

# Using MVDA with Stoichiometric Balances to Optimize Amino Acid Concentrations in Chemically Defined CHO Cell Culture Medium for Improved Culture Performance

Taha Salim<sup>1</sup>, Gaurav Chauhan<sup>1</sup>, Neil Templeton<sup>2</sup>, and Wai Lam Ling<sup>1</sup>

<sup>1</sup>Merck & Co Inc

<sup>2</sup>Merck and Co Inc West Point

July 8, 2021

## Abstract

Chemically defined (CD) media are routinely used in the production of biologics in Chinese Hamster Ovary (CHO) cell culture and provide enhanced raw material control. Nutrient optimized CD media is an important path to increase cell growth and monoclonal antibody (mAb) productivity in recombinant CHO cell lines. However, nutrient optimization efforts for CD media typically rely on multi-factorial and experimental design of experiment (DoE) approaches or complex mathematical models of cellular metabolism or gene expression systems. Moreover, the majority of these efforts are aimed at amino acids since they constitute essential nutrients in CD media as they directly contribute to biomass and protein production. In this study, we demonstrate the utilization of multi-variate data analytics (MVDA) coupled with amino acid stoichiometric balances (SBs) to increased cell growth and mAb productivity in efforts to reduce CD media development efforts. SBs measure the difference between theoretical demand of amino acids and the empirically measured fluxes to identify metabolic states of the cell. When coupled with MVDA, the statistical models were not only able to highlight key amino acids towards cell growth or productivity, but also provided direction on metabolic favorability of the amino acid. Experimental validation of our approach resulted in a 55% increase in total cell growth and about an 80% increase in total mAb productivity. Increased specific consumption of stoichiometrically balanced amino acids and decreased specific consumption of glucose was also observed in optimized CD media suggesting favorable consumption of desired nutrients and a potential for energy redistribution towards increased cellular growth or mAb productivity.

## Using MVDA with Stoichiometric Balances to Optimize Amino Acid Concentrations in Chemically Defined CHO Cell Culture Medium for Improved Culture Performance

Taha Salim

Senior Scientist

Merck & Co. Inc.

Kenilworth, NJ, USA

(201) 982-0330

Email: Tsalim786@gmail.com

Gaurav Chauhan

Principal Scientist

Merck & Co. Inc.

Kenilworth, NJ, USA

(240) 308-4173

Email: Gaurav.chauhan1@merck.com

Neil Templeton

Associate Principal Scientist

Merck & Co. Inc.

West Point, PA, USA

(434) 665-8060

Email: neil.templeton@merck.com

Wai Lam W. Ling<sup>1</sup>

Senior Principal Scientist

Merck & Co. Inc.

Kenilworth, NJ, USA

(646) 627-5525

Email: wai.lam.ling@merck.com

1. Wai Lam Ling has recently changed affiliations to Head of MS&T at Rocket Pharma (Cranbury, NJ, USA). Her updated email address is wling@rocketpharma.com

## Using MVDA with Stoichiometric Balances to Optimize Amino Acid Concentrations in Chemically Defined CHO Cell Culture Medium for Improved Culture Performance

\*Taha Salim <sup>a</sup>; \*Gaurav Chauhan <sup>a</sup>; Neil Templeton <sup>b</sup>; Wai Lam W. Ling <sup>a</sup>

<sup>a</sup> Merck & Co., Inc., Kenilworth, NJ, USA

<sup>b</sup> Merck & Co., Inc., West Point, PA, USA

\* Equal Contribution

### Abstract

Chemically defined (CD) media are routinely used in the production of biologics in Chinese Hamster Ovary (CHO) cell culture and provide enhanced raw material control. Nutrient optimized CD media is an important path to increase cell growth and monoclonal antibody (mAb) productivity in recombinant CHO cell lines. However, nutrient optimization efforts for CD media typically rely on multi-factorial and experimental design of experiment (DoE) approaches or complex mathematical models of cellular metabolism or gene expression systems. Moreover, the majority of these efforts are aimed at amino acids since they constitute essential nutrients in CD media as they directly contribute to biomass and protein production. In this study, we demonstrate the utilization of multi-variate data analytics (MVDA) coupled with amino acid stoichiometric balances (SBs) to increased cell growth and mAb productivity in efforts to reduce CD media development efforts. SBs measure the difference between theoretical demand of amino acids and the empirically measured fluxes to identify metabolic states of the cell. When coupled with MVDA, the statistical models were not only able to highlight key amino acids towards cell growth or productivity, but also provided direction on metabolic favorability of the amino acid. Experimental validation of our approach resulted in a 55% increase in total cell growth and about an 80% increase in total mAb productivity. Increased specific consumption of stoichiometrically balanced amino acids and decreased specific consumption of glucose was also observed

in optimized CD media suggesting favorable consumption of desired nutrients and a potential for energy redistribution towards increased cellular growth or mAb productivity.

## Keywords

Stoichiometric Balances

Multivariate Data-Analytics

Orthogonal Partial Least Squares

Amino Acid Consumption Rates

Cell Culture Media Development

## Introduction

Improved robustness in production of monoclonal antibodies (mAbs) from Chinese Hamster Ovary (CHO) cell culture processes depends on gaining better understanding of biological changes, interactions, and impact to product quality and yield. Advances in systems biology have led to the development of computational tools to compliment emerging technologies that can perform data-rich experimentation (Schaub, Clemens, Kaufmann, & Shulz, 2011; Sharma, Tripathi, & Mukherjee, 2016). As a result, the employment of computational tools such as empirical or mathematical modeling have gained acceptance as a potential path to provide a deeper understanding of cellular mechanisms to inform process decisions that result in more efficient experimental practices during pre-commercial biologics development (Insaidoo, Banerjee, Roush, & Cramer, 2017; Roush et al., 2020; Smiatek, Jung, & Bluhmki, 2020). For instance, computational efforts have shown great potential in optimizing nutrient concentrations within chemically defined (CD) media for CHO cells by modeling routine cell culture and metabolomic data to highlight key limiting nutrients (Cuperlovic-Culf & Culf, 2016). Optimizing base and feed media compositions is a critical step of process development as components in these media directly impact critical quality attributes and cellular productivity of mAbs (Li, Vijayasankaran, Shen, Kiss, & Amanullah, 2010). However, despite long-term benefits of raw material control and cost reduction, development of proprietary CD media is a resource-and time-intensive process which requires a significant number of experiments (Kelley, 2009). By leveraging historical data in computational algorithms, process development researchers are able to identify key components in media that can significantly reduce the complexity of large experimental designs and reduce the development time to optimize media (Torkashvand et al., 2015; Xing et al., 2011).

In this study, we explored the utilization of multivariate data analytics (MVDA) modeling to optimize amino acid concentrations in CD feed media with aims to increase cell growth and mAb productivity. While other components in CD media such as vitamins, trace elements, and co-factors are required for various cellular functions, amino acids are considered one of the most important component group in CD media representing about 74% of CHO cell mass (Carrillo-Cocom et al., 2015; Fan et al., 2015; Pan, Streefland, Dalm, Wijffels, & Martens, 2017; Templeton, Xu, Roush, & Chen, 2017a). Balancing and optimizing amino acid concentrations therefore is critical to ensure CD media can sustain cell growth and mAb productivity. Thus far, majority of laboratory-based experimental and computational approaches for optimizing amino acid levels in CD media have been built on supplementing the amino acids that were being consumed in excess (Dietmair et al., 2012; Duarte et al., 2014; Robitaille, Chen, & Jolicoeur, 2015a). A challenge with such an approach is that over supplementation and consumption of certain amino acids can lead to the production of harmful byproducts. For instance, varying the amount of essential amino acid concentrations can lead to the differential regulation of amino acid metabolic pathways that could ultimately result in increased formation of harmful byproducts such as lactate or ammonia (Alden et al., 2020; Morris et al., 2020; Park, Reimonn, Agarabi, Brorson, & Yoon, 2018; Pereira, Kildegard, & Andersen, 2018). Alternatively, supplementing amino acids in a time-dependent manner based on cell demand would ensure optimal consumption and cellular metabolism presumably resulting in better cell growth and mAb productivity.

In aims to understand the dynamics of amino acid metabolism, Xie and Wang described the relationship

between the theoretical demand and the actual consumption of amino acids as their stoichiometric balances (Xie & Wang, 1994b; Xie & Wang, 1994c). According to their work, the theoretical demand of an amino acid is the amount needed to maintain both biomass and mAb productivity per cell, and with a steady state approximation, the amount consumed is balanced stoichiometrically with the theoretical demand. However, recent work on CHO cell amino acid requirements showed that most cells do not behave in a steady state and that the production process by which nutrient feeds are provided can greatly modulate the dynamics of consumption (Kyriakopoulos, Polizzi, & Kontoravdi, 2013; Luo et al., 2020). Templeton et al described the existence of two possible metabolic states. Within the first state, the actual consumption of an amino acid exceeds the theoretical demand resulting in a catabolic state in which the cell can breakdown the excess nutrients for energy or biomass. In contrast, within the second state, the theoretical demand exceeds the actual consumption of an amino acid resulting in an anabolic state in which the cells must biosynthesize or acquire the required nutrients (Table 1) (Templeton, Xu, Roush, & Chen, 2017b). Time-dependent dynamics of amino acid stoichiometric balances can therefore highlight concentration constraints and metabolic requirements for different amino acids throughout the culture (Ahn & Antoniewicz, 2011; Pan, Dalm, Wijffels, & Martens, 2017a). Although a stoichiometric-based model can provide a systematic and mathematical approach to balancing component composition in CD media, it does not eliminate the resources, and time needed for experimental work (Kumar, Bhalla, & Rathore, 2014). Accordingly, computational MVDA approaches have shown potential in utilizing historical cell culture data and nutrient consumption rates to minimize the resource demand of lab-based experimental designs and execution (Suarez-Zuluaga, Borchert, Driessen, Bakker, & Thomassen, 2019).

MVDA methods rely on a dimensional reduction strategy in which information from a group of variables can be consolidated and summarized into latent variables that in turn, can help describe the overall distribution of experimental observations (i.e. culture time points, cell culture batches) (Mevik & Wehrens, 2007). Each latent variable is based on the individual contribution of variance from all of the variables within the group, and variable ranking statistics can therefore be used to highlight and sub-select a set of key variables towards specific response parameters (Prieto, Eriksson, & Tyrge, 2014). As a result, MVDA approaches can be used to analyze batch-to-batch variability, test the impact of culture performance from process variations, and identify key nutrients from cell culture media (Nitish & Rathore, 2011). In addition, statistical MVDA models in bioprocess rely on the “black box” approach in correlating extracellular metabolites and fluxes to cellular phenotypical behavior without mathematically modeling the intracellular complexities of cellular metabolism (Kanehisa & Goto, 2000; Noll & Henkel, 2020). Accordingly, MVDA models are easier to develop, do not require systematic constraints or curation of biological parameters, and are computationally efficient in determining key correlations (Rouiller et al., 2013).

Several MVDA algorithms that are routinely used within bioprocess development include Principal Component Analysis (PCA), Partial Least Squares (PLS) Regression, and variations of PLS such as Orthogonal Partial Least Squares (OPLS) (Mevik & Wehrens, 2007; Yamamoto et al., 2009). For instance, Kirdar et al utilized PCA to diagnose similarities and differences between small-scale and large-scale manufacturing batches for biopharmaceutical products (Kirdar, Conner, Baclaski, & Rathore, 2007). Dietmair et al utilized PLS to determine a range of metabolites in the culture medium that correlate with the cell growth rate (Dietmair et al., 2012). In most cases, MVDA models have provided support in upstream bioprocess development, process characterization, platform development, and more recently, media development that drive specific responses (Behrouz et al., 2020; MKS Umetrics AB, 2013; Pybus et al., 2014; Worley & Powers, 2013). Therefore, we hypothesized that amino acid stoichiometric balances (SB) used in combination with MVDA models, could not only highlight key amino acids, but also provide information on cell culture performance characteristics. While SB may not be able to identify specific intracellular pathways or enzyme activation due to amino acid changes, it may however serve as an effective surrogate to extrapolate the intracellular “black box” activities that result in the detected culture outcomes. Furthermore, we postulate that supplementing stoichiometrically desired amino acids that are positively correlated to cell growth or mAb productivity would help improve the desired cell culture performance. The amino acid enrichment ‘cocktails’ can then be designed and established based on SB and MVDA for a defined or training set of

culture runs and may be used as feed supplements for future or subsequent culture runs.

In support of our hypothesis, we trained an MVDA model using a 25-batch study in which cells were supplemented with various amounts of a high nutrient feed in aims to identify metabolite stoichiometric balances predictive of cell growth or mAb productivity. Model-derived key stoichiometric balances were translated into amino acid cocktail feeds and supplemented at model-informed time points. In addition to the feed media used in the training set, the model-generated amino acid cocktail feeds were also tested as supplements for two other chemically-defined feeds at lower nutrient concentrations, referred to as moderate and low nutrient feeds. Moderate and low nutrient feeds contained 40% and 17% of the carbon concentration relative to the high nutrient feed level. Experimental validation of our platform and model-directed amino acid supplementation showed about a 20% increase in cumulative viable cells (CVC) and a 30% increase in mAb titer within the supplemented high nutrient condition, about a 30% increase in CVC and 80% increase in titer within the supplemented moderate nutrient condition, and about a 55% increase in CVC and 80% increase in titer within the supplemented low nutrient condition. Results from our study provide evidence towards a novel method to help improve and streamline CD media development for bioprocess.

## Materials and Methods

### *Cell Line and Expansion*

A GS CHO cell line expressing a proprietary monoclonal antibody (mAb) was used for this study. The seed train was initiated from a vial thaw and scaled up in shake flasks to generate enough cells for production bioreactors. Cell expansion was carried out in CD CHO medium (Gibco, NY) and routinely passaged to maintain exponential growth. The shake flasks were maintained in shaking incubators at 36.5°C, 5% CO<sub>2</sub>, 70% humidity, and 100 RPM.

### *Cell Culture and Media*

A total of 25 14-day fed-batch production cell culture conditions were used to build a training dataset for the model. For all 25 batches, CHO cells were cultured in AMBR® 15 (Sartorius Stedim, GE) mammalian bioreactors seeded at 2E6 cells/mL with an initial working volume of 13 mL. Bioreactors were operated at dissolved oxygen of 30%, pH setpoint of 6.95±0.15 controlled by either CO<sub>2</sub> or base addition, agitation at 900 RPM, and incubated at 36.5°C. Control cultures were fed with a high nutrient feed based on a previously defined process. The remaining cultures were fed varying amounts of the same high nutrient feed following the same feeding schedule to obtain a varying distribution of growth and productivity profiles for the training dataset. Glucose was supplemented back to 6 g/L if extracellular glucose levels fell below 3 g/L. Similarly, glutamic acid was supplemented as 3 mM bolus additions if the extracellular glutamic acid levels fell below 4 mM.

To validate the MVDA models for cell growth and mAb productivity, an experiment containing 46-batches was performed. Three baseline controls were designed in which cells were fed with three CD feed media corresponding to high, moderate, and low nutrient levels. In addition to the ranging nutrient feeds (high, moderate, and low), key amino acids identified from either the growth or production model were formulated into amino acid cocktail feeds and supplemented to the cultures at model-informed time points.

Each cocktail feed contained a 100X concentration of the specific amino acids with pH adjusted via 10N NaOH to maintain solubility. Previous studies indicated that AMBR® 15 bioreactors incur a 1 – 1.5% of evaporation rate (v/v per day), which can lead to inaccurate quantification of cell growth and mAb productivity due to the small working volume (Hsu, Aulakh, Traul, & Yuk, 2012). Therefore, 150 µL of sterile water was supplemented daily to compensate for evaporation. To compensate for osmotic and pH differences of amino acid cocktails, control solutions with similar osmolality and pH to the most concentrated amino acid cocktail feed were added at the same volume. In addition, a control was included to confirm the impact of basal media dilution.

Cell culture samples were taken on alternate days up to and including day 14 for offline analysis of VCD, cell viability, cell diameter, pH, pCO<sub>2</sub>, extracellular metabolites (lactate, ammonia, glutamate, glutamine, and

glucose), mAb titer, and amino acids. Viable cell density, viability, and cell diameter were measured using the trypan blue exclusion method on a CEDEX HiRes cell counter (Roche Diagnostics GmbH, Mannheim, Germany). The ABL80 blood gas analyzer (Radiometer, Denmark) was used to capture offline pH and pCO<sub>2</sub> values. The extracellular metabolites were measured using the RX IMOLA Clinical Chemistry Tester (Randox, UK). Antibody titer was quantified using a previously described protein-A reverse-phase high-performance liquid chromatography (HPLC) method (Hoshan et al., 2019). Amino acids were measured using Waters Acquity ultra-performance liquid chromatography (UPLC) H-Class Bio System with an AccQ-Tag Ultra Derivatization Kit (Waters, MA). The amino acid standard kit (A9781) was obtained from Millipore Sigma (St. Louis, MO).

### *Statistical Analysis and Empirical Modeling Methods*

Experimental conditions were performed in biological replicates and resulting data is presented with the standard error of the mean. CVC and mAb titer were normalized to the maximum CVC and titer of the respective process control conditions. Graphical analysis, standard error calculations, ANOVA, and one-way student's t-tests were performed in Microsoft Excel and SAS JMP (SAS, USA), respectively.

SIMCA-P+ (Sartorius Stedium, GE) was used for MVDA modeling and detailed methods for PCA and PLS regression are described by Wold et al (MKS Umetrics AB, 2013). However, in short, multi-variate methods begin with dimensional reduction where high-dimensional datasets can be reduced to a lower dimensional space and be explained by fewer variables (i.e. latent variables). The latent variables are calculated by unit normalization of the data followed by a projection of the normalized data on a lower dimensional space, and lastly finding the directions of the greatest variance within that space or the eigen vector (Mevik & Wehrens, 2007). The total variable contribution towards the greatest variance in any given direction is described by the first latent variable or in the case of PCA, the first principal component. Every subsequent component within a PCA model is orthogonal or perpendicular to the preceding component and is aimed at explaining the total variance of a dataset. Accordingly, each component is a summation of the individual variable contributions or loadings towards the variance. Interestingly, Wold et al found that the sum of every variable contribution across all the components in a model could be represented as a variable importance to projection (VIP) which can provide a heuristic multivariate ranking system (Akarachantachote, Chadcham, & Saithanu, 2013; Prieto et al., 2014). As a result, dimensional reduction models can be used to not only provide a holistic distribution of batches or even individual observations, but also identify key variable contributions that explain the distribution between observations and ultimately, pose as potential targets for optimization efforts.

In contrast, the calculation of latent variables is modified when using supervised approaches such as PLS, as the goal for latency is the direction of the greatest co-variance between the explanatory variables and the response variable(s). Since the first predictive component in a PLS model is not the first principal component, PLS models become very powerful when there is a high degree of collinearity between the explanatory variables and the response variable. However, when the variables are nonlinear in behavior, such as in the case of amino acid stoichiometric balances, the goodness of fit ( $R^2$ ) and the goodness of prediction ( $Q^2$ ) of a PLS model are significantly impacted as a high degree of information is unaccounted for in the explanatory variable space. To circumvent this issue, Orthogonal Partial Least Squares (OPLS) was used in which the first component or the predictive component is forced to be the first principal component and every subsequent orthogonal component aims to explain the remaining co-variance between the explanatory and response variables. Consequently, OPLS models yield better predictions and increased model interpretability towards nonlinear response variables as a greater degree of information in the explanatory variable space is considered (Bylesjo M et al., 2006; Yamamoto et al., 2009). Accordingly, OPLS models built with stoichiometric balances were found to have higher  $Q^2$  values than PLS models (data not presented).

Lastly, to incorporate the time-dependent contributions of the amino acid consumption rates and resulting stoichiometric balances, the training set data matrix was transformed to a batch-level model (BLM) format. In a BLM format, every variable at every day measured across the batches was formatted to become an independent variable. As a result, each row of a BLM data matrix represented a single batch with every

cell culture variable expanded by the number of timepoints it was measured as additional columns, forming a wider and shorter data matrix. The benefits of the BLM format over the untransformed dataset format were two-fold: (1) each variable in the model was able to provide a specific time-dependent contribution that was representative of a particular instance in a batch and (2) increased precision with the OPLS algorithm as each component of the model is a weighted average of all the variables and increased number of variables provided a greater reliability towards the calculation of that weighted average (Vajargah, Sadeghi-Bazargani, Mehdizadeh-Esfanjani, Savadi-Oskouei, & Farhoudi, 2012; Worley & Powers, 2013; Worley & Powers, 2015). For instance, all 20 amino acid stoichiometric balances that was captured from the 25 training batches were measured every other day from day 0 to day 14 and interpolated for the unmeasured days resulting in 15 observations per batch or 375 observations across all 25 batches. But when transformed into a BLM format, the 20 original stoichiometric balance variables were expanded to 300 variables capturing the stoichiometric balance at each time point (day 0 – day 14) and the observations collapsed to 25 rows representing one observation per batch. The complexity and interpretation of the resulting matrix was drastically reduced into latent variables using the dimensional reduction thus highlighting the benefit of MVDA modeling capabilities and its ability to retain biological time-dependent information for a large set of variables to rapidly identify key modulators to improve bioprocess development efforts.

## Results

OPLS models were built to predict cell growth or mAb expression as a function of amino acid stoichiometric balances. Since cell growth was measured as viable cell density (VCD) throughout the culture, CVC was calculated to quantify the total cells that were present within a given time interval. VCD profiles of CHO cell culture typically follow four growth phases, namely, lag phase, log phase, stationary phase, and death phase (Dutton, Scharer, & Moo-Young, 2006). Similarly, in the training dataset, VCD profiles of the 25 batches also reflected the same four phases but varied drastically in peak cell densities (black lines in Fig. 1a). Peak VCD was observed at day 9 for majority of the training batches with the exception of a few that peaked near day 7 representing those batches that received a lower percentage of the high nutrient feed. In contrast, CVC profiles increase over time as they represented a growing sum of total in the culture (green lines in Fig. 1a). Intuitively, modeling the time at which peak VCD was reach or the end of the log phase would highlight the time-dependent contributions by amino acids from the start of the culture which would produce a model with narrower utility. Since >75% of the training runs showed day 9 as peak VCD and the remaining runs also achieve highest total cells on day 9, the CVC at day 9 was chosen as the response variable for the growth model. A similar OPLS model was created to predict titer but in contrast to VCD, mAb titer was measured as the cumulative total concentration of mAb within the culture at any given time. Since peak mAb titer was observed by the end of all cultures, day 14 titer was chosen as the response variable for the production model (Fig. 1b).

To measure the variability and generalization of a model prediction, a relatively large distribution space was required in the training dataset. Accordingly, the amino acid SBs from the 25 training batches in a BLM format were analyzed by PCA. Each observation or score of the PCA model in BLM format, which represented a single batch, was graphically analyzed in a score plot and the 375 amino acid SBs were analyzed in a loadings plot to identify any collinearity or dependencies of the variables within the datasets that could possibly bias the prediction (Fig. 1c and Fig. 1d). The PCA model explained 38.8% variance in the in the first component and 14.7% variance in the 2<sup>nd</sup> component (Fig. 1c). Although only two components are graphed, a 5-component model was built to ensure greater than 70% of variance was captured to represent the majority of the dataset. However, based on the first two components alone, the distribution of the batches showed a random dispersion and lack of any specific clustering. In addition, the variable loadings plot did not highlight any time-dependent grouping, suggesting a minimal collinearity or internal bias within the training dataset in terms of amino acid SBs (Fig. 1d). The generalized distribution of the variables provided a strong potential for the OPLS model to learn and predict across a varying space for future batches. The reliability of the PCA model was further justified by the criteria to remove any outliers that could cause internal biases. Accordingly, a 95% Hotelling's T<sup>2</sup> ellipse was provided as a confidence interval around the dataset to identify any batches that deviated from the majority (Rencher, 1993).

To ensure a strong fit for both models without a significant loss of predictive power, the OPLS model for day 9 CVC was built with 1 predictive component and 9 orthogonal components resulting in a  $R^2$  of 0.912 and  $Q^2$  of 0.726 (Supplementary Fig. S1a). The requirement of additional orthogonal components was further reflected by the diverse spread of CVC profiles throughout the 25 training batches ranging between 25E6 – 45E6 cell-days per mL at day 9. Similarly, there was a large distribution of day 14 titer ranging from 0.2 to 1.05 relative titer values (Fig. 2b). Accordingly, the OPLS model for day 14 Titer had 1 predictive component and 6 orthogonal components with  $R^2$  of 0.832 but a  $Q^2$  of 0.422. Although the predictive power of the production model to generalize to diverse future batches was not as strong as that of the growth model, it was able to highlight information on key variables for media optimization (Supplementary Fig. S1b).

### *Identifying Variable Importance with Stoichiometric Balances*

Since amino acid SBs provided information on the cellular metabolic state in terms of amino acid consumption, model-predicted key amino acid additions at specific time points provided the potential to improve cell growth and mAb productivity. However, to identify key stoichiometric balances from the growth model and the mAb production model, three aspects had to be considered: (1) a variable ranking heuristic that could provide a mathematical method to evaluate the importance of each variable at a given time point; (2) a directionality coefficient that could provide information on each variable’s relationship towards the response variable; and (3) the direction of the stoichiometric balance since positive and negative balances represented two distinct metabolic states. First, since each component in an OPLS model is described by a weighted contribution of each variable in the dataset, the cumulative squared sum of all the weighted contributions of each variable across all the components was calculated to represent the variable importance to projection (VIP). Since the VIP is a squared sum always resulting in a positive magnitude, the VIP was used as a ranking system for all the variables in terms of importance to the predictability of the model. Moreover, Powers et al described that the average VIP value compared across all variables is typically around 1 (Powers et al., 2020). Accordingly, VIP values greater than 1 were selected to be variables that significantly contributed towards the predictability of the response variable whereas those with a VIP less than 0.5 were presumed to have nominal contributions to the overall model. Variables with VIP values between 1 and 0.5 added to the accuracy and reliability of the model however did not significantly contribute to the predictive power of the model. Second, correlative directionality for each variable was averaged across all the components and represented as the magnitude and directionality of each variable’s coefficient. In such a case, positively correlated values represented those stoichiometric balances that would increase co-dependently with the response variable regardless of the actual magnitude of the variable. Lastly, since stoichiometric balances could exist either as a positive value representing greater consumption than the theoretical demand or a negative value representing a lack of consumption compared to the theoretical demand, the directionality of the stoichiometric balance was used to distinguish between nutrient rich and nutrient limited conditions. However, since the model-selected amino acid SBs were the weighted sums across all the batches, the directionality sign of the amino acid SBs was based on the process control from the training dataset. Accordingly, Table 2 shows the directionality signs of all the amino acid SBs from the process control condition and provides the reference values for biomass and mAb composition for each amino acid.

Based on the selection factors for variable importance, three distinct experimental criteria emerged for SBs to validate the growth and production models. The first Criterion consisted of positively correlated SBs with a VIP greater than 1 and a positive stoichiometric balance sign representing amino acids that are being favorably consumed greater than the theoretical demand. The second Criterion consisted of positively correlated stoichiometric balances with a VIP greater than 1 and a negative stoichiometric balance sign representing amino acids that are being favorably consumed less than the theoretical demand. The third Criterion consisted of positively correlated stoichiometric balances with a VIP less than 0.5 and a negative stoichiometric balance sign also representing amino acids that are being favorably consumed less than the theoretical demand but deemed unimportant by the model. For each Criterion however, the sign of the magnitude of the stoichiometric balance was representative of the process control cultures of the training dataset as opposed to the average across all the training batches as the goal of the model was to improve



cell growth and mAb production beyond the current benchmark (Table 1). In all cases however, negatively correlated stoichiometric balances were disregarded since removing nutrient components from an existing chemically defined medium poses a greater operational challenge than supplementing additional nutrients. Therefore, the scope of this study focused only on positively correlated stoichiometric balances.

For both the growth model and production model, scaled positive coefficients from OPLS were plotted for each time-dependent stoichiometric balance grouped by amino acid (Fig. 2). For each plot, the variables displayed included those that fell within the VIP factor for each Criterion and the variables highlighted included those that met the stoichiometric balance sign factor for each Criterion (Growth Model: Orange Bars; Production Model: Green Bars). For each model and Criterion, amino acid cocktail feeds were developed for the corresponding day based on the highlighted amino acid SBs. Interestingly, most variables for both the growth and the production models for Criterion 1 had positive stoichiometric balance signs (Fig. 2a and Fig. 2b). This was representative of the high nutrient feed conditions since amino acids can routinely be saturated within the extracellular environment. However, Criterion 1 was designed to measure the effectiveness of providing an increased concentration of amino acids already being consumed beyond the theoretical demand. For Criterion 2, only a few variables were highlighted as important based on negative stoichiometric balance signs. For instance, only alanine (days 1, 2, 4, 5, 6, and 7), cystine (days 3, 4, 6, 7, and 8), and glycine (days 1, 2, 3, and 7) were identified as important for the growth model (Fig. 2c) whereas the same variables with the addition of lysine (day 9) and methionine (day 9) were identified as important in addition to the former three for the production model (Fig. 2d). Interestingly, the high crossover of amino acids identified between the production model and the cell growth model supports the notion that increased total cells would produce increased antibody. Criterion 3, on the other hand, was primarily designed to measure the heuristic property of the VIP value and thus, served as a negative control. Only those stoichiometric balances were selected that provided minimal contribution to the predictive power of the model. The selected Criterion 3 amino acids were like that of Criterion two but were found important on different days providing additional justification that stoichiometric balances can also help highlight when a specific nutrient demand is needed by the cells (Fig. 3e and Fig. 3f).

### *Experimental Validation with Amino Acid Supplementation*

To validate the OPLS Model for each Criterion, experimental conditions were designed similarly to the high nutrient process control with the addition of amino acid cocktail feeds. CHO cell cultures fed with amino acid cocktails based on Criterion 1 showed a nominal increase in total cell growth at day 9 for both models (Growth Model vs Control: no significant change; Production Model vs Control: no significant change) (Fig. 3a). Day 14 CVC was also analyzed to assess if the early additions of amino acid up to day 9 would have an impact on the overall cell growth of the culture. Similar to day 9, there was only a slight increase of less than 10% in totals cells from the growth model ( $p = 0.1712$ ) suggesting that increased supplementation of amino acids that were already being consuming in excess based on the theoretical demands did not further drive cell growth. In some instances, overfeeding nutrients can have deleterious effects on the cell culture (Salazar, Keusgen, & von Hagen, 2016). For instance, the growth model showed a significant drop in titer near 20% ( $p < 0.05$ ) and a slight drop in titer for the production model ( $p = 0.159$ ) (Fig. 3b). For comparison, day 9 titer was also analyzed to determine if the growth model for Criterion 1 had any early impact on productivity. Unfortunately, no deviation from the control was observed (Growth Model vs Control: no significant change; Production Model vs Control: no significant change).

In contrast, Criterion 2 was designed to test those amino acids that were being consumed below their theoretical demand. However, since Criterion 2 was built on a basis of high nutrient feeds in the training dataset, we postulated that information gained could also support various phases of media development such as moderate nutrient or low nutrient feeds. Therefore, three feed media conditions for the growth and production models were included with biological replicates for the validation run. Similarly to Criterion 1, amino acid cocktail feeds were prepared and supplemented on model-informed days. For statistical comparisons of the model conditions to the control, CVC and relative titer values were only considered from day 8 to day 14 because a majority of the cultures showed similar CVC and mAb titer during the first half of the culture.

Although day 9 CVC did not show a significant change for high nutrient conditions, day 14 CVC showed a modest increase of about 10% for the growth model and about 20% for production model (Fig. 4). A more significant increase was observed within the moderate nutrient and low nutrient conditions when supplemented with the same amino acid cocktail feeds. Increase in CVC was noticed as early as day 7 for both models in the moderate nutrient condition reaching about a 35% increase in total cells by day 14 for the growth model ( $p < 0.001$ ) and about 30% increase for the production model ( $p < 0.05$ ). The low nutrient condition also showed an increase in CVC as early as day 7 but relative cell growth propagated to near 55% increase in total cells for both models by day 14 (Growth Model vs Control:  $p < 0.01$ ; Production Model vs Control:  $p < 0.05$ ). Interestingly, the relative cell growth rates were only slightly impacted by the amino acid supplementations (Fig. S3). The growth rate for high nutrient conditions showed a minor increase at day 3 during the log-phase of growth but also maintained the slight increased up to day 7 ( $p < 0.05$ ) which could have been the driver to the 20% increase in CVC. Similarly, the growth rates for moderate and low nutrient conditions also slightly increased, but unlike the high nutrient condition, the growth rate remained elevated for a longer duration up to day 9. Moderate nutrient conditions had a 5% increase in growth rate at day 3 in the growth model whereas the low nutrient condition had a 5% increase around day 5 reflecting a delayed log-phase but overall greater increase in CVC relative to the control.

The increase in cell growth for both models was further reflected in mAb productivity. Although the high nutrient growth model did not result in any significant increase in titer, the production model yielded about 25% increased productivity ( $p = 0.163$ ) (Fig. 4). The moderate and low nutrient productivity in contrast reflected a significant increase in cell growth by producing about 80% more relative titer by day 9 and 60% more relative titer by day 14 when compared to control cultures ( $p < 0.001$ ). Although low nutrient conditions had a higher boost in cell growth, it also resulted in a similar 60% increase in productivity by day 14 suggesting an upper threshold for mAb production by this cell line with those particular feeds. The large increases in mAb titer for moderate and low nutrient conditions were further reflected by the significant increases in relative specific productivity (qP) (Fig. S2). For example, the moderate nutrient condition showed a 20% increase in qP for the growth model ( $p < 0.05$ ) and showed a near 30% increase for the production model during the stationary phase ( $p < 0.01$ ). In contrast, the low nutrient condition showed about a 40% increase in qP starting from day 7 ( $p < 0.05$ ) suggesting the increased mAb production was not just an artefact of increased cell growth.

To further understand the differences between the resulting cell growth and mAb productivity phenotypes observed, glucose consumption rates were analyzed to assess how the model-driven conditions and supplementation of presumably limiting amino acids could alter cellular metabolism. Posed differently, was differential glucose consumption artefact of the increased percentage of cells within each model condition or a result of altered metabolic activity at the cellular specific level? Total glucose consumption within high nutrient feeds did not show any change in overall consumption whereas under moderate and low nutrient conditions, there was about a 2 g/L decrease in peak glucose consumption (Fig. 5). In contrast, the specific glucose consumption rates highlighted a more significant difference between control and model conditions within moderate and low nutrient feeds. Growth and production model conditions resulted in almost a 1000 pmol/cell\*day decrease in specific glucose consumption in moderate feed and about a 2000 pmol/cell\*day decrease in low nutrient feeds ( $p < 0.01$ ). Accordingly, the marked increase in cell growth and mAb production for moderate and low nutrient conditions when compared to the high nutrient condition was further reflected by the larger magnitude of specific glucose consumption. Taken together, the decreased specific glucose consumption rates in model-driven conditions suggests that cells within these conditions presumably consumed other nutrients such desired amino acids for bioproduction and energy generation that ultimately resulted the cells to be in a higher metabolic state to produce more mAb.

Since the amino acid cocktails were supplemented in addition to process feed media and prepared at 100x concentrations, modulations in feed volume, pH, and osmolality were warranted. Specific control conditions for each model at all three nutrient feed levels for Criterion 2 were included within the experiment to understand if the increase in CVC or productivity was in fact due to the amino acids and no other cell culture variations. The solution/pH control was designed for the high nutrient feed only and was supplemented with

base at the same volume and pH as the amino acid cocktail with time points representative of either of the growth or production models. Similarly, osmolality control cultures were fed with a NaCl solution with the same osmolality and volume as the highest osmolality among the amino acid feed cocktails. High nutrient feed cultures supplemented with the solution/pH control feed showed about a 20% decrease in day 14 CVC and day 14 titer (Fig. S3). Due to the high pH levels, supplementation of the solution/pH control could have increased cell death and a decline in productivity. Comparably, cultures supplemented with the osmolality control feed for high nutrient conditions also showed about a 20% drop in total cell growth and productivity ( $p < 0.35$ ). However, moderate and low nutrient conditions supplemented with the osmolality feed control did not show any significant differences from the control cultures suggesting that the accumulation of additional osmolality by the cultures in high nutrient conditions was an artefact of the already nutrient rich environment and not of the control feed. Therefore, the effects of osmolality and the pH change from the amino acid cocktail feeds did not contribute to the overall increase in cell growth or productivity, but it was in fact a result of the key amino acids supplementation.

The growth and production models for Criterion 3 represented those nutrients that were consumed less than their theoretical demand but were not deemed important by the model VIP statistic. Presumably, these nutrients were considered not to significantly impact cell growth or productivity within all three nutrient conditions. Although day 9 CVC was chosen as the response variable, relative cell growth rates from Criterion 2 showed that the peak cell growth occurred prior to day 7. This is further reflected in the experimental results for Criterion 3 growth model where there was insignificant change to total cells measured in response to day 8 amino acid supplementation for all three nutrient feed levels (Fig. 6a). Similarly, growth model conditions for all three nutrient levels resulted in about a 20% decrease in relative titer when compared to the control (Fig. 6b). Overall, the production model resulted in about a 5% increase in total cells for high and moderate nutrient conditions, and about a 15% decrease in total cells for the low nutrient conditions when compared to the control. Relative titer on the other hand showed a 10% increase for the high nutrient condition ( $p < 0.3$ ) and about a 40% increase for both moderate and low nutrient conditions ( $p < 0.05$ ). Although 40% increase was presumed to be significant, it was not as significant as the 90% increase in the analogous cultures within Criterion 2 ( $p < 0.001$ ). We presume that the large increase in titer from moderate and low nutrient feed conditions in Criterion 3 was due to the low nutrient levels in the culture and that any additional nutrients could provide a positive effect.

### *Driving Consumption of Amino Acids*

Although amino acid supplementation for the high nutrient condition had minimal impact, the larger increases in cell growth and productivity within Criterion 2 demonstrated an approach to identify limiting nutrients in the moderate and low nutrient controls. However, to determine if the limiting nutrients were indeed the model derived amino acids, consumption rates were compared between model conditions and the respective control within each nutrient feed level from Criterion 2 conditions (Fig. 8). The time dependent differences in amino acid consumption rates between the process controls for each nutrient feed level were represented as either a positive (green) difference reflecting an increase in consumption or a negative (red) difference reflecting a decrease in consumption (Fig. 7). For classification purposes, the amino acids supplemented to the model-driven conditions were highlighted and were grouped as either glucogenic or ketogenic. Glucogenic amino acids can provide a carbon source to the cells to biosynthesize glucose for metabolism whereas ketogenic amino acids provide a carbon source to generate ketone bodies which can serve as a secondary energy source in oxidative metabolism (Brosnan, 2003; Noguchi et al., 2010).

Previously, it has been postulated that amino acid consumption is largely dependent on the concentration of the corresponding amino acid (Fan et al., 2015). Indeed, this was true for the high nutrient feed conditions in which not only the model-supplemented amino acids, but almost all amino acids showed an increase in consumption up to and including day 11 for both models. The few with reduced consumption included glutamic acid, which showed a marked decrease in consumption across all time points after day 3. Although glutamic acid is a nonessential amino acid for CHO cells, we posit that this decrease in glutamic acid consumption is driven by a shift to increase consumption of other supplemented nonessential amino acids,

such as glycine and alanine. However, due to the limited increase in cell growth and mAb productivity from these conditions, overall increased consumption did not shift the cells to a higher metabolic state, suggesting a saturation state for cells and a point of maximum contribution by amino acids.

In contrast, moderate and low nutrient conditions showed a more drastic difference in the consumption patterns when compared to the high nutrient conditions. Most noticeably, there was a marked decrease in the consumption of ketogenic amino acids and increased consumption of glucogenic amino acids suggesting a greater need for energy and biosynthesis due to the limited nutrient conditions. Furthermore, for moderate nutrient conditions the growth model-derived amino acids had increased consumption earlier in the culture and production model-derived amino acids such as methionine were consumed more in the latter half of the culture, supporting the distinct difference in metabolic requirements between the cell growth phase and the high mAb-producing stationary phase. In the low nutrient condition, there was a decrease in majority of the amino acids within the growth model but increase in glutamic acid and alanine in the latter half of the culture. In the low nutrient production model, methionine and eventually lysine showed increased consumption resulting in increased mAb productivity. In all cases, model-derived amino acids were generally consumed more when supplemented according to Criterion 2 than process control cultures for each respective nutrient condition. Accordingly, these amino acids were potential limiting nutrients that pushed cells into a higher metabolic state when supplemented.

## Discussion

Herein we present a novel approach to utilize amino acid stoichiometric balances as inputs into MVDA models such as the orthogonal partial least squares algorithm to balance key amino acids in chemically-defined cell culture media supplements to increase CHO cell growth and productivity. However, the utilization of stoichiometric balances in chemometric approaches is not novel. Experimental stoichiometric models date as early as 1994 where Xie and Wang designed a cell culture medium development strategy in which the key nutrients required in biomass and product generation could be described mathematically by stoichiometric coefficients (Xie & Wang, 1994b; Xie & Wang, 1994c). Since increased cell mass is a necessary requirement to produce the desired concentration of key biologic products such as mAbs, the specific cell growth rate ( $\mu_{\text{growth}}$ ) and specific productivity (qP) become two key governing parameters within the stoichiometric equation. The resulting stoichiometric balance would help derive the theoretical demand of each nutrient, and thus, Xie and Wang showed that various nutrients could be balanced within complex CD media without the necessity of multi-factorial DoEs.

For instance, in a follow up study, Xie and Wang used their stoichiometric model to decrease ammonia and lactate formation in hybridoma cells by balancing the basal medium nutrient levels. In this model, they analyzed the relationship between glutamine and nonessential amino acids as a function of lactate and ammonia production rates. The modified medium resulted in a 4-fold decrease in ammonia production and almost a 10-fold decrease in lactate formation from the hybridoma cells with the added benefit of increased cell growth and product formation (Xie & Wang, 1994a). Val et al utilized the stoichiometric model to balance nucleotide sugars as a function of cell growth and specific productivity to understand nucleotide sugar demand towards glycosylation profiles in CHO cells (Del Val, Polizzi, & Kontoravdi, 2016). Xie and Nyberg et al further utilized the stoichiometric modeling approach in CHO cells by designing a serum-free feed medium with balanced nutrients to improve cell growth and IFN $\gamma$  productivity as well as glycosylation efficiency for IFN $\gamma$  (Xie et al., 1997).

Similarly, we derived stoichiometric balances to improve cell growth and productivity as a function of all 20 amino acids. In agreement with the original method, the stoichiometric coefficients or theoretical demands for amino acids derived in our model were based on their contributions towards biomass and mAb productivity as a function of specific growth rate ( $\mu_{\text{growth}}$ ) and specific productivity (qP). Accordingly, Xie and Wang's original method also considered the cell death rate, thus resulting in gross growth rate (Xie & Wang, 1994c). Cell death rate becomes particularly important during the late-stationary and death phases of the CHO cell culture where viable cell densities and viabilities start to decline (Balandras et al., 2011). The increased death rate during the latter half of the culture is attributed to several factors including increased concentration

of harmful byproducts such as  $\text{NH}_3$ , increased packed cell volume, and rapid nutrient elimination (Pan, Dalm, Wijffels, & Martens, 2017b; Pascoe, Arnott, Papoutsakis, Miller, & Andersen, 2007). Death rates as a result, are either calculated quantitatively by measuring differential levels of extracellular LDH accumulation that is released from lysed or apoptotic cells or theoretically by adjusting the net growth rate by a factor correlating to viability decline (Martínez, Bulté, Contreras, Palomares, & Ramírez, 2020; Martins Conde Pdo, Sauter, & Pfau, 2016; Templeton et al., 2017). Although the net growth rate was used in our calculation for theoretical demand, the growth model only predicted important variables up to day 9 in the culture at which point the viabilities of the cultures were relatively similar suggesting minute differences between gross and net growth rates. The resulting stoichiometric balance solution was obtainable by taking the difference between the empirical consumption and the theoretical demand of every amino acid. In all cases however, the underlying assumption behind stoichiometric balances was that the intracellular metabolism and complex gene expression machinery could be generalized into a “black box” in which extracellular fluxes of nutrients from the media could be directly related to biomass and product generation.

Statistical models in bioprocess have also relied on the “black box” approach in correlating extracellular metabolites and fluxes to cellular phenotypical behavior (Kroll, Hofer, Ulonska, Kager, & Herwig, 2017; Lee & Gilmore, 2006). In addition, multivariate approaches also offer the added benefit of being able to highlight key variables that contribute towards the prediction of specific response variables (Akarachantachote et al., 2013; Gangadharan et al., 2019; Hassan, Farhan, Mangayil, Huttunen, & Aho, 2013). Applied together with the stoichiometric model, the complexity of balancing all 20 amino acids in CD media can be greatly reduced by balancing only the specific amino acids deemed important by the statistical model. Furthermore, statistical models are easier to develop, do not require systematic constraints or curation of biological parameters, and are computationally efficient in determining key correlations (Kim, Rocha, & Maia, 2018; Martins Conde Pdo et al., 2016). More recently however, there has been keen interest on obtaining a deeper understanding of the effects on the metabolic framework by stoichiometric balancing via mechanistic modeling (Sha, Huang, Wang, & Yoon, 2018). Such models can elucidate the intricacies of the intracellular pathways that are extrapolated within statistical models. Thus, mechanistic models can help identify causal linkages within correlation structures. In a modeling review, Traustason et al described several computational methods aimed at optimizing amino acid concentrations for CD media in CHO cell culture. Among them include stoichiometric models such as metabolic flux analysis (MFA) or flux balance analysis (FBA) as well as kinetic models (Traustason, B., Cheeks, & Dikicioglu, 2019). Robitaille et al developed a dynamic model combining the benefits of MFA and kinetic models to explain the relationship between extracellular fluxes of amino acids, extracellular concentrations, and the intracellular flux responses towards that consumption. With a hybrid steady-state and kinetic model, a dynamic relationship between nutrient uptake and cellular response in terms of biomass generation and mAb production was determined (Robitaille, Chen, & Jolicoeur, 2015b).

In another example, combination of FBA and genome-scale modeling with the utilization of amino acid stoichiometric balancing revealed that the addition of nonessential amino acids (NEAAs) have a positive impact on CHO cell biomass production (Traustason, Bergthor, 2019). CHO cells are reported to contain seven NEAAs including glycine, alanine, asparagine, aspartic acid, glutamic acid, proline, and serine (Fomina-Yadlin et al., 2014). These seven amino acids are nutrients that CHO cells can directly biosynthesize whereas the remaining 13 essential amino acids must be obtained by nutrient supplementation from the CD media (Salazar et al., 2016). Presumably, the supplementation of NEAAs can redirect the energy needed from biosynthesis of amino acids to other biological processes such as cell growth or protein synthesis. Moreover, several essential amino acids are also used in biosynthesis for NEAAs and other metabolic intermediates, suggesting that supplementation of NEAAs could not only free up energy but also free up nutrients to support the biosynthesis of other cellular processes (Duarte et al., 2014; Harcum & Lee, 2016). Accordingly, the significant increases of cell growth and mAb productivity that were observed within moderate and low nutrient conditions for both the growth and production models in Criterion 2 supported the energy conservation theory of NEAAs since two of the three amino acids supplemented for cell growth were NEAAs. Supplementation of alanine and glycine also showed increased consumption in all model conditions which

also resulted in a decreased specific consumption of glucose further suggesting that the conserved energy was shifted towards cell growth and increased cellular metabolism. In the case of mAb productivity, methionine and lysine was also supplemented. Although not counted within NEAAs, methionine plays an important role in protein production as it is the universal initiator of protein synthesis and thus, crucial to cellular uptake and metabolism to increase protein production. In addition, methionine has shown to improve the cellular redox state of cells by using the ubiquitous methionine sulfoxide cycle to provide an antioxidant defense system (Lim, Kim, & Levine, 2019). Accordingly, we presume that the supplementation of NEAAs, critical protein synthesis amino acids, and increased cell growth all contributed towards the significant increase in mAb productivity within moderate and low nutrient model conditions.

In the case of high nutrients feeds, although there was a marked increase in cellular consumption of almost all amino acids, there was only a minute increase in cell growth and mAb productivity, and we presumed that a saturation point in the culture was reached. However, the notion of a saturation point with respect to extracellular amino acid concentrations was supported by Salazar et al who described that supplementing concentrations of amino acids that are generally highly consumed in mammalian cell cultivation does not always lead to improvement in cell culture, but rather can lead to undesired effects and the accumulation of harmful waste products (Salazar et al., 2016). Similarly, Templeton et al experimentally validated previous MFA models with  $^{13}\text{C}$  labeling studies *in vitro* and showed that protein expression does not always correspond the activation of specific metabolic pathways, but rather, even fluxes tend to eventually saturate (Templeton et al., 2017b). Taken together, high nutrient conditions resulted in greater consumption but maintained similar specific glucose consumption rates suggesting that cellular energy was not repurposed to increased cell growth or mAb productivity, rather, the cells had reached their maximum capacity to metabolize amino acids.

## Conclusion

The methods presented in this study were consolidated as a single platform approach to rapidly optimize amino acid concentrations as CD media supplements (Fig. 8). Since amino acids are one of the most important building blocks for cellular biomass and protein production, they also comprise an essential component group in CD media. In addition, since the relative amino acid requirement is similar across CHO cell lines and similar between various mAbs, the amino acid modeling platform presented here can be applied to various CHO cell cultures expressing various mAbs (Szélliová et al., 2020). Nevertheless, Xie and Wang's original stoichiometric model was inclusive of all media nutrients and thus, our empirical modeling approach and methods to calculate stoichiometric balances could be generalized to other nutrients as well, for example, vitamins, co-factors, trace elements, organic, and inorganic salts among others. Although raw measurements of nutrient concentrations can be directly input into MVDA approaches (Fig. 8 (1a)), calculation of secondary variables such as cell growth rate, specific productivity, and consumption rates can help formulate stoichiometric balances which in turn can serve as enhanced inputs to derive causal linkages (Fig. 8 (1b-1c)). Within the modeling platform, MVDA dimensional reduction algorithms such as PLS or OPLS can be utilized to understand the correlation between the dynamics of time-dependent stoichiometric balances and key response variables such as cell growth or mAb productivity. VIP statistics and correlation coefficients from multivariate algorithms can help generate importance and directionality criteria for the input variables. Key nutrients towards specific response variables or cellular phenotypical states can be further translated into experimental designs for validation studies (Fig. 8 (2a)). However, in the case of stoichiometric balances, a feature selection decision tree approach can be implemented to choose the appropriate stoichiometric balances for process developing. Bolded here is the positive correlated balances that are negative in value which potentially represent limiting nutrients as shown in Criterion 2 (Fig. 8 (2b-2c)). However, negatively correlated stoichiometric balances could also be considered for media formulation development. To support multivariate statistical models, experimental validation is necessary to not only verify the performance of the model, but also measure the success criteria of the desired phenotypical state. Model outputs can dictate supplementation with modified nutrient feeds to be validated in a targeted DoE, which in turn can generate new cell culture data to refine against the MVDA model (Fig. 8 (3 – 4)).

Since the prediction of the statistical model is largely based on the variance of the training dataset, high variance datasets can provide a larger prediction space for the model and a reduction in model bias. However, models training on datasets with high variance rely heavily on the upper and lower bounds of the variance space and thus, any validation datasets that fall outside of the training dataset space can lead to poor success. Therefore, as new processes are adapted based on model-driven conditions, future datasets can be combined with training datasets to iteratively improve the model (Fig. 8 (5)) (Luo et al., 2020). For instance, important amino acid stoichiometric balances based on high nutrient feeds were also fed to the moderate and low nutrient conditions resulting in significant increases in cell growth and mAb productivity. However, in future iterations the newly generated validation datasets could help improve the original training dataset and help increment towards an improved biologics production process and optimized CD media.

## References

- Ahn, W. S., & Antoniewicz, M. R. (2011). Metabolic flux analysis of CHO cells at growth and non-growth phases using isotopic tracers and mass spectrometry. *Metabolic Engineering*, 13 (5), 598-609. doi:<https://doi.org/10.1016/j.ymben.2011.07.002>
- Akarachantachote, N., Chadcham, S., & Saithanu, K. (2013). Cutoff threshold of variable importance in projection for variable selection. *International Journal of Pure and Applied Mathematics*, 94 (3)
- Alden, N., Raju, R., McElearney, K., Lambropoulos, J., Kshirsagar, R., Gilbert, A., & Lee, K. (2020). Using metabolomics to identify cell line-independent indicators of growth inhibition for chinese hamster ovary cell-based bioprocesses. *Metabolites*, 10 (5), 199. Retrieved from <https://pubmed.ncbi.nlm.nih.gov/32429145>
- Balandras, F., Olmos, E., Hecklau, C., Blanchard, F., Guedon, E., & Marc, A. (2011). Growth and death kinetics of CHO cells cultivated in continuous bioreactor at various agitation rates. *BMC Proceedings*, 5 Suppl 8 (Suppl 8), P101-6561-5-S8-P101. eCollection 2011. doi:10.1186/1753-6561-5-S8-P101 [doi]
- Behrouz, H., Molavi, B., Tavakoli, A., Askari, M., Maleknia, S., Mahboudi, F., & Khodadadian, M. (2020). Multivariate optimization of the refolding process of an incorrectly folded fc-fusion protein in a cell culture broth. *Current Pharmaceutical Biotechnology*, 21 (3), 226-235. doi:10.2174/1389201020666191002144424 [doi]
- Brosnan, J. T. (2003). Interorgan amino acid transport and its regulation. *The Journal of Nutrition*, 133 (6 Suppl 1), 2068S-2072S. doi:10.1093/jn/133.6.2068S [doi]
- Bylesjo M, Rantalainen, M., Cloarec, O., Nicholson, J. K., Holmes, E., & Trygg, J. (2006). OPLS discriminant analysis: Combining the strenghts of PLS-DA and SIMCA classification. *Journal of Chemometrics*, (20), 341-351.
- Carrillo-Cocom, L. M., Genel-Rey, T., Araiz-Hernandez, D., Lopez-Pacheco, F., Lopez-Meza, J., Rocha-Pizana, M. R., . . . Alvarez, M. M. (2015). Amino acid consumption in naive and recombinant CHO cell cultures: Producers of a monoclonal antibody. *Cytotechnology*, 67 (5), 809-820. doi:10.1007/s10616-014-9720-5 [doi]
- Cuperlovic-Culf, M., & Culf, A. S. (2016). Applied metabolomics in drug discovery. *Expert Opinion on Drug Discovery*, 11 (8), 759-770. doi:10.1080/17460441.2016.1195365
- Del Val, I. J., Polizzi, K. M., & Kontoravdi, C. (2016). A theoretical estimate for nucleotide sugar demand towards chinese hamster ovary cellular glycosylation. *Scientific Reports*, 6 , 28547. doi:10.1038/srep28547 [doi]
- Dietmair, S., Hodson, M. P., Quek, L. E., Timmins, N. E., Chrysanthopoulos, P., Jacob, S. S., . . . Nielsen, L. K. (2012). Metabolite profiling of CHO cells with different growth characteristics. *Biotechnology and Bioengineering*, 109 (6), 1404-1414. doi:10.1002/bit.24496 [doi]
- Duarte, T. M., Carinhas, N., Barreiro, L. C., Carrondo, M. J., Alves, P. M., & Teixeira, A. P. (2014). Metabolic responses of CHO cells to limitation of key amino acids. *Biotechnology and Bioengineering*, 111

(10), 2095-2106. doi:10.1002/bit.25266 [doi]

Dutton, R. L., Scharer, J., & Moo-Young, M. (2006). Cell cycle phase dependent productivity of a recombinant chinese hamster ovary cell line. *Cytotechnology*, *52* (1), 55-69. doi:10.1007/s10616-006-9041-4 [doi]

Fan, Y., Jimenez Del Val, I., Muller, C., Wagtberg Sen, J., Rasmussen, S. K., Kontoravdi, C., . . . Andersen, M. R. (2015). Amino acid and glucose metabolism in fed-batch CHO cell culture affects antibody production and glycosylation. *Biotechnology and Bioengineering*, *112* (3), 521-535. doi:10.1002/bit.25450 [doi]

Fomina-Yadlin, D., Gosink, J. J., McCoy, R., Follstad, B., Morris, A., Russell, C. B., & McGrew, J. T. (2014). Cellular responses to individual amino-acid depletion in antibody-expressing and parental CHO cell lines. *Biotechnology and Bioengineering*, *111* (5), 965-979. doi:10.1002/bit.25155 [doi]

Gangadharan, N., Turner, R., Field, R., Oliver, S. G., Slater, N., & Dikicioglu, D. (2019). Metaheuristic approaches in biopharmaceutical process development data analysis. *Bioprocess and Biosystems Engineering*, *42* (9), 1399-1408. doi:10.1007/s00449-019-02147-0 [doi]

Harcum, S. W., & Lee, K. H. (2016). CHO cells can make more protein. *Cell Systems*, *3* (5), 412. doi:https://doi.org/10.1016/j.cels.2016.11.007

Hassan, S., Farhan, M., Mangayil, R., Huttunen, H., & Aho, T. (2013). Bioprocess data mining using regularized regression and random forests. *BMC Systems Biology*, *7 Suppl 1* (Suppl 1), S5-0509-7-S1-S5. Epub 2013 Aug 12. doi:10.1186/1752-0509-7-S1-S5 [doi]

Hoshan, L., Jiang, R., Moroney, J., Bui, A., Zhang, X., Hang, T. C., & Xu, S. (2019). Effective bioreactor pH control using only sparging gases. *Biotechnology Progress*, *35* (1), e2743. doi:10.1002/btpr.2743 [doi]

Hsu, W. T., Aulakh, R. P., Traul, D. L., & Yuk, I. H. (2012). Advanced microscale bioreactor system: A representative scale-down model for bench-top bioreactors. *Cytotechnology*, *64* (6), 667-678. doi:10.1007/s10616-012-9446-1 [doi]

Insaiddoo, F., Banerjee, S., Roush, D., & Cramer, S. (2017). *Computational modeling in bioprocess development* doi:https://doi.org/10.1002/9781119031116.ch6

Kanehisa, M., & Goto, S. (2000). KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, *28* (1), 27-30. doi:gkd027 [pii]

Kelley, B. (2009). Industrialization of mAb production technology: The bioprocessing industry at a crossroads. *mAbs*, *1* (5), 443-452. doi:9448 [pii]

Kim, O. D., Rocha, M., & Maia, P. (2018). A review of dynamic modeling approaches and their application in computational strain optimization for metabolic engineering. *Frontiers in Microbiology*, *9*, 1690. doi:10.3389/fmicb.2018.01690 [doi]

Kirdar, A. O., Conner, J. S., Baclaski, J., & Rathore, A. S. (2007). Application of multivariate analysis toward biotech processes: Case study of a cell-culture unit operation. *Biotechnology Progress*, *23* (1), 61-67. doi:10.1021/bp060377u [doi]

Kroll, P., Hofer, A., Ulonska, S., Kager, J., & Herwig, C. (2017). Model-based methods in the biopharmaceutical process lifecycle. *Pharmaceutical Research*, *34* (12), 2596-2613. doi:10.1007/s11095-017-2308-y [doi]

Kumar, V., Bhalla, A., & Rathore, A. S. (2014). Design of experiments applications in bioprocessing: Concepts and approach. *Biotechnology Progress*, *30* (1), 86-99. doi:10.1002/btpr.1821 [doi]

Kyriakopoulos, S., Polizzi, K. M., & Kontoravdi, C. (2013). Dynamic profiling of amino acid transport and metabolism in chinese hamster ovary cell culture. *BMC Proceedings*, *7* (Suppl 6), P97-6561-7-S6-P97. eCollection 2013. doi:1753-6561-7-S6-P97 [pii]



- Lee, K. M., & Gilmore, D. F. (2006). Statistical experimental design for bioprocess modeling and optimization analysis: Repeated-measures method for dynamic biotechnology process. *Applied Biochemistry and Biotechnology*, 135 (2), 101-116. doi:ABAB:135:2:101 [pii]
- Li, F., Vijayasankaran, N., Shen, A. Y., Kiss, R., & Amanullah, A. (2010). Cell culture processes for monoclonal antibody production. *mAbs*, 2 (5), 466-479. doi:12720 [pii]
- Lim, J. M., Kim, G., & Levine, R. L. (2019). Methionine in proteins: It's not just for protein initiation anymore. *Neurochemical Research*, 44 (1), 247-257. doi:10.1007/s11064-017-2460-0 [doi]
- Luo, Y., Lovelett, R. J., Price, J. V., Radhakrishnan, D., Barnthouse, K., Hu, P., . . . Ogunnaike, B. A. (2020). Modeling the effect of amino acids and copper on monoclonal antibody productivity and glycosylation: A modular approach. *Biotechnology Journal*, , e2000261. doi:10.1002/biot.202000261 [doi]
- 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. doi:10.3389/fbioe.2020.00279
- Martins Conde Pdo, R., Sauter, T., & Pfau, T. (2016). Constraint based modeling going multicellular. *Frontiers in Molecular Biosciences*, 3 , 3. doi:10.3389/fmolb.2016.00003 [doi]
- Mevik, B. H., & Wehrens, R. (2007). The pls package: Principal component and partial least squares regression in R. *Journal of Statistical Software*, 18 (2)
- MKS Umetrics AB (Ed.). (2013). *Multi- and megavariable data analysis* (3rd ed.). Sweden:
- Morris, C., Polanco, A., Yongky, A., Xu, J., Huang, Z., Zhao, J., . . . Yoon, S. (2020). Bigdata analytics identifies metabolic inhibitors and promoters for productivity improvement and optimization of monoclonal antibody (mAb) production process. *Bioresources and Bioprocessing*, 7 (1), 31. Retrieved from <https://doi.org/10.1186/s40643-020-00318-6>
- Nitish, B., & Rathore, A. (2011). Use of multivariate data analysis (MVDA) for generating process understanding from manufacturing data of biotech processes. *Proceedings of the Indian National Science Academy*, 77 , 133-142.
- Noguchi, Y., Nishikata, N., Shikata, N., Kimura, Y., Aleman, J. O., Young, J. D., . . . Stephanopoulos, G. (2010). Ketogenic essential amino acids modulate lipid synthetic pathways and prevent hepatic steatosis in mice. *PloS One*, 5 (8), e12057. doi:e12057 [pii]
- Noll, P., & Henkel, M. (2020). History and evolution of modeling in biotechnology: Modeling & simulation, application and hardware performance. *Computational and Structural Biotechnology Journal*, 18 , 3309-3323. doi:10.1016/j.csbj.2020.10.018 [doi]
- Pan, X., Dalm, C., Wijffels, R. H., & Martens, D. E. (2017a). Metabolic characterization of a CHO cell size increase phase in fed-batch cultures. *Applied Microbiology and Biotechnology*, 101 (22), 8101-8113. doi:10.1007/s00253-017-8531-y [doi]
- Pan, X., Dalm, C., Wijffels, R. H., & Martens, D. E. (2017b). Metabolic characterization of a CHO cell size increase phase in fed-batch cultures. *Applied Microbiology and Biotechnology*, 101 (22), 8101-8113. doi:10.1007/s00253-017-8531-y [doi]
- 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. doi:10.1007/s10616-016-0036-5 [doi]
- Park, S. Y., Reimonn, T. M., Agarabi, C. D., Brorson, K. A., & Yoon, S. (2018). Metabolic responses and pathway changes of mammalian cells under different culture conditions with media supplementations. *Biotechnology Progress*, 34 (3), 793-805. doi:10.1002/btpr.2623 [doi]

- Pascoe, D. E., Arnott, D., Papoutsakis, E. T., Miller, W. M., & Andersen, D. C. (2007). Proteome analysis of antibody-producing CHO cell lines with different metabolic profiles. *Biotechnology and Bioengineering*, 98 (2), 391-410. doi:10.1002/bit.21460 [doi]
- Pereira, S., Kildegaard, H. F., & Andersen, M. R. (2018). Impact of CHO metabolism on cell growth and protein production: An overview of toxic and inhibiting metabolites and nutrients. *Biotechnology Journal*, 13 (3), e1700499. doi:10.1002/biot.201700499 [doi]
- Powers, D. N., Trunfio, N., Velugula-Yellela, S. R., Angart, P., Faustino, A., & Agarabi, C. (2020). Multi-variate data analysis of growth medium trends affecting antibody glycosylation. *Biotechnology Progress*, 36 (1), e2903. doi:10.1002/btpr.2903 [doi]
- Prieto, B. G., Eriksson, L., & Tyrgg, J. (2014). Variable influence on projection (VIP) for orthogonal projections to latent structures (OPLS). *Journal of Chemometrics*, 28 (8), 2627.
- Pybus, L. P., Dean, G., West, N. R., Smith, A., Daramola, O., Field, R., . . . James, D. C. (2014). Model-directed engineering of "difficult-to-express" monoclonal antibody production by chinese hamster ovary cells. *Biotechnology and Bioengineering*, 111 (2), 372-385. doi:10.1002/bit.25116 [doi]
- Rencher, A. C. (1993). The contribution of individual variables to hotelling's T<sup>2</sup>, wilks' lambda, and R<sup>2</sup>. *Biometrics*, 49 (2), 479-489.
- Robitaille, J., Chen, J., & Jolicoeur, M. (2015a). 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. doi:10.1371/journal.pone.0136815 [doi]
- Robitaille, J., Chen, J., & Jolicoeur, M. (2015b). 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. doi:10.1371/journal.pone.0136815 [doi]
- Rouiller, Y., Perilleux, A., Collet, N., Jordan, M., Stettler, M., & Broly, H. (2013). A high-throughput media design approach for high performance mammalian fed-batch cultures. *mAbs*, 5 (3), 501-511. doi:10.4161/mabs.23942 [doi]
- Roush, D., Asthagiri, D., Babi, D. K., Benner, S., Bilodeau, C., Carta, G., . . . Zavalov, O. (2020). Toward in silico CMC: An industrial collaborative approach to model-based process development. *Biotechnology and Bioengineering*, doi:10.1002/bit.27520 [doi]
- Salazar, A., Keusgen, M., & von Hagen, J. (2016). Amino acids in the cultivation of mammalian cells. *Amino Acids*, 48 (5), 1161-1171. doi:10.1007/s00726-016-2181-8 [doi]
- Schaub, J., Clemens, C., Kaufmann, H., & Shulz, T. W. (2011). Advancing biopharmaceutical process development by system-level data analysis and integration of omics data. *Advanced in biochemical engineering biotechnology* (pp. 133-133-163)
- Sha, S., Huang, Z., Wang, Z., & Yoon, S. (2018). Mechanistic modeling and applications for CHO cell culture development and production. *Current Opinion in Chemical Engineering*, 22 , 54-61. doi:https://doi.org/10.1016/j.coche.2018.08.010
- Sharma, V., Tripathi, M., & Mukherjee, K. (2016). Application of system biology tools for the design of improved chinese hamster ovary cell expression platforms. *Journal of Bioprocess Biotech*, 6 (284)
- Smiatek, J., Jung, A., & Bluhmki, E. (2020). Towards a digital bioprocess replica: Computational approaches in biopharmaceutical development and manufacturing. *Trends in Biotechnology*, 38 (10), 1141-1153. doi:S0167-7799(20)30142-6 [pii]
- Suarez-Zuluaga, D. A., Borchert, D., Driessen, N. N., Bakker, W. A. M., & Thomassen, Y. E. (2019). Accelerating bioprocess development by analysis of all available data: A USP case study. *Vaccine*, 37 (47), 7081. doi:https://doi.org/10.1016/j.vaccine.2019.07.026"

- Széliová, D., Ruckerbauer, D. E., Galleguillos, S. N., Petersen, L. B., Natter, K., Hanscho, M., . . . Zanghellini, J. (2020). What CHO is made of: Variations in the biomass composition of chinese hamster ovary cell lines. *Metabolic Engineering*, 61 , 288-300. doi:S1096-7176(20)30101-4 [pii]
- Templeton, N., Smith, K. D., McAtee-Pereira, A. G., Dorai, H., Betenbaugh, M. J., Lang, S. E., & Young, J. D. (2017). Application of (13)C flux analysis to identify high-productivity CHO metabolic phenotypes. *Metabolic Engineering*, 43 (Pt B), 218-225. doi:S1096-7176(17)30033-2 [pii]
- Templeton, N., Xu, S., Roush, D. J., & Chen, H. (2017a). (13)C metabolic flux analysis identifies limitations to increasing specific productivity in fed-batch and perfusion. *Metabolic Engineering*, 44 , 126-133. doi:S1096-7176(17)30187-8 [pii]
- Templeton, N., Xu, S., Roush, D. J., & Chen, H. (2017b). (13)C metabolic flux analysis identifies limitations to increasing specific productivity in fed-batch and perfusion. *Metabolic Engineering*, 44 , 126-133. doi:S1096-7176(17)30187-8 [pii]
- Torkashvand, F., Vaziri, B., Maleknia, S., Heydari, A., Vossoughi, M., Davami, F., & Mahboudi, F. (2015). Designed amino acid feed in improvement of production and quality targets of a therapeutic monoclonal antibody. *PloS One*, 10 (10), e0140597. doi:10.1371/journal.pone.0140597 [doi]
- Traustason, B., Cheeks, M., & Dikicioglu, D. (2019). Computer-aided strategies for determining the amino acid composition of medium for chinese hamster ovary cell-based biomanufacturing platforms. *International Journal of Molecular Sciences*, 20 (21), 5464. doi: 10.3390/ijms20215464. doi:10.3390/ijms20215464 [doi]
- Traustason, B. (2019). Amino acid requirements of the chinese hamster ovary cell metabolism during recombinant protein production. *bioRxiv*, doi:10.1101/796490
- Vajargah, K. F., Sadeghi-Bazargani, H., Mehdizadeh-Esfanjani, R., Savadi-Oskouei, D., & Farhoudi, M. (2012). OPLS statistical model versus linear regression to assess sonographic predictors of stroke prognosis. *Neuropsychiatric Disease and Treatment*, 8 , 387-392. doi:10.2147/NDT.S33991 [doi]
- Worley, B., & Powers, R. (2013). Multivariate analysis in metabolomics. *Current Metabolomics*, 1 (1), 92-107. doi:10.2174/2213235X11301010092 [doi]
- Worley, B., & Powers, R. (2015). A sequential algorithm for multiblock orthogonal projections to latent structures. *Chemometrics and Intelligent Laboratory Systems : An International Journal Sponsored by the Chemometrics Society*, 149 (Pt B), 33-39. doi:10.1016/j.chemolab.2015.10.018 [doi]
- Xie, L., Nyberg, G., Gu, X., Li, H., Möllborn, F., & Wang, D. I. (1997). Gamma-interferon production and quality in stoichiometric fed-batch cultures of chinese hamster ovary (CHO) cells under serum-free conditions. *Biotechnology and Bioengineering*, 56 (5), 577-582. doi:10.1002/(SICI)1097-0290(19971205)56:53.0.CO;2-9 [doi]
- Xie, L., & Wang, D. I. (1994a). Applications of improved stoichiometric model in medium design and fed-batch cultivation of animal cells in bioreactor. *Cytotechnology*, 15 (1-3), 17-29. doi:10.1007/BF00762376 [doi]
- Xie, L., & Wang, D. I. (1994b). Fed-batch cultivation of animal cells using different medium design concepts and feeding strategies. *Biotechnology and Bioengineering*, 43 (11), 1175-1189. doi:10.1002/bit.260431123 [doi]
- Xie, L., & Wang, D. I. (1994c). Stoichiometric analysis of animal cell growth and its application in medium design. *Biotechnology and Bioengineering*, 43 (11), 1164-1174. doi:10.1002/bit.260431122 [doi]
- Xing, Z., Kenty, B., Koyrakh, I., Borys, M., Pan, S., & Li, Z. (2011). Optimizing amino acid composition of CHO cell culture media for a fusion protein production. *Process Biochemistry*, (46), 1423-1423-1429.
- Yamamoto, H., Yamaji H, Abe, Y., Harada, K., Waluyo, D., Fukusaki, E., . . . Fukuda H. (2009). Dimensionality reduction for metabolome data using PCA, PLS, OPLS, and RFDA with differential penalties to latent variables. *Chemometrics and Intelligent Laboratory Systems*, (98), 136-142.

## Figure Legends

**Fig. 1. Variance and distribution of training dataset:** A 25-batch CHO cell culture experiment was designed to obtain a training dataset to test the amino acid stoichiometric balance MVDA model in which the culture was supplemented with various amounts of high nutrient feeds within a defined fed-batch process. Data on cell growth and mAb titer from all batches was analyzed to obtain peak growth and productivity time points and a PCA model was generated to analyze the distribution and variance of training batches with respect to the amino acid stoichiometric balances. (a) VCD peaked between day 6 to day 9 across the 25 training batches with a majority peaking near day 9 (Black Lines). Similarly, CVC at each time point was also plotted along side VCD to show a similar distribution (Green dashed lines). Cell growth data up to day 9 represented the log-growth phase of the cell culture, thus, day 9 CVC was chosen as the response variable for the growth model. (b) Titer data was normalized to the peak titer value of the fed-batch process control. Since mAb titer represented an accumulation of extracellular secreted mAb in the spent media, peak titer occurred at the end of the process on day 14. Accordingly, day 14 titer was chosen as the response variable for the production model. (c). Distribution of the batches based on principle components 1 and 2, which cumulatively described 53.5% of the variance did not show any significant grouping or clustering of the batches thereby eliminating the possibility of any inherent bias in the data. (d) Similarly, the loadings plot of the time-dependent amino acid stoichiometric balances showed a wide distribution without any collinearity of variables at specific time points.

**Fig. 2. Criteria for Model-based Amino Acid Selection:** VIP values and coefficients from the OPLS models along with stoichiometric balance directionalities from the process control were utilized as selection factors in 3 unique criteria for amino acids. (a-b) Criteria 1 amino acid SBs represented positive correlation to either Day 9 CVC or Day 14 Titer, a VIP value greater than 1, and a positive stoichiometric balance sign. (c-d) Criteria 2 amino acid SBs represented positive correlation to either Day 9 CVC or Day 14 Titer, a VIP value greater than 1, and a negative stoichiometric balance sign. (e-f) Criteria 3 amino acid SBs represented positive correlation to Day 9 CVC, a VIP value less than 0.5, and a negative stoichiometric balance sign. Coefficient values were normalized and scaled to unit variance by the model, and only positively correlated amino acid SBs were selected. Since time-dependent amino acid SBs were used in the model, various selected time points that met each respective criterion were plotted. For each model and criteria, amino acid cocktail feeds were developed grouped by day of importance. For instance, in the growth model for criteria 1 (a), Asp, Ser, Pro, and Val were highlighted as important on Day 3, and thus, a cocktail of all 4 amino acids was developed and fed on day 3 of the culture.

**Fig. 3. Criteria 1 Validation of Culture Performance with respect too Cell Growth and mAb Productivity:** Amino acids for the growth model (Day 9 CVC) and the Production model (Day 14 Titer) were supplemented to CHO cell culture batches that ran with the high nutrient fed-batch process according to the selection factors from criteria 1 ( $\text{Corr} > 1$ ,  $\text{VIP} > 1$ ,  $\text{SB} > 0$ ). Day 9 and Day 14 values for CVC and Titer were normalized to the maximum value of the process control (black). Error bars represent standard error to mean. (a) Amino acids supplementation for models based on criteria 1 did not result in a significant increase in total cells by day 9. Only the growth model showed a nominal increase of ~10% at day 14 when compared to control ( $p < 0.15$ ). (b) Similarly, no significant increase of mAb productivity was observed when compared to the process control for either model and in the case of the growth model, mAb productivity decreased by ~20% by day 14 ( $p < 0.05$ ).

**Fig. 4. Criteria 2 Validation of Culture Performance with respect too Cell Growth and mAb Productivity:** In addition to the high nutrient fed-batch process, model informed amino acid supplements based on criteria 2 ( $\text{Corr} > 1$ ,  $\text{VIP} > 1$ ,  $\text{SB} < 0$ ) were also tested on moderate and low nutrient fed-batch processes. Accordingly, Day 9 and Day 14 values for CVC and Titer were normalized to the maximum value of each respective process control (Gray Line). Error bars represent standard error to mean. (Relative CVC) Amino acids supplementation based on criteria 2 showed ~20% increase in total cells for high nutrient, but close to a ~30% increase in moderate nutrients conditions ( $p < 0.05$ ). The relative total cells was even more profound in low nutrient reaching near ~55% increase in total cells ( $p < 0.01$ ). (Relative Titer) Antibody titer

in contrast increased only for the production model for high nutrient feeds but drastically for both models in moderate and low nutrient feeds reaching close to ~80% increase in titer by day 9 and a 60% increase in titer by day 14 with respect to each process control ( $p < 0.001$ ).

**Fig. 5. A comparison of total glucose consumption and specific glucose consumption:** To determine if criteria 2 amino acid supplementation for both, the growth and the production models modified the metabolic state of the cells, the total glucose consumption of each culture was compared to the specific glucose consumption per cell. The total glucose consumption was calculated by the change in extracellular glucose concentrations between feed events in the process. Specific glucose consumption was calculated like total glucose consumption however, the difference was divided by the change in CVC between each time point. Although total glucose consumption showed a slight decrease in moderate and low nutrient conditions, a more profound decrease was noticed in the specific glucose consumption rates for moderate and low nutrients feeds suggesting that cells under moderate and low nutrient conditions shifted consumption for carbon source and energy from glucose to other preferred metabolites such as the stoichiometrically balanced amino acids.

**Fig. 6. Criteria 3 Validation of Culture Performance with respect too Cell Growth and mAb Productivity:** To test the effectiveness of the VIP value to select biologically important amino acids only, amino acids were supplemented according to the selection factors from criteria 3 ( $\text{Corr} > 1$ ,  $\text{VIP} > 0.5$ ,  $\text{SB} < 0$ ). Supplementation based on criteria 3 was assumed to result in a minimal change in cell growth and mAb productivity for both models (a) Accordingly, no significant change in D14 CVC was observed across all nutrient conditions, with some cases resulting in a decrease of cell growth. (b) In contrast, there was a significant increase in titer for both moderate and low nutrient conditions in the production model ( $p < 0.05$ ), presumably due to the lack of nutrient levels in those conditions. However, this increase was not as significant as supplementation with criteria 2 amino acids.

**Fig. 7. Driving consumption of model informed amino acids:** To validate the statistical model, it was important to assess if the model driven amino acid supplementations led to an increased consumption of supplemented amino acids and/or activated the consumption of other amino acids. Specific consumption rates were calculated from the change in extracellular amino acid levels between feed events per cell. The resulting specific consumption rates within model conditions were compared to the consumption rates of the respective process control for each nutrient level. Positive changes are reflected by green boxes and negative by red. Gray boxes indicate missing values that were not captured by amino acid assay for various reasons. In addition, amino acids were grouped between glucogenic and ketogenic groups.

**Fig. 8. Media optimization platform with Stoichiometric Balances:** The modeling approach presented in this study was a proof of concept to utilize stoichiometric balances for amino acids to rapidly optimize media development efforts. However, stoichiometric balances and the methods presented could also be applied to other macro-nutrients necessary in CD media. Accordingly, figure 8 presents our approach as a platform to iteratively improve various stages of media development. Utilization of MVDA can easily highlight key stoichiometric balances that can be filtered through a feature selection decision tree to identify the limiting nutrients. Experimental validation studies can then be conducted to assess the change in phenotypical behavior with the modified media. Lastly, the newly generated data can be combined with the original dataset to further inform and improve media development efforts.

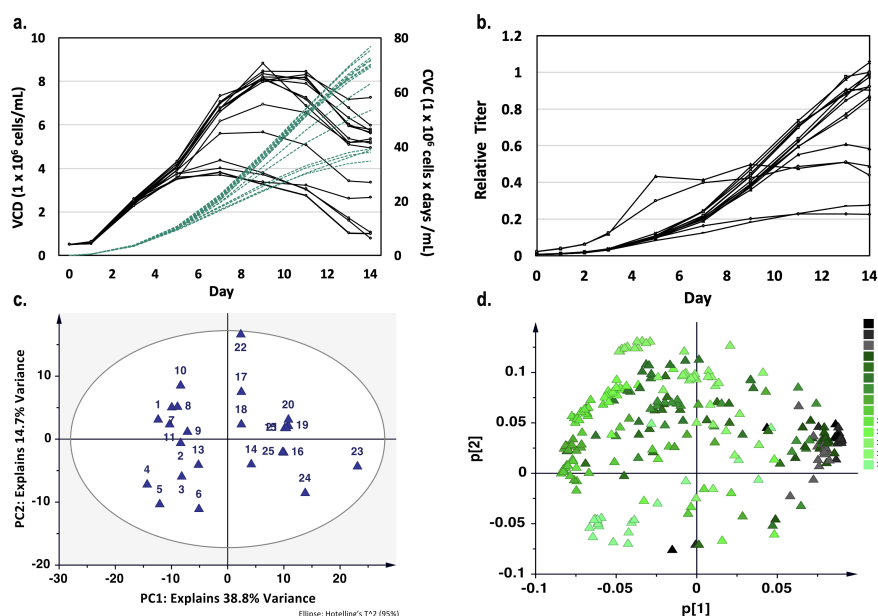
Supplementary Figure Text

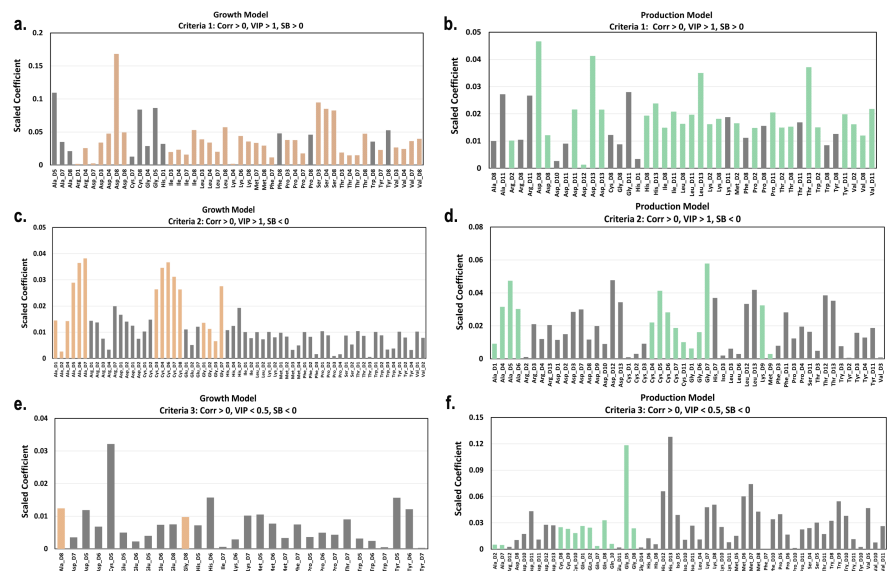
**Fig. S1. Training Model Statistics:** MVDA models using the Orthogonal Partial Least Squares (OPLS) algorithm in SIMCA-P+ were generated using 25 training batches. Models were built towards day 9 CVC or day 14 titer in the batch level model format. In addition to the goodness of fit ( $R^2$ ) parameter, a goodness of prediction ( $Q^2$ ) parameter was also generated that measured the cross-validated prediction accuracy. (a) Observed vs predicted plot for day 9 CVC showed a strong  $R^2$  of 0.912, and a strong  $Q^2$  of 0.726. (b). Observed vs predicted plot for day 14 titer showed a strong  $R^2$  of 0.832 but a predictable accuracy of 0.422. Although the  $Q^2$  was relatively lower, the day 14 titer model was able to highlight information on key

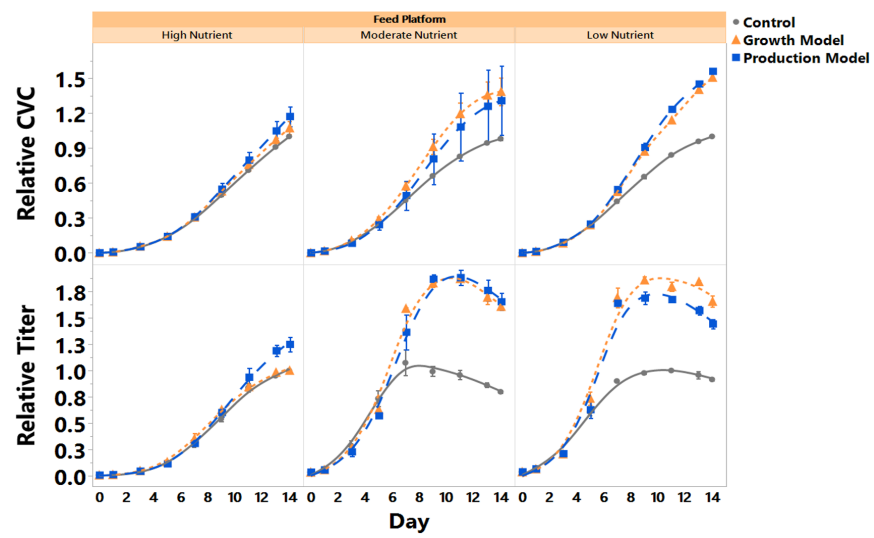
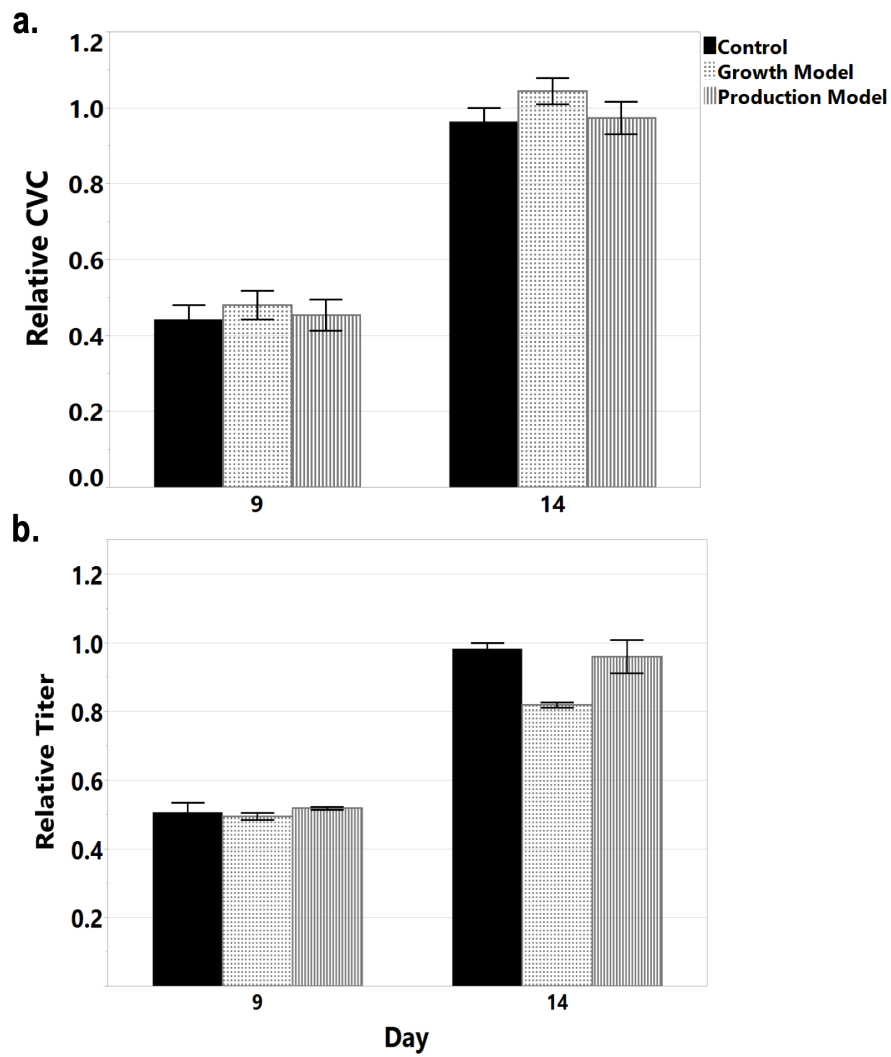
variables for media optimization.

**Fig. S2. Criteria 2 Relative Growth Rate and Relative Specific Productivity.** Relative growth rate for high nutrient condition showed a slight increase by day 3 and maintained the increase by day 7 which could explain the slight increase (~20%) in total cells for high nutrient feeds. Relative growth rates for moderate and low nutrient feeds also showed only slight increases, however remained slightly higher for a longer duration of time than the high nutrient condition. As a result, there was a more significant increase in total cells for moderate and low nutrient conditions. Similarly, relative productivity (qP) showed a more profound increase in moderate and low nutrient conditions. Moderate nutrient conditions showed a 20 increase in qP for the growth model ( $p < 0.05$ ) and showed near a 30% increase for the production model ( $p < 0.01$ ). In contrast, the low nutrient condition showed about a 40% increase in qP starting from day 7 ( $p < 0.05$ ) suggesting the increased mAb production was not only an artifact of increased cell growth, but a higher drive for these cells to produce more monoclonal antibody.

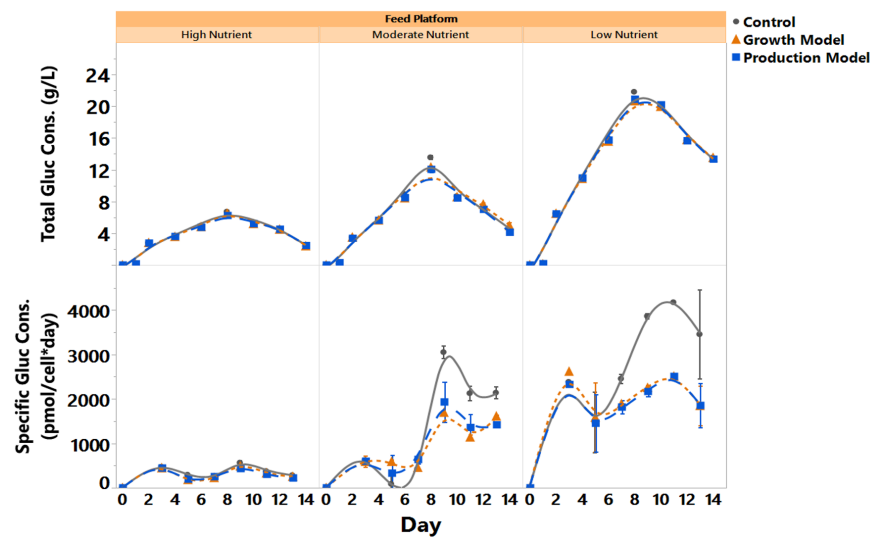
**Fig. S3. Cell Growth and Productivity Changes resulting from Process Controls:** Since amino acid cocktails were made at a 100x concentration for each of the specific amino acids, increased pH by NaOH was needed to solubilize the mixtures. In addition, highly concentrated cocktail mixtures also resulted in a high osmolality. Therefore, to measure the effects of high pH and high osmolality (osmo), a pH water solution was generated at the highest pH among the amino acid cocktail mixtures and fed to the cultures at the same volume as the cocktail mixture. In addition, a high osmo solution in water was generated using NaCl and fed at the same volume as the amino acid cocktail with the highest osmo. pH controls were only measured against the high nutrient conditions and resulted in a slight decrease of about 20% in total cells. Similarly, pH controls showed about a 30% decrease in relative titer. Osmo controls for high nutrient conditions showed similar trends as the pH controls but showed minimal differences for moderate and low nutrient conditions. The minimal overall effects of the pH control and the osmo control on total cells and relative titer suggest that the significant increases in cell growth and mAb production from criteria 2 cultures was attributed to the amino acids within the cocktail mixtures.

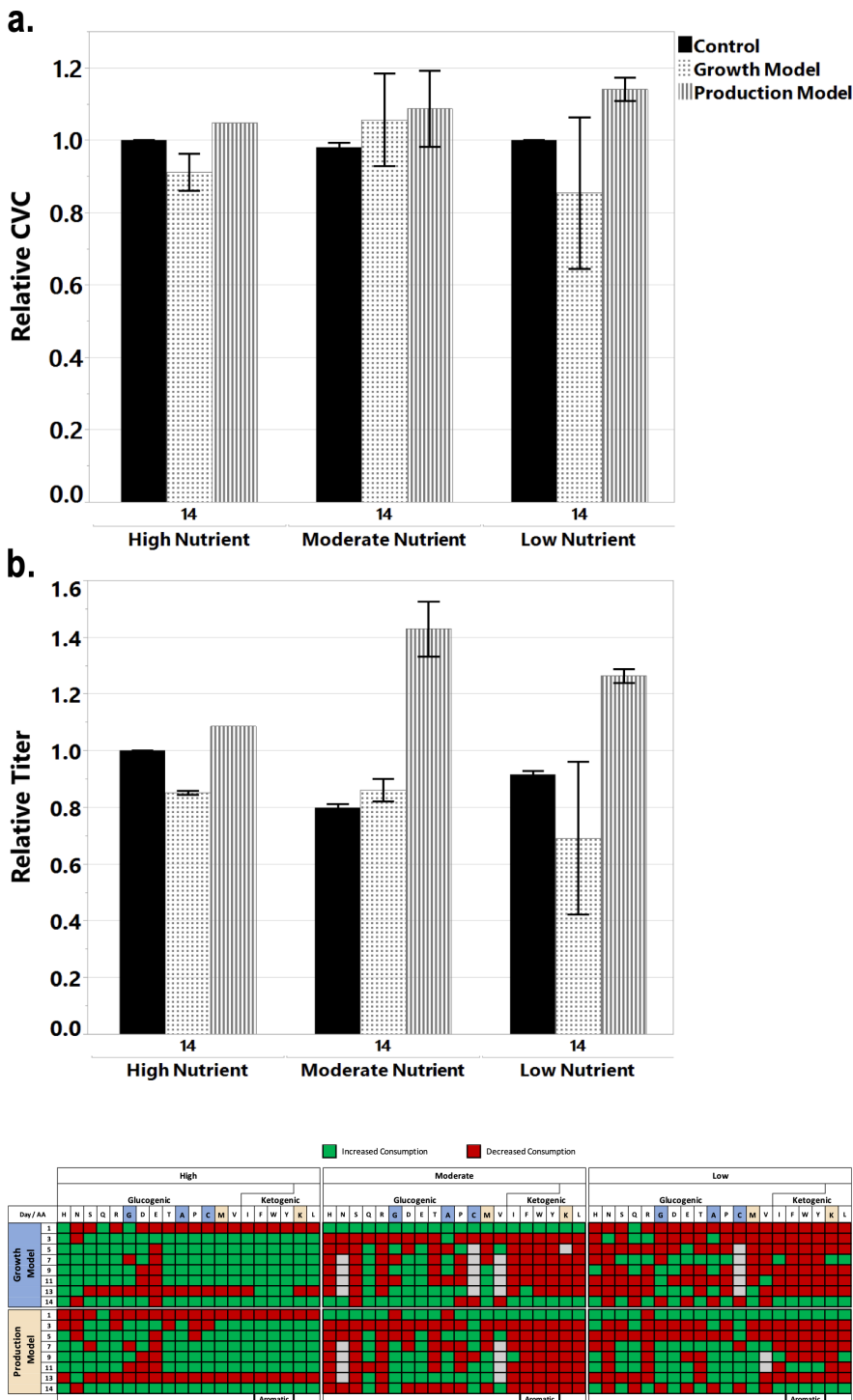


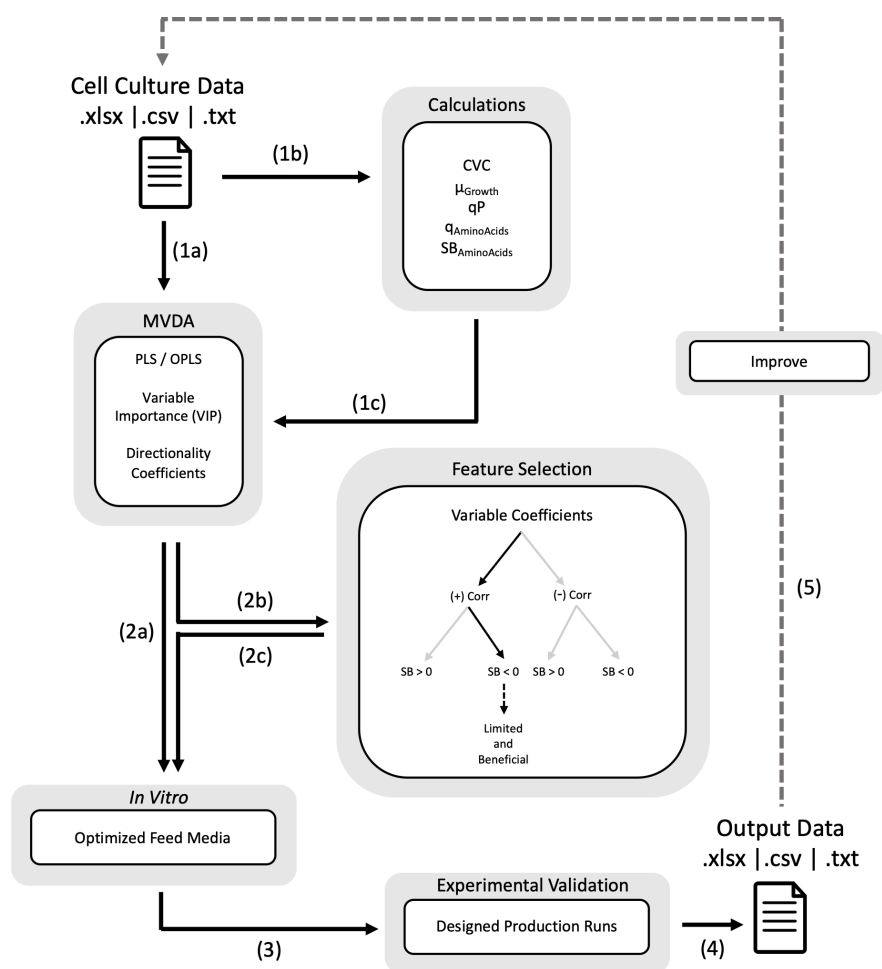












## Hosted file

Table 1.docx available at <https://authorea.com/users/424490/articles/529570-using-mvda-with-stoichiometric-balances-to-optimize-amino-acid-concentrations-in-chemically-defined-cho-cell-culture-medium-for-improved-culture-performance>

## Hosted file

Table 2.docx available at <https://authorea.com/users/424490/articles/529570-using-mvda-with-stoichiometric-balances-to-optimize-amino-acid-concentrations-in-chemically-defined-cho-cell-culture-medium-for-improved-culture-performance>

