

# Modeling the Effect of Amino Acids and Copper on Monoclonal Antibody Productivity and Glycosylation: A Modular Approach

Yu Luo, Robert J. Lovelett, J. Vincent Price, Devesh Radhakrishnan, Kristopher Barnthouse, Ping Hu, Eugene Schaefer, John Cunningham, Kelvin H. Lee, Raghunath B. Shivappa, and Babatunde A. Ogunnaike\*

In manufacturing monoclonal antibodies (mAbs), it is crucial to be able to predict how process conditions and supplements affect productivity and quality attributes, especially glycosylation. Supplemental inputs, such as amino acids and trace metals in the media, are reported to affect cell metabolism and glycosylation; quantifying their effects is essential for effective process development. We aim to present and validate, through a commercially relevant cell culture process, a technique for modeling such effects efficiently. While existing models can predict mAb production or glycosylation dynamics under specific process configurations, adapting them to new processes remains challenging, because it involves modifying the model structure and often requires some mechanistic understanding. Here, a modular modeling technique for adapting an existing model for a fed-batch Chinese hamster ovary (CHO) cell culture process without structural modifications or mechanistic insight is presented. Instead, data is used, obtained from designed experimental perturbations in media supplementation, to train and validate a supplemental input effect model, which is used to “patch” the existing model. The combined model can be used for model-based process development to improve productivity and to meet product quality targets more efficiently. The methodology and analysis are generally applicable to other CHO cell lines and cell types.

developing media recipes and for determining process operating conditions to meet production targets based on models.<sup>[1]</sup> Such models should capture precisely how process inputs (e.g., media recipes, operation conditions) affect the product attributes of interest, especially antibody productivity and glycosylation. N-linked glycosylation, a post-translational modification wherein oligosaccharides are enzymatically attached to the protein, is a critical quality attribute (CQA) monitored in mAb production. This is because the distribution of oligosaccharide structures (glycans) can have a significant impact on in vivo function of an antibody.<sup>[2–4]</sup> Since media composition has been shown to have a significant impact on glycan distribution, cell metabolism, and cell productivity,<sup>[5–7]</sup> in this paper, we develop a dynamic mathematical model that captures the effect of media composition on process dynamics during the manufacture of mAbs in a Chinese hamster ovary (CHO) cell culture process. We chose CHO cells as the mAb producer because they are the predominant host cell line used to manufacture mAbs,<sup>[8]</sup> and for illustrative purposes

## 1. Introduction

Regulatory authorities are increasingly requiring monoclonal antibody (mAb) manufacturers to employ *systematic* approaches for

only. The cell line used in this study is a glutamine synthetase (GS)-based CHO derivative incapable of natively producing glutamine. Our model was developed for a specific CHO cell line;

Y. Luo, R. J. Lovelett, D. Radhakrishnan, K. H. Lee, B. A. Ogunnaike  
University of Delaware, Chemical and Biomolecular Engineering  
150 Academy St, Newark, DE 19716, USA  
E-mail: ogunnaik@udel.edu

J. V. Price, K. Barnthouse, P. Hu, E. Schaefer, J. Cunningham  
Janssen Research and Development, Discovery,  
Product Development and Supply  
200 Great Valley Parkway, Malvern, PA 19355, USA  
R. B. Shivappa  
Takeda Pharmaceuticals, Biologics Process Development  
200 Shire Way, Lexington, MA 02421, USA

 The ORCID identification number(s) for the author(s) of this article can be found under <https://doi.org/10.1002/biot.202000261>

© 2020 The Authors. *Biotechnology Journal* published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

The copyright line for this article was changed on 6 October 2020 after original online publication.

DOI: 10.1002/biot.202000261

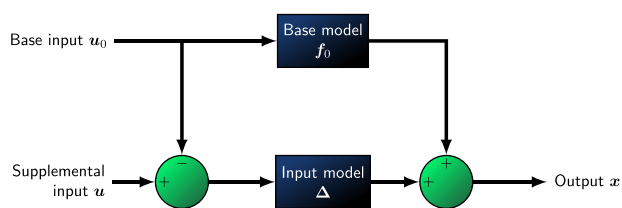


Figure 1. A schematic representation of the modular approach.

however, the modeling methodology and analysis are generally applicable to other cell lines and cell types.

We focused our analysis on asparagine, glutamate, and copper in the media because their impact on the cell culture process of manufacturing mAbs is known. Asparagine and glutamate play crucial roles in glycolysis and in the tricarboxylic acid (TCA) cycle—two major subprocesses in CHO cell metabolism. A high asparagine level can lead to a high ammonia level, which in turn can be detrimental to cell health, product quality, or both.<sup>[9,10]</sup> Copper has been shown to affect lactate metabolism, increase oxidative phosphorylation, and alter the glycan profile in mAb products or charge heterogeneity.<sup>[11–13]</sup> The input variables of our model therefore are the levels of asparagine, glutamate, and copper in media. Such a model, can be used to design new mAb manufacturing processes to meet desirable productivity and product quality targets by manipulating the levels of asparagine, glutamate, and copper in media.

While mechanistic or semi-mechanistic models that can simulate CHO cell mAb production or glycosylation dynamics accurately under predetermined process conditions exist<sup>[14–17]</sup> (see two reviews of the current state of mathematical modeling that appeared on the special issue in *Current Opinion in Chemical Engineering*<sup>[18,19]</sup> for their nearly exhaustive list of recent models), adapting them to new processes is challenging, mainly because this often requires modifying the structure of such models, which can be time-consuming. To address this problem, we take a fundamentally different approach in this paper. Our proposed modular modeling approach preserves the structure of an existing model,  $f_0$ , and augments it with a supplemental input model,  $\Delta$ , as illustrated in Figure 1. The supplemental input model,  $\Delta$ , describes how the process responds to the new inputs (asparagine, glutamate, and copper levels in media). Statistically designed orthogonal experiments were used to obtain data under the new process conditions, and the resulting data sets were used to characterize, in the form of the supplemental input model,  $\Delta$ , the new process information missing from the base model,  $f_0$ . Adding  $\Delta$  to  $f_0$  as a “patch” allows us to update the process dynamics when introducing not-yet-modeled process inputs without having to re-develop the base model,  $f_0$ . We assume that  $\Delta$  is structurally simple, and its contribution to overall model prediction does not dominate that of  $f_0$ .  $\Delta$  is meant to be used to adapt, not to replace,  $f_0$  efficiently for new processes. If the contribution of  $\Delta$  exceeds that of  $f_0$ , then it is time to develop a new base model.

Unlike the semi-mechanistic base model,  $f_0$ , the supplemental input model,  $\Delta$ , is data-driven and quantifies the effect of “known unknowns,” such as amino acids and trace metals, on cell metabolism and glycosylation when the actual mechanisms are not well understood. Since the development of  $\Delta$  is independent of the structure of  $f_0$ , the process of adapting an existing

model to new processes can be partitioned into two independent subtasks: i) training  $\Delta$  and ii) recalibrating  $f_0$ . The resulting augmented process model,  $f = f_0 + \Delta$ , can then be used for efficient, process-specific, and model-based process development to improve process performance and to meet product quality targets.

We illustrate this procedure with a multiscale base model,  $f_0$ , which describes process dynamics in two length-scales (macroscopic and microscopic) and two time-scales (slow and fast). The slow, macroscale model describes the dynamics of cell growth, metabolism, and mAb generation at the bioreactor level. The fast, microscale model, on the other hand, describes the dynamics of glycosylation at the molecular level where a series of enzymatic reactions occur on the antibodies. As a result of the disparate time-scales (hours and days for cell growth and minutes for glycosylation), the microscale model’s transient dynamics were assumed negligible compared with those of the macroscale model.

## 2. Experimental Section

### 2.1. Formulation of the Modular Model

The modular model consists of a semi-mechanistic base model,  $f_0$ , and a data-driven, supplemental input model,  $\Delta$ , designed to augment the base model, which as constructed, is incapable of accurately predicting the effect of asparagine, glutamate, and copper on process dynamics. In the technique, this base model is represented as a system of differential equations

$$\frac{dx}{ds} = f_0(x; \theta) \quad (1)$$

where  $x \in \mathbb{R}_{\geq 0}^n$  is a vector of  $n$  system state variables—cell densities, concentrations of nutrients, metabolites, mAbs, and distribution of glycans;  $s$  is a scalar of the independent time or position variable;  $\theta \in \mathbb{R}^p$  is a vector of  $p$  model parameters. The supplemental process input vector (not shown in  $f_0$ ),  $u \equiv [u_1, \dots, u_m]^T \in \mathbb{R}^m$ —asparagine, glutamate, and copper in the media ( $m = 3$ )—is kept constant at predetermined, baseline levels, represented by a vector  $u_0 \in \mathbb{R}^m$ .

The base model,  $f_0$ , alone cannot predict the process outputs accurately when the supplemental process inputs deviate from their baseline levels,  $u_0$ . Instead of modifying the structure of  $f_0$  in order to incorporate new inputs, we augment by adding  $\Delta \equiv [\Delta_1, \dots, \Delta_n]^T$  as follows

$$\frac{dx}{ds} = f(x, u_0, u; \Theta) = f_0(x; \theta) + \Delta(x, u - u_0; \beta) \quad (2)$$

where  $\Theta = \{\theta, \beta\}$  is a vector of all model parameters: the base model parameters,  $\theta$ , and the supplemental input model parameters,  $\beta$ , the dimension of which is yet to be determined.  $f = f_0 + \Delta$  is the augmented model that can predict the process outputs more precisely given the supplemental process inputs.

The supplemental input model determines the residual process dynamics,  $\Delta$ , based on how much the supplemental process inputs,  $u$ , deviate from their baseline levels,  $u_0$ . Let the dimensionless vector,  $u = u_0 \equiv [u_{0,1}, u_{0,2}, u_{0,3}]^T \equiv [-1, -1, -1]^T$ , represent the supplemental process inputs at the baseline levels, and  $u = [+1, +1, +1]^T$  represent their respective upper limits. The

supplemental input model,  $\Delta$ , is required to have the following characteristics:

- i)  $\Delta = 0$  when  $\mathbf{u} = \mathbf{u}_0$ . The effect of supplemental inputs is absent when asparagine, glutamate, copper are at their baseline levels. The process dynamics can be described solely by the base model,  $f_0$ .
- ii)  $\Delta$  is a linear function of  $\mathbf{u} - \mathbf{u}_0$  (see Section 3.1 for an empirical justification).
- iii)  $\Delta_i = 0$  when  $x_i = 0$ . This property ensures that the solution to Equation (2) is guaranteed to be nonnegative since concentration, cell density, and glycan distribution cannot fall below zero.

The simplest and most parsimonious form for  $\Delta$ , which also satisfies these properties is

$$\Delta_i(\mathbf{x}, \mathbf{u} - \mathbf{u}_0; \beta) = x_i \beta_{ij} (u_j - u_{j,0}) \quad (3)$$

which may be written in vector–matrix notation as

$$\Delta(\mathbf{x}, \mathbf{u} - \mathbf{u}_0; \beta) = \begin{bmatrix} x_1 & & \\ & \ddots & \\ & & x_n \end{bmatrix} \beta(\mathbf{u} - \mathbf{u}_0) \quad (4)$$

where  $\beta \in \mathbb{R}^{n \times m}$  is a matrix containing the supplemental input model parameters (recall that  $n$  is the number of state variables). The  $x_i$  term ensures the nonnegativity property by reducing  $\Delta_i$  to 0 when  $x_i = 0$ .

## 2.2. Base Model $f_0$

The multiscale base model,  $f_0$ , consists of a slow, macroscale cell culture model and a fast, microscale glycosylation model (Figure 2). The macroscale model describes the dynamic behavior at the bioreactor-level using a system of ordinary differential equations (ODEs)

$$\frac{d\mathbf{x}}{dt} = f_0(\mathbf{x}; \theta) = v(\theta)\mathbf{r}(\mathbf{x}; \theta) \quad (5)$$

where  $\mathbf{x} \in \mathbb{R}^{12}$  is a vector of 12 macroscale system state variables including nutrient, metabolite, mAb concentrations, and cell densities;  $\mathbf{r} \in \mathbb{R}^{10}$  is a vector of reaction, cell growth, and cell death rates;  $v \in \mathbb{R}^{12 \times 10}$  is the stoichiometry matrix;  $\theta \in \mathbb{R}^{35}$  is a vector of macroscale base model parameters. The inputs to the macroscale model are seeding cell density, nutrient concentrations of the basal medium, and the daily feeding and sampling volumes; the outputs are predicted mAb titer, viable cell density (VCD), total cell density (TCD), and metabolite concentrations over the course of a fed-batch run. The model structure and parameterization are presented in detail in the Supporting Information.

The outputs from the macroscale model are then sent as inputs to the microscale glycosylation model. The microscale model describes the glycosylation dynamics at the molecular level as new antibodies undergo the post-translational, enzymatic glycosylation process in the Golgi. The outputs of the microscale model are the glycan distributions as a function of time. The Golgi is

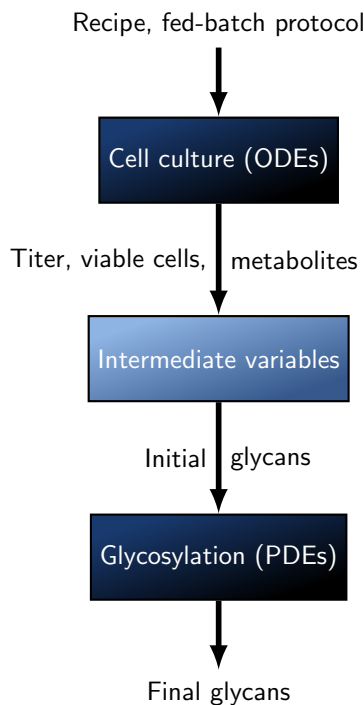


Figure 2. A schematic representation of the multiscale base model.

approximated as a plug flow reactor (PFR), therefore the system of partial differential equations (PDEs) is used below to describe the glycosylation dynamics inside the Golgi<sup>[14]</sup>

$$\frac{\partial \mathbf{x}}{\partial t} = -\frac{1}{\tau} \frac{\partial \mathbf{x}}{\partial z} + v\mathbf{r} \quad (6)$$

where  $z$  is position along the Golgi;  $\mathbf{x} \in \mathbb{R}^{30} \equiv [C_{M8}, C_{M7}, \dots, C_{FA2G2S2}]^T$  is a vector of 30 glycan concentrations (see the complete list of glycans in the Supporting Information);  $\tau = 22$  min is the residence time of antibodies inside the Golgi<sup>[14]</sup>;  $\mathbf{r} \in \mathbb{R}^{38}$  is a vector of glycosylation reaction rates;  $v \in \mathbb{R}^{30 \times 38}$  is the stoichiometric matrix for the glycosylation reactions (note that the notations  $\mathbf{x}$ ,  $\mathbf{r}$ , and  $v$  are reused for system state, reaction rates, and stoichiometry in the microscale model).  $\mathbf{x}_0 \in \mathbb{R}^{30} \equiv [\bar{C}, 0, 0, \dots, 0]^T$  is the vector of initial and boundary condition for Equation (6). Prior to entering the Golgi, antibodies are first glycosylated with glycan M8 inside the endoplasmic reticulum (ER).  $\bar{C}$  is the initial M8 concentration and also the average glycan concentration throughout the entire glycosylation process inside the Golgi because the reactions in question only change the type, not the amount, of glycans. Since  $\bar{C}$  is, by definition, the concentration of glycans from newly synthesized antibodies, it can be calculated as follows

$$\bar{C} = 2 \frac{\Delta N_{mAb}}{\Delta V} = \frac{2q_{mAb}}{Q} \quad (7)$$

where the number 2 is introduced to account for the two glycans per mAb;  $\Delta N_{mAb}$  is the amount of new antibodies entering the Golgi at a specific time;  $\Delta V$  is the total volume of new antibodies;

$Q = 1.12 \mu\text{m}^3 \text{min}^{-1}$  is the average flow rate of glycans through the Golgi<sup>[14]</sup>;  $q_{\text{mAb}}$  is the cellular productivity defined as

$$q_{\text{mAb}} \equiv \frac{1}{\text{VCD}} \frac{1}{V} \frac{1}{\text{MW}_{\text{mAb}}} \frac{d}{dt} (C_{\text{mAb}} V) \quad (8)$$

where  $\text{MW}_{\text{mAb}}$  is the average molecular weight of mAb (around 150 kDa).

Solving Equation (6) can be slow because of the large number of reactions involved and the usage of the finite difference method to approximate a system of PDEs with a series of systems of ODEs. On the other hand, the macroscale and microscale models describe dynamics of different time- and length-scales. The macroscale cell culture dynamics are on the order of hours and days while the microscale glycosylation dynamics are on the order of minutes, reaching steady-state quickly. Because of the time-scale difference, only the steady-state, instantaneous glycan profile is needed, which can be computed by letting  $\partial\mathbf{x}/\partial t = 0$  and converting Equation (6) from PDEs to ODEs

$$\frac{d\mathbf{x}}{dz} = f_0(\mathbf{x}; \boldsymbol{\theta}) = \tau \nu \mathbf{r}(\mathbf{x}; \boldsymbol{\theta}) \quad (9)$$

where the stoichiometry matrix,  $\nu \in \mathbb{R}^{30 \times 38}$ , is well documented in the literature<sup>[14]</sup>;  $\boldsymbol{\theta} \in \mathbb{R}^{44}$  is a vector of the microscale model parameters (see Supporting Information for detailed model structure and parameterization). The exit concentrations (i.e., the concentrations at  $z = L_{\text{Golgi}}$ ) are then normalized to obtain the *instantaneous* distribution of glycans on the *new* antibodies produced at a certain time point. To obtain the extracellular, *cumulative* glycan distribution, a common glycan measurement, the instantaneous glycan distribution over the culture time is integrated as

$$\mathbf{x}_{\text{ec}}(t) = \frac{\int_0^t \text{VCD}(t') q_{\text{mAb}}(t') \mathbf{x}(t') dt'}{C_{\text{mAb}}(t)} \quad (10)$$

where  $\mathbf{x}_{\text{ec}} \in \mathbb{R}^{30}$  is a vector of (cumulative) extracellular glycan fractions;  $\mathbf{x}$  is the vector of the instantaneous, intracellular mole fractions of glycans;  $C_{\text{mAb}}$  is the extracellular mAb concentration;  $t'$  is a dummy time variable of time used in the integral.

### 2.3. Experimental Design to Identify the Effect of Supplemental Inputs

Orthogonal experiments were designed to determine and quantify the effect of asparagine, glutamate, and copper levels in the feed supplements on mAb productivity and glycosylation. The resulting experimental data were then used to achieve a twofold objective: i) to quantify the effects these supplemental inputs in the form of the model,  $\Delta$ , and ii) to validate the model. The initial design of experiments (DoE) featured a  $2^3$  factorial set of experiments, where asparagine, glutamate, and copper concentrations are either at the baseline level (−1: asparagine at  $1.87 \text{ g L}^{-1}$ , glutamate at  $3.64 \text{ g L}^{-1}$ , and copper at  $1.58 \mu\text{g L}^{-1}$ ) or at an elevated level (+1: asparagine at  $3.68 \text{ g L}^{-1}$ , glutamate at  $6.95 \text{ g L}^{-1}$ , and copper at  $35.0 \mu\text{g L}^{-1}$ ). The eight conditions are labeled A01–A08 in **Table 1** with three replicates per condition for a total of 24 experiment runs. Subsequently, another  $2^3$  set of experiments, S01–S08, was designed for the purpose of generating data for

**Table 1.** Experimental design.

Experiment	Asparagine	Glutamate	Copper
A01	−1	−1	−1
A02	+1	−1	−1
A03	−1	+1	−1
A04	+1	+1	−1
A05	−1	−1	+1
A06	+1	−1	+1
A07	−1	+1	+1
A08	+1	+1	+1
S01	−1/3	−1/3	−1/3
S02	+1/3	−1/3	−1/3
S03	−1/3	+1/3	−1/3
S04	+1/3	+1/3	−1/3
S05	−1/3	−1/3	+1/3
S06	+1/3	−1/3	+1/3
S07	−1/3	+1/3	+1/3
S08	+1/3	+1/3	+1/3

model validation with three factors occupying a narrower design space than that of A01–A08 (more precisely, the range of each factor in S01–S08 is half of that from A01–A08). A01–A08 and S01–S08 represent 16 total unique experimental conditions; asparagine, glutamate, and copper are each varied at four different concentration levels. Originally, it was planned to use the data from experiments A01–A08 to train the model and then validate the model using the data from experiments S01–S08. However, such an arrangement, would lead to validating the model only once. Instead, a leave-*p*-out cross-validation approach was used by enumerating different combinations of training and validation data sets. As a result, the data sets from the two experimental designs were combined. Eighty percent of the data from the combined data sets were used to estimate the model parameters, and the remainder was used for model validation.

All experiments were carried out in fed-batch runs in 250 mL baffled-shake flasks (Corning, Oneonta, NY) with an initial working volume of 100 mL. The shake flasks were agitated on an orbital shaker at 125 rpm in an environmental chamber at  $36.5 \text{ }^\circ\text{C}$  and 5%  $\text{CO}_2$ . Cultures were seeded at  $0.5 \times 10^6 \text{ cells mL}^{-1}$  in a proprietary basal medium.

Each fed-batch run involved sampling and feeding procedures typically executed at the beginning of each culture day. A sample of the culture medium was taken daily to measure the concentrations of mAbs, cell densities, and metabolites, the glycan abundances, and the reactor operating conditions (pH, osmolality, etc.). Antibody titer was measured using a high-performance liquid chromatography (HPLC) with a prepacked Protein G immunodetection column (Applied Biosystems, Bedford, MA). VCD and TCD were measured using a Beckman Coulter Vi-Cell XR cell counter (Indianapolis, IN); cell culture metabolites (e.g., glucose, lactate, and ammonia) were measured using a Cedex Bio HT analyzer (Roche, Mannheim, Germany); glycan relative abundances were measured using a liquid chromatography–mass spectrometry (LC–MS) assay; pH and osmolality were measured using a Bioprofile Flex analyzer (Nova Biomedical, Waltham, MA).

**Table 2.** ANOVA results ( $p < .05^*$ ,  $p < .01^{**}$ ,  $p < .001^{***}$ ).

Output	Unit	Asn	Glu	Cu	Asn×Glu	Asn×Cu	Glu×Cu	Asn×Glu×Cu
VCD	10 <sup>6</sup> cells mL <sup>-1</sup>	-1.36**	0.953*	2.20***	0.718	1.22*	-1.18*	-0.583
TCD	10 <sup>6</sup> cells mL <sup>-1</sup>	0.119	-0.133	0.364*	-0.0325	-0.0475	-0.0575	-0.00417
Titer	g L <sup>-1</sup>	-0.977**	0.0133	2.62***	-0.107	0.252	0.195	-0.262
Lactate	g L <sup>-1</sup>	0.21***	-0.217***	-0.833***	0.02	-0.0833*	0.13**	0.0267
Ammonia	mM	0.0554***	-0.025***	-0.00806*	0.00244	-0.00307	-0.000993	-0.0019
Glucose	g L <sup>-1</sup>	0.0125	0.0842	-0.249***	-0.0408	-0.0775	-0.0125	0.0225
Glutamate	g L <sup>-1</sup>	0.416***	0.845***	0.0944*	-0.105*	0.00687	0.028	-0.0718
Glutamine	g L <sup>-1</sup>	0.119***	0.207***	0.227***	-0.0246	0.0514*	-0.00122	-0.00195
Alanine	g L <sup>-1</sup>	0.107*	0.157***	-0.524***	-0.00355	0.106*	0.0193	0.0662
Asparagine	g L <sup>-1</sup>	0.00711	0.00526	-0.0346***	0.00163	0.00158	9.70 × 10 <sup>-5</sup>	-0.00244
Aspartate	g L <sup>-1</sup>	0.226***	0.223***	0.207***	-0.0973	0.0689	-0.00762	-0.00112
Proline	g L <sup>-1</sup>	0.0229	0.0714*	-0.015	-0.0693	0.0938*	0.000777	0.014
Man5	1	-0.000161*	-8.31 × 10 <sup>-5</sup>	0.000585***	0.000126*	-9.25 × 10 <sup>-5</sup>	-5.53 × 10 <sup>-5</sup>	4.91 × 10 <sup>-5</sup>
G0F-GlcNAc	1	0.00021*	8.48 × 10 <sup>-5</sup>	0.000166*	-0.000187*	-2.52 × 10 <sup>-6</sup>	9.07 × 10 <sup>-5</sup>	-7.73 × 10 <sup>-5</sup>
G0	1	0.000478***	0.00023*	-0.000922***	-0.000291**	2.73 × 10 <sup>-5</sup>	0.000101	-4.05 × 10 <sup>-5</sup>
G0F	1	0.0187***	-0.00251	-0.00451	-0.00543	0.00473	0.00445	-0.00321
G1F	1	-0.0166***	0.0022	0.00402	0.00517	-0.0041	-0.0042	0.00296
G2F	1	-0.00193***	7.54 × 10 <sup>-5</sup>	0.00063*	0.000615*	-0.000594*	-0.00039	0.000307

Two feeds (feed A and feed B) were then added to the culture immediately after sampling. Sixteen different feed A medium recipes with different levels of asparagine, glutamate, and copper were created according to the two experimental designs in Table 1. Feed A was added to the reactor between culture days 3 and 13 based on the cell culture working volume; feed B containing a high level of glucose was added between culture days 3 and 16 such that the glucose concentration in the culture is at least 3 g L<sup>-1</sup> after feeding.

Note that the macroscale model in Equation (5) is only valid for constant-volume systems. Conditions in the fed-batch process are different from batch-mode operation conditions because of the periodic feeding and sampling activities mentioned above for the former, during which, the system state,  $\mathbf{x}$ , updates discontinuously and non-smoothly due to the volume expansion. Therefore, fed-batch runs are simulated in segments: i) the constant-volume, batch-mode dynamics are modeled by solving Equation (5), and ii) the system state is updated instantaneously after volume expansion according to the amount of samples taken from and feeds added to the culture. The overall model,  $f$ , was adjusted such that the state variable,  $\mathbf{x}$ , is updated daily according to the discrete-time dynamics equation below

$$\mathbf{x}(k+1) = \mathbf{g}(\mathbf{x}(k); \mathbf{u}, \Theta) \quad (k = 0, \dots, N-1) \quad (11)$$

where the state on the  $(k+1)$ th day,  $\mathbf{x}(k+1)$ , is a function of the previous-day state and inputs. The discrete-time dynamics function,  $\mathbf{g}$ , describes both the abrupt state change from volume expansion and the continuous evolution from constant-volume, batch-mode kinetics.

#### 2.4. Parameter Estimation

The model parameters,  $\Theta$ , are estimated by minimizing a function of the difference between cell culture and glycosylation

measurements on  $N$  culture days,  $\mathbf{x}(1), \dots, \mathbf{x}(N)$ , and the corresponding model predictions,  $\hat{\mathbf{x}}(1), \dots, \hat{\mathbf{x}}(N)$ , obtained from Equation (11). The specific objective function,  $V$ , used for parameter estimation is the weighted sum of squared residuals

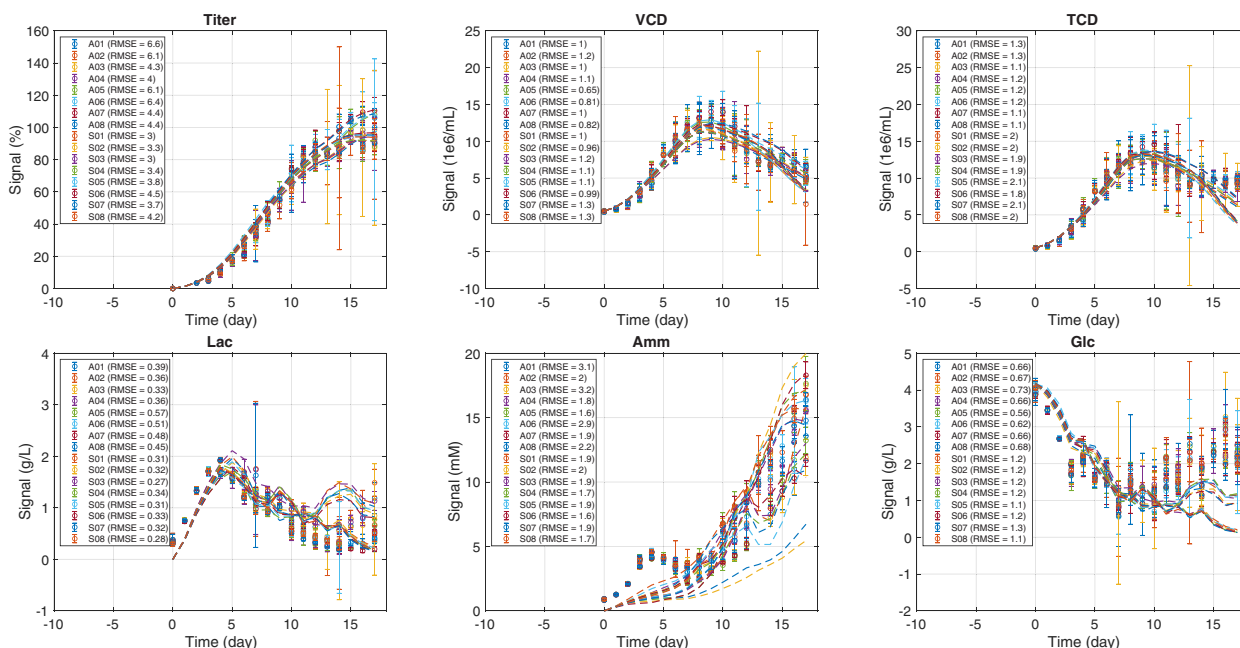
$$V(\Theta; \mathbf{u}, \mathbb{X}) = \sum_{i=1}^n \sum_{k=1}^N w_i^2 (x_i(k) - \hat{x}_i(k))^2 \quad (12)$$

where  $\mathbb{X} = \{\mathbf{x}_i(k) | k = 1, \dots, N\}$  is the training data set, and  $w_1, \dots, w_n$  are the weights chosen 1) to scale the  $n$  state variables,  $x_1, \dots, x_n$ , according to their disparate magnitudes and 2) to assign higher priority to variables of greater interest, such as mAb concentration and cell densities (see Supporting Information for the detailed weighting).

### 3. Results and Discussion

#### 3.1. Statistical Analysis

We performed an analysis of variance (ANOVA) to determine the statistical significance of the computed effects of asparagine, glutamate, and copper levels in media on the final, end-of-run (EoR) mAb titer, cell densities, metabolites, and glycan distribution. The results are unsurprising: all three factors showed significant effects ( $p < 0.05$ ) on mAb productivity, cell metabolism, or glycosylation (Table 2). The level of asparagine affected VCD, mAb titer, metabolites, and all the measured glycans (one can disregard the effect of asparagine on aspartate as asparagine is readily converted to aspartate via asparaginase in order for asparagine and aspartate to enter the TCA cycle<sup>[20,21]</sup>). Duarte et al.<sup>[22]</sup> found that continued asparagine availability led to a shift in nitrogen metabolism, which resulted in increased ammonia, glutamate, and glutamine secretion. These findings correspond to the ANOVA results in this study that also show a



**Figure 3.** In-sample validation of macroscale model (time series), error bars are the 95% confidence intervals, the root mean squared error (RMSE) is used to quantify the accuracy of model prediction.

significant and positive correlation between asparagine level and ammonia, glutamate, and glutamine levels. Similarly, the glutamate level affected VCD, lactate, ammonia, a few amino acids, and G0. Cultures supplemented with elevated levels of copper showed significant variations in all measured quantities except proline, G0F, and G1F. Specifically, copper has a significant, positive effect on titer and a significant, negative effect on lactate—a phenomenon consistently observed in other CHO cell culture processes.<sup>[12,13,23,24]</sup> A plausible explanation is that the increased copper level is known to drive lactate consumption. Copper deficiency reduces cytochrome c oxidase activity, limiting the ability of cells to produce ATP via oxidative phosphorylation. As a result, cells switch to aerobic glycolysis to generate ATP, causing increased lactate production, which affects other metabolic processes.<sup>[12,23]</sup>

The interactions (2- or 3-way) among asparagine, glutamate, and copper, on the other hand, did not show a consistently significant impact on either the mAb production or the final glycan distributions. Such a lack of significant higher-order interaction effects provides empirical support for the model formulation, in which  $\Delta$  in Equation (4) is represented as a linear function of  $\mathbf{u} - \mathbf{u}_0$ . It is possible for the interaction effects of the supplemental inputs on the process dynamics to be significant. One can modify the supplemental input effect model to capture such interaction effects by adding bilinear terms. We recommend always performing a statistical analysis to determine the appropriate structure of  $\Delta$ .

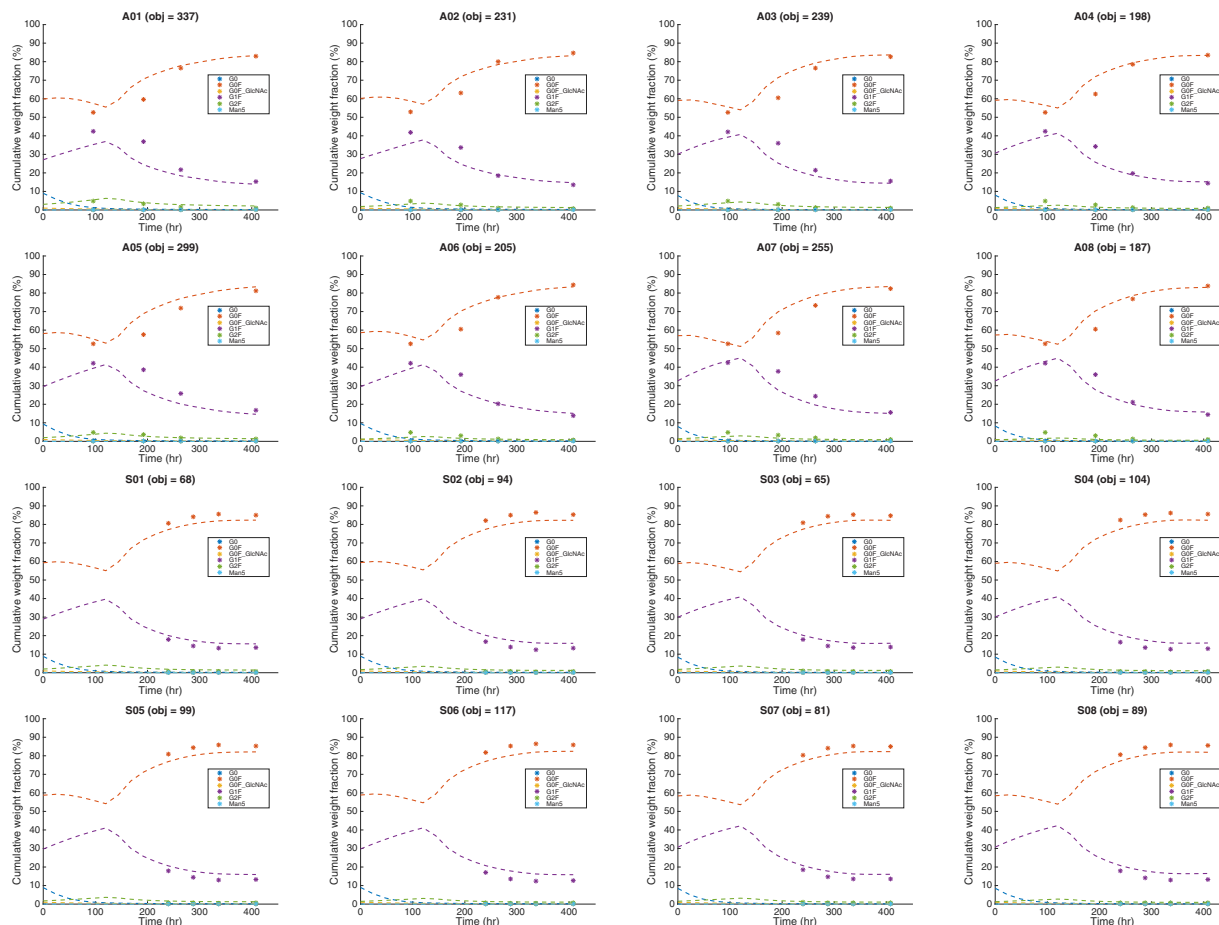
### 3.2. Model Validation

We validated the model prediction against data using two different approaches. First, we used all 16 data sets (A01–A08 and S01–S08 in Table 1) to train the model,  $f$ , and to test its perfor-

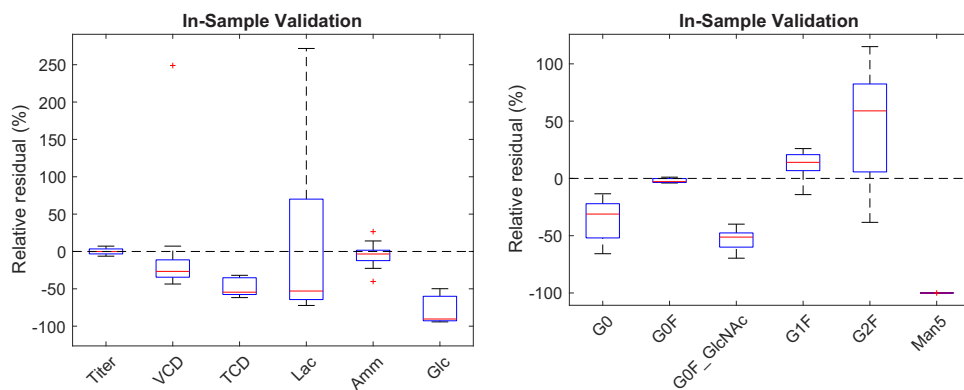
mance, that is, in-sample model validation (see the estimated parameters in the Supporting Information). We only performed an in-sample validation of the base model,  $f_0$ , because the main focus of this study is not the development of  $f_0$  and the estimation of its parameters; rather our focus is on using the modular modeling approach to adapt such a base model to new processes. **Figure 3** shows the model prediction (dashes) superimposed on measurements (dots) for mAb titer (scaled), VCD, TCD, lactate, ammonia, and glucose. **Figure 4** shows a comparison of the model prediction and measured glycan distribution for each unique experiment condition. Note the apparent shift in trend in glycan distribution where G0F appears to decline initially before starting to increase around culture day 5 (120 h). Due to low cell count and product concentration at the beginning of a fed-batch run, early glycan measurements are either unavailable or unreliable. While the model is consistent with data, we are unable to verify or disprove the trend definitively at this time. **Figure 5** is a plot of the EoR relative predictive residuals of titer, cell densities, metabolites, and glycans (the closer to 0 the better).

The in-sample validation shows that the model can predict accurately such variables as mAb titer, cell densities, and also G0F, G1F fractions, which are closely related to antibody productivity and product quality. By adjusting the weights in the objective function,  $V$ , in Equation (12), one can enhance the prediction accuracy for certain quantities while lowering the accuracy for others. Weights should be assigned based on the overall modeling or process development objective. In this case, lactate, ammonia, glucose, and some glycans were given relatively low weights because they are not directly related to mAb productivity or product quality.

Subsequently, to test how well the supplemental input model,  $\Delta$ , can make predictions under new conditions, we performed a leave- $p$ -out cross-validation exercise, that is, an out-of-sample validation. Leave- $p$ -out is an exhaustive cross-validation method



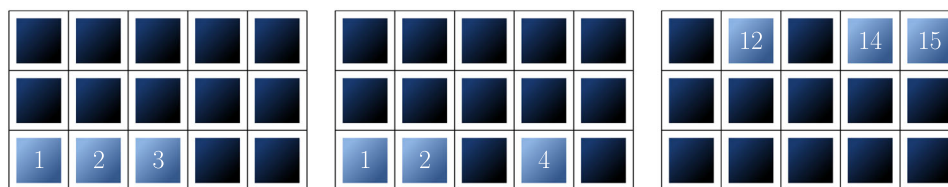
**Figure 4.** In-sample validation of microscale model (time series).



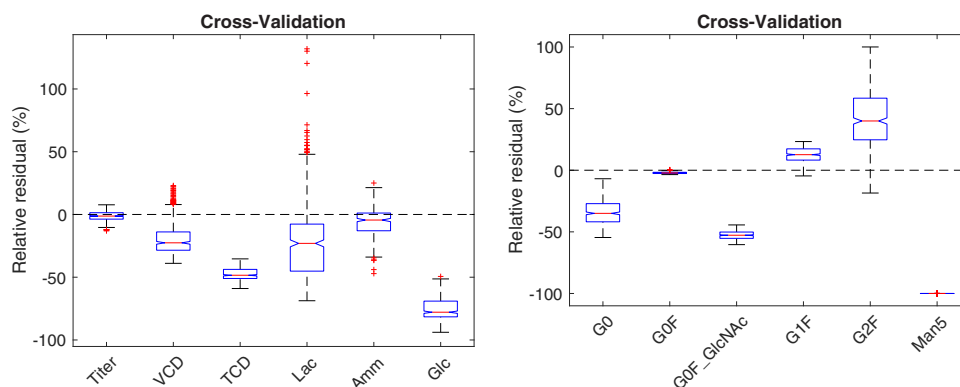
**Figure 5.** In-sample validation EoR relative residuals.

that is based on using all possible ways to divide the original data into training and validation data sets. Such a method involves training and validating the model ( $m$ ) times, where  $m = 15$  is the number of the data sets with the supplemental process inputs deviating from the baseline levels, and  $p$  is the number of data sets selected to validate the model, and the remaining ( $m - p$ ) is the number of data sets selected to train the model. Here, we set  $p = 3$  because the number of validation runs is sufficiently large

(455 times), and at the same time, the validation process can be completed within a reasonable time frame. In each of the 455 repeated training-testing runs, 12 out of a total of 15 data sets were selected to estimate the parameters,  $\beta$ , of the supplemental input model,  $\Delta$ , and the model was then tested against the remaining 3 data sets (**Figure 6**). The model prediction based on the testing set was recorded at the end of each validation run.



**Figure 6.** A schematic representation of the 1st, 2nd, and 455th runs of the leave-3-out cross-validation (the dark blue squares represent the training set, and the light blue squares represent the testing set).



**Figure 7.** Out-of-sample cross-validation EoR relative residuals.

**Figure 7** shows a plot of EoR relative residuals from all 455 cross-validation runs. As was the case with in-sample validation, the cross-validation also has low EoR relative residuals for mAb concentration, cell densities, and G0F, G1F fractions. The cross-validation performance implies that the supplemental input model,  $\Delta$ , is able to produce reasonably accurate predictions against new data.

#### 4. Conclusions

In manufacturing therapeutic mAbs, it is crucial to be able to predict how process inputs, such as media recipes and operating conditions, affect process performance attributes. Here, we study the effect of asparagine, glutamate, and copper levels in media on antibody productivity and glycosylation using a modular modeling approach. A supplemental input model,  $\Delta$ , was developed to capture these effects by augmenting an existing mechanistic model,  $f_0$ . The augmented model,  $f = f_0 + \Delta$ , is capable of producing accurate predictions under different asparagine, glutamate, and copper concentrations in media. We performed cross-validation to ensure that the predictions obtained from the supplemental input model,  $\Delta$ , can generalize to new data. This modular modeling approach can be used for efficient model development while avoiding the often time-consuming task of modifying a model's structure. The mechanistic component,  $f_0$ , is used to store structured information of the process kinetics and to confine model predictions to a reasonable region; the data-driven component,  $\Delta$ , on the other hand, is used to capture the “known unknowns”—effect of supplemental inputs on process dynamics that may not be well understood—to expand the prediction capability of the base model. While it is generally applicable to any supplemental inputs one chooses to add to a

model, the modular modeling approach is expected to be particularly useful in a model-based process development setting when the mechanistic understanding of supplemental process inputs is lacking, and where gaining such understanding might not be practical or necessary for production purposes. In other cases, when some mechanistic insights into how certain supplemental inputs affect the process outputs are available, one might prefer to adopt a more mechanistic approach (e.g., building flux-based models to use extracellular metabolite concentrations to predict intracellular nucleotide sugar levels, which are subsequently used to predict the glycan distribution<sup>[25]</sup>). It is the modeler's job to select the most appropriate modeling methodology after considering time constraint, data availability, process complexity, and other relevant factors.

The modular modeling approach has its limitations, however. First, the supplemental input model,  $\Delta$ , approximates the effect of supplemental process inputs by a linear function which may not capture the effect accurately, especially when the deviation from baseline is substantial. In such a case, resetting the baseline,  $u_0$ , may be necessary. Second, a data-driven model often has more parameters than a mechanistic model that describes the same process. As a result, the modular modeling approach may introduce more parameters to the original model than would otherwise be the case if the model structure was modified instead, creating a potential problem of model “overfitting”. Reducing the number of parameters based on sensitivity analysis of the supplemental input model may be needed to avoid overfitting.

We envision the following potential usage and expansion of the model. First, we plan to use the model to improve media recipe design. By formulating the design as an optimization problem, one can use the model to adjust amino acid and trace metal concentrations to meet productivity and product quality targets.



Second, we plan to introduce additional supplemental process inputs to the current model. Cell culture process operating conditions such as pH, dissolved oxygen, dissolved carbon dioxide, agitation, temperature, and pressure also affect process performance and product quality in important ways.<sup>[8]</sup> Components such as asparagine, glutamate, and copper in nutrient feeds, which we studied in this paper, are part of the media recipes and cannot be altered easily during a fed-batch run. In contrast, process operating conditions can be adjusted in real time, making them potential candidates as manipulated variables for on-line control of mAb productivity and glycosylation—a strategy that has not yet been attempted in the biopharmaceutical industry.<sup>[26]</sup> Third, we plan to include the cell line effect in the model. CHO cell dynamics are specific to the cell line used. Therefore, adapting an existing model to new cell lines requires capturing the effect of the specific cell line on the process dynamics. Unlike supplement concentrations or process operating conditions, the “cell line” is not a quantitative property and cannot be modeled as a supplemental input directly. Instead, quantifiable characteristics of a cell line, such as its specific productivity, may be modeled as the supplemental input to capture the cell line effect. A thorough literature review and relevant experiments are needed to identify such quantitative cell line proxies. Being able to model the cell line effect potentially allows us to use preliminary cell line development data to update the model for effective design and control of processes with new cell lines.

## Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

## Acknowledgements

The authors would like to thank Kendra Kester of Janssen R&D for providing quantitative analysis of amino acids and to Kerry Robinson of Janssen R&D for providing analytical characterization of glycan species for these studies.

## Conflict of Interest

The authors declare no conflict of interest.

## Keywords

biopharmaceutical manufacturing, glycosylation, model-based process development, multiscale model, parameter estimation

Received: June 1, 2020

Revised: July 22, 2020

Published online: September 23, 2020

- [1] FDA, *United States Department of Health and Human Services* **2004**.
- [2] R. Jefferis, *Biotechnol. Prog.* **2005**, *21*, 11.
- [3] A. Beck, E. Wagner-Rousset, M.-C. Bussat, M. Lokteff, C. Klinguer-Hamour, J.-F. Haeuw, L. Goetsch, T. Wurch, A. V. Dorselaer, N. Corvaia, *Curr. Pharm. Biotechnol.* **2008**, *9*, 482.
- [4] R. Dietmar, L. T. Max, *Glycobiology* **2015**, *25*, 1325, <https://doi.org/10.1093/glycob/cwv065>.
- [5] C. K. Crowell, G. E. Grampp, G. N. Rogers, J. Miller, R. I. Scheinman, *Biotechnol. Bioeng.* **2007**, *96*, 538.
- [6] P. Hossler, in *Genomics and Systems Biology of Mammalian Cell Culture*, Springer, Berlin Heidelberg **2012**, pp. 187–219. [https://doi.org/10.1007/10\\_2011\\_113](https://doi.org/10.1007/10_2011_113)
- [7] M. M. St Amand, D. Radhakrishnan, A. S. Robinson, B. A. Ogunnaike, *Biotechnol. Bioeng.* **2014**, *111*, 1957.
- [8] F. Li, N. Vijayasankaran, A. (Yijuan) Shen, R. Kiss, A. Amanullah, *mAbs* **2010**, *2*, 466, <https://doi.org/10.4161/mabs.2.5.12720>.
- [9] M. Gawlitzek, T. Ryll, J. Lofgren, M. B. Sliwowski, *Biotechnol. Bioeng.* **2000**, *68*, 637.
- [10] P. Xu, X.-P. Dai, E. Graf, R. Martel, R. Russell, *Biotechnol. Prog.* **2014**, *30*, 1457, <https://doi.org/10.1002/btpr.1957>.
- [11] T. Kaschak, D. Boyd, F. Lu, G. Derfus, B. Kluck, B. Nogal, C. Emery, C. Summers, K. Zheng, R. Bayer, A. Amanullah, B. Yan, *mAbs* **2011**, *3*, 577, <https://doi.org/10.4161/mabs.3.6.17959>.
- [12] S. Nargund, J. Qiu, C. T. Goudar, *Biotechnol. Prog.* **2015**, *31*, 1179.
- [13] I. H. Yuk, S. Russell, Y. Tang, W.-T. Hsu, J. B. Mauger, R. P. S. Aulakh, J. Luo, M. Gawlitzek, J. C. Joly, *Biotechnol. Prog.* **2015**, *31*, 226, <https://doi.org/10.1002/btpr.2004>.
- [14] I. Jimenez del Val, J. M. Nagy, C. Kontoravdi, *Biotechnol. Prog.* **2011**, *27*, 1730.
- [15] C. Kontoravdi, S. P. Asprey, E. N. Pistikopoulos, A. Mantalaris, *Comput. Chem. Eng.* **2007**, *31*, 392.
- [16] F. J. Krambeck, M. J. Betenbaugh, *Biotechnol. Bioeng.* **2005**, *92*, 711.
- [17] P. N. Spahn, A. H. Hansen, H. G. Hansen, J. Arnsdorf, H. F. Kildegaard, N. E. Lewis, *Metab. Eng.* **2016**, *33*, 52.
- [18] S. Sha, Z. Huang, Z. Wang, S. Yoon, *Curr. Opin. Chem. Eng.* **2018**, *22*, 54, <https://doi.org/10.1016/j.coche.2018.08.010>.
- [19] C. Kontoravdi, I. del Val Jimenez, *Curr. Opin. Chem. Eng.* **2018**, *22*, 89, <https://doi.org/10.1016/j.coche.2018.08.007>.
- [20] J. Dean, P. Reddy, *Biotechnol. Bioeng.* **2013**, *110*, 1735.
- [21] H. Lin, R. W. Leighty, S. Godfrey, S. B. Wang, *Biotechnol. Prog.* **2017**, *33*, 891, <https://doi.org/10.1002/btpr.2472>.
- [22] T. M. Duarte, N. Carinhas, L. C. Barreiro, M. J. T. Carrondo, P. M. Alves, A. P. Teixeira, *Biotechnol. Bioeng.* **2014**, *111*, 2095.
- [23] J. Luo, N. Vijayasankaran, J. Autsen, R. Santuray, T. Hudson, A. Amanullah, F. Li, *Biotechnol. Bioeng.* **2012**, *109*, 146.
- [24] I. H. Yuk, J. D. Zhang, M. Ebeling, M. Berrera, N. Gomez, S. Werz, C. Meiringer, Z. Shao, J. C. Swanberg, K. H. Lee, J. Luo, B. Szperalski, *Biotechnol. Prog.* **2014**, *30*, 429.
- [25] S. Sha, Z. Huang, C. Agarabi, S. Lute, K. Brorson, S. Yoon, *Processes* **2019**, *7*, 227.
- [26] M. M. St. Amand, K. Tran, D. Radhakrishnan, A. S. Robinson, B. A. Ogunnaike, *PLoS ONE* **2014**, *9*, e87973, <https://doi.org/10.1371/journal.pone.0087973>.