

RESEARCH

Open Access



Bigdata analytics identifies metabolic inhibitors and promoters for productivity improvement and optimization of monoclonal antibody (mAb) production process

Caitlin Morris^{1†}, Ashli Polanco^{1†}, Andrew Yongky², Jianlin Xu^{2*}, Zhuangrong Huang¹, Jia Zhao², Kevin S. McFarland², Seoyoung Park², Bethanne Warrack³, Michael Reily³, Michael C. Borys², Zhengjian Li² and Seongkyu Yoon^{1*} 

Abstract

Recent advances in metabolite quantification and identification have enabled new research into the detection and control of titer inhibitors and promoters. This paper presents a bigdata analytics study to identify both inhibitors and promoters using multivariate data analysis of metabolomics data. By applying multi-way partial least squares (PLS) model to metabolite data from four fed-batch bioreactor conditions where feed formulation and selection agent concentrations varied, metabolites which exhibited the most significant impact on titer during cultivation were ranked from highest to lowest. The model outputs were then constrained to reduce the number of statistically relevant inhibitors or promoters to the top ten, which were used to conduct metabolic pathway analysis. Furthermore, a method is presented for identifying amino acids that prevent the accumulation of the inhibitors and/or enhance the formation of promoters during production. Finally, the metabolomics and pathway analysis results were integrated and validated with transcriptomics data to characterize metabolic changes occurring among different growth conditions. From these results, new feeding strategies were implemented which resulted in increased fed-batch production titer. Methodology from this work could be applied to future process optimization strategies for biotherapeutic production.

Keywords: Metabolomics, Multivariate data analysis, Titer improvement, Feed media optimization

Introduction

Monoclonal antibodies (mAbs) are traditionally manufactured in large-scale bioreactors which require cells, metabolites, nutrients, air, and other life-supporting compounds. As the cellular environment has a direct impact on the growth, metabolism, and protein production in

mammalian cells, establishment of process parameters which optimize yield and maintain consistent product quality is essential to the biotherapeutics industry. Throughout the cultivation of cells in a bioreactor, several samples are analyzed to assure that the reactor is operating properly, assess cell health, and analyze the consumption or formation of key metabolites that directly impact mAb quality (Dunn and Ellis 2005). In addition, multiple titer measurements, as well as other monitoring techniques (i.e., pH, temperature, % DO, CO₂) are employed which lead to an overwhelming amount of data left for analysis. Although these measurements ensure cells are growing and producing at an ideal rate, they do not

*Correspondence: jianlin.xu@bms.com; Seongkyu_Yoon@uml.edu

[†]Caitlin Morris and Ashli Polanco contributed equally to this manuscript

¹ Department of Chemical Engineering, University of Massachusetts Lowell, University Ave, Lowell, MA 01854, USA

² Biologics Process Development, Bristol-Myers Squibb, 38 Jackson RD, Devens, MA 01434, USA

Full list of author information is available at the end of the article

provide insight on the metabolism within the cells which can be influenced by changes in raw materials such as the cell culture media composition or nutrients supplemented during fed-batch processes.

Metabolomics enables the identification and examination of biochemical reactions and metabolic pathway activation occurring within living cells (Zhang et al. 2013). More specifically, metabolomics has been used to characterize the impact of small molecules (i.e., amino acids and metabolites) present in cell culture media on CHO cell growth behavior and productivity (Mohmad-Saberi et al. 2010; Chong et al. 2009). Levels of identifiable metabolites in cell culture media or feed can pose a significant impact on either protein titer or cell growth in a positive or negative manner, and are commonly referred to as promoters or inhibitors, respectively (Mohmad-Saberi et al. 2010; Mulukutla et al. 2017). This approach when integrated with gene expression analysis has shown to reveal connections among complex cellular regulatory pathways and the biologic response of cells to selection agents, genetic factors, and/or environmental changes (Ferrara et al. 2008; Cuperlović-Cul et al. 2010). Therefore, a useful bioprocessing optimization strategy would involve identifying, understanding, and manipulating the production and consumption of metabolites of interest to enhance cell productivity.

Traditionally, inhibitors and promoters are identified through extensive data analysis of metabolite profiling data derived from various instrument platforms including nuclear magnetic resonance (NMR) spectroscopy, mass spectrometry, chromatographic and column separation, vibrational spectroscopies such as Fourier-transform infrared (FTIR) and hybrid instruments such as gas chromatography mass spectrometry (GC-MS) and liquid chromatography mass spectrometry (LC-MS) (Khoo et al. 2007). Most of these platforms are insensitive (i.e., NMR), require chemical derivatization of biomolecules (i.e., GC-MS), or destroy the sample via laser acquisition (i.e., FTIR or Raman) which can be time consuming, costly, and may require volatile chemicals (Bradley et al. 2010; Griffin 2003; Dippel 2011). LC-MS has proven to be an effective and robust platform for identifying and quantifying metabolites in mammalian cell culture that requires no derivatization of samples. The analysis of LC-MS metabolomic data can be used to classify inhibitors, for example, through identifying correlations between low producing cells and the metabolites that accumulated in that bioreactor culture. One such inhibitor that was identified through trials of experimentation is lactate, which is known to be a by-product of cellular metabolism and inhibitor to cell growth and titer during late-stage culture (Gagnon et al. 2011). Many companies have adapted strategies to control lactate accumulation

(Gagnon et al. 2011) but very few have adapted strategies that address the various other growth and/or titer inhibitors and promoters present during fed-batch cultures.

Strategies for promoting higher titers and avoiding the accumulation of inhibitors are currently being explored by research and development teams across the industry. One such method involves monitoring and supplementing amino acid levels in real time, ensuring that concentrations within the bioreactor remain within specified limits that prolong culture time while maintaining antibody product quality (Powers et al. 2019). The theory behind this approach is that the inhibitors and promoters that are being formed in the bioreactor are derived from amino acids (Mulukutla et al. 2017). This is supported by metabolic pathway analysis which shows the correlation between a given amino acid and the associated metabolite (Mulukutla et al. 2010; Mulukutla et al. 2017). In addition, integration of transcriptomics analysis with metabolite profiling and metabolic pathway analysis can serve as validation method (Bradley et al. 2010; Lei et al. 2011). The following research presents a novel method to identify both titer inhibitors and promoters through multivariate analysis of metabolomics data. By altering specific amino acid concentrations in the feed media, cell culture performance LC-MS data can serve as an input for chemometric models that identify key metabolites that influence cell productivity. In this study, chemometric model outputs led the pathway enrichment analysis to identify metabolic pathways activated that lead to increased titer, and findings were validated with transcriptomic differential gene expression analysis. Then, a new feed medium was designed based on amino acid level changes that the original model predicted would improve titer. Finally, a validation bioreactor run confirmed that the model accurately identified specific metabolite promoters and inhibitors that impact titer.

Materials and methods

Cell line, media, and cell culture

A CHO-K1 GS knockout cell line was used for expression of a proprietary recombinant monoclonal antibody. Proprietary chemically defined seed, basal, and feed media were used in this study unless otherwise specified. Preliminary cell culture experiments were conducted with four different production culture conditions based on the level ($1\times$ or $4\times$) of selection agent methionine sulfoximine (MSX) in the seed train. All the production cultures used the same proprietary basal medium containing no MSX. The feeding conditions include a standard feed formulation and the standard feed supplemented with four amino acids that had been previously identified as being useful in improving productivity. The initial experimental conditions are described in Table 1.

Table 1 Initial experimental conditions

Condition	Reactor label	Expansion medium	Basal medium	Feed medium
1	V201, V202 and V203	Seed medium with 1X MSX	Basal medium	Feed medium
2	V204 and V205	Seed medium with 1X MSX	Basal medium	Feed medium with an increased concentration of Ser, Thr, Tyr, and Lys
3	V206, V207 and V208	Seed medium with 4X MSX	Basal medium	Feed medium
4	V209 and V210	Seed medium with 4X MSX	Basal medium	Feed medium with an increased concentration of Ser, Thr, Tyr, and Lys

Table 2 Standard bioreactor conditions

Parameter	Condition
Seeding density	1.5×10^6 cells/mL
Working volume	3.3 L
Temperature	36.5 °C
Dissolved oxygen target	40%
Agitation	260 rpm
pH range	7.1 ± 0.5

Each experimental condition was tested in duplicate or triplicate in 5-l bioreactors under standard conditions (Table 2) over the course of 14 days, resulting in 10 total sample sets. Feeding at 3.5% (v/v) of initial volume was started when appropriate viable cell density (VCD) was achieved and fed daily thereafter. Additional glucose was supplied to maintain its concentration at a proprietary level. The pH was controlled through the addition of CO₂ gas to decrease pH or addition of 1 M Na₂CO₃ base to increase pH as needed. Daily samples were drawn from the bioreactors to measure cell density and viability, while supernatant and cell pellets were collected and stored as well. Viable cell density and viability were quantified off-line using a Vi-CELL XR automatic cell counter (Beckman Coulter). The supernatant samples were used to determine titers, metabolite concentrations, and amino acid concentrations. A Protein A high-performance liquid chromatography (HPLC) was used to measure the protein titer (Xu et al. 2018). For bioreactor samples, off-line pH, pCO₂, and pO₂ were detected using a Bioprofile pH/Ox analyzer (Nova Biomedical). The cell pellet samples were used for RNA-sequencing analysis.

UHPLC-mass spectrometric metabolomics data collection and analysis

Frozen supernatant samples were thawed at room temperature for 2 h and gently vortexed. A 50- μ L aliquot of each sample was transferred to the corresponding well of

an Axygen 96-well plate (Corning Life Sciences, Edison, NJ). 150 μ L of ice-cold methanol containing 0.1% formic acid and d5-glutamic acid, d3-carnitine, d8-phenylalanine, d5-hippuric acid, d16-sebacic acid, d4-palmitic acid, d3-octanoyl carnitine, and d4-deoxycholic acid as internal standards were added to all samples. These internal standards were used to ascertain that the mass spectrometer performance was stable during the analysis of the set. The plate was vortexed for 1 min and then spun for 10 min at 5000 rpm in a Beckman Coulter Allegra 25 centrifuge with a TA 10.25 rotor (Beckman Coulter Inc., Indianapolis, IN). 50 μ L of supernatant was transferred to a new 96-well plate for reversed phase LC-MS analysis and a second 96-well plate was similarly prepared with another 50- μ L aliquot for hydrophilic interaction liquid chromatography mass spectrometry (HILIC-MS) analysis. The reversed phase and HILIC plates were evaporated to dryness under nitrogen.

Reversed phase LC-MS

The reversed phase plate was reconstituted with 90:10 water:methanol. Samples were analyzed by LC-MS using a Nexera ultra-high-performance liquid chromatography (UHPLC or UPLC) (Shimadzu Scientific Instruments, Columbia, MD) interfaced to an Exactive Plus ion trap mass spectrometer with a heated electrospray ionization (HESI) source (ThermoFisher Scientific, San Jose, CA). Chromatographic separations were achieved employing a 2.1×150 mm, 1.7 μ m, Acquity BEH C18 column (Waters, Milford, MA) with gradient elution at 0.6 mL/min. The column temperature was maintained at 65 °C. Mobile phase A was water with 0.1% formic acid and mobile phase B was 98:2 acetonitrile:water with 0.1% formic acid. Mobile phase A was held at 99% for 0.5 min and then a three-step linear gradient was formed from 1% to 20% mobile phase B over 2.5 min, to 60% mobile phase B in 1 min and then to 100% phase B in 3 min. The final composition was held for 2 min before returning to the initial conditions. Positive and negative electrospray

ionization (ESI) data were acquired (separate injections) with a mass accuracy within 5 ppm at 35,000 resolution. A single 10- μ L injection was used for each ionization mode. Instrumental settings follow maximum injection time 10 ms, capillary temperature 275 °C; tube lens voltage 175 V; ESI spray voltage 3.7 kV; sheath gas 60 arbitrary units (arbs).

Hilic-ms

The HILIC plate was reconstituted with 25:75 methanol:acetonitrile. Samples were analyzed by LC-MS as described above. Chromatographic separations were achieved by employing a 2.1 \times 150 mm, 1.7 μ m, Acquity BEH Amide column (Waters, Milford, MA) with gradient elution at 0.3 mL/min. The column temperature was maintained at 65 °C. Mobile phase C was 95:5 water:acetonitrile with 10 mM ammonium acetate and 0.05% ammonium hydroxide. Mobile phase D was acetonitrile with 0.05% ammonium hydroxide. A linear gradient was formed from 5% to 63% mobile phase C over 3.5 min. The final composition was held for 5 min before returning to the initial conditions.

Peak areas for LC-MS quantification of 150 metabolites were calculated using Component Elucidator, a software package developed at Bristol-Myers Squibb (Hnatyshyn et al. 2013). Quantification of amino acid concentrations in supernatant was conducted using a UPLC system. The supernatants were diluted in Milli-Q water containing a standard. The derivatization reagent (Waters AccQ•Tag Ultra Derivatization Kit, Cat#03836), 6-aminoquinolyl-N-hydroxysuccinimidyl carbamate, was added into the diluted samples and vortexed at 55 °C for 10 min. The derivatized amino acids were detected at 260 nm as they eluted from the column.

Chemometric analysis of metabolomic data

Final metabolite concentrations as well as daily pH, viable cell density profile, and viability profile were compiled into a dataset with approximately 140 rows by 2400 columns. Preliminary analysis of the bioreactor runs was performed using SIMCA[®] software (version 15, Umetrics, Umeå, Sweden) via principal component analysis (PCA) from which a score plot summarizes the relationship between the different reactor conditions. A partial least squares (PLS) model was then applied to determine the effect of given input variables on a designated output (Wu et al. 2010). In this way, the relationships between multiple inputs and multiple outputs can be explained (Bylesjö et al. 2006).

From the PLS model, a batch evolution model (BEM) and a batch level model (BLM) were constructed so that the entire bioreactor dataset could be analyzed as separate batches, with each batch being a different reactor

condition (V201 through V205). The BEM produces control charts for each variable which can show time-based trends that can be used to identify potential avenues for further investigation. These control charts display the normal process “trajectory” or “path” a batch may take which enables the user to identify batches that do not evolve in the normal way by tracking deviations in the control chart (Svante and Michael 1977). Therefore, a BEM can be used to detect early batch deviations and can provide avenues for batch maturity prediction.

Once the raw data from the BEM is processed, a BLM is generated via batch decomposition. The BLM searches for correlations between the data matrix X and the output matrix Y (Svante and Michael 1977), then decomposes the data provided to the BEM into a set of several matrices. Each matrix contains the data for one parameter (such as metabolite concentration, pH, etc.) for every reactor for every time point. This differs from the BEM which contains the data for all parameters in one matrix. In this dataset, it was found from the BEM control charts that there were deviations in the titer values for specific bioreactor conditions. Thus, potential effects of the other amino acids and metabolites on titer were investigated further using the BLM.

To determine how metabolites, amino acids, and other bioreactor conditions influence titer during each bioreactor run, variable influence on projection (VIP) plots and coefficient correlation matrices were generated from the BLM. These plots quantify the influence of an input variable on an output variable and determine the relative positive or negative correlation among input and output variables, respectively. To reduce the number of metabolites being analyzed via pathway analysis, a VIP cut-off value of 1.35 was applied, where a VIP value greater than 1.0 indicates a variable that is statistically significant. Utilizing this VIP cut-off value, the top 10 metabolites that correlated to titer were considered as pathway candidates for pathway enrichment analysis. A coefficient correlation matrix was used to distinguish a positive correlation coefficient as promoter-like behavior or a negative correlation coefficient as inhibitor-like behavior for each metabolite influence on final titer and each amino acid influence on the metabolites of interest. Correlations among the top 10 metabolites and the 20 amino acids were carried out using a PLS-DA model. Then, a VIP plot was generated (not shown) with a cut-off criterion of 1*** rank the amino acids in terms of its effect on the concentration of the given metabolite was formed.

Pathway enrichment analysis

Chosen pathway candidates were inputted into Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway mapping tool (<https://www.genome.jp/kegg/>) using

Cricetulus griseus (Chinese hamster) as the species to determine the metabolic pathways that are impacted (Kanehisa and Goto 2000; Kanehisa et al. 2016). In addition, amino acid metabolism pathways were assessed using KEGG to identify upregulated pathways in samples with higher titer. Then a pathway network was manually constructed with the pathways identified. A list of some of the important pathways is provided in Additional file 1: Table S1. Pathway analysis was then validated with transcriptomics data.

Transcriptomics

RNA sequencing (RNAseq) was performed to identify genes that were up-regulated and down-regulated within each of the bioreactors to gain a more in-depth understanding of the biological mechanisms that impact titer. This dataset was used to evaluate the metabolomics approach to identifying inhibitors and promoters by correlating specific changes in CHO cell titer to differential gene expression data. Cell pellet samples for RNA-Seq analysis were collected at day 6 and day 10, corresponding to exponential growth and stationary phase, respectively. Approximately 5×10^6 cells were separated from culture broth by centrifugation at 1000 rpm for 10 min. Then the cell pellets were immediately frozen in (dry ice or liquid nitrogen) and stored at -80 °C. For the detailed information of mRNA isolation and sequencing, one can refer to Sha et al. (2018).

Gene differential expression and function analysis

The results of RNA-Seq by Expectation Maximization (RSEM) in the form of expected counts of gene levels were imported to an interactive interface for data examination and differential expression called DEBrowser (Kucukural et al. 2018). Principal component analysis (PCA) was then applied to cluster samples with trimmed mean of M-values (TMM) method and analyze the top 1000 most variant genes between the compared sample groups. Using DESeq 2 (Love et al. 2014), differential expression (DE) analysis was performed between the sample groups as well. DESeq 2 analysis provides an output of differentially expressed genes (DEGs) listed by their relative p-adj values and fold change. Positive and negative log-fold change values determined if specific genes were upregulated or downregulated between the compared sample groups.

RNAseq results were analyzed in various ways. First, gene expression across the tested bioreactor conditions was compared by performing PCA on the top 1000 most variant genes expressed. This analysis allowed for clustering of gene expression variations based on day collected, selection agent concentration, and amino acid concentration. RNA-sequence data were further analyzed in

two-condition comparisons that determine the top genes that were up-regulated and which genes were down-regulated based on the fold change in gene expression. KEGG database of biological systems was utilized to determine the reaction networks associated with upregulated genes in higher titer batches (Kanehisa et al. 2017). GO-enrichment software was used to confirm pathways that the top identified up- or down-regulated genes were associated with. From there, pathway analysis was confirmed via the relationship between the upregulated genes and the role of amino acids as well as metabolites in metabolic pathways that impact protein production.

Data integration

To effectively compare the results from all three methodologies, it was necessary to convert the results to an integrated table listing consistencies among findings. To this end, each set of results (either a set of genes or metabolites) from each omics technique was associated with the metabolic pathways that are most impacted during protein production. This type of data processing is a potential analysis tool that will guide future process optimization. In this way, an experimental proposal could be more strongly supported by such a metabolic pathway analysis tool before the actual experiment is performed. The results of the integrated data analysis will be presented in “Results” section.

Validation experiment

Based on the combined assessment of the BLM, metabolic networks, and transcriptomics data, a new feed medium was designed by altering the levels of specific amino acids that were found to correlate the most with the top 10 identified metabolites that influence final titer. Their respective concentrations were adjusted based on the model either to enhance or inhibit titer production. Feeds were altered by increasing or decreasing specific amino acids by 100%. Figure 6 summarizes the bioreactor conditions used to perform the validation experiment.

Results

Multivariate data analysis

Following the completion of the parallel bioreactor run composed of the conditions listed in Table 1 and bioreactor settings listed in Table 2, a large dataset was generated. This dataset included viable cell densities, metabolite and amino acid concentrations, and titer values. This dataset was analyzed in SIMCA via a decomposition methodology.

The decomposition of a raw data set into a BEM as performed by SIMCA is shown in Fig. 1. SIMCA decomposes multiple data points comprising several batches into a BLM based on the 3-dimensional matrix composed

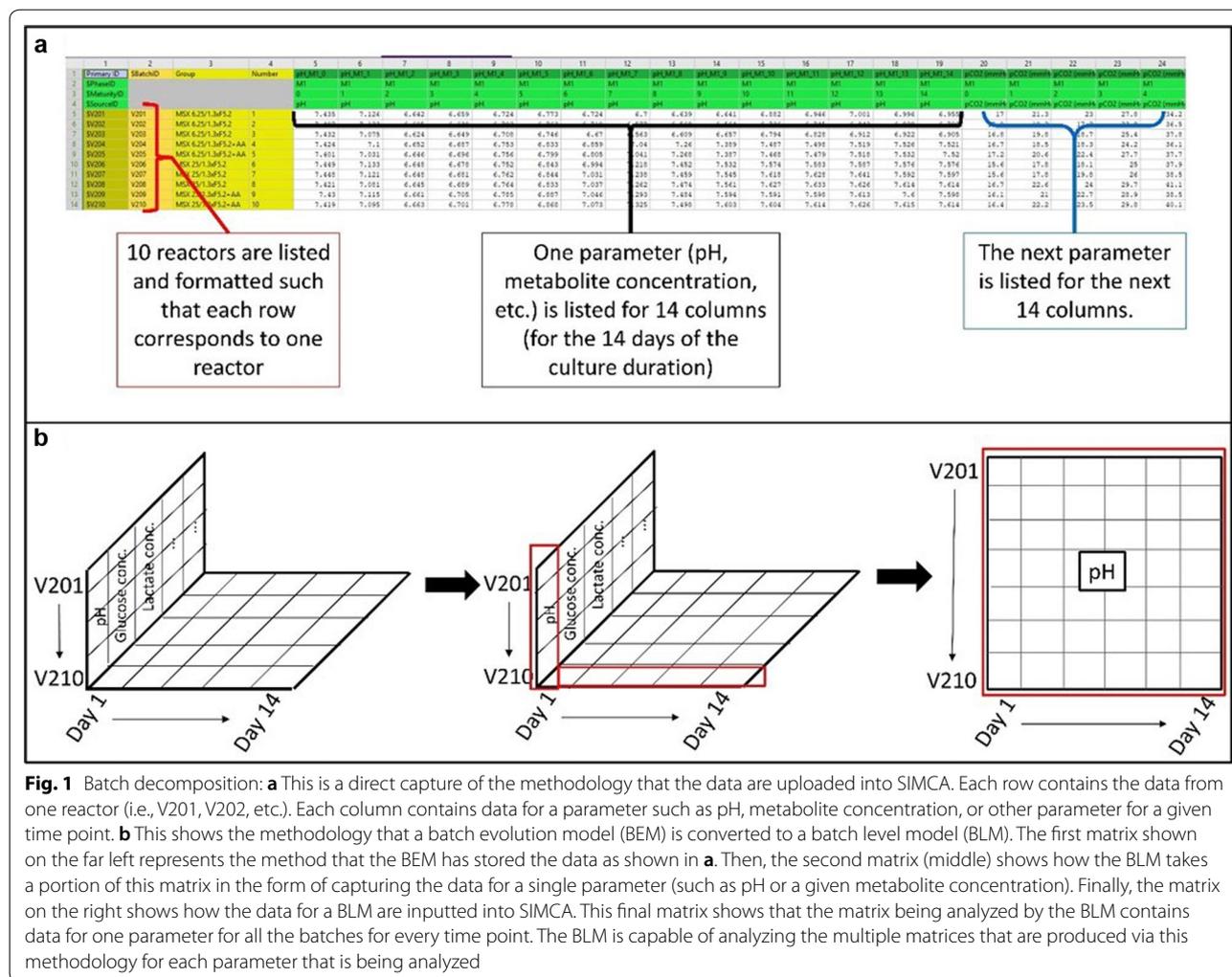


Fig. 1 Batch decomposition: **a** This is a direct capture of the methodology that the data are uploaded into SIMCA. Each row contains the data from one reactor (i.e., V201, V202, etc.). Each column contains data for a parameter such as pH, metabolite concentration, or other parameter for a given time point. **b** This shows the methodology that a batch evolution model (BEM) is converted to a batch level model (BLM). The first matrix shown on the far left represents the method that the BEM has stored the data as shown in **a**. Then, the second matrix (middle) shows how the BLM takes a portion of this matrix in the form of capturing the data for a single parameter (such as pH or a given metabolite concentration). Finally, the matrix on the right shows how the data for a BLM are inputted into SIMCA. This final matrix shows that the matrix being analyzed by the BLM contains data for one parameter for all the batches for every time point. The BLM is capable of analyzing the multiple matrices that are produced via this methodology for each parameter that is being analyzed

of the different batches (V201 through V210), the time-point at which samples were collected in days (1 to “n” (or 0 to 14 in this case)), and input variables (metabolite concentration data). An example of how the raw data is processed by SIMCA is shown in Fig. 1a. Figure 1b displays how SIMCA breaks down the batch information to compare the time series information per batch for one variable at a time. Starting from the left in Fig. 1b, the three-dimensional matrix containing cell culture performance data for each day is shown. Then, moving to the right side of Fig. 1b, SIMCA decomposes a portion of this 3D matrix and forms 2D matrices that contain timewise data for each parameter across batches (Wu et al. 2010). The BLM can then analyze multiple matrices for each parameter at once, which was advantageous for such a large dataset.

Graphical outputs from the BEM model are shown in Fig. 2. The proximity of batch scores on the score plot confirms the similarities among the performance of

bioreactors that were operated under the same conditions. The duplicate and triplicate bioreactor conditions tend to cluster, and their relative final titer values are designated by a color spectrum. The score plot dot colors indicate that bioreactor V205 (Condition 2) had the highest final titer.

Clustering patterns and titer color gradient of the score plot imply that amino acid addition to the low MSX bioreactor condition caused an increase in final titer. Thus, as a starting point of the analysis, metabolomics data were analyzed to determine how and why these specific amino acids help to improve the titer. The hypothesis was the added amino acids were either *constraining* the production of inhibitory metabolites or *promoting* the production of metabolites that stimulate protein production.

Half of the top ten identified metabolites from the VIP plot are characterized as fatty-acid-like compounds (as denoted by the C# DCA (dicarboxylic acid) metabolite names) (Fig. 3). Other metabolites in the top ten

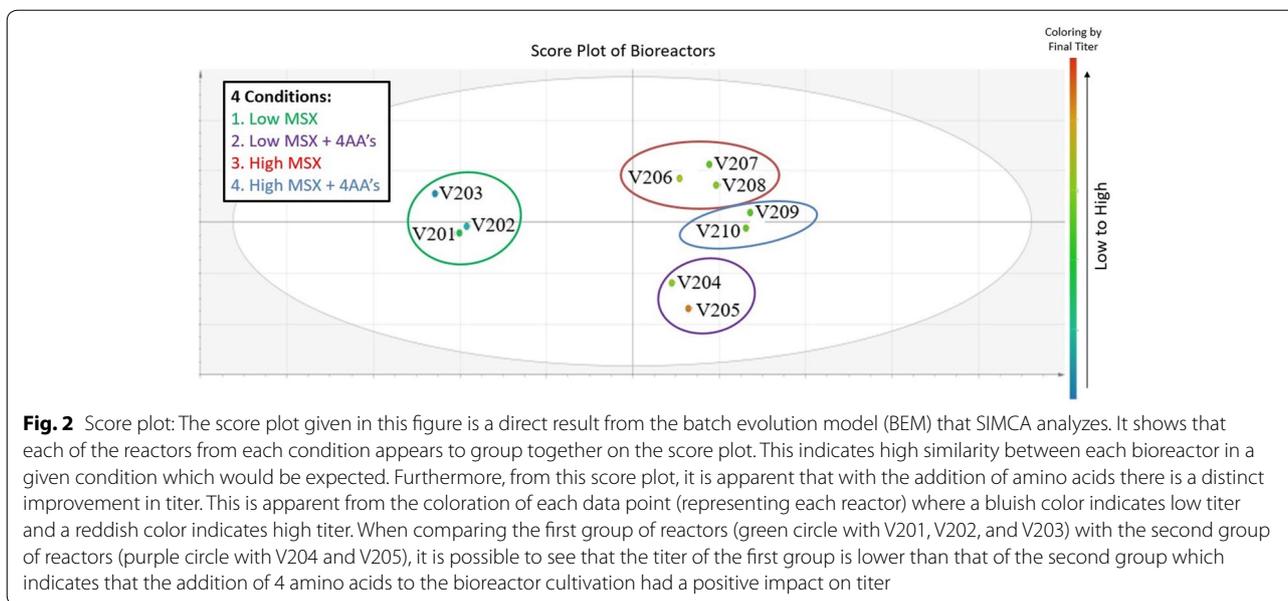


Fig. 2 Score plot: The score plot given in this figure is a direct result from the batch evolution model (BEM) that SIMCA analyzes. It shows that each of the reactors from each condition appears to group together on the score plot. This indicates high similarity between each bioreactor in a given condition which would be expected. Furthermore, from this score plot, it is apparent that with the addition of amino acids there is a distinct improvement in titer. This is apparent from the coloration of each data point (representing each reactor) where a bluish color indicates low titer and a reddish color indicates high titer. When comparing the first group of reactors (green circle with V201, V202, and V203) with the second group of reactors (purple circle with V204 and V205), it is possible to see that the titer of the first group is lower than that of the second group which indicates that the addition of 4 amino acids to the bioreactor cultivation had a positive impact on titer

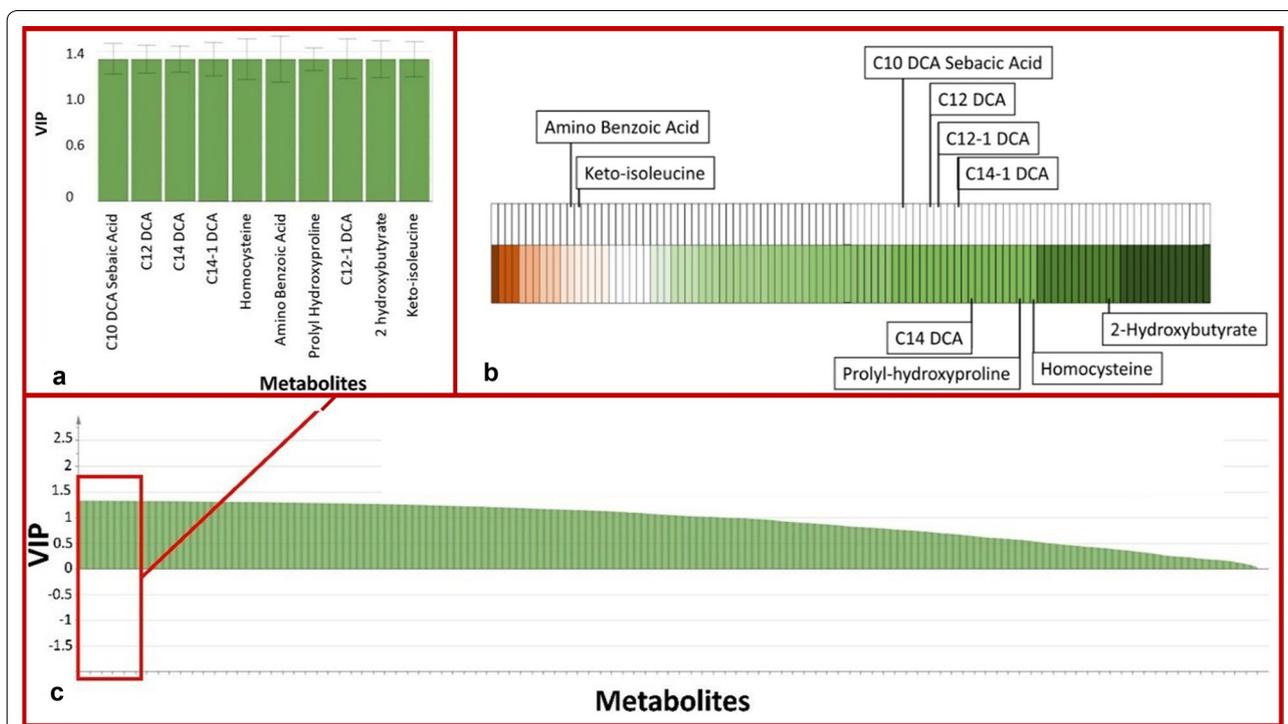


Fig. 3 VIP Plot and Correlation Heat Map generated from BLM model (R^2X (cum) = 0.635, R^2Y (cum) = 0.992, and Q (cum) = 0.957). The VIP plot generated from the BLM model using a PLS approach includes all the metabolites analyzed in the study (c). For simplicity, a cut-away of the VIP plot displays the top 10 metabolites in closer detail (a). A correlation matrix in the form of a heat map (b) displays the relative metabolite correlations to titer as positive (green) or negative (red). Highlighted in boxes are the locations of the top ten metabolites identified from the VIP plot

included some that have been previously identified (see discussion) and others that have not yet been highlighted as important to control during cell culture. All but two of the top ten metabolites had positive correlation

coefficient with respect to titer, indicating promoter-like behavior, with the exception for amino benzoic acid and keto-isoleucine which had negative correlation coefficient values. Characterized by pathway analysis, the top

identified metabolites are involved in several different pathways across CHO metabolism including the metabolism of lipids and transport of inorganics (as identified by KEGG). As these metabolites are not easily found in the central metabolism pathways (which includes glycolysis and the citric acid cycle) or pathways for degradation or production of amino acids, it is unlikely that they would be easily identified as regulators of titer without the use of MVDA. Overall, the metabolites listed would be difficult to pinpoint without prior knowledge. Furthermore, the process of identifying the amino acids that correlate with these metabolites would have been laborious.

Subsequent VIP and coefficient plots (not shown) generated from a PLS-DA model built using the concentration data versus time for the 20 amino acids and each metabolite of interest provided a list of the amino acids ranked by their significance to each metabolite's formation (+) or inhibition (-). For example, for C10 DCA Sebacic Acid, alanine, aspartate, and glutamine scored the highest values on the VIP plot, yet alanine correlated positively (+), aspartate correlated negatively (-), and glutamate correlated negatively (-) with the concentration of this metabolite. Therefore, to promote the formation of this metabolite *only*, a feed should be designed with higher alanine and lower aspartate/glutamate concentration. Table 3 lists the top ten identified, their type of correlation (positive or negative) with final titer, and the top three amino acids that were found to be correlated to each metabolite.

Transcriptomics analysis

Transcriptomic analysis was carried out on 1 × and 4 × MSX samples collected at different time points (days 0, 6, and 10) to support the metabolomic results from the initial screening. Initial PCA (Fig. 4a) shows that both 1 × and 4 × MSX seed trains have highly correlated gene

expression at day 0. By day 6, all sample conditions still cluster very tightly, indicating similar gene expression at the end of the exponential growth phase. At day 10, gene expression varies the most, especially among 1 × MSX samples with and without supplemented amino acids (Fig. 4b). This observation at day 10 supports the claim that the added amino acids do not impact cell growth and titer until the stationary phase is reached (Templeton et al. 2017). However, amino acid addition did not seem to impact gene expression of 4 × MSX samples as it did 1 × MSX samples. This is most likely due to the increased selection stringency and inhibition of enzyme activity in both the glutamine synthetase (GS) and γ -glutathione synthetase (GSH) pathways, which occurs in a selection agent concentration-dependent manner (Fan et al. 2013). Gene expression changes for 1 × MSX samples at day 10 were further analyzed to determine function of added amino acids in CHO cell productivity and growth.

Differential gene expression (DGE) analysis comparing the "1 × MSX + AAs" condition that produced the higher titer (V204 and V205) to the control batches (V202 and V203) revealed a large number of upregulated and down-regulated genes as displayed in the volcano plot (Fig. 5). From the DGE output, a list of the top 1000 upregulated and down-regulated genes in V204/V205 (based on fold change and p-agj values) was generated and this list of genes was utilized to conduct further pathway analysis using KEGG.

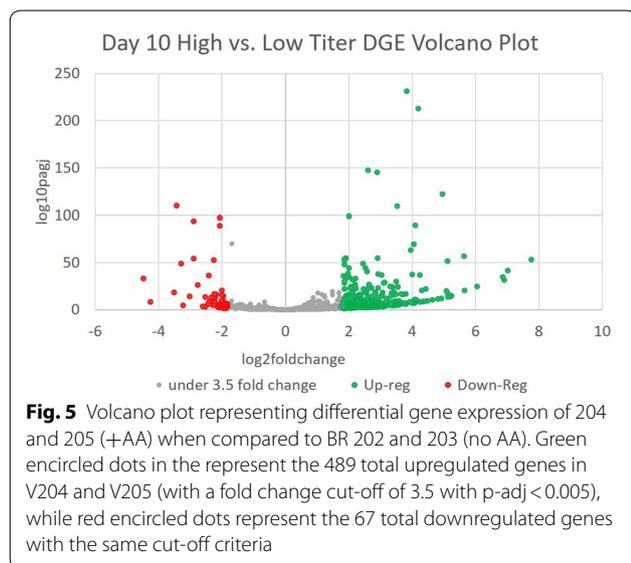
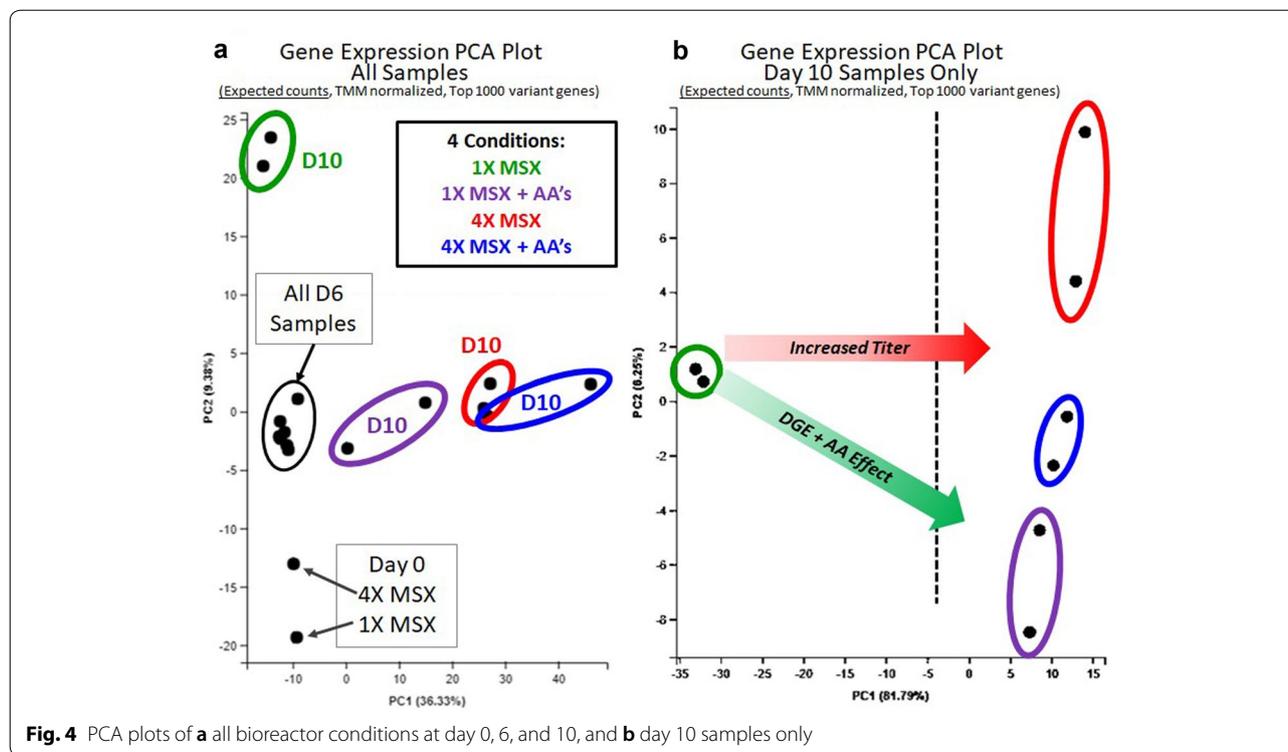
KEGG database search for CHO cell metabolism pathways involving supplemented AA (Ser, Lys, Thr, Tyr)

There are three pathways which are involved in the metabolism of the four amino acids of interest, including the glycine, serine, threonine metabolism pathway, tyrosine metabolism pathway, and lysine degradation pathway. Within each of these pathway maps, upregulated genes

Table 3 Metabolite list with characterization and corresponding amino acids

Order by VIP	Metabolite	Correlation with titer	Top amino acids ^a
1	C10 DCA sebacic acid	Positive	1. Ala (+) 2. Asp (-) 3. Glu (-)
2	C12 DCA	Positive	1. Ala (+) 2. Asp (-) 3. His (+)
3	C14 DCA	Positive	1. Asp (-) 2. Ala (+) 3. Ile (-)
4	C14-1 DCA	Positive	1. Ala (+) 2. Asp (-) 3. His (+)
5	Homocysteine	Positive	1. Ala (+) 2. Asp (-) 3. His (+)
6	Amino benzoic acid	Negative	1. Ala (+) 2. Ile (+) 3. Leu (+)
7	Prolyl-hydroxyproline	Positive	1. Ala (+) 2. Ser (-) 3. Pro (-)
8	C12-1 DCA	Positive	1. Ala (+) 2. Asp (-) 3. His (+)
9	2-Hydroxybutyrate	Positive	1. Thr (+) 2. Trp (-) 3. Ala (+)
10	Keto-isoleucine	Negative	1. Glu (+) 2. Ser (-) 3. Thr (-)

^a The sign next to each amino acid indicates the correlation that amino acid has with the given metabolite (i.e., a "+" sign indicates that the amino acid has a positive correlation or promoter-like behavior with the given metabolite)



present in higher titer samples were located to determine the function of the supplemented amino acid of interest. Table 4 outlines the pathways, identified upregulated genes based on gene expression fold change and p value in V205 (the highest titer condition), as well as the identified gene function during metabolism.

Four genes (GCDH, ECHS1, HADH, OGDH) related to the lysine degradation pathway in CHO cells were upregulated in amino acid-supplemented cultures. The genes involved function specifically to convert Glutaryl-CoA into Acetyl-CoA which is consumed by the citrate cycle (TCA cycle). Two genes (FAH and FAHD1) related to tyrosine metabolism are upregulated in amino acid supplemented cultures, specifically genes for enzymes that convert 4-fumaryl acetoacetate and 3-fumaryl-pyruvate to fumarate, which is a precursor to the TCA cycle. Finally, one gene (SDSL) involved during glycine, serine, and threonine metabolism pathway is upregulated which functions in two ways: (1) conversion of serine to pyruvate, a precursor for TCA cycle or pyruvate metabolism, or (2) conversion of threonine to 2-oxobutanoate, a precursor for the valine, leucine, isoleucine biosynthesis pathway which is a precursor pathway to the TCA cycle. In addition, several genes that regulate the TCA cycle were upregulated as well, such as PDHB, DLAT, PCK2, and OGDH.

In addition to genes related to amino acid metabolism and the TCA cycle being upregulated in amino acid-supplemented cultures, several key genes involved in fatty acid metabolism and degradation were upregulated as well, which supported the metabolomics MVDA findings. Important genes to note due to their high positive fold change values and low p values are ECHS1 and

Table 4 KEGG pathway analysis summary

Gene	Description	Fold change (p value)
Gly, Ser, Thr metabolism		
Sds1	N-Sulphoglucosamine sulphohydrolase 1	+1.89 (7.31E - 3)
Tyrosine metabolism		
Fadh1	Acylpyruvase FAHD1, mitochondrial	+9 (9.47E - 3)
Fah	Fumarylacetoacetase isoform X3	+5.54 (8.95E - 34)
Lysine degradation		
Echs1 ^{a,b}	Enoyl-CoA hydratase 1	+6.08 (8.22E - 9)
Ogdh ^c	2-oxoglutarate dehydrogenase	+3.02 (1.48E - 56)
Hadh ^{a,b}	Hydroxyacyl-coenzyme A dehydrogenase	+1.96 (6.31E - 13)
Gcdh ^a	Glutaryl-CoA dehydrogenase	+1.28 (1.11E - 1)
TCA cycle		
Pck2	Phosphoenolpyruvate carboxykinase 2	+3.09 (3.84E - 10)
Pdhb	Pyruvate dehydrogenase	+1.75 (1.43E - 6)
Dlat	Dihydrolipoamide S-acetyltransferase	+1.13 (1.91E - 1)
Fatty acid metabolism/degradation		
Oxsm ^b	3-oxoacyl-ACP synthase	+3.44 (1.28E - 5)
Acadsb ^a	acyl-CoA dehydrogenase	+1.44 (1.02E - 3)

^a Genes also involved in fatty acid degradation pathway

^b Genes also involved in fatty acid metabolism

^c Genes also involved in TCA cycle. Fold change values included here only represent differentially expressed genes in V205, the highest titer condition overall

OXSM, which metabolize fatty acids (i.e., C12 DCA and C10 DCA, specifically) into acetyl-CoA in mitochondria before entering the TCA cycle and metabolize acetyl-CoA into acyl carrier proteins, respectively. Overall, many of the upregulated genes promoted direct or indirect activation of the TCA cycle at day 10, which could provide enhanced energy consumption during the stationary phase in cell culture and explains the increases in cell growth and productivity.

Integrated data analysis and feed media design

The integration of MVDA analysis and transcriptomics is a complex process as results described in previous sections are not directly comparable. Therefore, findings from each technique were integrated into a table which classifies significant consistencies in analysis results so that reasonable conclusions can be drawn. Integrated analysis of metabolomic and transcriptomic data for key metabolic pathways revealed that abundance of the metabolites and amino acids are virtually consistent with changes in transcript levels of catalytic enzyme genes within the same pathway. For example, the top identified metabolite that positively correlated with titer, C10 DCA Sebacic Acid, was found to have a positive correlation with alanine concentration, whereas aspartate and glutamate levels were found to have highly negative correlations with the same metabolite. This could be interpreted as the metabolic shift that drives the formation

of C10 DCA Sebacic Acid leads to alanine production, while aspartate and glutamate are being consumed at a high rate. Pathway enrichment analysis located C10 DCA Sebacic Acid in the metabolism of lipids pathway of CHO cells. Transcriptomic results supported these findings, as specific genes involved in the same fatty acid metabolism pathway are shown to be upregulated in the highest titer-yielding condition (see Additional file 1: Table S1). Thus, to enhance the productivity of the cells, the recommended course of action would be to increase the levels of amino acids that positively correlate with promoter-like metabolites yet are not considered metabolic waste products. Careful analysis was performed for each of the top ten metabolites until a final recommendation for a new feed medium formulation that increases (or inhibits) cell productivity was made. For the positive, high extreme condition (based upon Table 3) metabolites that were able to increase the production of *only* metabolites which correlate positively with titer or decrease the production of *only* metabolites which correlate negatively with titer were selected for further review. In this instance, it was found that His, Glu, Leu, Met, and Tyr were correlated with positive changes with titer whereas Asp, Cys, Ile, Ser, Thr, and Trp were correlated with negative changes in titer. On the other hand, some metabolites such as Ala were found to correlate with both a positive change in titer (see C10 Sebacic Acid, C12 DCA, C14 DCA, etc.), but Ala also correlated with an increase in a metabolite

that was found to be negatively correlated with titer (see Amino Benzoic Acid). Thus, with conflicting correlations, further changes in the alanine concentration were not recommended. A validation bioreactor run incorporating the new feed tested the hypothesis.

Validation run results

The aim of this experiment was to determine whether the productivity of the cells could be improved further by adjusting feed media formulations based on the MVDA model output. In the new feed conditions, amino acids that were found to correlate most strongly with the putative inhibitors and promoters were either reduced or increased in concentration twofold. This is the first iteration of a process that can be used to optimize feed formulations for enhanced productivity. The feed conditions listed in Fig. 6 were used to test the MVDA results. It is important to note that the “control” condition contained the same amino acid supplement that was found to increase the titer during the initial run. Thus, this validation experiment was used to instigate an additional increase in titer over what these four amino acids were able to provide.

The “high extreme” was determined by directly using the amino acid recommendations from the MVDA analysis, which were to increase concentrations of histidine, glutamine, leucine, methionine, and tyrosine. The “low extreme” was set up by implementing amino acids’ concentrations that should promote the formation of inhibitors (i.e., the opposite recommendations from the MVDA results, which were aimed at limiting the concentration of inhibitors being formed), and decreasing concentrations of aspartate, cysteine, isoleucine, serine, threonine, and tryptophan in the original feed were recommended for this condition.

The validation experimental results showed that the titer of the “high extreme” which did in fact produce

higher titer (8.2%) compared to the control condition. On the other hand, the “low extreme” condition which was anticipated to yield poor results did in fact yield a lower titer (−7.1%) compared to the control condition. Figure 6 summarizes the titer results from this experiment. The titer measurements have been normalized to protect proprietary information. The p values were used to validate the statistical significance of the titer values of each condition. The p values were calculated based on 4 titer measurements (from 2 experimental runs for each condition and 2 measurement techniques for titer) using a normal distribution, two tails, the standard deviation of the four measurements, the mean value of the titer, and a test statistic (based on the entire population of titer values from all three conditions). The test statistic provides an avenue to compare all three of the conditions and incorporate the differences among the conditions.

Discussion

The metabolites detected from the analysis include some that have been previously identified as promoters or inhibitors, such as 2-hydroxybutyrate, which had also been identified as an inhibitor in another study (Mulukutla et al. 2017), and homocysteine, which was previously identified as an inhibitor (Mulukutla et al. 2017), but is described above as a promoter in this study. The rationale for why Table 2 shows conflicting results with previously identified inhibitors is twofold. First, inhibitors and promoters may be cell line or process specific especially if a cell line has been genetically modified to improve titer or prevent the accumulation of certain metabolites or if a specialized process is being used to deter lactate formation. Second, the above analysis involves developing an overall correlation between all the metabolites and the titer, and thus, this analysis considers potential interactions between multiple metabolites that may cause a metabolite to have inhibitory behavior

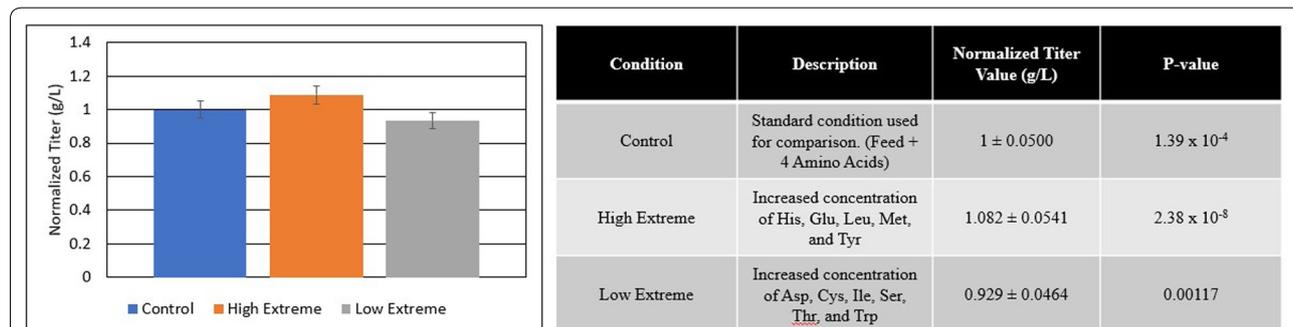


Fig. 6 Validation results: The validation experiment demonstrated that the “high extreme” and “low extreme” conditions were able to provide a relative increase or decrease (respectively) in titer. The control condition that was used includes the four additional amino acids that were previously identified. Thus, an overall increase in titer can be observed based upon a previously optimized feeding condition

when in the presence of other metabolites. Therefore, a questionable metabolite like homocysteine may appear to have inhibitory behavior when it is analyzed alone in spiking studies, but it may actually shift to behaving as a promoter while in the presence of other metabolites. These inconsistencies could be further explored through additional bioreactor testing involving more feed conditions which will lead to improvement of the MVDA model's predictive ability.

For the other metabolites presented in this analysis, specifically the dicarboxylic acid (DCA) components, it is hypothesized that these metabolites enhance titer. In literature, some success in enhancing titer has been shown for adding carboxylic acids to CHO cell cultures (Camire et al. 2017). In addition, other metabolites that were identified in this study alone include amino benzoic acid and keto-isoleucine which have a negative impact on titer whereas prolyl-hydroxyproline has a positive impact on titer. This study presents a method to analyze the overall effect of a metabolite in the presence of all other metabolites at concentrations that are present during an actual cultivation.

The proposed methodology in this paper was effective in finding metabolites that correlated positively or negatively with titer. Furthermore, application of process control changes via feed media optimization further increased titer. The feed media were optimized by increasing amino acids that were found to be correlated with the formation of positively correlating metabolites (with titer) or the prevention of the formation of negatively correlating metabolites (with titer). Further iteration of this process could lead to substantial titer increases after few experimental bioreactor runs.

MVDA analysis revealed the inhibitor and promoter metabolites were easily identifiable. The first four promoters identified were all DCA compounds, which are involved in fatty acid metabolism and can be correlated with energy production. Thus, with more of these DCA compounds being formed, it may be assumed that the cells are producing ample amounts of energy and must, therefore, be more metabolically productive. Moreover, the "High Extreme" validation condition aimed to enhance the production of these compounds yielded a higher titer, which supports the hypothesis that these compounds may be promoters. To our knowledge, DCA compounds have not been noted in previous works as promoters, whereas 2-hydroxybutyrate is a previously identified inhibitor (Mulukutla et al. 2017).

Alleged inhibitors and promoters from this study did correlate with the results from the transcriptomic analysis, leading to further confidence in the model. Genes that were found to be differentially upregulated in the higher producing condition were shown through pathway

analysis to be involved in amino acid-related metabolic pathways as well as energy production pathways. In addition, the identified upregulated genes translate to enzymes involved in the biosynthesis or degradation of fatty acids, which were the top metabolites identified from the MVDA model to influence titer.

This methodology could be useful for various other bioreactor experiments aimed at improving CHO cell productivity through feed media optimization. The model set-up is based on the data derived from the bioreactor experiment and is not based on the actual bioreactor conditions, and therefore this model could be applied to larger scale bioreactors and various cell lines. Also, as confidence in the model's ability to predict and identify titer promoter and inhibitors increases, the need for validation via transcriptomics analysis will no longer be necessary. In this way, the model is very versatile and may be able to produce positive results at an industrial scale.

Conclusion

The ability to improve titer through simple manipulations such as feed adjustments is a massive area of research today. The work presented in this paper illustrates how bigdata analytics can analyze massive amounts of biological data. The analysis method can also be used to determine which metabolites are acting as inhibitors or promoters and which metabolites are acting as putative inhibitors and promoters of cell growth and productivity. This methodology has been used in practice to increase titer through simple feed manipulations and with few bioreactor experiments. Thus, the need for full factorial designs of experiments can be reduced and time and money can be saved. In conclusion, the bigdata analytics provides a new avenue to explore in the effort to push productivity limits higher.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s40643-020-00318-6>.

Additional file 1: Table S1. Integrated data analysis results.

Acknowledgements

The authors are grateful for the financial support of this work provided by Bristol-Myers Squibb. Satorius Analytics provided the SIMCA license.

Authors' contributions

CM performed the multivariate data analysis including building all data-driven models and determining the correlations between amino acids and metabolites. CM was a major contributor to the writing of the manuscript. AP performed all the transcriptomics data analysis and was a major contributor to the writing of the manuscript. CM and AP contributed equally to the writing of the manuscript. AY provided insight from BMS and assisted in the writing of the manuscript. JX was in charge of defining the project, outlining the manuscript, and assisted in the writing of the manuscript. ZH assisted in the data analysis. JZ, KSM, SP, BW, MR, MB, and ZL were involved in the raw data analysis of the data that were inputted into data analysis tool and its generation at BMS. SY

was in charge of defining the project and outlining the manuscript. He contributed to the writing of the manuscript. All authors read and approved the final manuscript.

Funding

Bristol Meyers Squibb provided the funding for this study as it pertained to the understanding of proprietary information.

Availability of data and materials

The datasets generated and/or analyzed during the current study are not publicly available due the use of proprietary information but are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

Not applicable.

Author details

¹ Department of Chemical Engineering, University of Massachusetts Lowell, University Ave, Lowell, MA 01854, USA. ² Biologics Process Development, Bristol-Myers Squibb, 38 Jackson RD, Devens, MA 01434, USA. ³ Research and Development, Bristol-Myers Squibb, Princeton, MA, USA.

Received: 23 January 2020 Accepted: 25 May 2020

Published online: 01 June 2020

References

- Bradley SA, Ouyang A, Purdie J, Smitka TA, Wang T, Kaerner A (2010) Fermentanomics: monitoring mammalian cell cultures with NMR spectroscopy. *J Am Chem Soc* 132:9531–9533. <https://doi.org/10.1021/ja101962c>
- Bylesjö M et al (2006) OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification. *J Chemom* 20(8–10):341–351. <https://doi.org/10.1002/cem.1006>
- Camire Joseph, Kim Dongjoo, Kwon Soonjo (2017) Enhanced production of recombinant proteins by a small molecule protein synthesis enhancer in combination with an antioxidant in recombinant chinese hamster ovary cells. *Bioprocess Biosyst Eng* 40(7):1049–1056
- Chong WP, Goh LT, Reddy SG, Yusufi FN, Lee DY, Wong NS, Heng CK, Yap MG, Ho YS (2009) Metabolomics profiling of extracellular metabolites in recombinant Chinese Hamster Ovary fed-batch culture. *Rapid Commun Mass Spectrom* 23:3763–3771. <https://doi.org/10.1002/rcm.4328>
- Čuperlović-Culm M, Barnett DA, Culm AS, Chute I (2010) Cell culture metabolomics: applications and future directions. *Drug Discov Today* 15(15):610–621
- Dippel B (2011) <http://www.raman.de/htmlEN/home/advantageEn.html>. Accessed 25 Sept 2012
- Dunn Warwick B, Ellis David I (2005) Metabolomics: current analytical platforms and methodologies. *Trends Anal Chem* 24(4):285–294. <https://doi.org/10.1016/j.trac.2004.11.021>
- Fan Lianchun, Frye Christopher, Racher Andrew (2013) The use of glutamine synthetase as a selection marker: recent advances in Chinese hamster ovary cell line generation processes. *Pharm Bioprocess* 1:487–502. <https://doi.org/10.4155/pbp.13.56>
- Ferrara CT, Wang P, Neto EC, Stevens RD, Bain JR, Wenner BR, Attie AD (2008) Genetic networks of liver metabolism revealed by integration of metabolic and transcriptional profiling. *PLoS Genet* 4(3):e1000034. <https://doi.org/10.1371/journal.pgen.1000034>
- Gagnon M et al (2011) High-end PH-controlled delivery of glucose effectively suppresses lactate accumulation in CHO fed-batch cultures. *Biotechnol Bioeng* 108(6):1328–1337. <https://doi.org/10.1002/bit.23072>
- Griffin JL (2003) Metabonomics: NMR spectroscopy and pattern recognition analysis of body fluids and tissues for characterisation of xenobiotic toxicity and disease diagnosis. *Curr Opin Chem Biol* 7:648–654. <https://doi.org/10.1016/j.cbpa.2003.08.008>
- Hnatyshyn S, Shipkova P, Sanders M (2013) Expedient data mining for nontargeted high-resolution LC-MS profiles of biological samples. *Bioanalysis* 5(10):1195–1210. <https://doi.org/10.4155/bio.13.86>
- Kanehisa M, Goto S (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* 28:27–30
- Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M (2016) KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res* 44:D457–D462
- Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K (2017) KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res* 45:D353–D361
- Khoo SHG, Al-Rubeai M (2007) Metabolomics as a complementary tool in cell culture. *Biotechnol App Biochem* 47:71–84. <https://doi.org/10.1042/BA20060221>
- Kucukural A, Yukselen O, Ozata DM, Moore MJ, Garber M (2019) DEBrowser: interactive differential expression analysis and visualization tool for count data. *BMC Genom* 20:6
- Lei Z, Huhman DV, Sumner LW (2011) Mass spectrometry strategies in metabolomics. *J Biol Chem* 286:25435–25442
- Love MI, Huber W, Anders SJGB (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15(12):550. <https://doi.org/10.1186/s13059-014-0550-8>
- Mohamad-Saberi SE, Hashim YZ, Mel M, Amid A, Ahmad-Raus R, Packeer-Mohamed V (2010) Metabolomics profiling of extracellular metabolites in CHO-K1 cells cultured in different types of growth media. *Cytotechnology* 28(9):476–484. <https://doi.org/10.1007/s10616-012-9508-4>
- Mulukutla BC, Khan S, Lange A, Hu WS (2010) Glucose metabolism in mammalian cell culture: new insights for tweaking vintage pathways. *Trends Biotechnol* 28(9):476–484
- Mulukutla BC et al (2017) Identification and control of novel growth inhibitors in fed-batch cultures of Chinese Hamster Ovary Cells. *Biotechnol Bioeng* 114(8):1779–1790. <https://doi.org/10.1002/bit.26313>
- Powers DN, Wang Y, Fratz-Berilla EJ et al (2019) Real-time quantification and supplementation of bioreactor amino acids to prolong culture time and maintain antibody product quality. *Biotechnol Progress*. <https://doi.org/10.1002/btpr.2894>
- Sha S, Bhatia H, Yoon S (2018) An RNA-seq based transcriptomic investigation into the productivity and growth variants with Chinese hamster ovary cells. *J Biotechnol* 271:37–46
- Svante W, Michael S (1977) SIMCA: a method for analyzing chemical data in terms of similarity and analogy. *ACS Symp Ser* 52:243–282
- Templeton N, Smith KD, McAtee-Pereira AG, Dorai H, Betenbaugh MJ, Lang SE, Young JD (2017) Application of (13)C flux analysis to identify high-productivity CHO metabolic phenotypes. *Metab Eng* 43(Pt B):218–225
- Wu Z, Li D, Meng J, Wang H (2010) Introduction to SIMCA-P and its application. In: Eposito Vinzi V, Chin W, Henseler J, Wang H (eds) *Handbook of Partial Least Squares*. Springer Handbooks of Computational Statistics, Springer
- Xu J, Rehmann MS, Xu X, Huang C, Tian J, Qian NX, Li ZJ (2018) Improving titer while maintaining quality of final formulated drug substance via optimization of CHO cell culture conditions in low-iron chemically defined media. *mAbs* 10(3):488–499
- Zhang A, Sun H, Xu H, Qiu S, Wang X (2013) Cell metabolomics. *OMICS* 17(10):495–501. <https://doi.org/10.1089/omi.2012.0090>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.