} = \frac{1}{Y_{X/P_{(j)}}} r_{X_{(i)}}, \tag{8}$$

where $Y_{X/S_{(j)}}$ is the X yield from $S_{(j)}$ (cells/g) and $Y_{X/P_{(j)}}$ is the X yield from $P_{(j)}$(cells/g).

However, even when in CHO cell metabolism modeling, Monod's law and its inhibition variant are the most used, it will be advisable to consider another kinetics expression already used in bioprocess modeling to find an adequate description of most of the variables' effect of the phenomena described. Fig. 8 shows examples

of kinetics expressions used in bioprocess modeling. The Moser Equation adds the parameter λ to Monod's law that describes the growth rate to the limiting substrate concentration. The Contois Equation states the proportionality between the effective saturation constant and the biomass concentration X. At high X, μ is inversely proportional to X. This is sometimes used to represent a diffusion limitation in flocculating or immobilized biomass (Snape et al., 2008). Expressions that combine the limiting and inhibitory effect for the same substrate, as is the case of the Haldane, Edwards, Webb, and Luong models, could be very useful in describing metabolic processes from substrates that have an inhibiting effect on the cell but can be utilized as a carbon source, as is the case for lactate in CHO systems. According to Edwards (1970), some mathematical expressions proposed to describe product inhibition can be borrowed to correlate substrate inhibition. For example, the Luong (1987) model was obtained from the Levenspiel (1980) model, while the Edwards (1970) model was derived from the Aiba et al. (1968) model. The Levenspiel (1980) and Aiba et al. (1968) models describe the inhibitory effect of the individual products. The logistic equation was initially proposed by the UK sociologist Thomas Malthus to describe "the law of population growth" at the end of the 18th century (Malthus, 1986), which was later used by the Belgian mathematician Pierre François Verhulst (Verhulst, 1838) to describe the biological population kinetics, in particular the self-limiting growth (Xu, 2020). According to this expression, the specific growth rate decreases linearly with an increasing cell population (X) and reaches zero at the maximum population. Indeed, this behavior is observed in most CHO systems. The use of this mathematical expression for modeling CHO cell metabolism has already been evaluated by Shirsat et al. (2015), obtaining better results than when Monod's law is used. In addition, the logistic equation has also been used to describe product inhibition (Contois, 1959; Fujimoto, 1963) (Fig. 8).

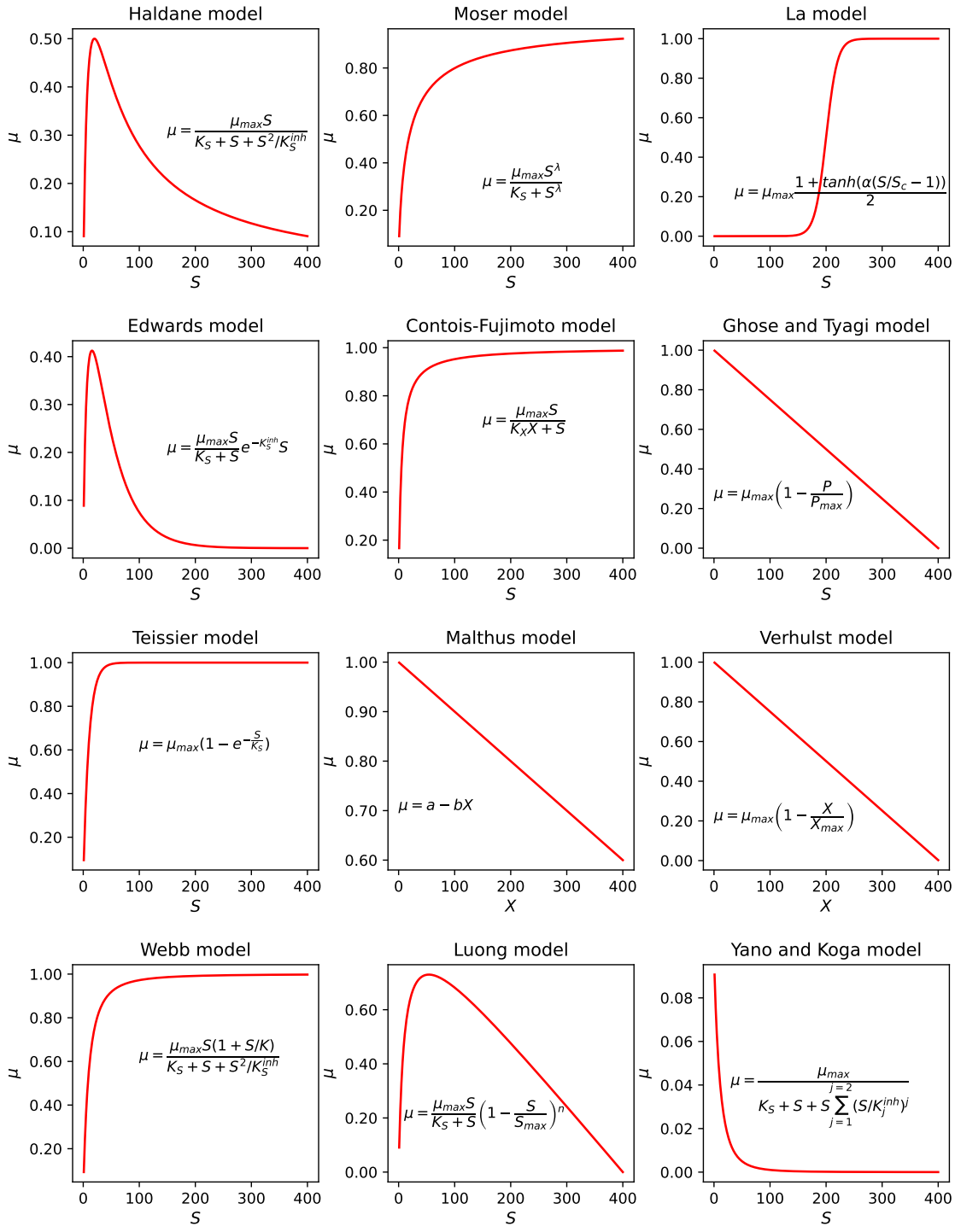


Fig. 8: Example of kinetics expressions used for modeling substrate limitation and by-products inhibition in bioprocesses (Verhulst, 1838; Teissier et al., 1936; Contois, 1959; Fujimoto, 1963; Webb, 1963; Yano and Koga, 1969; Edwards, 1970; Ghose and Tyagi, 1979; Malthus, 1986; Luong, 1987; Moser, 1988; Selişteanu et al., 2007; La et al., 2020).

However, even when this mathematical expression describes a similar behavior to that reported for CHO cell metabolism, it does not consider the metabolic shift. It is more related to the by-product inhibition accumulation. Nevertheless, the use of this mathematical expression is attractive because of the aspects mentioned above. The stepwise function proposed by La et al. (2020) is defined by two parameters: the shift value S_c defines the concentration value at which transition occurs, and the α parameter defines the sharpness of this transition. This powerful function can be used for describing switching between metabolic pathways, limiting substrate, and inhibiting by-product effects by manipulating the S_c and α values. It allows the value at which switching occurs and the rate at which it occurs around that value to be set independently. In this article, we have selected a set of functions that can tackle most situations. To go further, we recommend the review proposed by Mulchandani and Luong (1989), where the applicability and limitations of several functions are analyzed.

4.2. Kinetics of glutamine metabolism

The utilization and synthesis of glutamine can be represented as a chemical equilibrium, where the glutamine substrate is in equilibrium with ammonia and glutamate (Eq. 9):

$$Glutamate + NH_4^+ + ATP \overset{GS}{\leftrightharpoons} Glutamine + ADP \tag{9}$$

The equilibrium constant is then expressed as:

$$K_{Gln} = \frac{[Gln][ADP]}{[NH_4^+][ATP][Glu]} \tag{10}$$

As can be corroborated with Eq. 9 the glutamine formation from glutamate and ammonia requires energy in the form of ATP. Thus, the glutamine production rate

can be expressed as follows:

$$\frac{dGln}{dt} = k_{Gln}(K_{Gln}[NH_4^+][ATP][Glu] - [Gln][ADP]), \tag{11}$$

where k_{Gln} is the specific equilibrium displacement rate (s^{-1}). Batstone et al. (2002) suggested setting the specific equilibrium displacement rate one order of magnitude higher than the highest biological rate constant to reduce model stiffness. The reaction equilibrium markedly favors synthesis; the equilibrium constant at pH 7.0 and 37 ℃was 1200 (Levintow and Meister, 1954).

4.3. Kinetics of physical and chemical processes

For considering pH prediction in the model, the hydronium concentration must be estimated:

$$pH = -log[H_3O^+] \tag{12}$$

To estimate hydronium concentration, the main acid-base compounds involved in the CHO cell metabolism should be considered: ammonium, lactic acid, and carbon dioxide. The chemical equilibrium system shown in Table 1 allows us to set the following ordinary differential equation system:

Table 2: Kinetics expressions corresponding to main pH contributors

No.	pH contributor	Kinetics expression
1	Ammonium	$\frac{d[H_3O^+]}{dt} = k_{H_3O^+}(K_{a,NH_4^+}[NH_4^+] - [NH_3][H_3O^+])$
2	Lactate	$\frac{d[H_3O^+]}{dt} = k_{H_3O^+}(K_{a,HLac}[HLac] - [Lac^-][H_3O^+])$
3	Carbon dioxide	$\frac{d[H_3O^+]}{dt} = k_{H_3O^+}(K_{a,H_2CO_3}[CO_2] - [HCO_3^-][H_3O^+])$
		$\frac{d[H_3O^+]}{dt} = k_{H_3O^+}(K_{a,HCO_3^-}[HCO_3^-] - [CO_3^{2-}][H_3O^+])$

Where $k_{H_3O^+}$ is the specific equilibrium displacement rate (s^{-1}).

$$\frac{d[H_3O^+]}{dt} = \left(\frac{d[H_3O^+]}{dt}\right)_{NH_4^+} + \left(\frac{d[H_3O^+]}{dt}\right)_{HLac} + \left(\frac{d[H_3O^+]}{dt}\right)_{CO_2} \tag{13}$$

In biological systems, ammonia stripping takes place when ammonia is produced in the presence of continuous aeration. For a better understanding, the ammonia equilibrium should be first analyzed. According to the chemical equilibrium of Eq. 14, the production of ammonium via metabolic activity shifts the equilibrium, favoring free ammonia production.

$$NH_4^+ + H_2O \leftrightharpoons NH_3 + H_3O^+ \tag{14}$$

The free ammonia is liberated from the bioreactor liquid medium phase by aeration. Due to the low partial pressure of ammonia in the atmospheric air, the ammonia air flow saturation occurs instantaneously when it enters the biological reactor throughout the diffusion system. The free ammonia production rate can be expressed as follows:

$$\frac{d[NH_3]}{dt} = k_{NH_3}(K_{a,NH_4^+}[NH_4^+] - [NH_3][H_3O^+]), \tag{15}$$

where k_{NH_3} is the specific equilibrium displacement rate (s^{-1})(with the same rule proposed by Batstone et al. (2002) to reduce the system stiffness).

In diffused air systems, saturation with ammonia occurs within the first few millimeters of the ascent of the air bubbles through the liquid. Because this saturation process is almost instantaneous, the measured duration of ammonia desorption in the diffused air system is, in effect, a measure of the volume of air needed to achieve the observed removal of ammonia (Srinath and Loehr, 1974). Therefore, the ammonia removal process can be expressed using the following mathematical expression:

$$\frac{d[NH_3]}{dt} = k_{NH_3,st}[NH_3] \tag{16}$$

According to Srinath and Loehr (1974), the specific ammonia removal rate by aeration can be calculated using the following equation:

$$k_{NH_3,st} = 0.021e^{[1.93x10^{-4}\frac{q_{Air}}{V}+0.062(T-5)]}, \tag{17}$$

where $k_{NH_3,st}$ is the specific ammonia removal rate (h^{-1}), q_{Air} is the air flow injected to the bioreactor (cm^3/min), V is the bioreactor volume (l) and T is the temperature inside the bioreactor (℃). The Eq. 17 considers the airflow intensity and the temperature influence on the ammonia removal process.

4.4. Critical analysis of the published CHO mechanistic models

In the literature, two main model approaches can be found to describe CHO cell metabolism:

1. Considering the overall biological reactions (Xing et al., 2010; Jimenez del Val et al., 2016; López-Meza et al., 2016; Yahia et al., 2021),

2. Separating anabolism and catabolism reactions, considering ATP/ADP and NADH/NAD^+ electron transporters (Nolan and Lee, 2011).

Both approaches can describe the glycosylation process (Jimenez del Val et al., 2016; Kotidis et al., 2019). The protein quality depends on glycosylation, which occurs within the endoplasmatic reticulum and Golgi apparatus. This process takes place along the protein secretory pathway. It involves attaching an oligosaccharide chain to an amino acid residue, primarily asparagine (N-linked) or serine/threonine (O-linked glycosylation) (Galleguillos et al., 2017). The pros and cons of these models are thoroughly analyzed in Table 3. In most of these models, Monod's law and its variant for inhibition are used to describe the metabolic process rates, where glucose and glutamine or glutamate are the main limiting substrates (Xing et al., 2010; Jimenez del Val et al., 2016; López-Meza et al., 2016). In contrast, lactate and ammonia are the main inhibitory substrates (Xing et al., 2010; Kotidis et al., 2019). Lactate, an inhibitory product of CHO cell metabolism, becomes a limiting substrate during the stationary phase (Nolan and Lee, 2011; Kotidis et al., 2019).

Table 3: Pros and cons of mechanistic CHO models

Model description	Pros	Cons
Xing et al. (2010): Model based on Monod's law and its variant for inhibition.	• Protein production independent of cell growth, • μ is used as an emerging factor of metabolic shift, • Adequate prediction of the main model variables.	• No *Lac* utilization for cell growth during the stationary phase.
Nolan and Lee (2011): Model based on empirical kinetic rates describing intracellular catabolic and anabolic reactions.	• Protein production modeled independently of cell growth, • *Lac* utilization for cell growth during the stationary phase, • Adequate prediction of the main model variables.	• No decay process inclusion, • No NH_4^+ inhibitory effect inclusion, • Complex and over-parameterized model.
Jimenez del Val et al. (2016): Model linking *mAbs* glycosylation with cell secretory capacity based on Monod's law and its variant for inhibition.	• Integration of protein glycosylation with cellular secretory capacity, • Protein production modeled independently of cell growth, • Adequate prediction of the main model variables.	• No *Lac* inhibitory effect inclusion, • No *Glu* limiting effect inclusion, • No NH_4^+ production inclusion, hence its inhibiting effect.
López-Meza et al. (2016): Model based on Monod's law for cell growth and Luedeking-Piret model for *mAbs* production.	• Satisfactory prediction of μ, X_v, and protein concentration over time.	• No consideration of *Lac* and NH_4^+ production and their inhibiting effect, • No *Lac* utilization for cell growth, • Protein production associated with the cell growth process.
Kotidis et al. (2019): Model describing the influence of glycosylation precursor feeding on CHO cell metabolism based on Monod's laws, its variant for inhibition, and empirical rates.	• *Lac* production used as a trigger of metabolic shift. • Inclusion of *Lac* and NH_4^+ inhibiting effects and glycosylation, • *Lac* utilization for cell growth during the stationary phase, • Adequate prediction of the main model variables.	• Protein production associated with the cell growth process.
Yahia et al. (2021): Model based on mixed Monod's law and its variant for inhibition.	• Adequate prediction of protein production over time.	• No consideration of *Lac* and NH_4^+ production and their inhibiting effect, • No *Glc* and *Lac* utilization for cell growth and protein production, • Protein production associated with the cell growth process.

On the other hand, some empirical expression rates are also used to describe metabolic reactions involving intermediate metabolites, as in modeling central carbon metabolism (Nolan and Lee, 2011). Different stoichiometry and kinetics are assumed when the temperature is changed to trigger the metabolic shift (Jimenez del Val et al., 2016; López-Meza et al., 2016). In contrast, they are assumed to be constant when the metabolic shift occurs spontaneously (Xing et al., 2010; Kotidis et al., 2019). The metabolic shift is crucial in CHO cell metabolism modeling as it connects the exponential growth and stationary phases. Therefore, accurately predicting the metabolic shift is a genuine necessity for modeling. Yet this prediction remains an open challenge as this phenomenon is poorly understood. The most common indicators used in modeling for connecting both metabolic phases are temperature (Jimenez del Val et al., 2016), lactate concentration (Kotidis et al., 2019), and specific growth rate (μ) as an emerging indicator (Xing et al., 2010). Indeed, μ depends on limiting substrates, inhibiting by-products, and the temperature, which includes the main factors associated with this phenomenon as reported in the literature (Fig. 7). Several models describe the process of protein production associated with cell growth, where protein is considered a metabolite of cell growth, especially during the exponential phase (López-Meza et al., 2016; Kotidis et al., 2019; Yahia et al., 2021). The MFA (Ahn and Antoniewicz, 2011; Sengupta et al., 2011; Templeton et al., 2013) and modeling studies (Xing et al., 2010; Kotidis et al., 2019) suggest that there is always a significant increase in protein production following a substantial decrease in specific growth rates (either by an increase in the concentration of inhibitory metabolites or by a reduction in temperature). This finding demonstrates the antagonism between the two metabolic processes. It should also be noted that protein production is directly proportional to the concentration of viable cells. Consequently, to accurately predict protein production, it is crucial to consider cell growth and death processes. Cell death, influenced mainly by the toxic impact of

ammonia, plays a key role during the stationary phase (Xing et al., 2010).

The exponential phase is characterized by rapid cell growth and low protein production. For this reason, some authors omit this metabolic pathway in the exponential phase submodel. However, the yield of protein production from glucose is relatively low, requiring a significant amount of substrate for its synthesis, which can considerably influence the prediction of glucose consumption over time, even at very low protein production rates. Therefore, despite the low protein production, including it in the model is crucial.

Another important challenge of CHO cell metabolism models is the large number of kinetics and stoichiometric parameters involved, many of which are impossible to determine experimentally. In the case of models based on central carbon metabolism, although stoichiometry is known a priori, the kinetics involved are complex due to the number of intermediate processes, which in most cases include parameters and variables that are difficult to quantify. To summarize this literature review, the following aspects can be highlighted:

- The models offer poor investigation and description of metabolic pathways,
- In most cases, lactate use for cell growth is not considered,
- Criteria used for metabolic shift prediction are still not adequate,
- Two main pitfalls are observed: complex over-parameterized models and considerably simple models, both of which have little application in industry,
- The direct experimental determination of stoichiometric parameters is very difficult due to the simultaneous processes such as cell growth and protein production.

5. Numerical solution and calibration

5.1. Computational solutions of ODEs

The models under consideration in this work assumes the bioreactor to be perfectly stirred. The set of coupled equations is then a set of ODEs (Ordinary Differential Equations) of the generic form:

$$\begin{cases} u'(t) = F(u(t)) & \text{(18a)} \\ u(t_0) = u_0 & \text{(18b)} \end{cases}$$

Where $u(t)$ is the vector of unknowns and $F(u)$ is a vector of functions, the same size as u, giving the time-derivative of vector u.

Even though the equations are coupled and nonlinear, their computational solution is generally quite simple, using classical iterative methods to solve ODEs. The solution is obtained at discrete times t_n ($u_n = u(t_n)$). The linear *k-step* methods, an important family of solutions, can be defined as a general expression:

$$\sum_{j=0}^{k} \alpha_j u_{n+j} = h \sum_{i=0}^{k} \beta_j F_{n+j} \quad with \quad \alpha_k = 1 \tag{19}$$

Where h is the time step and $F_{n+j} = F(u_{n+j})$.

Expression (19) is a recurrent expression that allows the value $t(t_n) = u(t_n + h)$ to be computed from all previous values up to time t_n. $b_k = 0$ for explicit methods and $b_k \neq 0$ for implicit methods. For example, the simple forward (explicit) Euler method is obtained with $k = 1, \alpha_0 = -1, \beta_1 = 0$, and $\beta_0 = 1$. Using $\beta_1 = 1$ and $\beta_0 = 0$ instead gives the backward (implicit) Euler method. Expression (19) also includes the multistep explicit Adams–Bashforth and implicit Adams–Mouton methods (Butcher, 2000). Multistep methods aim to increase the order of the method (rate of convergence when h decreases). The Runge–Kutta methods are one-step

methods that use another strategy to increase the convergence order: in a single step, the function F is evaluated several times to obtain a Taylor expansion of the desired order.

The reader should, however, be aware that the solution order is not the single criterion for choosing a method. Formulations resulting from biological assumptions might be challenging to solve. This is the case, for example, with formulations involving several compartments (bioreactor, cells, organelles) and with very stiff functions, such as the switching varying over very narrow concentration levels (for example, the La et al. (2020) function shown in Fig. 8). In such cases, the system has components that vary on very different time scales, which poses challenges in selecting a suitable time step for numerical integration methods, resulting in stiff ODEs. The concept of convergence of stiff ODEs was introduced by Dahlquist (1963). He introduced the concept of A-convergence and proved that:

- A-convergence is possible only with implicit methods,
- A-convergence is not possible for methods above order 2.

It is important to keep this in mind when using the algorithms implemented in generic tools, such as the package *ode* in *R*, the class *scipy.integrate.ode* in *Python* or the existing *Matlab* solvers, to solve equations (18a) and (18b). The reader could refer to published works (Butcher, 2000; Cash, 2003) to pick up the suitable options for these solvers. All these solvers perform quite well for standard problems. As ODEs are much less demanding than PDEs (Partial Differential Equations), the CPU time usually remains very low (in the order of some seconds or even less). One point of attention, though, is that the system of equations may have components with very different orders of magnitude. Due to variations on very small quantities, the evolution of some variables may not be correctly taken into account in the convergence criterion, leading to a bad control of the time step and erroneous results.

For very severe configurations, more sophisticated algorithms, if possible coded in low-level languages (Fortran or C) for their efficiency once compiled, should therefore not be excluded a priori. Once compiled, these tools could be embedded in high-level tools such as the ones cited above. Even with classical problems, the efficiency of such solvers in terms of CPU is also likely to open new routes for the usage of mechanistic modeling: i) using these tools in an optimization loop requiring many solutions to be computed, ii) online parameter identification or iii) hybrid modeling for example.

Several methods were proposed to efficiently solve stiff ODEs (Butcher, 2000; Abdulle and Medovikov, 2001; Cash, 2003; Fatunla, 2014; Lebedev, 2017). The family of exponential integrators (exponential Rosenbrock-type integrators) are certainly among the most efficient ones (Cox and Matthews, 2002; Tokman, 2006; Caliari and Ostermann, 2009; Hochbruck et al., 2009; Carr et al., 2013). To derive these exponential methods, the Jacobian of F at point u_n, noted J_n, is first introduced in equation (18a):

$$u'(t) = F_n + J_n(u(t) - u_n) + R(u(t)) \tag{20}$$

Where R is the remainder.

Using the integration factor $e^{-F_n t}$, equation (20) becomes

$$u(t_n + h) = u_n + (e^{J_n h} - I)J_n^{-1}F_n + \int_{t_n}^{t_n+h} e^{J_n(t_n+h-t)}R(u(t))dt \tag{21}$$

All exponential methods rely on suitable methods to evaluate:

- the second term of the left-hand side, which needs to compute $\varphi(z) = \frac{exp(z)-1}{z}$ where z is a matrix,
- the third terms of the left-hand side, where R can simply be neglected, evalu-

ated assuming J to vary linearly over the time step, or by several evaluations, such as with a Runge–Kutta method.

5.2. Calibration and validation

The calibration process of a mathematical model is essential for its validation and future application (Rajamanickam et al., 2021). It consists of deterministic calculation of model parameter values consistent with data (Dawkins et al., 2001). The most appropriate strategy for model calibration is to experimentally determine as many model parameters as possible (direct determination)(González-Hernández et al., 2022). The remaining parameters can be taken from the literature if their value does not change significantly from one system to another (usually parameters with little influence on the model) or can be estimated by inverse analysis through an optimization process that searches for the unique combination of model parameters that allows a minimum deviation between the model output and the experimental data (indirect determination)(González-Hernández et al., 2022). Sometimes, when the model is simple, depending on the experience of modelers, this process can be carried out manually (Sin et al., 2008). The quality of the calibration process will depend on the quantity and quality of the data (Boudreau and McMillan, 2007; Abt et al., 2018) (economic resources availability, user competency, software access...). Therefore, the data collection process must be carefully carried out with an optimal and well-defined experimental design considering the variation of as many parameters as possible to obtain representative data of the phenomena described in the model (Rajamanickam et al., 2021). If historical data is used, much attention should be paid to the data curation process and selecting representative data. Finally, it is essential to remark that the model, once calibrated, must be tested against a validation database under different operating conditions.

5.2.1. *Parameter estimation by inverse analysis method*

The inverse analysis can be performed through an optimization procedure, looking for a minimum value using objective or multi-objective functions:

$$\begin{aligned} &Min : \{f_1(x), f_2(x), f_3(x), ..., f_n(x)\} \\ &subject\ to\text{: } x \in S, \end{aligned} \tag{22}$$

where x is the solution corresponding to the minimization process of n objective functions in the subspace S of the calibration parameter ranges.

Multi-objective optimization is essential when faced with real-world optimization problems since, in most cases, they are conditioned by multiple conflicting objectives with different levels of importance (Deb, 2014). In CHO models, two main metabolic phases are described mathematically, which present significant stoichiometric and kinetic differences (Jimenez del Val et al., 2016). When CHO cells are used as a host for protein production in the pharmaceutical industry, more importance is given to the stationary growth phase, characterized by a strong production of antibodies. In this case, multi-objective functions are helpful, as they allow for reinforcing the significance of specific phenomena variables during the calibration process, if necessary. This reinforcement can be achieved when the factor or variable's importance is given as a weighting average of all subfunctions:

$$F(x) = \sum_{i=1}^{n} w_i f_i(x) \quad \text{with: } \sum_{i=1}^{n} w_i = 1, \tag{23}$$

where w_i is the weighting factor, whose value is chosen by the decision maker according to his constraints (Florez et al., 2023).

For example, the mean relative error (MRE) between the experimental data (viable cells, glucose, glutamate, glutamine, lactate, ammonium, etc.) and the model prediction. This objective function can be adapted to weigh the importance of the

experiments, variables, and data points considered during the calibration process:

$$F(x) = \frac{1}{lmn}\sum_{i=1}^{\ell} w_i \sum_{j=1}^{m} w_j \sum_{k=1}^{n} w_k \left| \frac{y^{(i)}_{e(k,j)} - y^{(i)}_{m(k,j)}}{y^{(i)}_{e(k,j)}} \right| = \begin{cases} MRE, & \text{if } w_i, w_j, w_k = 1 \\ \neq MRE, & \text{if } w_i, w_j, w_k \neq 1 \end{cases}, \tag{24}$$

where $y^{(i)}_{e(j,k)}$ is the experimental data value $-k$ of variable $-j$ in experiment $-i$, $y^{(i)}_{m(j,k)}$ is the model output data value $-k$ of variable $-j$ in experiment $-i$, w_k is the weight assigned to data value $-k$, w_j is the weight assigned to variable $-j$ and w_i is the weight assigned to experiment $-i$.

An alternative approach involves using self-guided fitness functions that speed up the optimization process while reducing the risk of getting stuck in a local minimum. As an example, we can propose a fitness function that effectively amalgamates accuracy, utilizing the mean relative error (MRE), with variable trends (via Pearson's correlation coefficient (r)):

$$fitness = \frac{1}{3}(\underbrace{MRE}_{\text{Accuracy}} + \underbrace{(1-r)}_{\text{Trend}} + \underbrace{|MRE - (1-r)|}_{\text{Equilibrium}}) \tag{25}$$

where:

$$r = \frac{1}{lm}\sum_{i=1}^{i=l}\sum_{j=1}^{j=m} \frac{n(\sum_{k=1}^{k=n} y_e y_m) - (\sum_{k=1}^{k=n} y_e)(\sum_{k=1}^{k=n} y_m)}{\sqrt{[n\sum_{k=1}^{k=n} y_e^2 - (\sum_{k=1}^{k=n} y_e)^2][n\sum_{k=1}^{k=n} y_m^2 - (\sum_{k=1}^{k=n} y_m)^2]}} \tag{26}$$

In equation 25, the equilibrium term ensures an equitable contribution between accuracy and trends. This type of function significantly accelerates the convergence of the optimization process while decreasing the probability of becoming trapped in a local minimum.

5.2.2. Calibration challenges

During calibration, the model simulations are compared to experiments (Villaverde et al., 2022). A typical experimental data set consists of a number of batch or fed-batch trials, for which the information collected consists of growing conditions and time series of certain variables (in-line and/or offline measurements). This can represent a huge data set but with few contrasted conditions. The system is then likely to be over-determined, which is necessary to gain accuracy and counterbalance experimental noise/variability. This allows the system to be projected onto a smaller space, where the solution is optimized. However, these conditions are not sufficient: the series of experiments must also be able to test each parameter, obtaining significant variations in the measured variables when these parameters are modified. Indeed, one major challenge in bioprocess modeling is addressing over-parameterization (Mowbray et al., 2023), often resulting from an excessive number of parameters and/or highly correlated, that cannot be directly determined by experiments (Barz et al., 2015; Abt et al., 2018). Modelers often use inverse analysis but encounter difficulties due to the impossibility of generating experimental data that adequately captures the effect of all model parameters. Parametric Sensitivity Analysis is a widely adopted technique for this purpose (Kyriakopoulos et al., 2018). It effectively reduces the parameter calibration subspace by excluding less influential parameters and, in some cases, omitting less significant phenomena, to obtain a determinate system.

Another method to address over-parameterization consists of developing simplified models (Sha et al., 2018), called lumped models. This method models the set of metabolic pathways within a cellular compartment, such as glycolysis, the Krebs cycle, or oxidative phosphorylation, as a single global metabolic pathway (La et al., 2020). Alternatively, it could also be considered the integration of these closely related compartments as a unified entity through global biochemical reactions involv-

ing extracellular metabolites, a commonly employed practice in bioprocess modeling (Xing et al., 2010; Jimenez del Val et al., 2016; Kotidis et al., 2019).

Modelers also face challenges when introducing new parameters without prior references in the literature. These parameters often pose significant difficulties for direct experimental determination, making it necessary to employ inverse analysis. Consequently, a new challenge arises: establishing calibration bounds for these parameters. In such situations, using numerical derivatives can offer valuable insights, allowing the understanding of how these parameters affect the fitness function. This process may require a tedious iterative procedure: the user analyzes the results at each stage and decides which parameter to select for evaluation at the next iteration stage, until the desired results are achieved.

5.3. Optimization methods for minimizing the objective function

First, note that the objective of this study is not to compare optimization methods but to mention the most used methods and explain their advantages and disadvantages for the calibration of mechanistic models. Model calibration by inverse analysis requires simple and robust mathematical optimization methods. Swarm intelligence and evolutionary computation (SIEC) are now some of the most popular optimization methods employed in scientific research (Bansal et al., 2019; Kumar et al., 2019). These methods use a stochastic approach, allowing one to solve many complex problems without demanding many mathematical properties (e.g., convexity, continuity, or the explicit definition of the objective function) (Bansal et al., 2019). Particle swarm optimization (PSO) is among the most popular and successful swarm intelligence algorithms (Bansal et al., 2019). A great advantage of these algorithms is their parallelization capacity, considering that they are based on populations (fitness function evaluation) that considerably reduce time consumption but require a powerful computational capacity depending on the complexity of the fitness function. Evolutionary Computation (EC), mainly used to solve optimization

problems, comprises a series of problem-solving techniques based on the principles of biological evolution (e.g., natural selection and genetic inheritance) that allow for finding optimal global solutions (Bansal et al., 2019).

Mechanistic models present a great complexity considering the considerable number of stoichiometric and kinetic parameters that may be involved. Generally, the calibration process is carried out under non-stationary operating conditions, and multiple local minima may be encountered. For these reasons, heuristic optimization methods such as bio-inspired algorithms are mainly used for this task.

Several population-based, fitness-oriented, and variation-driven evolutionary algorithms have been proposed in the last century (Yu and Gen, 2010). These algorithms evolve using different strategies by employing common genetic operators such as selection, mutation, and reproduction, which depend on individual structures defined by an environment (Khaparde et al., 2022). In particular, in the area of evolutionary algorithms, the genetic algorithm (GA) and differential evolution (DE) algorithms are the most popular in the scientific community (Chaudhary et al., 2019). Currently, bio-inspired methods such as particle swarm optimization (PSO) have seen a remarkable application in solving several engineering problems.

Genetic Algorithm

The genetic algorithm, inspired by natural selection, is one of the most popular and widely used algorithms in various research areas due to its ease of implementation and convincing concepts (Omidinasab and Goodarzimehr, 2020; Goodarzimehr et al., 2022). Genetic algorithms are stochastic mathematical optimization methods based on the processes of natural selection and Darwinian survival of the fittest (Wang, 2003). New populations are produced by iteratively using genetic operators (e.g., chromosomal representation, selection, crossover, and mutation) on the individuals of a population. GAs are highly parallel, based on individual populations

of optimal candidate solutions that can be evaluated simultaneously. However, one of the main limitations of this method is its premature convergence, as they are sometimes trapped in local minima. Some researchers have suggested increasing diversity through selection pressure to avoid this problem (Katoch et al., 2021).

Differential Evolution

The differential evolution algorithm is a combinatorial algorithm based on populations of individuals. Like the GA, it allows the resolution of complex real-world optimization problems, which, in most cases, are reduced to the search for the global minimum of non-differentiable, discontinuous, and nonlinear objective functions (Lilla et al., 2013). Similar to other evolutionary algorithms, DE is a method that performs a stochastic search using a population of candidate solutions, applying mutation, crossover, and selection operators that drive the population toward better solutions in the optimization space (Georgioudakis and Plevris, 2020). DE differs from traditional evolutionary algorithms in generating new candidate solutions, employing a greedy generation scheme (Vesterstrom and Thomsen, 2004). The DE algorithm is simple but robust, governed by few control parameters, and its structure facilitates parallel computation with high convergence speed. Even when DE has few control parameters, their adjustment remains difficult, and their inappropriate manipulation can lead to premature convergence or stagnation, being key strategies chosen for the mutation operator. Therefore, to improve the performance of this algorithm a self-adaptive parameter setting technique is required (Khaparde et al., 2022). In this sense, new advanced DE variants with adaptive and self-adaptive control parameters have been developed (Eiben et al., 1999; Georgioudakis and Plevris, 2020).

Particle Swarm Optimization

The particle swarm optimization method was introduced in the mid-1990s by Kennedy and Eberhart (1995). This method is among the most widely used bio-inspired algorithms for solving optimization problems that are loosely inspired by foraging flocks of birds (Couceiro and Ghamisi, 2016). According to this analogy, each bird (considered a particle in the algorithm) uses its memory and the knowledge acquired by the swarm while searching for the best available food source (Venter and Sobieszczanski-Sobieski, 2003). This algorithm is governed by three fundamental operators: memory, inertia, and socialization. Subsequently, PSO has been extensively used to solve real-world problems in various biological and medical applications, computer graphics, and music composition (Sedighizadeh and Masehian, 2009). However, the main weakness of this optimization method is its propensity to converge to local optima prematurely (Banks et al., 2007; Houssein et al., 2021). One way to mitigate these problems is to work with optimization parameter ranges that are as narrow as possible.

Hybrid Optimization Methods

Hybrid optimization methods have become increasingly popular in recent years, especially in artificial intelligence, as they combine desirable properties of two or more optimization methods to increase their performance by mitigating their individual weaknesses (Thangaraj et al., 2011). It is important to remark that most of these algorithms have been modified by several authors to reduce the time consuming and facilitate the convergence process (Trivedi et al., 2015; Abidin, 2018; Aguitoni et al., 2018; Chaudhary et al., 2019). These algorithms have been integrated, taking advantage of the best of each one. Such connections may be implemented in many ways: (1) The use of a less accurate algorithm to initialize a more accurate algorithm (2) The parallel operation of two or more independent algorithms, with timely communication between them, to exchange the best individual solutions, (3)

The sequential execution of several algorithms in a single algorithm, (4) The combination of genetic operators from different optimization methods in each iteration, and others (Dziwiński and Bartczuk, 2019).

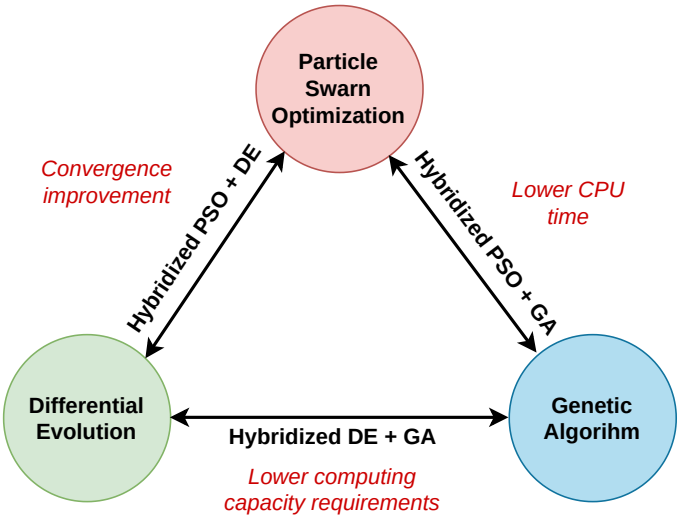


Fig. 9: Optimization methods hybridization

To achieve better performance (convergence speed, accuracy, and global optimization ability) compared to individual methods, several hybrid approaches appear in the literature, including DE and PSO (Hendtlass, 2001; Zhang and Xie, 2003; Talbi and Batouche, 2004; Hao et al., 2007; Das et al., 2008; Vaisakh et al., 2009; Garcia-Guarin et al., 2019; Dash et al., 2020; Li et al., 2021), PSO and GA (Mousa et al., 2012; Esmin et al., 2013; Samuel and Rajan, 2015; Ali and Tawhid, 2017; Dziwiński and Bartczuk, 2019; Pu et al., 2019; Omidinasab and Goodarzimehr, 2020; Fu et al., 2021; Goodarzimehr et al., 2022), and GA and DE (Trivedi et al., 2015, 2016; Abidin, 2018; Aguitoni et al., 2018; Chaudhary et al., 2019; Fathy et al., 2020).

Fortunately, most of these algorithms are already included in optimized and verified packages in popular programming tools used by the scientific community, such as MATLAB, Python, R, and others, which considerably facilitates the work of modelers. However, the selection of the optimization method depends on the modelers' experience, the fitness function's complexity, and computational capacity. We

strongly recommend that readers use the bio-inspired optimization method they know best, which is easy to handle with few control parameters, allowing them to obtain fast, efficient, and accurate results.

6. Challenges and prospects of mechanistic modeling of CHO cells

6.1. Needs, concerns and improvement gaps

Mechanistic modeling in the industry is rapidly expanding, finding applications in experimental design, scale-up, data analysis, product development, quality control, optimization, and decision-making, among other key areas (Hallow et al., 2010; Kuepfer et al., 2012; Helmlinger et al., 2017; Xing et al., 2023). In the pharmaceutical industry, mechanistic modeling holds its prestige for its ability to convert process data into enhanced information to understand the process better, guide decision-making, and facilitate the development of digital and automated technologies (Kuepfer et al., 2012; Sha et al., 2018; Narayanan et al., 2020). Examples of mechanistic modeling applications in CHO cell systems can be mentioned: e.g., Paul et al. (2019) improved the volumetric productivity of the CHO cell system operated in fed-batch mode by optimizing temperature and pH shifts by applying a mechanistic model, achieving an increase of 20% in the final product concentration; Kotidis et al. (2019) developed a mechanistic model (discussed earlier in section 4.4) that was successfully utilized to design and optimize the feeding strategy, achieving antibody concentrations exceeding 90% when compared to the control, without compromising the integral of viable cell density or the final antibody titer; Craven et al. (2014) used a nonlinear predictive model demonstrating its ability to achieve fixed-setpoint closed-loop glucose concentration control in a CHO cell system. As mentioned in this review article, the modeling of CHO cell metabolism has evolved considerably in recent decades, leading to a better understanding of the phenomena and a mathematical description. This paper focused on operational models as candidates for

optimization and improved control/command of the production process. Simple and complex models can be found in the literature. In most cases, these models are over-parameterized, which makes their practical application difficult, as the stoichiometry and kinetics change with the CHO strain, requiring recalibration of the model parameters. These over-parameterized models require much more information about the system for their application, which is hindered by the widespread use of optimized commercial basal mediums for CHO culture, of which, in most cases, the exact composition is still unknown. Moreover, in CHO systems, the direct experimental determination of certain stoichiometric parameters is very complex due to the simultaneous activation of several metabolic pathways involving the same substrate, such as cell growth and protein production from glucose or lactate (Figs. 4–6). The inverse analysis is a powerful tool for estimating these parameters during calibration. On the other hand, simple models are insufficient to predict CHO cell metabolism due to the omission of essential processes or variables in the model.

Metabolic reconstructions at the genomic scale are complex and often challenging to study in great detail. Recently, an interesting approach proposed by Martínez et al. (2022) simplifies the model by choosing subsystems based on the topology of the metabolic model, thus preserving the essential biochemical reactions.

Finding the right balance between the complexity and operationality of CHO modeling is part of the research gap that must be bridged. This determination requires that the metabolic pathways included in the model should be fed by relevant parameters. For example, this strategy allowed the metabolic shifts to be predicted in the case of yeast (González-Hernández et al., 2022). In our opinion, one of the main challenges for future improvements of the CHO model is the relevant prediction of spontaneous metabolic shifts from the exponential to the stationary phase.

Including organelles in the cell metabolism, leading to a compartmented model, and including balances of energy (ATP/ADP) and electron carriers ($NADH/NAD^+$)

are two promising possibilities to get a model able to spontaneously predict the metabolic shifts as a function of time or growth conditions (La et al., 2020). Although several studies suggest the importance of these redox metabolites in controlling variables, their inclusion in the models is complex by the fact that these metabolites are found at such low levels (on the order of μM), which makes it difficult to quantify them, and thus to determine the stoichiometry and associated kinetics (Nolan and Lee, 2011).

Glycosylation is another process of great interest to the pharmaceutical industry as it is crucial in biological activity and stability, increases the half-life, and reduces the immunogenicity of protein therapeutics (Kuriakose et al., 2016). Although this phenomenon has already been successfully coupled with a mechanistic model (Jimenez del Val et al., 2016; Kotidis et al., 2019), it involves many enzymatic reactions, which could lead to the over-parametrization of the model. Machine learning and molecular modeling could be an appropriate combination to obtain a better description of the structure, quality, and function of the protein obtained by glycosylation.

Over longer time horizons, one can imagine that a comprehensive description of the metabolic pathways could be included in operational tools. This description would bridge the significant gap that remains between the knowledge gained by metabolic flux analyses and operational simulation. Eventually, genome-scale models, together with the incorporation of enzyme constraint and enzyme kinetics, could produce predictive models from the genomes of specific strains (Price et al., 2004; Davidi and Milo, 2017). This development would be a major step forward in bridging the gap between metabolic engineering and the control and command of high-performance microbial strains. The GECKO project, for example, proves that this will certainly be possible in a not-to-far future (Sánchez et al., 2017; Domenzain et al., 2021).

6.2. The digital twin at the crossroads of mechanistic modeling and data science

Control command represents one of the most sought-after modeling applications, to improve the final product's concentration and quality. In our opinion, it is essential to maintain the mechanistic model as the core of the control system due to its prediction potential. To this end, there is a notable trend towards using hybrid models for control commands, particularly with new online sensors that immediately collect sufficient and relevant data to machine learning tools. Our research team is currently immersed in an ambitious project called CALIPSO, with the primary goal of halving development time and doubling productivity.

At short time horizons, the fast-growing field of data science is about to change how to build and tune mechanistic models and will address many of the above-mentioned limitations. In particular, the various building blocks presented in detail will be designed and improved concomitantly rather than sequentially. Over the past decade, machine learning has spread across many areas of engineering science. Autonomous cars face recognition, and weather forecasting without solving the equations of physics are probably the most popular examples of the success of machine learning. Machine learning has also spread to the field of bio-modeling (Pozzobon et al., 2021). However, in most of the studies in these review papers, machine learning is seen as an alternative to mechanistic modeling (Baker et al., 2018). In this sense, machine learning can cope with complex situations, provided the training database is large enough. The predictive capability is restricted to the domain paved by the database.

Rather than using machine learning instead of mechanistic modeling, we believe both areas are now mature enough to benefit from the best of both worlds: a predictive model capable of adapting to different products. Such a digital twin approach would work offline or online (Zhang et al., 2020; Yang and Chen, 2021). In this hybrid approach, the mechanistic model would remain at the heart of the interactions

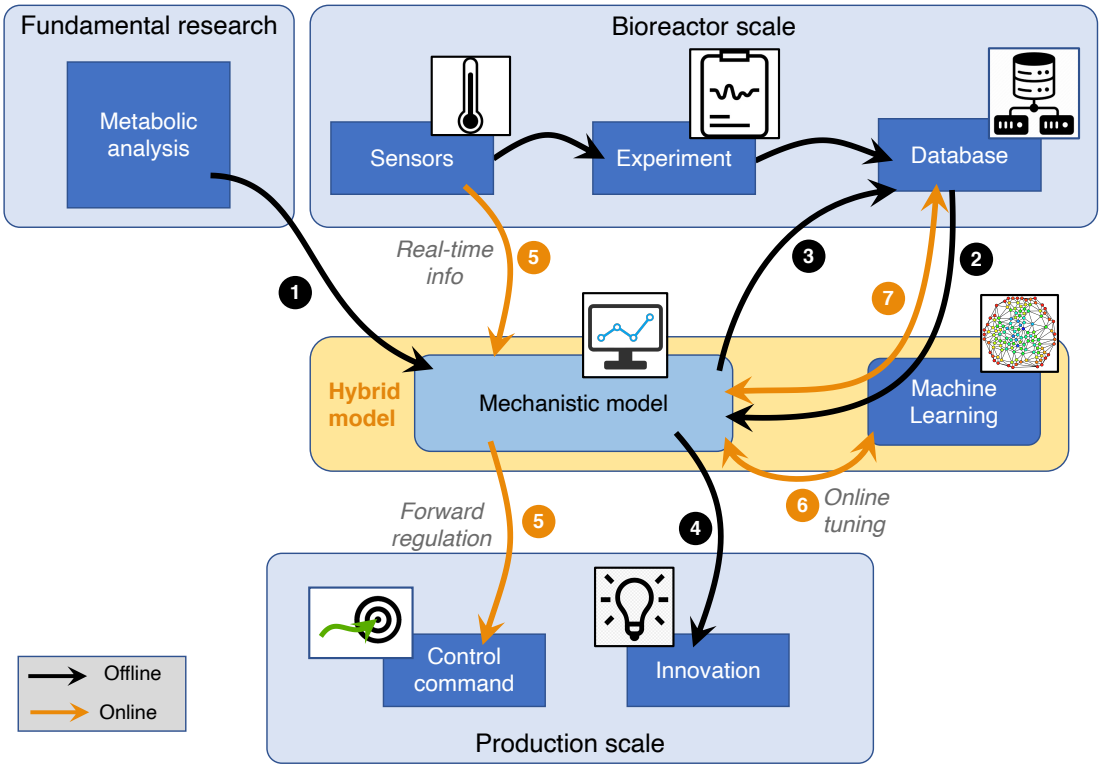


Fig. 10: The concept of hybrid modeling: using the synergies between experiment, mechanistic modeling, data science, and machine learning to fill the gap between modeling and process optimization (control command and innovation) by the concept of digital twin.

(Fig. 10). Hybrid models not only deliver more precise predictive outcomes but also exhibit superior resilience and extrapolation abilities (Narayanan et al., 2019, 2021).

The offline work (black arrows in Fig. 10) starts with constructing the mechanistic model, assembling all building blocks detailed in this paper. The metabolic description to be included in the model comes from fundamental studies (arrow 1). This is the cornerstone of mechanistic modeling. Besides, an experimental database must be generated from fermentation tests performed at the bioreactor level for various conditions using refined instrumentation (in-line and offline measurements). This database is primarily used for identifying the model parameters by inverse analysis, as explained in section 5. However, machine learning can also be used to test

model assumptions, either to reduce the formulation in case of overestimation or to recommend improvements (additional metabolic pathways, additional activating/inhibiting factors...) in particular, to represent non-standard data (this is why arrow 2 passes through the machine learning box). Once calibrated, the mechanistic model can be used offline to extend the database (arrow 4) or to imagine innovations (arrow 5: new production protocols, new strategies to overcome problems...).
The online use of the model (orange arrows of Fig. 10) includes at first the classical approach: using real-time information gathered by sensors together with the mechanistic model for an efficient control/command of the bio-process (arrows 5). As the model is predictive, it is capable of forward regulation, which is a significant advantage in reducing the culture time and making an early decision, for instance, to stop a batch when its trajectory can no longer be recovered. To achieve better performance in process control, machine learning can be used in two ways:

- By using the complete set of information collected up to the present time, machine learning can be used to tune online the model parameters. This is a crucial advantage for bioprocesses (arrow 6),

- by continuous queries to the database during cultivation to confirm whether the current batch is atypical or inconsistent with the information provided to the model. Identification of an atypical batch can reveal, for example, contamination (arrow 7).

In this last step of complexity, the mechanistic model and the database work in synergy as a hybrid approach in which information provided by the mechanistic model can benefit from the database, with queries triggered and analyzed by machine learning to complement the model. This process would allow, for example, the use of simulations that are impossible to carry out in real-time, or for comparing previous situations to detect specific issues (technical problems, sensor failure,

product anomaly, etc.).

7. Conclusions

This review article details the various building blocks that must be assembled to produce a mechanistic model of CHO cells for protein production. We have intentionally focused on operational models that can be used at a production scale. The starting point is the set of metabolic pathways that must be provided to the model. Once this input is completed, the activation and inhibition parameters and the reaction rates must be defined to provide a self-contained model capable of reproducing the dynamics of a bioreactor. One main concerns is finding the right balance between complexity and predictive capability. This balance includes the choice of the metabolic description, the database quality able to fit the parameters, and the prediction capabilities of the model in forward regulation.

The main facts, recommendations, and prospects in the papers include:

- To obtain a robust tool, over-parameterized models must be absolutely avoided,
- We recommend using lactate for cell growth during the stationary phase as it is not considered enough in the published works,
- Inverse analysis stands out as a powerful tool for model calibration and validation to overcome the complex direct experimental determination of stoichiometric parameters due to concurrent processes, such as cell growth and protein production,
- Indicators used for metabolic shift prediction still need to be improved, which motivates further investigations,
- Compartmental models, able to account for balances in organelles and the balance of energy and electron carriers, are two promising ways to predict the metabolic shift better,

- The intense use of machine learning and hybrid models, both offline and online, will shape the future of mechanistic modeling,
- In the long term, the considerable gap that remains between metabolic engineering and command and control should be bridged.

8. Acknowledgments

Communauté urbaine du Grand Reims, Département de la Marne, Région Grand Est and European Union (FEDER Champagne-Ardenne 2014-2020, FEDER Grand Est 2021-2027) are acknowledged for their financial support to the Chair of Biotechnology of CentraleSupélec and the Centre Européen de Biotechnologie et de Bioéconomie (CEBB). This work was also co-funded by BPI France as part of the Calipso project.

CRediT author statement

Y. González-Hernández: Conceptualization, Methodology, Formal analysis, Modeling, Visualization, Writing-Original draft preparation, Writing–Review and Editing. P. Perré: Conceptualization, Methodology, Formal analysis, Modeling, Writing–Review and Editing, Supervision, Funding acquisition.

References

A Gibbons, L., Rafferty, C., Robinson, K., Abad, M., Maslanka, F., Le, N., Mo, J., Clark, K., Madden, F., Hayes, R., et al., 2022. Raman based chemometric model development for glycation and glycosylation real time monitoring in a manufacturing scale cho cell bioreactor process. Biotechnol. Progr. 38, e3223.

Abdulle, A., Medovikov, A.A., 2001. Second order chebyshev methods based on orthogonal polynomials. Numer. Math. 90, 1–18.

1157 Abidin, D., 2018. A hybrid genetic-differential evolution algorithm (hybgade) for a
1158 constrained sequencing problem, in: 2018 International Conference on Artificial
1159 Intelligence and Data Processing (IDAP), IEEE. pp. 1–6.
1160 Abt, V., Barz, T., Cruz-Bournazou, M.N., Herwig, C., Kroll, P., Möller, J., Pört-
1161 ner, R., Schenkendorf, R., 2018. Model-based tools for optimal experiments in
1162 bioprocess engineering. Current opinion in chemical engineering 22, 244–252.
1163 Aguitoni, M.C., Pavão, L.V., Siqueira, P.H., Jiménez, L., Ravagnani, M.A.d.S.S.,
1164 2018. Heat exchanger network synthesis using genetic algorithm and differential
1165 evolution. Comput. Chem. Eng. 117, 82–96.
1166 Ahn, W.S., Antoniewicz, M.R., 2011. Metabolic flux analysis of cho cells at growth
1167 and non-growth phases using isotopic tracers and mass spectrometry. Metab. Eng.
1168 13, 598–609.
1169 Aiba, S., Shoda, M., Nagatani, M., 1968. Kinetics of product inhibition in alcohol
1170 fermentation. Biotechnol. Bioeng. 10, 845–864.
1171 Ali, A.F., Tawhid, M.A., 2017. A hybrid particle swarm optimization and genetic
1172 algorithm with population partitioning for large scale optimization problems. Ain
1173 Shams Eng. J. 8, 191–206.
1174 Altamirano, C., Illanes, A., Becerra, S., Cairó, J.J., Gòdia, F., 2006. Considerations
1175 on the lactate consumption by cho cells in the presence of galactose. J. Biotechnol.
1176 125, 547–556.
1177 Altamirano, C., Paredes, C., Illanes, A., Cairo, J., Godia, F., 2004. Strategies
1178 for fed-batch cultivation of t-pa producing cho cells: substitution of glucose and
1179 glutamine and rational design of culture medium. J. Biotechnol. 110, 171–179.
1180 Antoniewicz, M.R., 2013. Dynamic metabolic flux analysis—tools for probing tran-
1181 sient states of metabolic networks. Current opinion in biotechnology 24, 973–978.
1182 Antoniewicz, M.R., 2021. A guide to metabolic flux analysis in metabolic engineer-
1183 ing: Methods, tools and applications. Metab. Eng. 63, 2–12.

1184 Baker, R.E., Pena, J.M., Jayamohan, J., Jérusalem, A., 2018. Mechanistic models
1185 versus machine learning, a fight worth fighting for the biological community? Biol.
1186 lett. 14, 20170660.
1187 Banks, A., Vincent, J., Anyakoha, C., 2007. A review of particle swarm optimization.
1188 part i: background and development. Nat. Comput. 6, 467–484.
1189 Bansal, J.C., Singh, P.K., Pal, N.R., 2019. Evolutionary and swarm intelligence
1190 algorithms. volume 779. Springer.
1191 Barz, T., Körkel, S., Wozny, G., et al., 2015. Nonlinear ill-posed problem analysis
1192 in model-based parameter estimation and experimental design. Computers &
1193 Chemical Engineering 77, 24–42.
1194 Batstone, D.J., Keller, J., Angelidaki, I., Kalyuzhnyi, S., Pavlostathis, S., Rozzi, A.,
1195 Sanders, W., Siegrist, H.a., Vavilin, V., 2002. The iwa anaerobic digestion model
1196 no 1 (adm1). Water Sci. Technol. 45, 65–73.
1197 Boudreau, M.A., McMillan, G.K., 2007. New directions in bioprocess modeling and
1198 control: maximizing process analytical technology benefits. ISA.
1199 Bree, M.A., Dhurjati, P., Geoghegan Jr, R.F., Robnett, B., 1988. Kinetic mod-
1200 elling of hybridoma cell growth and immunoglobulin production in a large-scale
1201 suspension culture. Biotechnol. Bioeng. 32, 1067–1072.
1202 Brunner, M., Doppler, P., Klein, T., Herwig, C., Fricke, J., 2018. Elevated pco2
1203 affects the lactate metabolic shift in cho cell culture processes. Eng. Life Sci. 18,
1204 204–214.
1205 Brunner, M., Kolb, K., Keitel, A., Stiefel, F., Wucherpfennig, T., Bechmann, J.,
1206 Unsoeld, A., Schaub, J., 2021. Application of metabolic modeling for targeted
1207 optimization of high seeding density processes. Biotechnol. Bioeng. 118, 1793–
1208 1804.
1209 Budge, J.D., Roobol, J., Singh, G., Mozzanino, T., Knight, T.J., Povey, J., Dean,
1210 A., Turner, S.J., Jaques, C.M., Young, R.J., et al., 2021. A proline metabolism

selection system and its application to the engineering of lipid biosynthesis in chinese hamster ovary cells. Metab. Eng. Commun. 13, e00179.

Butcher, J.C., 2000. Numerical methods for ordinary differential equations in the 20th century. J. Comput. Appl. Math. 125, 1–29.

Cacciatore, J.J., Chasin, L.A., Leonard, E.F., 2010. Gene amplification and vector engineering to achieve rapid and high-level therapeutic protein production using the dhfr-based cho cell selection system. Biotechnol. Adv. 28, 673–681.

Caliari, M., Ostermann, A., 2009. Implementation of exponential rosenbrock-type integrators. Appl. Numer. Math. 59, 568–581.

Calmels, C., McCann, A., Malphettes, L., Andersen, M.R., 2019. Application of a curated genome-scale metabolic model of cho dg44 to an industrial fed-batch process. Metab. Eng. 51, 9–19.

Carr, E.J., Turner, I.W., Perré, P., 2013. A variable-stepsize jacobian-free exponential integrator for simulating transport in heterogeneous porous media: Application to wood drying. J. Comput. Phys. 233, 66–82.

Carrillo-Cocom, L., Genel-Rey, T., Araíz-Hernández, D., López-Pacheco, F., López-Meza, J., Rocha-Pizaña, M., Ramírez-Medrano, A., Alvarez, M., 2015. Amino acid consumption in naive and recombinant cho cell cultures: producers of a monoclonal antibody. Cytotechnology 67, 809–820.

Cash, J., 2003. Efficient numerical methods for the solution of stiff initial-value problems and differential algebraic equations. Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences 459, 797–815.

Chaudhary, D., Tailor, A.K., Sharma, V.P., Chaturvedi, S., 2019. Hygade: hybrid of genetic algorithm and differential evolution algorithm, in: 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), IEEE. pp. 1–4.

Chen, G., Hu, J., Qin, Y., Zhou, W., 2021. Viable cell density on-line auto-control

in perfusion cell culture aided by in-situ raman spectroscopy. Biochem. Eng. J. 172, 108063.

Contois, D., 1959. Kinetics of bacterial growth: relationship between population density and specific growth rate of continuous cultures. Microbiology 21, 40–50.

Costa, A.R., Rodrigues, M.E., Henriques, M., Azeredo, J., Oliveira, R., 2010. Guidelines to cell engineering for monoclonal antibody production. Eur. J. Pharm. Biopharm. 74, 127–138.

Couceiro, M., Ghamisi, P., 2016. Particle swarm optimization, in: Fractional order darwinian particle swarm optimization. Springer, pp. 1–10.

Cox, S.M., Matthews, P.C., 2002. Exponential time differencing for stiff systems. J. Comput. Phys. 176, 430–455.

Craven, S., Whelan, J., Glennon, B., 2014. Glucose concentration control of a fed-batch mammalian cell bioprocess using a nonlinear model predictive controller. Journal of Process Control 24, 344–357.

Dahlquist, G.G., 1963. A special stability problem for linear multistep methods. BIT Numer. Math. 3, 27–43.

Dang, C.V., 2010. Glutaminolysis: supplying carbon or nitrogen or both for cancer cells? Cell cycle 9, 3884–3886.

Das, S., Abraham, A., Konar, A., 2008. Particle swarm optimization and differential evolution algorithms: technical analysis, applications and hybridization perspectives, in: Advances of computational intelligence in industrial systems. Springer, pp. 1–38.

Dash, J., Dam, B., Swain, R., 2020. Design and implementation of sharp edge fir filters using hybrid differential evolution particle swarm optimization. AEU - Int. J. Electron. Commun. 114, 153019.

Davidi, D., Milo, R., 2017. Lessons on enzyme kinetics from quantitative proteomics. Curr. Opin. Biotechnol. 46, 81–89.

Dawkins, C., Srinivasan, T.N., Whalley, J., 2001. Calibration, in: Handbook of econometrics. Elsevier. volume 5, pp. 3653–3703.

Dean, J., Reddy, P., 2013. Metabolic analysis of antibody producing cho cells in fed-batch production. Biotechnol. Bioeng. 110, 1735–1747.

Deb, K., 2014. Multi-objective optimization, in: Search methodologies. Springer, pp. 403–449.

Domenzain, I., Sánchez, B., Anton, M., Kerkhoven, E.J., Millán-Oropeza, A., Henry, C., Siewers, V., Morrissey, J.P., Sonnenschein, N., Nielsen, J., 2021. Reconstruction of a catalogue of genome-scale metabolic models with enzymatic constraints using gecko 2.0. BioRxiv .

Domján, J., Pantea, E., Gyürkés, M., Madarász, L., Kozák, D., Farkas, A., Horváth, B., Benkő, Z., Nagy, Z.K., Marosi, G., et al., 2022. Real-time amino acid and glucose monitoring system for the automatic control of nutrient feeding in cho cell culture using raman spectroscopy. Biotechnol. J. , 2100395.

Dziwiński, P., Bartczuk, Ł., 2019. A new hybrid particle swarm optimization and genetic algorithm method controlled by fuzzy logic. IEEE Trans. Fuzzy Syst. 28, 1140–1154.

Edwards, V.H., 1970. The influence of high substrate concentrations on microbial kinetics. Biotechnol. Bioeng. 12, 679–712.

Eiben, Á.E., Hinterding, R., Michalewicz, Z., 1999. Parameter control in evolutionary algorithms. IEEE Trans. Evol. Comput. 3, 124–141.

Esmin, A.A., Matwin, S., et al., 2013. Hpsom: a hybrid particle swarm optimization algorithm with genetic mutation. Int. J. Innov. Comput. Inf. Control. 9, 1919–1934.

Fan, L., Kadura, I., Krebs, L.E., Hatfield, C.C., Shaw, M.M., Frye, C.C., 2012. Improving the efficiency of cho cell line generation using glutamine synthetase gene knockout cells. Biotechnol. Bioeng. 109, 1007–1015.

Fan, L., Kadura, I., Krebs, L.E., Larson, J.L., Bowden, D.M., Frye, C.C., 2013. Development of a highly-efficient cho cell line generation system with engineered sv40e promoter. J. Biotechnol. 168, 652–658.

Fathy, A., Abd Elaziz, M., Alharbi, A.G., 2020. A novel approach based on hybrid vortex search algorithm and differential evolution for identifying the optimal parameters of pem fuel cell. Renew. Energy 146, 1833–1845.

Fatunla, S.O., 2014. Numerical methods for initial value problems in ordinary differential equations. Academic press.

Feidl, F., Garbellini, S., Luna, M.F., Vogg, S., Souquet, J., Broly, H., Morbidelli, M., Butté, A., 2019. Combining mechanistic modeling and raman spectroscopy for monitoring antibody chromatographic purification. Processes 7, 683.

Florez, D., Stéphan, A., Perré, P., Rémond, R., 2023. Probabilistic multi-objective optimization of wood torrefaction conditions using a validated mechanistic model. Fuel 335, 126932.

Fu, X., Sun, Y., Wang, H., Li, H., 2021. Task scheduling of cloud computing based on hybrid particle swarm algorithm and genetic algorithm. Clust. Comput. , 1–10.

Fujimoto, Y., 1963. Kinetics of microbial growth and substrate consumption. J. Theor. Biol. 5, 171–191.

Galleguillos, S.N., Ruckerbauer, D., Gerstl, M.P., Borth, N., Hanscho, M., Zanghellini, J., 2017. What can mathematical modelling say about cho metabolism and protein glycosylation? Comput. Struct. Biotechnol. J. 15, 212–221.

Garcia-Guarin, J., Rodriguez, D., Alvarez, D., Rivera, S., Cortes, C., Guzman, A., Bretas, A., Aguero, J.R., Bretas, N., 2019. Smart microgrids operation considering a variable neighborhood search: The differential evolutionary particle swarm optimization algorithm. Energies 12, 3149.

Georgioudakis, M., Plevris, V., 2020. A comparative study of differential evolution

variants in constrained structural optimization. Front. Built Environ. 6, 102.

Ghorbaniaghdam, A., Chen, J., Henry, O., Jolicoeur, M., 2014. Analyzing clonal variation of monoclonal antibody-producing cho cell lines using an in silico metabolomic platform. PloS one 9, e90832.

Ghose, T., Tyagi, R., 1979. Rapid ethanol fermentation of cellulose hydrolysate. ii. product and substrate inhibition and optimization of fermentor design. Biotechnol. Bioeng. 21, 1401–1420.

Gianchandani, E.P., Chavali, A.K., Papin, J.A., 2010. The application of flux balance analysis in systems biology. Wiley Interdiscip Rev Syst Biol Med . 2, 372–382.

González-Hernández, Y., Michiels, E., Perré, P., 2022. A comprehensive mechanistic yeast model able to switch metabolism according to growth conditions. Fermentation 8, 710.

Goodarzimehr, V., Omidinasab, F., Taghizadieh, N., 2022. Optimum design of space structures using hybrid particle swarm optimization and genetic algorithm. World J. Eng. .

Gutierrez, J.M., Feizi, A., Li, S., Kallehauge, T.B., Hefzi, H., Grav, L.M., Ley, D., Baycin Hizal, D., Betenbaugh, M.J., Voldborg, B., et al., 2020. Genome-scale reconstructions of the mammalian secretory pathway predict metabolic costs and limitations of protein secretion. Nat. Commun. 11, 68.

Hallow, D.M., Mudryk, B.M., Braem, A.D., Tabora, J.E., Lyngberg, O.K., Bergum, J.S., Rossano, L.T., Tummala, S., 2010. An example of utilizing mechanistic and empirical modeling in quality by design. Journal of Pharmaceutical Innovation 5, 193–203.

Handlogten, M.W., Zhu, M., Ahuja, S., 2018. Intracellular response of cho cells to oxidative stress and its influence on metabolism and antibody production. Biochem. Eng. J. 133, 12–20.

Hao, Z.F., Guo, G.H., Huang, H., 2007. A particle swarm optimization algorithm

with differential evolution, in: 2007 international conference on machine learning and cybernetics, IEEE. pp. 1031–1035.

Hartley, F., Walker, T., Chung, V., Morten, K., 2018. Mechanisms driving the lactate switch in chinese hamster ovary cells. Biotechnol. Bioeng. 115, 1890–1903.

Hefzi, H., Ang, K.S., Hanscho, M., Bordbar, A., Ruckerbauer, D., Lakshmanan, M., Orellana, C.A., Baycin-Hizal, D., Huang, Y., Ley, D., et al., 2016. A consensus genome-scale reconstruction of chinese hamster ovary cell metabolism. Cell systems 3, 434–443.

Helmlinger, G., Al-Huniti, N., Aksenov, S., Peskov, K., Hallow, K.M., Chu, L., Boulton, D., Eriksson, U., Hamren, B., Lambert, C., et al., 2017. Drug-disease modeling in the pharmaceutical industry-where mechanistic systems pharmacology and statistical pharmacometrics meet. European Journal of Pharmaceutical Sciences 109, S39–S46.

Hendtlass, T., 2001. A combined swarm differential evolution algorithm for optimization problems, in: International conference on industrial, engineering and other applications of applied intelligent systems, Springer. pp. 11–18.

Hochbruck, M., Ostermann, A., Schweitzer, J., 2009. Exponential rosenbrock-type methods. SIAM J. Numer. Anal. 47, 786–803.

Hong, J.K., Nargund, S., Lakshmanan, M., Kyriakopoulos, S., Kim, D.Y., Ang, K.S., Leong, D., Yang, Y., Lee, D.Y., 2018. Comparative phenotypic analysis of cho clones and culture media for lactate shift. J. Biotechnol. 283, 97–104.

Houssein, E.H., Gad, A.G., Hussain, K., Suganthan, P.N., 2021. Major advances in particle swarm optimization: theory, analysis, and application. Swarm Evol. Comput. 63, 100868.

Huang, Z., Lee, D.Y., Yoon, S., 2017. Quantitative intracellular flux modeling and applications in biotherapeutic development and production using cho cell cultures. Biotechnol. Bioeng. 114, 2717–2728.

Ivarsson, M., Noh, H., Morbidelli, M., Soos, M., 2015. Insights into ph-induced metabolic switch by flux balance analysis. Biotechnol. Progr. 31, 347–357.

Katoch, S., Chauhan, S.S., Kumar, V., 2021. A review on genetic algorithm: past, present, and future. Multimed. Tools Appl. 80, 8091–8126.

Kauffman, K.J., Prakash, P., Edwards, J.S., 2003. Advances in flux balance analysis. Curr. Opin. Biotechnol. 14, 491–496.

Kennedy, J., Eberhart, R., 1995. Particle swarm optimization, in: Proceedings of ICNN'95-international conference on neural networks, IEEE. pp. 1942–1948.

Khaparde, A., Alassery, F., Kumar, A., Alotaibi, Y., Khalaf, O., Pillai, S., Alghamdi, S., 2022. Differential evolution algorithm with hierarchical fair competition model. Intell. Autom. Soft Comput. 33, 1045–1062.

Kim, M.S., Kim, N.S., Sung, Y.H., Lee, G.M., 2002. Biphasic culture strategy based on hyperosmotic pressure for improved humanized antibody production in chinese hamster ovary cell culture. In Vitro Cell. Dev. Biol. 38, 314–319.

Kingston, R.E., Kaufman, R.J., Bebbington, C., Rolfe, M., 2002. Amplification using cho cell expression vectors. Curr. Protoc. Mol. Biol. 60, 16–23.

Kotidis, P., Jedrzejewski, P., Sou, S.N., Sellick, C., Polizzi, K., Del Val, I.J., Kontoravdi, C., 2019. Model-based optimization of antibody galactosylation in cho cell culture. Biotechnol. Bioeng. 116, 1612–1626.

Kuepfer, L., Lippert, J., Eissing, T., 2012. Multiscale mechanistic modeling in pharmaceutical research and development. Advances in Systems Biology , 543–561.

Kumar, S., Nayyar, A., Paul, A., 2019. Swarm intelligence and evolutionary algorithms in healthcare and drug development. CRC Press.

Kuriakose, A., Chirmule, N., Nair, P., 2016. Immunogenicity of biotherapeutics: causes and association with posttranslational modifications. J. Immunol. Res. 2016.

Kyriakopoulos, S., Ang, K.S., Lakshmanan, M., Huang, Z., Yoon, S., Gunawan, R., Lee, D.Y., 2018. Kinetic modeling of mammalian cell culture bioprocessing: the quest to advance biomanufacturing. Biotechnol. J. 13, 1700229.

Kyriakopoulos, S., Kontoravdi, C., 2014. A framework for the systematic design of fed-batch strategies in mammalian cell culture. Biotechnol. Bioeng. 111, 2466–2476.

La, A., Du, H., Taidi, B., Perre, P., 2020. A predictive dynamic yeast model based on component, energy, and electron carrier balances. Biotechnol. Bioeng. 117, 2728–2740.

Lebedev, V., 2017. How to solve stiff systems of differential equations by explicit methods, in: Numerical methods and applications. CRC Press, pp. 45–80.

Levenspiel, O., 1980. The monod equation: a revisit and a generalization to product inhibition situations. Biotechnol. Bioeng. 22, 1671–1687.

Levintow, L., Meister, A., 1954. Reversibility of the enzymatic synthesis of glutamine. J. Biol. Chem. 209, 265–280.

Li, H., Xiang, S., Yang, Y., Liu, C., 2021. Differential evolution particle swarm optimization algorithm based on good point set for computing nash equilibrium of finite noncooperative game. AIMS Math. 6, 1309–1323.

Li, M.Y., Ebel, B., Paris, C., Chauchard, F., Guedon, E., Marc, A., 2018. Real-time monitoring of antibody glycosylation site occupancy by in situ raman spectroscopy during bioreactor cho cell cultures. Biotechnol. Progr. 34, 486–493.

Lilla, A., Khan, M., Barendse, P., 2013. Comparison of differential evolution and genetic algorithm in the design of permanent magnet generators, in: 2013 IEEE International Conference on Industrial Technology (ICIT), pp. 266–271.

Liste-Calleja, L., Lecina, M., Lopez-Repullo, J., Albiol, J., Solà, C., Cairó, J.J., 2015. Lactate and glucose concomitant consumption as a self-regulated ph detoxification mechanism in hek293 cell cultures. Appl. Microbiol. Biotechnol. 99, 9951–9960.

López-Meza, J., Araíz-Hernández, D., Carrillo-Cocom, L.M., López-Pacheco, F., del Refugio Rocha-Pizaña, M., Alvarez, M.M., 2016. Using simple models to describe the kinetics of growth, glucose consumption, and monoclonal antibody formation in naive and infliximab producer cho cells. Cytotechnology 68, 1287–1300.

Luo, Y., Kurian, V., Ogunnaike, B.A., 2021. Bioprocess systems analysis, modeling, estimation, and control. Curr. Opin. Chem. Eng. 33, 100705.

Luong, J., 1987. Generalization of monod kinetics for analysis of growth data with substrate inhibition. Biotechnol. Bioeng. 29, 242–248.

MacDonald, M.A., Nöbel, M., Roche Recinos, D., Martínez, V.S., Schulz, B.L., Howard, C.B., Baker, K., Shave, E., Lee, Y.Y., Marcellin, E., et al., 2022. Perfusion culture of chinese hamster ovary cells for bioprocessing applications. Critical reviews in biotechnology 42, 1099–1115.

Malthus, T.R., 1986. An essay on the principle of population (1798). The Works of Thomas Robert Malthus, London, Pickering & Chatto Publishers 1, 1–139.

Martínez, V.S., Buchsteiner, M., Gray, P., Nielsen, L.K., Quek, L.E., 2015. Dynamic metabolic flux analysis using b-splines to study the effects of temperature shift on cho cell metabolism. Metabolic Engineering Communications 2, 46–57.

Martínez, V.S., Dietmair, S., Quek, L.E., Hodson, M.P., Gray, P., Nielsen, L.K., 2013. Flux balance analysis of cho cells before and after a metabolic switch from lactate production to consumption. Biotechnol. Bioeng. 110, 660–666.

Martínez, V.S., Saa, P.A., Jooste, J., Tiwari, K., Quek, L.E., Nielsen, L.K., 2022. The topology of genome-scale metabolic reconstructions unravels independent modules and high network flexibility. PLOS Computational Biology 18, e1010203.

Martínez-Monge, I., Comas, P., Triquell, J., Casablancas, A., Lecina, M., Paredes, C., Cairó, J., 2019. Concomitant consumption of glucose and lactate: A novel batch production process for cho cells. Biochem. Eng. J. 151, 107358.

Marx, N., Eisenhut, P., Weinguny, M., Klanert, G., Borth, N., 2022. How to train

your cell-towards controlling phenotypes by harnessing the epigenome of chinese hamster ovary production cell lines. Biotechnol. Adv. , 107924.

Matasci, M., Hacker, D.L., Baldi, L., Wurm, F.M., 2008. Recombinant therapeutic protein production in cultivated mammalian cells: current status and future prospects. Drug Discov. Today Technol. 5, e37–e42.

McHugh, K.P., Xu, J., Aron, K.L., Borys, M.C., Li, Z.J., 2020. Effective temperature shift strategy development and scale confirmation for simultaneous optimization of protein productivity and quality in chinese hamster ovary cells. Biotechnol. Prog. 36, e2959.

Min Lee, G., Koo, J., 2009. Osmolarity effects, chinese hamster ovary cell culture. Encyclopedia of industrial biotechnology: bioprocess, bioseparation, and cell technology , 1–8.

Moser, A., 1988. The principles of bioprocess technol., in: Bioprocess Technol.. Springer, pp. 13–65.

Mousa, A., El-Shorbagy, M.A., Abd-El-Wahed, W.F., 2012. Local search based hybrid particle swarm optimization algorithm for multiobjective optimization. Swarm Evol. Comput. 3, 1–14.

Mowbray, M.R., Wu, C., Rogers, A.W., Rio-Chanona, E.A.D., Zhang, D., 2023. A reinforcement learning-based hybrid modeling framework for bioprocess kinetics identification. Biotechnology and Bioengineering 120, 154–168.

Mulchandani, A., Luong, J., 1989. Microbial inhibition kinetics revisited. Enzyme Microb. Technol. 11, 66–73.

Mulukutla, B.C., Gramer, M., Hu, W.S., 2012. On metabolic shift to lactate consumption in fed-batch culture of mammalian cells. Metab. Eng. 14, 138–149.

Narayanan, H., Luna, M.F., von Stosch, M., Cruz Bournazou, M.N., Polotti, G., Morbidelli, M., Butté, A., Sokolov, M., 2020. Bioprocessing in the digital age: the role of process models. Biotechnology journal 15, 1900172.

Narayanan, H., Seidler, T., Luna, M.F., Sokolov, M., Morbidelli, M., Butté, A., 2021. Hybrid models for the simulation and prediction of chromatographic processes for protein capture. Journal of Chromatography A 1650, 462248.

Narayanan, H., Sokolov, M., Morbidelli, M., Butté, A., 2019. A new generation of predictive models: The added value of hybrid models for manufacturing processes of therapeutic proteins. Biotechnology and bioengineering 116, 2540–2549.

Noh, S.M., Sathyamurthy, M., Lee, G.M., 2013. Development of recombinant chinese hamster ovary cell lines for therapeutic protein production. Curr. Opin. Chem. 2, 391–397.

Noh, S.M., Shin, S., Lee, G.M., 2018. Comprehensive characterization of glutamine synthetase-mediated selection for the establishment of recombinant cho cells producing monoclonal antibodies. Sci. Rep. 8, 1–11.

Nolan, R.P., Lee, K., 2011. Dynamic model of cho cell metabolism. Metab. Eng. 13, 108–124.

Nolan, R.P., Lee, K., 2012. Dynamic model for cho cell engineering. J. Biotechnol. 158, 24–33.

Omidinasab, F., Goodarzimehr, V., 2020. A hybrid particle swarm optimization and genetic algorithm for truss structures with discrete variables. J. Appl. Comput. Mech. 6, 593–604.

Orth, J.D., Thiele, I., Palsson, B.Ø., 2010. What is flux balance analysis? Nat. Biotechnol. 28, 245–248.

Pan, X., Dalm, C., Wijffels, R.H., Martens, D.E., 2017a. Metabolic characterization of a cho cell size increase phase in fed-batch cultures. Applied microbiology and biotechnology 101, 8101–8113.

Pan, X., Streefland, M., Dalm, C., Wijffels, R.H., Martens, D.E., 2017b. Selection of chemically defined media for cho cell fed-batch culture processes. Cytotechnology 69, 39–56.

Paul, K., Rajamanickam, V., Herwig, C., 2019. Model-based optimization of temperature and ph shift to increase volumetric productivity of a chinese hamster ovary fed-batch process. Journal of bioscience and bioengineering 128, 710–715.

Pereira, S., Kildegaard, H.F., Andersen, M.R., 2018. Impact of cho metabolism on cell growth and protein production: an overview of toxic and inhibiting metabolites and nutrients. Biotechnol. J. 13, 1700499.

Pozzobon, V., Levasseur, W., Guerin, C., Perré, P., 2021. Nitrate and nitrite as mixed source of nitrogen for chlorella vulgaris: fast nitrogen quantification using spectrophotometer and machine learning. J. Appl. Phycol. 33, 1389–1397.

Price, N.D., Reed, J.L., Palsson, B.Ø., 2004. Genome-scale models of microbial cells: evaluating the consequences of constraints. Nature Reviews Microbiology 2, 886–897.

Pu, H., Song, T., Schonfeld, P., Li, W., Zhang, H., Hu, J., Peng, X., Wang, J., 2019. Mountain railway alignment optimization using stepwise & hybrid particle swarm optimization incorporating genetic operators. Appl. Soft Comput. 78, 41–57.

Rajamanickam, V., Babel, H., Montano-Herrera, L., Ehsani, A., Stiefel, F., Haider, S., Presser, B., Knapp, B., 2021. About model validation in bioprocessing. Processes 9, 961.

Reinhart, D., Damjanovic, L., Kaisermayer, C., Kunert, R., 2015. Benchmarking of commercially available cho cell culture media for antibody production. Appl. Microbiol. Biotechnol. 99, 4645–4657.

Ritacco, F.V., Wu, Y., Khetan, A., 2018. Cell culture media for recombinant protein expression in chinese hamster ovary (cho) cells: history, key components, and optimization strategies. Biotechnol. Progr. 34, 1407–1426.

Rodrigues, M.E., Costa, A.R., Henriques, M., Azeredo, J., Oliveira, R., 2012. Comparison of commercial serum-free media for cho-k1 cell growth and monoclonal antibody production. Int. J. Pharm. 437, 303–305.

Romann, P., Kolar, J., Tobler, D., Herwig, C., Bielser, J.M., Villiger, T.K., 2022. Advancing raman model calibration for perfusion bioprocesses using spiked harvest libraries. Biotechnol. J. , 2200184.

Romanova, N., Schmitz, J., Strakeljahn, M., Grünberger, A., Bahnemann, J., Noll, T., 2022. Single-cell analysis of cho cells reveals clonal heterogeneity in hyperosmolality-induced stress response. Cells 11, 1763.

Samuel, G.G., Rajan, C.C.A., 2015. Hybrid: particle swarm optimization–genetic algorithm and particle swarm optimization–shuffled frog leaping algorithm for long-term generator maintenance scheduling. Int. J. Electr. Power Energy Syst. 65, 432–442.

Sánchez, B.J., Zhang, C., Nilsson, A., Lahtvee, P.J., Kerkhoven, E.J., Nielsen, J., 2017. Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints. Mol. Syst. Biol. 13, 935.

Santos, R.M., Kessler, J.M., Salou, P., Menezes, J.C., Peinado, A., 2018. Monitoring mab cultivations with in-situ raman spectroscopy: The influence of spectral selectivity on calibration models and industrial use as reliable pat tool. Biotechnol. Progr. 34, 659–670.

Schinn, S.M., Morrison, C., Wei, W., Zhang, L., Lewis, N.E., 2021. A genome-scale metabolic network model and machine learning predict amino acid concentrations in chinese hamster ovary cell cultures. Biotechnol. Bioeng. 118, 2118–2123.

Schwarz, H., Mäkinen, M.E., Castan, A., Chotteau, V., 2022. Monitoring of amino acids and antibody n-glycosylation in high cell density perfusion culture based on raman spectroscopy. Biochem. Eng. J. 182, 108426.

Sedighizadeh, D., Masehian, E., 2009. Particle swarm optimization methods, taxonomy and applications. Int. J. Comput. Sci. Eng. 1, 486.

Selişteanu, D., Petre, E., Răsvan, V.B., 2007. Sliding mode and adaptive sliding-mode control of a class of nonlinear bioprocesses. Int. J. Adapt. Control Signal

Process. 21, 795–822.

Sengupta, N., Rose, S.T., Morgan, J.A., 2011. Metabolic flux analysis of cho cell metabolism in the late non-growth phase. Biotechnol. Bioeng. 108, 82–92.

Sha, S., Huang, Z., Wang, Z., Yoon, S., 2018. Mechanistic modeling and applications for cho cell culture development and production. Curr. Opin. Chem. Eng. 22, 54–61.

Shirsat, N., Mohd, A., Whelan, J., English, N.J., Glennon, B., Al-Rubeai, M., 2015. Revisiting verhulst and monod models: analysis of batch and fed-batch cultures. Cytotechnology 67, 515–530.

Sin, G., De Pauw, D.J., Weijers, S., Vanrolleghem, P.A., 2008. An efficient approach to automate the manual trial and error calibration of activated sludge models. Biotechnology and bioengineering 100, 516–528.

Snape, J.B., Dunn, I.J., Ingham, J., Přenosil, J.E., 2008. Dynamics of environmental bioProcesses: modelling and simulation. John Wiley & Sons.

Srinath, E.G., Loehr, R.C., 1974. Ammonia desorption by diffused aeration. J. Water Pollut. Control Fed. , 1939–1957.

Talbi, H., Batouche, M., 2004. Hybrid particle swarm with differential evolution for multimodal image registration, in: 2004 IEEE International Conference on Industrial Technology, 2004. IEEE ICIT'04., IEEE. pp. 1567–1572.

Teissier, G., et al., 1936. Quantitative laws of growth. Ann. Physiol. Physicochim. Biol. 12, 527–586.

Templeton, N., Dean, J., Reddy, P., Young, J.D., 2013. Peak antibody production is associated with increased oxidative metabolism in an industrially relevant fed-batch cho cell culture. Biotechnol. Bioeng. 110, 2013–2024.

Thangaraj, R., Pant, M., Abraham, A., Bouvry, P., 2011. Particle swarm optimization: hybridization perspectives and experimental illustrations. Appl. Math. Comput. 217, 5208–5226.

Tokman, M., 2006. Efficient integration of large stiff systems of odes with exponential propagation iterative (epi) methods. J. Comput. Phys. 213, 748–776.

Torres, M., Zúñiga, R., Gutierrez, M., Vergara, M., Collazo, N., Reyes, J., Berrios, J., Aguillon, J.C., Molina, M.C., Altamirano, C., 2018. Mild hypothermia upregulates myc and xbp1s expression and improves anti-tnfα production in cho cells. PloS one 13, e0194510.

Trivedi, A., Srinivasan, D., Biswas, S., Reindl, T., 2015. Hybridizing genetic algorithm with differential evolution for solving the unit commitment scheduling problem. Swarm Evol. Comput. 23, 50–64.

Trivedi, A., Srinivasan, D., Biswas, S., Reindl, T., 2016. A genetic algorithm–differential evolution based hybrid framework: case study on unit commitment scheduling problem. Inf. Sci. 354, 275–300.

Tsao, Y.S., Cardoso, A., Condon, R., Voloch, M., Lio, P., Lagos, J., Kearns, B., Liu, Z., 2005. Monitoring chinese hamster ovary cell culture by the analysis of glucose and lactate metabolism. J. Biotechnol. 118, 316–327.

Tsopanoglou, A., del Val, I.J., 2021. Moving towards an era of hybrid modelling: advantages and challenges of coupling mechanistic and data-driven models for upstream pharmaceutical bioprocesses. Curr. Opin. Chem. Eng. 32, 100691.

Vaisakh, K., Sridhar, M., Murthy, K.L., 2009. Differential evolution particle swarm optimization algorithm for reduction of network loss and voltage instability, in: 2009 World Congress on Nature & Biologically Inspired Computing (NaBIC), IEEE. pp. 391–396.

Jimenez del Val, I., Fan, Y., Weilguny, D., 2016. Dynamics of immature mab glycoform secretion during cho cell culture: An integrated modelling framework. Biotechnol. J. 11, 610–623.

Venter, G., Sobieszczanski-Sobieski, J., 2003. Particle swarm optimization. AIAA J. 41, 1583–1589.

Verhulst, P.F., 1838. Notice sur la loi que la population suit dans son accroissement. Corresp. Math. Phys. 10, 113–129.

Vesterstrom, J., Thomsen, R., 2004. A comparative study of differential evolution, particle swarm optimization, and evolutionary algorithms on numerical benchmark problems, in: Proceedings of the 2004 congress on evolutionary computation (IEEE Cat. No. 04TH8753), IEEE. pp. 1980–1987.

Villaverde, A.F., Pathirana, D., Fröhlich, F., Hasenauer, J., Banga, J.R., 2022. A protocol for dynamic model calibration. Briefings in Bioinformatics 23, bbab387.

W Eyster, T., Talwar, S., Fernandez, J., Foster, S., Hayes, J., Allen, R., Reidinger, S., Wan, B., Ji, X., Aon, J., et al., 2021. Tuning monoclonal antibody galactosylation using raman spectroscopy-controlled lactic acid feeding. Biotechnol. Progr. 37, e3085.

Wahrheit, J., Niklas, J., Heinzle, E., 2014. Metabolic control at the cytosol–mitochondria interface in different growth phases of cho cells. Metab. Eng. 23, 9–21.

Wang, S.C., 2003. Genetic algorithm, in: Interdisciplinary computing in java programming. Springer, pp. 101–116.

Webb, J.L., 1963. Enzyme and metabolic inhibitors .

Xing, Z., Bishop, N., Leister, K., Li, Z.J., 2010. Modeling kinetics of a large-scale fed-batch cho cell culture by markov chain monte carlo method. Biotechnol. Progr. 26, 208–219.

Xing, Z., Nguyen, T.B., Kanai-Bai, G., Yamano-Adachi, N., Omasa, T., 2023. Construction of a novel kinetic model for the production process of a cva6 vlp vaccine in cho cells. Cytotechnology , 1–15.

Xu, P., 2020. Analytical solution for a hybrid logistic-monod cell growth model in batch and continuous stirred tank reactor culture. Biotechnol. Bioeng. 117, 873–878.

Xu, S., Jiang, R., Mueller, R., Hoesli, N., Kretz, T., Bowers, J., Chen, H., 2018. Probing lactate metabolism variations in large-scale bioreactors. Biotechnol. Progr. 34, 756–766.

Yahia, B.B., Malphettes, L., Heinzle, E., 2021. Predictive macroscopic modeling of cell growth, metabolism and monoclonal antibody production: Case study of a cho fed-batch production. Metab. Eng. 66, 204–216.

Yang, N., Guerin, C., Kokanyan, N., Perré, P., 2024. Raman spectroscopy applied to online monitoring of a bioreactor: Tackling the limit of detection. Spectrochim. Acta A Mol. Biomol. Spectrosc. 304, 123343.

Yang, R., Chen, J., 2021. Mechanistic and machine learning modeling of microwave heating process in domestic ovens: A review. Foods 10, 2029.

Yang, W., Zhang, J., Xiao, Y., Li, W., Wang, T., 2022. Screening strategies for high-yield chinese hamster ovary cell clones. Front. Bioeng. Biotechnol. 10.

Yano, T., Koga, S., 1969. Dynamic behavior of the chemostat subject to substrate inhibition. Biotechnol. Bioeng. 11, 139–153.

Yilmaz, D., Mehdizadeh, H., Navarro, D., Shehzad, A., O'Connor, M., McCormick, P., 2020. Application of raman spectroscopy in monoclonal antibody producing continuous systems for downstream process intensification. Biotechnol. Progr. 36, e2947.

Yousefi-Darani, A., Paquet-Durand, O., Von Wrochem, A., Classen, J., Tränkle, J., Mertens, M., Snelders, J., Chotteau, V., Mäkinen, M., Handl, A., et al., 2022. Generic chemometric models for metabolite concentration prediction based on raman spectra. Sensors 22, 5581.

Yu, X., Gen, M., 2010. Introduction to evolutionary algorithms. Springer Science & Business Media.

Zagari, F., Jordan, M., Stettler, M., Broly, H., Wurm, F.M., 2013. Lactate metabolism shift in cho cell culture: the role of mitochondrial oxidative activ-

ity. New Biotechnol. 30, 238–245.

Zalai, D., Koczka, K., Párta, L., Wechselberger, P., Klein, T., Herwig, C., 2015. Combining mechanistic and data-driven approaches to gain process knowledge on the control of the metabolic shift to lactate uptake in a fed-batch cho process. Biotechnol. Progr. 31, 1657–1668.

Zhang, J., Petersen, S.D., Radivojevic, T., Ramirez, A., Pérez-Manríquez, A., Abeliuk, E., Sánchez, B.J., Costello, Z., Chen, Y., Fero, M.J., et al., 2020. Combining mechanistic and machine learning models for predictive engineering and optimization of tryptophan metabolism. Nat. commun. 11, 1–13.

Zhang, J.H., Shan, L.L., Liang, F., Du, C.Y., Li, J.J., 2022. Strategies and considerations for improving recombinant antibody production and quality in chinese hamster ovary cells. Front. Bioeng. Biotechnol. 10.

Zhang, W.J., Xie, X.F., 2003. Depso: hybrid particle swarm with differential evolution operator, in: SMC'03 Conference Proceedings. 2003 IEEE International Conference on Systems, Man and Cybernetics. Conference Theme-System Security and Assurance (Cat. No. 03CH37483), IEEE. pp. 3816–3821.

Zhang, X., Garcia, I.F., Baldi, L., Hacker, D.L., Wurm, F.M., 2010. Hyperosmolarity enhances transient recombinant protein yield in chinese hamster ovary cells. Biotechnol. Lett. 32, 1587–1592.