Received: 18 August 2021 | Revised: 26 October 2021 | Accepted: 13 November 2021

DOI: 10.1002/bit.27994

ARTICLE

BIOTECHNOLOGY AND BIOENGINEERING WILEY

A genome-scale nutrient minimization forecast algorithm for controlling essential amino acid levels in CHO cell cultures

Yiqun Chen[1] | Xiao Liu[1] | Ji Young L. Anderson[2] | Harnish Mukesh Naik[1] | Venkata Gayatri Dhara[1] | Xiaolu Chen[1] | Glenn A. Harris[2] | Michael J. Betenbaugh[1]

[1]Department of Chemical and Biomolecular Engineering, Johns Hopkins University, Baltimore, Maryland, USA

[2]Research and Development, 908 Devices Inc., Boston, Massachusetts, USA

Correspondence

Michael J. Betenbaugh, Department of Chemical and Biomolecular Engineering, Johns Hopkins University, Baltimore, MD 21218, USA.
Email: beten@jhu.edu

Funding information

National Science Foundation, Grant/Award Number: 1624684

Abstract

Mammalian cell culture processes rely heavily on empirical knowledge in which process control remains a challenge due to the limited characterization/understanding of cell metabolism and inability to predict the cell behaviors. This study facilitates control of Chinese hamster ovary (CHO) processes through a forecast-based feeding approach that predicts multiple essential amino acids levels in the culture from easily acquired viable cell density data. Multiple cell growth behavior forecast extrapolation approaches are considered with logistic curve fitting found to be the most effective. Next, the nutrient-minimized CHO genome-scale model is combined with the growth forecast model to generate essential amino acid forecast profiles of multiple CHO batch cultures. Comparison of the forecast with the measurements suggests that this algorithm can accurately predict the concentration of most essential amino acids from cell density measurement with error mitigated by incorporating off-line amino acids concentration measurements. Finally, the forecast algorithm is applied to CHO fed-batch cultures to support amino acid feeding control to control the concentration of essential amino acids below 1–2 mM for lysine, leucine, and valine as a model over a 9-day fed batch culture while maintaining comparable growth behavior to an empirical-based culture. In turn, glycine production was elevated, alanine reduced and lactate production slightly lower in control cultures due to metabolic shifts in branched-chain amino acid degradation. With the advantage of requiring minimal measurement inputs while providing valuable and in-advance information of the system based on growth measurements, this genome model-based amino acid forecast algorithm represent a powerful and cost-effective tool to facilitate enhanced control over CHO and other mammalian cell-based bioprocesses.

KEYWORDS

amino acid control, CHO cell, fed-batch culture, model predictive control

1 | INTRODUCTION

Modern bio-pharmaceutical processes harness mammalian cells such as the Chinese Hamster Ovary (CHO) cells to produce therapeutic proteins of high commercial value (Kildegaard et al., 2013). In these processes, large-scale productions are usually achieved through fed-batch or perfusion operations (Bibila & Robinson, 1995; Hiller et al., 2017; Huang et al., 2010; Konakovsky et al., 2017; Meuwly et al., 2006). Successful optimization in fed-batch and perfusion strategies requires careful considerations of nutrient levels to supply sufficient levels of amino acids and other substrates to maintain high growth rate and productivity, while avoiding accumulation of insoluble or harmful and inhibitory by-products caused by over-feeding (Konakovsky et al., 2017; Lu et al., 2013). Optimization of fed-batch or perfusion processes can be based on empirical approaches such as the widely used Design of Experiments to find the optimal experimental parameters for certain defined cell lines, instruments and materials (Konakovsky et al., 2017), and dynamic feeding control approaches that rely on analytical technologies or well-designed algorithms to address the real-time nutrient demand of cells or to keep process parameters within the optimal range (André et al., 2017; Gagnon et al., 2011; Wlaschin & Hu, 2006). For example, the HIP-DOG (HI-end pH-controlled Delivery Of Glucose) method reduces the accumulation of lactate by feeding glucose according to the detected rising of pH in culture (Gagnon et al., 2011; Hiller et al., 2017) while another approach uses oxygen uptake rate measurements as an effective input to control glutamine feed rates (Aehle et al., 2011). In addition, technologies using a Raman spectroscopy probe or near infrared can be applied to control nutrient concentrations to achieve on-line monitoring and control of glucose levels in CHO cell cultures (Berry et al., 2016; Li et al., 2018). Currently, the precise monitoring and control of amino acid levels in mammalian cell cultures have been proven to be beneficial but is more challenging due to the typical requirement of laborious and often time-consuming daily off-line measurements (Mulukutla et al., 2017, 2019). The tediousness of these approaches has encouraged users to develop alternative control approach that may be easy to implement to maintain amino acid levels at desirable levels in animal cell culture.

While many current feeding and control strategies are based on empirical or sophisticated analytical and measurement techniques, alternative and complementary successful approaches have utilized modeling tools to guide process control (Ben Yahia et al., 2015; Craven et al., 2014; Teixeira et al., 2007). Mathematical models can be an effective tool for guiding mammalian culture operations due to their capability to describe the behavior of complex systems, making it possible to predict future states or optimal process parameters without conducting time-consuming experiments (Ben Yahia et al., 2015). Models designed for cell culture process control are expected to be able to describe enough details about cell physiology or behaviors to accurately monitor and predict the process state but at the same time simple enough to allow fast and straightforward process control (Sommeregger et al., 2017). Such models can include mechanistic models such as substrate-based Monod-type kinetic models or complex metabolic models based on genome scale representations, regression models such as polynomial models, as well models based on data analytics approaches (Almquist et al., 2014; André et al., 2017; Bayrak et al., 2015; Martínez et al., 2020). Monod-type growth models for mammalian cells can be highly effective in describing cell growth but their development can be laborious due to the requirement to obtain the kinetics parameters for numerous inputs in the media including glucose and many or all amino acids. Simple Monod-type equations taking into account only one or a few key nutrients (e.g., glucose) are possible but they may lack the robustness needed for describing the impact of other inputs such as amino acids (Goudar, 2012; López-Meza et al., 2016).

Metabolic models, as another type of mechanistic models, are typically useful for investigating metabolism and reaction fluxes within an organism. Furthermore, genome-scale metabolic models can be used to predict the specific growth rate of CHO cells by solving a linear program with more than 20 inputs of uptake and secretion rates of metabolites, via the commonly used "biomass objective function" approach (Chen et al., 2019; Hefzi et al., 2016). However, in cell culture operations, cell density is often much easier to measure as compared to many other measurements used as model inputs. As an alternative, our group has demonstrated that genome-scale metabolic models can be used to predict amino acid concentrations from cell growth or density data via an "essential nutrient minimization" (ENM) approach that can be useful in mammalian bioprocesses including CHO cells. This approach can identify and generate more than 10 essential amino acid outputs from one easy-to-measure cell density input for CHO cell culture (Chen et al., 2019). Thus, instead of being used to predict growth rates, the genome-scale metabolic models can also potentially be applied to forecast the nutrient concentration levels throughout the duration of batch and fed batch cell culture processes. However, to apply the model for predicting nutrient uptake rates in a predictive framework, cell growth and cell densities must be estimated for at least one future time intervals. For this additional parameter, regression models can be applied due to their simplicity and capability to describe time- and condition-dependent system behaviors, in different biological and other systems (Jaqaman & Danuser, 2006). For example, a regression model-based approach was used to predict metabolite uptake-secretion rates metabolic uptake and secretion rates from specific growth rates of CHO cells (Ben Yahia et al., 2017). Growth dynamics of cells can be readily described using empirical-based regression models including exponential and logistic approaches, whose model construction can account both cell proliferation and environmental limitations (Goudar et al., 2005; Goudar, 2012; Henry et al., 2008). Due to their limited requirements and relative accuracy, exponential and logistic models can be appropriate candidates for simple, effective growth estimators as part of a cell culture process characterization and control algorithms (Naderi et al., 2011).

It is believed that the development of biomanufacturing processes will very likely involve the combination of real-time analytical monitoring systems and mathematical models known as virtual monitoring systems (Park et al., 2021). In this study, we focus on applying modeling tools to

guide mammalian cell culture process control by predicting key process parameters including cell growth and amino acid concentrations. The model-based tool can be readily implemented; in other words, such a tool will require a limited number of measurement inputs that are easily acquirable and be able to predict and control a large number of process parameters which are more challenging and time consuming to measure. To accomplish this goal, we have integrated the genome-scale metabolic model-based ENM uptake rate prediction algorithm developed previously by our group with a robust regression model for predicting cell growth into a control algorithm that forecasts and controls the future concentration of essential amino acids from viable cell density (VCD) measurements. In contrast to the conventional control strategies, this forecast-control algorithm is designed to implement nutrient control by predicting future nutrient needs of the cultures and offers the capability to easily control species-specific essential amino acids levels at a desired optimum without the strict requirement of detailed analytical measurements by utilizing rapidly acquired cell density estimates together with a model of cell physiology. As a demonstration of the capability of this approach, we evaluate the reliability of this algorithm in forecasting amino acid levels in CHO batch cultures and demonstrate its applicability in controlling amino acids in CHO fed-batch cultures. First, cell growth and essential amino acid level forecast accuracy of the developed algorithm is evaluated using CHO batch culture datasets collected from two different sources by inputting only the VCD data. Then, the forecast algorithm is used to guide the feeding of three amino acids in CHO fed-batch cultures and controls their concentration at low levels before the death phase using initial amino acid concentration and VCD as the only inputs. This approach can be especially useful when particular amino acids are to be maintained at low levels to optimize culture performance (Mulukutla et al., 2017; Mulukutla et al., 2019). Overall, the forecast algorithm exhibits reliable capabilities for predicting and controlling essential amino acid levels in CHO cell batch and fed-batch culture processes.

2 | RESULTS AND DISCUSSION

2.1 | Logistic function well describes the growth behavior of CHO cell batch cultures

To achieve the goal of forecasting amino acid levels from VCD measurements in CHO cell cultures, a growth prediction method is first required to estimate the effective growth rate and VCD at specific future time points. Equation fitting using exponential and logistic growth functions were considered as approaches to forecast the growth of CHO cells and the feasibility of this method was tested with several batch growth datasets for CHO cells from two different published studies. The first study published previously by our group (Chen et al., 2019) includes growth and concentration data of eight essential amino acids for an immunoglobulin G (IgG)-producing CHO-K1 cell line grown in 125 ml shake flasks and cultured in five different media sets (dataset A1–A5). The second study (Martínez et al., 2015) reports growth and the concentration profiles for 22 metabolites including eleven essential amino acid from four batch cultures of an IgG-producing CHO-K1-derived cell line grown in 1 L shake flasks at standard conditions (37°C) and also at reduced temperatures (32°C) (dataset B1–B4).

Initially, experimentally measured growth profiles were compared against exponential and logistic growth curves generated by a least-squares fitting to investigate if the two selected growth functions can effectively describe CHO cell growth. An additional constraint which enforces the curve fitting to cross the initial data point ("fixed initial") was also evaluated for both equations. For cases without the "fixed initial" constraint, curve fitting using the exponential equation produces R^2 values between 92.6% and 99.6% for datasets A1–A5, and between 87.4% and 92.8% for datasets B1–B4, as shown in Figure 1 and Figures S1–S8. Alternatively, logistic curves fitted without fixing the initial data points demonstrated R^2 values >99% for most of the datasets except A5 and B2, which still exhibited >98% R^2 values. This observation indicates that compared to the exponential equation, the logistic equation appears to describe the growth behavior of both CHO cell cultures more effectively for all of the representative batch culture cases. As an example, based on the fitting results for datasets B1 to B4 (Figure 1 and Figures S6–S8), the exponential function does not account for the decrease in growth rate occurring after the early growth periods observed for those cases, however, this shift in growth rate is well captured by the logistic equation. On the other hand, for datasets A2–A5 in which the growth profiles show exponential behavior, the logistic function fits equally well since this function is also capable of describing exponential growth profiles.

For all cases the curves fitted with the "fixed initial" constraint result in R^2 values lower than that of the curves fitted without this constraint, with a more significant impact on exponential curve fitting as compared to logistic curve fitting. Specifically, for exponential fitting, R^2 values for datasets A1–A5 is lowered from a range between 81.7% and 99.1% (compared to 92.6%–99.6%) and is reduced to between 62.1% and 73.6% for datasets B1–B4 (compared to 87.4%–92.8%; see Figure 1 and Figures S1–S8). Alternatively, R^2 values for logistic curve fitting drops only marginally with the constraint, resulting in a decrease in R^2 values of less than 0.8% for all the different cases. This implies that despite its advantage of reducing the number of parameters needed to fit (N_0 in Equation (2), Section 4), forcing the curve to cross the initial data point also exerts a negative impact on the flexibility of curve fitting for CHO cell growth curves.

In addition to exponential and logistic equations, others such as polynomial equations are also able to describe known growth behavior and such equations fits may be used as an extrapolation approach to forecast the future state of cell growth (López-Meza et al., 2016). However, despite the capability to fit the growth profile well, there are a few disadvantages which can make polynomial fits challenging for mammalian cell growth forecasting. For example, it is difficult to set or adjust the degree of the polynomial to avoid underfitting and overfitting when a different number of data points are available at different phases of the culture. Extrapolating via polynomial curve fitting, especially for high-order polynomial equations, can be very sensitive to the last point of regression, and thus extrapolation of future time points can be significantly biased if the last

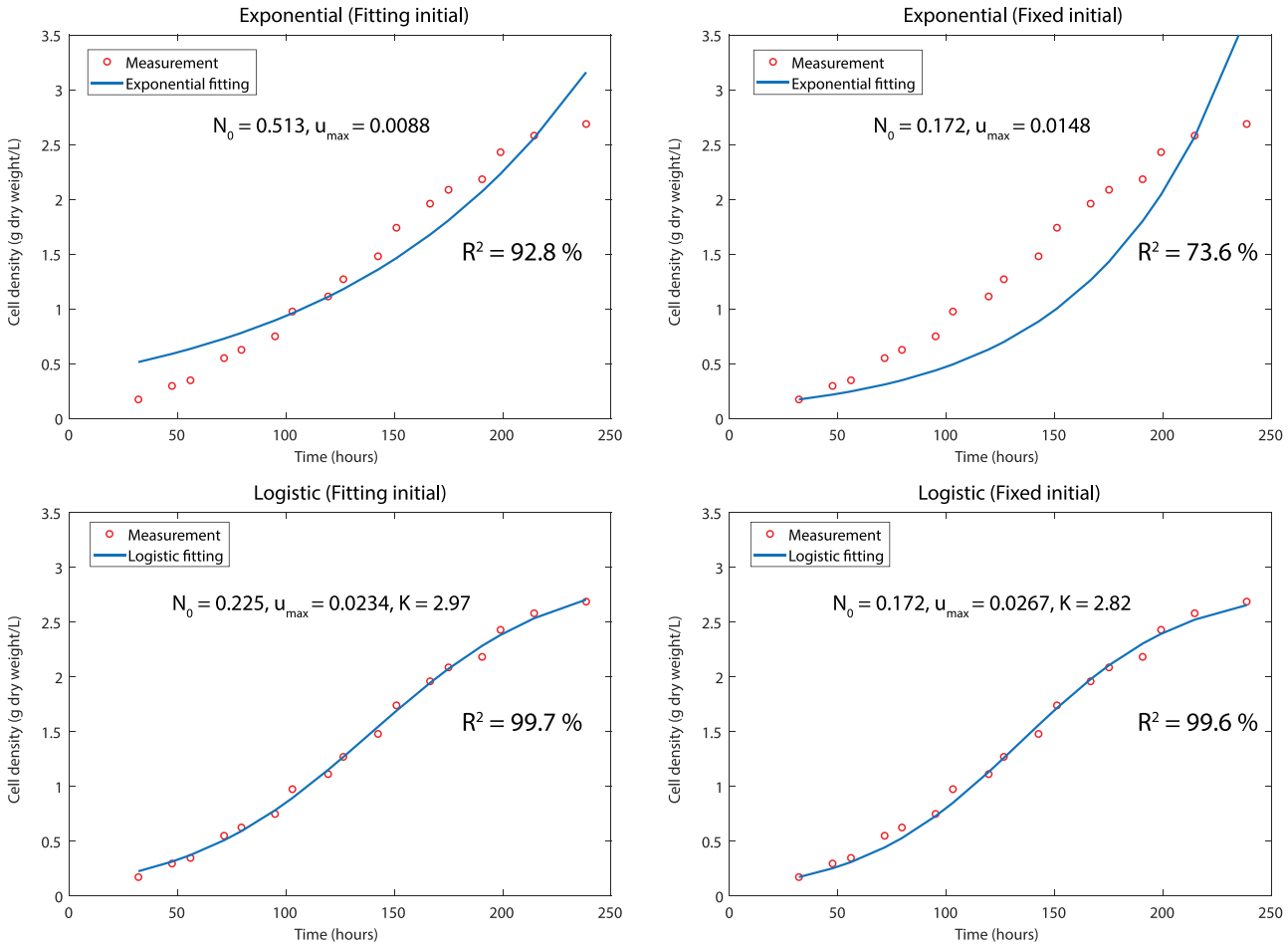

FIGURE 1 Experimental growth data versus curves fitting using exponential and logistic functions for dataset B1. "Fixed initial" stands for curved fitted with the constraint that the fitting curves have to cross the initial data point, and "fitting initial" stands for fitting without this constraint. Results for other datasets are shown in Supporting Information Figures

point is not measured accurately (Hawkins, 2004). As demonstrated in previous studies (Goudar et al., 2005; Goudar, 2012), polynomial fits are also likely to be affected by outliers and fail to characterize the initial and final stage of the CHO cell culture which makes it problematic to effectively estimate specific growth or metabolite exchange rates at current and future time points. Thus, rather than implementing fitting approaches that solely depict the trend of change of measured data points, logistic and exponential equations which characterize the cell behavior with biologically meaningful parameters represent preferable options for the regression-based growth prediction approach in the current study.

2.2 | An algorithm combining growth prediction and ENM approach forecasts VCD and essential amino acid levels in batch cultures

After confirming the feasibility of using an equation fitting approach to predict CHO cell growth, the next step is to implement this approach to develop a forecast algorithm that predicts future cell density as well as amino acid levels from current cell density measurements for CHO cell culture processes. Such an algorithm was designed by combining growth prediction methods which estimate the cell density changes, with the ENM and static optimization approach which utilize the genome-scale metabolic model for CHO cells to estimate essential amino acid consumption according to the increase in cell density (see Section 4). In addition to the exponential and logistic equation fitting, a "two-point extrapolation" approach was also considered and tested assuming the effective growth rate is constant from the previous time point to the next time point. Unlike the exponential fitting and logistic fitting approaches which require at least three and four initial VCD measurements to generate a fitting curve, the two-point extrapolation method needs only two measurements to predict the next time point. Shown in Figure 2 is a comparison between measured and algorithm-forecasted profiles of VCD and amino acid levels for dataset B4 using the logistic curve fitting approach, given only the amino acid concentrations at the initial time point as input (see Figures S9–S34 for other datasets and exponential and two-point growth prediction approaches). Also, to demonstrate and compare the prediction approaches in forecasting

the culture behaviors, the mean absolute percentage error (MAPE) of VCD and all amino acid levels forecasting results for every different dataset and approach were calculated relative to measured values and are listed in Tables 1 and 2. For datasets A1–A5 in which most of the cell density data points follow exponential behavior, the forecast VCD and amino acid level profiles are all generally comparable to measurements (shown in Figure 2, Tables 1 and 2), however, no clear conclusion can be drawn as to which one of the three approaches (exponential fitting, logistic fitting or two-point extrapolation) generates the smallest MAPE. However, due to the fact that the number of minimal data points required to initiate each algorithm are different, the number of available data points for comparison are limited and vary between the three methods, and thus the MAPE values for these datasets in group A may not well represent the accuracy of the forecasting results. For datasets B1–B4 which describe longer culture processes and contain more frequent measurements, differences between the three approaches are more evident. Table 2 illustrates that the exponential fitting approach leads to significantly higher MAPE values than the other two approaches for all four B group datasets, due to the inability of the exponential function to predict the sigmoidal shape of the actual growth curves, as demonstrated by Figure 1, Figures S24, S27, S30, and S33. The logistic fitting approach results in the lowest MAPE values for datasets B1 and B4, and the MAPE values with the two-point approach providing slightly lower values than that of the logistic fitting approach for datasets B2 and B3. Indeed, the logistic fitting method predicts the growth trend and the corresponding amino acid levels of B group datasets well, as demonstrated in Figure 2, Figures S25, S28, and S31. Thus, the amino acid forecasting algorithm with all the three growth prediction approaches is able to predict essential amino acid levels at future time points for all the nine cases of batch cultures albeit with different levels of prediction effectiveness. Among the three growth prediction approaches, the logistic curve fitting method generates the forecasted profiles that best describe the actual culture behaviors for all considered cases; the two-point extrapolation method also shows

TABLE 1 Mean absolute percentage error (MAPE) of VCD forecasting results for the nine datasets

Dataset	MAPE of forecasted viable cell density		
	Two-point extrapolation	Exponential equation fitting	Logistic equation fitting
A1	24.67%	36.28%	29.62%
A2	39.93%	22.16%	22.47%
A3	29.68%	27.81%	31.75%
A4	60.48%	35.23%	32.11%
A5	37.10%	23.77%	30.00%
B1	13.45%	14.73%	5.76%
B2	18.85%	16.28%	9.29%
B3	9.40%	18.93%	6.38%
B4	11.26%	20.43%	4.35%

Abbreviation: VCD, viable cell density.

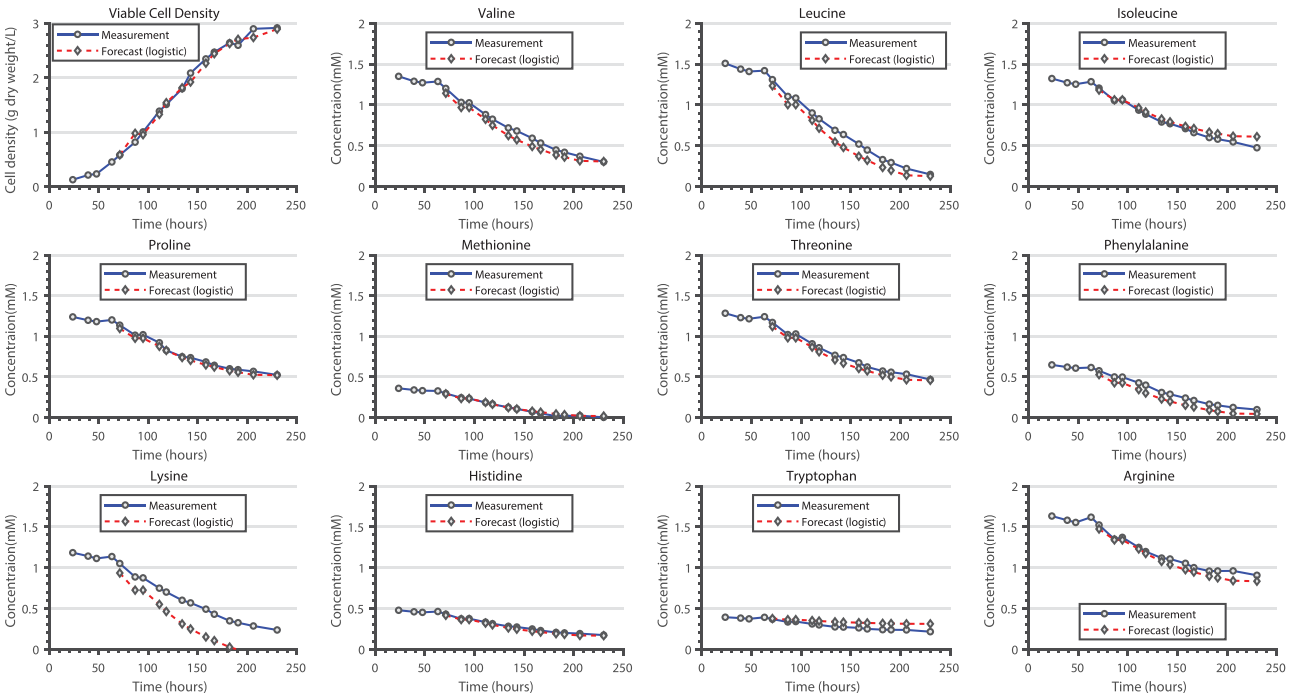

FIGURE 2 Forecast versus measurements of viable cell density and amino acid profiles for batch datasets B4, via logistic curve fitting growth prediction approach without forcing the fitted curves to cross initial data points. Each mark represents the forecast result for the time point with cell density measurements from all the past time points as inputs. Forecast results were generated assuming no amino acid concentration measurement is available as algorithm inputs

TABLE 2 Mean absolute percentage error (MAPE) of amino acid level forecasting results for the nine datasets without concentration measurement data as inputs

Dataset	MAPE of forecasted amino acid level results		
	Two-point extrapolation	Exponential equation fitting	Logistic equation fitting
A1	6.31%	13.31%	9.60%
A2	19.45%	32.19%	40.55%
A3	10.66%	11.40%	16.35%
A4	12.81%	22.76%	17.41%
A5	20.16%	32.13%	28.55%
B1	22.31%	33.29%	12.25%
B2	24.52%	57.18%	26.81%
B3	51.92%	104.69%	56.98%
B4	42.57%	117.46%	21.96%

good forecast capability while the exponential fitting method does not perform as well as the other two approaches especially when cells exhibit lag or stationary phase transitions.

Although both are considered to be effective growth prediction approaches, the logistic curve fitting and two-point extrapolation methods have different advantages when applied to the actual cell culture processes. The sigmoidal shape of logistic function makes it possible to estimate when the culture may enter stationary phase, which provides guidance to the altering feeding strategies as the cells emerge slowly from the exponential growth phase. It can be observed that cell growth profiles from most of the published CHO batch, fed batch as well as perfusion mode culture data demonstrate shapes that can be well fitted by logistic equations before cultures reaching the death phase (Bielser et al., 2018; Chen et al., 2014; Farges et al., 2008; Goudar, 2012; Gutierrez et al., 2020; Hiller et al., 2017; Huang et al., 2010; Kotidis & Kontoravdi, 2019; López-Meza et al., 2016; Meuwly et al., 2006; Pan, Streefland, et al., 2017; Robitaille et al., 2015; Wiegmann et al., 2019; Xing et al., 2010; Yongky et al., 2019). On the other hand, the two-point extrapolation approach exhibits the greatest flexibility to handle various kinds of growth behaviors because, unlike the equation fitting methods, its predictions are not predicated on the cells exhibiting any kind of defined growth patterns. Indeed, the two-point extrapolation forecast approach has been implemented successfully with nutrient minimization models by our group to control nitrate level for *Chlorella vulgaris* cultures (Li et al., 2019). Such approaches based on two consecutive points are often used to estimate specific rates due to their simplicity, but such methods can be overly sensitive to measurement error and may lead to high variability when used to predict CHO cell culture (Goudar et al., 2005; Goudar, 2012). In actual culture processes, the choice between the two forecast approaches may depend on how frequently cell density data input is acquired. In particular, logistic curve fitting is preferred when a long-term forecast is to be performed or if VCD measurements are taken relatively infrequently, because logistic curve fitting can depict the long-term fate of cell growth from even a few sporadic data points. The two-point extrapolation approach may be favored when cell density measurement data is frequently obtained and short-term culture conditions are to be estimated, since this approach can best capture the dynamic changes occurring for cell growth during the elapsed time between each iteration thanks to its high sensitivity.

Furthermore, an examination of the comparison between measurements and predictions in Figure 2 indicates that the forecast results for certain amino acids fit the experimental data more effectively than others, such as lysine in the B group datasets. Certain amino-acid specific deviations negatively impact the MAPE values and may be due to amino acid composition discrepancies between the biomass function in the genome-scale model and the specific cell line applied (Széliová et al., 2020), additional contributions arising from highly secreted products such as IgG not accounted for or with different amino acid compositions, or bias or inaccuracies in the experimental measurements of specific amino acid levels. Thus, to improve prediction accuracy and reliability, it is important to ensure key model parameters, assumptions, and measurements are realistic, accurate and representative, including the amino acid composition of biomass and recombinant protein, protein productivity, cell dry weight for specific cell lines and amino acid determinations in the culture.

2.3 | Forecast reliability can be improved by in-process amino acid concentration measurements

In the above section, we evaluated the applicability of the amino acid forecast algorithm in CHO batch cultures assuming only cell count measurements are available with time and amino acid levels known only at the beginning of the experiment. It has been shown that such continuous predictive model simulations without any correction from reliable in-process experimental data and measurements may lead to accumulations of error due to imperfect model predictions or measurement bias, especially in extended culture processes (Chen et al., 2019). Therefore, we next wanted to investigate how performance of this algorithm may be improved in situations where measurements of amino acid levels are acquired and available throughout the culture process and these can be used as additional algorithm inputs. These amino acid concentration measurements can, along with the VCD data, potentially serve as valuable corrective inputs for the predictive algorithm. However, most amino acid concentration measurements are taken off-line and such data may not be available immediately after the culture samples are taken. However, incorporation of these values can be beneficial to the accuracy of the model predictions even if the measurements become available later in the culture process. Thus, to investigate if the algorithm can be utilized in an industrial CHO culture setting, the forecast simulations for datasets A and B groups were conducted again, assuming that at a given time point, the amino acid concentration measurements from

the previous time point are available as inputs to the algorithm (in Section 4, Figure 6, "new available measurement at t_{n-1}") along with the current cell density measurement to forecast future time points. Shown in Figure 3 are the forecast profiles of datasets B4 using the logistic fitting approach with amino acid data correction, and Table 3 lists the corresponding MAPE values for all approaches and datasets. Besides the better forecast profile with measurement data correction (Figure 3 vs. Figure 2), the resulting MAPE values listed in Table 3 for almost every approach and dataset are improved compared to the values in Table 2, meaning that incorporating the amino acid concentration measurements as additional inputs, even if delayed by one measurement time point, can have a significant positive effect on the reliability of this forecast algorithm. For most cases, the improvement in forecast accuracy is most evident later in the culture periods when errors begin to accumulate. As an example, significant improvement in lysine forecast results in group B datasets can be achieved, exhibiting much closer alignment to the measured profiles compared to the uncorrected predictions.

Thus, being able to access delayed or less frequent measurement data of amino acids can still benefit the reliability of the forecast approach significantly, since even one accurate measurement at a particular time point can eliminate the error build-up from all the previous forecast results. In this way, incorporating at-line or off-line amino acid concentration measurements can represent an effective strategy to complement our model-based forecast algorithm and its applications.

2.4 | Forecast-guided feeding successfully controls amino acid level throughout the CHO fed-batch culture process

After demonstrating that the amino acid forecast approach can be used to predict concentration levels of multiple amino acids in batch

TABLE 3 Mean absolute percentage error (MAPE) of amino acid level forecasting results for the nine datasets with concentration measurement data as inputs

Dataset	MAPE of forecasted amino acid level results (with concentration measurement input)		
	Two-point extrapolation	Exponential equation fitting	Logistic equation fitting
A1	6.82%	6.86%	10.06%
A2	13.49%	22.54%	28.86%
A3	13.37%	13.88%	22.99%
A4	11.76%	16.19%	19.17%
A5	17.92%	21.48%	28.14%
B1	11.86%	7.03%	6.36%
B2	21.52%	8.62%	10.81%
B3	33.90%	31.71%	15.69%
B4	34.69%	17.19%	7.05%

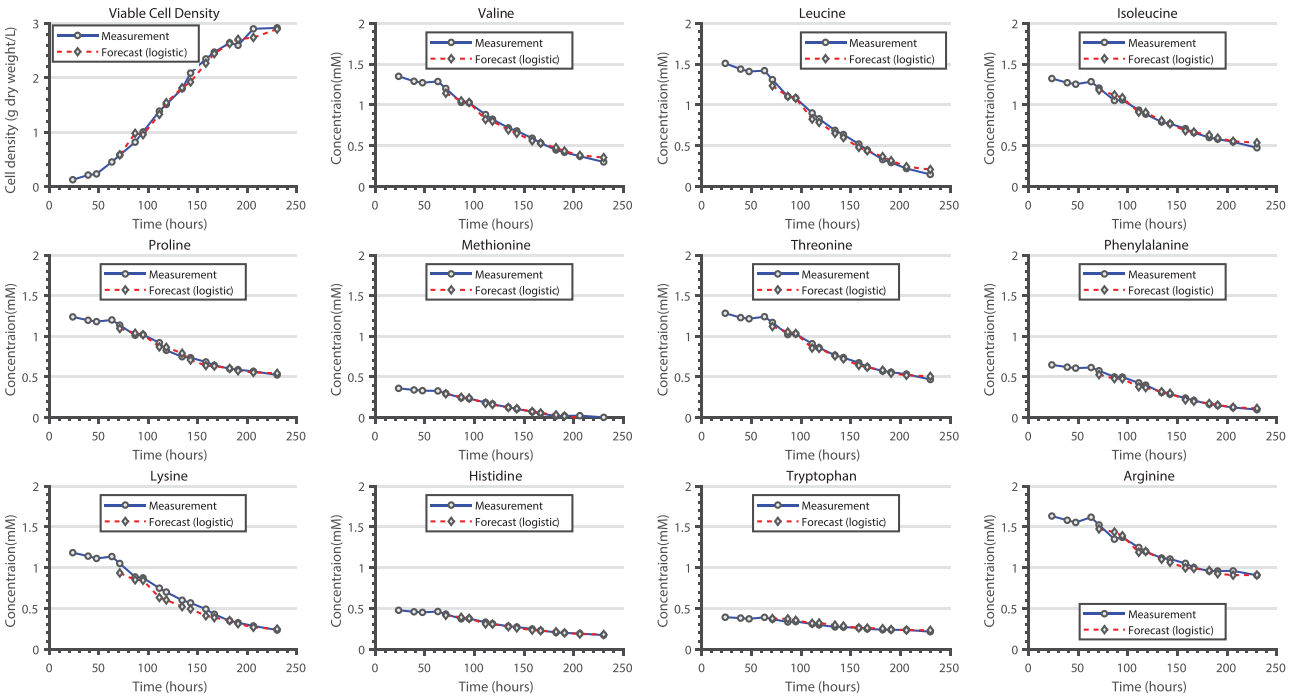

FIGURE 3 Forecast versus measurements of viable cell density and amino acid profiles for batch datasets B4, via logistic curve fitting growth prediction approach without forcing the fitted curves to cross initial data points. Each mark represents the forecast result for the time point with cell density measurements from all the past time points as inputs. Forecast results were generated also assuming amino acid concentration measurements are available as algorithm inputs one time point after the cell density is measured

cultures, we next investigated if our forecasting algorithm could be applied to dictate amino acid feeding levels during fed-batch CHO cultures. A proof-of-concept experiment was designed and conducted to control the level of three amino acids—leucine, lysine and valine at 0.6 mM throughout the fed-batch culture—as shown in Figure 4a. Leucine and valine were selected to be controlled at 0.6 mM to examine the potential of applying the forecast algorithm to control the level of those two branched-chain amino acids (BCAAs) between 0.5 and 1 mM as suggested in previous studies as a means to reduce the accumulation of growth-inhibiting byproducts (Mulukutla et al., 2017; Mulukutla et al., 2019). Lysine was selected as another representative essential amino acid to control in CHO culture. Initially, an empirical-based fed-batch shaker flask culture strategy was designed (see Section 4 and Figure 4a) for an IgG-producing CHO glutamine synthetase (GS) cell line with chemically defined media and defined, preset daily feed volumes. Then, a corresponding forecast-guided fed-batch culture approach was designed to control the concentration of leucine, lysine and valine at 0.6 mM at each sampling time, using customized basal and feed media without the three amino acids. VCD measurements were taken every 12 h and the total cell number was used as input to the forecast algorithm to take volume change into account. The concentrations of leucine,

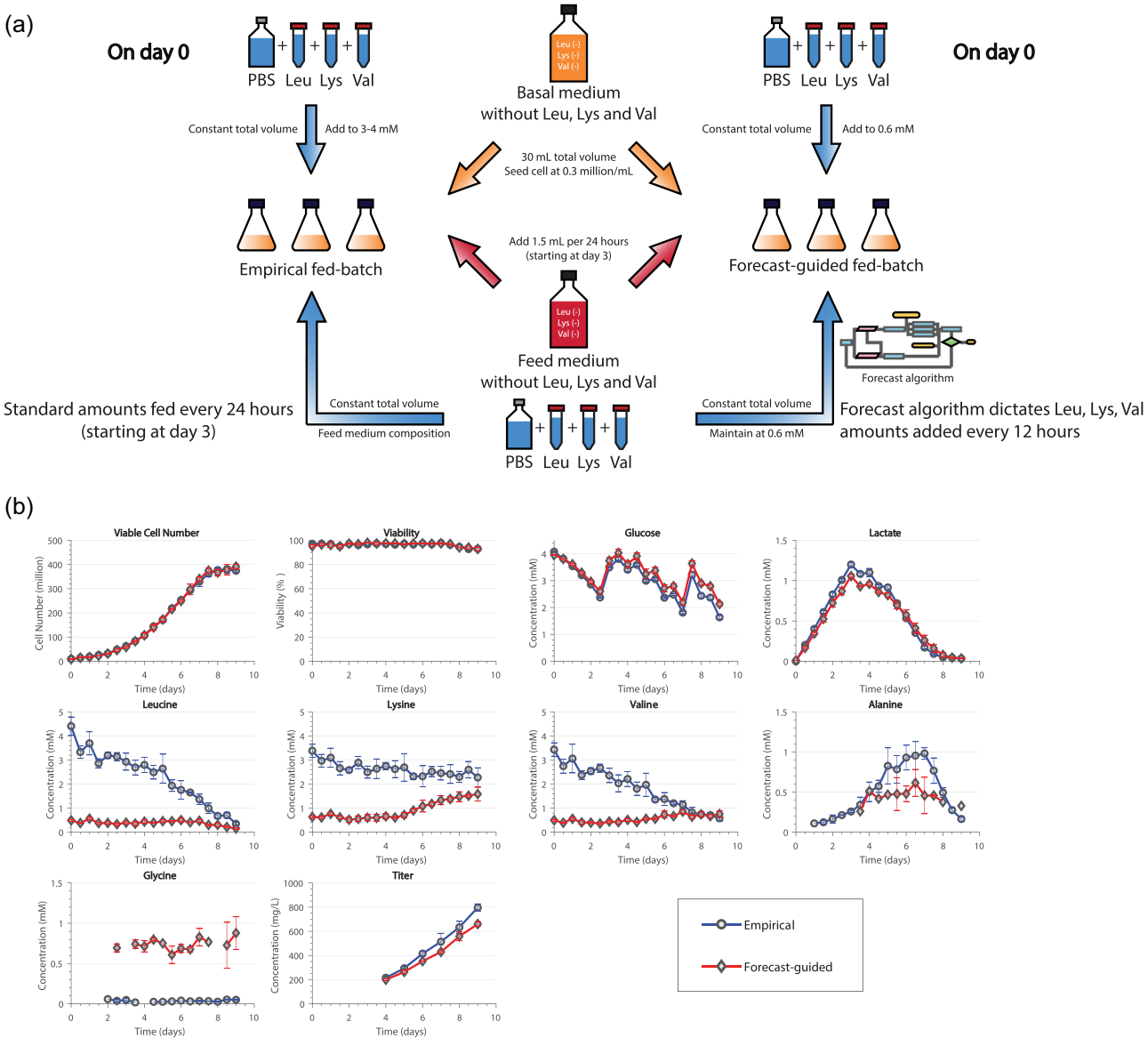

FIGURE 4 Design of the proof-of-concept algorithm-guided amino acid feeding control experiment and measurement results. (a) Visualization of fed-batch CHO cell culture experiment design. CHO cultures were divided in two different groups using empirical and forecast-based approaches to feed leucine, lysine and valine. (b) Comparison of cell number, metabolite concentration and IgG titer in empirical and forecast-guided CHO cell fed-batch cultures. For forecast-guided cultures, concentration of leucine, lysine and valine were designed to be controlled at 0.6 mM at every sampling time. Error bars represent standard deviation of the triplicate samples. CHO, Chinese hamster ovary; IgG, immunoglobulin G

lysine and valine at the following sampling time were forecasted by the algorithm based solely on the measured cell number and initial amino acid levels, and the amount of the three amino acids was calculated and added to the flasks as needed to supply the cells while also maintaining a 0.6 mM concentration at the start of each feeding point. Amino acid concentration at all time points was forecasted using the logistic equation fitting approach, except for the third and fourth time points which were forecasted via two-point extrapolation and the exponential equation fitting approach, respectively. To ensure that the fitting approach does not introduce bias, a check of the R^2 value was performed at each time point after every logistic curve fitting. The two-point extrapolation approach is used instead if R^2 is less than 95% for the logistic curve fit, however, such a case was not encountered throughout the experiment. To test the forecast-guided feeding algorithm and to investigate the effect of controlling the three amino acids at low levels, CHO cells were seeded into empirical culture triplicate and forecast-guided culture triplicate for comparison. Customized basal and feed media without leucine, lysine and valine was used in both culture groups, and the three amino acids were supplemented. For the forecast-guided group, the amounts of amino acids were added based on model prediction and maintenance of a 0.6 mM initial cycle level every 12 h after cell density measurements. For the empirical cultures, defined volumes of amino acid stock solution were added daily based on the original feed composition and predetermined feeding scheme as is common in fed batch processes. Differing volumes of phosphate buffered saline (PBS) were supplied to the two groups to balance the volume change caused by amino acid feeding. The fed-batch culture was stopped when viable cell number started to drop. The amino acid composition of the IgG molecule was considered in the forecast algorithm by generating a corresponding IgG production reaction in the CHO genome-scale model. According to historical data, productivity of IgG for this cell line ranges from about 1×10^{-5} pg/gDW/h to 3.5×10^{-5} pg/gDW/h in various fed-batch conditions, thus the IgG production reaction flux constraint in the model was estimated to be 1.5×10^{-5} pg/gDW/h from Days 0 to 4 and 3×10^{-5} pg/gDW/h after Day 4. Averaged dry weight of the CHO cells was measured to be 430.1 pg/cell for cells cultured in similar conditions as the empirical and forecast-guided culture conditions for Days 3–7 (Figure S61), with no significant change over time or difference between the two culture conditions. Shown in Figure 4b is the comparison of viable cell number, viability, glucose, lactate, leucine, lysine, valine and IgG titer measured for the empirical and forecast-guided fed-batch culture samples. For the empirical cultures, leucine, lysine and valine levels are higher than 3 mM at the beginning and decrease over time. Meanwhile, measurements for forecast-guided culture shows that concentrations of leucine and valine were successfully maintained below 1 mM till the end of Day 9 of the culture while the concentration of lysine gradually increases from Days 5 to 9, indicating overfeeding after Day 5. Leucine concentration decreases after Day 8 indicating higher consumption than expected.

Both empirical and forecast-guided fed-batch cultures exhibit a logistic growth curve, with very similar shapes and peak cell number on Day 9 to indicate comparable growth behaviors. However, some minor differences are evident between the two culture groups including lower lactate production rates (before Day 3), slightly higher lactate consumption rates (after Day 3), and slightly elevated glucose levels for the forecast-guided group. The concentration profiles of most of the other metabolites are comparable between the empirical and forecast-guided cultures except for alanine and glycine (Figure S62), in which peak alanine concentration is higher for the empirical culture and glycine levels are higher for the forecast-guided culture. This observation is in agreement with the results of previous whole animal studies in which BCAA restriction or inhibition of BCAA metabolism results in elevation of glycine levels and decreases in alanine levels in Zucker rat metabolism (White et al., 2016, 2020). These studies report that glycine has an inverse relationship to alanine and BCAA in the skeletal muscle of Zucker rats, which can be explained by the consumption of glycine to generate alanine in the pyruvate-alanine cycle in presence of elevated BCAA levels (White et al., 2016, 2020). Specifically, in the pathway shown in Figure 5a, elevated level of BCAAs results in the catabolism of BCAAs together with the generation of glutamate, which induces the conversion from pyruvate to alanine in a partner transamination reaction. The strong relationship between alanine and BCAA oxidation is also reported in rat skeletal muscle studies (O'Connell, 2013). These BCAA degradation products will subsequently be further degraded and feed metabolites into the tricarboxylic acid (TCA) cycle. Glycine is believed to be a source of pyruvate via the serine dehydratase reaction which also generates NH_4 to be used in the GS reaction (White et al., 2016; White et al., 2020). The reduction in glycine levels in the empirical case may be the result of an enhanced demand for pyruvate for regeneration of α-ketoglutarate needed in BCAA together with alanine by-products (Figure 5a). Thus, in our study, the elevated alanine levels and reduced glycine levels observed in empirical cultures compared to the forecast-guided cultures can be similarly explained by this pathway given the enhanced levels of two BCAAs (leucine and valine) available in these cultures compared to controls.

Surprisingly, the IgG titers are also slightly reduced in the forecast-control group. This may be led by the downregulation of the mechanistic target of rapamycin (mTOR) pathway due to low leucine concentrations maintained in the forecast-guided culture, since the activity of mTOR pathway is known to be strongly related to protein productivity and leucine is a particularly important amino acid regulating mTOR signaling (Dreesen & Fussenegger, 2011; Edros et al., 2014; Kim et al., 2017; Son et al., 2019; Wang et al., 2008). Despite the improved peak cell density and overall higher titer resulting from controlling eight amino acids (including leucine) at low levels, cell specific protein productivity is lower relative to the non-controlled condition as previously noted by (Mulukutla et al., 2017). It is reasonable to hypothesize that simultaneously controlling multiple amino acids at low levels may benefit growth by reducing inhibitory byproducts but may also reduce protein productivity due to low availability of certain amino acids such as leucine. Thus, an optimal level of amino acids should be chosen to limit feeding of amino acids while maintaining IgG and other recombinant protein productivities.

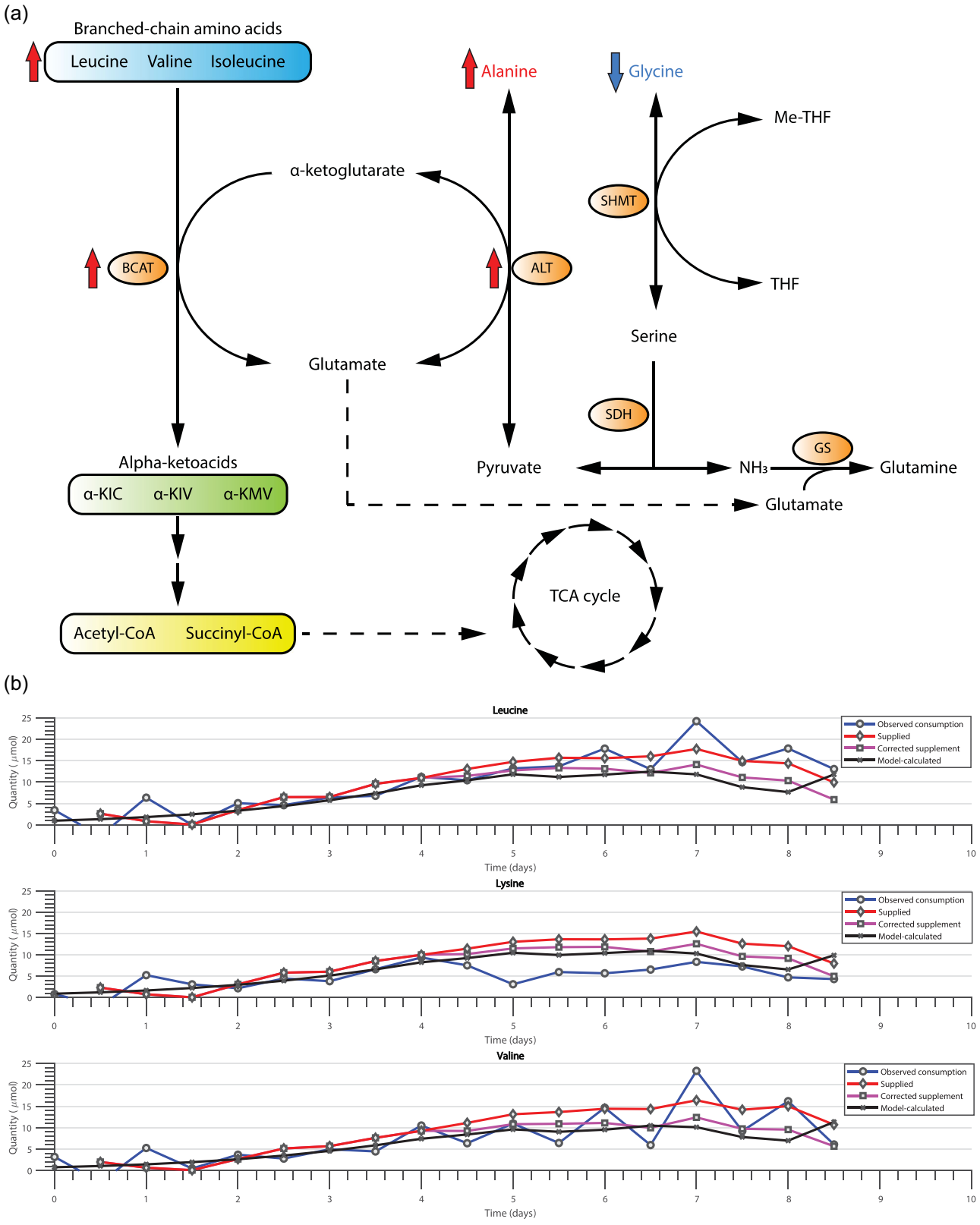

FIGURE 5 (See caption on next page)

We next examined the possible causes of the deviation in observed amino acid concentrations from model predictions. First of all, as with any control algorithm, the predictability depends on the model input used for dictating proper addition of nutrients. Titer measurements following the experiments indicated that, IgG productivity after Day 4 was approximately 1.6×10^{-5} pg/gDW/h and 1.3×10^{-5} pg/gDW/h on average for the empirical and forecast-guided group, respectively, which is about 50% of the assumed productivity used in the initial forecast algorithm. Thus, the increase of lysine concentration in the forecast-guided culture after Day 4 may be due, at least in part, to overfeeding caused by an overestimation of IgG productivity. Second, as mentioned previously, minor error build-up from model predictions without amino acid concentration measurements as correction may also lead to deviations observed in the late stage of the cultures. To examine the impact of these deviations, the actual quantity of amino acids consumed every 12 h estimated from the measurements and feeding history was calculated ("observed consumption") and compared with the amount that was experimentally supplied ("supplied"), as well as the amount that would have been supplied if the measured IgG productivity was instead used as input ("corrected supplement"). In addition, the calculated quantity of amino acids the cultures would consume via implementing the ENM uptake rate approach based on the genome-scale model using measured VCD and IgG productivity as inputs was also determined ("model-calculated"). All four of the aforementioned observed and simulated amino acid quantity profiles for forecast-guided cultures are plotted and presented in Figure 5b. First of all, for all the three amino acids, the actual consumption level matches well with both the experimentally supplied amount and the ENM model-calculated amount before Day 5, suggesting that the forecast-guided feeding approach can effectively predict and control the level of all three amino acids during the first half of the fed-batch culture. After Day 5 when overestimated IgG productivity was used as model input, the experimentally fed amount is higher than the theoretical minimum amount and the "corrected supplement" feeding amount, demonstrating the impact of using inaccurate IgG productivity as input. For all three amino acids, the "corrected supplement" amount is comparable to the ENM "model-calculated" cell consumption amount, indicating that the logistic-based forecast algorithm can accurately predict the supplement quantity for the next time increment.

The leucine and valine profiles exhibit similar consumption/supplement profiles, in which the estimated consumption values for a number of time points after Day 6 are higher than the model-calculated amounts. This observation may be due to the enhanced catabolic reactions of BCAAs during the late stage of the cultures. Indeed, metabolic byproducts of L-leucine and L-valine are detected in the Day 8.5 spent media samples (data not shown). Furthermore, it has been reported that catabolic byproducts of leucine and valine can accumulate to high concentration in culture media on Day 7 of fed-batch cultures, suggesting high activity of BCAA breakdown happening at late stages of fed-batch cultures (Mulukutla et al., 2017, 2019).

Interestingly, the amount of observed lysine consumption from Days 5 to 7 is even lower than the "model-calculated" amount which represents the minimal consumption amount required to sustain the observed growth rate and protein productivity. This overestimation of lysine consumption may be related to the depletion of asparagine that occurs beginning on Day 5. Indeed a previous study on the asparagine starvation response in CHO cell cultures indicates that asparagine depletion has a significant impact on intracellular lysine metabolism and reduces the intracellular lysine pool (Seewöster & Lehmann, 1995). Thus, limitation in asparagine after Day 5 in the forecast-guided cultures may also lead to the decrease in the intracellular lysine pool in CHO cells and may potentially reduce the overall apparent lysine uptake rate as observed in the forecast-guided fed-batch experiment.

As a result, this proof-of-concept amino acid level control experiment demonstrates the potential of a forecast-based control strategy to effectively limit specific amino acids at desired levels in CHO cell cultures while maintaining cell growth performance. With viable cell number as the sole input measurement, this forecast-guided feeding strategy represents a straight-forward, efficient and cost-effective model predictive control system suitable for fed-batch and potentially perfusion mammalian cell culture processes. All that is required to implement this feeding approach is the initial

FIGURE 5 Visualization of relevant amino acid pathways and analysis of amino acid consumption. (a) Metabolic pathway illustrating the catabolism of BCAA and its relationship with glycine consumption and alanine generation in the empirical cultures compared to forecast-guided cultures. Presumably, BCAA degradation via BCAT induces the generation of glutamate and the following formation of alanine in the transamination reaction catalyzed by ALT. Glycine is assumed to be the source of pyruvate to generate alanine. BCAT, branched-chain amino acid aminotransferase; ALT, alanine aminotransferase; SHMT, serine hydroxymethyltransferase; SDH: serine dehydratase; GS, glutamine synthetase; α-KIC, α-ketoisocaproic acid; α-KIV, α-ketoisovaleric acid; α-KMV, α-keto-β-methylvalerate; Me-THF, 5,10-Methylenetetrahydrofolate; THF, tetrahydrofolate. (b) Quantity of consumed/supplied L-leucine, L-lysine, and L-valine by/to the CHO forecast-guided culture estimated by measurement average, feeding history and simulation. "Observed consumption" stands for the observed amount consumed by the cultures. "Supplied" stands for the experimentally added amount. "Corrected supplement" stands for the amount that would have been supplied if the measured IgG productivity was used as model input. "Model-calculated" stands for the quantity of amino acid which the CHO cultures are expected to consume, calculated by the ENM approach using CHO genome-scale model given measured viable cell density and recombinant protein productivity. For consumption, each point represents the amount of amino acid consumed in the following 12 h. For supplement, each point represents the amount supplied at that time point. CHO, Chinese hamster ovary; ENM, essential nutrient minimization; IgG, immunoglobulin G

concentration of amino acids, biomass composition, cell size, and IgG productivity and composition. According to a study comparing the biomass and amino acid composition of multiple CHO cell lines and the CHO genome-scale model, no significant difference was observed between various CHO cells (Széliová et al., 2020). In contrast to biomass composition, size and dry weight of mammalian cells including CHO cells is known to vary between cell lines and may also increase during fed-batch processes (Alhuthali et al., 2021; Edros et al., 2014; Pan et al., 2019; Pan, Dalm, et al., 2017; Pan, Streefland, et al., 2017; Széliová et al., 2020). The relevance of knowing IgG productivity is evident in the limitations of the algorithm in accurately forecasting the concentration of all the three amino acid depletion rates after Day 5, when recombinant protein production becomes a key contributor (Figure S63). Alternatively, extrapolating late-stage culture parameters at the beginning of culture can also be challenging especially for the transition from exponential to stationary phase. As presented in Figures S64, S65, and S66, fitting of an accurate and stable maximal growth capacity parameter (K) at early culture time points is difficult while estimating the other two logistic parameters (N_0 and μ_{max}, see Methods) is straightforward, suggesting that users should increase sampling and algorithm iteration frequency and avoid long forecast time horizons if possible to minimize feeding control error. In the future, the availability to take more rapid at-line or off-line measurements will be beneficial since those measurements can significantly improve the forecasting capacity of this predictive control approach. Indeed, amino acid concentration measurements can help correct model prediction error accumulated over time or help to account for catabolic activity of certain amino acids in different cell lines or culture conditions while measurements of titer can help estimating the cell-specific protein productivity at different phases of the culture so that the forecast model can be adjusted accordingly to generate more accurate predictions.

This forecast algorithm could be expanded to a broader range of applications ina future version. For example, forecast of the final stage of fed-batch culture in which viable cell number and viability decrease is not included. To account for culture behavior during late fed-batch stage, future expansion of the current forecast control algorithm can incorporate additional models or parameters which effectively describe death behavior of the cells. The most straightforward approach is to introduce a death rate parameter that turns regular logistic equation to a four-parameter logistic equation, which is demonstrated to be effective in modeling CHO cell batch and fed-batch cultures (Goudar et al., 2005). Second, the currently used ENM-based uptake rate prediction approach may only accurately predict the consumption of about ten amino acids specifically essential for CHO cells, leaving the other ten non-essential nutrients not modelled. The consumption of most non-essential nutrients cannot be accurately predicted based on the same assumption as the ENM approach because their consumption serves many purposes, and they can be interconverted. Such metabolic characteristics may also be dependent on specific cell lines as well as culture conditions, making the forecast of multiple non-essential nutrients in culture challenging and requiring additional analytical technologies or kinetic and data-driven model configurations (Hagrot et al., 2019; Park et al., 2021; Schinn et al., 2021). Indeed future forecast algorithm can combine a number of additional tools and methods to describe cell growth and nutrient consumption behavior throughout the entire culture period.

3 | CONCLUSIONS

Model predictive control approaches require a dynamic process model with the capability to effectively describe process performance. In this study, an algorithm is developed to forecast future essential amino acid levels in CHO cell cultures from VCD measurements, by combining a growth regression model and a genome-scale-model-based nutrient minimization approach. Logistic equation fitting is selected as an effective regression model due to its advantages in depicting the profiles of VCD observed across various culture datasets. The logistic-based forecast algorithm is able to predict the levels of most essential amino acids with VCD as the sole measured input, while off-line amino acid concentration measurements can serve as secondary inputs to correct error build-up. Finally, when used in CHO fed-batch cultures, the forecast algorithm is able to guide the feeding of amino acids and successfully controls the concentration of three selected essential amino acids at low levels throughout the culture period with conserved growth performance. This algorithm promises to transform our ability to control input costs in mammalian cell cultures and as a result lower cultivation costs surrounding key nutrients and limit the accumulation of undesirable metabolic byproducts that can result from overfeeding of amino acids. Unlike control approaches based on analytical technologies, this model-based approach facilitates CHO bioprocesses by providing in-advance estimation or reference information to guide the operation and design of better process control strategies such as feeding schemes or media composition. The forecast algorithm can be further expanded in the future through incorporation of additional model elements as more datasets and information regarding cellular physiology and metabolism becomes available that enables us to effectively model and control the performance of mammalian cell cultivation systems.

4 | METHODS

4.1 | Growth prediction methods by equation fitting and two-point extrapolation

Growth kinetics of CHO cells is characterized by exponential equation and logistic equation in this study. Exponential equation describes cell growth as:

$$N_x = N_{x,0} \times exp(u_{max} \times (t - t_0)) \quad (1)$$

And three-parameter logistic equation can be expressed as:

$$N_x = \frac{K}{1 + (K - N_{x,0})/N_{x,0} \times exp(u_{max} \times (t_0 - t))} \quad (2)$$

Here $N_{x,0}$ is the number of cells (counted in million cells, or total dry weight in grams) at initial culture time t_0(h), N_x is the number of cells at a later culture time t . u_{max} is the maximal specific growth rate (1/h) and K is the maximal growth capacity (million cells or gram) which represents the upper limit of sigmoidal growth curve. Equation parameters are determined by fitting the VCD data over different time, via least squares method. Exponential equation requires at least three data points and logistic equation requires at least four to estimate all the parameters.

Two-point extrapolation approach predicts VCD at the future time point from the two latest time points (one current time point and its previous time point) by exponential growth function, assuming the effective growth rate is constant across three time points. Based on Equation (1), effective growth rate u_{eff} during the previous and current time point $t_1$and $t_2$can be determined by:

$$u_{eff} = ln(N_{x,2} - N_{x,1})/(t_2 - t_1) \quad (3)$$

Here $N_{x,1}$ and $N_{x,2}$ are the VCD at t_1 and t_2. Then the VCD $N_{x,3}$ at a future time point t_3 can be estimated by:

$$N_{x,3} = N_{x,2} \times exp(u_{eff} \times (t_3 - t_2)) \quad (4)$$

4.2 | Amino acid consumption rate and concentration change predictions

Essential amino acid consumption rates are predicted from specific growth rates using the ENM approach as described in (Chen et al., 2019). Specifically, estimated specific growth rate is set as a constraint in the CHO-K1 genome-scale model (Hefzi et al., 2016), and each nutrient uptake rate is individually minimized while uptake rates of all the other nutrients are set to be unlimited. Specific growth rates used as the input of ENM can be estimated for the time period of forecast (from current time point to forecast time point) using exponential or logistic equation. If two-point extrapolation or exponential fitting approach was used to predict growth, effective growth rate estimated using Equation (3) or fitted maximal growth rate can be used as ENM input, respectively. If logistic equation fitting approach was used, average effective growth rate at the "current" and "future" time point should be used as ENM input, where the effective growth rate at a particular time point can be calculated by:

$$u_{eff} = \frac{dN_x}{dt} \times \frac{1}{N_x} = u_{max} \times \left(1 - \frac{N_x}{K}\right) \quad (5)$$

Changes in concentration of amino acid i can be calculated with the ENM-solved uptake rates r_i(millimole per million cells per hour, or millimole per gram dry weight per hour) by:

$$C_{i,2} = C_{i,1} - \int_{t1}^{t2} (r_i N_x) dt \quad (6)$$

For two-point extrapolation or exponential equation fitting, Equation (6) becomes:

$$C_{i,2} = C_{i,1} - r_i N_{x,0} \times \frac{exp(ut_2) - exp(ut_1)}{u \times exp(ut_0)} \quad (7)$$

$$u = u_{eff} \text{ for two point extrapolation}$$

$$u = u_{max} \text{ for exponential fitting}$$

For logistic equation fitting:

$$C_{i,2} = C_{i,1} - \frac{r_i K}{u_{max}} \times \ln\left(\frac{(K - N_0) \times \exp(u_{max} t_0) + N_0 \times \exp(u_{max} t_2)}{(K - N_0) \times \exp(u_{max} t_0) + N_0 \times \exp(u_{max} t_1)}\right) \quad (8)$$

Here $C_{i,1}$ and $C_{i,2}$ is the concentration of amino acid (mmol/L) i at time t_1 and t_2 This approach uses the steady-state rate solution r_i at the beginning of each time interval to solve for concentration changes during the interval by integration, which is known as the static optimization approach. Conversion between unit of million cells per hour and gram dry weight per hour is performed assuming cell dry weight equals to 216.1 (Hefzi et al., 2016) and 430.1 pg per cell for analysis with datasets group A1–B4 and forecast-guided control experiment.

4.3 | Essential amino acid level forecasting algorithm

The complete amino acid level forecasting algorithm shown in Figure 6 was designed by combining growth prediction and amino acid uptake prediction methods mentioned in above sections. First of all, the algorithm begins with at least two, three, or four time points with corresponding VCD data for two-point extrapolation, exponential equation fitting and logistic equation fitting method to forecast the VCD at the next time point. Then, the VCD at the future time point and effective growth rate of the corresponding time period are estimated by Equations (1), (2), (3), (4), and (5). With the genome-scale model and the ENM approach, uptake rates of essential amino acids can be predicted from the effective growth rate and the future amino acid level predicted by static optimization approach with Equations (7) or (8) and values of initial amino acid levels. When applying this algorithm to continuous cell culture processes, the above procedures run in an iterative manner and the next iteration begins when a new VCD measurement becomes available. Starting from the second iteration, cell growth prediction for future time point is performed with the updated growth profile which includes the latest VCD measurement. Amino acid concentration levels at each iteration are estimated again with the newest VCD measurement using static optimization approach, and the future amino acid concentrations are then forecasted based on the re-estimated current amino acid concentrations.

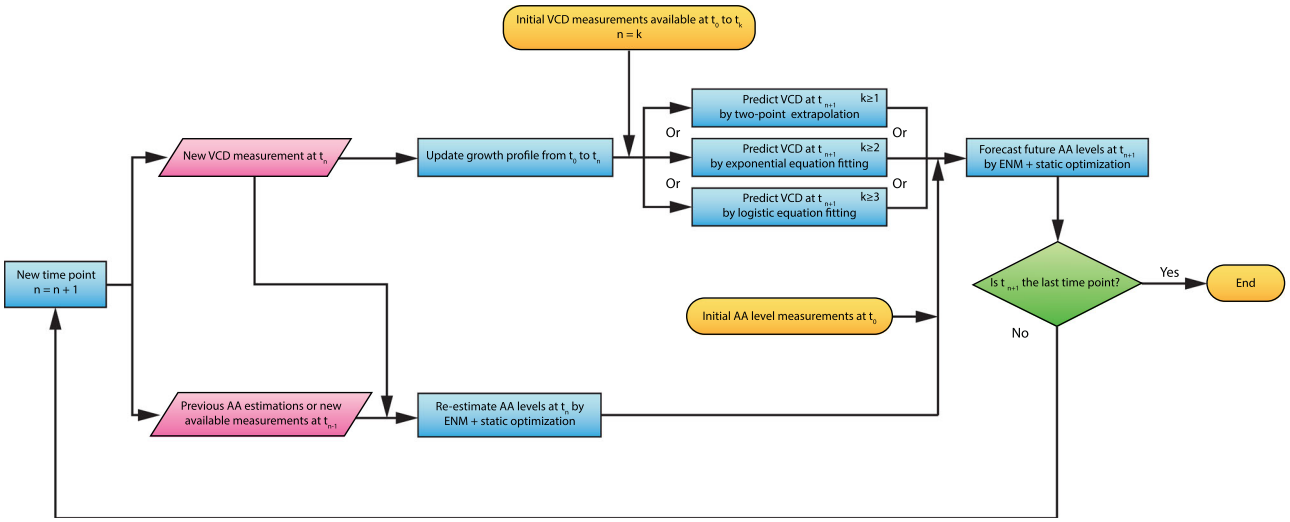

FIGURE 6 Flow chart demonstrating the essential amino acid level forecasting algorithm. Future viable cell density can be estimated by two-point extrapolation, exponential equation fitting or logistic equation fitting depending on the appropriateness in particular scenarios and available number of time points and cell density measurements ($k \geq 1, 2, 3$)

4.4 | Cell culture and analytics

An IgG-producing CHOZN® GS cell line and customized chemically defined basal and feed media without L-glutamine, L-leucine, L-lysine, and L-valine were acquired from MilliporeSigma and were used to perform the algorithm-guided amino acid feeding control experiment. CHO cells were seeded at 0.3×10^6 cells/ml in 30 ml basal media and cultured in 125 ml shaker flasks (Thermo Fisher Scientific) in incubators operating at 37°C, 125 RPM and 5% CO_2 level. Every 12 h, flasks were supplied with certain volume of 60 mM L-leucine, L-lysine, and L-valine stock solution in PBS to reach desired concentrations, and different volume of PBS was added to the flasks to ensure total volume added to each flask is constant. A total of 1.5 ml of feed media A and B were added to the flasks every day starting from Day 3 and same amount of glucose was supplied to all flasks on Days 2.5 and 7. About 300–500 µl of culture samples were collected every 12 h. VCD, averaged cell size and viability were determined using Cellometer Auto T4 automated cell counter (Nexcelom) with trypan blue staining approach. Glucose and lactate concentrations were measured using a YSI 2950D biochemical analyzer (Yellow Spring Instrument). For empirical-based feeding flasks, the amount of amino acid supplied is determined to resemble the original media formulation. For algorithm-guided feeding control flasks, the amount of amino acid supplied every 12 h was calculated according to the forecast algorithm using measured VCD profile as input, aiming to control the concentrations at 0.6 mM after 12 h. Another batch of CHO cells were cultured in the same conditions as empirical-based and algorithm-guided feeding flask groups, and averaged cell dry weight was estimated by weighing the freeze-dried 50 million of cells collected on Days 3–7 of these flasks.

Metabolite profiling of spent media samples was done using GC-MS (Agilent 7890A GC with 5977B MSD) based analytical methods (Long & Antoniewicz, 2019). Specifically, spent media samples were dried under nitrogen at 37°C and incubated at 37°C in 2 wt% methoxyamine dissolved in pyridine for 90 min. Samples were derivatized using MTBSTFA + 1% tBDMS and incubated at 60°C for 30 min. Finally, samples were centrifuged at 14,000 rpm for 5 min and 50 µl of sample was transferred to a GC injection vial for GC-MS analysis. Concentration of amino acids and a few other metabolites were measured after the end of culture runs using Rebel bioprocess analyzer (908 Devices) following the corresponding protocol. IgG titer was measured via an affinity chromatography approach using POROS A protein A column (Thermo Fisher Scientific) and an HPLC system (Agilent). Solution of 50 mM of sodium phosphate with 150 mM sodium chloride at pH = 7.2 was prepared as binding buffer. Due to the high concentration of dextran sulfate presented in culture, 0.7 M of L-arginine solution at pH 3.0 (adjusted with hydrochloric acid) was used as elution buffer as suggested in (Kim & Park, 2015). The running flowrate is constant at 2.0 ml/min with a stepwise elution gradient.

4.5 | Data acquisition and algorithm implementation

The original non-normalized cell density and amino acid concentration data from (Chen et al., 2019) was used as datasets A1–A5 in this study. Cell density and amino acid concentration profiles in datasets B1–B4 were estimated from figures in (Martínez et al., 2015). Genome-scale metabolic models of CHO cells were acquired from (Hefzi et al., 2016). The entire algorithm was coded and implemented all using MATLAB version 2018b.

ACKNOWLEDGEMENTS

This work was supported by the Industry–University Cooperative Research Center Program under U.S. National Science Foundation

(Grant Number: 1624684). We thank the Advanced Mammalian Biomanufacturing Innovation Center (AMBIC) member companies for their financial support and mentoring.

DATA AVAILABILITY STATEMENT

The data that supports the findings of this study are available in the supplementary material of this article.

ORCID

Yiqun Chen http://orcid.org/0000-0003-3074-9605
Venkata Gayatri Dhara http://orcid.org/0000-0002-5748-8161
Michael J. Betenbaugh http://orcid.org/0000-0002-6336-4659

REFERENCES

Aehle, M., Schaepe, S., Kuprijanov, A., Simutis, R., & Lübbert, A. (2011). Simple and efficient control of CHO cell cultures. *Journal of Biotechnology*, *153*(1), 56–61. https://doi.org/10.1016/j.jbiotec.2011.03.006

Alhuthali, S., Kotidis, P., & Kontoravdi, C. (2021). Osmolality Effects on CHO Cell Growth, Cell Volume, Antibody Productivity and Glycosylation. https://doi.org/10.3390/ijms22073290

Almquist, J., Cvijovic, M., Hatzimanikatis, V., Nielsen, J., & Jirstrand, M. (2014). Kinetic models in industrial biotechnology—Improving cell factory performance. *Metabolic Engineering*, *24*, 38–60. https://doi.org/10.1016/j.ymben.2014.03.007

André, S., Lagresle, S., Da Sliva, A., Heimendinger, P., Hannas, Z., Calvosa, É., & Duponchel, L. (2017). Developing global regression models for metabolite concentration prediction regardless of cell line. *Biotechnology and Bioengineering*, *114*(11), 2550–2559. https://doi.org/10.1002/bit.26368

Bayrak, E. S., Wang, T., Cinar, A., & Undey, C. (2015). Computational modeling of fed-batch cell culture bioreactor: Hybrid agent-based approach. *IFAC-PapersOnLine*, *48*(8), 1252–1257. https://doi.org/10.1016/j.ifacol.2015.09.140

Berry, B. N., Dobrowsky, T. M., Timson, R. C., Kshirsagar, R., Ryll, T., & Wiltberger, K. (2016). Quick generation of Raman spectroscopy based in-process glucose control to influence biopharmaceutical protein product quality during mammalian cell culture. *Biotechnology Progress*, *32*(1), 224–234. https://doi.org/10.1002/btpr.2205

Bibila, T. A., & Robinson, D. K. (1995). In pursuit of the optimal fed-batch process for monoclonal antibody production. *Biotechnology Progress*, *11*(1), 1–13. https://doi.org/10.1021/bp00031a001

Bielser, J., Wolf, M., Souquet, J., Broly, H., & Morbidelli, M. (2018). Perfusion mammalian cell culture for recombinant protein manufacturing—A critical review. *Biotechnology Advances*, *36*(4), 1328–1340. https://doi.org/10.1016/j.biotechadv.2018.04.011

Chen, N., Bennett, M. H., & Kontoravdi, C. (2014). Analysis of Chinese hamster ovary cell metabolism through a combined computational and experimental approach. *Cytotechnology*, *66*(6), 945–966. https://doi.org/10.1007/s10616-013-9648-1

Chen, Y., McConnell, B. O., Gayatri Dhara, V., Mukesh Naik, H., Li, C., Antoniewicz, M. R., & Betenbaugh, M. J. (2019). An unconventional uptake rate objective function approach enhances applicability of genome-scale models for mammalian cells. *NPJ Systems Biology and Applications*, *5*(1), 25. https://doi.org/10.1038/s41540-019-0103-6

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*(4), 344–357. https://doi.org/10.1016/j.jprocont.2014.02.007

Dreesen, I. A. J., & Fussenegger, M. (2011). Ectopic expression of human mTOR increases viability, robustness, cell size, proliferation, and antibody production of chinese hamster ovary cells. *Biotechnology and Bioengineering*, *108*(4), 853–866. https://doi.org/10.1002/bit.22990

Edros, R., McDonnell, S., & Al-Rubeai, M. (2014). The relationship between mTOR signalling pathway and recombinant antibody productivity in CHO cell lines. *BMC Biotechnology*, *14*, 15. https://doi.org/10.1186/1472-6750-14-15

Farges, B., Chenu, S., Marc, A., & Goergen, J. L. (2008). Kinetics of IFN-γ producing CHO cells and other industrially relevant cell lines in rapeseed-supplemented batch cultures. *Process Biochemistry*, *43*(9), 945–953. https://doi.org/10.1016/j.procbio.2008.04.025

Gagnon, M., Hiller, G., Luan, Y., Kittredge, A., DeFelice, J., & Drapeau, D. (2011). High-End pH-controlled delivery of glucose effectively suppresses lactate accumulation in CHO Fed-batch cultures. *Biotechnology and Bioengineering*, *108*(6), 1328–1337. https://doi.org/10.1002/bit.23072

Goudar, C. T. (2012). Computer programs for modeling mammalian cell batch and fed-batch cultures using logistic equations. *Cytotechnology*, *64*(4), 465–475. https://doi.org/10.1007/s10616-011-9425-y

Goudar, C. T., Joeris, K., Konstantinov, K. B., & Piret, J. M. (2005). Logistic equations effectively model mammalian cell batch and fed-batch kinetics by logically constraining the fit. *Biotechnology Progress*, *21*(4), 1109–1118. https://doi.org/10.1021/bp050018j

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., Faustrup Kildegaard, H., Min Lee, G., Palsson, B. O., Nielsen, J., & Lewis, N. E. (2020). Genome-scale reconstructions of the mammalian secretory pathway predict metabolic costs and limitations of protein secretion. *Nature Communications*, *11*(1), 68. https://doi.org/10.1038/s41467-019-13867-y

Hagrot, E., Oddsdóttir, H. Æ., Mäkinen, M., Forsgren, A., & Chotteau, V. (2019). Novel column generation-based optimization approach for poly-pathway kinetic model applied to CHO cell culture. *Metabolic Engineering Communications*, *8*, e00083. https://doi.org/10.1016/j.mec.2018.e00083

Hawkins, D. M. (2004). The problem of overfitting. *Journal of Chemical Information and Computer Sciences*, *44*(1), 1–12. https://doi.org/10.1021/ci0342472

Hefzi, H., Ang, K. S., Hanscho, M., Bordbar, A., Ruckerbauer, D., Lakshmanan, M., Orellana, C. A., Baycin-Hizal, D., Huang, Y., Ley, D., Martinez, V. S., Kyriakopoulos, S., Jiménez, N. E., Zielinski, D. C., Quek, L., Wulff, T., Arnsdorf, J., Li, S., Lee, J. S., & Lewis, N. E. (2016). A consensus genome-scale reconstruction of Chinese hamster ovary cell metabolism. *Cell Systems*, *3*(5), 434–443. https://doi.org/10.1016/j.cels.2016.10.020

Henry, O., Kwok, E., & Piret, J. M. (2008). Simpler noninstrumented batch and semicontinuous cultures provide mammalian cell kinetic data comparable to continuous and perfusion cultures. *Biotechnology Progress*, *24*(4), 921–931. https://doi.org/10.1002/btpr.17

Hiller, G. W., Ovalle, A. M., Gagnon, M. P., Curran, M. L., & Wang, W. (2017). Cell-controlled hybrid perfusion fed-batch CHO cell process provides significant productivity improvement over conventional fed-batch cultures. *Biotechnology and Bioengineering*, *114*(7), 1438–1447. https://doi.org/10.1002/bit.26259

Huang, Y., Hu, W., Rustandi, E., Chang, K., Yusuf-Makagiansar, H., & Ryll, T. (2010). Maximizing productivity of CHO cell-based fed-batch culture using chemically defined media conditions and typical manufacturing equipment. *Biotechnology Progress*, *26*(5), 1400–1410. https://doi.org/10.1002/btpr.436

Jaqaman, K., & Danuser, G. (2006). Linking data to models: Data regression. *Nature Reviews Molecular Cell Biology*, *7*(11), 813–819. https://doi.org/10.1038/nrm2030

Kildegaard, H. F., Baycin-Hizal, D., Lewis, N. E., & Betenbaugh, M. J. (2013). The emerging CHO systems biology era: Harnessing the 'omics revolution for biotechnology. *Current Opinion in*

Biotechnology, *24*(6), 1102–1107. https://doi.org/10.1016/j.copbio.2013.02.007

Kim, B. G., & Park, H. W. (2015). Arginine as an eluent overcomes the hindrance of monoclonal antibody quantification by dextran sulfate in protein A affinity chromatography. *Biotechnology Progress*, *31*(6), 1536–1541. https://doi.org/10.1002/btpr.2164

Kim, J. H., Lee, C., Lee, M., Wang, H., Kim, K., Park, S. J., Yoon, I., Jang, J., Zhao, H., Kim, H. K., Kwon, N. H., Jeong, S. J., Yoo, H. C., Kim, J. H., Yang, J. S., Lee, M. Y., Lee, C. W., Yun, J., Oh, S. J., & Kim, S. (2017). Control of leucine-dependent mTORC1 pathway through chemical intervention of leucyl-tRNA synthetase and RagD interaction. *Nature Communications*, *8*(1), 732. https://doi.org/10.1038/s41467-017-00785-0

Konakovsky, V., Clemens, C., Müller, M. M., Bechmann, J., & Herwig, C. (2017). A robust feeding strategy to maintain set-point glucose in mammalian fed-batch cultures when input parameters have a large error. *Biotechnology Progress*, *33*(2), 317–336. https://doi.org/10.1002/btpr.2438

Kotidis, P., & Kontoravdi, C. (2019). Strategic Framework for Parameterization of Cell Culture Models. https://doi.org/10.3390/pr7030174

Li, C., Yelsky, J., Chen, Y., Zuñiga, C., Eng, R., Jiang, L., Shapiro, A., Huang, K., Zengler, K., & Betenbaugh, M. J. (2019). Utilizing genome-scale models to optimize nutrient supply for sustained algal growth and lipid productivity. *NPJ Systems Biology and Applications*, *5*(1), 33. https://doi.org/10.1038/s41540-019-0110-7

Li, M., Ebel, B., Chauchard, F., Guédon, E., & Marc, A. (2018). Parallel comparison of in situ Raman and NIR spectroscopies to simultaneously measure multiple variables toward real-time monitoring of CHO cell bioreactor cultures. *Biochemical Engineering Journal*, *137*, 205–213. https://doi.org/10.1016/j.bej.2018.06.005

Long, C. P., & Antoniewicz, M. R. (2019). High-resolution 13C metabolic flux analysis. *Nature Protocols*, *14*(10), 2856–2877. https://doi.org/10.1038/s41596-019-0204-0

López-Meza, J., Araíz-Hernández, D., Carrillo-Cocom, L., López-Pacheco, F., Rocha-Pizaña, M., del Refugio, & 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*(4), 1287–1300. https://doi.org/10.1007/s10616-015-9889-2

Lu, F., Toh, P. C., Burnett, I., Li, F., Hudson, T., Amanullah, A., & Li, J. (2013). Automated dynamic fed-batch process and media optimization for high productivity cell culture process development. *Biotechnology and Bioengineering*, *110*(1), 191–205. https://doi.org/10.1002/bit.24602

Martínez, J. A., Bulté, D. B., Contreras, M. A., Palomares, L. A., & Ramírez, O. T. (2020). Dynamic modeling of CHO cell metabolism using the hybrid cybernetic approach with a novel elementary mode analysis strategy. *Frontiers in Bioengineering and Biotechnology*, *8*, 279. https://www.frontiersin.org/article/10.3389/fbioe.2020.00279

Martínez, V. S., Buchsteiner, M., Gray, P., Nielsen, L. K., & Quek, L. (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. https://doi.org/10.1016/j.meteno.2015.06.001

Meuwly, F., Weber, U., Ziegler, T., Gervais, A., Mastrangeli, R., Crisci, C., Rossi, M., Bernard, A., von Stockar, U., & Kadouri, A. (2006). Conversion of a CHO cell culture process from perfusion to fed-batch technology without altering product quality. *Journal of Biotechnology*, *123*(1), 106–116. https://doi.org/10.1016/j.jbiotec.2005.10.013

Mulukutla, B. C., Kale, J., Kalomeris, T., Jacobs, M., & Hiller, G. W. (2017). Identification and control of novel growth inhibitors in fed-batch cultures of Chinese hamster ovary cells. *Biotechnology and Bioengineering*, *114*(8), 1779–1790. https://doi.org/10.1002/bit.26313

Mulukutla, B. C., Mitchell, J., Geoffroy, P., Harrington, C., Krishnan, M., Kalomeris, T., Morris, C., Zhang, L., Pegman, P., & Hiller, G. W. (2019). Metabolic engineering of Chinese hamster ovary cells towards reduced biosynthesis and accumulation of novel growth inhibitors in fed-batch cultures. *Metabolic Engineering*, *54*, 54–68. https://doi.org/10.1016/j.ymben.2019.03.001

Naderi, S., Meshram, M., Wei, C., McConkey, B., Ingalls, B., Budman, H., & Scharer, J. (2011). Development of a mathematical model for evaluating the dynamics of normal and apoptotic Chinese hamster ovary cells. *Biotechnology Progress*, *27*(5), 1197–1205. https://doi.org/10.1002/btpr.647

O'Connell, T. M. (2013). The Complex Role of Branched Chain Amino Acids in Diabetes and Cancer. https://doi.org/10.3390/metabo3040931

Pan, X., Alsayyari, A. A., Dalm, C., Hageman, J. A., Wijffels, R. H., & Martens, D. E. (2019). Transcriptome analysis of CHO cell size increase during a fed-batch process. *Biotechnology Journal*, *14*(3), 1800156. https://doi.org/10.1002/biot.201800156

Pan, X., Dalm, C., Wijffels, R. H., & Martens, D. E. (2017). Metabolic characterization of a CHO cell size increase phase in fed-batch cultures. *Applied Microbiology and Biotechnology*, *101*(22), 8101–8113. https://doi.org/10.1007/s00253-017-8531-y

Pan, X., Streefland, M., Dalm, C., Wijffels, R. H., & Martens, D. E. (2017). Selection of chemically defined media for CHO cell fed-batch culture processes. *Cytotechnology*, *69*(1), 39–56. https://doi.org/10.1007/s10616-016-0036-5

Park, S., Park, C., Choi, D., Hong, J. K., & Lee, D. (2021). Bioprocess digital twins of mammalian cell culture for advanced biomanufacturing. *Current Opinion in Chemical Engineering*, *33*, 100702. https://doi.org/10.1016/j.coche.2021.100702

Robitaille, J., Chen, J., & Jolicoeur, M. (2015). A single dynamic metabolic model can describe mAb producing CHO cell batch and fed-batch cultures on different culture media. *PLoS One*, *10*(9), e0136815. https://doi.org/10.1371/journal.pone.0136815

Schinn, S., 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. *Biotechnology and Bioengineering*, *118*(5), 2118–2123. https://doi.org/10.1002/bit.27714

Seewöster, T., & Lehmann, J. (1995). Influence of targeted asparagine starvation on extra- and intracellular amino acid pools of cultivated Chinese hamster ovary cells. *Applied Microbiology and Biotechnology*, *44*(3), 344–350. https://doi.org/10.1007/BF00169927

Sommeregger, W., Sissolak, B., Kandra, K., von Stosch, M., Mayer, M., & Striedner, G. (2017). Quality by control: Towards model predictive control of mammalian cell culture bioprocesses. *Biotechnology Journal*, *12*(7), 1600546. https://doi.org/10.1002/biot.201600546

Son, S. M., Park, S. J., Lee, H., Siddiqi, F., Lee, J. E., Menzies, F. M., & Rubinsztein, D. C. (2019). Leucine signals to mTORC1 via its metabolite acetyl-coenzyme A. *Cell Metabolism*, *29*(1), 192–201. https://doi.org/10.1016/j.cmet.2018.08.013

Széliová, D., Ruckerbauer, D. E., Galleguillos, S. N., Petersen, L. B., Natter, K., Hanscho, M., Troyer, C., Causon, T., Schoeny, H., Christensen, H. B., Lee, D., Lewis, N. E., Koellensperger, G., Hann, S., Nielsen, L. K., Borth, N., & Zanghellini, J. (2020). What CHO is made of: Variations in the biomass composition of Chinese hamster ovary cell lines. *Metabolic Engineering*, *61*, 288–300. https://doi.org/10.1016/j.ymben.2020.06.002

Teixeira, A. P., Alves, C., Alves, P. M., Carrondo, M. J. T., & Oliveira, R. (2007). Hybrid elementary flux analysis/nonparametric modeling: Application for bioprocess control. *BMC Bioinformatics*, *8*(1), 30. https://doi.org/10.1186/1471-2105-8-30

Wang, X., Fonseca, B. D., Tang, H., Liu, R., Elia, A., Clemens, M. J., Bommer, U. A., & Proud, C. G. (2008). Re-evaluating the roles of proposed modulators of mammalian target of rapamycin complex 1 (mTORC1) signaling. *The Journal of Biological Chemistry*, *283*(45), 30482–30492. https://doi.org/10.1074/jbc.M803348200

White, P. J., Lapworth, A. L., An, J., Wang, L., McGarrah, R. W., Stevens, R. D., Ilkayeva, O., George, T., Muehlbauer, M. J., Bain, J. R., Trimmer, J. K., Brosnan, M. J., Rolph, T. P., & Newgard, C. B. (2016). Branched-chain amino acid restriction in Zucker-fatty rats improves muscle insulin sensitivity by enhancing efficiency of fatty acid oxidation and acyl-glycine export. *Molecular Metabolism*, *5*(7), 538–551. https://doi.org/10.1016/j.molmet.2016.04.006

White, P. J., Lapworth, A. L., McGarrah, R. W., Kwee, L. C., Crown, S. B., Ilkayeva, O., An, J., Carson, M. W., Christopher, B. A., Ball, J. R., Davies, M. N., Kjalarsdottir, L., George, T., Muehlbauer, M. J., Bain, J. R., Stevens, R. D., Koves, T. R., Muoio, D. M., Brozinick, J. T., & Newgard, C. B. (2020). Muscle-liver trafficking of BCAA-derived nitrogen underlies obesity-related glycine depletion. *Cell Reports*, *33*(6), 108375. https://doi.org/10.1016/j.celrep.2020.108375

Wiegmann, V., Giaka, M., Martinez, C. B., & Baganz, F. (2019). Towards the development of automated fed-batch cell culture processes at microscale. *Biotechniques*, *67*(5), 238–241. https://doi.org/10.2144/btn-2019-0063

Wlaschin, K. F., & Hu, W. (2006). Fedbatch culture and dynamic nutrient feeding. *Advances in Biochemical Engineering/Biotechnology*, *101*, 43–74. https://doi.org/10.1007/10_015

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. *Biotechnology Progress*, *26*(1), 208–219. https://doi.org/10.1002/btpr.284

Ben Yahia, B., Gourevitch, B., Malphettes, L., & Heinzle, E. (2017). Segmented linear modeling of CHO fed-batch culture and its application to large scale production. *Biotechnology and Bioengineering*, *114*(4), 785–797. https://doi.org/10.1002/bit.26214

Ben Yahia, B., Malphettes, L., & Heinzle, E. (2015). Macroscopic modeling of mammalian cell growth and metabolism. *Applied Microbiology and Biotechnology*, *99*(17), 7009–7024. https://doi.org/10.1007/s00253-015-6743-6

Yongky, A., Xu, J., Tian, J., Oliveira, C., Zhao, J., McFarland, K., Borys, M. C., & Li, Z. J. (2019). Process intensification in fed-batch production bioreactors using non-perfusion seed cultures. *Null*, *11*(8), 1502–1514. https://doi.org/10.1080/19420862.2019.1652075

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Chen, Y., Liu, X., Anderson, J. Y. L., Naik, H. M., Dhara, V. G., Chen, X., Harris, G. A., & Betenbaugh, M. J. (2022). A genome-scale nutrient minimization forecast algorithm for controlling essential amino acid levels in CHO cell cultures. *Biotechnology and Bioengineering*, 119, 435–451. https://doi.org/10.1002/bit.27994