



ARTICLE

Forecasting and control of lactate bifurcation in Chinese hamster ovary cell culture processes

John Schmitt¹  | Brandon Downey¹ | Justin Beller¹ | Brian Russell¹ | Anthony Quach¹ | David Lyon¹ | Meredith Curran² | Bhanu Chandra Mulukutla²  | Chia Chu³

¹Lonza, R&D, Bend, Oregon

²Pfizer, Bioprocess R&D, Andover, Massachusetts

³Pfizer, Bioprocess R&D, Chesterfield, Missouri

Correspondence

John Schmitt, Lonza, 64550 Research Road, Bend, OR 97702.

Email: john.schmitt@lonza.com

Abstract

Biomanufacturing exhibits inherent variability that can lead to variation in performance attributes and batch failure. To help ensure process consistency and product quality the development of predictive models and integrated control strategies is a promising approach. In this study, a feedback controller was developed to limit excessive lactate production, a widespread metabolic phenomenon that is negatively associated with culture performance and product quality. The controller was developed by applying machine learning strategies to historical process development data, resulting in a forecast model that could identify whether a run would result in lactate consumption or accumulation. In addition, this exercise identified a correlation between increased amino acid consumption and low observed lactate production leading to the mechanistic hypothesis that there is a deficiency in the link between glycolysis and the tricarboxylic acid cycle. Using the correlative process parameters to build mechanistic insight and applying this to predictive models of lactate concentration, a dynamic model predictive controller (MPC) for lactate was designed. This MPC was implemented experimentally on a process known to exhibit high lactate accumulation and successfully drove the cell cultures towards a lactate consuming state. In addition, an increase in specific titer productivity was observed when compared with non-MPC controlled reactors.

KEYWORDS

bioprocess predictive modeling, lactate accumulation, model predictive control

1 | INTRODUCTION

Since approval of the first biotherapeutic produced using recombinant DNA technology in 1982, biologics have grown significantly and encompassed approximately 25% of the total pharmaceutical market in 2016 (Haydon, 2017). Today it is estimated that biologics represent 30% of all major biopharmaceutical development pipelines (Lybecker,

2016). Though there are many clear benefits of expressing recombinant proteins using living cells, the complexity of cell culture does present challenges to upstream process control in a drug substance manufacture campaign. To help address the inherent biological complexity, analytics have evolved to allow measurement of more parameters that are indicative of cell culture performance. With increased emphasis on the development of in process analytics and the advancement in

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2019 The Authors. *Biotechnology and Bioengineering* Published by Wiley Periodicals, Inc.

auto-sampling capabilities, a modern bioreactor experiment can easily generate hundreds of data points in both online and offline measurements. Therefore, the tasks of data management and mining for correlative relationships between process inputs and cell culture performance have become increasingly daunting for process development scientists. Various multivariate analysis techniques, such as principal component analysis and partial least squares (PLS), were applied to analyze bioprocess data to identify parameters that are characteristic of highly productive cell culture processes (Charaniya et al., 2010; Le et al., 2012; Rathore et al., 2015). After identifying these key cell culture levers, computational models were built to predict important culture performance attributes, including viable cell density, viability, carbon source nutrient levels, and titer (Elif Seyma Bayrak, Cinar & Undey, 2015; Sokolov et al., 2015). Moreover, the predictive control approach was extended to protein attributes like glycosylation and charge variants to dial in the desired product quality (Downey et al., 2017; Jan Bechmann et al., 2015; Sommeregger et al., 2017; Zupke et al., 2015).

In this study, we demonstrated the ability to forecast and control cell culture process performance. The performance marker we used to demonstrate this ability was lactate concentration, a key marker of metabolic state and indicator of a favorable batch process. Prior work demonstrated the impact cell metabolic state has on product titer and product quality (Fan et al., 2015; Li, Wong, Vijayasankaran, Hudson, & Amanullah, 2012; Toussaint, Henry, & Durocher, 2016). Amino acid and glucose metabolism in fed-batch Chinese hamster ovary (CHO) cell culture affects antibody production and glycosylation (Fan et al., 2015). In addition, accumulation of a major byproduct of amino acid and glucose metabolism, lactate, can profoundly impair process performance. Various bioprocess strategies, such as glucose limiting and cofactor additives to the cell culture media, successfully suppressed lactate accumulation (Gagnon et al., 2011; Qian et al., 2011; Yuk et al., 2014). However, what differentiates the work presented here, is the ability to fine-tune processes to achieve desirable profiles. Lactate was an ideal target to develop the proof of concept, as it is easily measured and its metabolism is well understood. Here we sought to use historical process data to develop (a) a forecasting model for lactate accumulation behavior early in a run to determine the extent of the variability, (b) an MPC to control the lactate variability.

To construct both the forecast model and MPC, historical process time-course data were compiled from approximately 128 runs from five different CHO clonal lines. In this study, we demonstrated (a) process data early in run can be used to predict ultimate lactate behavior for each run, (b) process data can be used to construct an MPC to control lactate production in the face of disturbances. Applying the predictive controller in confirmation experiments, we were successful in suppressing lactate accumulation in bioreactors and observed beneficial improvement in specific productivity of the cell culture.

2 | MATERIALS AND METHODS

2.1 | Data compilation

Process data from 128 1-liter fed-batch runs were used for model development. These runs consisted of five different clones originating from the same parental CHO cell line. The same basal media and feeding strategy were used for all runs. Offline process parameters were measured daily using a Nova FLEX (Nova Biomedical, Waltham, MA). Online parameters were measured continuously using a BioNet control system (Broadley James, Irvine, CA). All time-course process parameters recorded for each run are summarized in the Appendix Table A1.

2.2 | Classification model development

The process data were used to develop a classification model capable of predicting the percent probability that a bioreactor run would end in a favorable or unfavorable lactate state. The model development workflow is summarized in Figure 1 and detailed in the following subsections.

2.3 | Data preprocessing

The collected experimental process data represent a cube of three dimensions defined by: batch runs, attributes (process time-course measurements), and process time. Offline and online process data recorded through Day 10 were interpolated or downsampled onto a uniform daily grid to enforce measurement consistency within and across batch runs. Missing attribute data for each run were linearly

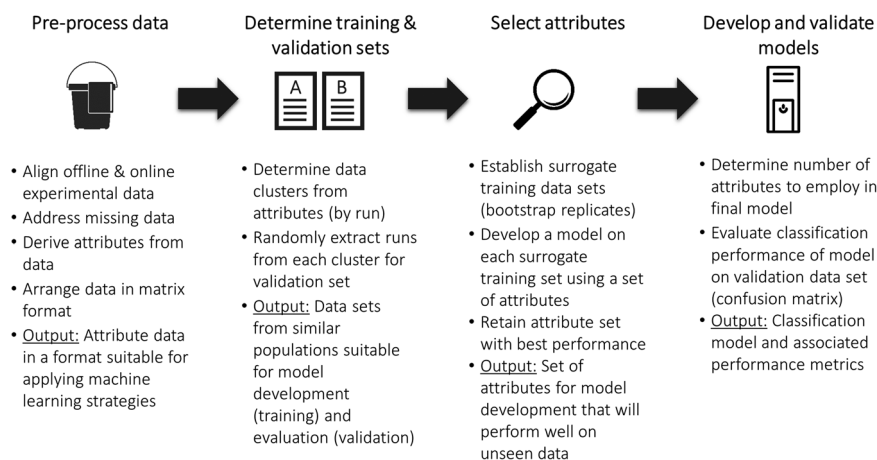


FIGURE 1 Overview of the classification model development workflow

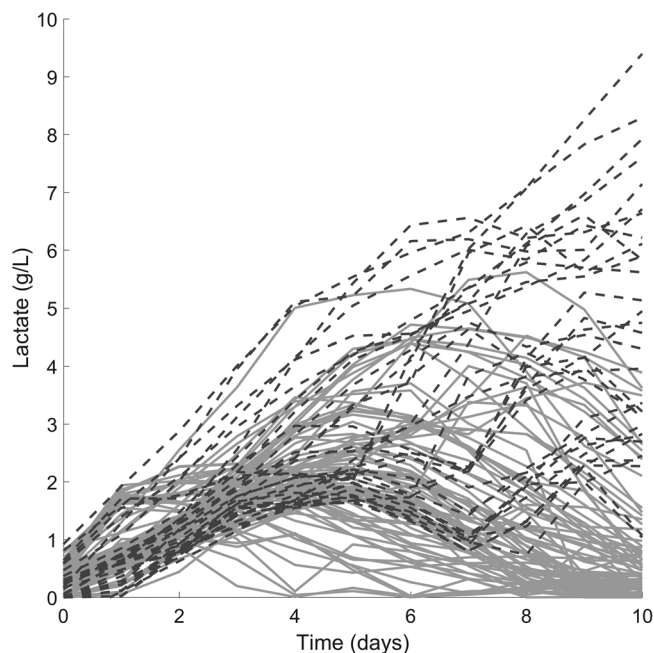


FIGURE 2 Lactate concentration profiles associated with the bioreactor runs considered in this study. Manually classified (dashed: unfavorable lactate end state, solid: favorable lactate end state)

interpolated or extrapolated from nearby time-course measurements. Aligned batch process time-course data were subsequently batch-unfolded into a design matrix, as in (Nomikos & MacGregor, 1995), for use in machine learning model development.

Process engineers manually evaluated the lactate profile of each run to determine the lactate end state (favorable/unfavorable). These manual classifications were used as the truth metric in the development of a classification model to accurately predict end-of-run lactate state. The lactate profiles of all runs and their associated lactate behavior classification are illustrated in Figure 2.

2.4 | Development of training and validation data sets

A split sample methodology was used in developing models and evaluating their performance on unseen data (Bishop, 2006). Data were split into a training and validation set, to identify model overtraining and performance. To better ensure that the training and validation data sets were comprised of similar run populations, a cluster feature selection algorithm was used on the favorable and unfavorable lactate runs separately to identify natural clusters of runs based upon the associated process measurements (Dy & Brodley, 2004). The validation data set was established by randomly extracting 30% of the runs from each resulting cluster for both the favorable and unfavorable lactate clustering results, with the remaining runs comprising the training data set. By establishing similar populations of runs in the training and validation data sets, differences in model performance on these two sets are more indicative of overtraining rather than differences between the population samples present in each set.

2.5 | Attribute selection

Each daily metabolite concentration and process condition is a potential attribute to employ for modeling purposes as a result of the batch unfolding process. For modeling strategies requiring attribute selection, a wrapper-based forward feature selection algorithm was used to determine the subset of attributes that produced a model with the best capability of distinguishing between runs with favorable and unfavorable lactate end states (Kohavi & John, 1997). As forward feature selection is known to overfit when applied to the training set directly, 400 bootstrap replicates were constructed from the training data set and used in the attribute selection process as a cross-validation strategy (Efron & Tibshirani, 1997; Hastie, Tibshirani, & Friedman, 2001). Oversampling of unfavorable lactate runs was performed in the construction of each bootstrap replicate to better ensure that the attributes selected could distinguish between favorable and unfavorable lactate classes rather than accurately identifying only the dominant class (favorable lactate; Chawla, 2005). As our goal was to determine the expected type of lactate profile as early as possible in each run, the feature selection process was limited to select from attributes present within 3 to 5 days from the start of a run.

2.6 | Model development and evaluation

The attribute selection strategy determined the set of attributes with the best classification performance for a particular modeling strategy in order of their importance. The final number of attributes to employ was identified as the number for which the difference between training and validation performance began to increase. This attribute set was subsequently used to create the final classification model using the entire training data set.

Performance of each classification model was evaluated for each data set via the confusion matrix (Davis & Goadrich, 2006). The confusion matrix summarizes model performance by detailing the number of favorable lactate (respectively, unfavorable lactate) runs correctly predicted as favorable lactate (resp., unfavorable lactate) as well as the number of runs incorrectly classified. Overall classification accuracy is defined as the number of runs correctly predicted divided by the total number of runs.

2.7 | Identification of critical process parameters

For final classification models comprised of a combination of models, a permutation strategy was used to determine the importance of an individual attribute to the classification prediction (Strobl, Boulesteix, Zeileis, & Hothorn, 2007). Specifically, individual attribute importance to the classification result was evaluated by permuting the values of the attribute in the validation data set 100 times. For an attribute important to the classification result, randomly permuting its value in the validation data set leads to an increase in the classification error metric as compared with the same metric for the nonpermuted data set. For each permutation of a particular attribute,

the difference between the cross-entropy classification error on the permuted and nonpermuted data sets was determined. Normalized attribute importance was subsequently computed as the mean of those differences divided by the standard deviation. Attributes with normalized predictor importance greater than unity were retained as important attributes.

2.8 | Process control

While classification models can provide an early indication of run end state, they do not address how to prevent unfavorable end states via run intervention. The second goal of this study was to develop a proof of concept model predictive controller (MPC) from the provided fed-batch process data and evaluate its effectiveness in driving runs to a favorable lactate state in both simulation and experiment.

2.9 | Development of dynamic reduced-order model of lactate concentration

Model-based control design requires developing a reduced-order model that adequately represents the input–output dynamics of the system to be controlled. Constructing this model requires identifying a set of manipulated variables that have a strong influence on the output(s) of interest. The explicit assumption in model development is that knowledge of the manipulated variable values, in conjunction with knowledge of the prior output values, is sufficient to predict future output behavior. The development of a linear multiple-input, single-output reduced-order model is motivated in this study both by the ability of linear models to locally represent nonlinear system dynamics and the effectiveness of linear control strategies in controlling slowly time-varying systems.

The provided process data were used to create a time-varying autoregressive exogenous (ARX) model that predicts future lactate concentration values from the knowledge of prior lactate and manipulated variable values (Zhu, 2001). The relationship between

inputs and outputs in a multi-input, single-output ARX formulation is of the form:

$$y(t) = - \sum_{i=1}^{n_a} a_i y(t-i) + \sum_{j=1}^{n_i} \sum_{i=1}^{n_b} b_{ji} u_j(t-n_k-i+1), \quad (1)$$

where $y(t)$ is the output/controlled variable, $u_j(t)$ represents one of n_i manipulated variables, n_k is the time delay, n_a is the number of poles, n_b is the number of zeros, and a_i and b_{ji} are coefficients to be determined via the identification process. In a time-varying ARX model, the coefficients representing the influence of each parameter change with time (i.e., day), such that the model is time-varying. The ARX model as written in (1) is a one-step-ahead predictor; the value for the output at day t is determined from prior values of the output as well as current and prior values of the manipulated variables. This model can be extended into a multistep ahead predictor by using the output prediction from the prior day along with prescribed values for the manipulated variables, such as would be determined by a control strategy, to predict future output values.

Model parameters (n_a , n_b , n_k , a_i , b_{ji}) were determined by minimizing the multistep 0.632 bootstrap root mean square prediction error across 50 bootstrap replicates drawn with replacement from training data set runs. In these multistep simulations, recorded process data were used for the manipulated values while predicted output values from (1) were used for subsequent prediction days. For each run, the residuals between recorded output lactate concentrations and the predicted values were retained for computing the multistep root mean square error. The combination of model parameters with the minimum multistep 0.632 bootstrap root mean square prediction error was retained.

2.10 | Development of a MPC

An MPC strategy was used to regulate lactate concentration using a set of manipulated variables. As illustrated in Figure 3, an MPC prescribes the values for the manipulated variables over a control horizon from the knowledge of the desired lactate concentration

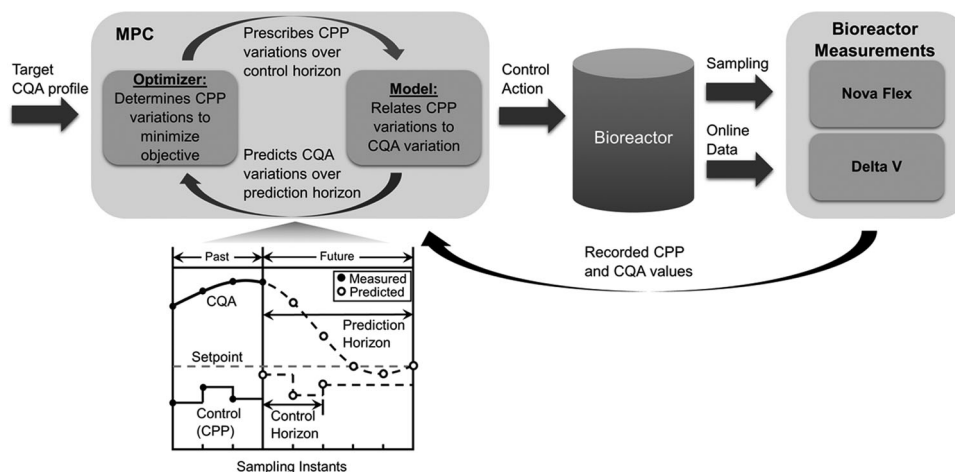


FIGURE 3 Overview of the model predictive control strategy

and prior values of the recorded manipulated variables and lactate concentration. The MPC employs the time-varying ARX model developed from historical process data to determine the values for the manipulated variables that will result in the lactate concentration reaching the desired value in the future. Lactate predictions are generated in a multistep fashion over the prediction horizon from a sequence of values for the manipulated variables over the control horizon. Optimal values for the manipulated variables are determined over the control horizon to minimize an objective function involving the deviation of the model output predictions from the desired trajectory over the prediction horizon. Once the optimal sequence of manipulated variable values is determined, only the first of these values are used in the bioreactor. At the next sampling instant, the lactate concentration is measured and the process repeats. Because the recorded, rather than predicted, lactate concentration is used in each subsequent optimization cycle, the prediction errors that can accumulate in a multistep ahead prediction are limited in their impact in controller implementation.

The design of an MPC requires specifying the number of design parameters to compute the objective function optimized during controller operation

$$J = \sum_{i=1}^P \{w_i^y (\hat{y}(t+i) - r(t+i))\}^2, \quad (2)$$

where P is the number of days in the prediction horizon; \hat{y} is the predicted value of the lactate concentration from the reduced-order model; r is the value of the lactate concentration for the desired reference trajectory; w_i^y is the weighting to be applied to the difference between the predicted output and the reference trajectory for each day in the prediction horizon.

The objective function penalizes differences in the predicted output from the reference trajectory. Different weightings can be used across the days of the prediction horizon if concern exists regarding multistep prediction accuracy of the reduced-order model far into the future. The optimal values for the manipulated variables over the control horizon are achieved by minimizing the objective function with respect to both bound and rate constraints on the manipulated variables.

3 | EXPERIMENTAL EVALUATION OF PROCESS CONTROL

3.1 | Cell lines and media

Clones derived from CHOK1SV® cells and stably expressing recombinant proteins were routinely cultured in suspension using commercially available CD-CHO AGT™. Inoculum trains were maintained in shake flasks in Kuhner incubators at 37°C, 5% CO₂, with no humidity control. Cells were regularly passaged to maintain exponential growth and expanded as needed to inoculate bench-scale bioreactors for experiments described herein.

3.2 | Fed-batch bioreactor operation

Two-liter scale glass bioreactors (Broadley James) were used to perform the fed-batch experiments. Cells were inoculated into Pfizer's proprietary production media formulation. Reactor cultures were fed at predetermined rates using Pfizer's proprietary nutrient feed. Bioreactor conditions such as pH, DO, and temperature set points varied according to the experimental plan. Culture pH was controlled using CO₂ sparge and base titrant addition. Dissolved oxygen was maintained at set points using oxygen sparge on demand. Culture temperature was controlled using a heating jacket. Concentrated glucose stock solutions were added as needed to maintain at least 0.5 g/L residual glucose concentration throughout the production run. Reactor experiments were performed for a 12-day duration.

3.3 | Implementation of MPC in bioreactors

The performance of the MPC was tested by experimentally adjusting the nutrient feed rate and pH set point values as prescribed by the controller. The MPC was integrated into a spreadsheet (Microsoft Excel) where the measured process inputs were entered, and the controller-prescribed action updated based on the newly entered process measurements. Control actions prescribed by the MPC were retrieved from the spreadsheet calculations and manually implemented. Culture pH was maintained via a local PID controller with a 0.15 deadband; changes in the pH set point of the local controller were implemented as prescribed by the MPC. The MPC-prescribed action was updated and implemented once every 24 hr, beginning after 72 hr culture duration.

4 | RESULTS

4.1 | Classification models accurately predict lactate end-state from early-run process data

Classification models were developed to predict the final lactate state (favorable/unfavorable) from process data present through a specified end day (days 3, 4, and 5). For each end day considered, the following classification models were developed: linear discriminant analysis (LDA), classification trees, LDA applied to partial least squares scores (PLS-LDA), support vector machines and logistic regression. Each individual model was computed from the batch-unfolded process data present in the training data set using functions (*fitcdiscr*, *fitctree*, *plsregress*, *fitcsvm*, *fitglm*) from the Matlab statistics and machine learning toolbox (R2016b). A class threshold probability of 0.5 (i.e., 50%) was used across classification models. Runs with a predicted unfavorable lactate probability greater than or equal to the 0.5 threshold were classified as an unfavorable lactate state and those below 0.5 classified as a favorable lactate state. Ensembles of individual model results were also considered with the ensemble unfavorable lactate probability determined as the median of the unfavorable lactate probabilities arising from the individual models.

TABLE 1 Fed-batch classification model performance on training and validation data sets

Day	Model(s)		# True negatives	# False positives	# False negatives	# True positives	Percent accuracy
3	PLS-LDA(13 components)	Training	63	1	7	17	91%
		Validation	25	3	4	8	83%
4	Ensemble PLS-LDA (12 components), Classification tree, LDA (18 attributes)	Training	64	0	1	23	99%
		Validation	26	2	3	9	88%
5	PLS-LDA (6 components)	Training	57	7	4	20	88%
		Validation	25	3	2	10	88%

Abbreviation: PLS-LDA: linear discriminant analysis applied to partial least squares scores.

No additional training was conducted to determine ensemble performance; all possible ensemble combinations were evaluated from previously constructed models and the ensemble with the highest classification accuracy was retained.

Models consistently yielding good classification accuracy across all end days included: PLS-LDA, LDA, classification trees, and ensembles of these models. Elements of the confusion matrices of the best performing classification models are presented in Table 1. As illustrated in Table 1, the classification models were able to accurately classify favorable and unfavorable lactate runs with validation accuracy ranging between 83% (Day 3) and 88% (Day 4 & 5). Though the Day 4 and 5 models achieved equivalent validation classification accuracy in total, the Day 4 ensemble model produced more consistent validation performance across clones. Identified attributes that correlated with the lactate end state for each model are presented in the Appendix in Table A2. Attributes commonly appearing across models include metabolites (glutamate, glucose, and glutamine) and attributes related to pH modulation (CO₂ sparge rate). Critically evaluating these attributes from a biological perspective suggests that glutamate concentration, tricarboxylic acid (TCA) cycle substrate supplementation and ammonium ion

concentration were associated with runs that ended in a favorable lactate state.

4.2 | Reduced-order dynamic models predict lactate evolution as a function of pH and nutrient feed volume

Considering their importance in achieving a favorable lactate end state and the ease of implementation via local control loops, nutrient feed volume, and pH set point were used as manipulated variables in these initial experiments to control lactate accumulation. The number of prior manipulated and output variable terms to employ in the dynamic model relating these variables was determined using bootstrapping of the runs in the training data set as a cross-validation strategy. The combination with the minimum bootstrap root mean square error in multistep ahead prediction was retained. As detailed further in the Appendix, the best model structure identified used three prior lactate values, a single prior nutrient feed volume value, and four prior values of the pH set point. The ARX model coefficients of (1) associated with the best model structure are presented in the Appendix in Table A3.

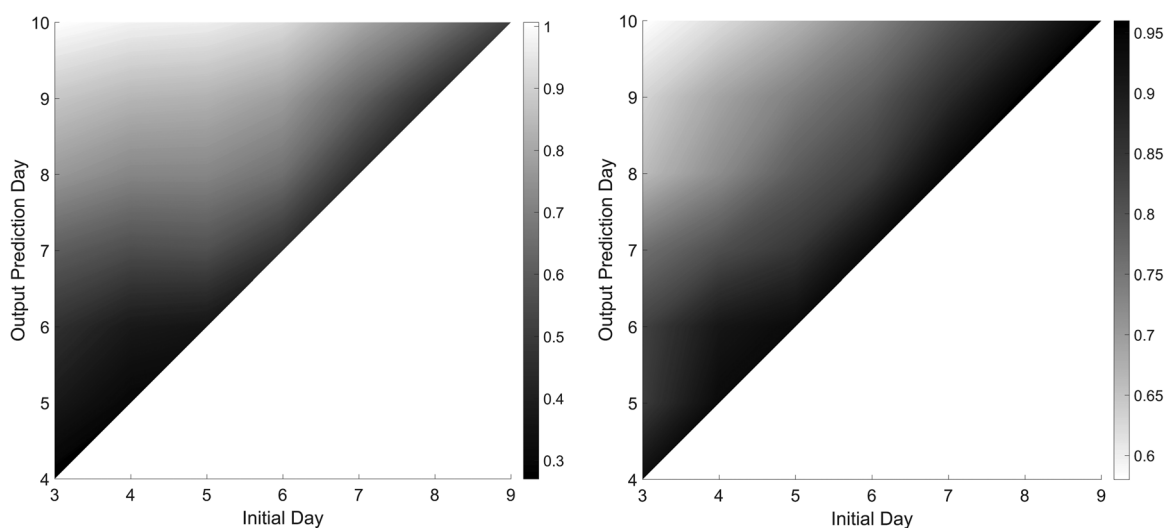


FIGURE 4 Filled contour plots of root mean square error in prediction (left) and coefficient of determination (R^2 , right) for one-step-ahead predictions (diagonal) and multistep ahead predictions (off-diagonal values) across all runs present in the validation data set. Darker colors indicate better performance in each plot

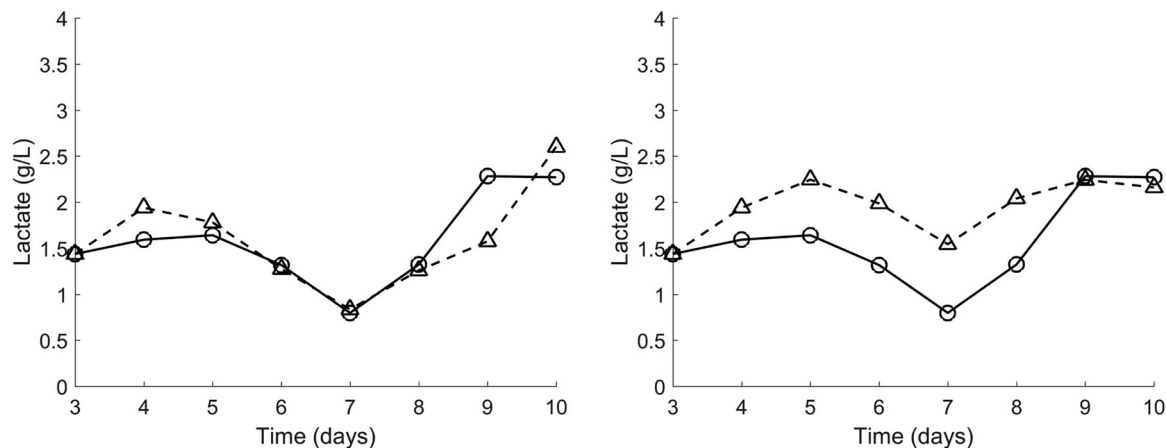


FIGURE 5 One-step (left) and multistep (right) model simulations of lactate concentration (g/L) for a bioreactor run present in the validation data set. Simulation initialized from experimental data provided through Day 3. Simulated and recorded values are represented by dashed and solid lines, respectively

ARX model performance was quantified by computing the one-step and multistep ahead root mean square error in prediction and associated coefficient of determination from each run end day. In each simulation, the experimental process data from the validation data set were used to generate the one-step and multistep ahead predictions. The predicted values for each output prediction day were used with the recorded lactate concentration values in computing the root mean square error and coefficient of determination. ARX model performance for the validation data set is illustrated in Figure 4, where values along the diagonal and off-diagonal represent one-step ahead and multistep ahead prediction performance, respectively.

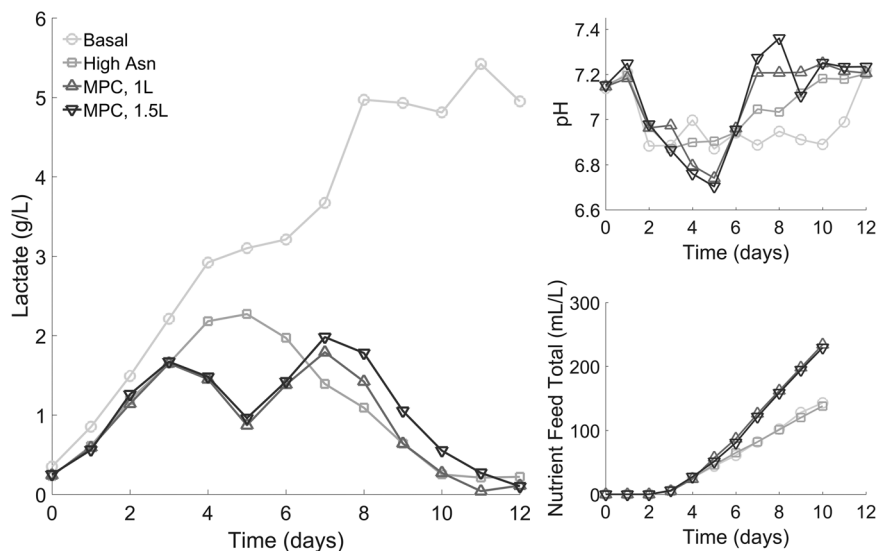
One-step and multistep ahead prediction performance is illustrated in Figure 5 for a simulation initiated from data present through Day 3. As illustrated in Figures 4 and 5, one-step-ahead predictions are more accurate than multistep ahead predictions over the course of the run, as prediction errors in multistep simulations cascade for predictions further into the future. Although prediction

errors accumulate in multistep simulations, trends in lactate concentration variation often match quite well across the prediction horizon. This, in conjunction with the model prediction accuracy results in simulation, identified the time-varying ARX model as the most promising for use in an MPC strategy.

4.3 | Model predictive control strategy drives cultures to a lactate consuming state

An MPC employing the time-varying ARX model was built in Matlab, with *fmincon* of the Matlab optimization toolbox (R2016b) used to minimize the cost function of Equation (2). Controller design parameters were initialized in simulation and tuned during preliminary experimental runs. Specifically, the desired lactate reference trajectory was set to zero for all days. The prediction and control horizons used were 7 days and 1 day, respectively. The prediction horizon was decreased after Day 3, as predictions were only required through Day 10. Values for manipulated variables after Day 10 were

FIGURE 6 Lactate concentration variation in response to changes in pH and nutrient feed volume prescribed by the MPC in the experiment. For MPC runs, control was implemented from the end of Day 3 onwards. Run conditions include: basal (no applied control, circles), feed supplemented with asparagine (no applied control, squares), MPC at 1 L working volume (triangles), MPC at 1.5 L working volume (inverted triangles). MPC: model predictive controller



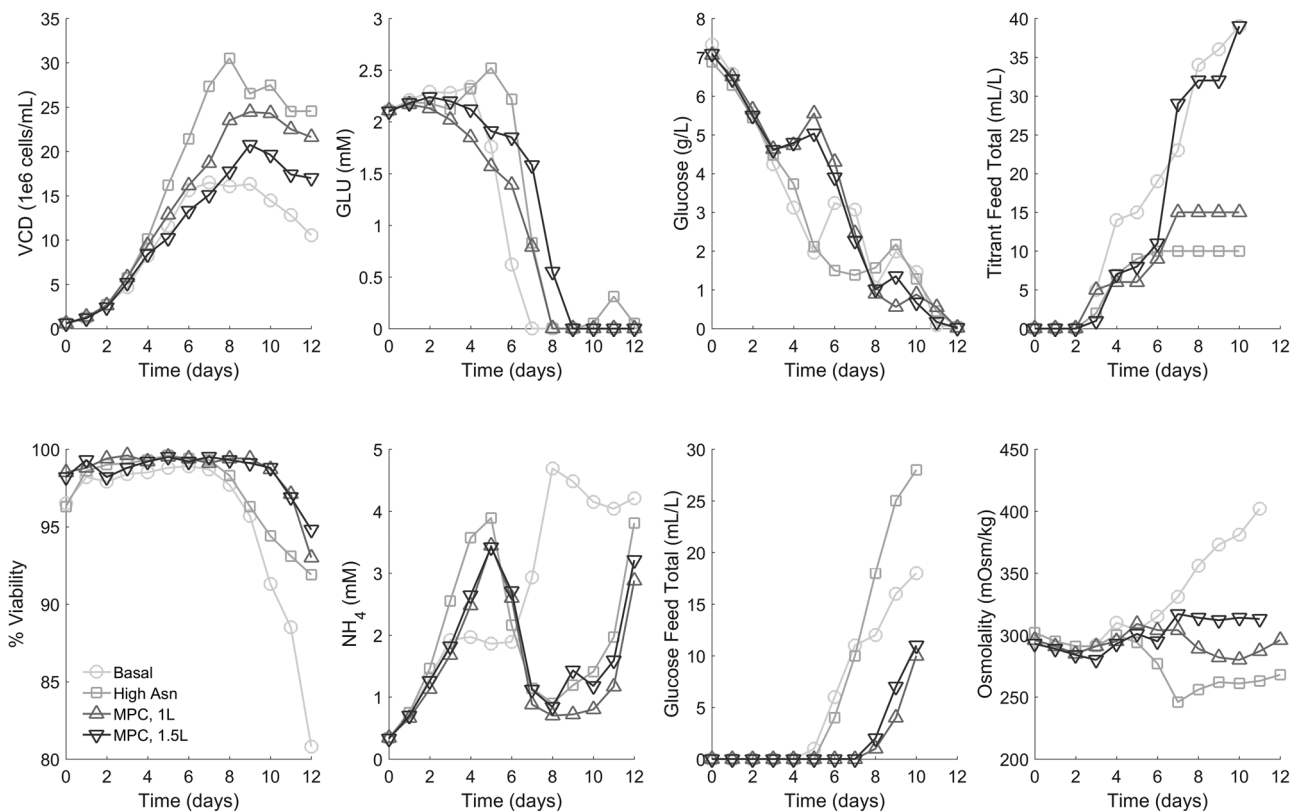


FIGURE 7 Supplementary cell culture measurements associated with control and MPC experiments. For MPC runs, control was implemented from the end of Day 3 onwards. Run conditions include: basal (no applied control, circles), feed supplemented with asparagine (no applied control, squares), MPC at 1 L working volume (triangles), MPC at 1.5 L working volume (inverted triangles). MPC, model predictive controller

maintained at the last controller-prescribed values. A long prediction horizon served to ensure that the full effect of variations in the manipulated variables through run end were considered, whereas, a short control horizon ensured aggressive control action in the manipulated variables. As prediction accuracy did not dramatically degrade over longer prediction horizons, all prediction errors were considered to contribute equally to the minimized cost function (i.e., all w_i^j of equation (2) were set to unity). Nutrient feed volume was constrained to remain between 1.8% and 3.6%, with maximum variations between days limited to $\pm 1.8\%$ on days 3 to 6 and $\pm 1.0\%$ otherwise. Bound constraints on pH were established at 6.7 and 7.2, with the maximum variation in pH between days set to ± 0.5 .

The resulting MPC was used in a series of experimental bioreactor runs to determine its efficacy in driving runs to a favorable lactate end state. Cell cultures used in the experiment were associated with a clone known to exhibit lactate accumulation in prior process development. While it is expected that lactate behavior under application of MPC would be similar for the other four clones used in the development of the predictive model used by the MPC, experimental verification of MPC performance for those clones remains an avenue for future work. Experimental MPC runs for the clone of interest were conducted alongside two control runs: a basal run with known lactate accumulation behavior and a second for which supplemental asparagine included in the feed had previously been identified via prior process development history to

achieve a favorable lactate end state under normal operating conditions. The performance of culture using this previously identified strategy serves as a basis of comparison for MPC performance in modifying lactate behavior. In this set of experiments, MPC-prescribed variations in pH set point and nutrient feed volume were used at the original reactor working volume (1 L) as well as a working volume of 1.5 L. As illustrated in Figure 6, both control runs performed as expected, with the basal and asparagine-supplemented runs ending in unfavorable and favorable lactate states, respectively. MPC runs, with controller-prescribed changes in pH set point and nutrient feed volume initiated at the end of Day 3, resulted in the cell culture achieving a favorable lactate end state with substantially lower lactate concentrations than the basal run.

Figure 7 illustrates other relevant cell culture measurements recorded for the control and MPC runs. As illustrated, MPC runs had a viable cell density less than the asparagine-supplemented run, but greater than the basal run. However, percent viability for both of the MPC runs was greater than that of either control run. The increase in nutrient feed volume prescribed by the MPC resulted not only in increased glucose concentration and total glucose feed volume but also in delayed depletion of glutamate as compared with the basal run.

The second set of experiments evaluated the ability of MPC to compensate for lactate-inducing disturbances in pH and glucose concentration. Elevated pH or glucose levels were used early in each

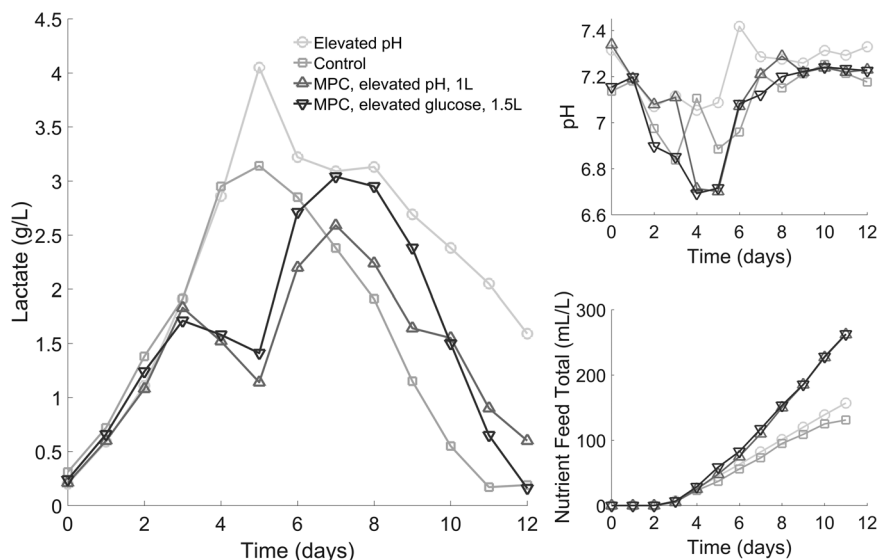


FIGURE 8 Lactate concentration variation in response to changes in pH and nutrient feed volume prescribed by the MPC in experiments with elevated pH and glucose concentrations. All experiments used feed supplemented with asparagine. For MPC runs, control was implemented from the end of Day 3 onwards. Run conditions include: elevated pH (no applied control, circles), only asparagine-supplemented (no applied control, squares), model predictive applied to culture with elevated initial pH level (triangles), MPC applied to culture with elevated initial glucose concentration (inverted triangles). MPC: model predictive controller

run to produce elevated lactate concentration levels. The asparagine-supplemented feed was used in all the runs of this experiment. Two control runs were used: one with normal pH and glucose levels and a second with elevated pH level (7.2 with 0.15 deadband). One MPC run used the same elevated pH level through Day 3 as in the corresponding control run while the second MPC run had an increased initial glucose concentration. As illustrated in Figure 8, while the elevated pH control run did not end in an unfavorable lactate state, it did evidence increased end lactate concentration compared with the control run with normal operating conditions. The MPC runs rejected the initial disturbances in pH and glucose, with both runs yielding lower end lactate concentrations than the elevated pH control run. Variations in other measured cell culture parameters followed similar trends to those evidenced in the initial experiments, as illustrated in the Appendix in Figure A1. In contrast to the prior runs, viable cell density for the MPC runs was similar to that evidenced for the control run without elevated pH. Increased nutrient feed volumes in the MPC runs resulted in increased ammonium ion concentration and delayed depletion of glutamate.

The MPC developed in this study was constructed specifically to reduce lactate accumulation, without regard to potential impacts on other process performance attributes. MPC impacts on a particular process performance attribute, cell productivity, were evaluated by processing retains to determine product titer and cell-specific productivity (Q_p). As illustrated in Figure 9, MPC runs for both unperturbed and perturbed initial conditions increased titer, compared to the control runs, by increasing cell-specific production. The observed difference in the Q_p , a calculated parameter based on volumetric titer and integral viable cell concentration, of the two MPC runs, was most likely due to mass transfer differences between

the two bioreactor working volumes which led to different peak cell densities as shown in Figure A1. But overall, the volumetric titers were consistent between the two MPC runs.

5 | DISCUSSION

In this project, we demonstrated the ability to construct an MPC to control lactate accumulation from historical process data commonly collected in a cell culture process. The MPC controlled lactate by forecasting its trajectory using metabolite and process data and then using this forecast to determine how to manipulate amino acid concentrations and pH, through the addition of nutrient feed and modulating the pH set point. While others have shown examples of using MPCs to control cell culture processes (Zupke et al., 2015), this study demonstrates that these controllers can be developed empirically using only commonly collected cell culture time-course data. In addition to enabling control of cell culture processes, this empirical path of developing a predictive model highlighted underlying biological mechanisms that led to further hypothesis generation and ultimately mechanistic understanding.

5.1 | Control of lactate in bioreactors is possible using models developed solely from commonly collected process data

One of the key obstacles to implementing MPC in cell culture processes is the amount and type of data required to construct the dynamic predictive model needed for the controller. Since lactate accumulation fundamentally stems from an overflow of pyruvate in cell metabolism,

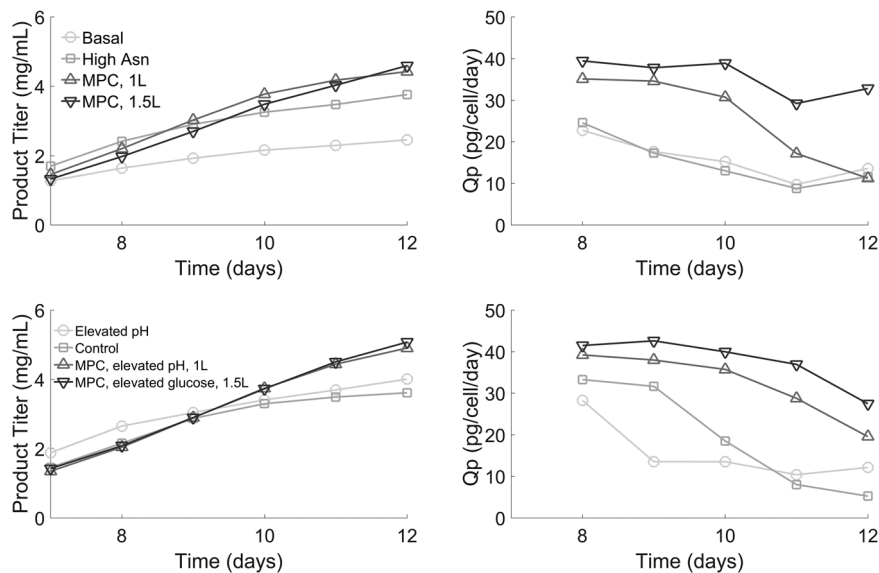


FIGURE 9 Product titer and Q_p for uncontrolled and controlled runs for unperturbed (top row) and lactate-inducing perturbations (bottom row) in initial conditions. Q_p : cell-specific productivity

we hypothesize that a data set incorporating many of the key inputs and outputs to cellular metabolism would be ideal for prediction. In fact, other studies have attempted to understand the causal factors of lactate accumulation through metabolic flux analysis (MFA; Chen, Bennett, & Kontoravdi, 2014; Verónica S. Martínez, Gray, Nielsen, & Quek, 2015; Wilkens, Altamirano, & Gerdtzen, 2011). Though useful for generating process understanding, MFA requires data that are expensive and time-consuming to generate. To construct predictive models, many cell culture data sets are required to ensure representation of the variability present in a given cell culture process. Due to a large number of data sets required, only the most commonly measured metabolites (i.e., those measured by a Nova bioanalyzer) are readily available for analysis. In this project, we successfully developed a forecasting model of lactate accumulation through the use of common, routinely captured measurements from cell culture processes. In addition, we identified a significant pattern associated with the forecasting of lactate behavior, and this allowed us to successfully control lactate in a model cell culture process. While having exhaustive cell culture measurements is ideal for control, this study illustrates that value can still be derived from commonly collected measurements provided that they are collected frequently with sufficient variation in process conditions to encompass the expected process variation of routine operation.

5.2 | Fundamental interpretation of model parameters and subsequent controller development allows establishment of a causal relationship between a manipulated variable and process output

We demonstrated the construction of an MPC using previously collected time-course data sets. Importantly, the predictions generated using this retrospective approach are purely correlations and do not establish causality by themselves. To create a predictive model that can be implemented in an MPC, the model must be sufficiently accurate in forecasting the process behavior as a function of an

independently manipulated control handle. This assumes a causal relationship exists between the independently manipulated control handle and the process behavior. Therefore, an additional step is required to convert a predictive model generated using this retrospective approach (the type that would be used in a process monitoring application) into one that can be used in an MPC. In this study, the causal link was established by first inspecting the importance of the input variables contributing to the prediction, generating hypotheses explaining this pattern based on cellular metabolism, and then identifying potential process parameters that could be manipulated in the cell culture process. The ability of the predictive model to highlight patterns that contribute to a desired process behavior allows quick identification of potential control handles, which otherwise would require a great deal of experimentation, experience, and existing published knowledge to identify.

This retrospective approach represents one of two possible ways for constructing an MPC. Building a predictive model using a retrospective approach requires a large amount of data, ensuring the expected process variability present in the ultimate control scenario is captured. During the normal cell culture development process, enough data covering this expected variability may not be collected, thus not enabling the construction of an accurate predictive model. In contrast to the retrospective approach, there are approaches to collect data in a prospective manner, where the experimental data is collected in a manner specifically designed with MPC development in mind. Construction of large enough design of experiments to cover a wide design space, with sufficient replicates, may be able to generate the data necessary for the construction of a predictive model. In addition to this approach, a systems identification approach represents an accelerated means of producing the data needed to construct an MPC (Downey et al., 2017). A prospective approach also allows the experiment to be designed such that a causal relationship between selected manipulated variables and the process output is specifically identified. Furthermore, an initial retrospective modeling exercise can be used to help identify the manipulated variables for the prospective controller development

studies, and therefore increasing the success of identifying a causal input/control handle.

5.3 | Model outputs and constructed control strategy implicates amino acid consumption

Although the modeling outputs from this experiment helped identify the control handles used to control lactate accumulation, they also highlight important underlying biological mechanisms leading to lactate production within these CHO cell lines. Under normal physiological conditions, pyruvate derived from the breakdown of glucose is commonly shuttled into the TCA cycle. Generally speaking, the conversion of pyruvate into lactate is an energetically unfavorable process, and in mammalian physiology is only carried out in the absence of oxygen (anaerobic) or to recycle cytosolic NAD⁺/NADH pools (Muller et al., 2012). This phenomenon of high lactate accumulation is observed in cancerous cells and tumors and is referred to as the “Warburg effect”. The “Warburg effect” is characterized by high consumption of glucose, high productivity of lactate, and a high rate of cell growth and division (Vander Heiden, Cantley, & Thompson, 2009). Indeed, the behavior of these cells in the absence of our developed control strategy resembles the behavior of a cell demonstrating the “Warburg effect”, and since these CHO cells are immortalized cells it is not fully surprising for them to exhibit a cancerous phenotype.

Through the attribute selection process, the pattern emerged highlighting the role amino acid consumption has on lactate accumulation. In particular, attribute selection highlighted increased glutamate consumption and NH₄⁺ production as indicators of a favorable lactate run. This suggests an important role of amino acid feeding in bioprocessing, that besides increasing the building blocks for protein synthesis, amino acids may also represent an important fuel source for CHO cells. In addition, the attribute selection process highlighted the ability for increased feeding of amino acids to reduce glucose consumption and therefore lactate production. Indeed, our ability to control lactate production by increasing the amount of amino acids supplied during fermentation further supports the role of amino acid consumption on lactate production.

5.4 | Biological implications of amino acid control of lactate accumulation

Two potential hypotheses that help explain the observation of increased amino acid consumption leading to decreased lactate production both revolve around central metabolism and primarily the role of the TCA cycle in converting amino acids into energy. One hypothesis suggests that the TCA cycle in these CHO cells is truncated under the normal feeding strategy, which causes the cells to rely on glycolysis and lactate fermentation for quick energy generation. When these depleted amino acids are replenished by feeding, as in our control strategy, cells can redirect more pyruvate into the TCA cycle instead of converting them to lactate and alanine (Duarte et al., 2014). A second hypothesis, though not mutually exclusive to the first hypothesis, concerns the ability of these cells to incorporate pyruvate produced in the cytosol as a product

of glycolysis into the TCA cycle. The mitochondrial pyruvate carrier is a transporter responsible for moving lactate from the cytosol and into the mitochondria (Herzig et al., 2012). The observation of high lactate accumulation in the absence of increased amino acid feeding suggests that the pyruvate produced through the breakdown of glucose during glycolysis is not being utilized in the mitochondria, and therefore converted to lactate. Future experimentation investigating the flux of pyruvate into the TCA cycle may shed light on the high lactate accumulating phenomena and potentially lead to a solution for the metabolic issue (i.e., cell line engineering) in addition to the implementation of a process control strategy.

5.5 | Predictive modeling and control in bioprocessing

In the age of “Big Data,” there is a push to ensure that industries are leveraging as much value out of their data as they can. Throughout the biomanufacturing industry, tremendous amounts of data are collected along the biotherapeutic development pipeline. Though the work we presented here focused on the construction of a control strategy for an upstream process, the development of predictive models may help enlighten other areas of the pipeline including cell line development, media development, downstream optimization, and structural-functional studies (i.e., CQAs). Through identification of attributes and their predictive power in process behavior, new hypotheses may be generated and new avenues of process control may be revealed. The biomanufacturing process is riddled with multivariate problem statements that make it difficult to develop control strategies. Through the use of MPC, solutions to complex manufacturing issues can be developed without full comprehension of the mechanisms driving the particular problem. This can yield hypotheses and unveil mechanisms unknown to even the most knowledgeable of scientists. However, challenges still exist regarding the implementation of MPC within a validated GMP environment. As MPC prescribes changes to process parameters in reaction to measurements quantifying culture behavior, it represents a departure from the recipe-based control strategies historically used. Implementation of MPC in a validated environment will require an appropriate definition of a design space for MPC operation. This design space can be translated into hard constraints in the objective function used by the MPC to prevent operation outside of the design space.

In conclusion, as more and more industries drive towards automation and complete process transparency, the need for biomanufacturers to establish the framework for MPC construction is clearer. The integration of MPCs in biomanufacturing will help to fully realize the lower cost, lower risk, and more integrated regulatory compliant processes of the future.

ACKNOWLEDGMENTS

Lonza has applied for a U.S. patent (Application Number 62/588,464) entitled “Process and system for propagating cell cultures while preventing lactate accumulation” covering the development of

predictive models and associated control strategies for quality attributes associated with cell culture.

ORCID

John Schmitt  <http://orcid.org/0000-0002-9137-9540>

Bhanu Chandra Mulukutla  <http://orcid.org/0000-0002-7515-9599>

REFERENCES

- Bishop, CM. (2006). *Pattern recognition and machine learning (Information Science and Statistics)*. Berlin, Heidelberg: Springer-Verlag.
- Charaniya, S., Le, H., Rangwala, H., Mills, K., Johnson, K., Karypis, G., & Hu, WS. (2010). Mining manufacturing data for discovery of high productivity process characteristics. *Journal of Biotechnology*, 147(3-4), 186–197. <https://doi.org/10.1016/j.jbiotec.2010.04.005>
- Chawla, NV. (2005). Data mining for imbalanced datasets: An overview. In O. Maimon, & L. Rokach (Eds.), *The data mining and knowledge discovery handbook* (pp. 853–867). Springer.
- Chen, N., Bennett, MH., & 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>
- Davis, J., & Goadrich, M. (2006). *The relationship between Precision-Recall and ROC curves*. Paper presented at the Proceedings of the 23rd international conference on Machine learning, Pittsburgh, Pennsylvania, USA. <https://dl.acm.org/citation.cfm?doid=1143844.1143874>
- Downey, B., Schmitt, J., Beller, J., Russell, B., Quach, A., Hermann, E., ... Breit, J. (2017). A system identification approach for developing model predictive controllers of antibody quality attributes in cell culture processes. *Biotechnology Progress*, 33(6), 1647–1661. <https://doi.org/10.1002/btpr.2537>
- Duarte, TM., Carinhas, N., Barreiro, LC., Carrondo, MJ., Alves, PM., & Teixeira, AP. (2014). Metabolic responses of CHO cells to limitation of key amino acids. *Biotechnology and Bioengineering*, 111(10), 2095–2106. <https://doi.org/10.1002/bit.25266>
- Dy, JG., & Brodley, CE. (2004). Feature selection for unsupervised learning. *Journal of Machine Learning Research*, 5, 845–889.
- Efron, B., & Tibshirani, R. (1997). Improvements on cross-validation: The 632+ bootstrap method. *Journal of the American Statistical Association*, 92(438), 548–560.
- Elif Seyma Bayrak, TW., Cinar, A., Undey, C. (2015, June). *Computational modeling of fed-batch cell culture bioreactor: Hybrid agent-based approach*. Paper presented at the 9th International Symposium on Advanced Control of Chemical Processes, Whistler, British Columbia, Canada.
- Fan, Y., Jimenez Del Val, I., Muller, C., Wagtberg Sen, J., Rasmussen, SK., Kontoravdi, C., ... Andersen, MR. (2015). Amino acid and glucose metabolism in fed-batch CHO cell culture affects antibody production and glycosylation. *Biotechnology and Bioengineering*, 112(3), 521–535. <https://doi.org/10.1002/bit.25450>
- Gagnon, M., Hiller, G., Luan, YT., 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>
- Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The elements of statistical learning*. New York, NY, USA: Springer New York Inc.
- Haydon, I. (2017). Biologics: The pricey drugs transforming medicine. <https://theconversation.com/biologics-the-pricey-drugs-transforming-medicine-80258>
- Herzig, S., Raemy, E., Montessuit, S., Veuthey, JL., Zamboni, N., Westermann, B., ... Martinou, JC. (2012). Identification and functional expression of the mitochondrial pyruvate carrier. *Science*, 337(6090), 93–96. <https://doi.org/10.1126/science.1218530>
- Jan Bechmann, FR., Gebert, L., Schaub, J., Greulich, B., Dieterle, M., Bradl, H. (2015). *Process parameters impacting product quality*. Paper presented at the 24th European Society for Animal Cell Technology Meeting, Barcelona, Spain.
- Kohavi, R., & John, GH. (1997). Wrappers for feature subset selection. *Artificial Intelligence*, 97(1-2), 273–324. [https://doi.org/10.1016/s0004-3702\(97\)00043-x](https://doi.org/10.1016/s0004-3702(97)00043-x)
- Le, H., Kabbur, S., Pollastrini, L., Sun, Z., Mills, K., Johnson, K., ... Hu, WS. (2012). Multivariate analysis of cell culture bioprocess data—lactate consumption as process indicator. *Journal of Biotechnology*, 162(2-3), 210–223. <https://doi.org/10.1016/j.jbiotec.2012.08.021>
- Li, J., Wong, CL., Vijayasankaran, N., Hudson, T., & Amanullah, A. (2012). Feeding lactate for CHO cell culture processes: Impact on culture metabolism and performance. *Biotechnology and Bioengineering*, 109(5), 1173–1186. <https://doi.org/10.1002/bit.24389>
- Lybecker, K. (2016). The biologics evolution in the production of drugs. <http://www.fraserinstitute.org>
- Muller, M., Mentel, M., van Hellemond, JJ., Henze, K., Woehle, C., Gould, SB., ... Martin, WF. (2012). Biochemistry and evolution of anaerobic energy metabolism in eukaryotes. *Microbiology and Molecular Biology Reviews*, 76(2), 444–495. <https://doi.org/10.1128/MMBR.05024-11>
- Nomikos, P., & MacGregor, JF. (1995). Multivariate SPC charts for monitoring batch processes. *Technometrics*, 37(1), 41–59.
- Qian, Y., Khattak, SF., Xing, Z., He, A., Kayne, PS., Qian, NX., ... Li, ZJ. (2011). Cell culture and gene transcription effects of copper sulfate on Chinese hamster ovary cells. *Biotechnology Progress*, 27(4), 1190–1194. <https://doi.org/10.1002/btpr.630>
- Rathore, AS., Kumar Singh, S., Pathak, M., Read, EK., Brorson, KA., Agarabi, CD., & Khan, M. (2015). Fermentanomics: Relating quality attributes of a monoclonal antibody to cell culture process variables and raw materials using multivariate data analysis. *Biotechnology Progress*, 31(6), 1586–1599. <https://doi.org/10.1002/btpr.2155>
- Sokolov, M., Soos, M., Neunstoecklin, B., Morbidelli, M., Butte, A., Leardi, R., ... Broly, H. (2015). Fingerprint detection and process prediction by multivariate analysis of fed-batch monoclonal antibody cell culture data. *Biotechnology Progress*, 31(6), 1633–1644. <https://doi.org/10.1002/btpr.2174>
- 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>
- Strobl, C., Boulesteix, A-L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics*, 8(25).
- Toussaint, C., Henry, O., & Durocher, Y. (2016). Metabolic engineering of CHO cells to alter lactate metabolism during fed-batch cultures. *Journal of Biotechnology*, 217, 122–131. <https://doi.org/10.1016/j.jbiotec.2015.11.010>
- Vander Heiden, MG., Cantley, LC., & Thompson, CB. (2009). Understanding the Warburg effect: The metabolic requirements of cell proliferation. *Science*, 324(5930), 1029–1033. <https://doi.org/10.1126/science.1160809>
- Verónica S. Martínez, MB., Gray, P., Nielsen, L., & Quek, L-E. (2015). Dynamic metabolic flux analysis using B-splines to study the effects of temperature shift on CHO cell metabolism. *Metabolic Engineering Communications*, 2, 46–57.
- Wilkens, CA., Altamirano, C., & Gerdtsen, ZP. (2011). Comparative metabolic analysis of lactate for CHO cells in glucose and galactose. *Biotechnology and Bioengineering*, 16(4), 714–724. <https://doi.org/10.1007/s12257-010-0409-0>
- Yuk, IH., Zhang, JD., Ebeling, M., Berrera, M., Gomez, N., Werz, S., ... Szperalski, B. (2014). Effects of copper on CHO cells: Insights from gene expression analyses. *Biotechnology Progress*, 30(2), 429–442. <https://doi.org/10.1002/btpr.1868>
- Zhu, Y. (2001). *Multivariable system identification for process control*. New York, NY: Elsevier Science Inc.

Zupke, C., Brady, L.J., Slade, P.G., Clark, P., Caspary, R.G., Livingston, B., ... Bailey, R.W. (2015). Real-time product attribute control to manufacture antibodies with defined N-linked glycan levels. *Biotechnology Progress*, 31(5), 1433–1441. <https://doi.org/10.1002/btpr.2136>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Schmitt J, Downey B, Beller J, et al. Forecasting and control of lactate bifurcation in Chinese hamster ovary cell culture processes. *Biotechnology and Bioengineering*. 2019;116:2223–2235. <https://doi.org/10.1002/bit.27015>