Titre: Title:	Robust Process Monitoring and Control of Critical Variables in the Production of SARS-CoV-2 Spike Protein
Auteur: Author:	Juan Sebastian Reyes Davila
Date:	2024
Type:	Mémoire ou thèse / Dissertation or Thesis
Référence: Citation:	Reyes Davila, J. S. (2024). Robust Process Monitoring and Control of Critical Variables in the Production of SARS-CoV-2 Spike Protein [Thèse de doctorat, Polytechnique Montréal]. PolyPublie. https://publications.polymtl.ca/59644/

URL de PolyPublie: PolyPublie URL:	https://publications.polymtl.ca/59644/
Directeurs de recherche: Advisors:	Olivier Henry, & Yves Durocher
Programme: Program:	Génie chimique

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

Robust Process Monitoring and Control of Critical Variables in the Production of SARS-CoV-2 Spike Protein

JUAN SEBASTIAN REYES DAVILA

Département de génie chimique

Thèse présentée en vue de l'obtention du diplôme de *Philosophiæ Doctor*

Génie chimique

Octobre 2024

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

Cette thèse intitulée :

Robust Process Monitoring and Control of Critical Variables in the Production of SARS-CoV-2 Spike Protein

présentée par **Juan Sebastian REYES DAVILA**

en vue de l'obtention du diplôme de *Philosophiæ Doctor*

a été dûment acceptée par le jury d'examen constitué de :

Michel PERRIER, président
Olivier HENRY, membre et directeur de recherche
Yves DUROCHER, membre et codirecteur de recherche
Gregory De CRESCENZO, membre
Véronique CHOTTEAU, membre externe

DEDICATION

“Many years later, in front of the thesis committee, Sebastian Reyes Dávila was to remember that distant afternoon when his grandmother took him to discover the sea.”

In loving memory of Carmen Mesa.

ACKNOWLEDGEMENTS

First, I would like to thank my supervisor Dr. Henry whom countless hours discussing objectives, goals and looking over research results were priceless in completing this phase of my life. I would also like to thank Dr. Pham for helping me polish and perfect every detail of a manuscript and presentation to make sure high scientific standards are met, with out a doubt this detail centered look helped me reach this point. I would also like to thank my dad who inculcated in me a great deal of curiosity about the universe and the importance science has on understanding it. I would like to thank my mom who always made sure to teach me how to take deep breaths and control my emotions whenever situations become difficult and how to hold on to persevere for a little longer. To both I owe my life and its thanks to their sacrifice I can write these present words. I would like to thank my first teacher, who taught me to read, write and arithmetic. I know my grandmother would have been happy reading these words and I am forever thankful to her memory. I would like to thank Lucas for the numerous hours spent discussing the best way to plot an experimental result. I would like to thank Raul for helping me structure and tune the AI applications to make them robust enough for implementation. I would like to also thank all the friends I made and kept from elementary school until graduate studies. Every interaction with them, every step, every laugh, was a grain of sand that added up. Lastly, I would like to thank the love of my life, my confidant, the person whom I have built a home away from home, Tanisha. She stood by me in the hard times and celebrated with me in the triumphs. Forever my successes will be in dedication to her, from now until death.

RÉSUMÉ

La pandémie de SARS-CoV-2 a représenté un défi pour l'industrie biopharmaceutique, nécessitant une production rapide de thérapeutiques tout en respectant des directives de fabrication strictes et des profils de sécurité. Dans ce contexte, les pools cellulaires se sont révélés prometteurs en raison de leur capacité à produire des composés thérapeutiques dans des délais réduits. Cependant, ces plateformes novatrices présentent des inconvénients. Par conséquent, une meilleure compréhension et optimisation sont nécessaires pour en faire une alternative viable. Cette thèse se concentre sur l'identification du suivi et du contrôle des variables critiques dans la production de la protéine de spicule du SARS-CoV-2, permettant ainsi le développement d'un processus de production robuste. Pour y parvenir, des techniques d'analyse de données multivariées (ACP et modélisation par apprentissage automatique) ont été appliquées pour comprendre les tendances globales au sein d'un ensemble de données de développement précoce du processus de pools cellulaires CHO produisant la protéine de spicule du SARS-CoV-2. À partir d'une solide compréhension qualitative, des modifications du processus ont été tentées pour optimiser la production de la protéine de spicule du SARS-CoV-2, ce qui a entraîné une plus grande longévité, une diminution de l'accumulation de lactate et une augmentation du rendement en titre. Enfin, dans le cadre de l'initiative PAT, un capteur logiciel multivarié est proposé pour le suivi de l'accumulation des déchets métaboliques, de la croissance cellulaire et de la production de la protéine de spicule du SARS-CoV-2. Cette approche de prédiction à un pas en avant des résultats de culture tout au long du processus de 17 jours d'alimentation en lots s'est avérée utile pour suivre des variables difficiles à déterminer, comme la protéine de spicule, qui dépendent d'analyses en fin de processus pour quantifier le rendement (gels semi-quantitatifs ou Elisa). Ce travail démontre en général l'importance de l'analyse de données multivariées et de l'apprentissage automatique dans le développement de plateformes de production de bio-pharmaceutiques, ainsi que dans le développement de nouvelles technologies de capteurs en phase avec les initiatives PAT.

ABSTRACT

The SARS-CoV-2 pandemic represented a challenge to the biopharmaceutical industry as fast paced production of therapeutics was needed while still maintaining the strict manufacturing guidelines and safety profiles. Within this context, cell pools have shown promise for their capacity to produce therapeutic compounds with reduced timelines. However, these novel platforms have their drawbacks. Consequently, increased understanding, and optimization is needed to make them a viable alternative. This thesis centers around identifying monitoring and controlling critical variables in the production of SARS-CoV-2 spike protein that allows for the development of a robust production process. To achieve this, multivariate data analysis techniques where applied (PCA and machine learning modeling) to gain an understanding of the global trends within an early process development dataset of CHO cell pools producing SARS-CoV-2 spike protein. From strong qualitative understanding, proposed process changes are tried in the optimization of SARS-CoV-2 spike protein production that resulted in increased longevity, decreased lactate accumulation and increased titer yield. Lastly, within the context of PAT initiative a multivariate soft sensor is proposed for the monitoring of metabolic waste accumulation, cell growth and SARS-CoV-2 spike protein production. This approach of one-step-ahead prediction of culture outcomes across the 17-day fed-batch process proved useful in tracking hard to determine variables, like the spike protein, which rely on end of process analytics to quantify yield (semi quantitative gels or Elisa). This overall work shows the important role multivariate data analysis and machine learning has in the development of biotherapeutic production platforms as well as developing novel sensing technologies in keeping with PAT initiatives.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

CHAPTER 1 INTRODUCTION

Research Problem

Currently mammalian cell lines like Chinese Hamster Ovary Cells (CHO) are the industry standard for antibody production because these organisms can be adapted to produce the protein of interest through various transient or stable gene expression strategies. Developing stable cell lines can often be a time-consuming process that requires numerous screening procedures to differentially select the best behaving cell population. This can clearly be an issue if the desired bio-therapeutic product is urgently needed as a response to a rapidly changing public health crisis, such as the one caused by the SARS-CoV-2 pandemic. Therefore, mammalian cell pools have been considered an alternative option to developing bio-therapeutic production processes that do not require several months of screening and selection experiments. Such cell pools, although not as homogenous in cell population as a cell line, can be employed to produce target proteins on a large scale. However, usually such cell pools are strongly affected by culture age, meaning with increasing age, target protein expression productivity is significantly diminished. Consequently, for cell pools, the production phase needs to start as soon as the cells have been adapted to suspension to maximize protein yield. It is thus clear that not only does the production phase need to start as soon as possible but it must also be highly optimized for the cellular pool to mitigate the inherent disadvantages that cell pool protein production has when compared to cell line protein production. The proposed work centers around the optimization of a CHO cell pool developed at the Human Health Therapeutics Research Center (NRC) in Montreal that can produce the SARS-CoV-2 spike protein, which has a

wide range of applications including but not limited to novel therapeutics, clinical diagnostics, and subunit vaccines. Importantly, the metabolic networks of mammalian cells are highly dependent on process conditions and un-optimized process parameters, or un-optimized feeding strategies can generally cause overflow metabolism in which inhibitory by products are accumulated within the reactor causing the cells to lose viability and productivity. To develop said strategies it is of paramount importance to identify how variables relate to each other in historical datasets and how protein production is impacted by key metrics for the newly developed CHO pools. Consequently, this work aims to create a framework in which historical bioprocessing datasets can be evaluated through multiway principal component analysis to understand relationship among variables. Consequently, protein production can be modeled and analyzed to identify possible strategies during the process development stage that can improve upon the yield of recombinant protein. Additionally, within the context of process analytical technologies for process monitoring of mammalian cell culture the development of a soft sensor capable of one step ahead prediction of metabolic activity (lactate accumulation, glucose consumption, ammonia accumulation) and recombinant protein production is detailed.

Hypothesis and Objectives

The overall objective of this work is to identify, monitor and control critical variables in the production of SARS-CoV-2 spike protein that allow for the development of a robust production process. To attain this objective the following specific objectives were defined:

1. Multivariate analysis of historical bioprocessing records of SARS-CoV-2spike protein production to understand global trends within the production platforms.

2. Modeling and interpretation of endpoint protein production with respect to key metrics utilizing nonlinear regression techniques.
3. Transfer historical process to a new bioreactor system and improve yields by optimizing aeration, feeding and dosing strategies.
4. Develop a multivariate soft sensor capable of one-step-ahead predictions of mammalian cell growth, lactate accumulation, ammonia accumulation, glucose consumption and titer.

Thesis Structure

Following this introduction, chapter two provides a detailed literature review regarding mammalian cell metabolism, cell pools, inducible systems and the link between nutrients waste and protein production. Quality by design strategies for the purpose of identifying design spaces (processing conditions) are detailed. Additionally, the process analytical technologies initiative is discussed as a framework to allow for improved monitoring of biotherapeutic processes. Industry standard and state of the art sensor technologies for these tasks are reviewed as well as the development of soft sensors which arise from the deconvolution of numerous data sources that can then be indirectly related to hard to determine variables. Chapter three covers the article titled “Multivariate data analysis (MVDA) of factors affecting the growth and productivity of stable CHO cell pools expressing SARS-CoV-2 spike protein in early process development” in which multivariate analysis of a historical bioprocessing record is detailed as well as the modeling of endpoint titers for process understanding. Chapter four encompasses the article titled “CHO Stable pool fed-batch process development of SARS-CoV-2 spike protein

production: impact of aeration conditions and feeding strategies" in which optimization of aeration strategies coupled with optimization of feeding regimen led to an improvement of spike protein production when compared to historical averages as well as experimental controls. Chapter 5 titled "A recurrent neural network for soft sensor development in CHO cell pools in fed-batch process for SARS-CoV-2 spike protein production" details the development of a data driven soft sensors capable of realizing one step ahead predictions of metabolic variables (lactate accumulation, ammonia accumulation, glucose consumption) and protein yield by relying on easily accessible online process data (oxygen sparged, carbon dioxide sparged, pH, temperature, dissolved oxygen level, integral of dissolved oxygen, base addition). Chapter six provides a general discussion of the results obtained in each chapter and the interrelation among them as well as future work that can be undertaken. Finally, chapter seven presents the main conclusions of the thesis and its accomplishments.

CHAPTER 2 LITERATURE REVIEW

Biopharmaceutical Industry

The worldwide biotechnology market in 2023 was estimated to be worth near 1.3 trillion American dollars [1]. From this market cap, an important portion was estimated to belong exclusively to the biopharmaceutical industry [2-4]. This fast-growing segment of the global biotechnology market is expected to be worth close to half a trillion dollars by 2025 alone [5]. Within this market, the important value-added products include, growth and coagulation factors, hormones, monoclonal antibodies, vaccines, among others [4].

Interestingly, monoclonal antibodies represent the most dominant segment of the biopharmaceutical market, given that they are used in the treatment of various chronic illnesses, like cancer and auto immune diseases [6]. With this in mind, increased R&D output has the potential to enlarge the market share of monoclonal antibody (mAb) manufacturing [4,6]. It must be noted that while global bio-pharmaceutical industry comes of age, patents of FDA approved biologics will become void. Consequently, non-brand companies can begin to produce generic versions of said biotherapeutics in the form of biosimilars. Such biosimilars are analogous to generic active compounds in the pharmaceutical industry [4,7].

It must be recognized that the SARS-CoV-2 pandemic has acted as a stimulant for advances in the biopharmaceutical industry given that a large number of biopharmaceutical compounds have been generated in order to tackle the global pandemic [4]. Importantly, within these products, monoclonal antibody treatments and novel vaccine platforms are included [8-11].

Mammalian cells, which is the organism of choice in the manufacturing of antibodies and other complex molecules, have historically yielded low protein titers. Mammalian cell culture has added manufacturing complexity when compared to microbial organism given that these cells are shear sensitive to both sparging and mixing, Additionally, complex media additives need to be added to demonstrate optimal growth [8]. Importantly, thanks to advances in process development as well as cell line engineering, antibody production has increased to yields as high as 10 g/L–15 g/L in fed-batch mode which represents orders of magnitude increase to a few decades ago [8,12,13]. As a direct consequence of expanded production capacity, 570 therapeutic mAbs have been tested for phase I-III clinical trials by bio-pharmaceutical organizations [9]. Of the total mAbs evaluated, 79 antibodies have been approved for commercial use by the Food and Drug Administration (FDA) [9]. Given that in the time between 2008 and 2021, 60% (48 out of 79) of the total approved mAbs have been created. Thus, it can be argued that increased understanding of mammalian cells as biopharmaceutical production platforms has directly increased mAb manufacturing capacity and capability. This is especially true when one considers that the first FDA approved antibody for human use was released in 1986 (IgG2a CD3) [9].

Given this sustained industry growth, the FDA established Quality by Design (QbD) and Process Analytical Technologies (PAT) initiatives to guide the biopharmaceutical industry towards better efficiency while maintaining maximum process safety within a timely manner. Additionally, because lot-to-lot variation routinely indicated that standard procedures were not as robust as previously thought [14-17]. The FDA created an initiative, titled “current Good Manufacturing Practices (cGMP) for the 21st century” [16]. This initiative placed emphasis in a Quality by Design approach rather than counting solely on

quality control through batch testing [4]. Quality by Design is a scientific, risk-based, approach that requires defining and identifying Critical Quality Attributes (CQA) of a therapeutic compound as well as characterizing a relevant design space [4].

Mammalian Cell and Metabolism

Mammalian cell metabolism during a culture process changes depending on substrate availability and other environmental parameters. In brief, mammalian cells use carbon and/or nitrogen as energy sources. As such, mammalian cell metabolism can be either oxidative in nature or glycolytic [18-20]. When acting via the glycolytic pathway, glucose is exhausted at high rates almost exclusively producing lactate as a by-product. Importantly, here only two adenosine triphosphates (ATP) are produced (see Figure 2.1). Lactate is generated from the conversion of pyruvate to lactate thanks to the lactate dehydrogenase (LDH) enzyme [18-20]. This transformation of pyruvate into lactate (through NADH oxidation) restricts the total oxidation potential of glucose to carbon-dioxide and water, through the oxidative pathway [18-20]. Consequently, this lactate accumulation creates a net carbon flux removed from the Krebs cycle also known as the tricarboxylic acid cycle (TCA). This flux diminishes energy generation and in lieu permits for the energy source to be employed in biomass formation. This phenomenon is routinely noticed in cancerous cells; therefore it's designated as Warburg effect [19,21,22]. In cancer cells, because of their unique characteristic of increased undifferentiated growth, the magnitude of the glucose uptake rate is usually orders of magnitude bigger when paralleled to noncancerous cells [4,21]. Since the mammalian cell lines used in bioprocessing are derived from immortalized

cell lines, it is rational to observe similarities with neoplastic cells regarding their metabolic behavior [4].

Two ATPs are generated by the main glycolytic pathway which pales in comparison to the thirty-six ATPs that can be created by the oxidative pathway under aerobic conditions [18-20]. As noted previously, the glycolytic pathway yields low energy efficiency. Consequently, mammalian cells in bioreactors must also employ the oxidative pathway for their net energy requirements. Importantly, the pyruvate that is created at the completion of the glycolytic pathway, see Figure 2.1, is the primary substrate into the TCA cycle. Such is the way that the two central pathways are intimately linked. However, it must be mentioned that amino acid catabolism can be another relevant substrate source that enters the Kreb cycle [18-20]. As an ilustration, glutamine is catabolized as an energy source, this catalytic process generates glutamate and ammonia [18-20]. Changes in metabolic activity from glycolytic dominating to TCA cycle dominating can fluctuate during a manufacturing process and can be guided by manipulating process conditions. As an example, mammalian cells cultivated in environments with low glucose can upregulate the oxidative pathway and, consequently, maximize energy (ATP) synthesis in the absence of readily available glucose [4,19].

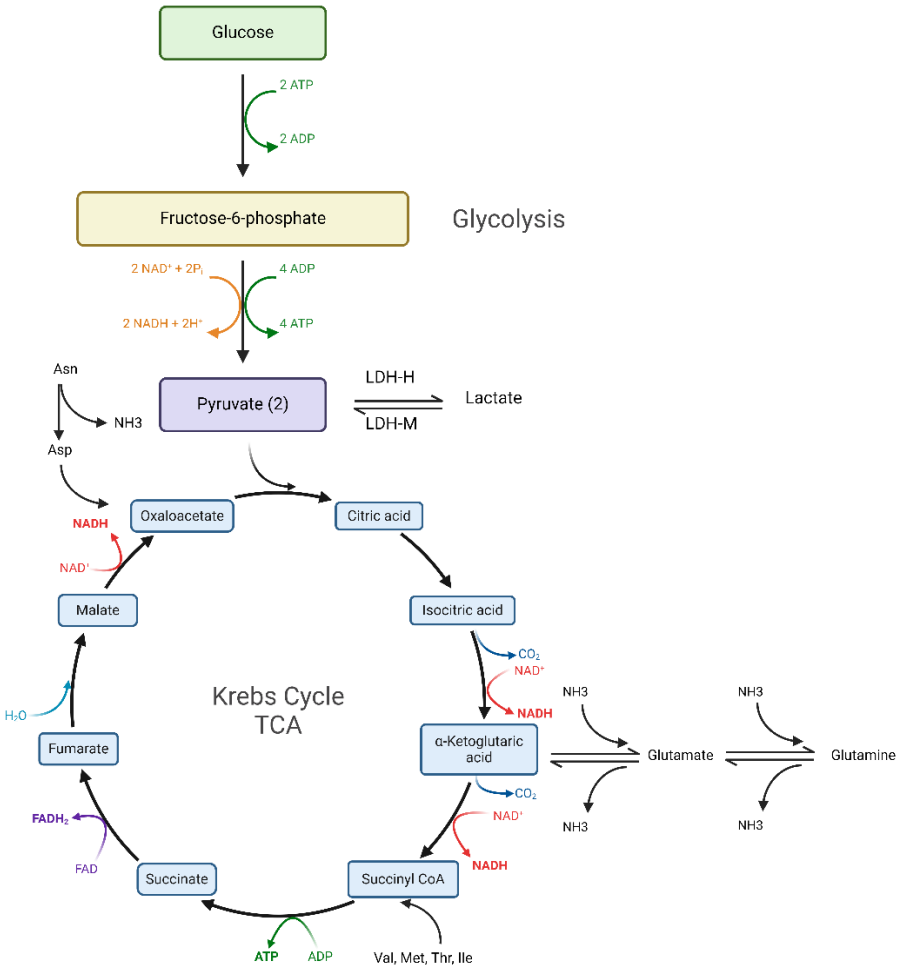


Figure 2.1. Diagram of the main glycolytic pathway and Kerb cycle.

Mammalian cell metabolism throughout a cell culture process may have sub-optimal metabolic behavior since the nutrient composition of basal media and supplemental feeds can lead to the buildup of toxic waste or nutrient depletion [18-20]. Consequently, this causes substrates to not be fully utilized in the generation of recombinant proteins or cell growth. As an example, 35-70% of the consumed glucose by cells is redirected into the generation of by-products [19]. This stark inefficiency hints at the presence of bottlenecks in the relevant metabolic pathways in addition to un-even flux distributions. Consequently, the metabolic wastefulness can generate waste buildup that directly harms culture outcomes by causing

decline in biomass growth, protein yield, and by adversely altering the product glycosylation profile [18-20]. Importantly, the accumulation of lactate is the primary toxic metabolic by product that routinely affects mammalian culture [18,19,23,24]. This is because it has been observed to restrain cell growth as well as promote early apoptosis. It has also been observed to reduce the protein productivity because lactate accumulation can increase in osmolality and change the pH of the medium [18]. In a Chinese Hamster Ovary (CHO) cell bioprocess run, two definite states with regard to lactate metabolism has been observed. A first phase characterized by lactate generation at the beginning of the process as a direct consequence of glucose uptake and consumption thanks to the glycolytic pathway that is concomitant to rapid biomass growth, and secondary lactate consumption phase that follows after the rapid cell growth [18,19,23,24]. Furthermore, an additional third phase has been noted which is distinguished by co-consumption of glucose and lactate [19]. Importantly, metabolic studies of mammalian cells have been able to establish a link between lactate consumption and increased titer [19,25]. It must be noted that the lactate absorption state is characteristically observed in cultures that have entered a stationary phase during the growth cycle. Therefore, lactate re-absorption served the purpose of reducing lactate accumulation in the medium and consequently limit the negative impact on culture outcomes [18,19,23,24]. This change between lactate phases is denominated lactate shift and it is thought to be originated by upregulation. This upregulation causes lactate to be transformed into pyruvate. Consequently, pyruvate is incorporated into the Kreb cycle. Alternatively, pyruvate can also be coupled to monocarboxylate transporters (MCT) which can enter or exit the cell in co-transport with H+ ions [18].

Apart from lactate other metabolic intermediates have also been known to accumulate in a cell culture run. These compounds have been identified as citrate, succinate, fumarate, and malate [19]. This accumulation occurs when the oxidative pathway is activated. Additionally, this phenomenon has been discerned to happen after the onset of feed addition. Therefore, hinting at the presence of metabolic bottlenecks [18-20]. Amino acids are known to play a relevant role in mammalian cell culture metabolism. Essential amino acids are supplied in both the basal medium and feed supplements (in the case of fed-batch processes). Alternatively, some amino acids can be generated de novo through biosynthetic pathways inside mammalian cells (i.e. nonessential amino acids). Amino acids can aid cell growth and are key in protein synthesis. At the same time, the catabolism of amino acids promotes the formation of Kreb cycle intermediates that can be used for energy production [18-20]. It must be noted however that once amino acids are supplied in excess to the cell culture, accumulation of the Krebs cycle intermediates, can cause accumulation of ammonium as a consequence of transamination and deamination reactions that produce ammonium ions [19]. Generally, ammonia is accumulated as a direct consequence of glutamine breakdown. Importantly, other amino acids like serine and threonine are also capable of producing ammonia thanks to transamination/deamination reactions [19]. It has been determined that buildup of ammonia has the potential to negatively impact both biomass growth and final protein titer and thus limiting this outcome is important [24,26,27]. One theory for this observation is that as the ammonia concentration increases in the medium, it alters the electrochemical gradient and thus acidifies intracellular compartments. Consequently, standard enzymatic activity is impaired, which then induces apoptosis [19,24,28]. Beyond ammonia and lactate accumulation varying amino acid concentrations at distinct stages of

the biomanufacturing culture run have also been positively correlated to cell growth inhibition and cellular death (apoptosis) [4]. For example, asparagine exhaustion has been observed to decrease cell growth. Conversely, the de novo production of alanine has been observed to increase ammonium accumulation as well as inhibit the TCA cycle [29]. Excess addition of lysine has been correlated with cell death. On the other hand, the catabolism of serine, methionine, leucine, threonine, tryptophan, tyrosine, glycine, and phenylalanine has been observed to generate intermediates that constrains cell growth [19,30,31]. Since the accumulation of intermediates is caused by non-optimally regulated pathways, the meticulous formulation of defined amino acid concentrations in cell culture basal media and feed additives can boost both cell growth and titer [19,31]. It has been observed that during the protein production phase a metabolic shift occurs such that the citric acid cycle is upregulated and thus cells are subjected to increased oxidative stress [19,32]. As a consequence of increased oxidative levels, glutathione is biosynthesized which then interacts with reactive oxygen species so as to diminish the toxic impact of said oxidative stress [19]. Due to this, glutathione has been observed to be and adequate marker of productivity given that it is strongly related to oxidative metabolism which itself is related to protein production states. [32,33]. Furthermore, it has been observed that the accumulation of phenylalanine-tyrosine derivatives, prompted by secondary offshoot routes due to the reduced expression of crucial enzymes in the primary breakdown pathway, can result in growth suppression. [34]. Interestingly, it has been observed that peak specific growth rate was associated with high lactate production and minimal TCA cycle activity [32,33]. Conversely, concomitant to the lactate switch peak growth rates were observed to be minimal while peak protein production was achieved. Suggesting that a highly oxidative

state of metabolism corresponded to peak antibody production while a highly glycolytic state corresponded to peak growth cycles [32,33].

Mammalian cell respiration is another relevant aspect of mammalian metabolism, since oxygen uptake rates (OUR) and carbon dioxide evolution rates (CER) are related to both cell growth and energy production. Because of this, the oxygen consumed by the cells during the bioprocessing run can be most comprehensively grasped as a substrate while the generated carbon dioxide is recognized as metabolic waste. Consequently, dissolved oxygen (DO) concentration must be constantly supervised and regulated as it plays a crucial role in sustaining cell viability. Crucially, findings indicate that DO can negatively affect glycosylation patterns, which are crucial in determining the proteins of interest pharmacodynamics [35]. It must also be maintained within a rational limit given that, high DO levels can bring about the generation of super-oxides or peroxides, that can be extremely detrimental to the cell membrane. Consequently, determining the optimal operating ranges for DO is instrumental for a successful biomanufacturing run [35]. On the other hand, dissolved CO_2 (dCO_2) is rapidly gaining acknowledgement as a critical process parameter [36,37]. This is in part because high dCO_2 values have been observed to negatively impact glycosylation profiles. Additionally, protein productivity and cell growth have also been determined to be significantly impacted, in a negative way, if dCO_2 values exceed 68 mmHg at bench scale and 179 mmHg at pilot scale [36]. This observation is generally thought to be caused by the harmful impact on internal pH and cellular metabolic processes that dCO_2 accumulation has on mammalian cells [36,37]. This effect can be thwarted by adding base into the system as to keep a constant pH. Nevertheless, this by itself cannot be the only solution given that a stepwise increases in osmolality (caused by repeated base

addition that counteracts CO_2 accumulation) can adversely influence cell culture outcomes as well [36,37]. One interesting observation in insect cultures has determined that at higher dCO_2 values decreased glutamine intakes are observed [38]. This is of note because despite reduced glutamine intake, glucose intake remained unchanged [38]. This observed phenomenon could suggest that ammonia and lactate accumulation were a consequence of non-glutamine amino acid catabolism caused by inefficiencies in the TCA cycle [38]. Consequently, dCO_2 measurements can be directly correlated with other metabolic fluxes to leverage a deeper understanding of the biomanufacturing run. Additionally, this knowledge tells us that low pH on its own does not generate the same level of impact on mammalian cell metabolism as when it is combined with high dCO_2, thus dissolved carbon dioxide is a significant factor to regulate independently, rather than solely controlling it indirectly through pH monitoring [4,38]. Sensors are regularly used to measure and regulate the previously indicated metabolic parameters. They are capable of directly assessing the concentration of the main accumulating byproducts like lactate and ammonia and detect changes in the glucose concentration that may be inhibitory to the culture outcomes [4]. Alternatively, metabolism can be indirectly measured through cellular respiration by observing changes in gas composition (CER, OUR). Through online assessment of these factors, it becomes possible to sensibly devise feeding strategies, process conditions and scaling methods [4]. This regular surveillance maintains an optimal environment in which the cells can grow and produce recombinant proteins. Continuous streams of data regarding medium composition or cellular metabolic activity can also be key when constructing dynamic feed on-demand strategies which are automatically triggered after it has been determined that important nutrients are becoming limiting. Rather than feeding

the cultures based on set pre-determined dates. Recent trends have evolved towards developing and optimizing soft sensors that can abstract information from various sensor sources rather than one single source [4]. This is important as it has been observed that OUR can continue to increase even after peak viable cell densities are obtained, indicating that a relationship exists between volumetric oxygen demand and volumetric production rates [39]. Specific oxygen consumption rates and specific protein production rates have also been observed to have a direct relationship, suggesting a close physiological connection among cellular respiration and product formation rates [40]. This link can be attributed to the fact that the metabolic activity of the cells has a vast effect on product formation. Given that respiratory activity correlates well to TCA fluxes in mammalian cells, it is reasonable to infer a comparable direct relationship between specific protein production and Kreb cycle activity. This was also found in a study where linear relationship between the energy production rate and OUR was observed [41,42]. Interestingly, it was also determined that energy production rate is in positive relation with recombinant protein production rate [43]. It was postulated that the specific OUR can be used to represent the activity of the TCA cycle and the energy metabolic state of the cells. Because of this, it was suggested that high values of specific OUR in the recombinant protein production phase indicates high specific ATP production rate through TCA cycle (if the Phosphate/Oxygen ratio is almost constant). This, in consequence, can lead to the increase in the specific recombinant protein production rate [41].

Studies have shown that, in general, larger cells consume oxygen at higher rates than smaller cells [44]. Additionally, observed dependence on protein content as a function of cell size has been determined [44]. Given this knowledge it is possible to hypothesize that as cell

size and protein content increases, it is expected that rate of oxygen consumption per cell also increases [44]. Indeed, linear correlations between cell volume and oxygen requirements have been observed [44]. Within a fed batch cell culture cycle, three distinct states can be discerned: a state of exponential growth, a stationary state, and a decline state. The shift from the growth stage to stationary stage predetermines the maximum viable cell density (VCD). On the other hand, the progression from stationary stage to decline stage inherently presets the lifespan of the culture [45]. Within the three characteristic cell cycles, it has also been found that it can be further broken up into a phase where the viable cell count increases exponentially while the volume per cell remains constant. A secondary phase where the cell size increase is nearly linear with respect to time, while the viable cell concentration stays approximately constant. Next there is a short stationary phase where the viable biomass concentration remains in near constant state while the total viability remains above 80% [45]. Lastly, a decline stage in which the viability decreases below 80%. It was found that average specific productivity is two-fold higher in the cells that are going through the cell size increase phase suggesting there exists a correlation between the volume of the cell and its mAb productivity. This was furthermore evident when the productivity per cell is plotted against the cellular volume, leading to a direct linear relationship between the average cell volume and specific mAb productivity [45]. Importantly, the cellular protein formation rate decreased by 25% while the mAb formation rate increased by 250% during the cell size increase[45]. It has been previously found that the cell size is the major factor of productivity rather than the cell cycle phase as it was originally thought [46]. Additionally, it has been shown that the increase in specific productivity is linearly proportional to the increase in cell volume, this is caused by several

factors such as an increased expression of genes related to the mTOR signaling pathway [47]. Similarly in another study it was found that when the average cell diameter increased from 15.7 to 18.5 µm during a fed-batch process concomitant increase in specific respiration rates and specific protein production rates was also observed [39]. Importantly, it has been demonstrated that a lactate shift from accumulation to consumption is correlated with increased mitochondrial activity [25]. With this in mind, it can be hypothesized that cell size can be a bioprocess marker which may be important in the development of feeding strategies [44]. Given that biomass continues to grow (cell size increase) after the exponential growth stage, an adequate amount of nutrients is required not only for maintenance and cell size increase but also for recombinant protein production reactions. Given this trend of analyzing the importance of sell size increase in cell cultures, some publications have suggested that determining the viable cell volume is more relevant for process development than viable cell count [48,49]. As a result, many online techniques, such as bio-capacitance and OUR, show stronger correlations with viable cell volume than with viable cell count. A study comparing OUR to optical density, cell cycle analysis, viable cell count, packed cell volume and cellular count as a means to determine the end of exponential growth in CHO cell cultures was realized [50]. It was determined that on-line OUR and offline determination of intracellular nucleotide ratios were the best methods for determining the end of exponential growth [50]. Given the ease of calculation, it was suggested that OUR could be used to determine important trigger points as it was reliable for estimating point of change in growth rate and cell physiology.

Cell Pools and Inducible Systems

Transient transfection in suspension adapted cells is rapid and often generates adequate amounts of protein. However, for projects that require large amounts of culture, plasmid DNA necessary for transfection may become a limitation [51]. Thus, as an alternative, stable gene expression from a recombinant cell line has been employed. It has several advantages since it reduces the amount of plasmid DNA needed as well as simplifying volumetric scale up [51]. Stable gene expression can be further subdivided into stable transfected cell pool and stable clonal cell lines. CHO (Chinese ovary cells) and HEK293 (human embryonic kidney cells) cell lines are both commonly employed in stable gene expression since they can grow to high cell densities in suspension, are adaptable to serum free media, are easily transfected and have appropriate glycosylation profiles. Importantly, these cells can be engineered to have repressor systems that allow for controlled induced expression [51,52].

Transient gene expression involves the transfection of cells with one or more expression vectors. Cells are subsequently maintained in culture for 1-10 weeks until the end of the production phase [51]. On the other hand, stable gene expression involves the transfection of cells with one or more expression vectors and a selection phase of up to 3-4 weeks [10,51,52]. The surviving recombinant cells can be maintained as a heterogeneous pool for recombinant protein production or can be subjected to single cell cloning in order to generate stable cell lines. This involves a stringent screening procedure in order to obtain superior clones that have increased protein productivity and stability. This screening procedure elongates the timeline before a production run can be realized as can be evidenced by Figure 2.2.

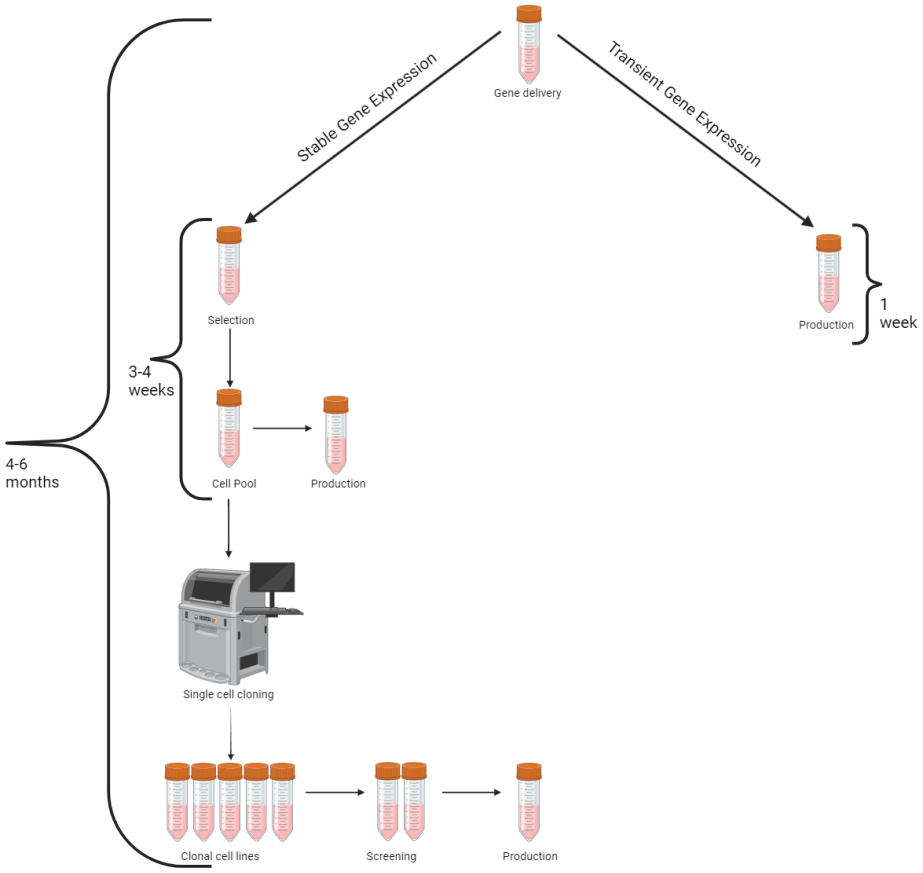


Figure 2.2. Stable and transient gene expression strategies (adapted from 49).

Because of the diverging timelines, transient transfection is typically realized in screening stages of a project while stable clonal cell lines are generated only when clinical candidates have been selected. An important draw back in transient transfection production processes is that there is no guarantee that the next transient transfection will give proteins the same post translational modifications, which is a big concern in terms of safety and regulatory agencies [10,51,52].

Stable transfection pools house mixed populations of cells which have different expression levels. Thus, due to this heterogeneity, the cell pools have lower expression levels when compared to the final selected clonal cell lines. A critical advantage of stable transfected pools is that they have reproducible productivities without the need of time-consuming

screening procedures. This has the drawback that transfected cell pool's productivity is affected by culture age. With increasing age, productivity is observed to decline [51,52]. Thus, once the cells are adapted to suspension growth, the production phase should start immediately to allow for the production of the highest possible yield. This is because cell pools are made up of-mixed populations that have different growth rates and expression levels. Importantly recent advances have shown that stable pools can have similar outcomes to stable cell lines further suggesting that this approach can be used in fast to market applications [8].

The two most important expression systems for the CHO cell line are the DHFR-based methotrexate (MTX) system and the glutamine synthetase (GS) based methionine sulfoximine (MSX) system. These expression systems are chosen on their ability to deliver high productivity along with acceptable product quality [53]. GS is a universal housekeeping enzyme that catalyzes the hydrolysis reaction:

Glutamate + ATP + NH_3 → Glutamine + ADP + Phosphate.

It is the only enzyme capable of synthesizing glutamine de novo and is also able to regulate toxic levels of ammonia [53]. This reaction can be inhibited by MSX through interaction with the GS enzyme. In liver cells, GS is responsible for the removal of ammonia while in plant and bacterial cells, it facilitates the fixation of ammonia so that glutamine can be used as a nitrogen source [53]. Glutamine plays an important role in the metabolism of CHO cells since it serves both as an energy source and as a nitrogen donor [54]. The GS system was originally designed to be used in cells that have no endogenous GS enzymes. However, since CHO cells are glutamine prototroph. As such, either an appropriate inhibitor (MSX) is added, genome editing is realized in order to knock out the cell's GS capability or both can be

done concomitantly in order to improve on its individual effects [54,55]. Because MSX is an analogue of glutamate it is capable of irreversibly inhibiting GS through binding and its consequent phosphorylation. Interestingly, MSX is also capable of inhibiting γ-glutamylcysteine synthetase, the first enzyme of both glutathione (GSH) synthetic pathways [53,55]. This enzyme plays an important role in maintaining an appropriate environment within the endoplasmic reticulum to allow for protein folding and secretion. Inhibition of this enzyme causes cells to overexpress GSH. Essentially, it is entirely possible, that the dual inhibition of GS and GSH not only selects cell lines by inducing glutamine auxotrophy (GS inhibition), but selects cells with a better inherent capacity for protein secretion (GSH inhibition) [53]. Additionally, over expression of the GS system reduces ammonia accumulation in the culture medium which also extends the overall life cycle of the cell culture [55]. The expression systems are compromised by vectors that contain the GS gene and efficient transcription promoters for the gene of interest [53]. In consequence, highly productive cell lines are selected by using an MSX concentration that inhibits growth and selecting the cell clones that manage to grow well. Thus, the cell line selected over expresses GS in order to generate sufficient glutamine for cellular metabolism and nucleotide synthesis. Since the GS gene and the gene of interest are genetically linked through the vector, selection for GS gene integration into a transcriptionally active locus results in co integration of the gene of interest in the same locus.

Inducible systems have also been widely used for mammalian cell platforms since they allow the expression of cytotoxic compounds, but more importantly, they allow for the separation between the growth phase and the production phase which can then be optimized

independently [56]. Numerous inducible systems for mammalian cells have been developed by implementing heavy metals, steroid hormones, insect hormones and heat shock [56]. However, all of these methods have been observed to have severe limitations such as leakiness and pleiotropic effects. Because of this, modifications have been realized to the Tet switch in order for it to be easily adapted for mammalian cells [57]. In terms of the cumate inducible system, three configurations have been developed. The repressor configuration uses CymR to repress transcription from a mammalian promoter. Once cumate is added to the mammalian cell culture, a conformational change is enacted on CymR which prevents it from binding to the DNA promoter sequence and thus it causes inhibition of the repression. In the activator configuration, CymR is fused to an activation domain (VP16) in order to form a chimeric trans-activator (cTA) that can consequently bind to promoters. The activation of this system is regulated by the addition of cumate [57]. With cumate, it is unable to bind, while when cumate is absent, the system is on the off state thus initiation of activation can occur. Lastly, in the reverse activator configuration, rCymR is fused with VP16 to form the chimeric molecule rcTA which interacts with operator binding sites to begin activation [57]. Once again cumate changes the conformation of the chimeric molecule. When no cumate is present, rcTA is not able to bind to the operator sites, while in the presence of cumate, rcTA is able to properly bind to the operator binding sites and thus initiate activation [57].

Process Analytical Technologies Initiative

In order to direct the expanding biomanufacturing industry toward enhanced production efficiencies, all the while maintaining adherence to process safety records, quality by design process analytical technologies has been established [14-17,58]. Lot-to-lot variance in product quality demonstrated that process robustness was not as high as it was once thought [14-17,58]. These lot variations have caused less product to be commercialized and thus increased drug manufacturing costs. Additionally, because of globalization, quality control guidelines have become harder to enforce [16]. To counteract the changing regulatory climate, the FDA created an initiative called current Good Manufacturing Practices (cGMP) for the 21st century [16]. These new quality guidelines placed more weight on Quality by Design approach (QbD) instead of postproduction quality control batch testing. QbD is a holistic, scientific, risk-based approach that defines and identifies the Critical Quality Attributes (CQA) of a biopharmaceutical. Additionally design spaces need to be adequately defined for the production process to be accepted [16]. By formulating biopharmaceutical compounds and production processes around the defined CQA, the manufacturer can then monitor its process platform to ensure consistent product quality [17]. The CQAs are established through in vitro and animal model studies that guide the characterization of the biopharmaceutical compound. Once the definitions of the CQA are established, the development of a specific manufacturing process (design space) whose outcome will be the desired products with the adequate attributes is needed [17]. Consequently, the design space is characterized early into the process development cycle. Here, temperature ranges, pH, feeding points, feed amount, mixing, dissolved oxygen concentration are characterized thoroughly [17]. The characterization process is realized with design experiments that

evaluate the impact that one or multiple variables, and changes to the respective variables, have upon the CQA that were previously defined. Consequently the goal is to find acceptable operating conditions in which the approved target attributes are achieved [17]. Importantly, biopharmaceutical compound production process utilizes multiple steps that may be parallel or serial to each other. Consequently, the development of a design space must necessarily take into account possible variations in process conditions as a form of risk mitigation [14-17,58].

Given that characterization of the design space and monitoring/controlling the CQA have become fundamental tenants of QbD, Process Analytical Technologies (PAT) have become important tools [17]. PAT can help analyze the CQA during the biomanufacturing process. Either in-line or at-line sensor data can then be analysed so as to make real time adjustment on process parameters [16,17,58]. In theory, these techniques can be deployed at every stage of the biomanufacturing process, from upstream cell culture to downstream purification and formulation [16,17]. Once the design space has been clearly defined the regulatory filing also requires the identification of acceptable ranges for the CQA. These CQA are then monitored to ensure the process does not deviate from the specified design space [16,17]. Consequently, it is good practice to expand the defined design space such that process flexibility is established within the regulatory filling. This is beneficial as process changes that may occur during the production process have already been taken into consideration during the approval process diminishing the need for additional testing for regulatory compliance. Conversely operating outside the defined design space because of process changes or raw materials variations requires additional fillings and approval from regulatory agencies to prove compliance [17]. This flexibility built in to the QbD guidelines

is important because it means that process improvement can be undertaken during the product cycle with out additional regulatory approval. Consequently, an industry as historically conservative as the biopharmaceutical industry is encouraged to improve upon and innovate on its process platforms by adopting technologies as they emerge and mature such that process monitoring is enhanced without additional regulatory burden. This conceptual arrangement can be visualized in Figure 2.3. Here the knowledge space is a non-design space that requires regulatory approval to operate in [4]. The design space is the approved process while the control space are the process configuration of the biomanufacturing process [4].

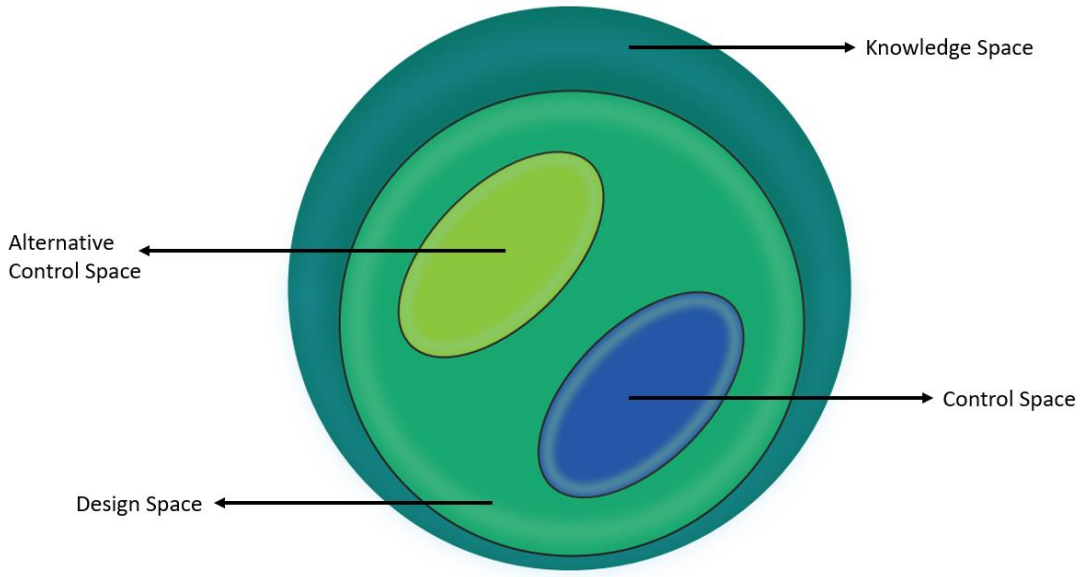


Figure 2.3. Design Space.

Bioreactor Modes of Operation

To produce biopharmaceutical compounds, there are different types of modes of operation like batch, fed-batch, concentrated fed-batch, and perfusion. In batch production, cells are cultured in a bioreactor with only basal medium and no further feed supplements [59,60]. Consequently, as cells proliferate within the vessel, nutrient concentrations decrease, while metabolic by-products and titer accumulate in the spent medium [59]. Despite the drawbacks of batch cultures, it's a fairly uncomplicated arrangement to implement that does not require complex control loops to add feed or remove by-products [61]. On the other hand, in fed-batch arrangement, substrates are added continuously or periodically (bolus) to the bioreactor in order to supplement media components that have been depleted [59,61-65]. For relevant substrates to be added correctly, the measurements of key metabolic products and substrate concentration are needed. This arrangement is extensively utilized in the biopharmaceutical industry given that it is optimal for the manufacturing of non growth associated products while also being a strong alternative to more complex continuous regimens, such as perfusion [59,61-65] One important caveat is that the process must be designed with increased reactor space in mind as to permit medium addition. For fed-batch operation, two relevant modes exist: constant feeding and constant substrate concentration [60]. With the latter strategy, a constant growth rate can be maintained during the start of the process. Therefore, the viable cell density in the bioreactor will increase exponentially as a function of time. Nevertheless, it also implies that the reactor must be supplied exponentially with substrates. Even with highly concentrated feed, this generates significant volume shifts within the bioreactor. As a direct consequence, maintaining the feed rate constant is the most relevantly used method even though the culture process may

become substrate limited without adequate monitoring [60]. Therefore, industry relies mostly on periodic feeding (bolus) thanks to the simplicity of the method and high efficiency in culture outcomes [60-63]. The batch and fed-batch modes are depicted in Figure 2.4.

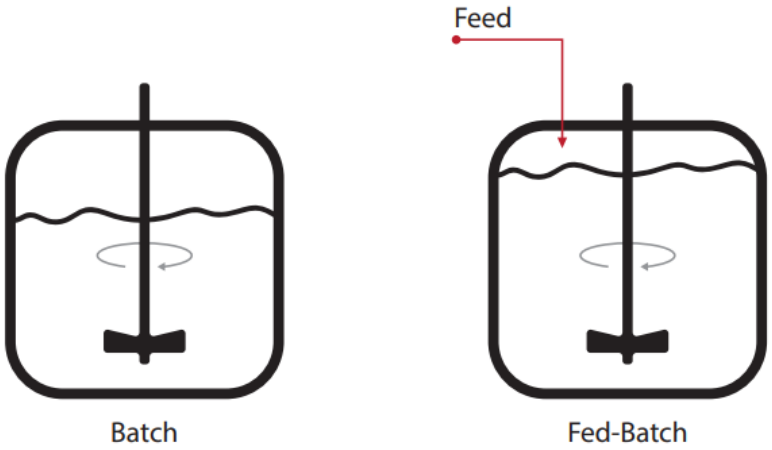


Figure 2.4. Batch And Fed Batch bioreactor operation

Given that bolus addition of concentrated feed and glucose can create large oscillations of nutrients within the reactor slow continuous additions of feed have been suggested in the literature [66,67]. Interestingly it was determined that continuous feeding eliminated glucose fluctuations while also allowing cellular expansion in mammalian cell spheroids [66]. Additionally in a CHO K1 cell line, continuous feeding demonstrated lower lactate and ammonia accumulation while also improving antibody yields by 10% [67]. Suggesting that avoiding nutrient oscillations associated with bolus addition of feed could be a way to reducing metabolic waste product accumulation.

Dissolved oxygen (DO) levels also need to be studied within a range to gauge the impact it may have on cell behaviour and even glycosylation patterns, as it has been found that specific productivity of recombinant protein is diminished in hypoxic (3%) and hyperoxic

(200%) conditions [68]. Hyperoxia (240%) in a U-937 cell line was shown to alter metabolic waste accumulation along with a strong growth inhibition effect [69]. Additionally a CHO cell line exposed to 380% DO saturation showed reduced growth and increased glutathione concentration when compared to a 20% DO control [70]. It has been suggested that the observed reduced growth may be caused by DNA strand breakage caused by the hyperoxic conditions [71]. Interestingly, for embryonic stem cells a study of DO ranges demonstrated that DO levels of 10-20% outperformed 5% and 30% due to a combination of factors [72]. Namely, the hypoxic conditions at 5% made it such that growth and metabolic behaviour was sub optimal while 30% DO required increased sparging that may have added unnecessary shear stress to the cells and as such created adverse conditions that worsened performance [72]. Demonstrating that in both stem cells and immortalized cell lines DO set points have to be adjusted so as to not be too low or too high. Which aeration strategies are deployed in order to control the level of oxygen in the reactor have also shown to be important [73]. It was determined that the utilization of F-68 was important when utilizing direct gas sparging strategies [73,74]. Interestingly it has also been found that high entrance velocities (above 60 m/s) in sparging strategies early in on the production (day 5) can cause irreparable harm to cells in a CHO cell line [75]. Alternatively, for a NS0 cell line the entrance velocity at which adverse outcomes were observed was above 20 m/s [76]. The most utilized strategies rely on diminishing the maximum amount of airflow used such that entrance velocity is diminished while still having sufficient air flow that allows for carbon dioxide striping. Additionally, increasing agitation rates to aid in the transfer of oxygen can also help in reducing the amount of oxygen flow that is needed to sustain a DO level [75]. Studies of stress activation in CHO cell lines demonstrated that hydrodynamic stress thresholds

and response mechanisms differ depending on the stress origin (agitation or sparging) [77]. For example, agitation related stress had a temporary decrease in viability after stress induction followed by a recovery in viability after the cells adapted to new conditions [77] Alternatively, sparging related stress demonstrated a monotonic decrease in viability. Following mRNA transcriptome analysis, it was determined that separate activation mechanisms exist depending on the stress source with almost no overlap [77]. In miniature bioreactor models it was found that active gas sparging that oscillated with respect to a DO control loop had decreased viability and increased lactate accumulation when compared to only overhead air flow. Suggesting that gas sparging has a significant impact on culture outcomes [78].

Monitoring Sensors in Bioprocessing

Optical chemosensors, denominated optodes, work because of the interaction between the analyte of interest and an indicator that is immobilized in a hydrophobic matrix at the sensor tip [79-82]. Changes in optical properties (photoluminescence intensity, absorption or reflection) are detected by a photodiode when the indicator is illuminated by a diode through optical fiber. The changes in the indicator are consequently correlated with the concentration of the analyte [79-82]. Such optical sensors can either be used in situ for stirred tank reactors through standard ports or, alternatively, they can also be minimized for low volume conditions to a patch form such that they may also be applied in deep well plates or shake flasks [79-82]. Importantly, patch form optical sensors have been applied broadly in single use bioreactors for DO, pH and carbon dioxide monitoring given their ease of sterilization through gamma radiation [79,80]. One significant advantages that optical

chemosensors have over their electrochemical standards is the fact that no direct contact between sample of interest and electronics is necessary [81]. In addition, no analyte is consumed during the measurement process and consequently the concentration of the analyte remains constant. As an example electrochemical DO sensors actively consume oxygen during the measurement process which is in stark contrast to the DO optodes which are largely non-invasive [83]. It was determined that optical and electrochemical DO sensors share a person correlation of 98.7% showing that both sensor types are in agreement with measured DO values [84]. By utilizing a similar mechanism, disposable inline optical sensors for glucose monitoring have been designed [85]. Here, a crosslinked glucose oxidase layer coats an optical DO sensor developed by PreSens. Changes in the oxygen concentration within the crosslinked enzyme is tracked, which should be inversely proportional to the glucose levels in the sample [85]. Importantly, the crosslinked layer could be tuned for different applications just by changing the enzyme that reacts with the analyte of interest [85].

Optical carbon dioxide sensors work through measuring changes in the optical properties of a pH bicarbonate system that reacts to the presence of carbon dioxide [79,80,83]. This is because a fluorometric or colorimetric pH sensitive indicator is incorporated within the buffer. Importantly, this pH buffer indicator system is isolated from the bioreactor media by a carbon dioxide permeable membrane [79,80,83]. Since the pH equilibrium of a bicarbonate system is quantitatively described by the Henderson-Hasselbalch equation, it is feasible to calculate the carbon dioxide partial pressure given a change in pH. The reaction time for these sensors is in the order of minutes given that it is highly dependent on diffusion phenomenon. Additionally, frequent pH buffer changes are required given that the optical

properties are dependent on the ionic strength of the buffer [79,80,83]. To counteract the slow response times, pH sensitive indicators may be immobilized within a hydrophobic membrane structure while the bicarbonate pH buffer is replaced with a lipophilic quaternary ammonium hydroxide [79-81]. Here changes in fluorescence can be measured and directly correlated to carbon dioxide concentrations [79-81]. Given the interrelationship, optical pH sensors operate akin to carbon dioxide optodes since the variable of interest is the change in pH. The sensors can either be constructed around absorbance or fluorescence dyes that are covalently immobilized in cellulose matrixes [79-81]. It must be noted that modern pH optodes largely utilize fluorescent dyes since absorbance-based measurements are not as sensitive as fluorescence-based measurements [79-81].

Spectroscopic sensors depend on the interplay between electromagnetic waves and the relevant variables [86,87]. These electromagnetic waves can interact through scattering, emission or absorption. As it discernible from Figure 2.5, the wavelengths employed exist within a wide range such as far infrared (FIR), mid infrared (MIR), near infrared (NIR) ultraviolet visible (UV/Vis) and Raman (400 nm-1000 nm) [86,87]. Since NIR spectra is less defined, NIR spectroscopy is generally used as qualitative monitoring of a bioprocess rather than quantitative applications that are routinely seen in MIR spectroscopy [80,86-89]. Importantly since the spectral data captured in NIR and MIR spectroscopy can be complex, multivariate analysis techniques are frequently employed. For qualitative analysis of the data and interpretation principal component analysis (PCA) is commonly used [88,90]. Alternatively, quantitative models require reference data sets that correlates spectral signals to relevant process variables. Such models may utilize artificial neural networks (ANN), partial least squares regression (PLS), support vector machine regression (SVM) or

multiple least squares regression (MLS) so as to correlate measured NIR/MIR spectra do the analytical measurements [88,90]. When contrasting NIR with MIR spectroscopy, it has been discovered that MIR exhibits significantly greater accuracy in predicting individual analytes, while NIR is proficient in estimating concentrations of multiple analytes in a sample [91]. For example, with NIR spectra and a PLS model seven different parameters in parallel where monitored, specifically: osmolality, glucose levels, protein yield, packed cell volume, integrated packed cell volume, VCD and IVCC [92]. The reason that NIR can monitor several parameters concomitantly (at a lower accuracy) is because absorption bands and absorption coefficients are much wider and lower (respectively) when compared to MIR. Consequently, in MIR spectroscopy glucose and lactate can be detected at much higher accuracies and at lower concentrations [91]. Low cost MIR probes have been developed to monitor viable cell density, lactate dehydrogenase, antibodies, lactate, glutamate concentration in an at-line system [93]. Noteworthy is the fact that antibody concentrations could only be adequately detected at concentrations above 0.4 mg/L while viability had an error of 9.8% at ranges between 20-95%. MIR has also been employed in the monitoring of product quality and impurity [94].

UV/Vis spectroscopy utilizes wavelengths in the range of 10-700 nm[86,90,95]. Absorbance of UV/Vis light is confined to molecular function groups denominated chromophores whose electrons are excited by UV/Vis interaction [86,90,95]. The correlation between concentration of chromophores, light path and light absorption can be described through Beer-Lambert law [86,90,95]. It must be noted that differentiation between proteins is difficult if relying on UV spectra alone and consequently, it requires prior purification steps [86,90,95]. Interestingly, UV spectroscopy has been previously applied for biomass estimation through turbidity-

cell concentration estimations [96]. It was observed that correlations are linear during the exponential phase of the cell culture but during the plateau and death phase of the cell culture turbidity estimation cannot adequately differentiate between viable and dead cells [95,97]. To counteract this an ANN have been used alongside UV spectroscopy so as to monitor offline glutamine, glutamate and viable cell concentrations [98].

Fluorescence spectroscopy is another relevant monitoring technology for the production of biotherapeutics. The mechanism of action relies on the fact that multiple biological compounds (amino acids, enzymes, cofactors, vitamins) within the bioreactor broth have fluorescent properties [99]. Historically, fluorescent sensors have been based on a single wavelength pair limiting process analysis to a single fluorophore [86,87]. Recently, 2D fluorescence spectroscopy has been created where several wavelengths are utilized so as to analyze several analytes during the bioprocessing run [100]. This novel approach has been used along side PCA and PLS modeling in CHO cell cultures so as to monitor viable cell counts and recombinant protein production [101,102]. A similar methodology has been applied in Baby Hamster Kidney cell culture processes to readily monitor viable cells and antibody yields [103]. When utilizing PCA with PLS of the 2D fluorescence spectra it was determined that differentiation between viable, dead and lysed cell populations could be monitored concomitantly during a fed-batch process [104]. This is of note as differentiating among cellular populations within a fed batch process can be a difficult task.

Raman spectroscopy is another important technology that has matured enough to be applied in bioprocessing. It relies on detecting the inelastic scattering of monochromatic light that happens when incident light interacts with the molecules within the bioreactor broth [79,86,90,95,105]. When the incident light (monochromatic) interacts with the vibrational

frequencies of the molecules of interest most of the light is scattered without suffering a change in frequency. That being said, a small fraction of the scattered light is shifted from its original frequency. This phenomenon is denominated Raman scattering [95,105]. The resulting shift in wavelength is dependent on the vibrating chemical bonds that caused the Raman scattering [95,105]. Consequently, by capturing the light that is a result of Raman scattering, information with respect to the rotational and vibrational characteristics of the molecules of interest can be obtained. Importantly, Raman spectra is insensitive to water which is an important feature in bioprocessing monitoring [79,86,90,95,105]. Careful consideration must be taken when choosing the appropriate monochromatic light frequency so as to minimize possible fluorescence interference for the purpose of bioprocess monitoring [79,86,90,95,105]. Raman spectra along with PLS modeling has successfully monitored in real time glucose, glutamine, lactate, ammonium, glutamate and viable cell density [106]. Previous work had centered on achieving the same monitoring applications in an off-line fashion [107]. The ability to monitor in real time metabolic indicators (glutamate, ammonia, glutamine, glucose, lactate concentrations) and cell growth in CHO cells in fed-batch cultures have been shown to be scale independent (3L and 15L) [108]. Further studies have shown that PLS modeling of Raman spectra can be utilized at developmental scales (3L and 200 L) as well as at a clinical manufacturing scale (2000L) [109]. During this implementation, lactate, glucose and osmolality could be sufficiently modeled across scales. However raw viable cell density and total cell density data showed some scale dependent behaviour which caused variations in the model and thus limited across scale predictions [109]. Interestingly, PLS modeling of Raman spectral data for recombinant protein monitoring was shown to be viable from early-stage product development up to

end stage 5000L bioreactors of a CHO cell fed batch process across 37 bioprocessing runs. Relative errors between 2.1% and 3.3% were found indicating good accuracy of the methodology [110]. So as to build a robust generic PLS model capable of predicting lactate, glucose, ammonium, glutamate, viable cell density and total cell density across 5-10L scales. Different CHO cell lines were utilized to generate the chemometric model [111]. It must be noted that some issues with automated feeding strategies that center around maintaining glucose at a low set point by utilizing Raman spectroscopy for glucose monitoring have been observed. This is because measurement error can be around 0.3 g/L to 0.5 g/L thus maintaining a glucose concentration below such values can be unreliable [112]. To counteract such issue measuring both lactate and glucose has been proposed. Consequently, with the proposed strategy glucose addition is stopped when lactate exceeds a predetermined set point [112]. Alternatively, glucose was fed when lactate was detected to be beneath a predetermined set point. Such strategy led to an 84% increase in final titer [112]. Additionally, it has also been shown that just by controlling glucose near a 2 g/L set point with Raman monitoring glycosylation profiles could be tuned such that a reduction in glycation was observed from 9% to 4% when comparing the Raman controlled strategy to a standard bolus feed addition strategy [113]. Noteworthy however, is the fact that since glucose was tightly controlled at 2 g/L in the bioreactor media, the cells consumed less glucose and consequently produced less lactate. This in turn caused an increase in pH across process time and meant that less base then expected was added [113]. As such it can be observed that careful consideration with respect to cascading effects caused by changes in the process must be considered and detected quickly in the early process development cycle.

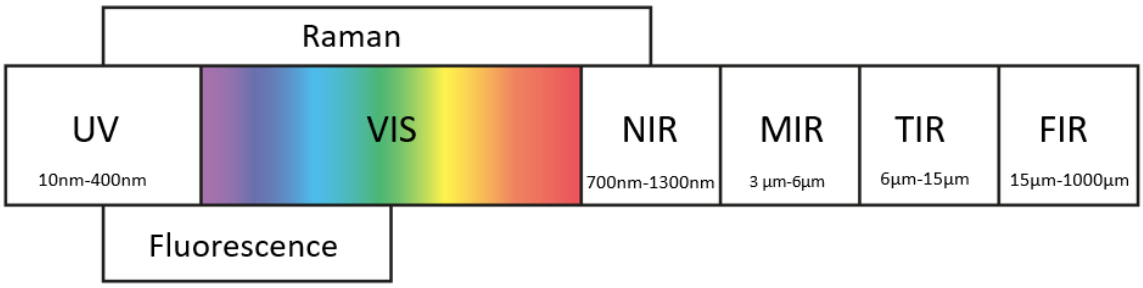


Figure 2.5. Spectroscopy spectrum.

As it can be discerned from Figure 2.5, the spectroscopy spectrum is wide and various technologies overlap with each other. For this reason it can be postulated that spectroscopic methods are not competitive but rather complementary to each other [4]. That being said, Raman spectroscopy has become the gold standard since it has the unique ability to measure multiple analytes at the same time with good enough accuracy.

Another important type of spectroscopy is dielectric spectroscopy, which has matured enough to be widely used in monitoring the evolution of viable cell concentrations across culture time. The working principal behind the sensor is the continuous estimation of passive dielectric properties of cells within the bioreactor medium with sterilisable probes [114-117]. The estimation of said variables are possible because the cell is insulated by a phospholipid layer that has very low conductivity when compared to the cytoplasm inside the cell that is a complex mix of water, salts, proteins, nucleic acids that are highly conductive. As a result, when an oscillating electric field is applied to cells in suspension cellular polarization occurs as a consequence of intracellular and extracellular ions moving towards the direction of the electrode of opposite charge only to be stopped by the phospholipid layer [114-117]. Therefore, only cells that have an intact membrane exhibit

accumulation of charge (capacitance) and as a result permittivity values that are measurable. On the other hand, cells with a compromised phospholipid membrane cannot polarise and no permittivity values are detected. As a direct result of these dielectric properties viable cells can be estimated within a bioreactor. Importantly, not all cells have the same diameter nor does the diameter remain constant during a culture run and as a consequence the measured permittivity is more an estimation of biovolume/biomass [114-117]. Importantly, fragments and or solids that are not part of biomass are not capable of polarising given the lack of an uncompromised phospholipid layer. It must be noted that the dielectric properties of cells demonstrated frequency dependant permittivity values [114-117]. That is to say, permittivity measurements of a cell decrease as the frequency of the oscillating electric field increases [114-117]. This phenomenon is denominated dispersion and three are relevant and identifiable for bioprocessing, they are α, β and γ. The α, dispersion is observed at low frequencies (below 10 kHz) and are most commonly associated to diffusion processes of ionic species within a solution. The β dispersion, is observed at medium frequencies (between 0.1 Hz and 10 MHz). This dispersion is associated with the polarisation that occurs among the insulating phospholipid layer the two conductive layers made up of cytoplasm and aqueous extracellular solution. Last but not least, γ dispersion is detected at very high frequencies (above 1 GHz) and is largely caused by the reorientation of water molecules that are exposed to high frequency alternating electric fields [114-117]. As it can be discerned within the context of mammalian cells in suspension operated in a stirred tank bioreactor, β dispersion is used for the characterization of viable cell concentrations given that this specific dispersion is associated with the polarisation of electric charge cased by the non conductive phospholipid membrane [114,115]. Because of the

physical mechanisms henceforth established, permittivity values correlate well with viable cell concentrations either with linear regression and theoretical based cole-cole models during the growth phase and assuming cellular diameter remained constant [118]. Additionally, during the culture, there was a noticeable correlation between the evolving cell diameter and shifts in the characteristic frequency. It is worth noting that alterations in cell membrane properties are also believed to potentially impact this parameter [118]. The Cole-Cole model works by representing how a material responds to an electric field over a range of frequencies. It uses a mathematical equation to describe the impedance of the material, which includes both resistance and reactance. This equation accounts for the fact that biological materials exhibit both immediate and delayed responses to electric fields. By fitting experimental data to this model, researchers can extract important information about the material's properties, such as the distribution of cell sizes and their health, making it a valuable tool in biophysics and biomedical research. Nutrient availability has also been found to be monitored through permittivity measurements because at every feeding event, permittivity events were detected to increase while declines in permittivity signals were correlated to states of nutrient depletion [118]. Additionally, during the exponential phase (which is when cell radius remains more or less constant) good correlations with OUR were obtained. This may suggest that metabolic activity may also be indirectly reflected by dielectric spectroscopy [118]. It must be noted that although linear models are able to relate dielectric spectroscopy data to cell counts during the exponential phase of a cell culture, accuracy in viable cell density estimation decreases during the decline phase [116]. To counteract this fact multivariate approaches like PLS or orthogonal partial least squares (OPLS) modeling have been utilized [116,119-122]. Thanks to this multivariate methods,

estimation of VCD rather than VCV could be realized across the entire fed-batch cycle [122]. To test the multivariate models the fed-batch cultures were subjected to dilutions (30 % vol) of the cell culture broth at various points of the process. It was observed that the decrease in cellular concentrations as a direct cause of manual dilutions were detected by the model [122]. Interestingly, it has been determined that in insect cells, dielectric spectroscopy was able to quantify cell growth. This knowledge could then be used to determine the most appropriate point in the culture to infect cells with baculovirus so as to maximize the yield of β galactosidase [123]. Changes in cell density could also be monitored in real time so as to evaluate the progression of infection in the bioreactor [123]. Since infection is characterized by an increase in cell volume, the captured permittivity signals reflect the phenomenon as bio-capacitance values spike following infection. It has also been proposed that permittivity signals are more accurate for VCD estimation of adherent cells in micro carriers than offline alternatives [117]. Further work has been realized to develop a universal PLS model that can be easily transferred across cell lines without compromising the ability of the model to accurately predict viable cell densities, especially in the declining phase of a fed-batch culture [124]. Importantly, since permittivity signals can monitor the changes in viable cell concentrations during a bioreactor run, these permittivity values can be used as a proxy for estimating cell growth. This allows for the dynamic adjustment of feeding rates to align with biocapacity signals [125]. This is important as it simplifies the integration of online signals into feed algorithms. In the aforementioned study, constant re-calculation of growth rates was realized automatically so as to determine bolus feeding frequency [125]. When compared to manual bolus feeding the experimental results showed that dynamic feeding attained higher viable cell density values and

increased titer yields[125]. Accordingly, dynamic feeding strategies that were coupled with feedback control adjusted feed rates in response to changes in cellular growth. This method suggested that such strategies can avoid over or underfeeding cell cultures across the fed-batch process[125]. As a testament to the maturity dielectric spectroscopy, Biogen compared measured permittivity signals in 5L, 200L, 315L and 15000L bioreactor scales [126]. Observing no scale dependency in the signal trends. As a result of this consistency across scales it was possible to automate seeding strategies within the manufacturing floor by relying on permittivity signals. Furthermore, utilizing the permittivity data allowed for the development of dynamic feeding strategies that compute and deliver feed volumes as necessary, as opposed to the conventional practice of calculating a daily feed volume and administering it in a single batch [126]. Hence it was determined that the resulting feeding strategy was responsive to actual culture behaviour and therefore minimized the risk of process failure caused by underfeeding [126]. Additionally, within the same study, they also observed that permittivity signals can be used as a marker for a change in salt balance within a mammalian cell process. A change in potassium phosphate concentration within the media (that was not previously known) caused drastic changes in the metabolic behaviour of the cells. This issue was only fixed thanks to noticing the shift in permittivity signals when compared to historical processes [126]. Subsequent Biogen studies demonstrated the feasibility of relating integral of cell growth-based feed amount with the integral of bio-capacitance data [127]. Thanks to this a linear correlation between the integral of the bio-capacitance signal and the total feed amount was established for the entire process. Therefore, it was possible to directly employ permittivity signals in the control of feed addition [127]. To test the robustness of the strategy, bioreactors where intentionally

seeding at low viable cell concentration densities. The feeding strategy that utilized the integral of capacitance adjusted the feed, avoiding overfeeding while the fixed feed strategy did not change the total amount of feed added and consequently created an overflow metabolism (increased ammonium and lactate concentrations) that led to cell growth inhibition [127]. Because of this, higher titers and higher product purities were attained in the bio-capacitance based feeding strategy since the fixed feed process altered glycation and trisulfide formation [127]. Consequently, it was determined that by introducing permittivity data into process control the robustness of a process could be improved thereby providing consistency in productivity and product quality despite process variations. Another interesting approach noted in the literature is that of determining glucose uptake rate measurements by relying on permittivity signals to estimate viable cell density and glucose concentration measurements. Using these glucose uptake rate measurements prediction of feed requirements for the next sampling day were realized [128]. This led to optimally fed cultures across the fed-batch cycle in terms of waste product build up as lactate profiles improved when compared to a standard bolus strategy.

Similar to their optode counterparts, pH, DO, and carbon dioxide sensors can be potentiometric, conductometric voltametric or amperometric which work by detecting charged species or variations of electrical phenomena during chemical reactions [129]. Temperature sensors also fall within this category and they are a stable of processing runs as temperature has to be tightly controlled along a set point that varies depending on the organism of choice (mammalian cells are generally operated at 37 ℃ during the exponential phase while insect cell lines have optimal growth at 28 ℃) [130]. Importantly, it is routine that cultures are subjected to process induced shift during the culture run, namely

pH shifts and temperature shifts [131]. Thus, there exists a reliance on accurate sensors so as to maximalize the impact that such shifts have on process outcomes. Other noteworthy sensors are electrochemical enzymatic arrangements that can monitor glucose and lactate on-line [132]. Such sensors may be advantageous as they reduce the need for sampling which always adds risk of contamination. However, given that they rely on enzymatic reactions the sensors are only usable for up to 21 days meaning they can only be used in a single fed-batch cycle and are unusable to long term processes like perfusion [132]. Since accumulation of dissolved carbon dioxide has been observed to have significant impacts on both product quantity and quality there has been work centered around monitoring the values of dissolved carbon dioxide across culture time [133]. Two equivalent CHO fed-batch cultures were realized in parallel, one tightly controlled at 10% dissolved carbon dioxide and another culture without dissolved carbon dioxide control (reached a maximum of 20%). It was observed that the tightly monitored and controlled culture yielded higher recombinant protein titer and had a longer productive phase [133]. Importantly, pH values were comparable between both cultures suggesting that pH control alone is not sufficient for maximum process optimization [133]. Additionally, optimized sparging strategies that take into account dedicated dissolved carbon dioxide feedback loops, for the purpose of gas only pH control, have been reported in the literature to improve batch-to-batch reproducibility [36]. This allowed cultures to not require base addition for pH regulation which means no osmolality fluctuations related to pH feedback loops [36]. Dissolved carbon dioxide sensors have also been utilized during scale up of bioreactors. One such strategy relied on maintaining a constant ratio between kLa (O2)/ kLa (CO2) across reactor scales [134]. It was observed that by utilizing this ratio as a scale up criteria dissolved carbon dioxide profiles remained

comparable across reactor scales. Thus it was suggested that this strategy can help maintain consistency in product quality and quantity when doing technology transfer projects [134]. Other applications include using the probe to determine carbon dioxide removal rates [135]. This then allows researchers to test out various aeration strategies and specific power inputs that are most effective at striping out carbon dioxide from the bioreactor [135].

The determination of OUR in bioreactors for mammalian cell culture is primarily realized through three standard methods: the dynamic method, the stationary liquid mass balance method, and the global mass balance method [136].

$$O.U.R. = \frac{C_{\mathrm{L}}(t_0) - C_{\mathrm{L}}(t_{\mathrm{f}})}{t_{\mathrm{f}} - t_0} + \frac{\int_{t_0}^{t_{\mathrm{f}}} (-K_{\mathrm{des}} \cdot C_{\mathrm{L}}(t)) \cdot dt}{t_{\mathrm{f}} - t_0} \qquad (1)$$

$$O.U.R. = \frac{G_{\mathrm{m}} \cdot \left(yO_{2,\mathrm{in}} - yO_{2,\mathrm{out}}\right)}{V_{\mathrm{L}}} \qquad (2)$$

$$O.U.R. = O.T.R. = k_{\mathrm{L}}a \cdot \left(C_{\mathrm{L}}{}^{*} - C_{\mathrm{L}}\right) \qquad (3)$$

1) The dynamic method relies on the cyclical measurement of the dissolved oxygen extinction profile when air supply is stopped. The constant of desorption (K_{des}) must be previously determined in cell-free media keeping the same culture conditions while the real oxygen concentration in the liquid phase (C_L) can be readily measured with a DO probe, t_0 and t_f represent the initial time of measurement and the final time of measurement respectively. The DO concentration decreases due to cellular oxygen need and thus OUR directly correlates with the gradient (slope) of the decay curve [136].

2) The global mass balance (GMB) approach relies on estimating the oxygen concentrations at the inlet and outlet of the bioreactor while the DO is kept constant. Given

that this method is dependent on gas phase estimation of oxygen concentration the used devices can be complex and expensive [136]. For example, mass spectrometers, paramagnetic sensors, acoustic analyzers can be used. Additionally, zirconium dioxide or galvanic cell off-gas analyzer have become popular as of late [136].

3) This Liquid mass balance method relies on the estimation of the OUR by means of calculating OTR. This is true because if the DO is held constant, an estimation of OTR must equal OUR. For this purpose, the theoretical oxygen concentration in equilibrium with the liquid phase can be estimated. The real oxygen concentration in equilibrium with the liquid phase can be monitored with DO probes. The oxygen concentration in equilibrium with the liquid phase (C_L*) can be calculated using Henry's law while the k_La must be determined in real time given that changes in flow rates or stirring (common parameters for DO control) can change its value [136].

It was also shown that DO fluctuation had a greater negative impact on performance with respect to titer production and lactate accumulation than maintaining a DO setpoint [137]. Additionally, estimated specific oxygen uptake rates (qOURs) were lower in DO fluctuation lots when compared with DO stable lots [137]. Another study showed that DO fluctuations can impact glycosylation profiles of produced antibodies [138]. Consequently, a shift away from the gold standard of dynamic method has taken place.

Beyond OUR estimation it has also been shown that oxygen flow closely mirrored viable cell concentration which is not surprising given that oxygen demands should increase with increased OUR [139]. This in turn should result in increasing viable cell concentration. Specific oxygen uptake rates were calculated and determined to be relatively constant in the first half of the cultures while a relative increase was observed in the second half of the

cultures. This change was later correlated to a shift in metabolism from lactate production to consumption. To further explore this correlation, an averaged Respiratory Quotient (RQ =CTR/OUR) was calculated [139]. During the lactate production phase, RQ>1 was determined to be the norm while during the consumption phase RQ<1 was found. In theory exclusive use of glucose yields an RQ value near 1 while the oxidation of more reduced molecules like amino acids or fatty acids can result on an RQ below 1 given that more oxygen is required to fully oxidise these molecules. Given that at the start of the culture, glycolysis is the main metabolic pathway (anaerobic fermentation), a relatively lower specific oxygen uptake rate is expected and consequently a higher RQ [139]. On the other hand, during the lactate consumption phase the TCA cycle is the dominant metabolic pathway. Consequently, a higher specific oxygen uptake rate is expected leading to a lower RQ. Because of the strong linearity between oxygen and glucose, it was also postulated that the relationship between the total oxygen consumption rate and the total glucose consumed can serve as a control method to sustain the glucose level in the basal media at a chosen target value. Thus, OUR estimation or oxygen flow requirements can be used as a key parameter in process monitoring and control [139].

Recent trends towards developing soft sensors for online estimation of biomass in mammalian cell cultures have been established [140]. OUR values have been known to correlate not only with viable cell counts but also with metabolic turnover rates. The developed soft sensor is made up of two distinct models that predict in terms of VCD and VCV based on a multi-linear regression [140]. Interestingly, VCD has been found to lack important information about cellular volume. For this reason, viable cell volume (VCV) has been determined to relate more information regarding biomass information.

Consequently, VCV can lead to more precise correlations with OUR [140]. This is key because cell size can have a direct impact on oxygen demand. Because of this, larger cells can have increased oxygen requirements when compared to smaller cells. The authors considered cell number to be insufficient as an estimator of biomass because increased cell growth is always accompanied by increase in cell numbers and cell volume. Additionally, they postulate that even small changes in mean cell diameter during the production process can result in large differences of cell volume [140]. Since mammalian cell volume differs not only between cell lines but also throughout the process time, viable cell volume serves as a better indicator of biomass. The models were capable of adequately describing biomass change when compared to offline measurements. Importantly, the online biomass estimation can be used to estimate specific consumption rates [140]. As mentioned, OUR can provide information regarding the metabolic state of cultures. In fact, its linear relationship with substrates makes it a strong candidate for developing dynamically fed cultures [141]. On the other hand, the culture transition from the plateau phase to the death phase can be clearly detected in the online monitoring by a decline in oxygen consumption and specific oxygen consumption rates which can prove useful in avoiding overfeeding towards the end of a cell culture cycle [141]. Glucose consumption on the basis of stoichiometric relationships between accumulative oxygen consumption and accumulative glucose consumption is the common strategy for dynamically fed cultures using OUR [142]. Importantly, the relationship between glucose and oxygen consumption was not constant over time and thus was manually re-calculated with offline glucose samples as to ensure the estimation did not deviate over time [142]. Thanks to this strategy sufficient nutrients were available for cell growth as well as maintaining glucose and glutamine at low levels as to reduce the glucose

consumption rate, and consequently the production rate of lactic acid [142]. A similar result was obtained in perfusion cultures of hybridoma cells. By employing stoichiometric ratios that were previously calculated, the amount of nutrients consumed was estimated from hourly oxygen-consumption-rate measurements [143]. OUR monitoring of adenoviral vector production using HEK-293S has been used to automate the transition from batch to perfusion mode [144]. It demonstrates the reliability of the dynamic nutrient feeding strategy. Indeed, the perfusion rate varied according to a given OUR value and was capable of maintaining glucose concentration at a given level [144]. Importantly, the final adenoviral titer obtained in this process clearly improves the basal batch process by 4 times the total amount [144].

Most recently, OUR estimation has been used as a tool to keep glucose concentration near 20 mM in fed batch processes [145]. This is done by estimating biomass concentrations from OUR readings through specific oxygen uptake rate estimation. Feeding was started once the OUR biomass indicator indicated a density of 3 million cells/mL. The cultures would reach a maximum viable cell concentration of 13 million cells/mL [145]. This automated strategy of feed addition presented a 102% increase in cell number when compared to a standard batch process. This massive increase in cell density was concomitant with a 124% increase in final product titer concentration and a 64% increase in volumetric productivity. The glucose concentration was kept within a narrow range thanks to the linear relationship between glucose consumption and oxygen consumption [145]. Another attempt at using OUR as a control strategy revealed that the parameter in question has a close relationship with viable cell density even when exponential growth transitioned into cellular death [146]. It was noted that maximum OUR was attained before maximum VCD, possibly anticipating

limiting conditions in the culture. For feed additions, the proposed method relied on predicting glucose consumption for the next time intervals [146]. Other studies have shown that at high cell densities (above 10 million cells per mL), the OURs no longer exhibits a clear relationship with the VCC. On the other hand, when viable cell volume was used to correlate oxygen uptake rate throughout the culture, a less scattered and more accurate correlation was found regardless of the viable cell density [147]. However, changes in respiration rates were observed and thus biomass estimation could be made with linear regression if the culture was partitioned at the point of respiratory shift. This allowed for prediction of biomass in a validation dataset with an error of 14%. A similar strategy was used for capacitance measurements and it was determined that permittivity data better reflected viable cell volume than viable cell count [147]. It was hypothesized that the switch in specific respiration rates was caused by a shift in the metabolic state of the cells which may be caused by a truncated TCA cycle. In a truncated cycle, less energy is produced, and less oxygen is consumed. Importantly, real time metabolic transition could be tracked over time and future combinations of on-line permittivity values and respiratory measurements to allow for more advanced process optimization and control of metabolic states could be a future area of soft sensor research [147].

A feedback controller was developed as to switch the glucose pump on if the predicted cumulative on-line glucose concentration generated with the integral of OTR correlation was higher than the cumulative glucose added to the vessel [148]. This control strategy diminished glucose variations that were observed during bolus addition. Importantly, the adaptive nature of the dynamic feeding strategy tracked the variation of glucose consumption rates and ensured that the glucose levels were maintained within the culture

without the need for manual adjustments [148]. Lastly, OUR was used to monitor the metabolic activity of the culture as well as to detect nutrient limitations [40]. Thanks to this on-line signal based on a liquid mass balance, a feeding strategy was developed to obtain different specific protein production (qP) profiles, Steep decrease in OUR was linked to the exhaustion of tyrosine in the medium as was confirmed by spent broth analysis. With this knowledge supplementation of tyrosine was realized [40]. Therefore, respiratory activity of the cells recovered and OUR showed an increase. This pattern of depletion and supplementation was repeated throughout the process. Importantly, it was observed that the exhaustion of tyrosine led to a decrease in OUR, it also decreased the specific productivity of the culture. To further determine the impact of supplementation on the OUR profile, an experiment was designed where a supplementary feed containing tyrosine (and other essential amino acids) was added to the cell culture. This was compared to a standard un-supplemented fed batch. The supplementary feed was added immediately after observing an initial decrease of the OUR profile [40]. The on-line monitoring of cellular respiration, which detected early onset of nutrient limitations maintained a high specific productivity when compared to un-supplemented cultures, which decreased in specific productivity over time due to nutrient limitations. Thus, it can be concluded that either OUR or oxygen flows can give underlaying information about cellular volume or metabolic activity.

Soft Sensors

The term soft sensor describes the union between hardware and software implemented models so as to collect information about an ongoing process that would be unattainable under different circumstances by relying solely on hardware sensors. Consequently, it describes the synergy between software tools and hardware technologies for the purpose of discerning phenomena within a bioreactor [149-152]. Thus, soft sensors leverage accessible online data so as to deduce numerical data about complex process variables that are impossible to measure directly within a sterile system or can only be currently measured at very low frequency with analytical tests [153]. Consequently, soft sensors can become useful tools in terms of monitoring and control implementations within the biopharmaceutical industry [149-152]. A sophisticated soft sensor, thoughtfully designed to accommodate the needs of its stake holders, is theoretically expected to decrease the requirement of daily surveillance and maintenance work. Moreover, soft sensors should improve interpretability of the results from bioreactor runs by leveraging the capability of the model to establish connections between various crucial variables [4]. With this in mind, soft-sensors are optimal candidates for the PAT initiative so as to contribute in the path towards automated control [4,153]. Crucially, soft sensors can be classified into three groups: data driven sensors, model driven sensors, and hybrid models [4,100].

Model-driven sensors involve mechanistic models rooted in engineering principals such as mass or energy balances. As a result, they provide insights into the process inherently embedded in biological insights [100]. These models can incorporate known culture conditions like media composition and performance indicators (protein yield, lactate production, glucose consumption, cell growth) so as to establish concrete models. Due to

this characteristics, model-driven sensors can leverage known kinetic equations that captures the dynamic variation of variables [154]. In summary, these types of soft-sensors can incorporate, thermodynamic restrictions, transport phenomena and reaction kinetics into the model which greatly facilitates interpretability [154]. Nevertheless, model-driven soft-sensors must undergo thorough stages of parameter identification, uncertainty assessment and sensitivity analysis so as to adequately evaluate and validates the constructed models. It is important to highlight that when the constructed model-driven sensor is shown to be replicable and reliable, it can greatly enhance the understanding of the manufacturing process [4,154]. Model-driven soft sensors can be classified into two clear cut categories: steady-state models and dynamic models. The latter type of models rely on dynamic balances and kinetic assumptions to adequately articulate rate expressions as a function of state variables [155]. Alternatively, steady-state models emanate from mass and heat transfer phenomena. Examples of steady-state models include Flux Balance Analysis (FBA) or Metabolic Flux Analysis (MFA) which are stoichiometric-rooted techniques commonly employed to characterize cellular metabolism [156-158]. They are also valuable for estimating intracellular fluxes by utilizing known extracellular analyte or production rates as model restrictions. Because the quasi-steady-state assumption for intracellular metabolites is crucial, these models are deemed static in nature [154]. In contrast, kinetic models are typically formulated as a set of ordinary differential equations (ODEs) that can depict dynamic alterations in metabolite concentrations, titer, cellular growth across culture time [149-152]. Due to this, cellular death and growth can be explicitly linked to alterations in the concentration of pertinent nutrients and metabolic waste products. Furthermore, protein expression has been linked to both cell growth and amino acid metabolism [154,159,160].

Dynamic models can be designed with different degrees of intricacy, contingent upon the assumptions formulated by the researchers concerning the culture system within the reactor. This varied level of complexity can be adjusted by taking into account the heterogeneity within cell populations or by recognizing the presence of established cellular compartments and their respective behaviours. Conversely, dynamic models can be streamlined when reactions are combined into rate limiting steps [154,159,160]. Due to these considerations, mode-driven sensors can be very complex to develop.

Hybrid models, also referred to as grey box models, constitute another pertinent category of soft sensors. These soft sensors can be viewed as an amalgamation of data driven soft sensors and model driven soft sensors, effectively harnessing the advantages of both methods [149-152]. Several architectural strategies are available for crafting hybrid models, broadly categorized into three groups: i) Calibration ii) Composition iii) Transformation. Calibration architectures employ black box (data driven) models to minimize errors in mechanistic models. Composition architectures utilize black box models to estimate unknown terms within a mechanistic model. Finally, a transformation approach leverages mechanistic models to create data rich environments from which to train a black box model [161]. Illustrations of such approached include state observers which integrate dynamic modeling (white box) and data-driven modeling (black box). This is realized by updating the state estimates derived from noisy measurements and gradually reducing the estimation error with each iteration [4]. This necessitates the distinct assumption of linear dynamics within the process and a gaussian distribution for the error terms. Under such assumptions a Kalman filter can be employed. However, considering that the process dynamics within a bioprocess are inherently non-linear, the extended Kalman filter may be utilized. This

approach achieves a piecewise linearization through a first-order Taylor series expansion [162]. Another significant variant of the Kalman filter that is extensively applied to non-linear systems is the unscented Kalman filter. This technique utilizes an unscented transform to circumvent the necessity for a Taylor series expansion of the system of equations so as to linearize the model [163]. This approach can be beneficial as the unscented transforms enables the use of non-linearizable functions as a state observer. Consequently, black box functions such as support vector machine regression can be employed to establish a relationship between an online sensor output and a non-online variable [163]. It is essential to acknowledge that the accuracy of a hybrid sensor (grey box model) can be greatly influenced by the accuracy of the mechanistic model integrated into the grey box mode. Therefore extensive validation of the mechanistic model is imperative so as to ensure its capability to effectively represent the process [164]. Extended Kalman filters have been implemented in rAAV production processes where online bio-capacitance signals and an unstructured mechanistic model are utilized in conjunction with neural ordinary differential equations so as to estimate cell-specific rates. This configuration yields a soft-sensor capable of continuously estimating other state variables that are otherwise measured at low sampling frequencies [165]. Similarly, hybrid extended Kalman filters incorporating PLS model to estimate specific rates within the model have also been devised [166]. Hybrid models can also characterize the biological system using a mechanistic framework but articulate cell-specific rates through black box statistical expressions [149-152]. Here, the mechanistic framework inherently limits the solution space of the model. Therefore the statistical cell-specific rate expressions can be readily determined [167]. In this framework, the inputs into a mechanistic model can be provided a PLS or an Artificial Neural Network (ANN)

prediction derived from multi-wavelengths spectra or multivariate parameters [168,169]. In scenarios with limited data, such as rAAV production, employing unstructured mathematical models enables the generation of cell line specific data. This information-rich environments can then be utilized to train non linear regression techniques such as recurrent neural networks (RNN) which are exceedingly good with time varying data [170].

Multivariate analysis techniques such as PLS, principal component regression (PCR), and non-linear regression methods like ANN and SVMR are also employed in data-driven soft sensor development. These techniques aim to establish relationships between input features and the desired variables [171-175]. These nonlinear models can play a crucial role in comprehending mammalian cell culture, especially considering that the relationship between metabolic indicators and process variables remain unknown or are highly cell line specific [4]. PLS and ANN regression are commonly employed for the analysis of spectral data, particularly NIR/MIR/Raman/Fluorescence spectroscopy. In this context, spectral data serves as input and is correlated with outputs such as substrate concentrations, biomass accumulation, cell viability and/or titer [124,176]. These models enable the prediction of crucial process variables that may not be directly discernible from spectral signals or multi-sensor data alone but from the deconvolution of datasets generates by such sensors. In theory this presents an advantage over mechanistic models, as online measurements (such as pH, oxygen sparging data, base addition, DO control curves, process temperature, dissolved carbon dioxide sparging, OUR, bio-capacitance signals, Raman spectra, etc.) not directly coupled to cellular growth and metabolic indicators (protein expression, lactate accumulation, etc.) can be used to aid in the prediction of those variables. Alternative methods applied in bioprocessing include multiple linear regression, k-nearest neighbors,

regression tress, ensemble approaches (Gradient Boosting Machine, Extreme Gradient Boosting, Adaptive Boosting, Random Forest), and Gaussian process regression [177-179]. As mammalian cell culture exhibits both time-dependent variation and a multivariate nature, techniques designed for sequence forecasting have been employed. While ANNs can capture dynamics of non-linear systems like cell culture runs, RNNs, a sub class of ANNs, excel in capturing the internal temporal dependencies of a system. These architectures are especially beneficial for making predictions of relevant state variables several time steps ahead [180]. Recently, these networks have been employed to predict biomass growth both before and after transfection in an rAAV production process [181]. This involved using features such as cumulative oxygen sparged and DO values, integral of DO. The time-related variance of these features was then related to cellular growth. Additionally, RNN models for multivariate estimation of mammalian cell culture data (total cell density, viable cell density, viability, lactate concentration in the reactor broth, glucose concentration in the reactor broth and titer) have also been developed [182]. Data driven techniques such as PCA can be employed for exploratory data analysis and qualitative understanding of the data. The most common approach in bioprocessing is to use PCA as it can help visualize data, compare datasets and reduce dimensionality [178]. PCA is particularly useful for identifying correlations among variables. It is a process that transforms a large dataset with collinearity into a low-dimensional space of uncorrelated principal components. Each principal component explains a certain portion of the overall variance within the dataset [178]. Crucially, the connection between principal components and the initial process variables can be established through eigen vectors providing interpretability of the outcomes [183]. PCA has been commonly employed to assess the effectiveness of scaling

operations across different reactor dimensions [184-186]. Statistical tests leveraging PCA have been created to ascertain the extent of resemblance between the scaled-up process and the laboratory-scale process [187]. Considering bioprocessing data is a three-dimensional dataset, encompassing diverse experiments with variables evolving over culture time (lot number, process parameters, sampling time points), unfolding of data becomes essential. PCA, inherently designed for two-dimensional dataset's requires specific methods for unfolding, including batch-wise, time-wise and variable-wise unfolding. Collectively, these methods are referred as multiway PCA [188,189]. Batch-wise unfolding facilitates a direct comparison across different batches, albeit at the expense of losing granularity along the time dimension [183,190]. Nevertheless, by comparing batches distinct experimental clusters can be visualized providing insights into how various process parameters were altered between conditions. While PCA is commonly employed on datasets from late-stage process development or manufacturing to asses outliers and enhance process robustness, investigations indicate that employing comparable techniques on early process development datasets can unveil details such as uncontrolled variance, experimental flaws and gaps in knowledge [191,192]. This methodology facilitated the establishment of a framework enabling exploratory data analysis to assess the influence of carbon dioxide concentrations and osmolarity on antibody production [183]. PLS can also be employed to discern broader patterns in the dataset by reducing the original dimensionality[178]. However, a crucial distinction is that PLS established a connection between a feature vector (X) and a response vector (Y), which can subsequently be used in regression tasks for outcome prediction. This approach has found extensive use in bioprocessing scale-up, allowing the assessment and tracking of variable importance across different scales [193]. As

an example, PLS served as a regression method for forecasting monoclonal antibody production, facilitating a comparison between 3L scale process and a 2000L scale process[194]. By analyzing PLS loading plots and adjusting process conditions of the 3L bioreactor, a comparable process was established at a smaller scale, exhibiting similar behaviours and outcomes when compared to the 2000L process. Despite its high interpretability, as it involved dimensionality reduction on the feature vector and response vector followed by a linear regression with the non-correlated latent variables, the technique is constrained by the assumption of linearity [178]. Considering the inherent non-linearity of biological processes, less interpretable but more effective methods can be employed through machine learning models, particularly when aiming to create soft-sensors. Non-linear SVMR has been applied in large scale reactors to predict titer based on several measured process variables [172]. Lactate metabolism was determined to be a pivotal process indicator forecasting the endpoint protein yield, underscoring the significance of regulating glycolytic fluxes during seed train inoculation on large scale reactors [172]. Tree-based models serve as potent tools capable of precisely capturing non-linearity inherent in biological systems while maintaining a high level of interpretability [178]. In this context, the data can be partitioned into intervals and a decision tree can delineate the response variables outcome based on the observed ranged of various input variables. As this strategy may risk overfitting, a random forest approach can be implemented. Random Forest techniques involve running an ensemble of decision trees concurrently and subsequently averaging the predictions from each decision tree. This collective approach mitigates the overfitting challenges associated with individual decision trees. These algorithms have been employed in microbial hydrogen production, demonstrating improved model performance compared

to regularized regression [177]. These methods have been applied in the modeling of Raman spectroscopy, providing an alternative to the established PLS gold standard [195]. Extreme gradient boosting algorithms also rooted in decision tree-based methods, distinguish themselves from random forest by constructing trees sequentially instead of in parallel. This sequential approach enables a gradual enhancement of the decision tree by continually re-weighting trees that accurately predict outcomes at were poorly modeled by previous trees [178]. Given that this objective function can be optimized with a learning rate, it can be more susceptible to overfitting. Therefore, caution should be exercised when constructing and tuning the model. Furthermore, machine learning regression techniques designed to predict process outcomes can be developed not only to forecast future outcomes based on monitored variables but also to comprehend the importance of variables within the model [196]. To achieve these goals, SHapley Additive exPlanation (SHAP) was devised based on the concept of the Shapley value. The Shapley value, rooted in game theory, aids in fair profit allocation to different stakeholders by assessing their individual contributions to the outcome [197]. Within the realm of machine learning, each feature can be linked to a stake holder and the payout corresponds to the outcome of the model. In essence, the Shapely value for each feature delineates its contribution to the prediction of the model for a specific datapoint. SHAP values have consistently demonstrated reliable results, and the utilization of SHAP dependency plots provide a valuable summary of the model [196,198,199]. Consequently, advancements in explainable AI have contributed to making model interpretation less opaque.

In figure 2.6, it is evident how various sensors discussed in this literature review serve as pivotal sources of data streams for the development of soft sensors. As emphasized, these

soft sensors exhibit diverse methodologies-mechanistic, data driven, and hybrid models-each carrying its own set of advantages and disadvantages. The ongoing initiative in PAT underscores the pivotal role that soft sensors will play in bioprocess monitoring. Black box models have demonstrated remarkable proficiency in integrating spectroscopic data with offline analytics to yields quantitative predictions. Conversely, mechanistic-based soft sensors provide interpretable insights into process dynamics and metabolic rates, serving as a foundation for control algorithm. Finally, hybrid-based soft sensors hold the promise of combining the robust modeling capabilities of complex non-linear techniques with interpretable model architecture with knowledge-based methods.

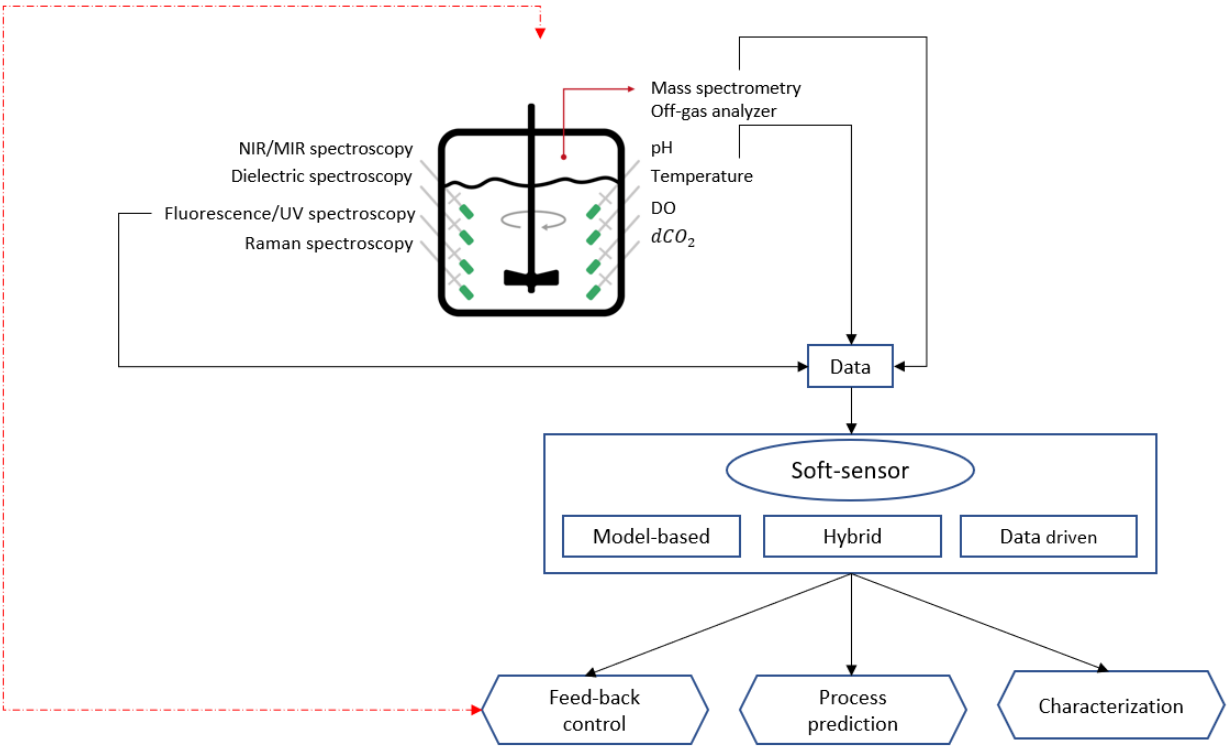


Figure 2.6. Overview of the integration of diverse hardware sensor data into soft sensor configurations.

CHAPTER 3 ARTICLE 1: MULTIVARIATE DATA ANALYSIS OF PROCESS PARAMETERS AFFECTING THE GROWTH AND PRODUCTIVITY OF STABLE CHINESE HAMSTER OVARY CELL POOLS EXPRESSING SARS-COV-2 SPIKE PROTEIN AS VACCINE ANTIGEN IN EARLY PROCESS DEVELOPMENT

The article titled "Multivariate data analysis of process parameters affecting the growth and productivity of stable Chinese Hamster Ovary cell pools expressing SARS-CoV-2 spike protein as vaccine antigen in early process development" was published in *Biotechnology Progress* on March 28, 2024, and covers objectives 1-2 of this thesis. It demonstrates the utility of applying multivariate data analysis and explainable AI techniques to historical process records so as to gain macro insights within the data and generate future process development starting points that can be validated experimentally. I was responsible for conceptualization; methodology; data curation; investigation; validation and writing – original draft.

Sebastian-Juan Reyes

Department of Chemical Engineering, Polytechnique Montreal, Montreal, H3T 1J4, Quebec, Canada

Human Health Therapeutics Research Centre, National Research Council Canada, 6100 Royalmount Avenue, Montréal, H4P 2R2, Quebec, Canada

Lucas Lemire

Department of Chemical Engineering, Polytechnique Montreal, Montreal, H3T 1J4, Quebec, Canada

Human Health Therapeutics Research Centre, National Research Council Canada, 6100 Royalmount Avenue, Montréal, H4P 2R2, Quebec, Canada

Raul-Santiago Molina

Proelium S.A.S, Carrera 16 C #153, Bogotá, 110131, Bogotá DC, Colombia

Marjolaine Roy

Human Health Therapeutics Research Centre, National Research Council Canada, 6100 Royalmount Avenue, Montréal, H4P 2R2, Quebec, Canada

Helene L′Ecuyer-Coelho

Human Health Therapeutics Research Centre, National Research Council Canada, 6100 Royalmount Avenue, Montréal, H4P 2R2, Quebec, Canada

Present address: Biodextris, 525 Bd Cartier W, Laval, H7V 3S8, Quebec, Canada

Yuliya Martynova

Human Health Therapeutics Research Centre, National Research Council Canada, 6100 Royalmount Avenue, Montréal, H4P 2R2, Quebec, Canada

Brian Cass

Human Health Therapeutics Research Centre, National Research Council Canada, 6100 Royalmount Avenue, Montréal, H4P 2R2, Quebec, Canada

Robert Voyer

Human Health Therapeutics Research Centre, National Research Council Canada, 6100 Royalmount Avenue, Montréal, H4P 2R2, Quebec, Canada

Yves Durocher

Human Health Therapeutics Research Centre, National Research Council Canada, 6100 Royalmount Avenue, Montréal, H4P 2R2, Quebec, Canada

Olivier Henry

Department of Chemical Engineering, Polytechnique Montreal, Montreal, H3T 1J4, Quebec, Canada

Phuong Lan Pham

Human Health Therapeutics Research Centre, National Research Council Canada, 6100 Royalmount Avenue, Montréal, H4P 2R2, Quebec, Canada

Correspondence concerning this article should be addressed to: Phuong Lan Pham at phuonglan.pham@nrc-cnrc.gc.ca and Olivier Henry at olivier.henry@polymtl.ca

Abstract

The recent Covid19 pandemics revealed an urgent need to develop robust cell culture platforms which can react rapidly to response to this kind of global health issue. We demonstrated in this article that CHO stable pools can be a vital alternative to quickly provide gram amounts of recombinant proteins required for early-phase clinical assays. In this study, we analyze early process development data of recombinant trimeric spike protein Cumate-inducible manufacturing platform utilizing Chinese Hamster Ovary (CHO) stable pool as a preferred production host across three different stirred-tank bioreactor scales (0.75L, 1L, and 10L). The impact of cell passage number as an indicator of cell age, methionine sulfoximine (MSX) concentration as a selection pressure, and cell seeding density was investigated using stable pools expressing 3 variants of concern (VOC) namely Wuhan (Wu), Delta (De), and Beta (Be). Multivariate data analysis (MVDA) with Principal Component Analysis (PCA) and batch-wise unfolding technique was applied to evaluate the effect of Critical Process Parameters (CPP) on production variability. Once understanding of variable relationship was gained, a Random Forest (RF) model was developed to forecast protein production. In order to further improve process understanding, the RF model was analyzed with Shapley value dependency plots so as to determine what ranges of variables were most associated with increased protein production. Increasing longevity, controlling lactate build-up, and altering pH deadband are considered promising approaches to improve overall culture outcomes. The results also

demonstrated that these pools are in general stable expressing similar level of spike proteins up to cell passage 11 (approx. 31 cell generations). This cell passage number allows to expand enough cells required to seed large volume of 200 – 2000L bioreactor.

Keywords: MVDA, Random Forest, CHO stable pool, Cumate induction, SARS-CoV-2 trimeric spike protein, fed-batch bioreactor production

Introduction

Chinese Hamster Ovary (CHO) cells are currently the industry standard when it comes to producing recombinant therapeutic proteins (RTP). This is because CHO cells are able to produce human-like proteins that have adequate glycosylation profiles. CHO cells can come in two flavors, cell lines derived from single clones and stable pools. Cell line development requires numerous stringent screening procedures that select the best performing cell clone.[51] These tests are generally carried out at the micro-liter scale to evaluate growth, protein yield, and Critical Quality Attributes (CQA). Such screening procedures can take months to complete and consequently these strategies are sub-optimal when reacting to rapidly evolving public health crisis, such as the global pandemic caused by the SARS-CoV-2.[200,201] To fast-track the cell line creation to large-scale protein production, mammalian cell stable pools have been considered an attractive alternative.[202] This is because even though the cell pools are less homogeneous than their cell line counterparts, they can still be utilized to produce recombinant proteins at a sufficiently large-scale. It is noted however that cell pools can be susceptible to cell age effects and thus for stable pools to be a viable alternative to cell lines to provide enough materials for toxicology study and Phase 1 clinical trials, the process must be understood and optimized.[52] Optimizing pool process intensification can rely on Design-Of-Experiment

(DOE) to identify process conditions that enhance culture performance in a given production scale. Once sufficient data are generated across various production systems and with different pools, exploratory MVDA can be utilized to sort through the data to understand what parameters promote good performance or impact adversely the production process.[4] Additionally, soft sensors that predict process outcomes can also be developed not only to predict future yields based on monitored variables but to also understand the variable importance within the model.[196] When variable impact on model is combined with intimate process knowledge, insights can be gained to create representative models or improve process performance.

In terms of large dataset analysis, Principal Component Analysis (PCA) and Partial Least Square (PLS) are the two tools frequently used in biomanufacturing. PCA is generally employed as the first step of MVDA to explore the large data structure and provide process dynamics. PLS can be applied as the second part of MVDA focusing on process optimization and forecasting. In the first part of this paper, we will focus on the use of PCA to reduce data dimensionality thus facilitate data visualization and comparison.[203] PCA can uncover correlations among variables or show relationships between outcomes and variables to ultimately detect trends and outliers. PCA is a process that transforms a large dataset with collinearity into a low dimensional space of new uncorrelated variables so-called Principal Components (PC) in such a way that each PC explains a certain portion of the overall variance within the dataset.[203,204] In biomanufacturing, PCA has been readily used to evaluate the success of scale-up/scale-down and process impacts on glycosylation profiles.[186,205-208] Bioprocessing data can be considered as a three-dimensional dataset that consists of various experiments where all

the variables vary through culture time (lot number, process parameters, sampling time points). The first step in PCA is data unfolding. There are numerous approaches such as batch-wise, time-wise, and variable-wise unfolding. Collectively, these methods are called Multiway PCA (MPCA).[209,210] Batch-wise unfolding allows the direct comparison between different batches although granularity along the time dimension is lost.[204,211] However, by comparing among batches, different experimental clusters can be visualized and then understood depending on how various process parameters were changed between the conditions. Although PCA is generally used on late process development or manufacturing datasets to evaluate outliers and improve process robustness, studies have shown that applying similar techniques to early process development datasets can be beneficial revealing information such as uncontrolled variance and experimental flaws.[192,212] As mentioned above, Partial Least Square (PLS) is a similar tool that transforms the original dataset to latent variables, thus reducing the original dimensionality.[203,213] However, a key difference is that PLS relates a feature vector (X) to a response vector (Y) which can then be utilized in regression to predict outcomes. This tool has also been applied widely in bioprocessing scale-up.[214] For instance, PLS was used as a regression technique to predict monoclonal antibody production to compare a 3L and a 2000L scale process.[215] Through interpreting the PLS loading plots and altering the process conditions of the 3L bioreactor, a comparable process was created at small-scale that displayed similar behaviors and outcomes when compared to the 2000L process. Thus, the 3L bioreactor can be a scale-down model to predict the 2000L performance. Even though the PLS technique is highly interpretable (carry out dimensionality reduction on feature vector X and response vector Y and then realize a linear regression with the non-correlated latent variables), it has the

downside of utilizing a linearity assumption.[203] When PLS was coupled with amino acid stoichiometric balances, an approach to rapidly optimize amino acid concentrations in chemically defined media was developed. Here, PLS was deployed to comprehend the relationship between the dynamics of time-dependent stoichiometric balances and critical response variables like cell growth or monoclonal antibody (mAb) productivity.[216] Important nutrients that impacted specific response variables or cellular phenotypical states could be detected and further translated into experimental designs for validation studies.[216] Given that biological processes are nonlinear in nature, less interpretable but better performing techniques can be utilized through Machine Learning (ML) models (such as Support Vector Machine Regression), especially if the goal is to develop a soft sensor.[217] Such technique has been applied in large-scale bioreactors to predict titer based on features.[172] Lactate metabolism was found to be a Key Process Indicator (KPI) in terms of predicting final protein yield and suggested the importance of controlling glycolytic fluxes during seed train inoculation at large-scale. Tree-based models are also powerful tools that can accurately model nonlinearity among biological systems.[203] Here, data can be segmented into ranges and a decision tree can represent the outcome of a response variable depending on what ranges various input variables are observed at. Since this approach can lead to overfitting, which is an undesirable machine learning characteristic, Random Forest (RF) can be realized. RF techniques are able to run a collection of decision trees in parallel and then return the average prediction of each decision tree. Since it averages out multiple regression trees, RF can overcome the overfitting drawbacks that are observed with single decision trees.[218] Such approaches have also been utilized in the modeling of Raman spectroscopy and shown to be an alternative to the PLS gold standard.[219] Extreme Gradient

Boosting algorithms (XGBoost) are also decision tree-based methods and differ with RF in the sense that trees are constructed sequentially rather than in parallel. This allows for a gradual improvement of the decision tree by continuously re-weighting trees that correctly predict outcomes that were previously poorly modeled by previous decision trees.[203] Since this objective function can be optimized with a learning rate, it can be more prone to overfitting and as such, care must be taken when building and tunning the model. Alternative approaches applied in bioprocessing include multiple linear regression, k-nearest neighbors, Gaussian process regression, classification and regression tree, and ensemble approaches (Gradient Boosting Machine, Adaptive Boosting).[179,203,218]

MVDA tools can help improve our understanding of a process which is key to develop a robust production platform capable of being transferred to different cell types for various target proteins. It is also known that machine learning models can play a role in modeling the interaction between variables and outcomes. This, in turn, provides a pipeline for transferring knowledge gained during the early process development to the manufacturing stage where soft sensor prediction capability will be greatly increased given that the process will be locked in to a specific range within the Design Space (DS); the latter is needed to assure predefined Critical Quality Attributes (CQA).[4] Such tools are part of the Process Analytical Technology (PAT) initiative and an important aspect of Biopharma 4.0 manufacturing which aims to improve process understanding.[4]

The current paper focuses on applying Multiway Principal Component Analysis (MPCA) to examine early process development data with the purpose of expanding process knowledge and improving conditions that maximize SARS-CoV-2 trimeric spike protein production using stable pools instead of stable clones to accelerate development timelines.

Further modeling with Random Forest (RF) demonstrated that endpoint product titer can be predicted utilizing key cumulative and endpoint process values. In-depth analysis of the model demonstrated that improving overall longevity of the cell culture as well as limiting lactate build-up are key variables that if tunned appropriately with process related changes could improve spike protein yields. It is worth mentioning that SARS-CoV-2 spike protein is a difficult-to-express protein due to its structural complexity. Any upstream process development to improve its yield would be valuable to help manufacture this potential vaccine antigen.[3]

Materials and Methods

Stable CHO Cell Pool and Small-Scale Cell Culture Conditions

Four stable CHO cell pools expressing SmT1 trimeric spike proteins namely Wuhan (Wu), Wuhan Tagless (WuTL), Delta (De), and Beta (Be) variant were generated as described previously.[201] Pool cells were thawed and grown in BalanCD CHO Growth A medium (Fujifilm/Irvine Scientific) supplemented with 50 µM MSX (L-Methionine sulfoximine, Sigma-Aldrich) and 0.1% (w/v) Kolliphor P188 surfactant. 125-mL (20 mL working volume) shake flasks without baffles (Corning) were used for cell maintenance and passage. The flasks were shaken at 120 rpm (25 mm orbital diameter) in an incubator regulated at 37ºC, 5% CO_2, and 75% relative humidity. Cells were passaged every 2 or 3 days to keep a maximum viable cell density between 2-3x10^6 cells/mL.

Cell Culture Analytical Methods

Cell density, viability, main metabolites (glucose, lactate, ammonia) were measured utilizing the previously reported methodology.[200,220] Briefly, cell counts were realized with Innovatis Cedex (Roche) or ViCell Blue (Beckman Coulter) automated cell counter using trypan blue dye exclusion assay. Key metabolites such as glucose, lactate, and ammonia were determined using the Vitros 350 Chemistry System (Orthoclinical Diagnostics). Volumetric protein titers were estimated using TGX Stain-free SDS-PAGE gels (Bio-Rad) quantification method.

Bioreactor Fed-batch Process

The bioreactors were seeded at 0.2 x 10^6 cells/mL (Low-seed) or 0.4 x 10^6 cells/mL (High-seed) and cultivated for 17 days in the fed-batch mode. Temperature downshift (from 37ºC to 32ºC) was realized 3 days after seeding. A pH shift (from 7.05 ±0.05 to 6.95 ±0.05) was performed on all batches 2 days after seeding. Induction was conducted with 4-Isopropylbenzenecarboxylate (Cumate, ArkPham), concomitantly with the temperature downshift. Cultures were fed with BalanCD CHO Feed 4 (Fujifilm/Irvine Scientific) and supplemented with glucose to maintain the concentration above 17 mM (3 g/L) for the next sampling point. Samples were taken from the bioreactors on days -3, -2, -1, 0, 3, 5, 7, 10, 12, and 14 days post-induction (dpi) for off-line analysis, while feeding was realized in a bolus dosage from 0 dpi onward. Table 3.1 shows a summary of studied process conditions. The impact of seeding density (low vs. high), cell age through cell passage number (passage 5, 8, 11), MSX concentration (50 vs. 125 µM) on process outputs was examined.

Volumetric Power Input (P/V) indicating the relationship between agitation speed and culture volume was set in a range between 30-40 W/m^3 for the Multifors 0.75L (Infors) and BioFlo 1L (Eppendorf) systems. The final P/V value was decreased due to volume increase with feed events while keeping a same agitation speed. For the BioFlo 10L bioreactor (Eppendorf), a P/V range between 20-80 W/m^3 was explored. A dissolved oxygen (DO) setpoint of 40% (of air saturation) was chosen for the Multifors 0.75L and BioFlo 1L systems while for the BioFlo 10L bioreactor, DO of 40% and 60% were studied. A Kolliphor P188 surfactant concentration of 0.2% (w/v) was used for the Multifors 0.75L and BioFlo 1L whereas a variation from 0.2% to 0.6% (w/v) Kolliphor P188 was investigated in the BioFlo 10L system. In the Multifors 0.75L, micro-spargers (10 µm pore diameter) with an air cap were used in a cascade air/oxygen strategy. Cell passage number and MSX concentration were varied across batches. For the BioFlo 1L, micro-sparger, macro-sparger, and dual sparger composed of a micro-sparger and a macro-sparger were compared. A mix between air caped and no air caped strategies were also studied within this sub dataset. The BioFlo 10L studies encompassed a variety of process conditions such as air caps, sparger type, agitation rate, and number of impellers. It must be noted that for bioreactor runs that employed dual sparger configuration, the macro-sparger sent only air (air cap (AC)) while the micro-sparger injected flow of both oxygen and carbon dioxide as needed.

Table 3.1. Bioreactor Production Process Conditions

Bioreactor System	Pool	Seeding Density (10^6 cells/mL)	Cell passage number	MSX (µM)	P/V range (W/m^3)	DO (%)	Kolliphor P188 (%, w/v)	Sparger	Aeration strategy	Number of impellers
Multifors 0.75L (Infors)	Wuhan (Wu) Delta (De) Beta (Be) WuhanTL (WuTL)	Low: 0.2 High: 0.4	5 8 11	50 125	40-30	40	0.2	Micro	Air cap (AC) with Air/O_2 cascade	2
BioFlo 1L (Eppendorf)	Wuhan (Wu)	High: 0.4	5	50	40-30	40	0.2	Macro Micro Dual	Air cap (AC) No air cap	1 1 1
BioFlo 10L (Eppendorf)	Wuhan (Wu)	High: 0.4	5	50	80-20	40 60	0.2 0.6	Macro Micro Dual	Air cap (AC) No air cap	1 2 2

Dataset Structure and Batch-wise Unfolding Method

The dataset is made up of 59 batches (productions). It is worth mentioning we kept the terminology “batch” to indicate a production run as this term is widely used in data science. All productions (batches) were performed in fed-batch mode. Of the total 59 productions, 38 were conducted in the Multifors 0.75L, 13 batches in the BioFlo 1L, and 8

batches were realized in the BioFlo 10L. For the Multifors 0.75L parallel benchtop bioreactor platform, 14 batches were run with Wuhan pool, 12 batches were realized with Delta pool, 6 batches with Beta pool, and 6 batches with Wuhan Tagless pool. The other two systems (BioFlo 1L and 10L) used exclusively the Wuhan pool. Table 3.2 shows the variables considered in the MVDA. Product titer was not listed given that the Wuhan pool sub dataset has only endpoint titers while titer evolution profiles were available for Delta, Beta, and Wuhan Tag-less pools. It was decided to exclude titer as a variable to keep analysis consistent across pools. However, during the analysis of variable relationship, endpoint titers were considered for result discussion. Viability, viable cell density (VCD), residual glucose, lactate, and ammonia were measured following the same schedule as mentioned in the process conditions section above. Calculated values such as Integral Viable Cell Concentration (IVCC), glucose consumed per day, and cell specific metabolic rates (glucose, lactate, ammonia) were estimated using the same procedure across all batches. Online data from the bioreactor runs were also added into the dataset. pH, cumulative volumetric base addition, cumulative oxygen flow, and cumulative carbon dioxide flow were calculated into daily averages such that direct comparison with the sampling day's data could be made. The cumulative gas flows were normalized with respect to the bioreactor volume so as it represents the cumulative gas volume per liquid volume per minute (VVM) for each gas. Batch-wise unfolding of the data (Figure 3.1) was realized in such a manner, so each row represents a batch, and each column represents a variable at a given sampling timepoint (-3 to 14 dpi, Day post-induction).

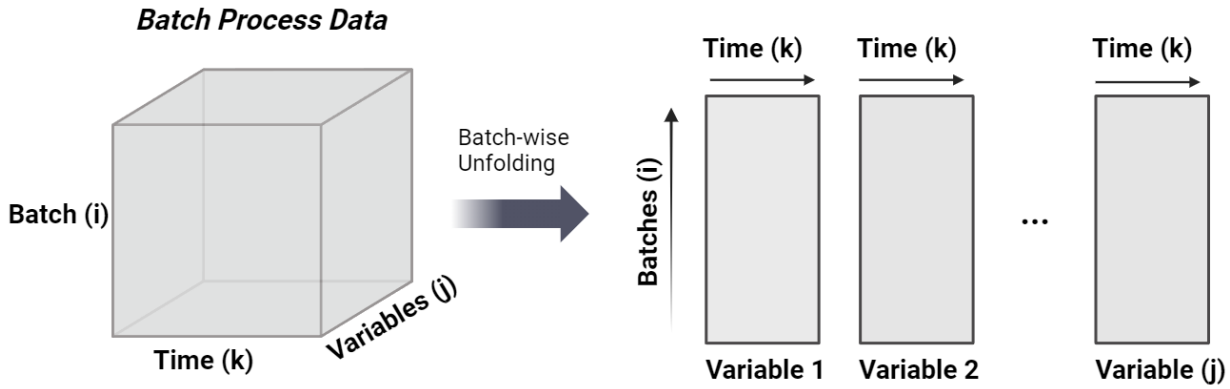


Figure 3.1. Batch-wise unfolding from 3D dataset to 2D data arrangement.

Each row represents a batch (i). Time (k) and variables (j) are presented as columns and arranged in a cyclical mode such that every variable is a block from time 0 to k.

For this unfolding method, each experiment (batch) becomes a score in the principal component (PC) plot effectively collapsing the time dimension while there is a loading value for each sampling point for every variable. Loadings (Annex 1S-18S) provide information about how each variable contributes to the formation of the principal components and can help interpret the data structure and relationships among variables within the dataset. Consequently, loading plots can help understand what variables are driving the position of a given score (batch experiment) in the principal component (PC) axis. Loadings can be positive or negative representing the direction of the relationship (positive loadings indicate positive correlations between the variable and the principal component while negative loadings represent negative correlations). Magnitude of the loading values are also relevant as they represent the strength of the relationship. Larger values suggest more significant influences of the variable on the principal component. It must be noted that since all the runs were performed to investigate specific process

parameters (cell passage number, seeding density, MSX concentration, pool type), the objective of the MVDA is to infer global trends within the dataset and determine KPI (Key Process Indicator). This is especially important since there is a lack of published information regarding inducible pools and much less large dataset analysis of such production platform. To the best of our knowledge, this is the first report demonstrating the benefit of MVDA to help interpret the data obtained during the early bioreactor process development of an inducible stable CHO pool.

Table 3.2. Variables considered in Batch-wise Multiway PCA (MPCA).

Offline Measurements	Cell Growth	IVCC, cells*day/mL
		VCD, cells/mL
		Viability, %
	Metabolites	Residual Glucose, mM
		Lactate, mM
		Ammonia, mM
		Total Consumed Glucose (TCG), mM
	Cell specific metabolite rates	Glucose (qGlc), pmol/cell/day
		Lactate (qLac), pmol/cell/day
		Ammonia (qAmon), pmol/cell/day
Online Environmental Continuous Measurements	Gas Sparging	Oxygen, mL/min
		Air, mL/min
	pH control	pH profile
		Base Addition Volume, mL
		Carbon Dioxide, mL/min

MVDA approach, software, and package

Pre-processing of online data and analysis was carried out using **R** programming language. The **mdatools**[221] package was utilized for PCA while the **caret**[222] package was employed to build the Machine Learning (ML) models. Multi-steps were used to treat the data. First, Savitzky-Golay filtering[223] was carried out on pH data before calculating daily

averages for bioreactor online data (pH, oxygen flow, carbon dioxide flow, base addition) so as to filter out noise sensor data. For the remaining variables (base, oxygen and carbon dioxide sparging), root cause analysis was carried out to determine if abnormal behaviour could be explained by sensor faults and thus excluded from analysis. Second, daily averages and sampled variables were gathered in Excel spreadsheets (Microsoft) and arranged such that they matched in the time dimension. Daily averages of the online data were taken and for the sparge rates, since they are highly variable (on/off flow of pure oxygen and carbon dioxide), cumulative of said variables (oxygen and carbon dioxide flow) were taken so as to compare trends. Lastly, of the daily online bioreactor values, only values that matched with sampling days were taken so as to not bias the dataset (in the 17-day fed-batch process, there would be 10 offline sampling data values for each variable while there would be 17 values for each online variable). For the online variables, values starting from -2 dpi (culture day 1) were included since base addition and oxygen sparging are nonexistent in the first day of culture (-3dpi = culture day 0). Input data (IVCC, VCD, viability, lactate, ammonia, cumulative glucose consumed, residual glucose, base addition, cumulative oxygen sparged, cumulative carbon dioxide sparged, pH, qLac, qGlc, qAmon) were arranged in batch-wise unfolding so as to carry out PCA where score values and principal component axis generation can be considered the output for the purpose of data visualization. Loading plots (Annex S1-S18) were utilized to determine the driving factors behind score plot distributions.

Once correlations among variables were better understood through Multiway PCA, key parameters describing the behavior of each batch (IVCC, endpoint viability, max lactate, endpoint lactate, endpoint ammonia, cumulative glucose consumed, endpoint

residual glucose, total base addition, cumulative oxygen sparged, cumulative carbon dioxide sparged, endpoint pH, pool type, passage number) were calculated and used as input features to predict endpoint titers (output) using four modeling methods as described below. MSX concentration was excluded from the model because the extra MSX supplementation did not show impact as demonstrated in the Results and Discussion section below.

Support Vector Machine (SVM), Partial Least Square (PLS), Random Forest (RF), and Extreme Gradient Boosting (XGBoost) were utilized as regression methods to relate the input variables (features) to the output variable (titer). The total dataset used to generate the regression models was made up of 50 batches. It was split into training (80%) and test (20%) sub datasets. Model metrics were obtained for both training and test datasets. Each model was subjected to the same hyperparameter tuning strategy. Adaptive resampling of the tuning parameter grid was realized in such a way that the random search of hyperparameters is concentrated on values that are in the neighborhood of the optimal parameters by discarding settings judged sub-optimal. To assess the performance of the training regression models, bootstrapping was conducted using the training and test dataset separately. For each iteration (i = 100), a bootstrap sample was created by resampling the imputed sub dataset (training or test) with replacement. The regression model was then used to predict the target variable for the bootstrap sample. Key evaluation metrics including Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and the coefficient of determination (R^2) were evaluated at each iteration. By repeating this process multiple times, a distribution of these metrics was obtained, providing insights into the performance and variability of the pre-trained regression models on both the training and

test datasets. Best performing model based on RMSE, MAE, and R^2 was then analyzed with shapely value dependency plots to obtain more information about the impact each feature has on the prediction outcome and at what range of values said features had a positive or negative impact on prediction. The **R caret** package was utilized to build the models, **boot** package[224,225] was used to get bootstrapped statistics, and **fastshap** package[226] was employed to construct the Shapely value dependency plots.

Results and Discussion

Seeding cell density impact

To investigate the impact of seeding cell densities (SCD), two densities (low: 0.2×10^6 cells/mL and high: 0.4×10^6 cells/mL) were evaluated within 14 batches. Our previous data (not shown) demonstrated that higher density at induction contributes to accelerate production pace thus shortens process duration without affecting final titers. However, the superior seeding density bound needs to be controlled mainly because the existing feed regimen has been designed to support a specific range of viable cell density. Post-induction cell overgrowth due to high induction cell density may lead to premature cellular decline probably due to nutrient limitation occurred at high cell biomass.

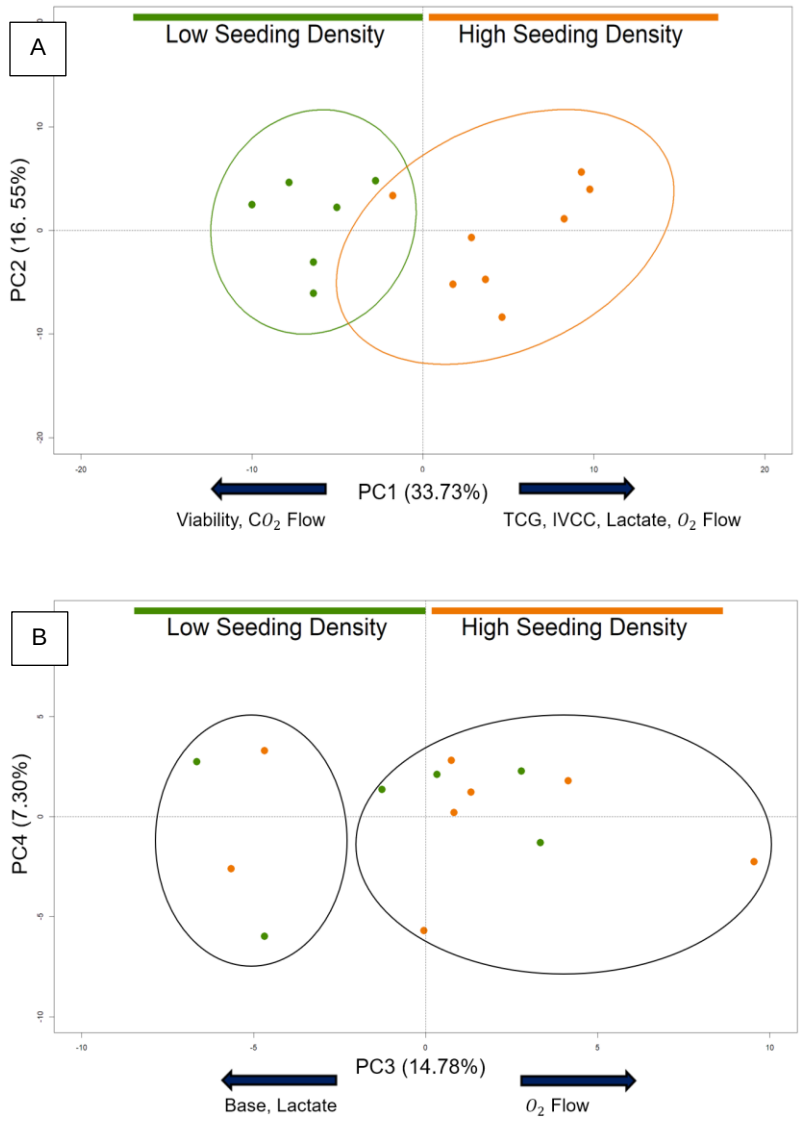


Figure 3.2. Principal Component (PC) scatter plots for Wu pool in Multifors 0.75L bioreactors.

A) PC1 vs PC2 scatter plot showing seeding density impact on the distribution of batch experiment scores. High seeding density batches are colored in orange while green indicates low seeding density batches, ellipses represent 95% Confidence Interval (CI), B) PC3 vs PC4 scatter plot showing high and low titer batches spread between both seeding densities. Left ellipse indicates low yield productions while right ellipse relates to high yield runs.

From Figure 3.2A, it is clear that two distinct clusters are created. The orange cluster represents productions seeded at high density while bioreactors seeded at low density are distributed within the green cluster. When evaluating what variables drive this phenomenon in the principal component 1 (PC1) axis, it is found that batches located on the positive PC1 axis have higher lactate accumulation (27.8 mM peak lactate vs. 23.7 mM peak lactate), increased peak VCD during the 3-day growth phase ($4.35x10^6$ c/mL vs. $2.98x10^6$ c/mL), and increased oxygen requirements (16.77 mL/min vs. 10.52 mL/min). These variables are intimately related with increasing biomass, which explains the segmentation based in seeding density. It is worth mentioning when the endpoint titers and viabilities are compared, no clear relationship with respect to seeding density can be attributed (Table 3.3).

Table 3.3. Endpoint product titers and viabilities

	Product Titer, mg/L	Viability, %
High Seeding Density (n=8)	1010±216.2	91.19±3.42
Low Seeding Density (n=6)	937±141.4	94.7±1.65

Figure 3.2B shows that some productions differ in the principal component 3 (PC3) axis. This component is primarily driven by addition of base to regulate pH (negative PC3) and increasing oxygen sparging flowrate (positive PC3). Given that links between oxygen consumption and protein production have been investigated[227], the final protein production of the batches at opposite ends of the PC3 axis was evaluated. It was determined that the points lying on near the origin and the right-hand side of the PC3 axis had an average

protein expression of 1064±137 mg/L (n=10) while the points lying on the left side of the PC3 axis had a final protein expression of 765±83 mg/L (n=4). Since the variation in protein expression happened in both high and low seeding density conditions, a deep dive into the time series data was required to explore the root cause of this variation. As such, time profiles were plotted (Figure 3.3). High performers (blue) were defined as batches that were above average in terms of final protein expression whereas low performers (red) were assigned as batches that had below-average protein expression.

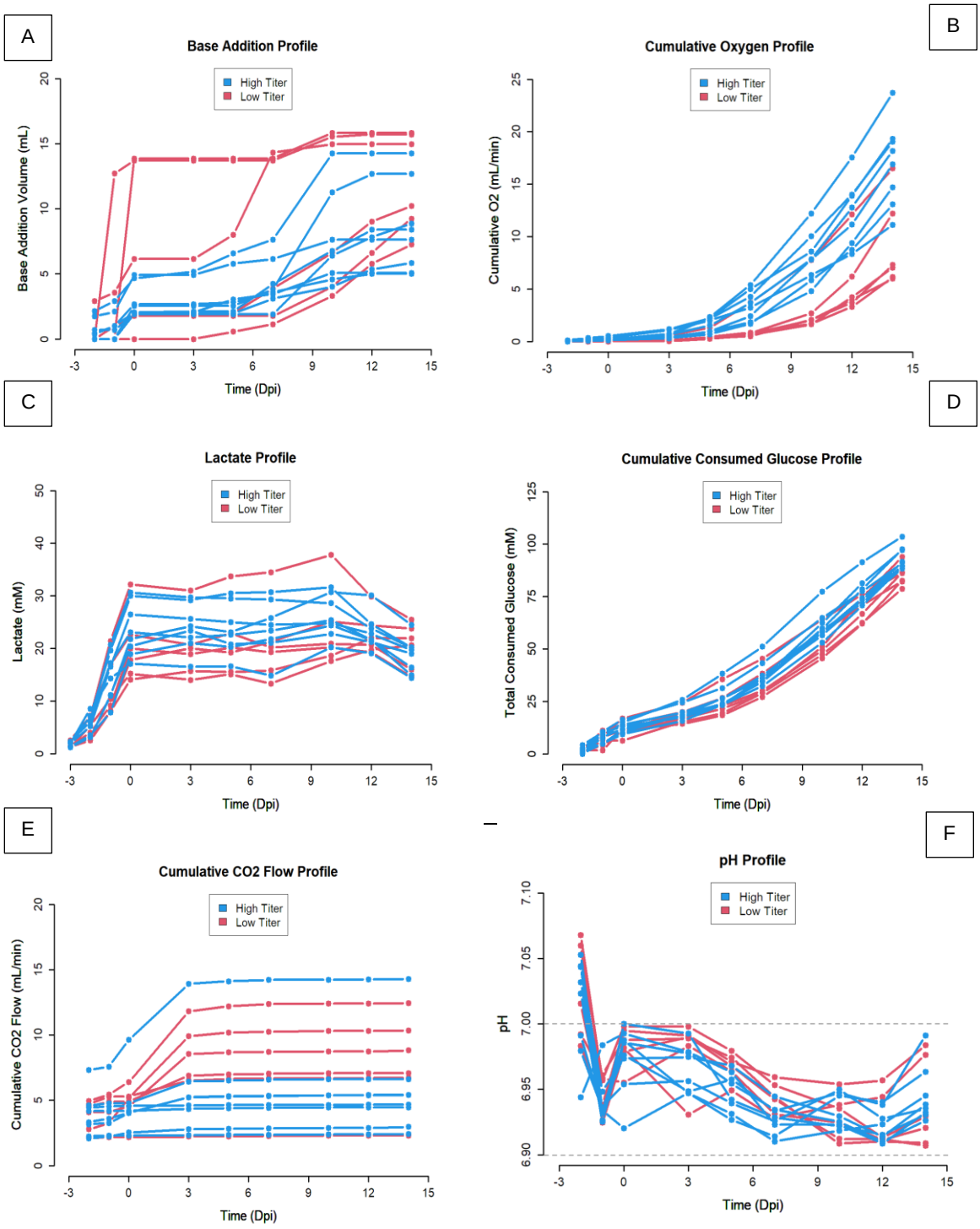


Figure 3.3. Color coded time profiles for the Wuhan pool experiments (n=14) performed in the Multifors 0.75L system.

A) Base Addition Volume, B) Cumulative Oxygen Sparging, C) Lactate Profile, D) Total Consumed Glucose (TCG), E) Cumulative Carbon Dioxide Sparging, F) Daily pH profile.

It is clear that high performing batches (blue) required generally lower base addition (Figure 3.3A) and higher oxygen sparging (Figure 3.3B) compared to low performing batches. It is unclear why two low yield batches displayed elevated base volume at the beginning of culture. The large sudden addition of base for these two cultures may have been caused by an error in priming the base lines for proper pH control at the culture start. This, in turn, could lead to an overly aggressive response from the PID controller. It was estimated that the total volume inside the base line is only 2 mL and thus total base volume uncertainty is near 12%. This low amount of base volume addition uncertainty was not considered to be impactful enough to merit excluding the two batch runs from further modeling. When evaluating the corresponding lactate profiles (Figure 3.3C), it is hard to conclude that lactate metabolism alone is enough to explain the variation in protein expression even though high titer cultures displayed less lactate production variability. Figure 3.3D shows that high performers had higher glucose consumption. It can be postulated that high glucose demand relates to higher levels of protein expression through TCA cellular respiration activity as evidenced by the increased oxygen requirements (Figure 3.3B). Higher overall CO_2 sparging rates were observed with low performing batches (Figure 3.3E). All the cultures were performed with pH setpoint±deadband of 7.05±0.05 (covered range: 7.0-7.1) until -1 dpi and pH was downshifted to 6.95±0.05 (range: 6.9-7.0) from -1dpi to 14 dpi. When looking at the daily average pH profile (Figure 3.3F), it can be discerned that the worst performing cultures had the largest deviations within the deadband, with values near the 6.9 or the 7.0 threshold. Increasing the pH deadband to cover a larger pH range (example: 6.8-7.2) could potentially diminish addition of carbon dioxide or base thus eventually improve process robustness.

Impact of bioreactor culture system

The Wu pool in the Multifors 0.75L system was compared to a dataset resulted from the BioFlo 1L system (Figure 3.4). The BioFlo 1L dataset centered around exploring impact of aeration conditions. The main goal was to find the appropriate aeration strategies that would re-create the results observed in the established Multifors 0.75L process as part of a technology transfer project.

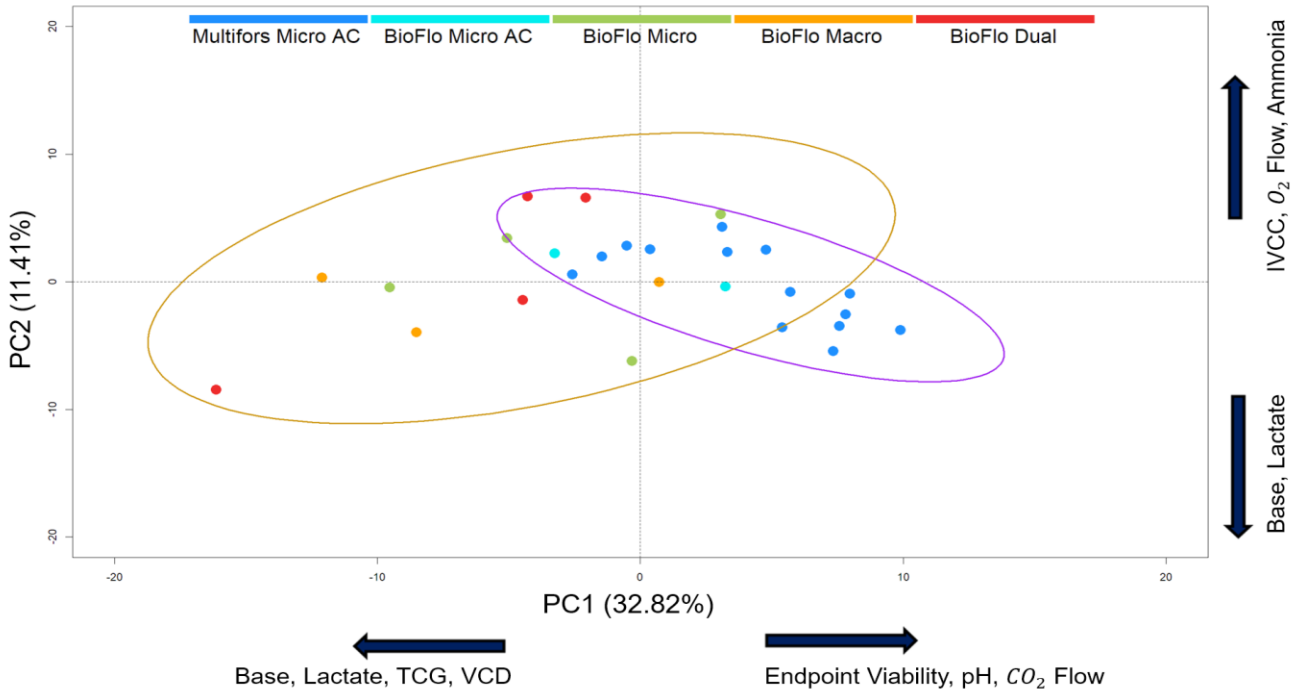


Figure 3.4. Principal component (PC) scatter plots of Wu pool in two bioreactor systems (Multifors 0.75L and BioFlo 1L).

PC1 vs PC2 scatter plot showing the same pool (Wuhan) with different aeration strategies. Orange batches (BioFlo Macro) employ macro-sparger in BioFlo 1L, green batches (BioFlo Micro) use micro-sparger in BioFlo 1L, red batches (BioFlo Dual) utilize dual sparger configuration in BioFlo 1L, blue batches (Multifors Micro AC (Air Cap)) employ

micro-sparger with air cap for the Multifors 0.75L system, and ocean blue batches (BioFlo Micro AC (Air Cap)) deploy air caped micro-sparger in the BioFlo 1L system. Brown ellipse represents the BioFlo 1L system (13 batches), and purple is assigned to the Multifors 0.75L system (14 batches). Ellipses represent 95% Confidence Interval (CI).

Figure 3.4 shows that the negative PC1 axis is strongly influenced by high levels of total consumed glucose (TCG), base addition, and lactate accumulation. Alternatively, the positive PC2 axis is driven by increased IVCC, oxygen sparging, and ammonia. For example, when comparing the left most culture with the right most culture, it is found that the left most culture shows a 2.54-fold increase in cumulative glucose consumption before temperature shift, large lactate accumulation (2.52-fold increase in endpoint lactate), 5.2-fold increase in base addition, and initial large cell growth (2.69-fold increase in IVCC before temperature shift). Conversely, the positive PC1 axis is influenced strongly by high CO_2 sparging, high growth phase pH, and high viability specifically towards the end of the batch (24% higher endpoint viability for the culture in the bottom right when compared to the culture in the bottom left). Importantly, the far left-bottom batch which was a dual sparger culture had a base addition of 42.8 mL which was 2.8-fold higher than the average base addition in the BioFlo 1L dataset (14 batches). Interestingly, ammonia accumulation in this left-bottom batch was lower (4.0 mM) when compared to the average ammonia accumulation of the BioFlo 1L dataset (5.5 mM). This points at the idea that batches towards the right side of the graph had less lactate accumulation and better longevity. This viability dependence explains why the negative PC1 axis is driven by large initial cell growth and high glycolysis/lactate metabolism as the cultures are unable to sustain the large biomass increase and are followed by a premature decrease in viability. It is clear that the BioFlo 1L system has a wider spread in the score plot when compared to the Multifors

0.75L system (evidenced by the spread in the confidence intervals of each reactor system). This makes sense given that the BioFlo 1L dataset focused on testing a variety of sparging and aeration strategies whereas the Multifors 0.75L dataset studied the impact of seeding densities, MSX concentration, and cell passage number under the same hydrodynamics conditions in regard to mixing (volume, agitation, and aeration). Since cell passage number's effect was determined to not be an important factor for the Wu pool nor MSX increased concentration after induction, seeding density is the main factor driving the variation of the overall spread for the Wu pool cluster. Importantly, this implies that aeration strategies have a strong impact on process outcomes given that the batch spread in the scatter plot from BioFlo 1L cultures was larger when compared to the Multifors 0.75L batch distribution extent.

From Figure 3.4, it is also worth noting that the five Multifors 0.75L cultures that do not overlap (and a sixth batch that exists at the boundary) within the 95% confidence interval ellipse of the BioFlo 1L system are low density cultures. These 6 batches are the only low-density cultures of the Multifors 0.75L dataset. Since seeding density was observed to have an important impact on culture outcome (Figure 3.2A), these six cultures are shown different from the rest of high-density batches. Interestingly, within the 95% confidence interval ellipse of the Multifors 0.75L cluster, seven batches from the BioFlo 1L system are found. These cultures use a micro-sparger to provide oxygen to the cells. This may suggest that sparging oxygen with micro-sparger best recreates the hydrodynamics environment of the Multifors 0.75L cultures; the latter use uniquely a micro-sparger. In regard to aeration strategy (with or without air cap), large variation was found in non-air caped batches, both with micro- and macro-sparger. Dual sparger also

shows variability that was mostly driven by fluctuations in lactate metabolism and its impacts (base addition). It can be postulated from the process development data that multivariate tools can be deployed to help pick aeration strategies that diminish variability and translate process conditions across bioreactors to facilitate scale-up.

Scale-up from the small-scale bioreactors (Multifors 0.75L & BioFlo 1L) to a BioFlo 10L system was next explored as shown in Figure 3.5.

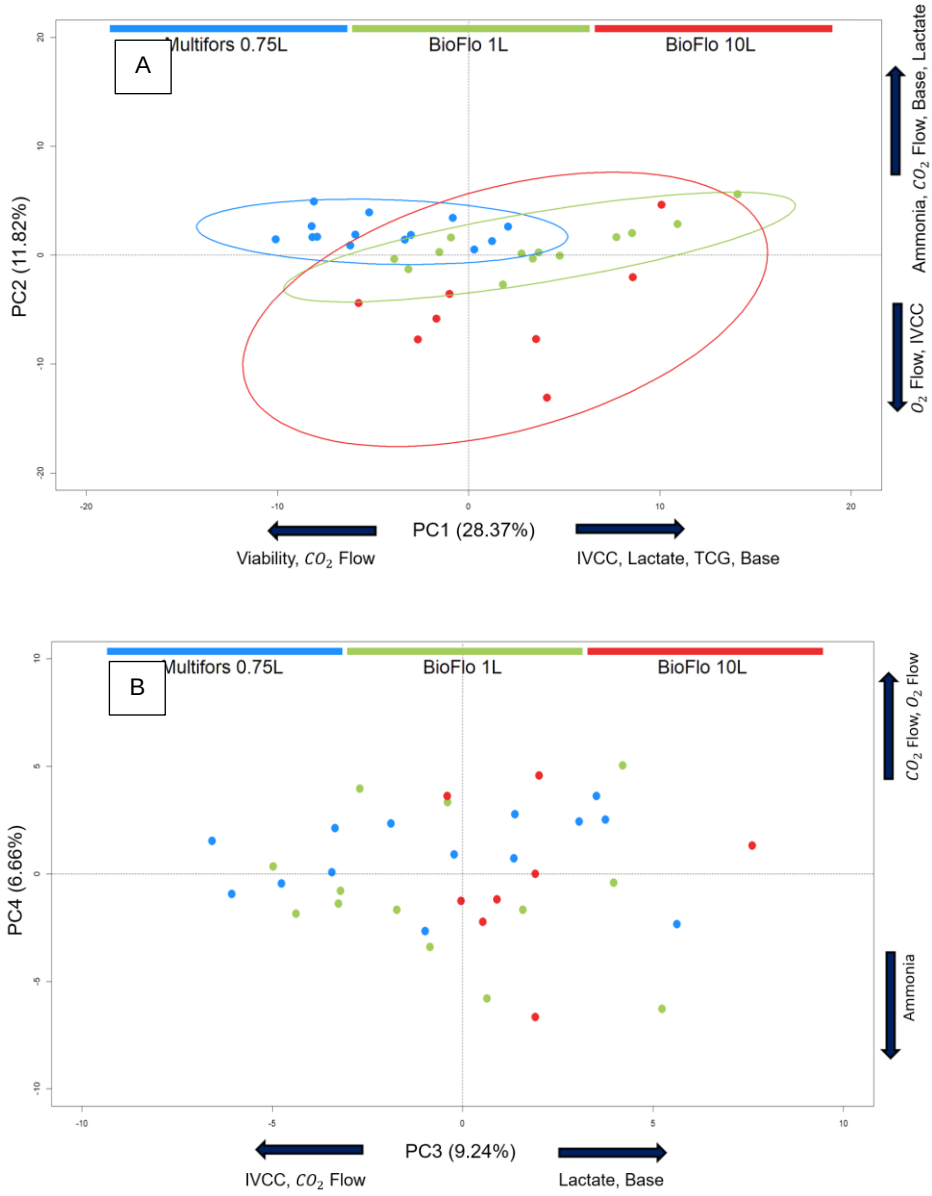


Figure 3.5. Principal component (PC) scatter plots of three bioreactor systems (Multifors 0.75L, BioFlo 1L and BioFlo 10L).

A) PC1 vs PC2 scatter plot and 95% confidence ellipses of Multifors 0.75L, BioFlo 1L, and BioFlo 10L. Ellipses represent 95% Confidence Interval, B) PC3 vs PC4 scatter plot of Multifors 0.75L, BioFlo 1L and BioFlo 10L.

Blue, green, red batches indicate the Multifors 0.75L, BioFlo 1L, and BioFlo 10L, respectively.

From Figure 3.5A, it can be observed that the spread of the BioFlo 10L system is significantly higher than the BioFlo 1L or the Multifors 0.75L. This is due to the fact that data was collected under varying experimental conditions such as differences in dissolved oxygen conditions (40% vs. 60%), sparger (macro, dual, micro), agitation speed, air cap, initial volume (5L vs. 7.5L), and impeller configuration (1 impeller vs. 2 impellers) as described in Table 3.1. The PC1 axis is driven by IVCC, lactate, total glucose consumption (TCG), and base addition in the positive direction while carbon dioxide and viability are the main driving factors in the negative direction. The low-density Multifors 0.75L cultures are centered in the left most negative axis whereas the high-density cultures are near the origin. BioFlo 1L and BioFlo 10L have equal seeding densities but varying hydrodynamics conditions. This difference in spread emphasizes the idea that the impact of hydrodynamics is much more important than other initial variables such as seeding densities (the spread of the Multifors 0.75L dataset is driven by variance in seeding density, Figure 3.2). The negative PC2 axis is driven by oxygen sparging and IVCC while the positive PC2 axis is strongly driven by ammonia, carbon dioxide, base addition, and lactate accumulation. There was a strong difference in sparging strategies between the 10L system and the two small-scale bioreactors that can explain the outliers in the 10L cluster. Four BioFlo 1L batches fall within the 95% confidence interval of the Multifors 0.75L data which again underscores the idea that batches with similar hydrodynamics (micro-sparger) across different bioreactors can be evaluated with MVDA. When examining the PC3 versus PC4 scatter plot (Figure 3.5B), one outlier (right-most batch) from the 10L system is evident. The positive PC3 axis is driven by increased lactate accumulation and base addition while the negative PC3 axis is strongly driven by IVCC and carbon dioxide sparging. The right-

most batch outlier utilized a single impeller with a macro-sparger which probably reduces oxygen transfer capability, and as such, large amounts of oxygen sparging was required while at the same time having issues maintaining its 40% DO setpoint. This suboptimal DO control could have impacted cell metabolism as it produced unusually high amounts of lactate (peak lactate was 98 mM) and consequently, required a lot of base addition (313.6 mL). This unusual behavior concluded in a 390 mg/L titer yield which is below the 732 mg/L average that the 10L bioreactors had. In the literature, similar adverse behaviours induced by inadequate DO controls have been reported.[228]

Cell age effect

As it was mentioned above, the expression stability of pools needs to be evaluated to ensure a commercial-scale production which requires an expanded seed train. Three pools stability was evaluated in the Multifors 0.75L system as part of an early process development objective (Figure 3.6).

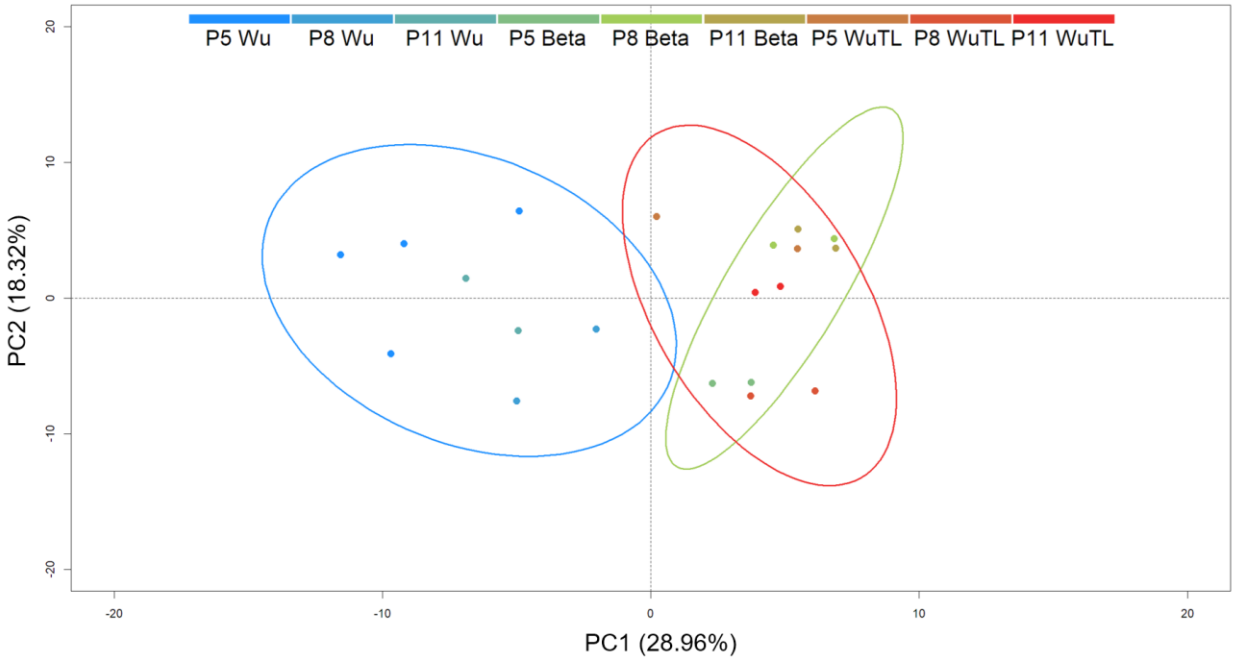


Figure 3.6. Scatter plot of Principal Component PC1 vs. PC2 of different cell passage numbers used for productions of three pool variants in Multifors 0.75L bioreactor.

Batches are color-coded based on their passage number and pool type (Wu, Beta, WuTL). Ellipses represent 95% Confidence Interval.

For the Wuhan pool, final spike protein yield remains high even at increasing passage number (P5 = 11 generations: 908±161.91 mg/L (n=4); P8 = 20 generations: 1252±158.39 mg/L (n=2); P11 = 31 generations: 971±250.31 mg/L (n=2)). Two-tailed t-tests (Table 1S) for the Wuhan pool comparing passage number impact show no significant difference between the three cell passage numbers (p value < 0.05). The Beta pool's final endpoint protein yield displays however a gradient behaviour such that P5 (455± 91.21 mg/L, n=2) > P8 (370±18.66 mg/L, n=2) > P11 (341±7.41 mg/L, n=2). Maximum lactate was observed to be different such that P5 had a higher peak lactate (50.15 mM) when compared to P8 (33.9 mM) and P11 (32.3 mM). Endpoint ammonia was higher in P5 (8.28 mM) compared to P8 (5.97 mM) and P11 (6.12 mM). Endpoint IVCC also demonstrated

cell age variance such that P5 ($1.14x10^8$ cell*day/mL) < P8 ($1.56x10^8$ cell*day/mL) < P11 ($1.69x10^8$ cell*day/mL). Interestingly, when detailing total oxygen sparged, it is clear that the cumulative average flow rates follow the IVCC trend such that P5 (18 mL/min) < P8 (26.58 mL/min) < P11 (27.27 mL/min). On the other hand, even WuTL endpoint protein production did not seem to be negatively impacted (P5 = 838.5±48.8 mg/L (n=2), P8 = 751.5±60.1 mg/L (n=2), P11 = 908±106.06 mg/L (n=2)), an inverse relationship between base addition and oxygen sparging was found. Passage 5 and passage 11 which had the higher protein production also had higher oxygen sparging and less base addition when compared to passage 8. A similar conclusion is reached when evaluating above-average batches and below-average batches in terms of final yield. The above-average (832 mg/L) batches required more oxygen sparging and needed less base addition. The difference was driven by increased lactate accumulation in the low performing batches. The best performing batches also had on average higher total glucose consumption demonstrating that the high metabolic activity was shared across glycolysis and other metabolic pathways that do not end in lactate accumulation.

When evaluating two-tailed t-tests (Table 2S) on key metrics, it can be said that although the passage number did not have statistical impact on final titer concentration, there was passage number related variation with respect to total glucose consumption for the WuTL pool. Similarly, although passage-related statistical significance was found in endpoint lactate and endpoint IVCC, no evidence for statistical impact on endpoint titers was determined in the Beta pools.

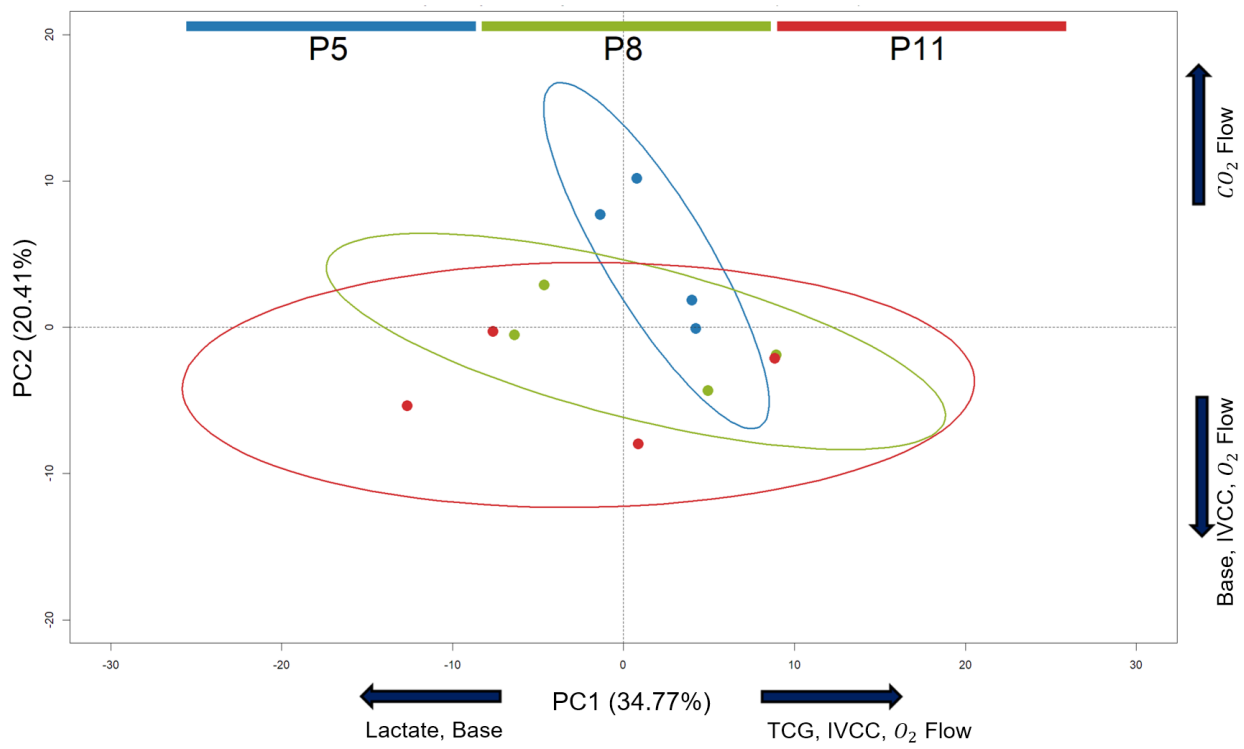


Figure 3.7. PC1 vs PC2 scatter plot showing cell age impact on Delta pool.

The blue, green, and red color represent batches performed with cell passage number P5, P8, and P11, respectively. Ellipses represent 95% Confidence Interval.

In stark contrast, when evaluating cell age impact on the Delta pool (Figure 3.7), passage number seems to have an impact on the spread of the scores such that higher passage number batches have more spread. Interestingly, it was determined from the loadings (Figure S13) that oxygen sparging and base addition are inversely related, while lactate and base are directly correlated. This is because spread in the PC1 axis is driven by base and lactate in the negative direction while glucose consumption, IVCC, and oxygen flow increase in the positive direction. Additionally, the positive PC2 axis is mostly driven by carbon dioxide sparging whereas the negative direction is driven by base addition. Since the scores did demonstrate a cell passage number's dependence, the average protein concentration for each passage was calculated. Titer of 600±112 mg/L, 361±31 mg/L, and

398±92 mg/L were estimated for P5 (n=4), P8 (n=4), and P11 (n=4), respectively. Additionally, endpoint ammonia showed cell age related behaviour since P5 (6.88 mM) had significantly lower ammonia accumulation when compared to P8 (8.79 mM) and P11 (10.82 mM). Table 3.4 shows a two-tailed t-tests comparing passage number impact. Only endpoint titers are statistically different. Taken together, it can be concluded that culture age does indeed play a role in culture outcomes for the Delta pool. This pool likely becomes unstable across time and thus considerations will be needed when scaling-up such as limiting the cell passage to 5 to preserve high titers.

Table 3.4. Two-tailed T-test of cell age's impact on different key variables for Delta pool

Values with asterisk * represent conditions in which statistical significance was found (p value < 0.05).

Delta pool	**P5 vs P8**	**P5 vs P11**
Base Volume	0.37	0.17
Endpoint Lactate	0.83	0.33
Total Oxygen Sparging	0.11	0.82
Endpoint Titer	0.01*	0.03*
Total Glucose Consumption	0.66	0.63
Endpoint IVCC	0.48	0.56
Max Lactate	0.53	0.27
Endpoint Ammonia	0.24	0.06

MSX concentration impact

The studied CHO pools in this work express Glutamine Synthetase (GS) gene implying MSX is required during the cell line generation process. Tian[229] showed that increasing the MSX concentration to certain level can lead to an improved overall protein yield. In this study, MSX supplementation up to 125 µM was investigated across the Wu, Delta, WuTL pools to determine if the increased selection pressure at the moment of induction has a positive impact on final protein yield.

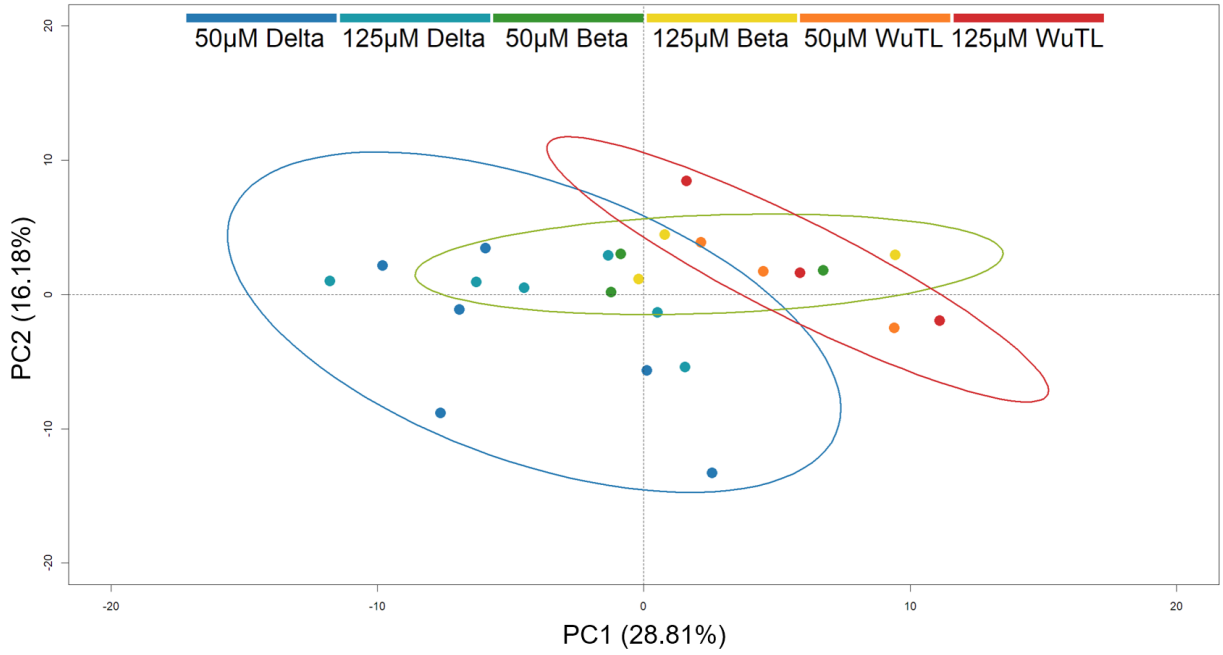


Figure 3.8. PC1 vs PC2 scatter plot showing the impact of MSX addition during induction on various pools.

Batches are colored with respect to pool and MSX addition. Blue and ocean green batches represent Delta pools with no extra MSX addition (50 µM Delta) or with addition of 75 µM MSX at induction (125 µM Delta in total), respectively. For Beta pools, green color represents the productions without extra MSX addition at induction (50 µM Beta) while

yellow batches are assigned to Beta pools with 125 µM MSX. Wuhan Tag-less (WuTL) productions were conducted without extra MSX addition at induction (50 µM WuTL, orange color) compared to red batches with 125 µM MSX. Colorful ellipses show 95% Confidence Interval for respective studied conditions.

As it can be seen from Figure 3.8, no discernible clustering is evident when analyzed based on MSX concentration. This suggests that extra addition of 75 µM MSX during induction (day 3 post-seeding) does not provide a clear impact. The visual clustering that occurs in the scores plot as evidenced by the 95% confidence ellipses is driven by differences in pool behavior. To further determine the impact of extra 75 µM MSX addition, two-tailed t-tests (Table 3S) were realized. No statistical differences were found in almost all the key variables between the base level MSX+ (50 µM MSX) and extra MSX ++ (125 µM MSX) conditions at the exception of endpoint ammonia ($p = 0.01 < 0.05$). The Delta cluster has the most spread profile probably due to the fact that this specific pool seemed to be significantly impacted with cell increasing age (Figure 3.7). Given that the Delta pool dataset had enough data points, an additional two-tailed t-test (Table S22) was conducted. It was determined that even if passage number is accounted for, MSX has no statistical impact on general cell culture variables except for the passage 11 batches; 35 mL of base was added with 50 µM MSX productions (MSX+) while only 8.9 mL of base was needed for 125 µM MSX bioreactors (MSX++). Taken together, it can be postulated that additional MSX supplementation at induction has no added value and thus can be avoided in the future when working with such pools.

Global pool analysis

Once every pool was analyzed separately, the four pools were then examined together to better understand pool related clustering and behavior that may not be self-evident when investigating individually.

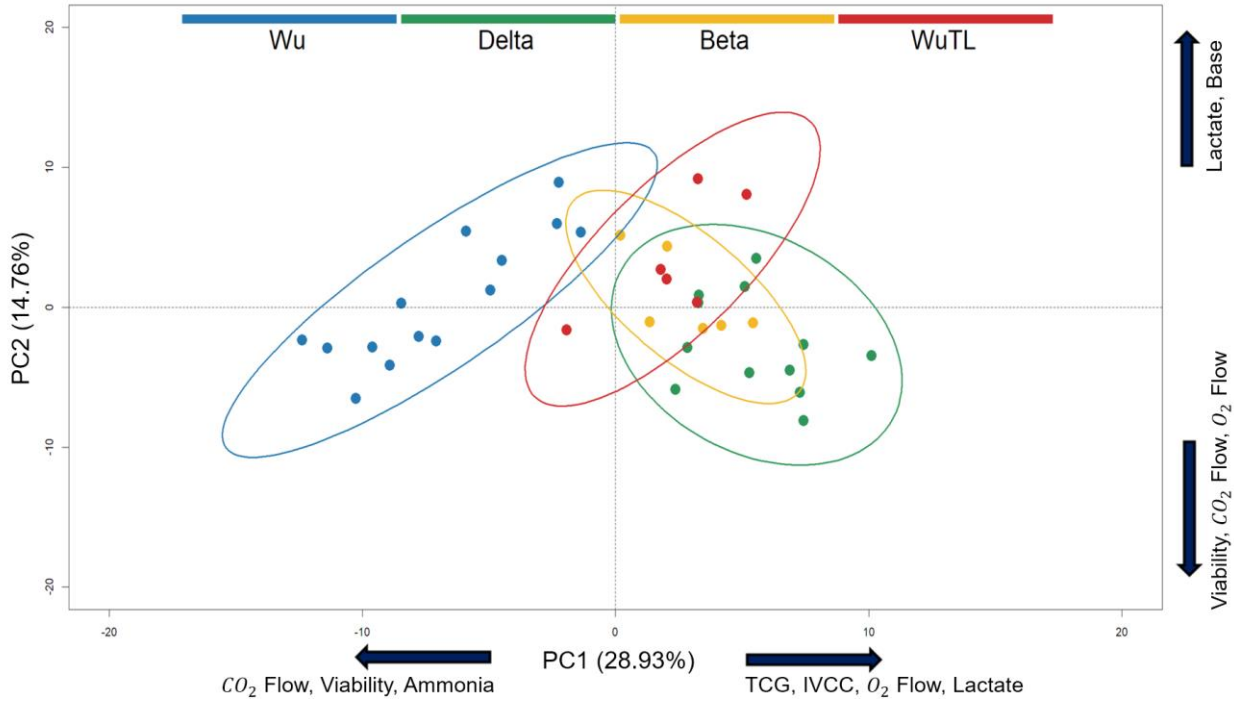


Figure 3.9. PC1 vs PC2 scatter plot of all the Multifors 0.75L experiments encompassing 4 pools (37 batches).

Wu pool colored in blue, Delta pool colored in green, Beta pool colored in yellow, and WuTL pool colored in red. Colorful ellipses represent 95% Confidence Interval (CI) for each respective pool.

From Figure 3.9, it is possible to discern that the pools tested in the Multifors 0.75L system exhibit distinct behaviour as evident by the extent in the distribution and spread variation in the PC1 vs PC2 graph. The Wuhan variant pool has a large spread in the PC1 axis that is driven by its high density versus low density seeding batches. This is evident

from the loading plots (Figure S17) in which it is clear that the driving factors are IVCC, lactate, total glucose consumption, and oxygen sparging requirements. It is worth mentioning that the Wu pool dataset is split into high-density and low-density seeding batches. Such a segmentation based on biomass is captured. On average, Delta pools had 1.7-fold higher IVCC when compared to the Wu pool which again explains how the Delta pool is centered in the positive PC1 axis while the Wu pool is spread out in the negative direction. This further demonstrates that every pool had different growth patterns and, in consequence, had different specific protein production characteristics. Delta pool had the most overall growth but provided the second poorest endpoint titer. The PC2 axis is heavily driven by lactate and base addition in the positive direction while oxygen sparging, carbon dioxide sparging, and viability in the negative direction. Even though most pools demonstrated an inverse relationship between base addition and oxygen sparging individually, it is not possible to carry out this analysis across pools. However, between batches within the same pool, analyzing these critical attributes does seem to hold as a predictor.

Endpoint Titer Modeling

Since many of the process conditions and measured variables were observed to have impact on culture behavior and a close relationship to final titer, the possibility of creating four machine learning models based on widely used linear (Partial Least Squares - PLS) and nonlinear regressors (Random Forest - RF, Support Vector Machine - SVM, Extreme Gradient Boosting - XGB) was explored. In order to build regression models, key variables were utilized. Table 3.5 summarises the important variables used to predict

batches not utilized during the training process. Endpoint viability can be considered an indicator of batch longevity whereas endpoint IVCC can be an indicator of accumulated biomass which has been observed to be a strong determinant in terms of total protein production given the strong link between the two variables (Protein yield = qP*IVCC). Endpoint residual glucose is a relevant parameter as it gauges metabolic activity at the end of the culture. Low endpoint viability indicates that the culture suffered a culture decline and thus was probably impacted in terms of its capacity to remaining metabolically active enough to produce protein in the end stage of the process. Cumulative sparged oxygen can also be understood as proxy parameter for metabolic activity. Given the important relationship between oxygen requirement and TCA cycle activity, it is coherent that oxygen sparging should be included in the model.[227] Peak lactate can be interpreted as a proxy measurement for maximum glycolytic activity while endpoint lactate can be understood as an estimation of lactate absorption which has been observed to be a good process indicator in CHO cell culture processes.[172] It can be observed as well that different pools have different importance on the protein yields. Endpoint ammonia should also be included within the model as high ammonia accumulation cultures maybe negatively impact cell culture longevity and consequently productivity. Total consumed glucose can be accepted as a proxy for overall metabolic activity (glucose can be consumed through glycolysis to yield lactate or it can be transformed to pyruvate to link with the Krebs (TCA) cycle (Figure 3.10)) and as such, it should also be represented within the model. As the Delta pool was observed to have a clear impact on protein production with increasing passage number, this information was also included in the modeling process. Total base addition can be understood as an indirect measurement of total lactate build-up and an indicator of pH

acidic profile. Total carbon dioxide sparged can be reasoned as a pH control indicator that contains information about the pH upper deadband and also indicates whether cells switch to lactate consumption phase. Lastly, endpoint pH which can be interpreted as a clear indicator of lactate consumption was also included within the model development process.

Table 3.5. Variables used for protein prediction modeling

Variables	Indicator
Endpoint Viability	Batch longevity
Peak Lactate Accumulation	Maximum Glycolysis Activity
Endpoint Lactate Concentration	Lactate consumption or production
Residual Glucose	Metabolic Activity
IVCC	Accumulative Cell Biomass
Endpoint Ammonia Concentration	Metabolic activity for glutamine synthesis and waste accumulation through amino acid deamination
Total Glucose Consumed	Metabolic Activity
Total Base Added	Lactate build-up
Total Average Oxygen sparged per day	Cellular respiration
Total Average Carbon Dioxide sparged per day	pH controlling profile for upper bound of pH deadband
Endpoint pH	Secondary indicator for lactate consumption (low pH indicates lactate accumulation while high pH relates to lactate consumption status)
Cell Passage Number	Pool stability
Pool Variant	Indicator of product specific nature

A dataset comprising 50 batches, encompassing both Multifors 0.75L and BioFlo 1L experimental runs, was split into training (84%) and test (16%) sub datasets. The BioFlo 10 L cultures were excluded from the modeling phase on the basis of high experimental variability without replicates (varying impeller configuration, varying sparger type, varying sparging strategy, varying DO set point). Furthermore, one culture from the BioFlo 1 L dataset was excluded from the modeling phase on the basis that respiratory tests were realized throughout the culture, thus potentially impacting the online values that are utilized as features in the model. From the considered data, a split was chosen randomly with the condition that it contains a high/low pro-duction batch performance. The features were regressed in function of endpoint titer to generate a model capable of predicting final yield given key process outcomes. As to give a fair chance to each model, the same strategy for tunning the hyperparameters (parameters used to control the learning process in machine learning). Here, adaptive resampling of the tuning parameter grid was realized in such a way that the random search of hyperparameters is concentrated on values that are in the neighborhood of the optimal parameters. This is done by discarding settings that are clearly sub-optimal. This approach has been observed to reduce training time.[230]

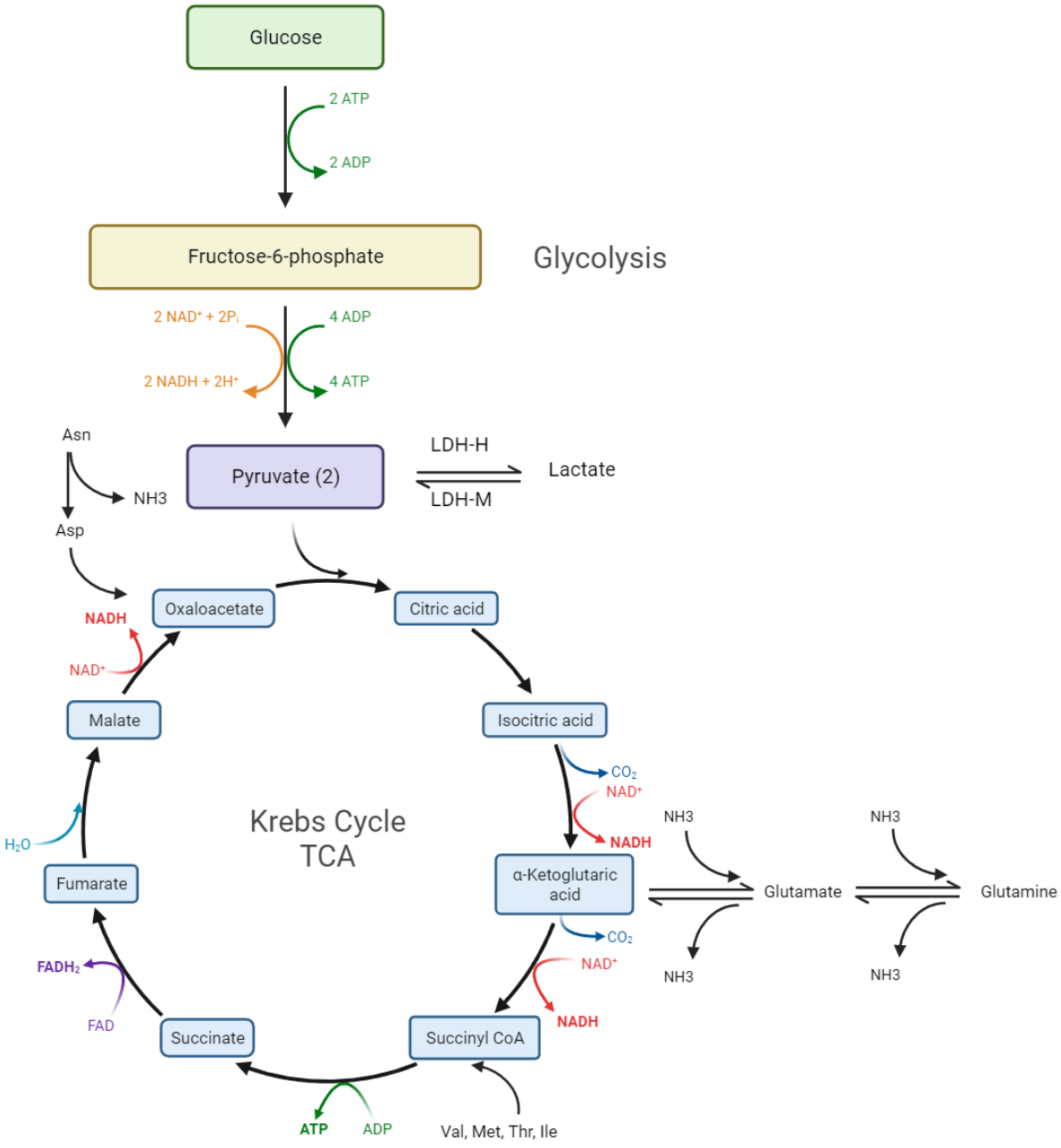


Figure 3.10. Schematic of the glycolysis pathway and TCA cycle.

2 ATPs are formed in the glycolytic pathway while 36 ATPs are generated through the oxidative phosphorylation pathway.

As it can be seen from Table 3.6 (bootstrapping results), the RF model was able to outperform SVM and PLS models in all the tested metrics (Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and R^2), for the test and training datasets. Overall metrics of the RF model performed similar in the training and test datasets suggesting that the model was generalizable across the dataset, this may be due to the fact that Random Forest algorithms are known to be robust to outliers and noise within datasets[231]. Importantly, XGB outperforms RF in the training dataset but performs worse-off in the test dataset possibly indicating a lack of generalization with the available data for this particular model. When detailing the 95% confidence intervals, it can be observed that SVM and RF are statistically different than PLS in terms of RMSE metric. Additionally, the RF model has the narrowest intervals (training and test for RMSE, MAE and R^2), when compared to the other 3 models.

Table 3.6. Mean and confidence intervals for training and test results for respective metric after bootstrapping

Training dataset	**Mean RMSE**	**Mean MAE**	**Mean R^2**	**95% CI RMSE**	**95% CI MAE**	**95% CI R^2**
SVM	95.92	66.93	0.88	61.56-125.98	43.84-82.58	0.82-0.96
RF	62.15	49.13	0.96	47.10-75.27	33.98-59.80	0.94-0.98
XGB	45.48	28.36	0.98	27.00-69.41	17.33-38,45	0.96-0.98
PLS	217.97	184.88	0.41	189.20-260.70	151.20-219.00	0.26-0.72
Test dataset	**Mean RMSE**	**Mean MAE**	**Mean R^2**	**95% CI RMSE**	**95% CI MAE**	**95% CI R^2**
SVM	80.35	74.33	0.92	61.80-97.42	52.39-89.30	0.84-0.98
RF	65.96	58.15	0.94	51.62-88.24	41.44-80.20	0.90-1.00
XGB	99.51	89.71	0.83	73.17-122.90	52.56-111.02	0.73-1.00
PLS	158.36	120.00	0.72	115.00-260.30	59.70-199.70	0.50-0.98

As it can be observed from Figure 3.11, good predictive capability is attained with the RF model, not only is the R^2 value high (0.95) but the majority (6/8) of the resulting predictions fall along the $R^2 = 1$ line which represents an ideal model of perfect prediction. When taking into account the confidence intervals of the regression line, we can see that for the total span of the data, the 95% confidence intervals contain the ideal model suggesting that despite the low data amount to test the model, predictions are statistically in line with an ideal model. Interestingly, outside the span of the data, confidence interval (CI) widens and begins to stop overlapping with the ideal model specifically within the 0-300 mg/L range. This is to be expected as there are no batches spanning this range thus no model prediction falling within this zone. Consequently, extrapolation of a linear model (regression line) outside the training and test range of the RF model is not an appropriate indicator for performance along these ranges.

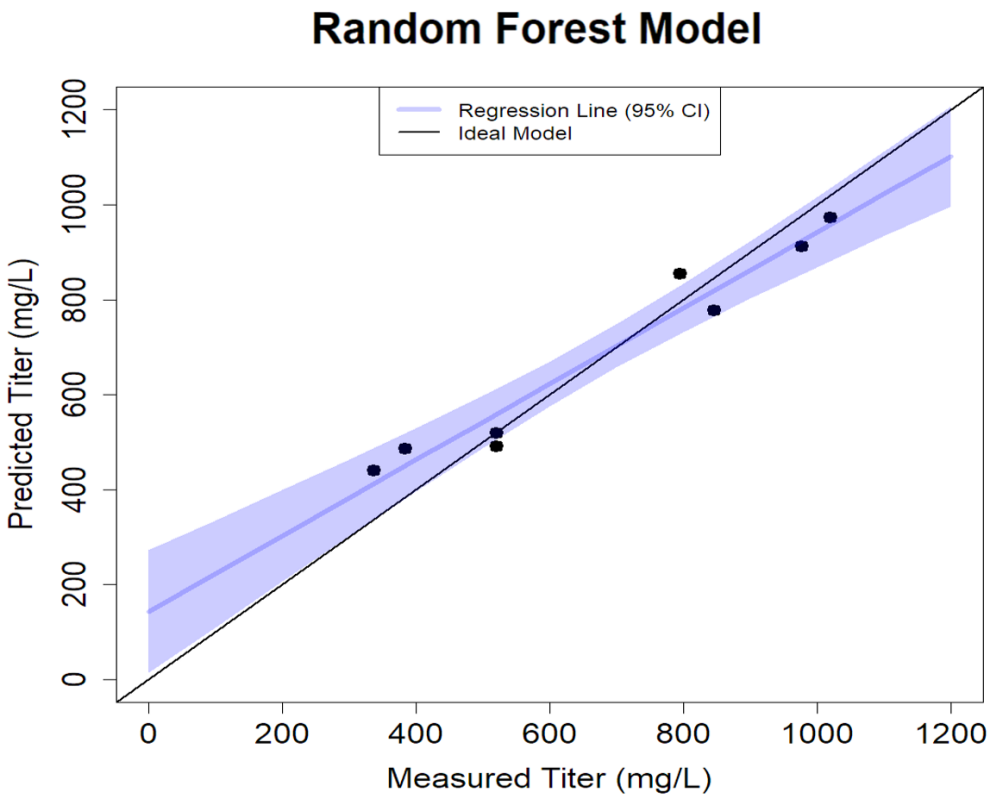


Figure 3.11. Measured versus predicted titer scatter plot.

R^2 = 1 line in the diagonal indicates ideal model and blue line shows regression line with 95% Confidence Interval (CI).

Recent approaches have centred around improving the interpretability of Machine Learning models.[232-234] This is especially important for industries where process understanding is a key requirement for regulatory approval, as is the case in the biotherapeutic industry. SHapley Additive exPlanation (SHAP) was developed using the idea of the Shapley value which is a notion in game theory that helps determine fair profit allocation to various stakeholders by evaluating their respective contribution to the outcome.[197] In the context of Machine Learning, each stakeholder can be understood as a feature and the payout is the outcome of the model itself. In summary, the Shapley value for each feature represents each feature contribution to the model's prediction of a

particular datapoint. This is estimated by calculating the average marginal contribution of a feature considering every possible combination. Consistent results have been shown with SHAP values, and SHAP dependency plots offer a helpful model summary.[196,198,233] From the generated dependency plots, various conclusions can be discerned. Features with positive magnitude SHAP values have a positive impact on the prediction, while negative magnitude values represent a negative impact. The bigger the magnitude of the SHAP values, the stronger the effect.

In Figure 3.12A, there seems to exist an optimal IVCC value for which good protein production is obtained. When coloring based on cell pools, the resulting behaviour seems to be cell pool dependent given that the Beta and Delta pools in general reached higher total IVCC but also had less cell specific protein productivity (qP). For the Wu pool, two clusters (one below 0 in the Y axis and another above 0 reaching 35) are formed. These two clusters correspond to high and low seeding densities, respectively. It can be concluded that for the two Wu pools (and presumably WuTL since the two high density clusters overlap for both pools), higher seeding density was concomitant with better protein yields. It may also suggest that the current feed regimen developed is unable to sustain higher cell densities and since Beta and Delta pools exhibited large biomass growth, an optimization of the feeding regimen could allow these pools to reach protein expression levels comparable to Wu and WuTL. When detailing Figure 3.12B, maximum lactate accumulation of 35 mM and beyond has a negative impact on endpoint protein yield. This observation holds across all pools given that the fast decrease in SHAP values is observed for Wu, Beta, and Delta pools. It is paramount to reduce lactate accumulation in a given culture to avoid increased osmolality due to base addition. One simple strategy that could

be implemented to control lactate accumulation is to replace bolus feeding with slow continuous feeding rate as it has been observed to diminish metabolic waste build-up by decreasing the variations in nutrient availability which might alter the metabolic behavior of the cell culture run.[67] From Figure 3.12C, it can be observed that high endpoint viability has a positive impact on final protein production. This impact rapidly turns negative once viability is below 85%. This suggests that increasing culture longevity and thus maintaining metabolic activity is critical. In order to avoid early cell culture decline, appropriate measures must be taken such as lowering osmolality impact (through less base addition by using only sparging gasses for effective pH control), decreasing hydrodynamic stress caused by shear damage and/or optimizing feeding strategies.[36,78,235] Feeding based on oxygen consumption rates or bio-capacitance measurements may be an attractive starting point to develop on-demand feeding strategies given the strong relationship that these signals have with viable cell volume and consequently, with metabolic activity, given that larger cells have consumed more oxygen.[39,44-46,121,236] Interestingly, when noting the SHAP endpoint pH dependency plot (Figure 3.12D), it is clear that endpoint pH of 6.93-6.95 had the most positive impact on final titers. It is worth mentioning that all the studied cultures existed within a deadband of ±0.05 between 6.9 and 7. Thus, cultures with endpoint pH of 6.93-6.95 represent processes in which no base addition or carbon dioxide was added in the last day of the fed-batch process. This could imply that unnecessary action upon cultures may be a net negative and thus pH deadband can be increased to ±0.2 around the 7.0 setpoint so as to avoid base addition and carbon dioxide sparging. From Figure 3.13A, it is discernible that even if there seems to be a pool dependence in terms of cumulative oxygen sparging (Delta and Beta cluster differently from Wu and WuTL pools), there are

signs of an optimal total oxygen sparging that is concomitant with high yields given the overlaps among pools and observing the fact that beyond 0.035 cumulative VVM, no further increase in SHAP values can be detailed. The Wu cluster near the -50 (Y axis) represents low density seeding cultures. This demonstrates the close link culture oxygen requirements have with viable cell density and viable cell volume.[140,237] For the SHAP cumulative glucose consumption dependency plot (Figure 3.13B), it can be observed that lower glucose consumption has a negative impact on protein production. This could be explained by either lower viable cell density or lower metabolic activity, both of which directly impact the culture capacity to achieve high titers. Alternatively, very high cumulative glucose consumption also had a negative impact on protein yields. This may be explained by the fact that cultures that consume glucose at high rates tend to have high lactate productions and thus adverse culture outcomes. This again points towards the idea that regulating the glucose intake of cells may be beneficial in terms of avoiding high lactate accumulation. From the endpoint ammonia dependency plot (Figure 3.13C), it is noticeable that ammonia concentrations beyond 8 mM, for the process studied, were generally associated with negative protein production prediction. This makes sense given that ammonia is another relevant by-product that can be a direct result of amino acid metabolism (Figure 3.10). Lastly, from Figure 3.13D, it can be observed how residual glucose serves as a proxy indicator of endpoint metabolic activity. Most notably, it can be understood as an inverse of the final viability measurement given that lower residual glucose values represent higher metabolic activity while high residual glucose values demonstrate cell cultures with little glucose metabolism and thus low overall metabolic activity.

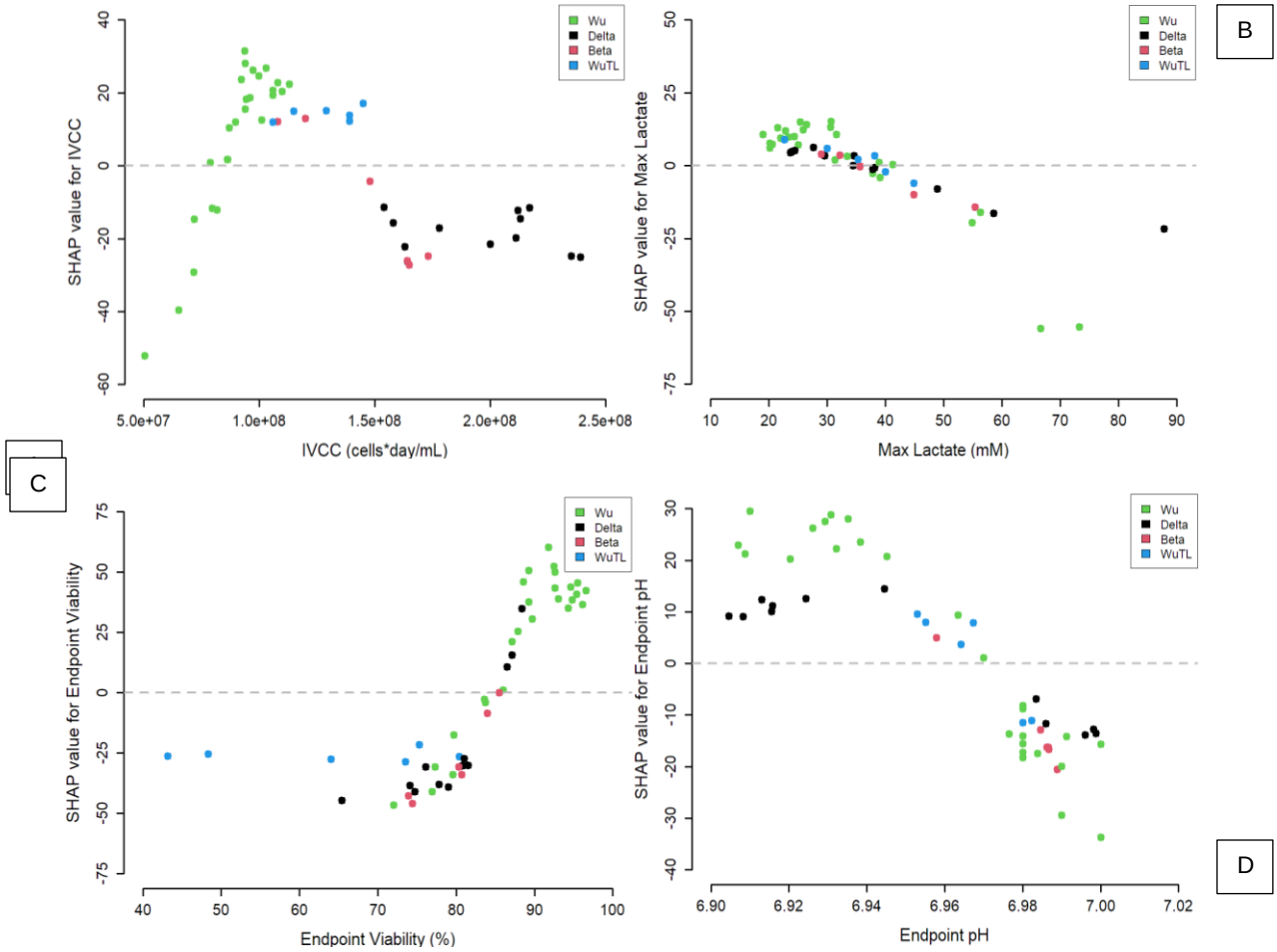


Figure 3.12. Shapley value dependency plots.

A) IVCC, B) Max Lactate, C) Endpoint Viability, D) Endpoint pH.

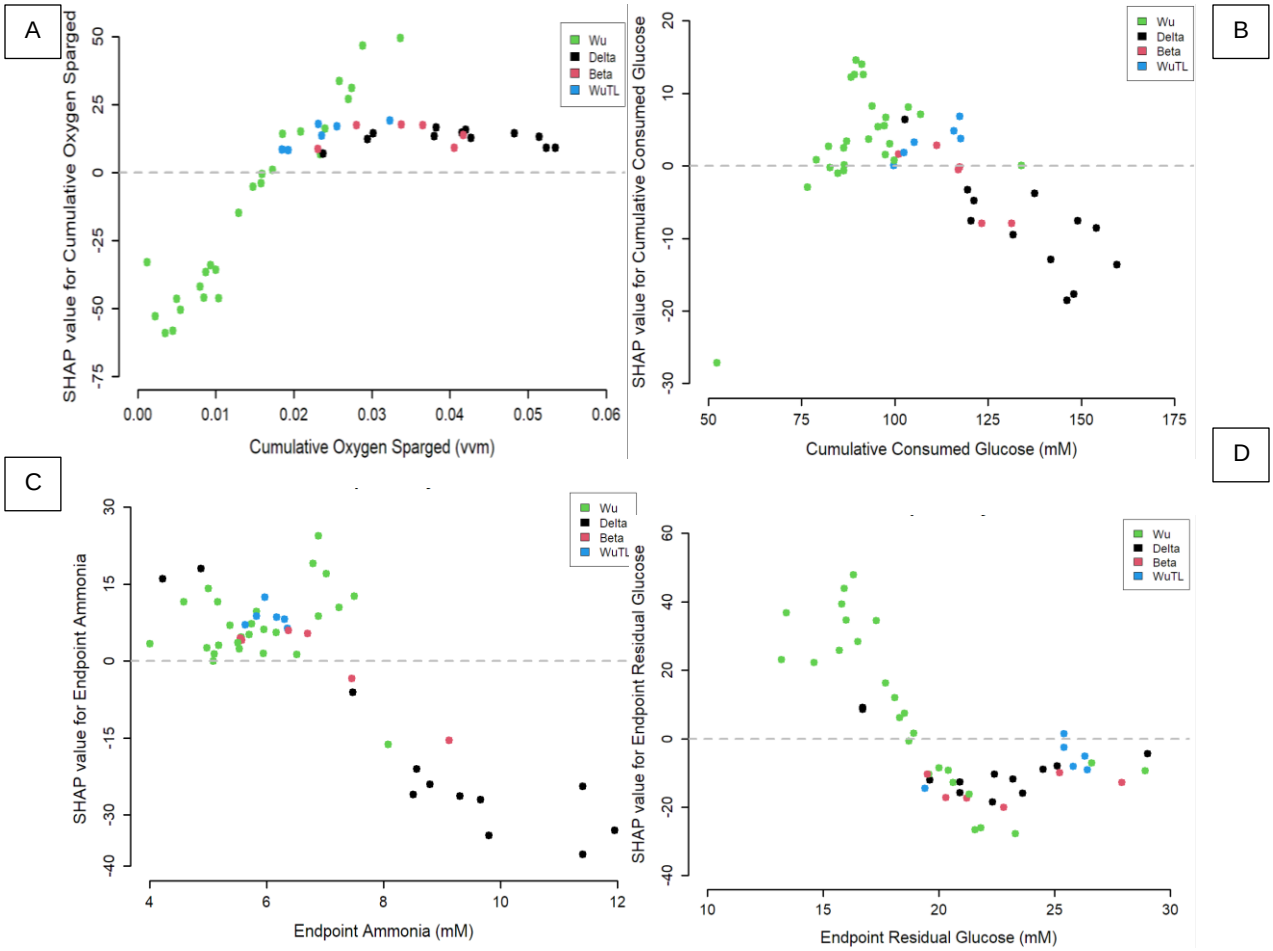


Figure 3.13. Shapley value dependency plots.

A) Cumulative Oxygen Sparged, B) Total Glucose Consumption, C) Endpoint Ammonia, D) Endpoint Residual Glucose.

Conclusion

MVDA tools can be utilized in early process development to generate wider insights that might otherwise be difficult to conclude through standard univariate analysis. In the context of this paper, it was possible to determine that although seeding density constraints the overall cell density available during the production phase, impacts on viability mean that there is a relative parity in terms on protein outcome by the end of the 17-day fed-batch process. Additionally, it was found that oxygen sparge requirements can be used as a process performance indicator to evaluate culture outcomes. It was observed that cultures with high sparge rates and low base addition were concomitant with high endpoint titers. On the other hand, by comparing the same pool across different systems, it was possible to determine which culture conditions reproduced similar behaviors such that representative processes are created. In this case, it was possible to determine which aeration conditions in the BioFlo 1L system best re-create the hydrodynamics conditions of the Multifors 0.75L system. Cell passage dependence on different pools was also readily observed through clustering in the PCA plots. This is key as it allows to determine which pools may be viable for larger scale production.

Difference in pool behaviour was also evaluated through clustering and careful analysis of the loading plots. It was observed that difference in lactate metabolism and cell growth (and the consequent derivatives such as base added volume, oxygen requirements, and longevity) were the main drivers in differentiating the culture outcomes. It was also demonstrated that a RF model, utilizing these key features, is able to capture the nonlinear relationships between the measured variables and the final protein yield in order to

generalize its predictive capabilities. These Machine Learning models can then be analyzed through SHAP dependency plots to recognize the interactions and given early process development goals, improve process understanding. To the best of the knowledge of the authors, this is the first time SHAP dependency plots have been applied for the purpose of CHO cell culture process performance analysis. This may prove to be a worthwhile strategy, given that the analysis of early process development datasets served to gain insights that can aid further process optimization. For example, it was concluded that increasing pH deadband may be beneficial to limit unnecessary base and carbon dioxide additions. This strategy may be used in tandem with slow continuous feeding to diminish sudden metabolic by-product build-up. Importantly, longevity was also determined to be a relevant factor, and as such, finding strategies that improve said longevity may also be adequate more than just increasing cell density but in detriment of viability. Such strategies could center around diminishing shear force, osmolarity stress, and improved feedings since these processing parameters (high shear, high osmolarity, inadequate feeding) have been observed to be important drivers of cellular apoptosis.[235,238,239]

Acknowledgements

This work was financially supported by the Pandemic Response Challenging Program (PRCP) of the National Research Council Canada (NRC).

CHAPTER 4 ARTICLE 2 : CHO STABLE POOL FED-BATCH PROCESS DEVELOPMENT OF SARS-COV-2 SPIKE PROTEIN PRODUCTION: IMPACT OF AERATION CONDITIONS AND FEEDING STRATEGIES

The article titled “CHO stable pool fed-batch process development of SARS-CoV-2 spike protein production: Impact of aeration conditions and feeding strategies” was accepted to *Biotechnology Progress* on September 6, 2024, and covers objective 3 of this thesis. It utilizes the process development insights gained thanks to objectives 1-2 of this thesis and applies them experimentally to improve culture outcomes. Importantly, aeration and feeding strategies are shown to be important parameters that can be tuned so as to maximize yields. Sensor technologies are also applied to soft sense metabolic shifts in real time. I was responsible for conceptualization; methodology; data curation; investigation; validation and writing – original draft.

Sebastian-Juan Reyes[1,2], Phuong Lan Pham[2], Yves Durocher[2], Olivier Henry[1]

[1] Department of Chemical Engineering, Polytechnique Montreal, Montreal, QC H3T 1J4, Canada

[2] Human Health Therapeutics Research Centre, National Research Council Canada, 6100 Royalmount Avenue, Montréal, QC H4P 2R2, Canada

Correspondence

Olivier Henry, Department of Chemical Engineering, Polytechnique Montreal, Montreal, QC H3T 1J4, Canada

E-mail: olivier.henry@polymtl.ca

Abstract

Technology scale-up and transfer are a fundamental and critical part of process development in biomanufacturing. Important bioreactor hydrodynamic characteristics such as working volume, overhead gas flow rate, volumetric power input (P/V), impeller type, agitation regimen, sparging aeration strategy, sparger type, and k_La must be selected based on key performance indicators (KPI) to ensure a smooth and seamless process scale-up and transfer. Finding suitable operational setpoints and developing an efficient feeding regimen to ensure process efficacy and consistency are instrumental. In this investigation, process development of a cumate inducible CHO stable pool expressing trimeric SARS-CoV-2 spike protein in 1.8 L benchtop stirred-tank bioreactors is detailed. Various dissolved oxygen levels and aeration air caps were studied to determine their impact on cell growth and metabolism, culture longevity, and endpoint product titers. Once hydrodynamic conditions were tuned to an optimal zone, various feeding strategies were explored to increase culture performance. Dynamic feedings such as feeding based on current culture volume, viable cell density (VCD), oxygen uptake rate (OUR), and bio-capacitance signals were tested and compared to standard bolus addition. Increases in integral of viable cell concentration (IVCC) (1.25-fold) and protein yield (2.52-fold), as well as greater culture longevity (extension of 5 days) were observed in dynamic feeding strategies when compared to periodic bolus feeding. Our study emphasizes the benefits of designing feeding strategies around metabolically relevant signals such as OUR and bio-capacitance signals.

KEYWORDS

Stirred-tank bioreactor, fed-batch process development, oxygen uptake rate, bio-capacitance, feeding strategy, process analytical technologies (PAT), SARS-CoV-2 vaccine antigen, CHO stable pool

Introduction

Currently, mammalian cell lines like Chinese Hamster Ovary (CHO) cells are the industry standard for monoclonal antibody and recombinant therapeutic protein (RTP) production because these organisms can be adapted to produce human-like proteins through various transient or stable gene expression strategies. Developing stable cell lines can often be a time-consuming process that requires numerous screening procedures to differentially select the best behaving cell clones.[51] This can clearly be an issue if the desired biotherapeutic product is urgently needed as a response to a rapidly changing public health crisis such as the one caused by the SARS-CoV-2 Covid 19 pandemic.[200] Therefore, mammalian cell pools have been considered as an interesting option to accelerating biotherapeutic production processes that do not require several months of screening and selection experiments to provide enough material for toxicology studies and Phase 1 clinical trials. Such cell pools, although inherently not as homogeneous in terms of cell population, can be employed to produce target proteins on a large scale.[201] Importantly, clones selected for production during the high-throughput screening phase are not necessarily guaranteed to work at larger scales due to varying culture environment and conditions, thus there is a case to be made to diminish excessive resource utilisation at micro liter scale and instead shift the focus towards representative scalable models. Here, cell pools can be utilized to fast forward to benchtop bioreactor scale in order to find adequate critical process parameters that are more representative of large-scale stirred tank bioreactor production. These pools can be affected by cell age, such that target protein expression productivity may significantly diminish with increased cell passage number.[52]

Thus, for stable pools to overcome their disadvantages and be a clear alternative to clonally derived cell lines, the production phase needs to start as soon as possible and the process itself has to be highly optimized.

Furthermore, the metabolism of mammalian cells is dependent on prevailing process conditions. Thus, unoptimized process parameters (e.g. suboptimized feeding strategies) can generally cause overflow metabolism in which inhibitory by-products are accumulated within the bioreactor causing the cells to lose prematurely viability and productivity. [240] Fed-batch is the dominant process used in biomanufacturing mostly due to its simplicity and efficacy. Feed bolus addition is largely developed and employed at large-scale. However, possible nutrient limitation or accumulation of primary inhibitor metabolites can induce early culture crash in large bioreactors. Feed on-demand can be an alternative option to avoid premature culture failure. For this to be possible, continuous streams of data regarding media composition or cellular metabolic activity can be key when constructing dynamic feeding strategies, which are automatically triggered when important nutrients are becoming limiting. Given this knowledge, instead of feeding cultures based on a predefined schedule, feeding strategies should be triggered or set based on biologically relevant measurements that vary over time and from culture to culture.[4] Due to the metabolic complexity of mammalian cells, recent trends have evolved towards developing and optimizing soft sensors that can abstract information from various sensor sources rather than one single measurement.[4] For example, feeding based on the integral of the bio-capacitance signal has been developed as a way to adjust feed rates with varying biomass.[241] Since bio-capacitance signals have a strong correlation with viable cell volume, such feeding has the added value of taking into account changes in cell size typically

observed during cultures.[236,242,243] With this strategy, automation of feeding protocol regardless of initial seeding density or bioreactor scale can be achieved. Bio-capacitance signals have also been used in conjunction with glucose measurements to estimate specific glucose consumption rate in real time that can then be used to forecast glucose consumption and supplement accordingly to diminish substrate variations.[126] It is worth noting that alterations in cell membrane properties are also believed to potentially impact this parameter.[118] Nutrient availability has also been found to be monitored through permittivity measurements because, at every feeding event, permittivity changes were detected to increase while declines in permittivity signals were correlated to states of nutrient depletion.[118] Additionally, during the exponential phase (which is when cell radius remains more or less constant), good correlations with oxygen uptake rate (OUR) were obtained. This may suggest that metabolic activity may also be indirectly reflected by dielectric spectroscopy.[118]

Other approaches have centered around measurement of the OUR due to the strong linear correlation between oxygen consumption and viable cell density during the exponential growth phase. OUR has thus been used to dictate glucose additions based on estimated viable cell densities[145] allowing to control glucose concentrations near a desired setpoint without large deviations. It has been postulated that the linear correlation between the cumulative oxygen consumption rate and the cumulative glucose consumed during the production phase can be used within a control strategy to maintain glucose level in the media at a given setpoint.[148,244] Such strategies were observed to diminish glucose concentration variations (under or overfeed) which are inevitably observed with bolus additions. OUR signals can also be used to detect nutrient limitations given that spent

media analysis revealed that decreases in respiratory rates were correlated with exhaustion of key amino acids. Thus, OUR can be used to supplement not only glucose but other important nutrients so as to optimize overall culture performance.[40] It has been observed that, during the protein production phase, a metabolic shift occurs such that the tricarboxylic acid cycle (TCA) is upregulated and thus cells are subjected to increased oxidative stress, suggesting that a highly oxidative state of metabolism corresponded to peak antibody production while a highly glycolytic state corresponded to peak growth cycles.[32,245] OUR can continue to increase even after peak viable cell densities are obtained indicating that a relationship exists between volumetric oxygen demand and volumetric protein production.[39] Specific oxygen consumption rates and specific protein production rates seem to have a direct relationship, suggesting a close physiological connection among cellular respiration and product formation rates.[40] Other studies also showed a linear relationship between the energy production rate and OUR.[41,42] Moreover, it was concluded that energy production rate is in positive relation with recombinant protein production rate.[43] Thus, the specific OUR can be used to represent the activity of the TCA cycle and the energy metabolic state of the cells. High values of specific OUR in the recombinant protein production phase can suggest high specific ATP production rate through TCA cycle (if the phosphate/oxygen ratio is almost constant). This, in turn, can lead to the increase in the specific recombinant protein production rate.[41] Taken together, estimating OUR in real time can be beneficial for real time monitoring of cell culture performance.

Reliable estimation of OUR can be challenging and is usually performed through one of the following three standard methods: the dynamic method, the global mass balance (GMB) method, and the stationary liquid mass balance (LMB).[246] The dynamic method

relies on the cyclical measurement of the DO extinction profile when air supply is stopped. The DO concentration decreases due to cellular oxygen need and thus OUR is directly proportional to the slope of the decay curve. The GMB approach relies on estimating the oxygen concentrations at the inlet and outlet of the bioreactor while the DO is kept constant such that the oxygen transfer rate (OTR) is equal to OUR. Mass spectrometers, paramagnetic sensors or acoustic analyzers can be used.[246] The liquid mass balance (LMB) method depends on the estimation of $k_L a$ in real time which can be difficult given that changes in gas flow rates or stirring (common parameters for DO control) can change its value.[246] Of note, it has been found that even though OUR is strongly linked with viable cell density, viable cell volume can lead to more precise correlations with OUR,[140,237] as larger cells can have increased oxygen requirements when compared to smaller cells.[44] Importantly, as the cell diameter increases, the total biovolume of the culture increases while the cell count stays constant. This biovolume increase is reflected in both OUR demand and bio-capacitance measurements, but not on cell counts.

The framework presented in this article centers around technology transfer from a 0.75L Multifors 2 bioreactor (Infors) to a 1L DASGIP multisystem bioreactor (Eppendorf) of a CHO stable cell pool and its fed-batch production process that can manufacture the SARS-CoV-2 spike protein as a potential vaccine antigen. Inducible CHO stable pools have recently been shown to robustly express SARS-CoV-2 spike protein thanks to the cumate gene switch system derived from the cymene operon of Pseudomonas putida. [220,247] Briefly, rCymR is fused with an activation domain (VP16) to form the reverse cumate activator (rcTA) that induces transcription when cumate is present, as opposed to when it is absent. [57] Thus, cumate changes the conformation of the chimeric molecule. When no

cumate is present, rcTA is not able to bind to the operator sites, while in the presence of cumate, rcTA is able to properly bind to the operator binding sites and thus initiate activation. In general, bioreactor processes are characterized by scale-dependent and scale-independent parameters. Scale-independent parameters like temperature, pH, and substrate concentrations, are, in principle, parameters that can be matched across process scales.[248] It is known that the key to realizing process transfer is to maintain scale-independent process characteristics constant across systems. Conversely, scale-dependent parameters can vary considerably with reactor size and configuration. Therefore, process technology transfer is to find out which scale-dependent variable should be kept constant and more importantly, to determine a good design space for the other scale-dependent variables which is not adverse to the culture.[248] In industrial applications, the most common scale-up and process transfer strategies are constant P/V, constant kLa, and constant impeller tip speed.[249] Constant P/V criteria have been used successfully to transfer a base-free ambr15 process to a 200L pilot scale.[250] Additionally, this strategy has also shown to produce similar growth, productivity, and glycosylation profiles across various scales (ambr15, 3L, 50L and 500L).[251,252] It must be noted that constant P/V can translate to a large difference in the $k_L a$ values between two scales and as such, alterations in the gassing profiles may be required to maintain the DO setpoint.[253]

Within the context of our study, we aimed to delve the effect of aeration strategies and feeding regimens on overall process performance. The key process transfer strategy used was keeping an equal P/V range across both systems (Multifors 2 and DASGIP parallel bench-scale bioreactors) and ensuring tip speed was kept below a threshold of 1 m/s as suggested in the literature for working volumes below 1L [248]. Overhead flow rates

were chosen such that a constant flow with respect to initial volume is subjected to the bioreactors (0.033 vvm) to avoid carbon dioxide accumulation and to replicate the overhead flow of the 0.75L Multifors 2 system [254-256]. Another relevant parameter is the dissolved oxygen (DO) setpoint. This value must provide the adequate oxygen availability in the medium for cells to consume during their metabolic activity and thus is expected to decrease as biomass grows.[257] This relates to the importance of constantly supplementing oxygen such that it is controlled around a setpoint. If this value was to be critically low (5% DO), alterations in the ratio of glucose consumption and lactate production have been observed.[258] Conversely high DO setpoints (200% DO) can lead to the accumulation of reactive oxygen species such that a negative impact on cell growth is detected [259] which implies that finding the adequate range of DO operation is key. Even though fed-batch strategies are the industry standard,[4] bolus feeding-induced oscillations on CHO cell cycles have been noted which were caused by variations in nutrient concentrations. This suggests that alternatives to standard bolus practices could be explored.[260,261]

Taken together, optimizing nutrient supplementation and gas transfer conditions within the bioreactor are two critical processes that must be carried out to obtain a robust manufacturing platform. Our research shows that synergistic effects of feeding and hydrodynamics must be considered and thus, future optimization must be done simultaneously given the strong impact that both processes have on culture outcomes.

Materials and Methods

Bioreactor cell culture conditions

A cumate-inducible proprietary CHO-GS stable cell pool expressing SARS-CoV-2 trimeric spike protein (SmT1) was grown in BalanCD CHO Growth A medium (Fujifilm/Irvine Scientific) supplemented with 50 µM MSX (L-Methionine sulfoximine, Sigma-Aldrich) in 1.8L (initial working volume of 650 mL and a maximum working volume of 1100 mL) benchtop bioreactors (DASGIP parallel bioreactor system, Eppendorf). Corning shake flasks without baffles were used to generate seed trains. The flasks were shaken at 180 rpm (25 mm orbital diameter) in a ThermoFisher HERAcell 240i incubator with 5% CO_2 and 75% relative humidity. The bioreactors were seeded at $0.4x10^6$cells/mL and cultivated for 17 days. Temperature downshift (37°C to 32°C) was realized 3 days after seeding unless stated otherwise. A pH shift was also performed on all bioreactors 2 days after seeding (from 7.05±0.05 to 6.95±0.05). Induction was realized with 4-isopropylbenzenecarboxylate (Cumate, ArkPham) 3 days post-seeding. Cultures were fed with BalanCD CHO Feed 4 (Fujifilm/Irvine Scientific) and supplemented with glucose to maintain the concentration above 17 mM (3 g/L) on the next sampling and feeding event. 6 mL samples were taken from the bioreactors on days -3, -2, -1, 0, 3, 5, 7, 10, 12, and 14 dpi (day post-induction) for off-line analysis, while feeding was realized from 0 dpi onward unless stated otherwise. Metabolic measurements (residual glucose, lactate, ammonia) were performed using the Cedex Bio (Roche, Switzerland). The manual cell counts using a hemocytometer were observed to have an average relative standard deviation error of 8%. Titer estimation using the SDS-PAGE TGX (BioRad) gel method

had an average relative standard deviation error of 12%. Amino acid measurements were conducted following the AccQ-Tag Ultra Derivatization Kit protocol (Waters Corporation, USA) that employs an AccQ-Tag Ultra C18 column with the ACQUITY H-Class UPLC system (Waters Corporation, USA) and UV detection. Offline osmolarity measurements of supernatant samples were realized with osmoTECH from Advanced instruments. Culture permittivity and conductivity were measured online using the Aber Futura biomass capacitance probe which was calibrated to 0 using fresh media prior to seeding so as to track biomass across the culture run. An open pipe sparger (4 mm outer diameter, 2 mm inner diameter) was used in all the bioreactor conditions. The overhead flow was set at 0.033 VVM (Gas Volumetric Flowrate in L/H divided by Initial Liquid Volume in L). Agitation was set at 250 rpm utilizing a single 45° pitch-blade impeller so as to maintain a volumetric power input within the range of 20-30 W/m^3. Volumetric power input decreases throughout the culture as the volume increases with feed additions.

The static gassing out method was employed to estimate oxygen $k_L a$ values. First oxygen concentration was reduced to zero by nitrogen degassing. Then, gassing was reintroduced under process specific conditions. An optical Hamilton dissolved oxygen (DO) sensor recorded the saturation process, enabling the determination of $k_L a$ values. OTR was estimated through a BluSenses off-gas analyzer GmbH utilizing the global mass balance (GMB) approach which relies on estimating the oxygen concentrations at the inlet and outlet of the bioreactor while the DO is kept constant.

$$OTR = \frac{G_m * (yO_{2,in} - yO_{2,out})}{V_L}$$

OTR (mmol/(L*H)) represents the oxygen transfer rate, G_m is the total gas flow rate (L/H), $yO_{2,in}$ is the oxygen concentration (mol/L) at the inlet, $yO_{2,out}$ is the oxygen concentration (mol/L) at the outlet, and V_L is the total bioreactor volume.[246] Variation in volume caused by feeding and sampling is taken into account during the estimation of OTR.

Online raw signals (OUR, capacitance, gas flows) were treated using Savitzky-Golay filtering in R so as to reduce noise in the data. Data pre-processing, analysis and visualization were carried out in R.[223]

Study Schematic

Figure 4.1 summarises the studies performed in the process transfer stage of SARS-CoV-2 spike protein production. As mentioned above, the initial process has been established in the 0.75L Multifors 2 (Infors) bioreactors [201]. This process was transferred to the 1L DASGIP (Eppendorf) presented in this article. Two studies on aeration strategies have been conducted investigating the impact of DO and air cap optimization. Further improvement of production performance has been enabled through two subsequent experiments delving the impact of feeding strategy.

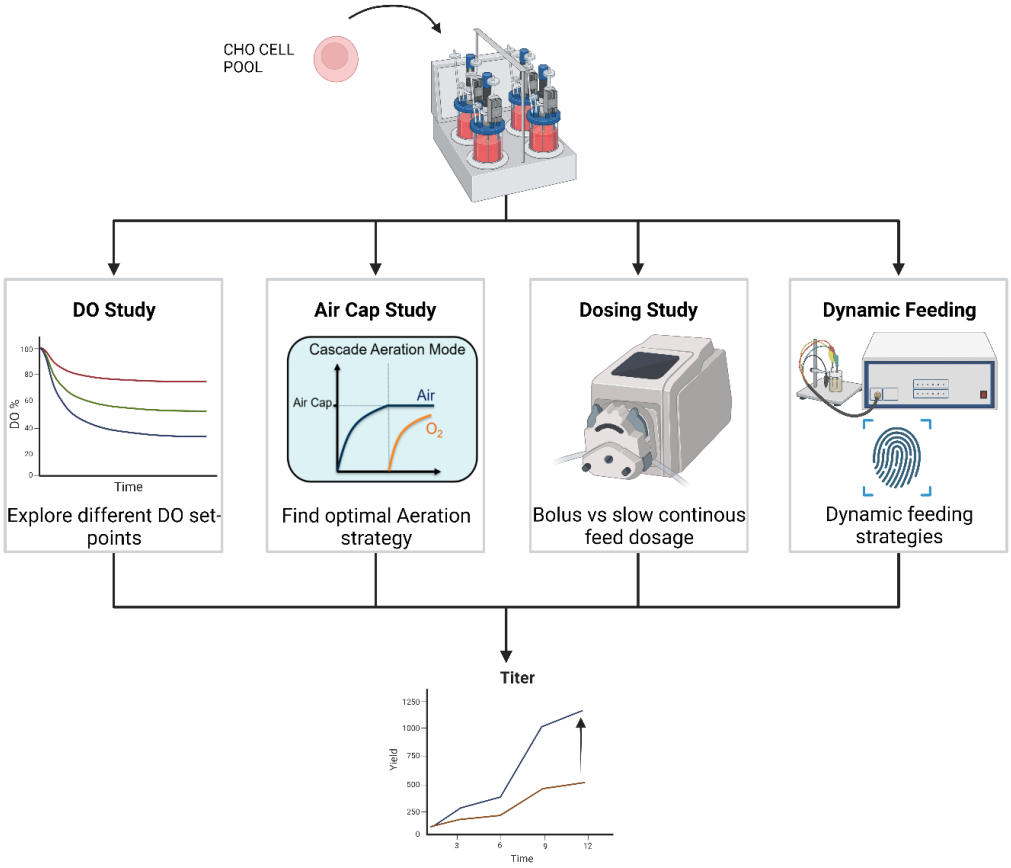


Figure 4.1. Experimental workflow.

DO study allowed to find optimal dissolved oxygen tension setpoint. Air cap study allowed to determine best performing aeration strategy while also further confirming DO setpoints. Dosing study determined the impact on culture outcomes when comparing bolus addition of feed versus slow pump addition of feed. Dynamic feeding study demonstrated feasibility of setting feed flow rates to track sensor signals.

The details of studied conditions are displayed in Table 4.1.

Table 4.1. Experimental conditions

Experiment number	Temperature shift day	Air Cap (L/H)	Dissolved Oxygen, %	Feeding	Dosage
1	0 dpi	1.5	90	Fixed feeding based on initial volume	Bolus
2	0 dpi	1.5	60	Fixed feeding based on initial volume	Bolus
3	0 dpi	1.5	40	Fixed feeding based on initial volume	Bolus
4	0 dpi	2.25	40	Fixed feeding based on initial volume	Bolus
5	0 dpi	3	40	Fixed feeding based on initial volume	Bolus
6	0 dpi	1	60	Fixed feeding based on current volume	Slow Pump
7	0 dpi	3	60	Fixed feeding based on current volume	Slow Pump
8	0 dpi	5	60	Fixed feeding based on	Slow Pump

				current volume	
9	3 dpi	1.5	60	Fixed feeding based on initial volume	Slow Pump
10	3 dpi	1.5	60	Fixed feeding based on current volume	Slow Pump
11	0 dpi	1.5	60	Fixed feeding based on current volume	Bolus
12	0 dpi	1.5	60	Fixed feeding based on initial volume	Bolus
13	2 dpi	1.5	60	Capacitance based feeding	Slow Pump
14	2 dpi	1.5	60	Constant feed per cell (CFPC) based feeding	Slow Pump
15	2 dpi	1.5	60	OTR based feeding	Slow Pump
16	0 dpi	1.5	60	Fixed feeding based on initial volume	Bolus (Control Process)

DO level impact

In this study, the air sparged flow rate was increased according to the cell growth until a limit (air cap of 1.5 L/H) followed by oxygen sparging as needed. Air caps were kept constants for all conditions while DO setpoints were varied (Table 4.1). Three concentrations of DO (40%, 60%, 90%) have been chosen.

Air caps and increased feed study

To evaluate the impact of air caps at a given DO level, a range between 1 L/H - 5 L/H (0.0015 - 0.0077 vvm) was chosen (Table 4.1). These values were selected based on the information that the $k_L a$ in frit spargers are 4-20 times higher when compared to drilled hole spargers at equal vvm [262]. It is worth mentioning that the 0.75L Multifors 2 system had an air cap that was set to 0.0033 vvm with a frit sparger (10 µm pore) and given the fact that the utilized sparger in the DASGIP system is an open pipe sparger, a range of 6-30 times more flow (in vvm) was explored. For the cultures realized at 40% DO, air caps of 1.5 L/H, 2.25 L/H, 3 L/H (0.038 vvm, 0.056 vvm, 0.075 vvm) were selected. For the 60% DO setpoint, the studied range for the air caps was 1 L/H, 3 L/H and 5 L/H (0.025 vvm, 0.075 vvm, 0.12 vvm). Since impact of DO levels (60% and 40%) was observed, air caps that fit the 6-30x vvm criteria were explored at each DO setpoint. Feeding was increased for the 60% DO conditions as viability was observed to be extended at this setpoint. The new feed condition (abbreviated as F+) was such that the total amount fed by the end of the culture (14 dpi) was a near 2-fold increase (240 mL of feed for control process and 425 mL of feed for increased feeding). Instead of feeding a fixed amount in

one rapid addition, the bioreactor receives a variable pre-set amount dependant on the current volume. Consequently, the feed volume added with peristaltic pumps (at a slow flowrate) to the bioreactor is gradually distributed between sampling days. The fixed feed volume percentages were designed to follow the increase and plateau of a cell culture run so as to mimic dynamic cellular kinetics (as observed in Figure S19). Given that osmolarity is known for Feed 4 (range of 780-930 mOsm/kg) as well as BalanCD basal medium (range of 290-310 mOsm/kg) and also feed volumes are known *a priori*, a 1.5-2-fold increase in Osm was expected with the F+ regimen. Consequently, to avoid causing instant excessive osmolarity increase (given that almost twice as much feed is added in the new feeding regimen F+ at 5, 7 and 10 dpi, the expected instant osmolarity increase from bolus additions would also be close to two times higher with respect to standard feeding F) and potentially stressful hydrodynamics variations (peak volume addition was 115 mL) during the cultures with bolus feeding, a slow continuous feeding was implemented using peristaltic pumps such that the feed amount is distributed equally at slow flow rates between sampling days. Doing so the osmolarity increase is expected to be distributed across a longer time frame (days rather than minutes). Cultures were terminated at 14 dpi as this was determined to be the optimal duration. However, if viability fell below 50% before reaching 14 dpi, cultures were terminated early. Conversely, high-performing cultures were extended beyond 14 dpi to assess viability beyond this period.

Bolus addition vs pump continuous feeding impact

Air caps (1.5 L/H), DO setpoints (60 %), were kept constant across cultures while feeding dosage was altered (pump vs bolus) for two different strategies (feed volume calculated based on initial volume and based on current volume). A temperature downshift was

realized on 3 dpi on the cultures with increased feeding (F+) to evaluate if further gains on productivity or cell growth could be obtained by exposing the cells to optimal growth temperature for a longer period of time given the fact that the new feed regimen allowed for higher nutrient availability. In the literature, temperature shift impacts on growth have been extensively characterized and modeled.[263]

Feed regimen variation study

The same DO setpoints (60%) and air caps (1.5 L/H) were used across cultures. Temperature downshift day was changed from 3 dpi to 2 dpi on the dynamically fed cultures as it was observed from capacitance measurements that the exponential growth phase seceded by 2 dpi (data not shown) thus it was postulated that by realizing the temperature shift one day earlier further gains in recombinant protein production could be realized. [263] Capacitance and oxygen transfer rate-based feedings rely on the integration of the online signal that is then transformed to a cumulative feed amount. This cumulative feed is differentiated so as to obtain the feed volume that is to be added on a given day. The feed pump flow rate is then adjusted accordingly. The constant feed per cell (CFPC) strategy relies on estimating the total amount of viable cells within the reactor at any given sampling event and adding enough feed such that each cell receives the same volumetric amount of feed. The three strategies (CFPC, OTR, Capacitance) were dosed in a slow constant speed pump fashion. The manual control feed addition methodology for OTR and capacitance encompassed the integration of the signal, followed by multiplication with a predetermined constant. This constant facilitated the transformation of units from the integrated variable to cumulative volume of feed. Subsequently, the daily updated cumulative feed underwent differentiation with respect to time, yielding a volume

earmarked for addition on a daily basis. The feed flow rate was adjusted based on this calculated volume, ensuring the dispensation of the designated volume within a 24-hour timeframe. In the CFPC feeding strategy, the feed flow rate was adjusted every sampling day so as to maintain the feed constant with respect to the number of cells at sampling day. This manual control strategy proved effective in the absence of a feedback control loop, showcasing a tailored approach to maintaining precise feed flow control. In the control process predetermined feed volume percentages (feed volume/Initial medium volume) are added on a bolus fashion every sampling day.

Results and Discussion

Impact of DO

As a first approach of evaluating adequate DO concentration, three different levels were explored: 40% DO, 60% DO, and 90% DO. In the literature, CHO cell cultures are normally cultivated at DO setpoints between 10% to 80% (% of air saturation) [264]. Prolonged exposure (10 or more days) to mildly hypoxic environments (20% DO setpoint) has elicited a similar hypoxic response as exposure of 1–3 days at 0.5–5% DO [265]. Therefore, 40% DO was chosen as the lower boundary to avoid mildly hypoxic environments that may be reached during DO control oscillations. Conversely, exposure to oxygen saturation levels above 200% DO has demonstrated alteration in the mitochondrial respiratory chain, lactate and alanine accumulation, and strong growth inhibition [266]. DO levels above 100% can impact adversely growth and yield, and even minor changes from 50% to 75% DO were observed to significantly impact cell culture performance [265]. Based

on the aforementioned information, the upper bound was chosen to be 90% DO and a representative midpoint of 60% DO was studied. These levels were set with a 1.5 L/H air cap in order to determine the impact on viability, integral of viable cell concentration (IVCC), viable cell density (VCD), and end point titers. Figure 4.2A shows that 40% DO caused a sharper decline in viable cell density by 12 dpi when compared to both 60% and 90% DO. A similar trend was observed when repeating the 40% and 60% DO conditions with an air cap of 4.2 L/H (Annex Figure S20). Figure 4.2B indicates that increasing the DO setpoint to 90 % allowed to sustain high cellular viability over a longer time period. However, this did not translate into improvement in IVCC, as this condition had consistently low viable cell counts across culture time (Figure 4.2C). It must be noted that an optimal condition was found such that the final yield of 60% DO bioreactor had a significantly higher endpoint titer when compared to 40% and 90% DO cultures (Figure 4.2D). Thus, choosing an adequate DO level along with the appropriate air cap is paramount to get the desired process outcomes. Similar behavior has been observed in other cell lines as it is known that DO setpoints can impact specific respiration rates and specific production rates.[259,267] Moreover, it has been observed that brief exposure to reactive oxygen species (ROS) in the growth medium can impede the proliferation of CHO cells without triggering cell death, while encouraging intracellular maintenance [268,269]. As a consequence, it has been postulated that the accumulation of ROS in CHO cells might interrupt the exponential growth phase, as cells strive to counteract additional intracellular ROS buildup and avert cell death.[270]. This hypothesis is in line with the observation that the 90% DO setpoint decreased cell growth (Figure 4.2A). It has been proposed that specific respiration rates are closely linked to specific protein production rates [40] which

could explain why different protein yields were detected for the different DO levels. This link can be attributed to the fact that the metabolic activity of the cells has a direct impact on recombinant protein formation. Since the respiratory activity of mammalian cells correlates well to TCA fluxes, linear relationships between specific protein production and TCA cycle activity have been reported [32,40,227,245,271].

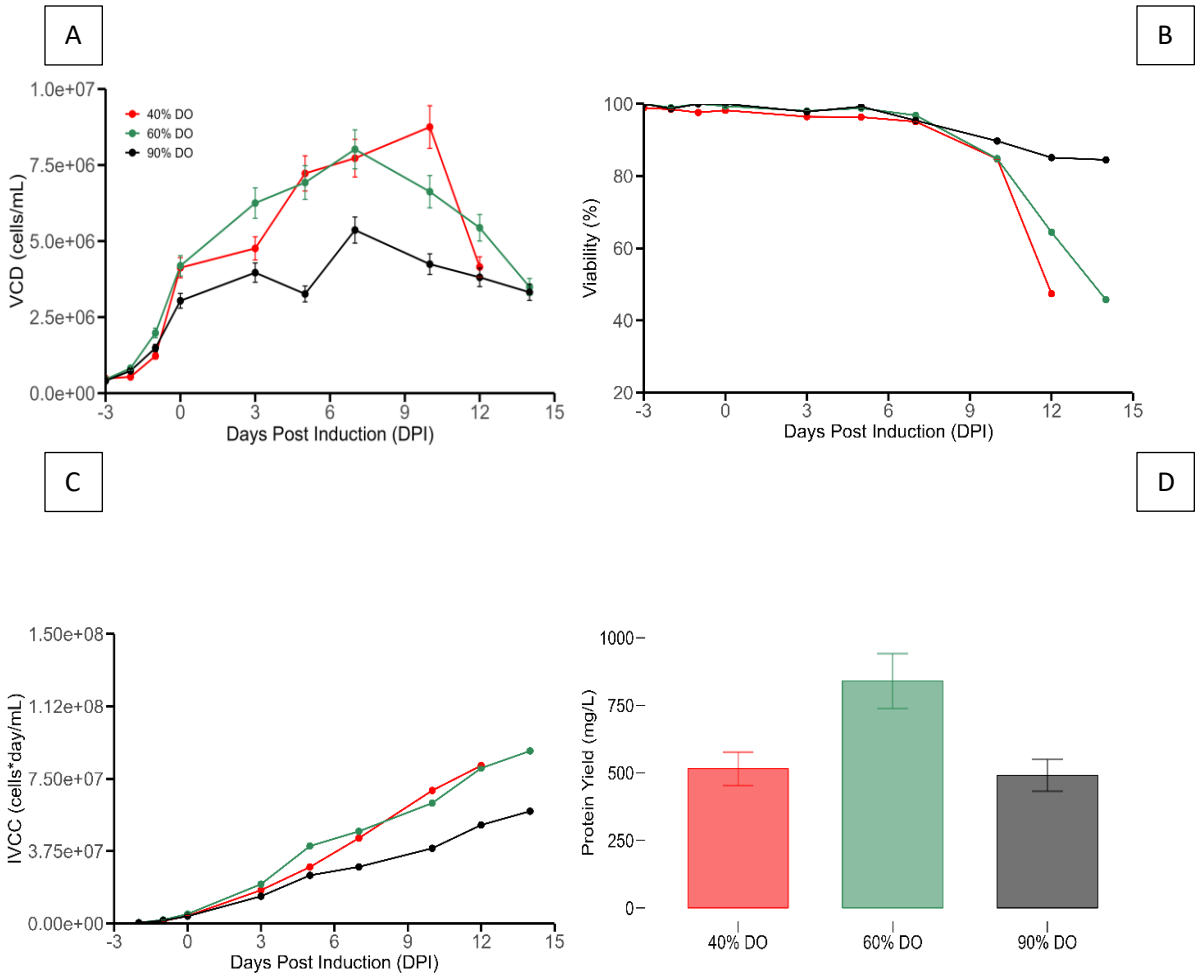


Figure 4.2. Impact of DO setpoint on A) viable cell density B) viability C) integral of viable cell concentration (IVCC) D) end point titers.

Red represents 40% DO, green represents 60% DO and black represents 90% DO. Feeding initiation was concomitant with a temperature shift (from 37°C to 32°C) and cumate induction at 0 dpi. Error bars represents the measurement error associated to each variable utilizing a representative average relative standard deviation error.

Impact of aeration strategy and increased feeding

For the cultures realized at a 40% DO level, increasing the air cap (1.5 L/H, 2.25 L/H, 3 L/H) had a negative impact on viability (Figure 4.3A). This impact on longevity also had an adverse impact on IVCC profile, given that the culture with the lowest air cap was able to reach higher cell concentrations (Figure 4.3B). This may be due in part to the fact that high aeration increases the amount of bubbles bursting at any given time, thereby augmenting the shear stress on the cells and thus adversely impacting culture performance [272,273] However, a high enough flow rate is needed to ensure sufficient CO_2 stripping. Importantly, dissolved CO_2 (dCO_2) concentrations above 68 mmHg at bench scale [250] have been observed to negatively impact protein productivity and cell growth. This observation is generally thought to be ca+used by the detrimental effects on internal pH and cellular metabolism that dCO_2 accumulation has on mammalian cells [274]. This effect can be thwarted by adding base into the system so as to keep pH constant. However, this by itself cannot be the only solution given that a stepwise increases in osmolality (caused by repeated base addition that counteracts dCO_2 accumulation) can adversely influence cell culture outcomes as well [275]. Additionally, high dCO_2 concentrations have also been observed to negatively impact glycosylation profiles [274]

As it can be seen in Figure 4.3A, at every air cap level, the combined 60% DO with gradually pumped feed (F+) outperformed the 40% DO with fixed feed (F) in terms of avoiding early cell culture termination. When looking at the IVCC profiles, again the increased feed and increased DO cultures outperformed all cultures except the 1.5 L/H with 40% DO culture (Figure 4.3B). This was presumably due to the higher shear stress conditions associated with the 3 L/H and 5 L/H cultures. Interestingly, we can see that the

60% DO condition with increased feeding and a 1 L/H air cap had a significant increase in end point viability (90% at 17 dpi) as well as a significant increase in IVCC (Figure 4.3A and 4.3B, respectively). When evaluating the corresponding lactate profiles, all the cultures with 40% DO levels and feed addition calculated based on initial volume reached a lactate production plateau and never achieved a consistent lactate consumption phase (Figure 4.3C). A similar conclusion can be drawn from the 60% DO with increased feeding in the two cultures with high air caps (3 and 5 L/H). Here a lactate re-production phase is initiated indicating high glycolytic activity. Interestingly, in the culture with low air cap, a consistent lactate absorption is discerned (Figure 4.3C). Here, once the plateau is reached, a slow decline until the end of the culture is observed indicating that the culture was able to enter a lactate consumption phase. This air cap related impact to the lactate metabolism could be explained by the fact that at higher air caps cells undergo greater environmental stress and thus have higher biosynthetic requirements for cellular reparation. Since one of the primary physiological functions of glucose is to provide the important building blocks for the biosynthesis of NADPH and nucleotides [227,270] and given the fact that NADPH plays a key role in the synthesis of macro molecules like fatty acids and amino acids [276,277], this may explain in part the higher glycolytic activities observed under higher shear stress conditions. A similar observation was realized in single-use miniature bioreactors in which high levels of bubble damage caused declines in cellular viability that were concomitant with increased glucose utilization and lactate accumulation [78].

When comparing the ammonia profiles, the increased pump feeding cultures (F+) have significantly more ammonia accumulation (Figure 4.3D). This is best elucidated by the fact that at 14 dpi, the concentration in the increased feed cultures is almost 2-fold

greater than the regular feed cultures. This may indicate that amino acid degradation took place because of greater amino acid consumption [278]. This high level of ammonia did not seem to adversely impact the culture outcomes although considerations are required as product quality may be impacted [279,280]. However, in other pools or cell lines, such increases in ammonia could be harmful and as such, better feeding strategies that avoid unnecessary nutrient additions must be developed. When detailing the final titers, it is clear that the 60% DO with low air cap (1 L/H) and high feed condition (F+) was the best performing culture in terms of protein production (Figure 4.3E). It is also worth noting that this yield is higher than 60% DO with low air cap and standard feeding (F) indicating that there is an advantage towards increasing the overall feed amount. Moreover, for the increased volume (F+), more total recombinant protein can be extracted at the point of harvest when compared to the lower feed condition (F) thus representing an economic benefit. It must also be stressed that any advantage gained by having 60% DO and increased feeding (F+) is lost when subjecting the culture to higher air caps (Figure 4.3E). This suggests that setting an inadequate air cap can be detrimental to culture outcomes and, thus, finding proper aeration conditions must be performed in tandem with adequate feeding strategies. Since no added benefit in terms of protein production was observed between 14 dpi and 17 dpi, subsequent cultures were terminated at 14 dpi.

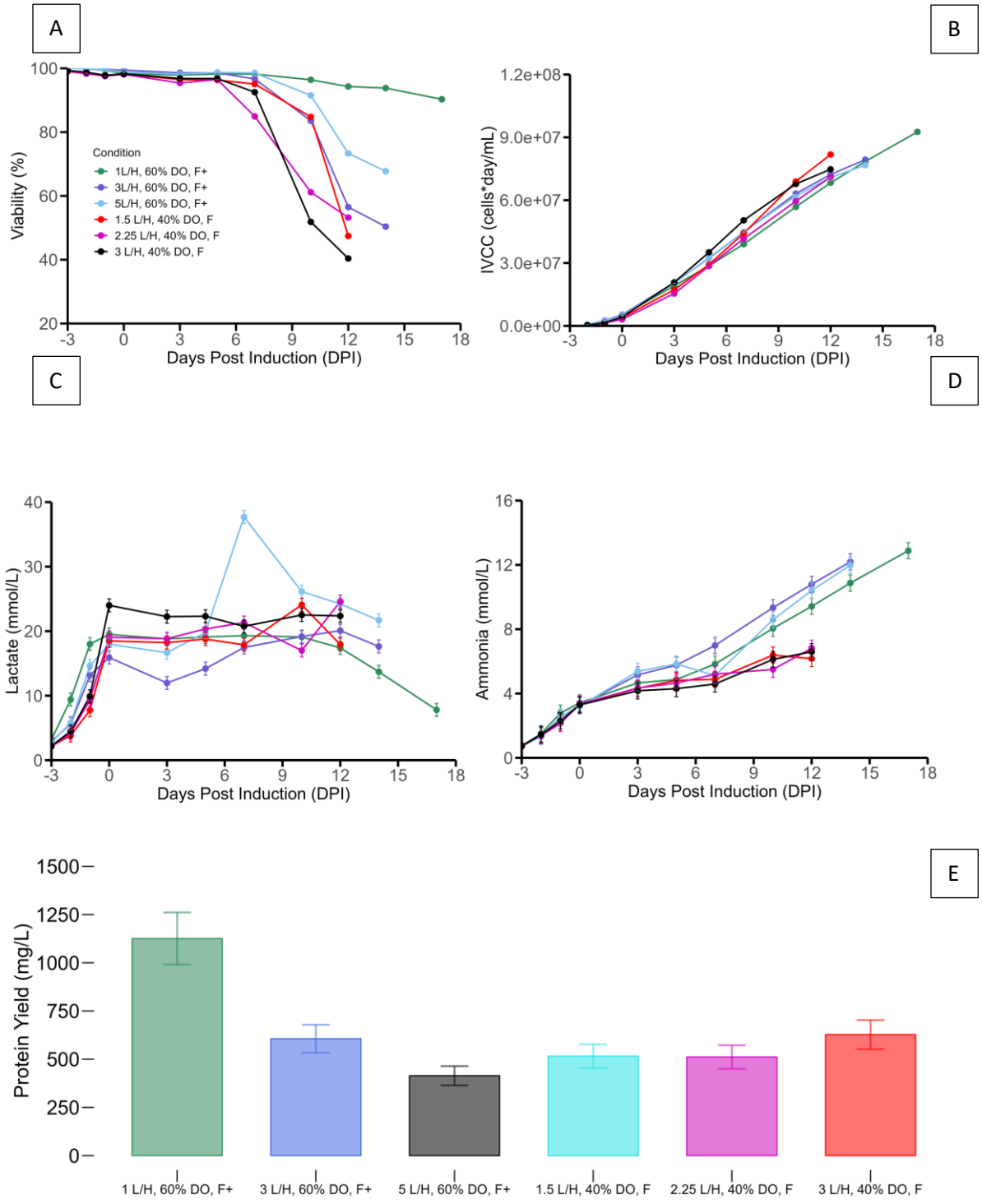


Figure 4.3. Impact of air caps and extra feeding on A) viability. B) IVCC. C) lactate accumulation. D) ammonia profiles. E) endpoint titers.

Feeding initiation was concomitant with cumate induction and temperature shift (from 37°C to 32°C) at 0 dpi for all the cultures. Feeding days correspond to 0, 3, 5, 7, 10, 12, 14 dpi. Error bars represents the measurement error associated to each variable utilizing a representative average relative standard deviation error.

Bolus feed addition vs slow continuous pump feeding strategy

An experiment comparing feed addition based on bolus dosing to slow pump dosing was devised to evaluate the impact of slow feed addition as it was not immediately clear if the increased feeding alone was responsible for the increased performance or if it was the interaction between the increased feeding and the slow addition process. A change in temperature shift for the slow pump feed condition was realized to determine if substantial gains could be made with the onset of nutrient addition. For all conditions, 60% DO was used with a 1.5 L/H air cap as low air caps were confirmed to be the best performing. As it can be visualized in the cultures with slow pump addition, regardless of the total amount of feed given (**control** feed calculated based on initial volume (control) and feed calculated based on **current** volume (CV)), both cultures had significantly higher viability by 14 dpi (Figure 4.4A). It must be stressed that the feeding volume based on current volume, by 14 dpi, resulted in 1.85-fold increase of the total feed added when compared to the feeding based on initial volume. Interestingly, the culture with increased feed (based on **current** volume) that was dosed in a bolus style had to be terminated early given that it suffered an early culture crash that was concomitant with a high bolus addition (110 mL) realized at 7 dpi (Figure 4.4A). The benefit of slow pump feeding is that it allows to increase the amount

of feed without generating adverse culture outcomes. Presumably, this culture crash is related to rapid increase in nutrient concentrations (glucose and amino acids) that, in turn, causes an abrupt change in osmolarity. This sudden osmolarity increase can be tolerated by the tested CHO pool within a threshold. In this case, the addition of 32.5 mL (between 0-5 dpi) by bolus did not immediately impact cell viability (**control** feed regimen) while the addition of 48.5 mL at 7 dpi (**control** feed) did result in a decrease in viability when measured at the next sampling point (Figure 4.4A). Importantly, when 110 mL was fed (regimen based on **current** volume) in a bolus fashion at 7 dpi, an even steeper decrease in viability was noted, thus demonstrating a step-wise decrease in viability given that ever-increasing amounts of feed (given in a bolus way) resulted in incrementally worse culture outcomes (Figure 4.4A). This observation may be due to various reasons. It could be due to a rapid increase in volume, critical decrease in nutrient concentrations between sampling days or variations in osmolarity. Since volume changes of 100 mL were not observed to drastically decrease k_La (k_La for 650 mL at 1.5 L/H air cap is 2.16 while the k_La for 750 mL at 1.5 L/H air cap is 1.96), and oxygen supplementation is anyway added as needed so as to maintain the DO constant, no impact of oxygen availability on cellular viability can be expected. Additionally, nutrient concentrations were not observed to be critically low when comparing sampling days as glucose concentration was always controlled to be above 17 mmol/L (3 g/L). Amino acid concentrations are not depleted during the production phase as seen in annex Figure 21S. When observing osmolarity measurements it was noted that at 7 dpi, the osmolarity levels were equivalent for regimen CV pump and regimen CV bolus (313 mOsm/kg for bolus and 310 mOsm/kg for pump). However, by 10 dpi, regimen CV bolus exhibited a spike in osmolarity to 427 mOsm/kg, while regimen CV

pump registered a comparatively lower value of 331 mOsm/kg. Given that both conditions were subjected to identical feeding and thus identical increases in nutrient concentrations by each sampling day, it is postulated that the observed increase in osmolarity results from decreased metabolic activity caused by cellular death. This decreased activity leads to decreased nutrient consumption and thus increased accumulation of nutrients in the media driving an increase in observed osmolarity. This is best exemplified by the differences in glucose consumption per day. For both regimen CV pump and CV bolus, glucose consumption per day is very similar (6.28 mmol/L*day and 6.63 mmol/L*day respectively) at 7 dpi. However, by 10 dpi, the glucose consumption per day begins to diverge such that for regimen CV pump the value is 7.81 mmol/L*day while for regimen CV bolus it is 5.2 mmol/L*day. This trend continues onto 12 dpi where glucose consumption remains high for regimen CV pump (5.92 mmol/L*day) and for regimen CV bolus a large decrease in glucose consumption per day is detected (2.19 mmol/L*day). This would explain the substantial difference in osmolarity measurements after 10 dpi where a large viability crash is detected. Consequently, it can be postulated that rapid osmolarity increase is a key factor driving the sudden decrease in cellular longevity. Since both cultures were fed the same total amount across sampling days, it stands to reason that the driver of cellular death is the difference in time in which the cultures were exposed to the increase in osmolarity. For the bolus culture, this osmolarity increase took place in the time frame of 1 hour or less, while for the pump fed cultures, the time frame is 48h or more. Similarly, for regimen control, osmolarity is nearly identical at 5 dpi (342 mOsm/kg for pump and 340 mOsm/kg for bolus) but begins to diverge by 7 dpi (344 mOsm/kg for pump and 387 mOsm/kg for bolus). By 12 dpi osmolarity in control regimen pump is 356 mOsm/kg while for control regimen

bolus it is 417 mOsm/kg. Concomitant with this change in osmolarity is a change in viability outcomes (93% for pump and 64% for bolus). Critically, proving the hypothesis that rapid variations in osmolarity are more negatively impactful to culture longevity when compared to slow increases in osmolarity (change in osmolarity spreads over several days) requires more frequent measurements. Such approach can be undertaken with conductivity measurements as positive correlations between osmolarity, and conductivity has been found in the literature [281,282]. As can be seen in Annex Figure 22S, conductivity measurements for bolus feed addition and pump feed addition show distinct patterns of stepwise increase (bolus feed addition) versus a slow linear increase (pump feed addition). Consequently, the observed difference of viability outcomes between bolus addition and pump addition may be explained by the fact that sudden bolus feed additions lead to rapid increase in conductance (and thus osmolarity), while slow pump addition generates a slow increase in conductivity measurements and by extension a slow increase in osmolarity that may allow the cells to adapt to the changing environment throughout the process. In the literature, it has been noted that hyperosmotic stress can drive cell death through necrosis and apoptosis [283,284]. Furthermore, it has been observed that CHO-S cells can be adapted to hyperosmotic conditions (>450 mOsm/kg) by repeated passaging (more than 10 times) while still remaining unaffected in terms of overall cell growth [285]. This approach was suggested to allow CHO cells to avoid being affected by rapid osmolarity increases caused by large bolus feed additions [285].

When evaluating the IVCC profile of the cultures, one can see that the increased feed with slow pump addition outperforms all the cultures (Figure 4.4B). This is due to the fact that viability and high cellular densities are maintained over the whole process. Lactate

profiles show that bolus additions do not facilitate a transition from lactate production to lactate consumption phase (Figure 4.4C). In fact, gradual increases and decreases in lactate concentrations are associated with every feeding event, indicating that the sudden bolus additions alter transiently the metabolic behavior of the cultures (Figure 4.4C). In contrast, the slow pump additions show a slow decrease in lactate concentration throughout the culture run. This could be explained by the fact that without the sudden changes in glucose concentrations, no sudden glycolytic influx is activated, and the culture is allowed to sustain a lactate absorption phase. This is important given that lactate absorption has been observed to be a key process indicator (KPI) for protein production and thus it is desirable behavior [172]. When evaluating ammonia concentration, it was shown that the increased feeding strategy (pump and bolus) leads to higher ammonia accumulation as previously observed (Figure 4.4D). This was detected regardless of dosage method (slow pumping or bolus addition). Despite the increase in ammonia accumulation, it cannot be implied that said outcome negatively impacted IVCC, longevity or yield. When discerning the final titer concentration, it is clear that increased feeding yields superior results, but only if it is dosed through slow pump addition (Figure 4.4E). Here, the sudden changes in nutrient concentrations and osmolality variations are avoided. In the literature, slow feeding strategies have been used with CHO-K1 cell lines for the production of monoclonal antibodies [67]. It was determined that when compared to bolus feeding methods, slow pump feeding did not show any advantage when the feed amount was low. However, with high feed amounts the pump method allowed for the reduction of metabolic-by-product buildup [67]. A similar strategy was utilized in large-scale expansion of mammalian cell spheroids. It was shown that slow pump feeding eliminated fluctuations in nutrients levels and

consequently improved cell growth [286]. Similar observations were realized in a CHO-S cell line where pump feed addition outperformed bolus feed addition in terms of IVCC and IgG production. However, viability was unaffected between both conditions [285]. A pulsating feeding strategy has also been applied to glucose and amino acid supplementation resulting in enhanced mAb production and increased endpoint viability [287]. Since these results show that slow pump feeding is applicable to CHO stable pools producing SARS-CoV-2 spike protein, it could be suggested that the advantages of slow feeding hold across a wide range of mammalian pools and cell lines and should be systematically assessed. The delayed temperature shift was observed to have a positive impact on IVCC profiles (Figure 4.4B vs Figure 4.3B) regardless of feeding strategy (based on initial volume or based on current volume). However increased yields were only observed in the cultures with increased feeding. Given that slow pump feed addition allowed for more aggressive total feed supplementation, it is possible that future gains with respect to specific protein production can be obtained since it has been observed that hyperosmotic stress on CHO cells can enhance specific protectivity [288]. Indeed, this phenomenon has also been observed in hybridoma cultures [289,290].

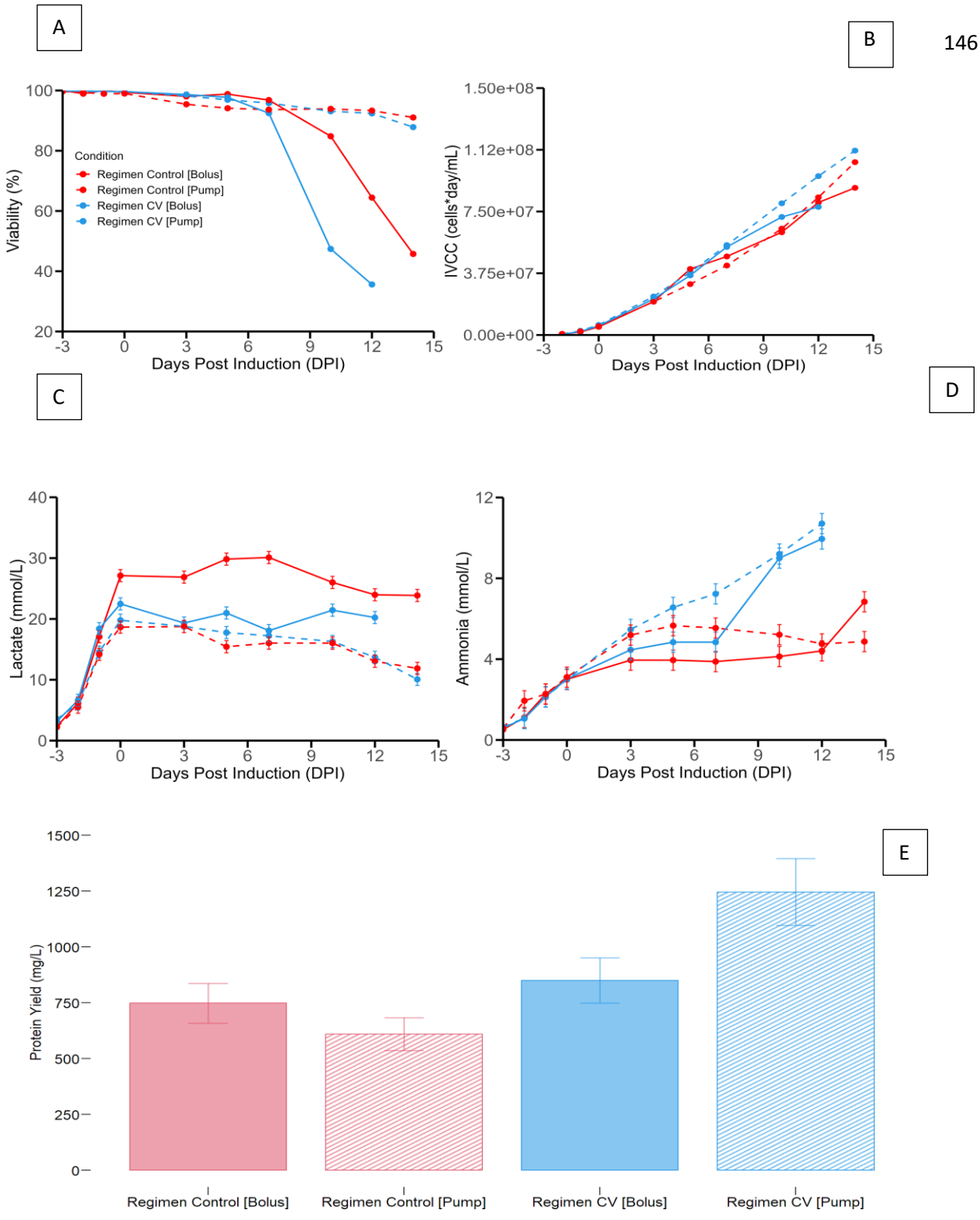


Figure 4.4. Pump versus bolus feeding impact on A) viability. B) IVCC. C) lactate accumulation. D) ammonia profiles. E) end point titers.

A) Viability, B) IVCC, C) Lactate accumulation, D) Ammonia profiles, E) End point titers. Solid red represents Regimen Control [Bolus]; Dashed red represents Regimen Control [Pump]; Solid blue represents Regimen CV [Bolus] and dashed blue represents Regimen CV [Pump]. Feeding initiation was concomitant with cumate induction at 0 dpi. A temperature shift (from 37°C to 32°C) was realized at 0 dpi for Regimen CV [Bolus] and Regimen Control [Bolus] while the temperature shift was realized at 3 dpi for Regimen Control [Pump] and Regimen CV [Pump]. Error bars represent the measurement error associated to each variable utilizing a representative average relative standard deviation error.

When detailing the amino acid concentrations in Figure S21, it can be discerned that His, Ser, Arg, Gly, Glu, Pro, Thr, Cys, Met, Ile, and Phe concentrations begin accumulating at greater amounts by 10 and 12 dpi in the CV Bolus condition when compared to the CV Pump condition. His, Ser, Arg, Thr, Cys, Met, Phe, Ile, Leu and Val are known to be consumed during the production phase and at the beginning of the decline phase in CHO cell cultures [32,291]. A simplified schematic of mammalian cell metabolism and the relationship between glycolysis and the TCA cycle can be viewed in Figure 4.5. An increase in the concentration of these amino acids in the CV Bolus was observed when compared to CV Pump. It makes sense as by 10 dpi, CV Bolus viability drops below 60% thereby diminishing the number of cells consuming the available amino acids in the feed and basal media.

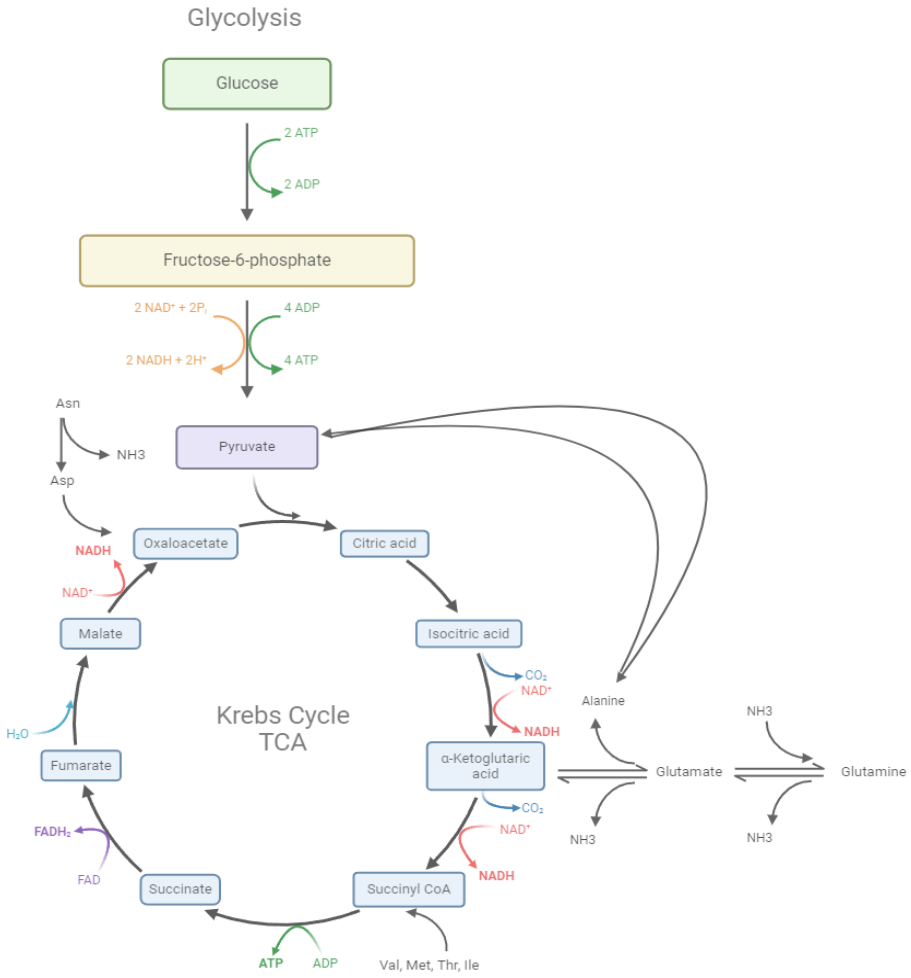


Figure 4.5. Schematic of mammalian cell metabolism

Glutamate is added to the feed so as to aid in glutamine synthesis for CHO-GS cell lines [292]. Its accumulation in the medium could indicate that glutamine synthesis is decreasing and consequently its metabolic activity is waning. Given that by 10 dpi the viability of CV Bolus drops below 60%, its reduced metabolic activity led to greater accumulation of glutamate when compared to CV Pump. Furthermore, CHO cells face challenges in synthesizing an adequate amount of proline to support essential cellular processes such as *de novo* protein synthesis and consequently proline is supplied via feed supplements [293]. As such, differential accumulation between Bolus and Pump conditions

may indicate inadequate metabolic activity in cellular maintenance and recombinant protein production which can be explained by the 60% viability drop by 10 dpi in the CV Bolus condition. Glycine is a product of serine metabolism which fuels 1C units metabolism [294]. This amino acid is present in low concentrations in both the feed supplied (0.67 mmol/L) and in the basal medial (1.41 mmol/L) and is involved in the synthesis of glutathione [278]. Since there have been hints at a potential connection between glutathione (GSH) levels and cellular productivity of recombinant proteins [295-297], it stands to reason that increased accumulation of glycine in the CV Bolus condition may be interpreted as lower GSH synthesis activity caused by the viability crash which led to a decrease in the oxidative metabolism that is generally associated with high protein production [245]. This is evident by the fact that peak pure oxygen flow in the CV Pump condition was 3.3 L/H at 12 dpi while for the CV Bolus condition it was 2.2 L/H at 6 dpi.

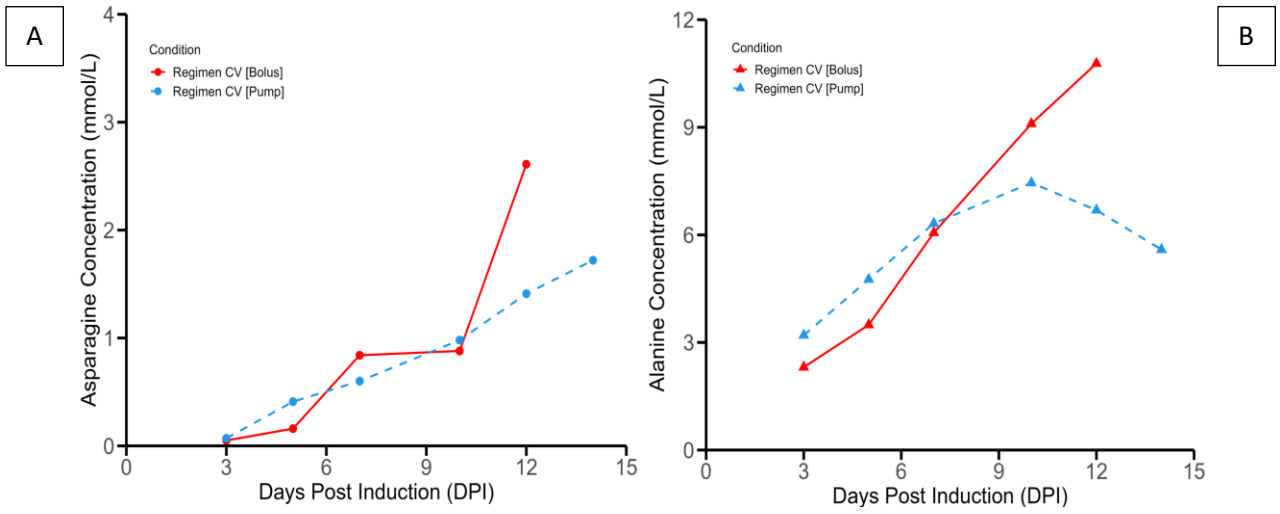


Figure 4.6. CV[Pump] and CV[Bolus] impact on A) Asparagine (Asn) concentration in the spent media and B) Alanine (Ala) amino acid concentration in the spent media.

Feeding initiation was concomitant with cumate induction at 0 dpi. A temperature shift (from 37°C to 32°C) was realized at 0 dpi for Regimen CV [Bolus] while the temperature shift was realized at 3 dpi for Regimen CV [Pump].

In Figure 4.6, it is possible to observe the concentration profiles for asparagine (Figure 4.6A) and alanine (Figure 4.6B). During the plateau phase, asparagine represents about 8% of incoming carbon source [245]. In fact, it has been determined that multiple TCA cycle intermediates (e.g. citrate, malate, succinate) derive substantial carbon from asparagine catabolism [298]. Consequently, it has been delimited as a key nutrient that has to be replenished to avoid rapid depletion [299,300]. A recent CHO metabolism review paper has noted that asparagine is significantly consumed during the production phase and the decline phase [32]. Therefore, it makes sense to observe differential accumulation in the media for the Regimen CV Bolus and Regimen CV Pump conditions. As the large bolus addition of feed is added to the CV Bolus condition at 7 dpi, the subsequent viability crash diminishes metabolic activity and consequently asparagine uptake requirements. Interestingly, the alanine profile (Figure 4.6B) is observed to diverge between CV Bolus and CV Pump conditions by 10 dpi. Generally, alanine is observed to accumulate in the extracellular media both in production phase and the decline phase [32,291,301]. There is no alanine in the feed formulation and a low amount in the basal media (0.53 mmol/L). It is hypothesized that the secretion of alanine into the cell culture serves as a strategy to mitigate ammonium toxicity by functioning as a reservoir for nitrogen [302]. It is thought that alanine biosynthesis serves as a stress response mechanism for metabolic waste accumulation [303]. Additionally, the metabolic interconnection between glutamine and alanine involves the glutamine-pyruvate transaminase reaction, wherein glutamine is transformed into α-ketoglutarate, and

the resulting amine group is transferred to pyruvate to produce alanine [304]. These transaminase reactions, including those mediated by glutamine-pyruvate and glutamate-aspartate transaminase, allow CHO cells to mitigate excess ammonia production, potentially contributing to the establishment of a more favorable cellular environment in culture [304]. Interestingly, in CHO cells, alanine concentrations have been observed to diminish during the production phase of a culture and as such it has been suggested that alanine is important in metabolic processes related to recombinant protein synthesis [305]. Tangentially, excessive alanine accumulation is known to be a negative in CHO cultures, as it can serve as an allosteric inhibitor of pyruvate kinase by signaling an abundance of intermediates from the tricarboxylic acid (TCA), thereby inhibiting pyruvate kinase and the TCA pathway [306], this can be best visualized in Figure 4.5. Consequently, it may be interpreted that alanine consumption observed in CV Pump condition underscores increased TCA cycle activity which can be further underlined by the higher oxygen (3.3 L/H > 2.2 L/H) requirements and protein production (1246 mg/L> 850 mg/L) when compared to the CV Bolus condition.

No clear impact regarding delayed temperature shift for regimen CV pump when compared to regimen CV bolus can be discerned. The main driver of difference can be attributed to the difference in dosage method as the difference in dosage method explains the difference in longevity and metabolic activity. For illustration, when comparing regimen CV pump to the previous set of experiments (1 L/H air cap, 60% DO with regimen CV pump, temperature shifted at 0 dpi), marginal gains in titer at 14 dpi (1193 mg/L vs 1246 mg/L) and increased IVCC at 14 dpi ($7.85x10^7$ cells*day/mL vs $1.12x10^8$ cells*day/mL) are detected. Here, the delayed temperature shift can be observed to allow

for more biomass accumulation and a marginal improvement in titer. Despite the temperature shift difference between these two conditions, longevity remains high at 14 dpi (viability is around 90%) and metabolic activity is also comparable at 14 dpi (glucose consumption of 6.02 mmol/L*day for the regimen CV pump culture and 5.33 mmol/L*day for the 1 L/H air cap, 60% DO with regimen CV pump, temperature shifted at 0 dpi culture).

Dynamic feeding strategies

Given that slow pump feeding was observed to be the optimal dosage, various dynamic feeding strategies were devised and assessed to optimize the process. These feeding strategies were set up so as to respond to biological signals rather than feed based on a set amount that was determined *a priori*. Constant feed per cell based on manual cell counts (CFPC), feeding based on capacitance signal (CAP), feeding based on OTR signal were compared to the standard process of feeding that relies on bolus additions calculated based on initial volume. For the three strategies, the onset of feeding was one day before induction (-1 dpi) since this was observed to be a key moment given that the cultures reach air cap at this point and thus require additional oxygen supplementation (Figure S23 in Annex contains a typical DO profile with online air cap and O_2 flow). This may indicate that the cultures are very metabolically active (probably through glycolysis path to rapidly produce 2 ATP per mole of glucose to support rapid cell growth), thus the cells are in need of additional nutrients. For the CAP and OTR feed regimens, a daily adjustment of the feed flow rate to align with the desired signal trajectory was realized while for the control and CFPC regimens feed additions are realized every sampling day.

For all three dynamic feeding strategies, culture longevity was maintained longer when compared to the standard process probably due to the dosage or the varying amount or a combination of both effects (Figure 4.7A). It is also clear from the viable cell density plot (Figure 4.7B) that after the temperature downshift at 2 dpi, cell concentrations for the CFPC, CAP, OTR conditions were maintained for longer periods of time. Importantly, for the three dynamic feeding strategies cell concentrations remain high after temperature shift while the control process continues growing after temperature shift, reaches a peak VCD value and subsequently crashes. Lactate again can be observed to enter a consumption phase soon after 2 dpi which was concomitant with the temperature downshift (Figure 4.7C). Ammonia build-up, even though higher than in the standard process, is overall less when compared to the previous regimen with feed volume calculated based on **current** volume. This could be an indication that, since the feeding strategies relied on biologically relevant signals, the degradation of amino acids in the media due to over feeding was diminished (Figure 4.7D). From the daily feed profile, one can see how the strategies were able to trace the relevant signal (Figure 4.7E). While CFPC relies on viable cell concentrations (Figure 4.7B), the feed based on capacitance signal is able to track the time evolution profile of the permittivity signal (Figure 4.8B) and the OTR-based feeding strategy is enabled based on the changes in OTR (Figure 4.8A). As it can be discerned from the endpoint titers (Figure 4.7F), the three dynamic feeding strategies outperformed the standard process such that capacitance-based feeding (CAP) > OTR based feeding > cell count based feeding (CFPC) > bolus feeding based on a fixed amount related to the initial volume. This ranking could be linked to the fact that both bio-capacitance and OTR-based feeding strategies are indirectly monitoring biovolume and its consequent metabolic

activity while the cell count based feeding is assuming that each cell should receive the same amount of feed per day. Thus, conceptually speaking, the two feeding strategies (CAP and OTR-based) are much more dynamic in terms of considering changes in metabolism or total biovolume. Regimen CFPC had a lot more feed addition between 2-6 dpi when compared to the best performing regimens (CAP and OTR). This sub optimal addition at the start of the production phase may have contributed to worse protein production given the fact that CFPC has lower endpoint protein production (762 mg/L) when compared to CAP (1300 mg/L) and OTR (1029 mg/L). It must be noted that when comparing the protein yield of regimen CV pump (1250 mg/L) to regimen CAP and OTR similar high performance is observed. Given that the feed profile of regimen CV pump was pre-set to track expected changes in cell counts during the cultivation process thus serving as a proto-dynamic feed regimen strategy. It may be hypothesized that the changing feed profile along with slow continuous addition of feed is one of the mayor explicatory reasons for the high protein yield results in regimen OTR, CAP and CV pump.

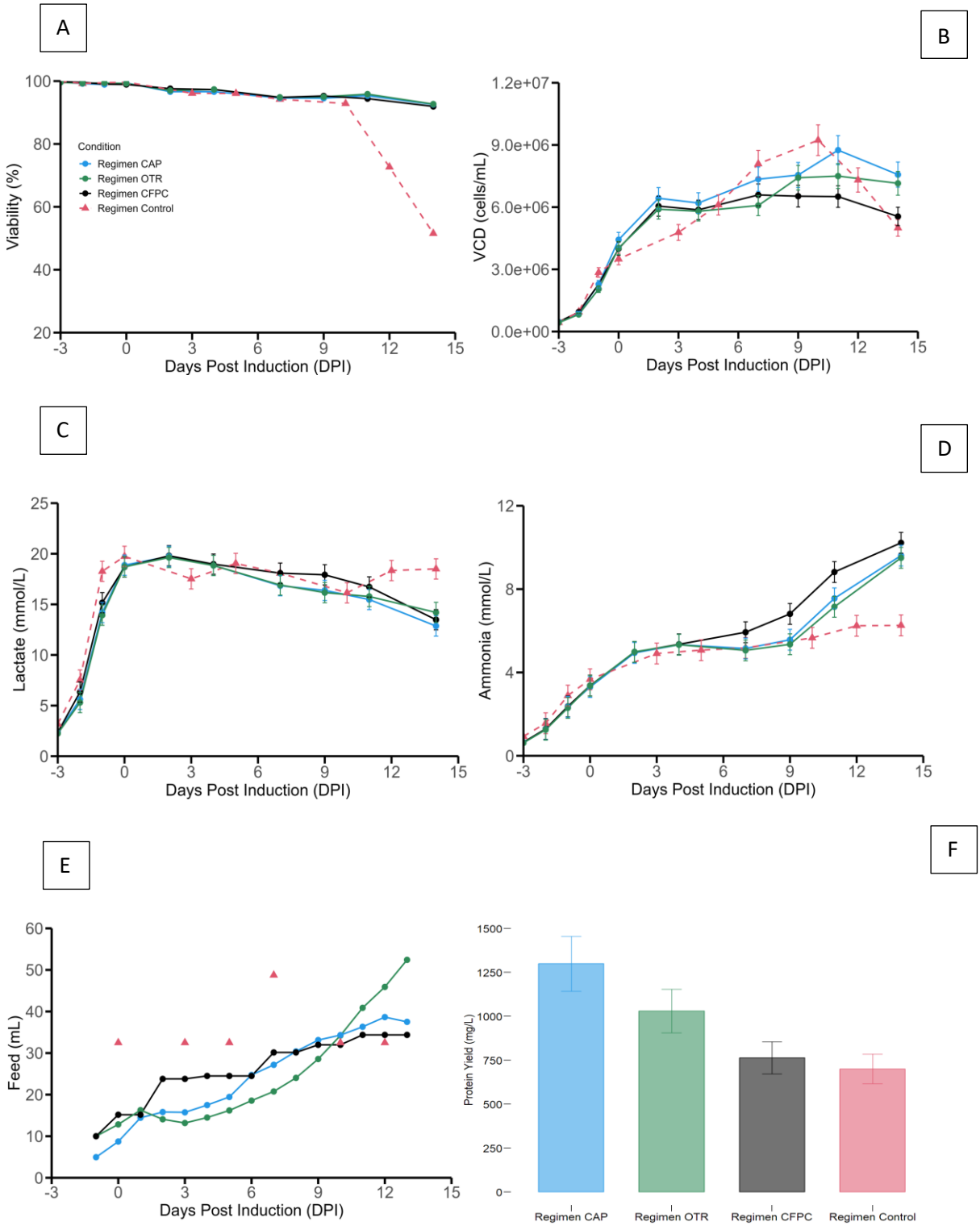


Figure 4.7. Impact of biologically relevant feeding strategies on A) viability. B) viable cell densities. C) lactate accumulation. D) ammonia accumulation. E) representation of daily feed volume additions. F) endpoint titers.

Initiation of feeding was concomitant with temperature shift and cumate induction at 0 dpi for the control regimen. For regimen OTR, CAP and CFPC, feeding was initiated at -1 dpi, cumate induction was realized at 0 dpi and temperature shift was realized at 2 dpi. Feeding days correspond to 0, 3, 5, 7, 10, 12 dpi for the control regimen and -1, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 dpi for regimens CAP, OTR and CFPC. Error bars represent the measurement error associated to each variable utilizing a representative average relative standard deviation error.

From the oxygen uptake rate plots (OTR=OUR due to a constant DO setpoint maintained in the bioreactors) (Figure 4.8A), it is clear that feeding based on capacitance signals induced higher overall oxygen requirements which are in part explained by a higher VCD that was sustained throughout the culture process. Additionally, the increased OUR may be linked with an increased protein expression as this culture was able to outperform all the tested feeding strategies (Figure 4.7E). In the cultures that were subjected to a 2-dpi temperature downshift (CFPC, CAP, OTR), a temporary decrease in OUR is observed from 2 dpi to 4 dpi indicative of a decreased metabolic activity probably due the decrease of temperature from 37°C to 32°C (a concomitant decrease in glucose consumed per day was also noted as shown in Annex Figure S24). This has been noted in the literature as decreases in temperature impact specific respiration rates [267]. It has been found that mild hypothermia when coupled with nutrient supplementation can increase specific protein production due to its inverse relationship with the fraction of cells in S phase during the cell cycle [307]. When overlaying the resulting OUR and capacitance signals, several points of interest can be noted. First, peak bio-capacitance and peak oxygen consumption measurements happen around the same points in time (7-10 dpi for the control regimen and 10-14 dpi for the CAP

regimen), indicating that both oxygen consumption and bio-capacitance are strongly related to viable cell volume as has been suggested in the literature [44,45,271,308,309]. Second, decrease in oxygen consumption precedes cell death and it is steeper than the decline in cell counts (Figure 4.8B). This could be due to the fact that once the cells are close to the decline phase, their metabolic activity slows down considerably. This is evident from the reduction in glucose consumed per day for the regimen control condition while daily glucose consumption rate remains high in the CAP condition (Annex Figure S24). Third, the increase of the bio-capacitance signal (from 6 pF/cm to 18 pF/cm) between 2 dpi and 12 dpi (Figure 4.8B) cannot be explained alone by the secondary increase in cell counts (from 6.5 million to 9 million) after induction (Figure 4.7B). This increase in cell counts was also accompanied by an increase in cell size. Historical records for this cell pool show radius increased from 14.9 µm at inoculation (-3 dpi) to 16.8 µm at 12 dpi. A similar observation was realized on a CHO cell line where the cellular diameter was determined increase from 13.6 µm before induction to 15.5 µm 9 days after induction [310]. For illustration, a 2 µm increase in radius represents a 1.43-fold increase in cellular volume [236,242,243]. Fourth, both OUR and bio-capacitance signals show similar trends for the bolus feeding process. Here both signals grow almost exponentially between 3 dpi and 9 dpi, which is in response to strong secondary growth phenomenon. In the pump fed cultures, the bio-capacitance and OUR curves also show similar behaviors. The capacitance signal shows a convex trend while the OUR signal displays a concave profile. This may in part be due to the fact that strong secondary growth is not observed. Since there is no strong secondary growth in this case, the increase in capacitance must necessarily be due to increase in cellular size while the OUR increase is due to both an increase in cell size and

an increase in metabolic activity. Nonetheless, both signals have an increasing trend thanks to the capacity of the signals to directly (capacitance) or indirectly (OUR) detect biovolume which can also serve as a proxy for metabolic requirements (bigger cells have higher oxygen requirements). As it can be seen from the titer production profiles (Annex Figure S25), a correlation is observed such that when the cultures are ordered in terms of peak oxygen requirements at the end of the culture process, they are consequently also ordered in terms of peak protein expression. This could be possibly explained by the fact that close relationships between recombinant protein expression and enhanced TCA cycle activity have been observed [227].

Interestingly, when evaluating the biovolume specific respiration rates (OTR/Capacitance) for the bolus-fed process and the continuously fed capacitance-based process, it is clear that biovolume specific respiration rates are higher in the slow pump process when compared to the bolus-fed process. Given the strong link between specific protein expression rates and specific respiration rates, it can be postulated that this is a key reason as to why higher protein titers were reached in the CAP regimen (Figure 4.8C).

When comparing the cumulative OUR versus the cumulative glucose consumption plot after 4 dpi - 5 dpi, there exists a change in the relationship between oxygen requirements and glucose utilization (Figure 4.8D). In essence, the total oxygen consumed increases at a higher rate (increasing slope) when compared to the increase in total glucose consumption. Since lactate begins to decrease after 4 dpi and protein expression begins to increase dramatically between 4 dpi and 12 dpi, it can be postulated that these two key nutrients (glucose and oxygen) are being fluxed into the TCA cycle to be used as part of protein synthesis. Metabolic shifts can also be monitored such that a change from highly

glycolytic metabolism to predominantly oxidative metabolism can be detailed. Given the positive correlation that has been observed between oxidative metabolism and protein production, it is a parameter to consider when analyzing OUR data. Thus, estimating OUR and bio-capacitance in real time and contrasting it with measured metabolic rates can serve as a basis for soft sensing changes in metabolism that in turn are related to process outcomes like levels of protein expression.

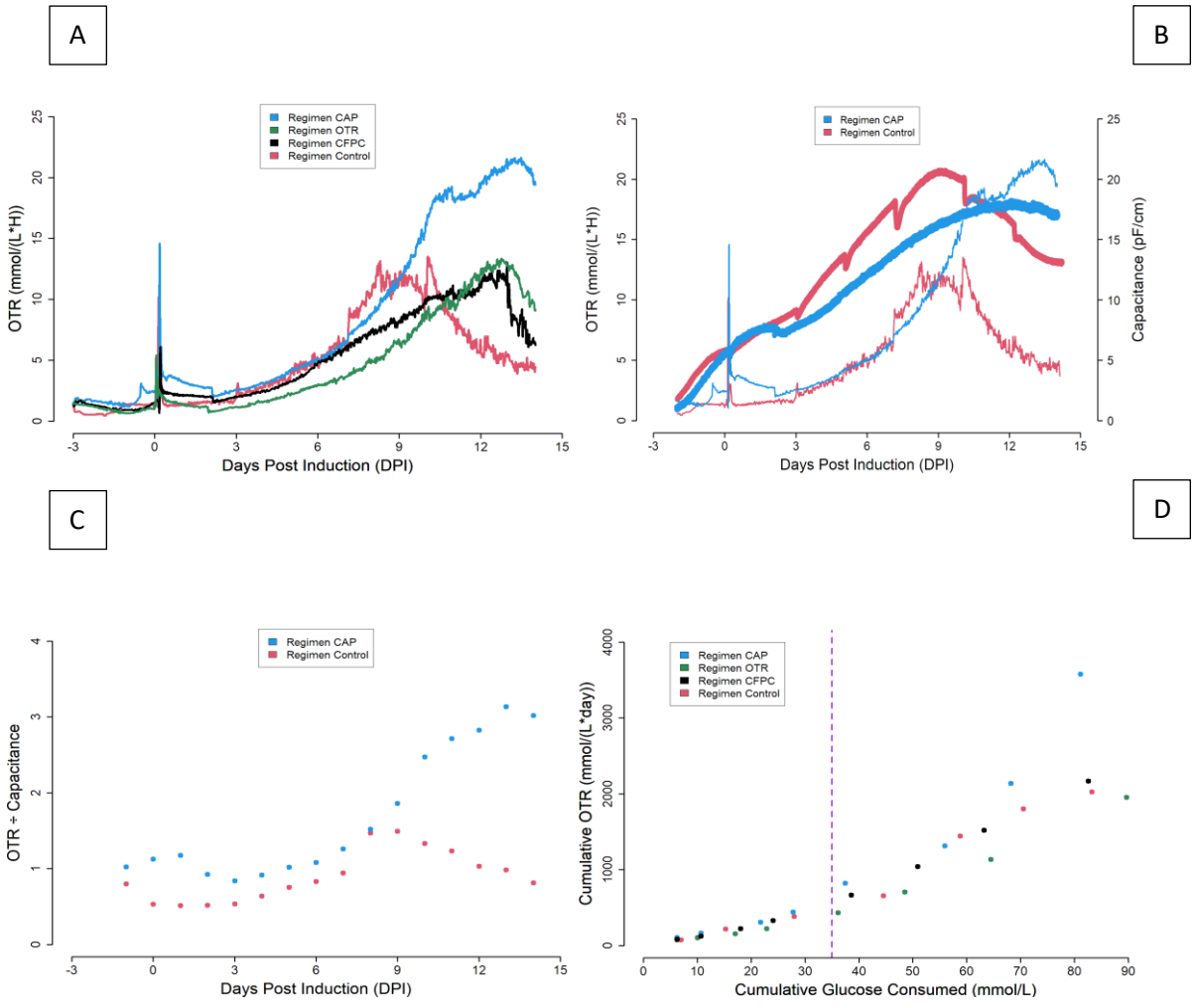


Figure 4.8. Biocapacitance and OTR online Data.

A) OTR measurements for the various feeding strategies. B) OTR and bio-capacitance signal overlay. Thin lines (red and blue) are the OTR signals while the thick lines

(red and blue) are bio-capacitance signals. C) Daily specific volumetric respiration rates (ratio of daily OTR measurements and daily capacitance measurements). D) Cumulative OTR versus cumulative glucose consumption scatter plot. Purple line separates 4 dpi from 7 dpi datapoints.

Conclusion

In the scope of this paper, improvement of culture outcomes in terms of longevity, IVCC and SARS-CoV-2 spike protein production was achieved through various aeration and feeding strategies. Firstly, it was observed that appropriate DO setpoints had important effects in terms of increasing endpoint viability and protein production yield. At the same time, it was determined that setting the appropriate sparging air caps also has a significant impact on cell culture kinetics such that adverse conditions (e.g. high air caps) impact not only cell culture longevity but also lactate accumulation profile. Presumably, this is a consequence of sub-optimal environmental hydrodynamics which alters the metabolic pathways of the cells. Although not shown within this paper, low air caps can also be detrimental given that insufficient carbon dioxide stripping can be observed especially when scaling the process. Thus, finding the adequate aeration conditions is paramount for a scalable process. It was also determined that increased feeding was only beneficial when coupled with slow pump addition as it diminished the cellular stress perhaps due to attenuating nutrient concentrations oscillations and exposing the cells to increased osmolarity at very low feed addition rates. Thus, it can be said that feed amount and dosage method are important components of feeding strategy development. This dosage strategy (slow pump) was also observed to allow for consistent lactate absorption which is known

to be a good process indicator in CHO cells. It was also determined that dynamic feeding strategies (oxygen uptake rate and bio-capacitance based) can increase titer when compared to the control process. Given the increase of monitoring technologies in the bio-therapeutical industry, it is then appropriate to begin designing strategies that react directly to measured signals as it can be easily automated using feedback control loops, thus diminishing the day-to-day workload that operators have. The suggested signals are bio-capacitance and OTR (if accurate DO control is achieved) given their close links to viable cell volume (biovolume) rather than cell counts and consequently their indirect relationships to increases in metabolic demands (larger cells have been observed to have increased respiratory demands). Alternatively, if no accurate DO control can be achieved OUR estimation be realized through standard respiration tests in which cyclical measurement of the DO extinction profile is realized when air supply is stopped [246]. An additional take-away from this research is that feeding strategies and processing parameters must be optimized in tandem since the positive impact of dosage and increased feeding can be undone by adverse aeration conditions and inappropriate DO setpoints. Future work can center around evaluating process related impacts on quality profiles of the SARS-CoV-2 spike protein (glycan monosaccharide analysis, human ACE2 affinity, trimerization status and purity, identification of the N- and C-terminal and thermal stability). Importantly, product quality attributes for this cell pool have been shown to be highly robust and comparable to cell line expression methods. [247]

CHAPTER 5 ARTICLE 3: A RECURRENT NEURAL NETWORK FOR SOFT SENSOR DEVELOPMENT USING CHO STABLE POOLS IN FED-BATCH PROCESS FOR SARS-COV-2 SPIKE PROTEIN PRODUCTION AS A VACCINE ANTIGEN

The article titled "A recurrent neural network (RNN) for soft sensor development using CHO stable pools in fed-batch process for SARS-CoV-2 spike protein production" was submitted to *Biotechnology and Bioengineering* on May 16, 2024, it covers objective 4 of this thesis. The development of a one-step-ahead soft sensor is detailed. Test metrics demonstrated strong linearity and low normalized root mean squared error (below 0.5). Soft sensor application for process monitoring is discussed and improvements based on increasing data availability is suggested. I was responsible for data curation, data analysis, model implementation and writing - original draft.

Sebastian-Juan Reyes[1,2], Robert Voyer[1], Yves Durocher[1], Olivier Henry[2], Phuong Lan Pham[1]

[1]Human Health Therapeutics Research Centre, National Research Council Canada, 6100 Royalmount Avenue, Montréal, H4P 2R2, Quebec, Canada

[2]Department of Chemical Engineering, Polytechnique Montreal, Montreal, H3T 1J4, Quebec, Canada

Correspondence

Phuong Lan Pham, Human Health Therapeutics Research Centre, National Research Council Canada, 6100 Royalmount Avenue, Montréal, H4P 2R2, Quebec, Canada.

Email: phuonglan.pham@nrc-cnrc.gc.ca

Olivier Henry, Department of Chemical Engineering, Polytechnique Montreal, Montreal, H3T 1J4, Quebec, Canada.

Email: olivier.henry@polymtl.ca

Abstract

Fed-batch recombinant therapeutic protein (RTP) production processes utilizing Chinese Hamster Ovary (CHO) cells can take a long period of time (> 10 days). Within this period not all features may be measured routinely and in fact some are only determined once the process is terminated, complicating in-process decision making. As a consequence, utilizing routine current day bioreactor online data to aid in next day predictions is a promising strategy. The article details the development of a proposed soft sensor that merges current day bioreactor online data and offline historical sampling data to generate predictions about the next day of the production process. This approach demonstrated the ability to track product titer, cell growth, key metabolites (lactate and ammonia), and cumulative glucose consumption across the 17-day process with low error (normalized root mean squared error (nRMSE) and normalized mean absolute error (nMAE) below unity). It was also demonstrated that the same model architecture could effectively soft sense product titer and metabolic profiles (glucose, lactate, ammonia) without having sampling day's offline data as inputs to the model. This suggests that the proposed model could act as a true soft sensor of hard-to-determine variables such as the trimeric SARS-CoV-2 spike protein that relies on end-of-process measurements to acquire the data (labor-intensive semi-quantitative SDS-PAGE gels or ELISA assay). Instantaneous specific glucose consumption rates (qGluc) were also predicted and showed good agreement with experimental measurements, further offering opportunities for online glucose control. Within the process analytical technologies (PAT) framework, it stands to reason that in the future work, the utilization of additional online sensor data such as oxygen uptake rate

(OUR) and bio-capacitance which stand outside standard measured variables (temperature, pH, DO, gas sparging, agitation) can further improve model performance supporting the promise of model architecture approach in analyzing data rich processes.

KEYWORDS

data driven sensor, recurrent neural network (RNN), soft sensor, process analytical technologies (PAT), SARS-CoV-2 spike protein, CHO stable pool, CHO fed-batch bioreactor production

Introduction

A soft sensor is the concomitant use of software-implemented models (soft) and hardware devices (sensor) to gather and gain new information about the process [311]. This is key because without the use of soft sensors, that is to say just exclusively using a sensor, it would be impossible to derive the same information [150,152,312,313]. At its core, these soft sensors are used with the explicit purpose of leveraging on-line data in order to infer quantitative information about complex process variables that are impossible to measure directly in a sterile system or can be measured at a very low sampling frequency [314]. Consequently, soft sensors can become useful tools in terms of monitoring and control applications within the biopharmaceutical industry [150,152,312,313]. A well-developed soft sensor that considers the need of its stakeholders should, in theory, results in a reduction of operational surveillance and maintenance work. Additionally, soft sensors should

increase the interpretability of the results of culture runs given the capacity of the models to relate various key variables to each other [311]. Given this promise, soft sensors are perfect candidates for the PAT initiative to contribute towards automated control [311,314]. Soft sensors can be split into three categories: model-driven sensors, data-driven sensors, and hybrid models [311,315].

Model-driven sensors involve mechanistic models that are based on engineering principles, like mass or energy balances. They can provide an understanding of the processes that is inherently ingrained in biological insights [315]. Such models are capable of introducing known culture conditions such as media composition and/or culture performance indicators (cell growth, titer) to set up tangible models. Because of these characteristics, model-driven sensors can exploit known kinetic equations that capture dynamic changes of relevant variables [154,316]. In essence, these soft sensors incorporate reaction kinetics, transport phenomena, and thermodynamic constraints into the model [154]. However, these types of soft sensors must go through rigorous phases of parameter identification, uncertainty, and sensitivity analysis to properly validate said models. It must be noted that in the case where the model is reproducible and reliable, biological interpretable information is provided to the end user that can increase the understanding of the production process [154,311]. Model-driven soft sensors can be split into two distinct categories: dynamic models and steady-state models. Dynamic models depend on balances and kinetic presumptions to suitably express rate expressions as functions of the state variables [155]. On the other hand, steady-state models originate from mass and heat transfer laws. Good examples of such steady-state models can be flux balance analysis (FBA) or metabolic flux analysis (MFA) which are stoichiometric-rooted techniques routinely used

to characterize cell metabolism [157,158,317]. They are also useful in estimating intracellular fluxes by leveraging known extracellular analyte consumption or compound production rates as model constraints. Given that the quasi-steady state presupposition for intracellular metabolites is key, such models are considered static in nature [154]. On the contrary, kinetic models are usually expressed as a series of ordinary differential equations (ODEs) that can describe dynamic changes in metabolite concentrations, cell density, and protein expression during the cell culture process [150,152,312,313]. Because of this, cell growth and cell death can be unambiguously linked to changes in concentration of relevant nutrients and metabolic by-products. In addition, protein expression has been linked to cell growth and amino acid metabolism [154,318,319]. As a direct consequence, dynamic models can be designed with varying levels of complexity conditioned on the assumptions made by the researcher regarding the culture system in the bioreactor. This diversified degree of complexity can be tuned by considering heterogeneity within cell population or by acknowledging the existence of known cellular compartments and their respective behaviours. On the other hand, dynamic models can be simplified if reactions are lumped to rate limiting steps [154,318,319]. Because of such caveats, model-driven sensors can be very complex and time consuming to develop [320].

Data-driven soft sensors utilize multivariate data analysis (MVDA) techniques such as partial least square (PLS), principal component regression (PCR), and non-linear regressions such as artificial neural network (ANN), and support vector machine regression (SVMR) in order to relate input features to predict desired variables [171-173,175,321-326]. These non-linear models are particularly useful in understanding mammalian cell cultures given the fact that a lot of the interactions between key metabolic and process variables remain

unknown or are highly cell line specific [311]. PLS regression and ANN are notably used to analyze spectral data. Under such conditions, the spectral data is used as input and linked to outputs such as substrate concentrations, biomass, cellular viability, or product titer [124,327]. Thanks to such models, it is possible to predict critical process parameters (CPPs) that are not available through the spectral signals or multi-sensor data alone but arise from the deconvolution of the datasets generated from such sensors. This, in theory, is an advantage over mechanistic models given that online measurements (temperature, pH, DO, oxygen flowrate, base addition, dissolved carbon dioxide flow rate, oxygen uptake rates, bio-capacitance signals, Raman spectral data, cell volume) are not directly coupled to cell counts. Metabolic parameters such as lactate production/consumption, ammonia production/accumulation, and glucose consumption can be utilized to help predict said variables (protein expression, cell growth, etc.). Alternative approaches applied in bioprocessing include multiple linear regression, k-nearest neighbors (KNN), regression trees, ensemble approaches (Gradient Boosting Machine, Extreme Gradient Boosting, Adaptive Boosting, Random Forest) and Gaussian process regression [328]. Since mammalian cell culture data is complex both in its time dependent variation and multivariate nature, methods developed for sequence forecasting have been applied. Even though ANN can capture dynamics of non-linear systems like cell culture runs, RNN is a subclass of ANN that better captures the internal temporal dependencies of a system. These architectures are particularly useful for making t-step ahead predictions of relevant state variables [329]. They have recently been applied in predicting biomass growth before and after transfection of a recombinant adeno-associated virus (rAAV) production process [181]. This was done by utilizing cumulative oxygen sparged, dissolved oxygen values, and

cumulative dissolved oxygen tension as features and relating their time related variance to cellular growth. RNN models for multivariate estimation of mammalian cell culture data (total cell density, viable cell density, viability, lactate, glucose, titer) have also been developed [330].

Hybrid models, known as grey box models, are another relevant class of soft sensors. These types of soft sensors can be considered to be a combination of data-driven soft sensors and mechanistic model-driven soft sensors. They have the capacity of utilizing the benefits of each method [150,152,312,313,316,320,331-336]. Various architectural techniques exist when developing hybrid models and they can fall in three general categories: i) Calibration ii) Composition iii) Transformation. Calibration architectures utilize black box models to reduce mechanistic model errors. Composition architectures utilize black box models to estimate unknown terms within a mechanistic model [320,337,338]. Lastly, a transformation approach utilizes mechanistic models to generate data rich environments from which training a black box model is possible [339]. Examples of these are state observers that integrate dynamic modelling (white box models) and data-driven modelling (black box models). This is realized by updating state estimates derived from noisy measurements and gradually reducing the estimation error with each iteration [311]. This is usually done assuming linear dynamics within the process and a Gaussian distribution for the error terms. Under such assumptions, a Kalman filter can be used. However, given that the process dynamics within a bioprocess are non-linear in nature, the extended Kalman filter may be applied. This method realizes a piecewise linearization through a first order Taylor series expansion [162]. Another important version of the Kalman filter that is widely used for non-linear systems is the unscented Kalman filter. This method employs an unscented

transform to avoid relying on a Taylor series expansion of the system of equations to linearize the model [340]. This method can be advantageous since the unscented transform allows non linearizable functions to be used as a state observer and thus black box techniques such as support vector machine regression (SVMR) can be utilized so as to relate an online sensor output to a non-online variable [340]. It must be noted that because the accuracy of a hybrid soft sensors (grey box models) can be significantly impacted by the accuracy of the mechanistic model embedded within the grey box model, the mechanistic model requires extensive validation to ensure it can successfully represent the process [164]. Extended Kalman filters have been applied in recombinant adeno-associated virus (rAAV) production processes in which online viable cell densities (bio-capacitance signal) and an unstructured mechanistic model are leveraged along with neural ODEs that estimate cell specific rates. This results in a soft sensor that is able to estimate continuously other state variables that are measured at low frequency [341]. Likewise, hybrid extended Kalman filters which utilize partial least squares (PLS) in order to estimate specific rates in the model have been developed [342]. Hybrid models can also describe the biological system by way of a mechanistic framework but define the cell specific rates through statistical expressions [150,152,312,313]. The mechanistic framework inherently constrains the solution space of the model and thus, the statistical cell-specific rate expressions can be automated [343]. Within this structure, PLS or ANN prediction resulting from multi-wavelength spectra or multivariate parameters can be fed as inputs into a mechanistic model [344,345]. XGBoost regressors, random forest regressors and multilayer perceptron (MLP) regressors have been used to estimate cell specific rates (specific growth rate, specific productivity, and specific cumulative glucose consumption) throughout a fed-batch process [338]. These updated

specific rates are then utilized as parameters in a mechanistic model to predict relevant culture outcomes (viable cell density, titer, and cumulative glucose consumption) [338]. Hybrid models have also been utilized for the prediction of key product quality attributes (CQA) (impurity levels, charge variants species, intact mass, total low molecule weight (LMW), and N-glycan profiling) by coupling a propagation model that describes the time evolution of cell culture variables (viable cell density, glucose, glutamine, glutamate, lactate, ammonia, cell viability, and titer) with a PLS model that regresses quality attributes as a function of cell culture variables and process conditions [346]. Alternatively, to increase the overall data richness of a production process, the utilization of unstructured mathematical models would allow the user to generate cell line relevant data so as to create information rich environments that can be purposed to train nonlinear deep learning regression techniques like recurrent neural networks (RNN) [347].

In this article, the development of a multivariate long-term time series forecasting model is detailed. This model is capable of realizing one step ahead predictions of key state variables (titer, viable cell density, cumulative glucose consumption, lactate, ammonia) by relying on both offline sampling data and bioreactor online data. This is key given that the offline sampling data is unevenly spaced with respect to time (every other day or every two days; e.g., total of 17 process days but only 10 measurements) thus reliance on online data (temperature, pH, base addition, DO, integral of DO, cumulative O_2 flow, cumulative CO_2 flow) is required to update predictions on days in which no offline sampling data was available. The model was developed on a rich dataset generated with multiple cumate inducible CHO-GS cell stable pools expressing variants of the SARS-CoV-2 spike protein with changes in process conditions (cell passage number, Methionine Sulfoximine (MSX)

supplementation at induction). The important advantage of this data driven method when compared to mechanistic or hybrid modeling is that the fully pre-trained model can be readily applied by non-experts as it needs no knowledge about boundary conditions or metabolic networks. This can thus be readily transferred to production processes without additional training on its operators. The model can qualitatively capture the dynamics of metabolic and protein expression profiles by relying exclusively on bioreactor online data and easily accessible viable cell counts throughout the whole 17-day process. This model advantageously soft senses hard-to-measure variables like the SARS-CoV-2 spike protein in a daily fashion.

Materials and Methods

Stable CHO-GS Cell Pool and Small-Scale Cell Culture Conditions

Three stable CHO-GS cell pools expressing SmT1 trimeric spike proteins namely Wuhan Tagless (WuTL), Delta (De), and Beta (Be) variant were generated as described previously [200,201,220]. Stable pool cells were thawed and grown in BalanCD CHO Growth A medium (Fujifilm/Irvine Scientific) supplemented with 50 µM MSX (L-Methionine Sulfoximine, Sigma-Aldrich) and 0.1% (w/v) Kolliphor P188 surfactant (Sigma-Aldrich). 125-mL (20 mL working volume) shake flasks without baffles (Corning) were used for cell maintenance. The flasks were shaken at 120 rpm (25 mm orbital diameter) in an incubator regulated at 37°C, 5% CO_2, and 75% relative humidity. Cells were passaged every 2 or 3 days to keep a maximum viable cell density between 2-3x10^6 cells/mL.

Cell Culture Analytical Methods

Viable and total cell density, cell viability, main metabolites (glucose, lactate, ammonia) were measured utilizing the previously reported methodology [200,201,220]. Briefly, cell counts were performed with Innovatis Cedex (Roche) or ViCell Blue (Beckman Coulter) automated cell counter using trypan blue dye exclusion assay. Key metabolites such as glucose, lactate, and ammonia were determined using the Vitros 350 Chemistry System (Orthoclinical Diagnostics). Volumetric protein titers were estimated using TGX Stain-free SDS-PAGE gels (Bio-Rad) quantification method. Table 5.1 summarises online and offline measurements.

Table 5.1. Process variables considered in the model.

Offline Measurements	Cell Growth	Viable Cell Density (VCD) (cells/mL)
	Metabolites	Lactate (mM)
		Ammonia (mM)
		Cumulative Glucose Consumed (cGC) (mM)
	Protein Production	Titer (mg/L)
Online Continuous Measurements	pH Control	pH profile
		Base Addition Volume (mL)
		Total Carbon Dioxide sparged (mL)
	Oxygen Requirement	Total Oxygen sparged (mL)
		Dissolved Oxygen (DO) (% of air saturation)

		Integral DO (%*Day)
	Temperature Control	Temperature (°C)

Fed-batch Cell Culture Process Conditions

All productions were conducted in parallel benchtop bioreactors 0.75 L Multifors 2 (Infors). Corning shake flasks were used to generate the seed trains. The bioreactors were seeded at 0.4 x 10^6 cells/mL and cultivated for 17 days. Temperature downshift (37°C to 32°C) was realized 3 days after seeding. A pH shift was conducted 2 days post-seeding (from 7.05 ± 0.05 to 6.95 ± 0.05). A dissolved oxygen (DO) set point of 40% (of air saturation) was chosen. Micro-spargers with a 0.0033 vvm (volume of gas per initial working volume per minute) air cap was implemented in a cascade air/oxygen strategy. Air flowrate was automatically increased to a selected maximum value (air cap) then remained constant to the end. Pure oxygen was injected as needed to maintain the DO setpoint. CO_2 and an in-house mix of $NaHCO_3$/NaOH were used to maintain pH in its selected deadband. Production induction was initiated with the addition of 4-Isopropylbenzenecarboxylate (Cumate, ArkPham). Cultures were fed with BalanCD CHO Feed 4 (Fujifilm/Irvine Scientific) and supplemented with glucose as needed to maintain glucose concentration above 17 mM (3.06 g/L) for the next sampling point. Samples were taken from the bioreactors on days -3, -2, -1, 0, 3, 5, 7, 10, 12, and 14 dpi (days post-induction) for off-line analysis, while feeding was realized in a bolus dosage from 0 dpi (induction day) onward. Cell passage number was varied across different batches (passage 5, 8, and 11) to study the impact of cell age on pool expression stability. It is known that cell pool is heterogenous composing of different cells with different expression level in contrast to

stable clone [348]. Therefore, it is critical to determine cell age operation window to avoid a significant expression loss when cell passage number increases [201]. Additionally, MSX supplementation (75 µM) at induction (0 dpi) was also investigated to evaluate the impact of high MSX concentration on production performance. High MSX concentration has enabled increased protein expression observed in our previous unpublished data. This effect was also shown with other group [229].

Dataset and data handling methodology

The dataset is made up of 21 production runs. 10 runs were performed with the Delta pool (De), 6 runs with Beta pool (Be), and 5 runs with the Wuhan Tag-less pool (Wu-TL). For all the productions, viable cell density, cumulative glucose consumption, lactate, ammonia, and titer were measured or calculated. Cumulative glucose consumption was estimated by adding up the glucose consumed between sampling days (difference between media glucose concentration after feed and measured glucose concentration in next sampling day) thus keeping track of total glucose consumption. Online data from the bioreactor runs were also added into the dataset. DO, integral of DO (DO_{int}), total oxygen sparged, pH, base addition, total carbon dioxide sparged, and temperature were monitored or calculated daily such that direct comparison with the sampling day data could be made. Integral of the DO curve (DO_{int}) was chosen as a variable to aid in giving the model information about possible changes in the DO profile between monitoring days. Integral of DO, O_2 sparge rates, and CO_2 sparge rates were estimated by calculating the area under the curve of each signal through trapezoidal rule for numerical integration:

$$DO_{int} = \sum_{i=1}^{n} \frac{(time_{i+1} - time_i) * (DO_{i+1} + DO_i)}{2}$$

$$Total\ Oxygen\ sparged = \sum_{i=1}^{n} \frac{(time_{i+1} - time_i) * (O_{2\_i+1} + O_{2\_i})}{2}$$

$$Total\ Carbon\ Dioxide\ sparged = \sum_{i=1}^{n} \frac{(time_{i+1} - time_i) * (CO_{2\_i+1} + CO_{2\_i})}{2}$$

Where *i* is the counter that ranges from *1* to *n* (from the first data point until the end of the data in the time series); $time_i$ represents the time associated with the i^{th} data point; $time_{i+1}$ is the time associated with the $(i+1)^{th}$ data point; DO_i represents the value of DO at the i^{th} data point; DO_{i+1} represents the value of DO at the $(i+1)^{th}$. Similarly, O_{2_i} and CO_{2_i} represent the respective gas flow value at the i^{th} data point, while O_{2_i+1} and CO_{2_i+1} signify the $(i+1)^{th}$ data point of the respective gas flows.

Details of the RNN methodology

Recurrent neural networks (RNN) are a family of neural networks for processing sequential data. There is a shared similarity with multilayer perceptron (MLP). The general network structure is represented by an input layer, one (or numerous) hidden layers, and an output layer. However, the RNN structure allows the model to carry over latent information from time step to time step thus capturing time varying profiles [349]. RNNs and MLPs have distinct differences that have significant implications for sequence learning. While MLPs can only map from input to output vectors, RNNs have the ability to map from the entire history of previous inputs to each output. This means that an RNN can leverage a memory of past inputs stored in its internal state to influence the network's output [350]. In fact, the

universal approximation theory applies to both RNNs and MLPs, but with different implications. For MLPs, it states that with enough hidden units, an MLP can approximate any measurable mapping from input to output. On the other hand, for RNNs, the equivalent result is that with a sufficient number of hidden units, an RNN can approximate any measurable sequence-to-sequence mapping [350]. RNNs are specifically designed for processing sequential data, just as convolutional neural networks (CNNs) which are specialized for processing grid-like data such as images [351]. RNNs excel at handling sequences of values, and they can scale to longer sequences compared to networks without sequence-based specialization. Additionally, most RNNs are capable of processing sequences of variable length, providing flexibility in handling diverse data inputs [351].

The latent information, which functions as a network memory, is captured within the hidden state (h_t), which is updated at each iteration as described in the equations below:

$$MLP_{hidden}:$$

$$h_t = W_{hidden} * \mathbf{\Phi}(W_{in} * [h_{t-1}, O_t, X_t])$$

$$MLP^i_{sampled}:$$

$$\dot{X}^i_{t+1} = W^i_{out} * \mathbf{\Phi}(W^i_{h\_in} * h_t)$$

MLP_{hidden}- multilayer perceptron that receives sampled data and online data; *t* – discrete time index; h_t – hidden state at time step *t;* W_{hidden} – hidden state weight matrix; **Φ** – standard hidden layer activation function in ANN (logistic, hyperbolic, tangent, sigmoidal, etc.); W_{in} – input weight matrix; h_{t-1} – hidden state at time step *t-1*; O_t – online variable at time t; X_t – sampled input vector at time t; $MLP^i_{sampled}$ – multilayer perceptron that projects the resulting hidden state to a sampled variable space and predicts the updated time series prediction for all *i* sampled variables (titer, viable cell density, lactate, ammonia,

cumulative glucose consumption); $\dot{X}_{t+1}^{i}$ – state vector predictions at a future time discrete index for all *i* sampled inputs (titer, viable cell density, lactate, ammonia, cumulative glucose consumption); W_{out}^{i} – intermediate hidden state to predicted variable weight matrix for all *i* sampled variables; **Φ** – standard hidden layer activation function in ANN (logistic, hyperbolic, tangent, sigmoidal, etc.); $W_{h\_in}^{i}$ – weight matrix to transition from a global hidden state to a variable specific hidden state for all *i* sampled variables; h_t – hidden state at time step *t* [329,330].

For the purpose of this paper, the rectified linear unit (ReLu) function was utilized. The hidden state serves as an internal representation of the network and captures information about previous inputs in the sequence. The hidden state is updated at each time step and serves as a way for the network to maintain information about the context or history of the input sequence. The initial hidden state h_0 was initialized with a zero vector which can be interpreted as carrying over no information from the past as the process had not yet been initialized. As can be seen on Figure 5.1.

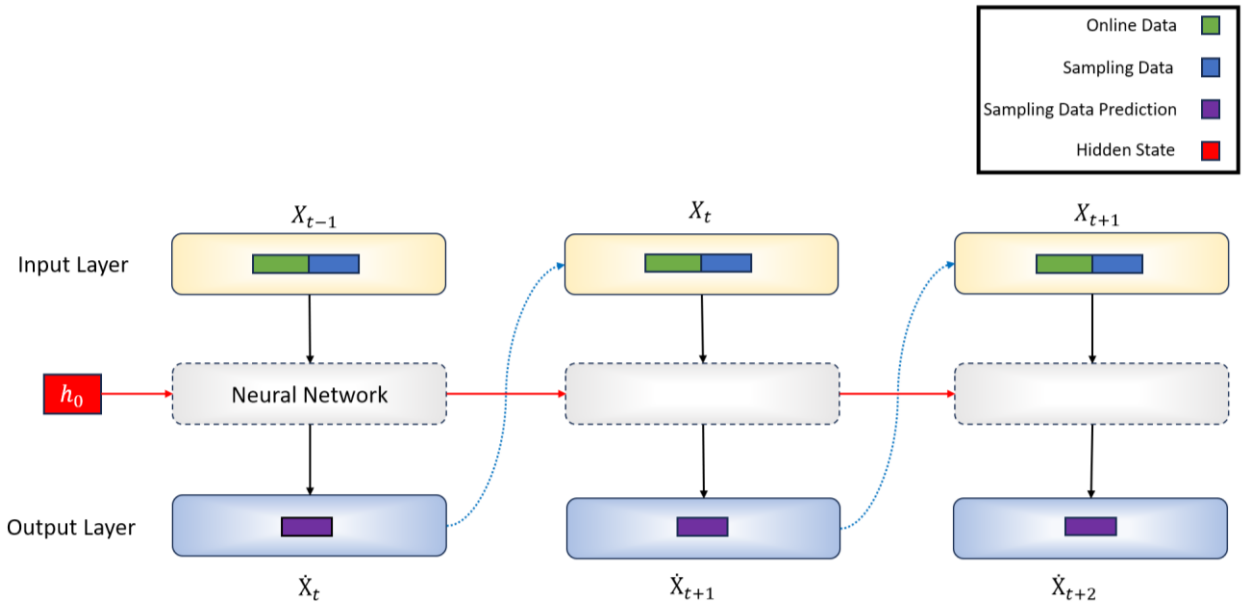


Figure 5.1. Soft sensor architecture for predicting next day sampling data.

Online bioreactor data (total oxygen sparged, total carbon dioxide sparged, pH, base addition, DO, integral of DO, temperature) and measured sampling data (lactate, ammonia, cumulative consumed glucose, viable cell density, titer) along with an initial hidden state are received as inputs to a neural network. This neural network outputs the next discrete time prediction of each sampled data along with an updated hidden state to be used in the next iteration. As can be seen in Figure 5.2. This process is repeated for all days during the production process (17 days). If there is no available sampling data, then the bioreactor online data and the previous output predictions (rather than the ground sampling data) will be used as inputs in the next iteration.

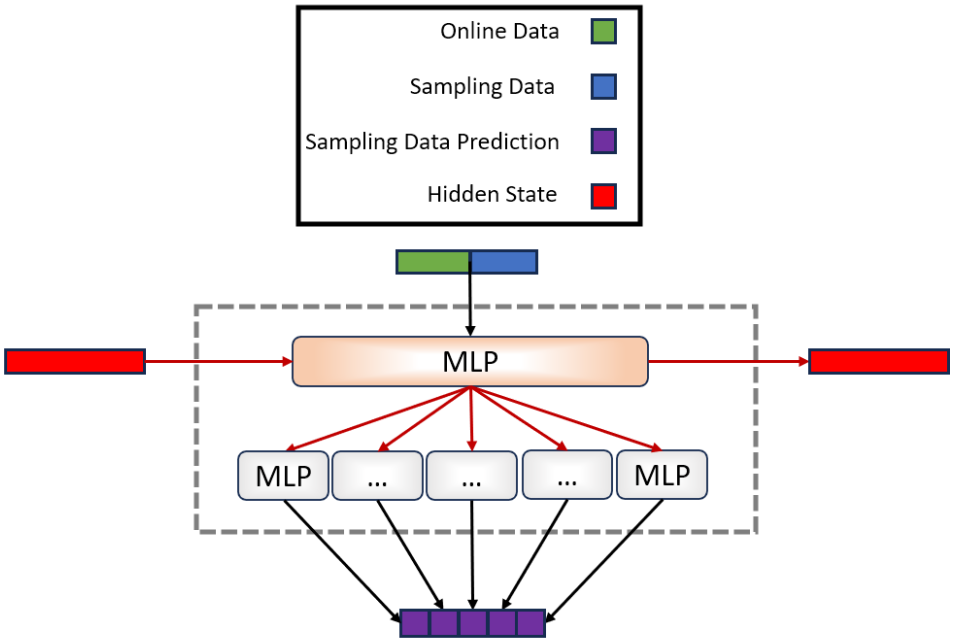


Figure 5.2. Internal neural network architecture.

Bioreactor online data, offline sampling data, and hidden state values are received as inputs to an MLP. The MLP outputs an updated hidden state which serves as input for several MLPs (an MLP for each measured variable). Each MLP projects the resulting hidden state to a sampled variable space and predicts the updated time series prediction for each variable. The hidden state is stored for the next iteration.

Data preprocessing was done with Pandas and Numpy which are important libraries of Python (version 3.9.13) programming [352,353]. Architecture design and training of the soft sensor were done with PyTorch [354]. The MLPs were all conformed by 256 hidden units. All MLPs utilized a rectified linear unit activation function (ReLu). Learning rate was set to 1.6×10^{-4}. The learning rate in neural networks is a hyperparameter that determines the size of the steps taken during the optimization process, or how quickly the model adjusts

its weights in response to the estimated error each time it updates. All input data were mean centered and standardized for each variable. Stochastic gradient descent (SGD) was employed as the optimizer with momentum of 0.97 and weight decay of 0.125. The training phase included 5000 epochs. An 80% training and 20% test split on the dataset was approximated. Training-test split was randomly realized such that 8 Delta cultures are used to train the network while 2 are left for testing. Similarly, 5 Beta cultures were used to train and 1 was left over to test. Lastly, 4 Wuhan Tag-less cultures were used to train and 1 culture was left over to test. Regarding model validation, root mean squared error (RMSE), mean absolute error (MAE) and coefficient of determination (R^2) were used as validation metrics to evaluate model performance for each predicted feature (glucose consumed per day, lactate, ammonia, viable cell density, titer). RMSE, MAE, and R^2 were first calculated for every batch prediction then averaged out across batches. When the RMSE and MAE of each feature for every batch are divided by the corresponding standard deviation of each feature of every batch, unbiased estimates (nRMSE and nMAE) can be calculated such that model performance can be easily compared across features.

$$RMSE = \sqrt{\frac{1}{n} * \sum_{i=1}^{n} (y_{ground\_truth} - y_{pred})^2}$$

$$MAE = \frac{1}{n} * \sum_{i=1}^{n} abs(y_{ground\_truth} - y_{pred})$$

$$nRMSE = \frac{\sqrt{\frac{1}{n} * \sum_{i=1}^{n} (y_{ground\_truth} - y_{pred})^2}}{Standard\_Deviation}$$

$$nMAE = \frac{\frac{1}{n} * \sum_{i=1}^{n} abs(y_{ground\_truth} - y_{pred})}{Standard\_Deviation}$$

Here, y_{ground_truth} is the real sampled data of a feature; y_{pred} is the prediction of a feature from the model; i represents the i^{th} value of a given feature in a batch; and n represents the total number of values of a feature in a batch; $Standard\_Deviation$ is the standard deviation of a batch for a given feature.

Results and Discussion

The prediction versus ground truth sampled data scatter plot for each of the 17 training batches is shown in Figure 5.3A clear linear relationship (R^2 values above 0.95 as indicated in Table 5.2) is seen in all feature predictions (titer, cumulative consumed glucose, ammonia, lactate, and VCD). The scatter plot results can center around the diagonal line which represents the R^2=1 ideal model. It must be noted that for lactate and VCD predictions (Figure 5.3D and 5.3E), there exists a slight deviation from the R^2=1 at higher concentrations. This dataset was centered on studying the impact of increase cell age on protein production, as well as determining if increased MSX (from 50 µm to 125 µM) at induction improved protein production outcomes. It was observed that passage impact was only an important factor for the Delta pool, while increased MSX was found to have no impact for all 3 pools [348].

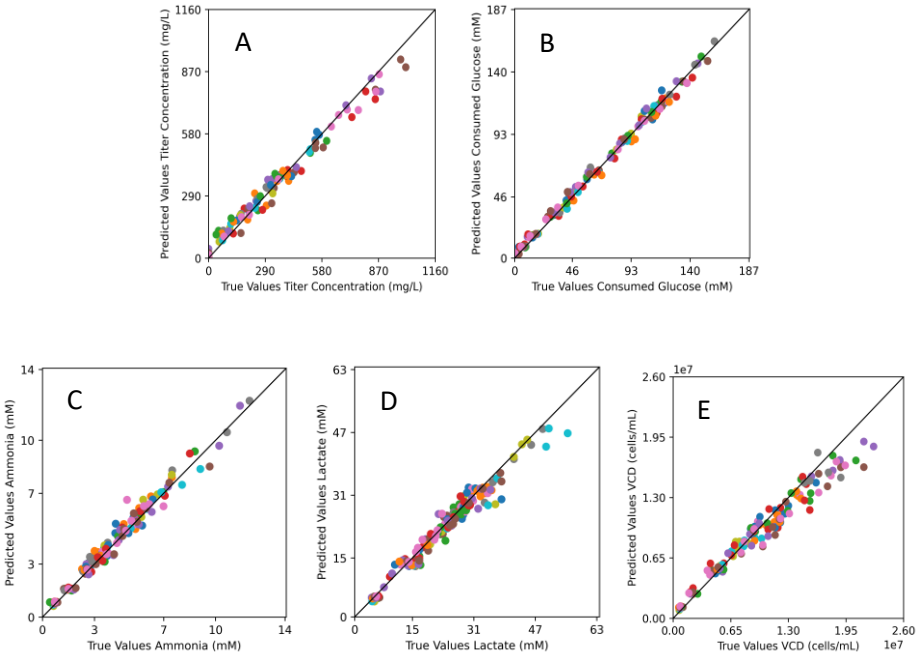


Figure 5.3. Model results for trained features in the training dataset comprised of 17 cultures (8 Delta pool batches, 5 Beta pool batches, and 4 WuTL pool batches).

Scatter plots of every batch prediction versus every batch sampled data for the whole training set. Diagonal line represents the R^2=1 ideal model. (a) product titer, (b) cumulative glucose consumed, (c) ammonia, (d) lactate, (e) viable cell density (VCD). Each color r E nts a batch culture (17 colors for 17 cultures) with 9 data points for each culture resulting in 153 points (17x9) on the graphs.

With respect to the time series progression of each feature (Annex Figure 26S), the model tracks sudden changes in titer values concomitant with the onset of induction (3 to 12 dpi) (Figure 26S, A). It is able to keep track of high glucose consumption during the growth phase (-2 to 0 dpi) followed by a decline in volumetric glucose consumption probably due to the temperature downshift at 0 dpi, then an increase in glucose consumption during the active production phase (3 to 14 dpi), (Figure 26S, B). Ammonia profile tracking (Figure 26S, C) is possible such that high rates of ammonia accumulation

are observed during the growth phase (-2 to 0 dpi) followed by a plateau in extracellular ammonia accumulation concomitant with slowed down metabolic activity probably due to the 0-dpi temperature shift from 37°C to 32°C. A shift to ammonia accumulation in the late stages of the production phase (7 to 14 dpi) was revealed. The model can also track rise in lactate accumulation (-3 to 0 dpi) during the growth phase and the onset of lactate consumption (5 to 7 dpi) during the protein production phase (Figure 26S, D). The training results show tracking of the exponential growth of viable cell densities (-2 to 3 dpi) and its resulting plateau and decline phase (7 to 14 dpi) (Figure 26S, E). All the models have similar normalized RMSE values indicating the proposed model is able to track every feature equally well.

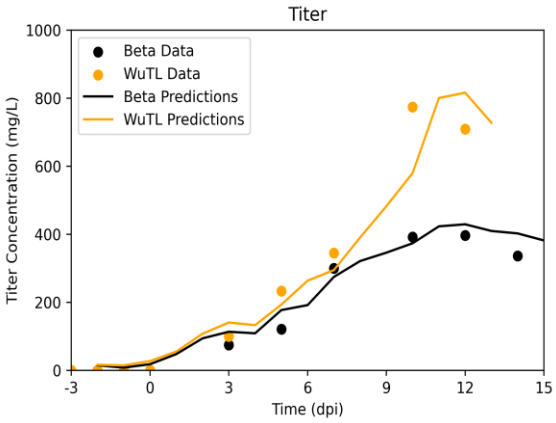
Next, to determine if the model can generalize, it was tested on 4 production runs that were not utilized in the training process. These hold-over cultures consisted of 2 Delta cultures, 1 Beta culture, and 1 WuTL culture (Figure 5.4). The WuTL production was terminated 2 sampling days earlier than normal (harvest at 12 dpi rather than 14 dpi). As it can be seen from Figure 5.4, time profile tracking of key features is possible even in non-sampling days. This is key given the fact that manufacturing process lasts 17 days in which it is unpractical to have every day sampled values especially for the long-to-measure outputs such as protein determined by SDS-PAGE which is the case for SARS-CoV-2 spike protein. For these non-sampling days, predictions are carried out by relying on previous day model predictions and the online data of the bioreactor (base addition, oxygen flow, DO, integral of DO, pH, temperature). In Figure 5.4A, it is clear that titer time series tracking is possible despite differences in pool protein production behavior. For example, the Beta and WuTL pool demonstrate vastly different endpoint titer results despite having

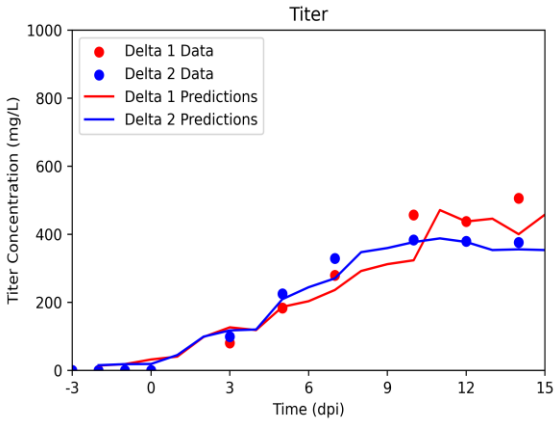
identical kinetics from 0 to 7 dpi (Figure 5.4A, left side). Both Delta pools demonstrate similar protein production behavior until 9 dpi from which Delta 1 culture slightly outperforms Delta 2 culture by close to 100 mg/L (Figure 5.4A, right side). This subtle increase in the end phase of the culture is also captured by the model. From Figure 5.4B, ammonia profiles can be detailed. It is clear that the Beta and WuTL pools exhibit different ammonia accumulation kinetics after the onset of protein production (0 dpi). Increased ammonia accumulation is evident in the WuTL culture following the onset of induction while the Beta pool has an accumulation lag between 0 and 7 dpi followed by a rapid ammonia accumulation after 7 dpi. These changes in kinetics are predicted by the model despite having no measured sampling values at the points in which the kinetics begin to diverge. Similarly, for the Delta cultures, both productions show a plateau in ammonia accumulation from 0 to 7 dpi and then an increase accumulation until 14 dpi (Figure 5.4B, right side). This behavior is tracked by the model even between 7 to 10 dpi in which no sampling data exists due to the 3-day weekend. From Figure 5.4C, it can be seen how the cumulative consumed glucose profile changes with culture time, namely the deceleration in cumulative consumption following temperature shift from 37°C to 32°C at 0 dpi. This reduction in glucose consumption can be linked to the cell growth slowdown that occurred at lower temperature. Only Delta 1 pool predictions overestimate (from 5dpi onwards) the total profile such that endpoint cumulative glucose consumed is overestimated by 20.5 mM (Figure 5.4C, right side). This may be caused by the fact that Delta 1 culture has high lactate accumulation (above 30 mM) (Figure 5.4D) and thus the model may be predicting the cumulative glucose consumption to be representatively higher. Figure 5.4D shows how lactate profiles can be tracked with the proposed model. For the Delta cultures, large

differences in peak lactate values are observed which in turn impacts the lactate re-absorption profile (5-14 dpi). These differences in kinetics for the same pool in two different culture runs are captured by the model. Moreover, it is worth mentioning that Delta 1 culture was performed with cells having a passage number 11 while Delta 2 culture was done with cell passage number 8. The large deviation in lactate concentration could be related to their different cell passage numbers as previously mentioned [348]. The latter work shows that Delta pools demonstrated increased variability with increased culture age. Whether pool age could influence cell metabolism in the case of Delta pool remains to elucidate. The lactate time courses of the Beta and WuTL pools are tracked throughout culture, even in the case where a sudden lactate re-production occurs after 10 dpi (WuTL) (Figure 5.4D). This increase lags the real value presumably because in the absence of measured lactate values, online signals like pH and base (that are known to relate to lactate concentration changes) did not show a marked change in response to the lactate accumulation (pH went from 6.96 at 10 dpi to 6.99 at 11 dpi to 6.97 at 12 dpi and base addition was never triggered as pH values stayed within the allowed deadband). Figure 5.4E shows that cell growth profiles can be tracked across the 17-day culture run. It is noteworthy that even though WuTL and Beta cultures had identical viable cell density profiles until 3 dpi, the Beta culture obtained a secondary growth spurt which is consequently tracked by the model while the WuTL culture remains in a plateau phase. Similarly, for the Delta cultures (Figure 5.4E, right side), identical viable cell density profiles are observed until 3 dpi from which Delta culture 2 enters a secondary growth phase while Delta culture 1 does not. Interestingly, both distinct profiles are captured by the model. This is important because when taking into account final titers, it suggests that

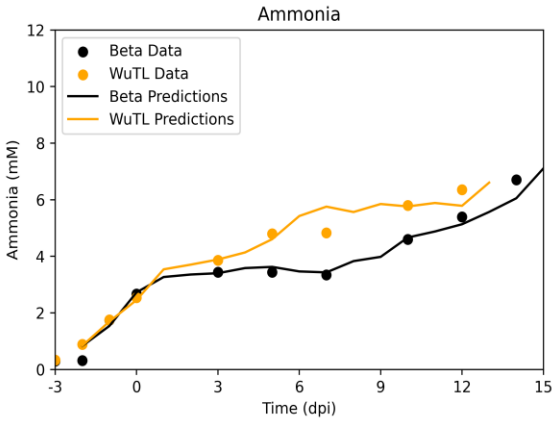
cultures that experienced secondary growth phases had lower endpoint titers. Consequently, monitoring that cell densities remain within a plateau phase during the protein production phase would be of interest. This argument can be supported by the development of a biphasic process strategy in which cells will be allowed to grow to certain high density then a process trigger such as lower temperature (31°C - 34°C from 37°C) or chemical addition (sodium butyrate, valeric acid) [355,356] will be introduced to keep cells in a biomass steady state while boosting the production.

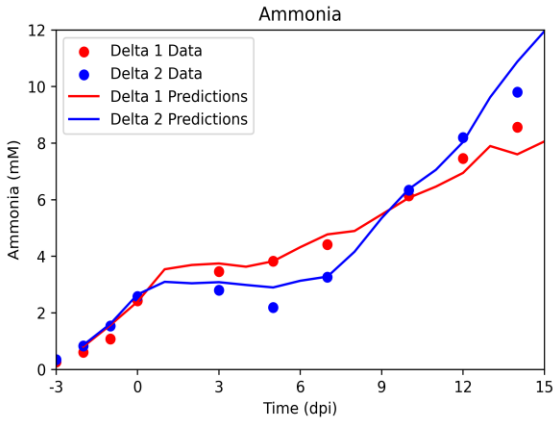
A


Titer
Beta Data
WuTL Data
Beta Predictions
WuTL Predictions
Titer Concentration (mg/L)
Time (dpi)

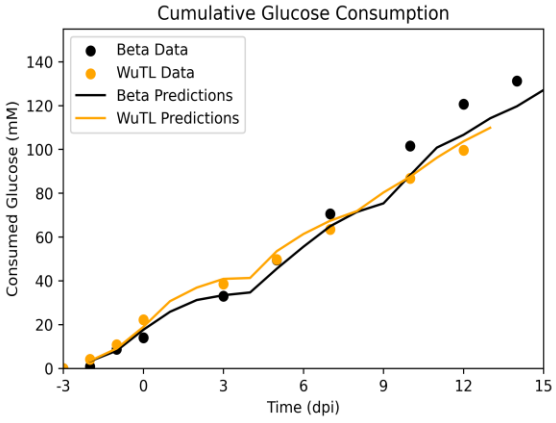

Titer
Delta 1 Data
Delta 2 Data
Delta 1 Predictions
Delta 2 Predictions
Titer Concentration (mg/L)
Time (dpi)

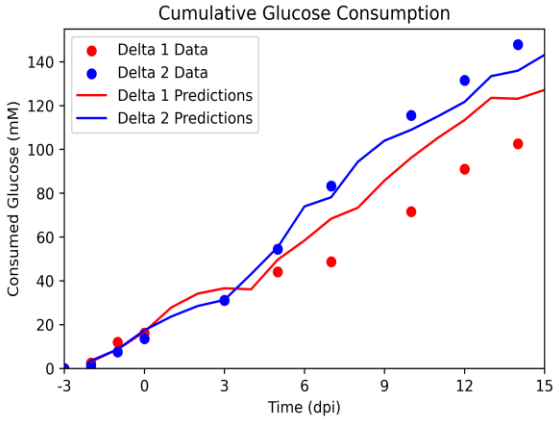
B


Ammonia
Beta Data
WuTL Data
Beta Predictions
WuTL Predictions
Ammonia (mM)
Time (dpi)


Ammonia
Delta 1 Data
Delta 2 Data
Delta 1 Predictions
Delta 2 Predictions
Ammonia (mM)
Time (dpi)

C


Cumulative Glucose Consumption
Beta Data
WuTL Data
Beta Predictions
WuTL Predictions
Consumed Glucose (mM)
Time (dpi)


Cumulative Glucose Consumption
Delta 1 Data
Delta 2 Data
Delta 1 Predictions
Delta 2 Predictions
Consumed Glucose (mM)
Time (dpi)

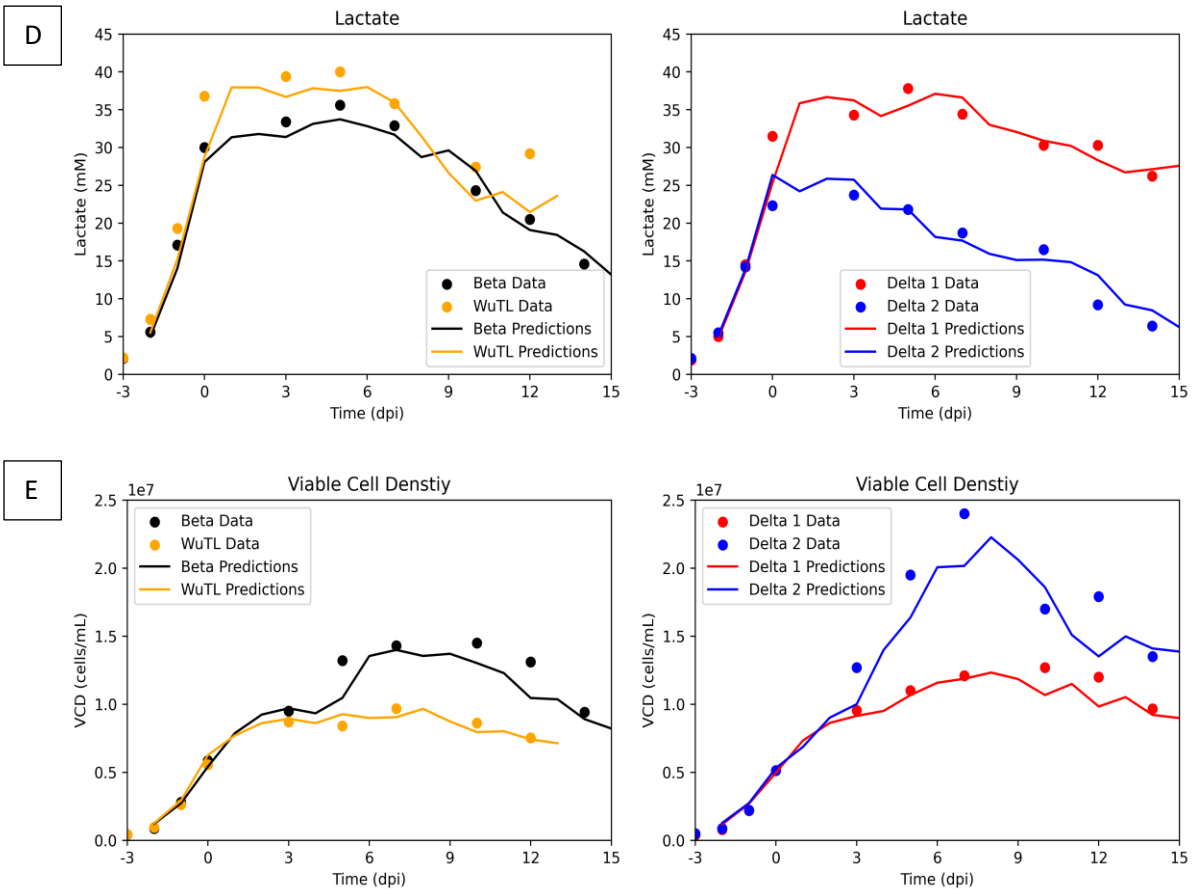


Figure 5.4. Model results for key features in the test dataset comprised of 4 cultures (1 Beta pool batch, 1 WuTL pool batch and 2 Delta pool batches).

Continuous lines are everyday predictions while dots are measured values. (a) product titer, (b) ammonia concentration, (c) cumulative glucose consumption, (d) lactate concentration, (e) viable cell density.

As it can be seen from Table 5.2, both train and test model predictions have strong linearity with respect to measured values ($R^2 \geq 0.9$). The lowest R^2 values in the test set correspond to lactate, titer, and VCD. These time series profiles displayed sharp variations (sudden lactate re-absorption or secondary growth phases) causing the model predictions

to lag measured values before converging again. When detailing the normalized RMSE and MAE, which take into account the standard deviation of the datasets, it can be seen that the predicted features have normalized error metrics below unity (<1). This demonstrates the applicability for the model to predict the time series variation of multiple features with highly nonlinear kinetics like cellular growth and lactate formation which can increase rapidly and then decline depending on the culture phase. The model's nRMSE metrics for VCD, titer, and lactate are in line with similar state-of-the-art applications of data-driven soft sensors for time series tracking [181,330].

Table 5.2. RMSE, MAE, nRMSE, nMAE, R^2 metrics for titer, ammonia, cGC (cumulative glucose consumed), lactate, VCD (viable cell density) for train and test datasets.

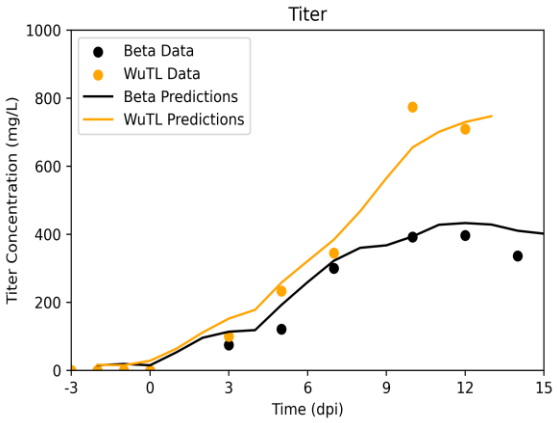
	Train RMSE	Train MAE	Train nRMSE	Train nMAE	Train R^2	Test RMSE	Test MAE	Test nRMSE	Test nMAE	Test R^2
Titer	33.55	27.71	0.13	0.16	0.99	51.50	38.82	0.24	0.18	0.96
Ammonia	0.36	0.28	0.14	0.18	0.97	0.39	0.27	0.18	0.12	0.98
cGC	3.35	2.80	0.06	0.08	0.99	7.95	6.20	0.20	0.16	0.99
Lactate	2.01	1.54	0.17	0.22	0.96	2.85	2.34	0.31	0.25	0.94
VCD	1235666	973513	0.19	0.24	0.97	1358272	1030309	0.26	0.20	0.96

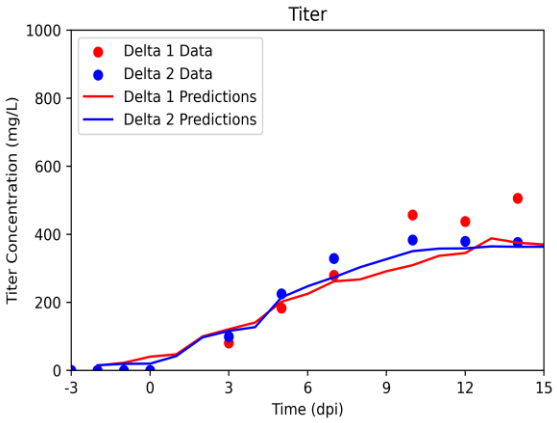
Given that the model relies on previous day predictions and current day online data to estimate next day features and that it was shown to work reasonably well during the non-sampling periods of the 17-day production process, it was hypothesized that the same model without alternative training for parameter tuning could be utilized to predict metabolic data and titer profiles in the absence of such data. Consequently, the model would rely on initial feature values which function as initial guesses, sampling day cell counts, and bioreactor online data to generate predictions of ammonia, cumulative consumed glucose, lactate, and titer throughout the 17-day process. The objective then was

to evaluate if the current model has suitable potential for true soft sensing capabilities which would be of interest for hard-to-detect variables such as non-antibody recombinant proteins. As it can be seen from Figure 5.5A and 5.4B, qualitative trends of titer and ammonia are tracked despite having no sampling day values of said features to aid in the updated prediction. Furthermore, glucose consumption (Figure 5.5C) is also tracked throughout the culture with only Delta 2 culture being overestimated at 14 dpi by 20 mM. Alternatively, lactate profiles for the Beta and WuTL pool are underestimated from induction until the end of the process (Figure 5.5D, left side). For the WuTL process, the lack of sampling lactate data did not allow the model to predict the sudden lactate re-production phase (10 dpi to 12 dpi). One of the reasons for this underestimation may be due to the fact that since key online data values like base addition are governed by a PID control, which is implemented with a pH deadband, indirect estimation of lactate may lag or be underestimated. One such example is the case of the WuTL and Delta 1 cultures. WuTL produces a higher peak lactate (40 mM > 37.8 mM) while having lower total base addition (12 mL < 13.3 mL). This pH deadband activation is then very connected to the lactate absorption which pushes the pH values away from the activating deadband limit. Additionally, predicting lactate re-production in the end phase of the culture within the context of pH deadbands also becomes difficult as the previous lactate consumption coupled with the base addition during the lactate production phase necessarily drives pH values closer to the upper edges of the pH deadband. Thus, any low amounts of lactate production will not be enough to activate base addition and will generally result in pH remaining closer to the upper edges of the pH deadband. For example, in the case of the WuTL culture from 7dpi until 12 dpi, daily average pH values are [6.96, 6.98, 6.99, 6.96,

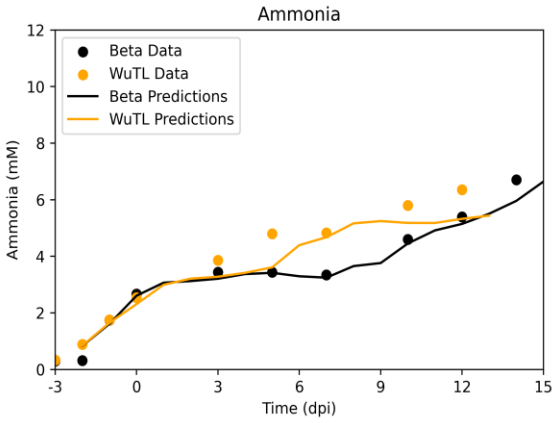
6.99, 6.99] and total carbon dioxide sparged values increased every day from 7dpi to 12 dpi [4603 mL, 4650 mL, 4680 mL, 5211 mL, 5866 mL, 6023 mL] to control the pH in the deadband (6.9-7.0). Similarly, although cell counts were added to the model every sampling day for prediction (Figure 5.5E), nRMSE and nMAE did increase when compared to the previous results (Figure 5.4E) in which all sampling data was available to aid in feature predictions. It suggests that the measured quantities regarding lactate accumulation, glucose consumption, ammonia accumulation, and protein production aided in next day prediction of cell culture growth dynamics.

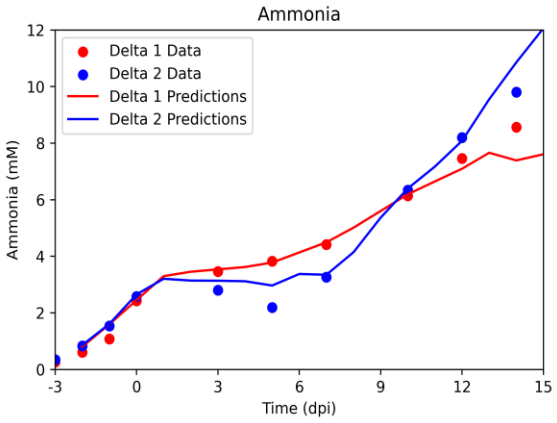
A


Titer
Beta Data
WuTL Data
Beta Predictions
WuTL Predictions
Titer Concentration (mg/L)
Time (dpi)

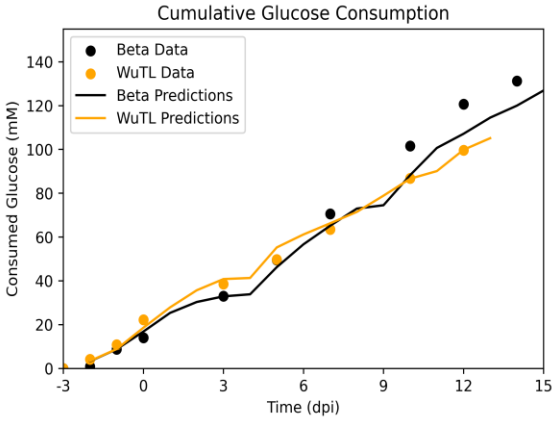

Titer
Delta 1 Data
Delta 2 Data
Delta 1 Predictions
Delta 2 Predictions
Titer Concentration (mg/L)
Time (dpi)

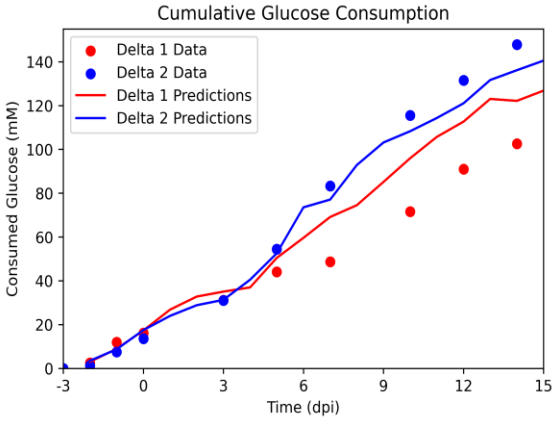
B


Ammonia
Beta Data
WuTL Data
Beta Predictions
WuTL Predictions
Ammonia (mM)
Time (dpi)


Ammonia
Delta 1 Data
Delta 2 Data
Delta 1 Predictions
Delta 2 Predictions
Ammonia (mM)
Time (dpi)

C


Cumulative Glucose Consumption
Beta Data
WuTL Data
Beta Predictions
WuTL Predictions
Consumed Glucose (mM)
Time (dpi)


Cumulative Glucose Consumption
Delta 1 Data
Delta 2 Data
Delta 1 Predictions
Delta 2 Predictions
Consumed Glucose (mM)
Time (dpi)

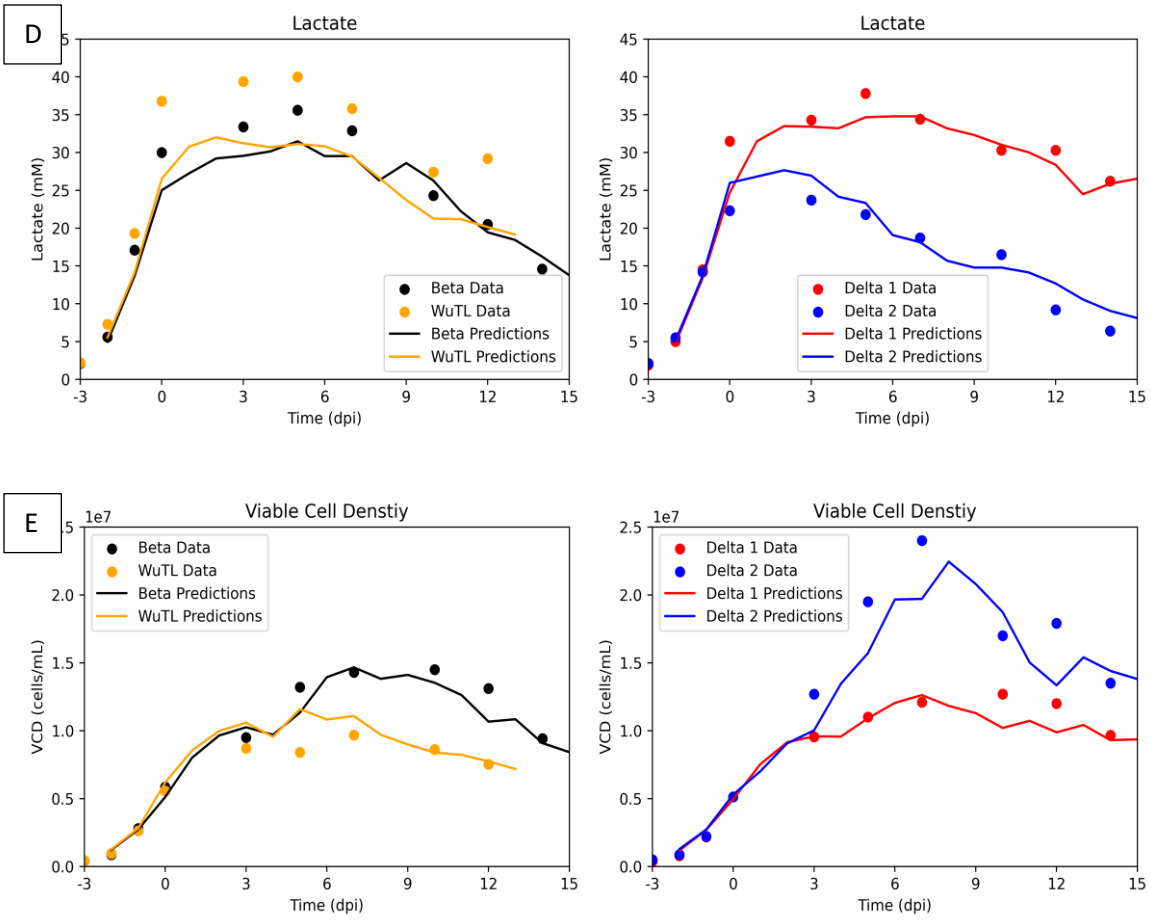


Figure 5.5. Model results for key features in the test dataset comprised of 4 cultures (1 Beta pool batch, 1 WuTL pool batch, and 2 Delta pool batches) without metabolic and titer sampling data.

Continuous lines are everyday predictions while dots are measured values. (a) product titer, (b) ammonia, (c) cumulative glucose consumed, (d) lactate, (e) viable cell density (VCD).

As it can be seen in Table 5.3, only VCD shows a reduction in R^2 indicating slight deviation in linearity from predictions. An increase in nRMSE and nMAE can be observed for lactate, ammonia, and VCD predictions indicating slight increase in error predictions. This suggests that the model architecture with cell count measurements (which are easy to

obtain manually let alone with dedicated cell counting machines) and online data from the bioreactor is robust enough to soft sense hard-to-detect variables (in this case metabolic variables and protein production). This is especially interesting within the context of difficult-to-quantify proteins such as the SARS-CoV-2 spike protein since quantification is generally done through ELISA assays or semi-quantitative SDS-PAGE once the process is finished [201]. Consequently, throughout the process, no knowledge regarding the trajectory of the titer values is known, hampering decision making and slowing down proposals for process improvement since retrospective analysis of titers needs to be realized before conclusions can be drawn. Alternatively, it is worth postulating whether said architectural approach can also aid in the time series evolution prediction of product glycoforms. Glycosylation profiles are key quality attributes that are generally only measured at endpoints due to the difficulty and cost of the process [357]. These glycosylation profiles are known to be impacted by several process conditions (e.g., pH changes, temperature shifts, osmolarity, dissolved oxygen tension) as well as cell line dependence thus having the capacity to evaluate their time series evolution during the production process before endpoint confirmation would be of interest for large scale manufacturing operations [358-360]. Assuming the proposed architecture can give reasonable qualitative predictions of glycosylation profiles, for a large-scale manufacturing process, only an initial capital and time expenditure with regard to obtaining the time series variation of the training data is required.

Table 7. RMSE, MAE, nRMSE, nMAE, R^2 metrics for titer, ammonia, cGC (cumulative glucose consumed), lactate, VCD (viable cell density) for the test dataset without utilizing sampling day metabolic and titer data in model predictions.

	RMSE	**MAE**	**nRMSE**	**nMAE**	$\boldsymbol{R^2}$
Titer	48.299	38.115	0.24	0.19	0.979
Ammonia	0.469	0.327	0.215	0.153	0.984
cGC	7.876	6.042	0.202	0.155	0.995
Lactate	3.883	3.367	0.411	0.355	0.951
VCD	1600083	1186287	0.335	0.243	0.938

Given that estimation of cell growth and cumulative glucose consumption rates were accurate even when no glucose consumption data was utilized in the **test** set, estimation of specific glucose consumption rates (qGluc) was performed. Since the model prediction represents cumulative glucose consumption up to any given day, the derivative of this output gives the glucose consumed every day (since every prediction is equally spaced in the time dimension, the derivative is calculated using the difference between consecutive datapoints such that $Consumed\_Glucose = cGC_{(i+1)} - cGC_{(i)}$). Given the VCD predictions, the integral of viable cell concentration (IVCC) can be estimated through the trapezoid rule of numerical integration. Consequently, differentiating these IVCC values allows for the estimation of ΔIVCC (ΔIVCC= $IVCC_{(i+1)} - IVCC_{(i)}$). If the consumed glucose estimated at every time point is then divided by the ΔIVCC at every point in time ($\Delta IVCC_i$), predicted specific glucose consumption rates ($qGluc_{pred}$) can be calculated.

$$qGluc_{pred} = \frac{Consumed\_Glucose_i}{\Delta IVCC_i}$$

The ground data of glucose consumed between sampling days ($Consumed\_Glucose_k$) can be divided by the $\Delta IVCC_k$ (change in IVCC between sampling days), then specific glucose consumption rates can be estimated. The subscript $_k$ indicates the ground data at sampling day $_k$.

$$qGluc_{Ground\_Truth} = \frac{Consumed\_Glucose_k}{\Delta IVCC_k}$$

In Figure 5.6, it is possible to see an overlay between the estimated specific glucose consumption rates and the predicted specific glucose consumption rates. For the 4 **test** cases, qualitative tracking of specific glucose consumption rates is possible showing a fast decrease from -1 to 0 dpi which represents the end of the growth phase and then near constant qGluc after temperature downshift (0 dpi). This qualitative tracking of specific glucose consumption rates (qGluc) is important as knowledge regarding the specific consumption represents knowledge regarding its internal metabolic state [361,362]. Additionally, having one day ahead predictions of specific glucose consumption rates represents a pseudo-online opportunity for nutrient control which has been tried utilizing various approaches (Raman spectroscopy, optical probes, oxygen monitoring) [148,363-365]. Here, glucose concentrations inside the bioreactor can be adjusted based on the prediction of specific glucose consumption rates such that glucose levels within the bioreactor are maintained within a given set point. Such monitoring and control strategies have shown promise in reducing product glycation [366]. Given the qualitative accuracy of tracking specific glucose consumption rates, a similar approach may be taken to track specific lactate production/consumption, specific ammonia accumulation/consumption, growth, and specific protein production rate. Additionally, since it was observed that indirect

measurements could be learned (qGluc), it can be postulated that by adding total cell density (TCD) prediction into the model, viability (viability=VCD/TCD) can be indirectly monitored throughout the manufacturing process.

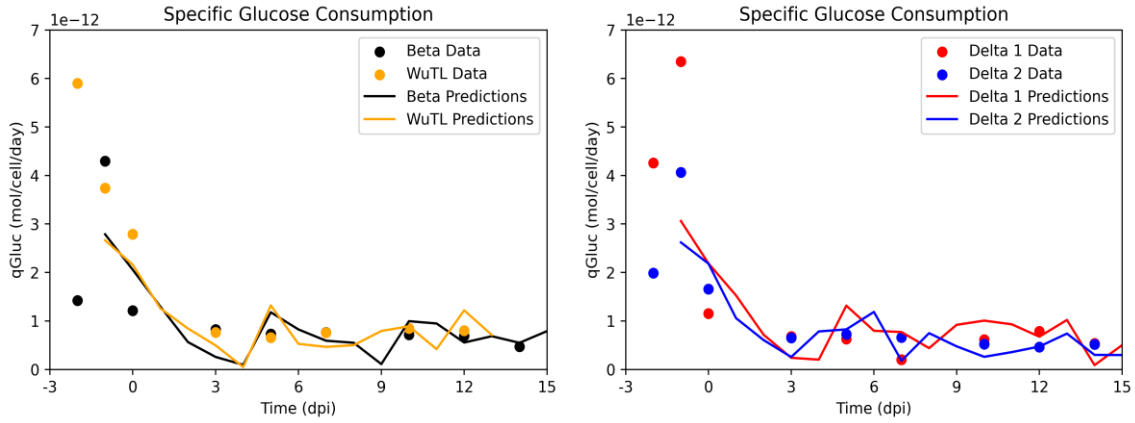


Figure 5.6. Specific glucose consumption rate (qGluc) estimation for test dataset comprised of 4 cultures (1 Beta pool batch, 1 WuTL pool batch, and 2 Delta pool batches). Continuous lines are everyday predictions while dots are measured values.

It is also worth noting that the online data available in this case study is routine online data such as temperature, pH, DO, oxygen injection, and carbon dioxide addition. Thus, it stands to reason that as more biologically relevant online sensor data is added to the processes and consequently the model, better predictive capabilities can be achieved. Such sensor data that could be utilized is oxygen uptake rates (OUR) and bio-capacitance signals [311]. OUR has been known to relate both to cellular growth and metabolic activity as peak antibody production has been associated with increased oxidative metabolism which, in turn, means an increased oxygen demand [40,227,271,367]. Alternatively, bio-capacitance signals have been observed to not only relate to total viable cells within the culture but also to total biovolume as the capacitance signal is dependent on the volume of

each cell [236,242,368]. This is key as cellular diameter has been observed to increase during a culture run [369] and consequently changes in cellular volume are expected. Cell volume also encodes information regarding cellular phase and recombinant protein production activity [45,46]. Given the close relationship between bio-capacitance and cellular volume, further improvements to the model's predictive capabilities could be made. Notably, sensor fusion (biocapacities, OUR, and fluorescence data) coupled with machine learning (ML) techniques have demonstrated strong monitoring capacity of relevant features such as cellular viability, cellular density, metabolites, and viral yield [370] further underscoring the potential for improved process forecasting with the proposed model when coupled with process analytical technologies (PAT).

CONCLUSION

The proposed soft sensor architecture can accurately (nRMSE, nMAE are below unity < 1 and $R^2 \geq 0.9$ for all features) predict product titer, total glucose consumption, ammonia, lactate, and viable cell densities, and for a long-term process (17 days) with ground sampling day data only available every other day or every two days to aid in next iteration predictions. To counter the lack of everyday sampling data, daily online data was also utilized into the model. The online data, although harder to directly interpret, still contains relevant information about culture phase (temperature downshifts denote process induced decrease in cellular growth to prime the process for protein production). In terms of PID controllers, it indirectly contains information about oxygen demand (DO, total oxygen sparge) and lactate metabolism (pH, base addition, total carbon dioxide addition). Interestingly, once the same model was applied in a test case where no ground sampling day data was given for titers, glucose consumption, ammonia, and lactate throughout the 17-day culture process, it was determined that the model was still effectively able to soft sense these hard-to-measure features (nRMSE and nMAE below unity < 1 and $R^2 \geq 0.9$ for all features). This is especially interesting in processes where the recombinant protein in question can be difficult to measure as is the case of the trimeric SARS-CoV-2 spike protein. In such cases, having a qualitative tracking of feature evolution can be of value. This was possible by considering the bioreactor data that is routinely available in all commercial bioreactor systems. Additionally, qualitative tracking of specific glucose consumption rates (qGluc) was enabled with the proposed method allowing for the possibilities of tight glucose control inside bioreactors by relying on the one day ahead

specific glucose consumption rates predictions and adjusting glucose addition to keep overall glucose concentration near a given setpoint. With this knowledge, it stands to reason that the proposed soft sensor can gain from further use of process analytical technologies (PAT) such as off-gas analyzers, bio-capacitance, and Raman spectroscopy signals from which biologically relevant signals can be related to the discretely measured features. Model robustness may also be tested by evaluating the proposed architecture in varying manufacturing platforms (HEK293 viral production, therapeutic antibody production, etc.).

AUTHOR CONTRIBUTIONS

Sebastian Juan Reyes: Data curation; data analysis; methodology; writing-original draft preparation. **Olivier Henry:** Supervision; writing–review and editing. **Phuong Lan Pham**: Experimental conceptualization; Supervision; writing–review and editing. **Robert Voyer, Yves Durocher**: Supervision; writing–review.

ACKNOWLEDGEMENTS

The work was financially supported by the Pandemic Response Challenge Program (PRCP) of the National Research Council Canada. The authors gratefully acknowledge Helene L'Ecuyer-Coelho, Yuliya Martynova, and Marjolaine Roy for their contribution in conducting bioreactor production runs. The expertise of Raul-Santiago Molina (Proelium S.A.S, Carrera 16 C #153, Bogotá, 110131, Bogotá DC, Colombia) in data analysis and modeling was greatly recognized.

CHAPTER 6 GENERAL DISCUSSION AND RECOMENDATION

General Discussion

CHO cells are the current industry standard to produce recombinant therapeutics and antibodies. This is because CHO cells offer a valuable platform for the production of glycoprotein products, as they confer human-like glycosylation features. Their capability for high protein productivity, achieving titers up to grams per liter, is attributed to well-established strategies for generating stable cell lines. Furthermore, these cells propagate efficiently in serum-free single-cell suspension cultures, facilitating easy scale-up. Importantly, CHO cells exhibit resistance to several human viruses, mitigating biosafety risks associated with their use. Additionally, they boast a proven regulatory track record, solidifying their standing in the manufacture of recombinant protein therapeutics [371-374]. Such recombinant protein expression is possible through either transient or stable gene expression strategies. It must be mentioned that developing stable cell lines is a time consuming and often expensive process that necessitates numerous rounds of screening procedures so as to select the best behaving cellular populations [51]. Within the context of rapidly changing public health crisis, like the one caused by the onset of the SARS-CoV-2 pandemic, there exists a benefit towards relying on fast tracked processes that still meet quality standards [10,51,52,200]. Consequently, mammalian cell pools have been contemplated as an alternative to cell lines. This approach potentially reduces overall timelines by 3-6 months [51,52]. Even though cell pools are not as homogeneous in cellular population as cell lines, they can still be employed to produce biotherapeutics on a large scale [51,52]. Clones chosen for production during the high-throughput screening phase may not necessarily

perform effectively at larger scales due to variations in culture environments and conditions. Consequently, there is a rationale for minimizing excessive resource usage at the microliter scale and redirecting focus toward representative scalable models. In this context, cell pools can be employed to fast-track the transition to the benchtop scale to identify suitable process parameters that reflect large-scale production. However, these pools are often strongly influenced by cell age, leading to a potential decrease in target protein expression productivity with increased passage numbers [52]. Thus, for stable pools to mitigate their drawbacks and emerge as a viable alternative to clonally derived cell lines, the production phase should commence at the earliest opportunity, and the process itself must undergo thorough optimization. Crucially, not many large datasets concerning mammalian cell pools in the process development stage have been published and much less in-depth analysis of said datasets utilizing exploratory analysis tools and machine learning. This is a great opportunity as with the accumulation of early process development data, a virtuous cycle can take hold such that increased understanding of the global trends within the data can help researchers target process improvements which then generates more data that can then be analyzed further. Additionally, process development datasets can be utilized as a pipeline for soft sensor development such that knowledge gained in model development with the varied datasets will greatly improve predictive capacity at the manufacturing stage, as process conditions are locked into a specific range within the explored design space. This is of note as soft sensors will continue to play an important role in Biopharma 4.0 for both online monitoring of cultures as well as process control. In this thesis, multiway PCA of CHO pools expressing the SARS-CoV-2 spike protein was realized and found that within pools there exists a direct relationship between oxygen

requirements and protein titer while it was also determined that there exists an inverse relationship between severe lactate accumulation and titer. Additionally, oxygen requirements and lactate accumulation/base addition were also observed to have an inverse relationship which makes sense if one interprets these results from the standpoint of glycolytic metabolism vs oxidative metabolism [32,33]. It was also evident from the exploratory data analysis that each pool had its own unique behaviour in terms of having different cell growth and protein production profiles, that may suggest tailoring of feed regimens to take into account the difference in biomass accumulations is needed so as to optimize yield (pools that on average demonstrated the most growth had also the lowest protein production rates on average). Passage dependency on protein production was only detected in 1 (Delta) of the 4 (Wuhan, Delta, Beta, Wutan-tagless) examined pools. Furthermore, Random Forest was applied, to the early process development dataset, so as to predict endpoint protein yield as a function of key summary variables. After carrying out bootstrapping to gauge the performance of the Random Forest model with the test dataset, low RMSE (65.96 mg/L) was reached as well as a high coefficient of determination (0.94), indicating strong linearity between test data and predictions. Importantly, by utilizing shapely dependency plots in the analysis of the machine learning model, further process insights were made. Namely, the negative impact that decreased endpoint longevity (below 80%) had on titer yields, the negative impact maximum lactate (>35 mM) accumulation had on titer. The positive relationship between increasing oxygen requirements and increased protein yield (up to 0.03 VVM). Endpoint pH above 6.95 was also associated with decreased yield, cultures with endpoint pH of 6.93-6.95 represent processes in which no base addition or carbon dioxide was added in the last day of the fed-

batch process. This could imply that unnecessary action upon cultures may be a net negative and thus pH deadband can be increased to ±0.2 around the 7.0 setpoint so as to avoid base addition and carbon dioxide sparging. Noteworthy, is the observation that cumulative glucose consumption below 80 mM were negatively related to protein outcomes, while cumulative glucose consumption above 120 mM were also associated with decreased protein yields. This suggests that controlling the glycolytic behaviour (not too low nor too high) of the pools is an important avenue for process improvement. Lastly, high values of endpoint residual glucose (> 19 mM) were also associated with decreasing protein production. This can be interpreted in tandem with endpoint viability as cultures that reach the end of the process with high viability remain metabolically active while cultures that reach with low viability towards the end of the process have low metabolic activity.

When transferring the Wuhan pool process off-site (from the National Resource Council to Polytechnique Montreal) to a DASGIP bioreactor several of the process alternatives were tested keeping in mind the idea that improving endpoint longevity, controlling maximal lactate build-up, tailoring feeding regimens to changing biomass values, and expanding pH deadbands were deemed important process variables with respect to maximizing protein production. First, different DO levels where tested and it was found that 60% DO both did not hamper cell growth and improved longevity while 90% DO improved longevity at the cost of overall cell growth. Presumably, because of the existence of increased ROS species (at 90% DO) which are known to limit CHO cell growth [268-270]. The 60% DO condition also represented a 1.62-fold increase in yield when compared to the 40% DO condition, which can be explained by the fact that the 40% DO condition had

low viability by 12 dpi while the 60% DO condition only reached similar values by 14 dpi thus allowing for two more days of protein production. Since the early process development dataset was developed using micro and macro spargers while the DASGIP system was operated with open pipe sparger and given the difference in oxygen transfer rate between spargers, a study dedicated to finding an adequate air cap was undertaken [375,376]. Additionally, feeding was also altered so as to increase the amount of nutrients during the production stage of the cultures. Crucially due to the increased feeding (by total volume), slow pump addition (speeds lower than 1 mL/h) was utilized as suggestions in the literature have shown that in CHO K1 cells, such dosing strategy can aid in reducing metabolic waste build up [67]. It was found that cultures with low air cap (1 L/h), 60% DO and increased feeding outperformed the standard 40% DO, bolus feeding process by 2-fold in protein production. Given the slow continuous dosage of the increased feed regimen it was not clear whether the improvements were caused by the slow continuous addition of feed, or by the increased feed, or by a combination of both factors. Consequently, a study comparing high vs low feeding conditions at bolus vs slow continuous dosing regimens was conducted. It was observed that when slow continuous dosing is applied regardless of feed amount, decreased lactate profiles are detected thus improving overall longevity as observed in the literature [67]. Alternatively, increased feeding had a 2-fold increase in yield (when controlling for slow continues dosing) suggesting that increased feeding during the production phase can aid in titer production optimization.

Further experiments involving improved feeding regimens relied on coupling feed flow rates to OUR and bio-capacitance signals on the rationale that these signals are maximal when protein production is maximal (stationary phase) and thus increased feeding at these

points could aid in protein production. It was found that OUR based feeding increased titer 1.47-fold when compared to a control process while biocapacitance feeding increased yield 1.85-fold when compared to a control process. Crucially, when detailing the resulting values of OUR/biocapacitance, which should give information regarding the specific oxygen consumption within the biovolume inside the reactor, it was clear that the cultures with higher OUR/ biocapacitance ratios during the protein production phase had better overall yields. Suggesting that the utilization of both signals can be a way to qualitatively soft sense protein production performance during a reactor run given the relationship between specific respiration rates and specific protein production rates [40-43]. Additionally, when plotting cumulative OUR against cumulative glucose consumption metabolic shifts can also be monitored such that a change from highly glycolytic metabolism to predominantly oxidative metabolism can be detailed. Given the positive correlation that has been observed between oxidative metabolism and protein production, it is a parameter to consider when analyzing OUR data [32,33]. Importantly, setting feed flow rated to follow OUR or biocapacitance signals is not only advantageous for developing dynamic feeding regimens but such strategies are also easily automatable such that automated feeding protocols can be created that do not require operator input.

It can be affirmed that many of the strategies suggested by interpreting the Random Forest model (controlling lactate-buildup, improving longevity, improving feed regimens on bases of biomass requirements) ended up being adequate starting points for process improvement, suggesting that the strategy applied can serve as a framework for future projects involving technology transfer. One noteworthy interpretation that did not yield significant results was expanding the pH deadband. As nearly identical IVCC and titer

values were obtained in conditions with a strict deadband (6.95±0.05) and wide deadband (7±0.2). This may be explained by the fact that in the DASGIP system an open pipe sparger was employed as opposed to micro or macro spargers in the NRC dataset. This may impact dissolved carbon dioxide profiles within the bioreactors and consequently alter pH control loops given the difference in carbon dioxide striping rates [375]. As previously mentioned, early process development datasets can serve as a pipeline for soft sensor development and it was further demonstrated that standard bioreactor online signals (pH, base addition, carbon dioxide sparging, oxygen sparging) indirectly hold information about culture kinetics. Indeed, both the Random Forest model and the Recursive Neural Network model utilized bioreactor signals to aid in protein prediction and both reached similar RMSE in test datasets (51.501 mg/L for RNN and 65.96 mg/L for RF). Importantly, the proposed RNN architecture can not only predict protein production across culture time, but it is also capable of predicting VCD, cumulative glucose consumption, ammonia accumulation and lactate accumulation, concomitantly. Thus, effectively turning standard reactor measurements into a 'smart' sensor capable of tracking multiple variables at the same time. This is especially useful for hard-to-detect variables such as SARS-CoV-2 spike proteins that rely on endpoint analytics (semi quantitative gels or ELISA assay) to quantify yields. This may suggest that the proposed model architecture may be applied in the temporal evolution tracking of glycosylation profiles which are also impacted by processing conditions [358,360,377]. It is noteworthy, that if standard bioreactor signals allowed for adequate prediction of key variables, then data rich processes which contain metabolically relevant online signals such as OUR, biocapacitance, dissolved carbon dioxide may further improve upon prediction metrics.

Recommendations

Several improvements that can be realized upon this work. One of such, lies in the automation of multiway PCA. Instead of relying on visual inspection of loading plots, sum, mean or median of the loadings of each variable can be applied to determine the predominant direction (positive or negative) as well as ranking variable loadings by magnitude. Indeed, it is possible to readily imagine an online tool which automatically outputs the PCA scatter plots with the top 4 ranked variables by magnitude for each principal component. Such tool would only require the researcher to upload his dataset already unfolded batchwise and a vector encoding conditions from which coloring each score by its respective condition is possible. For the machine learning models, considering sparger type may be an important variable as micro and macro spargers have different oxygen transfer rates and thus may impact on perceived oxygen requirements [375,376]. For this thesis when plotting cumulative oxygen sparge rates for all the 50 cultures no outliers could be determined or indeed directly related to difference in sparger. In fact, when sparger type was added as a feature to the model prediction outcomes worsened possibly due to the fact that in the total dataset, only 7 cultures utilized an aeration process that was not just a micro sparger. Additionally, these 7 cultures were realized in a different bioreactor (BioFLo rather than Multifors) system with different foam away addition protocols possibly further impacting perceived oxygen requirements. Nonetheless, for future applications taking into account sparger type so as to not bias perceived oxygen requirements may be important.

The addition of information regarding amino acid consumption and cell volume may also aid in improving model performance as these variables encode metabolically relevant

information [26,32,45,46]. It was experimentally observed that increased air caps lessen longevity and increased lactate buildup. Similarly, increasing RPM was also detected to decrease longevity. As it is known that different hydrodynamic stresses differentially impact cell cultures [77]. It would be worth exploring if the stress mechanisms in CHO pools when subjected to various hydrodynamic stress are any different than standard CHO cell lines stress responses. As detailed, increased feeding during the protein production phase served to improve overall protein production yields. However, this came at the cost of ammonia increases (> 8 mM), that the model suggested was adverse. Importantly, the model was trained with a single feed regimen so the increased ammonia accumulation that was observed to be adverse relates to observations with cultures that utilized that specific feeding strategy and can not be generalized for all feeding strategies. However, it remains to be seen if the observed ammonia increases with the proposed feeding strategies hamper different cell pools or cell lines than the ones tested experimentally. Considerations may be taken if it is found to be the case by feeding based on indirect or direct amino acid consumption monitoring [40,378-380]. Additionally, further yield improvements may be attained by exploring time of action events [144,381,382]. Given that inducible pools have several landmark events (induction point, temperature shift, feeding initiation), studying how to time these events to best optimize total protein production may be a promising strategy. For this purpose, OUR can be applied as it is a relevant on-line indicator of shifting metabolism and overall biomass inside a bioreactor [145,146,382]. Indeed, the OUR/biocapacitance signal may also be used as changes in specific respiration rates delineate metabolic transitions within the cell culture [147,383,384]. Furthermore, given the observation that specific respiration rates and specific protein production rates are linearly

correlated [40-42] and given the fact that process conditions can impact respiration rates [40,185]. Studying if further rational optimization of protein yields can be attained for these inducible pools by leveraging the known relationship between specific respiration rates and specific protein production rates is a promising strategy. Such experiments can center around improving process conditions such that specific respiration rates are increased consequently leading to an increase in specific protein production rates.

Difficulties were observed when predicting viability with the RNN model as oscillations in the output signal could be discerned. Most likely because culture kinetics for titer, ammonia accumulation, cell growth, glucose consumption, lactate accumulation are markedly different than viability, which is near 100% for most of the process and then declines rapidly. Consequently, this variable was excluded from the model. Importantly, an alternative way that viability could be incorporated back into the model is by predicting total cell density (TCD). Since VCD was adequately predicted it stands to reason that similar performance can be attained when predicting TCD. Consequently, by calculating the VCD/TCD ratio viability may be indirectly predicted. Furthermore, with knowledge regarding viability the soft sensor may be applied to automate end of culture events (prediction is below 60% viability threshold). Lastly, rather than relying exclusively on a data driven model (RNN) for soft sensor tracking of culture kinetics, a hybrid approach can be applied. For this purpose, recursive state observers can be deployed to combine dynamic metabolic modelling (white box models) and data-driven modelling (black box models). This is realized by updating state estimates derived from noisy measurements and gradually reducing the estimation error covariance. The unscented Kalman filter can also be used for this precise application. Here instead of a linearization of the system of equations (like in

an extended Kalman filter), a minimum number of points required to describe the probability distribution of the system is employed. These so-called sigma points are passed through the system of differential equations as to determine how the probability distribution changes at every time step. Thanks to the sigma points, the mean and variance of the system at every time step can be known (thanks to an unscented transform) and used to compare a given sensor input where its mean value and variance are also known. Thus, the mean and variance of both the system of equations and the sensor measurement can be used together to construct a dynamic estimator that can predict the state of the system. Such types of dynamic estimators are readily applied in microbial fermentations but some applications in cell culture have been published [166,385-387]. To successfully implement a Kalman Filter in CHO pools, two models are needed. One model is used to move the elements of the state vector within the state space from one time step $tk-1$ to the next time step [163]. The second model is the observation model that relates the state vector at time tk to the actual sensor measurement quantities at that time point [163]. For the state propagation model, an ordinary differential equations system describing the propagation of the initial state with time can be used. For the observer model, OUR data and biomass must be coupled in order to use the OUR signals as input for biomass estimation [388]. This observer model can either be a theoretical relationship, like the Luedeking-Piret which relates biomass to oxygen consumption through a yield and maintenance coefficient relationship, or a statistical relationship that relates biomass to oxygen consumption through non-linear regression models [163,389]. Thus, the hybrid model may be more robust in process tracking, especially when applying the same architecture to a different reactor system.

CHAPTER 7 CONCLUSION

This thesis provides a framework in which exploratory data analysis and machine learning modeling of protein titers can be realized on early process development datasets so as to gain information that serves the tasks of process optimization. Here, resulting model interpretation utilizing shapely dependency plots and variable correlations can serve as rational starting points when undertaking process development projects. In this thesis, such starting points for increasing SARS-CoV-2 yield were extending culture longevity, decreasing maximal lactate buildup, expanding pH deadband and setting feed regiments to track differences in biomass. Importantly, it was experimentally shown that extending longevity by adequately tunning air caps and finding appropriate DO set points in tandem improved culture outcomes. Additionally, changing from bolus feed to slow continuous feed addition improved upon maximal lactate reduction which also improved culture outcomes. Lastly, setting feed flow rates to track OUR and biocapacitance signal evolution so as to indirectly track biomass values also showed improvement when compared to a control process. As datasets grow a virtuous cycle of knowledge improvement can be created such that new experiments can be fed back into the original dataset and modeling/interpretation can be updated, such is the promise of this framework. Importantly, not only can early process development datasets be used for exploratory data analysis and model interpretation, but they can also serve as a starting point for soft sensor development. Here, leveraging the wide variety of data (consequence of exploring the design space), robust model architectures can be created which are expected to improve in predictive capability as the transition to manufacturing scale occurs, given that at the manufacturing phase process conditions are locked into a specific range within the explored design space. With this in mind, a novel RNN model was constructed that is

capable of tracking multiple hard-to-detect variables with high linearity (>0.9 coefficient of determination) and low nRMSE (below unity) across the entire fed-batch process by relying exclusively on standard online bioreactor signals. Thus, this work provides an architecture from which ‘smart’ process monitoring is possible for even the most standard of bioreactors. In summary, this work shows the great value that there is behind adequately storing and retroactively analysing early process development datasets so as to better prepare future process improvement projects that rely on both process intensification and/or process monitoring.

REFERENCES

1. precedenceresearch. Biotechnology Market Size | Share and Trends 2024 to 2034. Accessed Septemeber 29 2024, https://www.precedenceresearch.com/biotechnology-market#:~:text=The%20global%20biotechnology%20market%20size%20was%20valued%20at%20USD%201.38,USD%204.25%20trillion%20by%202033
.
2. Research GV. Biotechnology Market Growth Analysis Report, 2021-2028. https://www.grandviewresearch.com/industry-analysis/biotechnology-market
3. Intelligence M. BIOPHARMACEUTICALS MARKET - GROWTH, TRENDS, COVID-19 IMPACT, AND FORECASTS (2021 - 2026). https://www.mordorintelligence.com/industry-reports/global-biopharmaceuticals-market-industry
4. Reyes SJ, Durocher Y, Pham PL, Henry O. Modern Sensor Tools and Techniques for Monitoring, Controlling, and Improving Cell Culture Processes. *Processes*. 2022;10(2)doi:10.3390/pr10020189
5. mordorintelligence. Biopharmaceutical Industry Size & Share Analysis - Growth Trends & Forecasts (2024 - 2029). Accessed septiember 29 2024, https://www.mordorintelligence.com/industry-reports/global-biopharmaceuticals-market-industry
6. Markets Ra. Biopharmaceuticals Market By Type and Application

https://www.researchandmarkets.com/reports/4612776/biopharmaceuticals-market-by-type-and-application
7. Research GV. Biosimilars Market Size Worth $61.47 Billion By 2025. 2018, July;
8. Li F, Vijayasankaran N, Shen AY, Kiss R, Amanullah A. Cell culture processes for monoclonal antibody production. *MAbs*. Sep-Oct 2010;2(5):466-479. doi:10.4161/mabs.2.5.12720
9. Lu RM, Hwang YC, Liu IJ, et al. Development of therapeutic antibodies for the treatment of diseases. *Journal of biomedical science*. Jan 2 2020;27(1):1. doi:10.1186/s12929-019-0592-z
10. Joubert S, Stuible M, Lord-Dufour S, et al. A CHO stable pool production platform for rapid clinical development of trimeric SARS-CoV-2 spike subunit vaccine antigens. 2023;120(7):1746-1761. doi:https://doi.org/10.1002/bit.28387

11. Aggarwal K, Jing F, Maranga L, Liu J. Bioprocess optimization for cell culture based influenza vaccine production. *Vaccine*. 2011/04/12/ 2011;29(17):3320-3328. doi:https://doi.org/10.1016/j.vaccine.2011.01.081
12. Mahé A, Martiné A, Fagète S, Girod PA. Exploring the limits of conventional small-scale CHO fed-batch for accelerated on demand monoclonal antibody production. *Bioprocess and biosystems engineering*. Feb 2022;45(2):297-307. doi:10.1007/s00449-021-02657-w
13. Kunert R, Reinhart D. Advances in recombinant antibody manufacturing. *Applied microbiology and biotechnology*. Apr 2016;100(8):3451-61. doi:10.1007/s00253-016-7388-9
14. Lustri J. Process Analytical Technology-It's Not Rocket Science, but It Is Science, Math, Control, and IT. In Tech Magazine2015.
15. Uwe Kirschner REC, Rebecca Vangenechten, Kjell François. Process Analytical Technology: An industry perspective. https://www.europeanpharmaceuticalreview.com/article/3643/process-analytical-technology-pharma-industry/
16. Mhatre R, Rathore AS. Quality by Design: An Overview of the Basic Concepts. *Quality by Design for Biopharmaceuticals*. 2009:1-8.
17. Rathore AS, Winkle H. Quality by design for biopharmaceuticals. *Nat Biotechnol*. 2009// 2009;27doi:10.1038/nbt0109-26
18. Hartley F, Walker T, Chung V, Morten K. Mechanisms driving the lactate switch in Chinese hamster ovary cells. *Biotechnology and bioengineering*. Aug 2018;115(8):1890-1903. doi:10.1002/bit.26603
19. Pereira S, Kildegaard HF, Andersen MR. Impact of CHO Metabolism on Cell Growth and Protein Production: An Overview of Toxic and Inhibiting Metabolites and Nutrients. 2018;13(3):1700499. doi:https://doi.org/10.1002/biot.201700499
20. Galleguillos SN, Ruckerbauer D, Gerstl MP, Borth N, Hanscho M, Zanghellini J. What can mathematical modelling say about CHO metabolism and protein glycosylation? *Computational and Structural Biotechnology Journal*. 2017/01/01/ 2017;15:212-221. doi:https://doi.org/10.1016/j.csbj.2017.01.005
21. Locasale JW, Cantley LC. Metabolic flux and the regulation of mammalian cell growth. *Cell metabolism*. Oct 5 2011;14(4):443-51. doi:10.1016/j.cmet.2011.07.014
22. Mulukutla BC, Khan S, Lange A, Hu W-S. Glucose metabolism in mammalian cell culture: new insights for tweaking vintage pathways. *Trends*

in Biotechnology. 2010/09/01/ 2010;28(9):476-484. doi:https://doi.org/10.1016/j.tibtech.2010.06.005

23. Mulukutla BC, Gramer M, Hu W-S. On metabolic shift to lactate consumption in fed-batch culture of mammalian cells. *Metabolic Engineering*. 2012/03/01/ 2012;14(2):138-149. doi:https://doi.org/10.1016/j.ymben.2011.12.006

24. Ozturk SS, Riley MR, Palsson BO. Effects of ammonia and lactate on hybridoma growth, metabolism, and antibody production. *Biotechnology and bioengineering*. Feb 20 1992;39(4):418-31. doi:10.1002/bit.260390408

25. Zagari F, Jordan M, Stettler M, Broly H, Wurm FM. Lactate metabolism shift in CHO cell culture: the role of mitochondrial oxidative activity. *New Biotechnology*. 2013/01/25/ 2013;30(2):238-245. doi:https://doi.org/10.1016/j.nbt.2012.05.021

26. Fan Y, Jimenez Del Val I, Müller C, et al. Amino acid and glucose metabolism in fed-batch CHO cell culture affects antibody production and glycosylation. *Biotechnology and bioengineering*. Mar 2015;112(3):521-35. doi:10.1002/bit.25450

27. Schneider M, Marison IW, von Stockar U. The importance of ammonia in mammalian cell culture. *Journal of biotechnology*. May 15 1996;46(3):161-85. doi:10.1016/0168-1656(95)00196-4

28. Martinelle K, Häggström L. Effects of NH4+ and K+ on the energy metabolism in Sp2/0-Ag14 myeloma cells. *Cytotechnology*. Jan 1999;29(1):45-53. doi:10.1023/a:1008084622991

29. Duarte TM, Carinhas N, Barreiro LC, Carrondo MJ, Alves PM, Teixeira AP. Metabolic responses of CHO cells to limitation of key amino acids. *Biotechnology and bioengineering*. Oct 2014;111(10):2095-106. doi:10.1002/bit.25266

30. Mohmad-Saberi SE, Hashim YZ, Mel M, Amid A, Ahmad-Raus R, Packeer-Mohamed V. Metabolomics profiling of extracellular metabolites in CHO-K1 cells cultured in different types of growth media. *Cytotechnology*. Aug 2013;65(4):577-86. doi:10.1007/s10616-012-9508-4

31. Mulukutla BC, Kale J, Kalomeris T, Jacobs M, Hiller GW. Identification and control of novel growth inhibitors in fed-batch cultures of Chinese hamster ovary cells. *Biotechnology and bioengineering*. Aug 2017;114(8):1779-1790. doi:10.1002/bit.26313

32. Coulet M, Kepp O, Kroemer G, Basmaciogullari S. Metabolic Profiling of CHO Cells during the Production of Biotherapeutics. *Cells*. Jun 15 2022;11(12)doi:10.3390/cells11121929
33. Templeton N, Dean J, Reddy P, Young JD. Peak antibody production is associated with increased oxidative metabolism in an industrially relevant fed-batch CHO cell culture. 2013;110(7):2013-2024. doi:https://doi.org/10.1002/bit.24858
34. Mulukutla BC, Mitchell J, Geoffroy P, et al. Metabolic engineering of Chinese hamster ovary cells towards reduced biosynthesis and accumulation of novel growth inhibitors in fed-batch cultures. *Metab Eng*. Jul 2019;54:54-68. doi:10.1016/j.ymben.2019.03.001
35. O'Mara P, Farrell A, Bones J, Twomey K. Staying alive! Sensors used for monitoring cell health in bioreactors. *Talanta*. Jan 1 2018;176:130-139. doi:10.1016/j.talanta.2017.07.088
36. Hoshan L, Jiang R, Moroney J, et al. Effective bioreactor pH control using only sparging gases. *Biotechnology progress*. 2019;35(1):e2743. doi:10.1002/btpr.2743
37. Xing Z, Lewis AM, Borys MC, Li ZJ. A carbon dioxide stripping model for mammalian cell culture in manufacturing scale bioreactors. 2017;114(6):1184-1194. doi:https://doi.org/10.1002/bit.26232
38. Meghrous J, Khramtsov N, Buckland BC, Cox MMJ, Palomares LA, Srivastava IK. Dissolved carbon dioxide determines the productivity of a recombinant hemagglutinin component of an influenza vaccine produced by insect cells. 2015;112(11):2267-2275. doi:https://doi.org/10.1002/bit.25634
39. Huang YM, Hu W, Rustandi E, Chang K, Yusuf‐Makagiansar H, Ryll T. Maximizing productivity of CHO cell‐based fed‐batch culture using chemically defined media conditions and typical manufacturing equipment. *Biotechnology progress*. 2010;26(5):1400-1410.
40. Zalai D, Hevér H, Lovász K, et al. A control strategy to investigate the relationship between specific productivity and high-mannose glycoforms in CHO cells. *Applied microbiology and biotechnology*. Aug 2016;100(16):7011-24. doi:10.1007/s00253-016-7380-4
41. Lin J, Takagi M, Qu Y, Yoshida T. Possible strategy for on-line monitoring and control of hybridoma cell culture. *Biochemical Engineering Journal*. 2002/09/01/ 2002;11(2):205-209. doi:https://doi.org/10.1016/S1369-703X(02)00028-1

42. Dorresteijn RC, Numan KH, de Gooijer CD, Tramper J, Beuvery EC. On-line estimation of the biomass activity during animal-cell cultivations. *Biotechnology and bioengineering*. Jul 20 1996;51(2):206-14. doi:10.1002/(sici)1097-0290(19960720)51:2<206::Aid-bit10>3.0.Co;2-k
43. Lin J, Takagi M, Qu Y, Gao P, Yoshida T. Metabolic flux change in hybridoma cells under high osmotic pressure. *Journal of Bioscience and Bioengineering*. 1999/01/01/ 1999;87(2):255-257. doi:https://doi.org/10.1016/S1389-1723(99)89025-2
44. Wagner BA, V S, Buettner GR, Medicine. The rate of oxygen utilization by cells. *Free Radical Biology*. 2011;51(3):700-712.
45. Pan X, Dalm C, Wijffels RH, Martens DE. Metabolic characterization of a CHO cell size increase phase in fed-batch cultures. *Applied microbiology and biotechnology*. Nov 2017;101(22):8101-8113. doi:10.1007/s00253-017-8531-y
46. Lloyd DR, Holmes P, Jackson LP, Emery AN, Al-Rubeai M. Relationship between cell size, cell cycle and specific recombinant protein productivity. *Cytotechnology*. Oct 2000;34(1-2):59-70. doi:10.1023/a:1008103730027
47. Edros R, McDonnell S, Al-Rubeai M. The relationship between mTOR signalling pathway and recombinant antibody productivity in CHO cell lines. *BMC biotechnology*. Feb 17 2014;14:15. doi:10.1186/1472-6750-14-15
48. Frame KK, Hu WS. Cell volume measurement as an estimation of mammalian cell biomass. *Biotechnology and bioengineering*. Jun 20 1990;36(2):191-7. doi:10.1002/bit.260360211
49. Nielsen LK, Reid S, Greenfield PF. Cell cycle model to describe animal cell size variation and lag between cell number and biomass dynamics. *Biotechnology and bioengineering*. Nov 20 1997;56(4):372-9. doi:10.1002/(sici)1097-0290(19971120)56:4<372::Aid-bit3>3.0.Co;2-l
50. Schoenherr I, Stapp T, Ryll T. A comparison of different methods to determine the end of exponential growth in CHO cell cultures for optimization of scale-up. *Biotechnology progress*. Sep-Oct 2000;16(5):815-21. doi:10.1021/bp000074e
51. Hacker DL, Balasubramanian S. Recombinant protein production from stable mammalian cell lines and pools. *Current Opinion in Structural Biology*. 2016;38:129-136. doi:https://doi.org/10.1016/j.sbi.2016.06.005
52. Ye J, Alvin K, Latif H, et al. Rapid protein production using CHO stable transfection pools. *Biotechnology progress*. 2010;26(5):1431-7. doi:10.1002/btpr.469

53. Fan L, Frye CC, Racher AJJPB. The use of glutamine synthetase as a selection marker: recent advances in Chinese hamster ovary cell line generation processes. 2013;1(5):487-502.
54. Noh SM, Shin S, Lee GMJSR. Comprehensive characterization of glutamine synthetase-mediated selection for the establishment of recombinant CHO cells producing monoclonal antibodies. 2018;8(1):5361.
55. Lin P-C, Chan KF, Kiess IA, et al. Attenuated glutamine synthetase as a selection marker in CHO cells to efficiently isolate highly productive stable cells for the production of antibodies and other biologics. Taylor & Francis; 2019:965-976.
56. Mellahi K, Brochu D, Gilbert M, et al. Assessment of fed-batch cultivation strategies for an inducible CHO cell line. *Journal of biotechnology*. Jun 10 2019;298:45-56. doi:10.1016/j.jbiotec.2019.04.005
57. Mullick A, Xu Y, Warren R, et al. The cumate gene-switch: a system for regulated expression in mammalian cells. *BMC biotechnology*. Nov 3 2006;6:43. doi:10.1186/1472-6750-6-43
58. John D. Orr. George L. Reid I. An Introduction To Process Analytical Technology. https://www.pharmaceuticalonline.com/doc/an-introduction-to-process-analytical-technology-0001
59. Doran PM. *Bioprocess engineering principles*. Academic Press; 1995.
60. Shuler ML, Kargı F, DeLisa M. *Bioprocess engineering : basic concepts*. Third edition. ed. Prentice Hall; 2017.
61. O'Flaherty R, Bergin A, Flampouri E, et al. Mammalian cell culture for production of recombinant proteins: A review of the critical steps in their biomanufacturing. *Biotechnology Advances*. 2020/11/01/ 2020;43:107552. doi:https://doi.org/10.1016/j.biotechadv.2020.107552
62. Wiegmann V, Giaka M, Martinez CB, Baganz F. Towards the development of automated fed-batch cell culture processes at microscale. 2019;67(5):238-241. doi:10.2144/btn-2019-0063
63. Chen C, Wong HE, Goudar CT. Upstream process intensification and continuous manufacturing. *Current Opinion in Chemical Engineering*. 2018/12/01/ 2018;22:191-198. doi:https://doi.org/10.1016/j.coche.2018.10.006
64. Bielser J-M, Wolf M, Souquet J, Broly H, Morbidelli M. Perfusion mammalian cell culture for recombinant protein manufacturing – A critical review. *Biotechnology Advances*. 2018/07/01/ 2018;36(4):1328-1340. doi:https://doi.org/10.1016/j.biotechadv.2018.04.011

65. Wlaschin KF, Hu WS. Fedbatch culture and dynamic nutrient feeding. *Advances in biochemical engineering/biotechnology*. 2006;101:43-74. doi:10.1007/10_015
66. Weegman BP, Nash P, Carlson AL, et al. Nutrient regulation by continuous feeding removes limitations on cell yield in the large-scale expansion of mammalian cell spheroids. 2013;8(10):e76611.
67. Xiao S, Ahmed W, Mohsin A, Guo M. Continuous feeding reduces the generation of metabolic byproducts and increases antibodies expression in Chinese hamster ovary-K1 cells. *Life*. 2021;11(9):945.
68. Restelli V, Wang MD, Huzel N, et al. The effect of dissolved oxygen on the production and the glycosylation profile of recombinant human erythropoietin produced from CHO cells. 2006;94(3):481-494.
69. Scatena R, Messana I, Martorana GE, et al. Mitochondrial damage and metabolic compensatory mechanisms induced by hyperoxia in the U-937 cell line. *Journal of biochemistry and molecular biology*. Jul 31 2004;37(4):454-9. doi:10.5483/bmbrep.2004.37.4.454
70. Lin AA, Miller WM. Modulation of glutathione level in CHO cells. Effects of oxygen concentration and prior exposure to hypoxia. *Annals of the New York Academy of Sciences*. Oct 13 1992;665:117-26. doi:10.1111/j.1749-6632.1992.tb42579.x
71. Cacciuttolo MA, Trinh L, Lumpkin JA, Rao G. Hyperoxia induces DNA damage in mammalian cells. *Free radical biology & medicine*. Mar 1993;14(3):267-76. doi:10.1016/0891-5849(93)90023-n
72. Fernandes-Platzgummer A, Diogo MM, Lobato da Silva C, Cabral JMS. Maximizing mouse embryonic stem cell production in a stirred tank reactor by controlling dissolved oxygen concentration and continuous perfusion operation. *Biochemical Engineering Journal*. 2014/01/15/ 2014;82:81-90. doi:https://doi.org/10.1016/j.bej.2013.11.014
73. Velez‐Suberbie ML, Tarrant RD, Tait AS, Spencer DI, Bracewell DGJBp. Impact of aeration strategy on CHO cell performance during antibody production. 2013;29(1):116-126.
74. Apostolidis PA, Tseng A, Koziol M-E, Betenbaugh MJ, Chiang BJBEJ. Investigation of low viability in sparged bioreactor CHO cell cultures points to variability in the Pluronic F-68 shear protecting component of cell culture media. 2015;98:10-17.
75. Chaudhary G, Luo R, George M, et al. Understanding the effect of high gas entrance velocity on Chinese hamster ovary (CHO) cell culture

performance and its implications on bioreactor scale‐up and sparger design. 2020;117(6):1684-1695.
76. Zhu Y, Cuenca JV, Zhou W, Varma A. NS0 cell damage by high gas velocity sparging in protein-free and cholesterol-free cultures. 2008;101(4):751-760. doi:https://doi.org/10.1002/bit.21950
77. Sieck JB, Budach WE, Suemeghy Z, et al. Adaptation for survival: Phenotype and transcriptome response of CHO cells to elevated stress induced by agitation and sparging. 2014;189:94-103.
78. Betts J, Warr S, Finka G, et al. Impact of aeration strategies on fed-batch cell culture kinetics in a single-use 24-well miniature bioreactor. *Biochemical engineering journal*. 2014;82:105-116.
79. Beutel S, Henkel S. In situ sensor techniques in modern bioprocess monitoring. *Applied microbiology and biotechnology*. 2011/09/01 2011;91(6):1493-1505. doi:10.1007/s00253-011-3470-5
80. Biechele P, Busse C, Solle D, Scheper T, Reardon KJEiLs. Sensor systems for bioprocess monitoring. 2015;15(5):469-488.
81. Lam H, Kostov Y. Optical instrumentation for bioprocess monitoring. *Advances in biochemical engineering/biotechnology*. 2009;116:1-28. doi:10.1007/10_2008_50
82. Ulber R, Frerichs J-G, Beutel S. Optical sensor systems for bioprocess monitoring. *Anal Bioanal Chem*. 2003/06// 2003;376(3):342-348. doi:10.1007/s00216-003-1930-1
83. Harms P, Kostov Y, Rao G. Bioprocess monitoring. *Current Opinion in Biotechnology*. 2002/04/01/ 2002;13(2):124-127. doi:https://doi.org/10.1016/S0958-1669(02)00295-1
84. Hanson MA, Ge X, Kostov Y, Brorson KA, Moreira AR, Rao G. Comparisons of optical pH and dissolved oxygen sensors with traditional electrochemical probes during mammalian cell culture. 2007;97(4):833-841. doi:https://doi.org/10.1002/bit.21320
85. Tric M, Lederle M, Neuner L, et al. Optical biosensor optimized for continuous in-line glucose monitoring in animal cell culture. *Anal Bioanal Chem*. 2017/09/01 2017;409(24):5711-5721. doi:10.1007/s00216-017-0511-7
86. Claßen J, Aupert F, Reardon KF, Solle D, Scheper T. Spectroscopic sensors for in-line bioprocess monitoring in research and pharmaceutical industrial application. *Anal Bioanal Chem*. 2017/01/01 2017;409(3):651-666. doi:10.1007/s00216-016-0068-x

87. Kara S, Mueller JJ, Liese AJCO-ct. Online analysis methods for monitoring of bioprocesses. 2011;29(2):38-41.
88. Landgrebe D, Haake C, Höpfner T, et al. On-line infrared spectroscopy for bioprocess monitoring. *Applied microbiology and biotechnology*. 2010/09/01 2010;88(1):11-22. doi:10.1007/s00253-010-2743-8
89. Scarff M, Arnold SA, Harvey LM, McNeil B. Near infrared spectroscopy for bioprocess monitoring and control: current status and future trends. *Critical reviews in biotechnology*. Jan-Mar 2006;26(1):17-39. doi:10.1080/07388550500513677
90. Rolinger L, Rüdt M, Hubbuch J. A critical review of recent trends, and a future perspective of optical spectroscopy as PAT in biopharmaceutical downstream processing. *Anal Bioanal Chem*. 2020/04/01 2020;412(9):2047-2064. doi:10.1007/s00216-020-02407-z
91. Sandor M, Rüdinger F, Bienert R, Grimm C, Solle D, Scheper T. Comparative study of non-invasive monitoring via infrared spectroscopy for mammalian cell cultivations. *Journal of biotechnology*. 2013/12// 2013;168(4):636-645. doi:10.1016/j.jbiotec.2013.08.002
92. Clavaud M, Roggo Y, Von Daeniken R, Liebler A, Schwabe JO. Chemometrics and in-line near infrared spectroscopic monitoring of a biopharmaceutical Chinese hamster ovary cell culture: prediction of multiple cultivation variables. *Talanta*. Jul 15 2013;111:28-38. doi:10.1016/j.talanta.2013.03.044
93. Capito F, Zimmer A, Skudas R. Mid-infrared spectroscopy-based analysis of mammalian cell culture parameters. *Biotechnology progress*. Mar-Apr 2015;31(2):578-84. doi:10.1002/btpr.2026
94. Wasalathanthri DP, Rehmann MS, Song Y, et al. Technology outlook for real-time quality attribute and process parameter monitoring in biopharmaceutical development—A review. 2020;117(10):3182-3198. doi:https://doi.org/10.1002/bit.27461
95. Lourenço ND, Lopes JA, Almeida CF, Sarraguça MC, Pinheiro HM. Bioreactor monitoring with spectroscopy and chemometrics: a review. *Anal Bioanal Chem*. 2012/09/01 2012;404(4):1211-1237. doi:10.1007/s00216-012-6073-9
96. Ude C, Schmidt-Hager J, Findeis M, John GT, Scheper T, Beutel S. Application of an online-biomass sensor in an optical multisensory platform prototype for growth monitoring of biotechnical relevant microorganism and

cell lines in single-use shake flasks. *Sensors (Basel, Switzerland)*. Sep 17 2014;14(9):17390-405. doi:10.3390/s140917390
97. Edlich A, Magdanz V, Rasch D, et al. Microfluidic reactor for continuous cultivation of Saccharomyces cerevisiae. 2010;26(5):1259-1270. doi:https://doi.org/10.1002/btpr.449
98. Takahashi MB, Leme J, Caricati CP, Tonso A, Fernández Núñez EG, Rocha JC. Artificial neural network associated to UV/Vis spectroscopy for monitoring bioreactions in biopharmaceutical processes. *Bioprocess and biosystems engineering*. 2015/06/01 2015;38(6):1045-1054. doi:10.1007/s00449-014-1346-7
99. Teixeira AP, Oliveira R, Alves PM, Carrondo MJT. Advances in on-line monitoring and control of mammalian cell cultures: Supporting the PAT initiative. *Biotechnology Advances*. 2009/11/01/ 2009;27(6):726-732. doi:https://doi.org/10.1016/j.biotechadv.2009.05.003
100. Randek J, Mandenius C-F. On-line soft sensing in upstream bioprocessing. *Critical reviews in biotechnology*. 2018/01/02 2018;38(1):106-121. doi:10.1080/07388551.2017.1312271
101. Teixeira AP, Duarte TM, Carrondo MJT, Alves PM. Synchronous fluorescence spectroscopy as a novel tool to enable PAT applications in bioprocesses. 2011;108(8):1852-1861. doi:https://doi.org/10.1002/bit.23131
102. Teixeira AP, Duarte TM, Oliveira R, Carrondo MJT, Alves PM. High-throughput analysis of animal cell cultures using two-dimensional fluorometry. *Journal of biotechnology*. 2011/02/10/ 2011;151(3):255-260. doi:https://doi.org/10.1016/j.jbiotec.2010.11.015
103. Teixeira AP, Portugal CA, Carinhas N, et al. In situ 2D fluorometry and chemometric monitoring of mammalian cell cultures. *Biotechnology and bioengineering*. Mar 1 2009;102(4):1098-106. doi:10.1002/bit.22125
104. Claßen J, Graf A, Aupert F, Solle D, Höhse M, Scheper TJEiLS. A novel LED‐based 2D‐fluorescence spectroscopy system for in‐line bioprocess monitoring of Chinese hamster ovary cell cultivations—Part II. 2019;19(5):341-351.
105. Esmonde-White KA, Cuellar M, Uerpmann C, Lenain B, Lewis IR. Raman spectroscopy as a process analytical technology for pharmaceutical manufacturing and bioprocessing. *Anal Bioanal Chem*. 2017;409(3):637-649. doi:10.1007/s00216-016-9824-1

106. Abu-Absi NR, Kenty BM, Cuellar ME, et al. Real time monitoring of multiple parameters in mammalian cell culture bioreactors using an in-line Raman spectroscopy probe. 2011;108(5):1215-1221. doi:https://doi.org/10.1002/bit.23023
107. Xu Y, Ford J, Mann C, et al. *Raman measurement of glucose in bioreactor materials*. vol 2976. BiOS '97, Part of Photonics West. SPIE; 1997.
108. Whelan J, Craven S, Glennon B. In situ Raman spectroscopy for simultaneous monitoring of multiple process parameters in mammalian cell culture bioreactors. *Biotechnology progress*. Sep-Oct 2012;28(5):1355-62. doi:10.1002/btpr.1590
109. Berry B, Moretto J, Matthews T, Smelko J, Wiltberger K. Cross-scale predictive modeling of CHO cell culture growth and metabolites using Raman spectroscopy and multivariate analysis. *Biotechnology progress*. Mar-Apr 2015;31(2):566-77. doi:10.1002/btpr.2035
110. Li B, Ray BH, Leister KJ, Ryder AG. Performance monitoring of a mammalian cell based bioprocess using Raman spectroscopy. *Analytica Chimica Acta*. 2013/09/24/ 2013;796:84-91. doi:https://doi.org/10.1016/j.aca.2013.07.058
111. Webster TA, Hadley BC, Hilliard W, Jaques C, Mason C. Development of generic raman models for a GS-KOTM CHO platform process. 2018;34(3):730-737. doi:https://doi.org/10.1002/btpr.2633
112. Matthews TE, Berry BN, Smelko J, Moretto J, Moore B, Wiltberger K. Closed loop control of lactate concentration in mammalian cell culture by Raman spectroscopy leads to improved cell density, viability, and biopharmaceutical protein production. 2016;113(11):2416-2424. doi:https://doi.org/10.1002/bit.26018
113. Berry BN, Dobrowsky TM, Timson RC, Kshirsagar R, Ryll T, Wiltberger K. Quick generation of Raman spectroscopy based in-process glucose control to influence biopharmaceutical protein product quality during mammalian cell culture. *Biotechnology progress*. Jan-Feb 2016;32(1):224-34. doi:10.1002/btpr.2205
114. Flores-Cosío G, Herrera-López EJ, Arellano-Plaza M, Gschaedler-Mathis A, Kirchmayr M, Amaya-Delgado L. Application of dielectric spectroscopy to unravel the physiological state of microorganisms: current state, prospects and limits. *Applied microbiology and biotechnology*. 2020/07/01 2020;104(14):6101-6113. doi:10.1007/s00253-020-10677-x

115. Cannizzaro C, Gügerli R, Marison I, von Stockar U. On-line biomass monitoring of CHO perfusion culture with scanning dielectric spectroscopy. *Biotechnology and bioengineering*. Dec 5 2003;84(5):597-610. doi:10.1002/bit.10809
116. Párta L, Zalai D, Borbély S, Putics Á. Application of dielectric spectroscopy for monitoring high cell density in monoclonal antibody producing CHO cell cultivations. *Bioprocess and biosystems engineering*. 2014/02/01 2014;37(2):311-323. doi:10.1007/s00449-013-0998-z
117. Yardley JE, Kell DB, Barrett J, Davey CL. On-Line, Real-Time Measurements of Cellular Biomass using Dielectric Spectroscopy. *Biotechnology and Genetic Engineering Reviews*. 2000/08/01 2000;17(1):3-36. doi:10.1080/02648725.2000.10647986
118. Justice C, Brix A, Freimark D, et al. Process control in cell culture technology using dielectric spectroscopy. *Biotechnology Advances*. 2011/07/01/ 2011;29(4):391-401. doi:https://doi.org/10.1016/j.biotechadv.2011.03.002
119. Lemke J, Söldner R, Austerjost J. Online deployment of an O-PLS model for dielectric spectroscopy-based inline monitoring of viable cell concentrations in Chinese hamster ovary cell perfusion cultivations. *Engineering in life sciences*. Jun 2023;23(6):e2200053. doi:10.1002/elsc.202200053
120. Dabros M, Dennewald D, Currie DJ, et al. Cole–Cole, linear and multivariate modeling of capacitance data for on-line monitoring of biomass. *Bioprocess and biosystems engineering*. 2009/02/01 2009;32(2):161-173. doi:10.1007/s00449-008-0234-4
121. Opel CF, Li J, Amanullah A. Quantitative modeling of viable cell density, cell size, intracellular conductivity, and membrane capacitance in batch and fed-batch CHO processes using dielectric spectroscopy. *Biotechnology progress*. 2010;26(4):1187-1199. doi:https://doi.org/10.1002/btpr.425
122. Metze S, Blioch S, Matuszczyk J, et al. Multivariate data analysis of capacitance frequency scanning for online monitoring of viable cell concentrations in small-scale bioreactors. *Anal Bioanal Chem*. 2020/04/01 2020;412(9):2089-2102. doi:10.1007/s00216-019-02096-3
123. Zeiser A, Bédard C, Voyer R, Jardin B, Tom R, Kamen AA. On-line monitoring of the progress of infection in Sf-9 insect cell cultures using relative permittivity measurements. *Biotechnology and bioengineering*. Apr

5 1999;63(1):122-6. doi:10.1002/(sici)1097-0290(19990405)63:1<122::aid-bit13>3.0.co;2-i
124. Konakovsky V, Yagtu AC, Clemens C, et al. Universal Capacitance Model for Real-Time Biomass in Cell Culture. *Sensors (Basel, Switzerland)*. Sep 2 2015;15(9):22128-50. doi:10.3390/s150922128
125. Lu F, Toh PC, Burnett I, et al. Automated dynamic fed-batch process and media optimization for high productivity cell culture process development. *Biotechnology and bioengineering*. Jan 2013;110(1):191-205. doi:10.1002/bit.24602
126. Moore B, Sanford R, Zhang A. Case study: The characterization and implementation of dielectric spectroscopy (biocapacitance) for process control in a commercial GMP CHO manufacturing process. *Biotechnology progress*. May 2019;35(3):e2782. doi:10.1002/btpr.2782
127. Zhang A, Tsang VL, Moore B, et al. Advanced process monitoring and feedback control to enhance cell culture process production and robustness. *Biotechnology and bioengineering*. Dec 2015;112(12):2495-504. doi:10.1002/bit.25684
128. Konakovsky V, Clemens C, Müller MM, et al. Metabolic Control in Mammalian Fed-Batch Cell Cultures for Reduced Lactic Acid Accumulation and Improved Process Robustness. *Bioengineering (Basel)*. 2016;3(1):5. doi:10.3390/bioengineering3010005
129. Vojinović V, Cabral JMS, Fonseca LP. Real-time bioprocess monitoring: Part I: In situ sensors. *Sensors and Actuators B: Chemical*. 2006/04/26/ 2006;114(2):1083-1091. doi:https://doi.org/10.1016/j.snb.2005.07.059
130. Pamboukian MM, Jorge SAC, Santos MG, Yokomizo AY, Pereira CA, Tonso A. Insect cells respiratory activity in bioreactor. *Cytotechnology*. 2008;57(1):37-44. doi:10.1007/s10616-007-9118-8
131. Xu J, Tang P, Yongky A, et al. Systematic development of temperature shift strategies for Chinese hamster ovary cells based on short duration cultures and kinetic modeling. *MAbs*. 2019;11(1):191-204. doi:10.1080/19420862.2018.1525262
132. KDBIO. Flow Cell Biosensor for Glucose and/or Lactate. 2016
133. Becker M, Junghans L, Teleki A, Bechmann J, Takors R. The Less the Better: How Suppressed Base Addition Boosts Production of Monoclonal Antibodies With Chinese Hamster Ovary Cells. *Front Bioeng Biotechnol*. 2019;7:76-76. doi:10.3389/fbioe.2019.00076

134. Doi T, Kajihara H, Chuman Y, Kuwae S, Kaminagayoshi T, Omasa T. Development of a scale-up strategy for Chinese hamster ovary cell culture processes using the kLa ratio as a direct indicator of gas stripping conditions. 2020;36(5):e3000. doi:https://doi.org/10.1002/btpr.3000
135. Xing Z, Kenty BM, Li ZJ, Lee SS. Scale-up analysis for a CHO cell culture process in large-scale bioreactors. *Biotechnology and bioengineering*. Jul 1 2009;103(4):733-46. doi:10.1002/bit.22287
136. Martínez-Monge I, Roman R, Comas P, et al. New developments in online OUR monitoring and its application to animal cell cultures. 2019;103(17):6903-6917.
137. Hippach MB, Schwartz I, Pei J, Huynh J, Kawai Y, Zhu MM. Fluctuations in dissolved oxygen concentration during a CHO cell culture process affects monoclonal antibody productivity and the sulfhydryl-drug conjugation process. 2018;34(6):1427-1437. doi:https://doi.org/10.1002/btpr.2697
138. Fontova A, Lecina M, López-Repullo J, et al. A simplified implementation of the stationary liquid mass balance method for on-line OUR monitoring in animal cell cultures. 2018;93(6):1757-1766. doi:https://doi.org/10.1002/jctb.5551
139. Goh H-Y, Sulu M, Alosert H, Lewis GL, Josland GD, Merriman DE. Applications of off-gas mass spectrometry in fed-batch mammalian cell culture. *Bioprocess and biosystems engineering*. 2020/03/01 2020;43(3):483-493. doi:10.1007/s00449-019-02242-2
140. Wallocha T, Popp O. Off-gas-based soft sensor for real-time monitoring of biomass and metabolism in Chinese hamster ovary cell continuous processes in single-use bioreactors. *Processes*. 2021;9(11):2073.
141. Zhou W, Hu WS. On-line characterization of a hybridoma cell culture process. *Biotechnology and bioengineering*. Jun 20 1994;44(2):170-7. doi:10.1002/bit.260440205
142. Zhou W, Rehm J, Hu WS. High viable cell concentration fed-batch cultures of hybridoma cells through on-line nutrient feeding. *Biotechnology and bioengineering*. Jun 20 1995;46(6):579-87. doi:10.1002/bit.260460611
143. Europa AF, Gambhir A, Fu P-C, Hu W-S. Multiple steady states with distinct cellular metabolism in continuous culture of mammalian cells. 2000;67(1):25-34. doi:https://doi.org/10.1002/(SICI)1097-0290(20000105)67:1<25::AID-BIT4>3.0.CO;2-K
144. Gálvez J, Lecina M, Solà C, Cairó JJ, Gòdia F. Optimization of HEK-293S cell cultures for the production of adenoviral vectors in bioreactors using on-

line OUR measurements. *Journal of biotechnology*. Jan 2012;157(1):214-22. doi:10.1016/j.jbiotec.2011.11.007
145. Martínez-Monge I, Comas P, Triquell J, Lecina M, Casablancas A, Cairó JJ. A new strategy for fed-batch process control of HEK293 cell cultures based on alkali buffer addition monitoring: comparison with O.U.R. dynamic method. *Applied microbiology and biotechnology*. 2018/12/01 2018;102(24):10469-10483. doi:10.1007/s00253-018-9388-4
146. Casablancas A, Gámez X, Lecina M, Solà C, Cairó JJ, Gòdia F. Comparison of control strategies for fed-batch culture of hybridoma cells based on on-line monitoring of oxygen uptake rate, optical cell density and glucose concentration. 2013;88(9):1680-1689. doi:https://doi.org/10.1002/jctb.4019
147. Pappenreiter M, Sissolak B, Sommeregger W, Striedner G. Oxygen Uptake Rate Soft-Sensing via Dynamic kLa Computation: Cell Volume and Metabolic Transition Prediction in Mammalian Bioprocesses. Original Research. 2019-August-21 2019;7(195)doi:10.3389/fbioe.2019.00195
148. Goldrick S, Lee K, Spencer C, et al. On-Line Control of Glucose Concentration in High-Yielding Mammalian Cell Cultures Enabled Through Oxygen Transfer Rate Measurements. *Biotechnology journal*. Apr 2018;13(4):e1700607. doi:10.1002/biot.201700607
149. Luttmann R, Bracewell DG, Cornelissen G, et al. Soft sensors in bioprocessing: A status report and recommendations. 2012;7(8):1040-1048. doi:https://doi.org/10.1002/biot.201100506
150. Carrondo MJ, Alves PM, Carinhas N, et al. How can measurement, monitoring, modeling and control advance cell culture in industrial biotechnology? *Biotechnology journal*. Dec 2012;7(12):1522-9. doi:10.1002/biot.201200226
151. Sommeregger W, Sissolak B, Kandra K, von Stosch M, Mayer M, Striedner G. Quality by control: Towards model predictive control of mammalian cell culture bioprocesses. 2017;12(7):1600546. doi:https://doi.org/10.1002/biot.201600546
152. Luo Y, Kurian V, Ogunnaike BA. Bioprocess systems analysis, modeling, estimation, and control. *Current Opinion in Chemical Engineering*. 2021/09/01/ 2021;33:100705. doi:https://doi.org/10.1016/j.coche.2021.100705

153. Mandenius C-F, Gustavsson R. Mini-review: soft sensors as means for PAT in the manufacture of bio-therapeutics. 2015;90(2):215-227. doi:https://doi.org/10.1002/jctb.4477
154. Sha S, Huang Z, Wang Z, Yoon S. Mechanistic modeling and applications for CHO cell culture development and production. *Current Opinion in Chemical Engineering*. 2018/12/01/ 2018;22:54-61. doi:https://doi.org/10.1016/j.coche.2018.08.010
155. Rathore AS, Nikita S, Jesubalan NG. Digitization in bioprocessing: The role of soft sensors in monitoring and control of downstream processing for production of biotherapeutic products. *Biosensors and Bioelectronics: X*. 2022/12/01/ 2022;12:100263. doi:https://doi.org/10.1016/j.biosx.2022.100263
156. Martínez VS, Dietmair S, Quek L-E, Hodson MP, Gray P, Nielsen LK. Flux balance analysis of CHO cells before and after a metabolic switch from lactate production to consumption. 2013;110(2):660-666. doi:https://doi.org/10.1002/bit.24728
157. de Falco B, Giannino F, Carteni F, Mazzoleni S, Kim DH. Metabolic flux analysis: a comprehensive review on sample preparation, analytical techniques, data analysis, computational modelling, and main application areas. *RSC advances*. Sep 5 2022;12(39):25528-25548. doi:10.1039/d2ra03326g
158. Martínez-Monge I, Albiol J, Lecina M, et al. Metabolic flux balance analysis during lactate and glucose concomitant consumption in HEK293 cell cultures. *Biotechnology and bioengineering*. Feb 2019;116(2):388-404. doi:10.1002/bit.26858
159. Tang P, Xu J, Louey A, et al. Kinetic modeling of Chinese hamster ovary cell culture: factors and principles. 2020;40(2):265-281.
160. Kyriakopoulos S, Ang KS, Lakshmanan M, et al. Kinetic Modeling of Mammalian Cell Culture Bioprocessing: The Quest to Advance Biomanufacturing. 2018;13(3):1700229. doi:https://doi.org/10.1002/biot.201700229
161. Solle D, Hitzmann B, Herwig C, et al. Between the Poles of Data-Driven and Mechanistic Modeling for Process Operation. 2017;89(5):542-561. doi:https://doi.org/10.1002/cite.201600175
162. Tuveri A, Pérez-García F, Lira-Parada PA, Imsland L, Bar N. Sensor fusion based on Extended and Unscented Kalman Filter for bioprocess

monitoring. *Journal of Process Control*. 2021/10/01/ 2021;106:195-207. doi:https://doi.org/10.1016/j.jprocont.2021.09.005
163. Simutis R, Lübbert AJB. Hybrid approach to state estimation for bioprocess control. 2017;4(1):21.
164. Khodarahmi M, Maihami V. A Review on Kalman Filter Models. *Archives of Computational Methods in Engineering*. 2023/01/01 2023;30(1):727-747. doi:10.1007/s11831-022-09815-7
165. Iglesias Jr CF, Xu X, Mehta V, et al. Monitoring the Recombinant Adeno-Associated Virus Production using Extended Kalman Filter. 2022;10(11):2180.
166. Narayanan H, Behle L, Luna MF, et al. Hybrid‐ekf: Hybrid model coupled with extended Kalman filter for real‐time monitoring and control of mammalian cell culture. 2020;117(9):2703-2714.
167. Tsopanoglou A, del Val IJJCOiCE. Moving towards an era of hybrid modelling: advantages and challenges of coupling mechanistic and data-driven models for upstream pharmaceutical bioprocesses. 2021;32:100691.
168. Narayanan H, Sokolov M, Morbidelli M, Butté AJB, bioengineering. A new generation of predictive models: The added value of hybrid models for manufacturing processes of therapeutic proteins. 2019;116(10):2540-2549.
169. Ohadi K, Legge RL, Budman HMJB, bioengineering. Development of a soft‐sensor based on multi‐wavelength fluorescence spectroscopy and a dynamic metabolic model for monitoring mammalian cell cultures. 2015;112(1):197-208.
170. Iglesias Jr CF, Ristovski M, Bolic M, Cuperlovic-Culf MJB. rAAV Manufacturing: The Challenges of Soft Sensing during Upstream Processing. 2023;10(2):229.
171. Zavala-Ortiz DA, Denner A, Aguilar-Uscanga MG, Marc A, Ebel B, Guedon E. Comparison of partial least square, artificial neural network, and support vector regressions for real-time monitoring of CHO cell culture processes using in situ near-infrared spectroscopy. *Biotechnology and bioengineering*. Feb 2022;119(2):535-549. doi:10.1002/bit.27997
172. Le H, Kabbur S, Pollastrini L, et al. Multivariate analysis of cell culture bioprocess data—lactate consumption as process indicator. *Journal of biotechnology*. 2012;162(2-3):210-223.
173. Aehle M, Simutis R, Lübbert A. Comparison of viable cell concentration estimation methods for a mammalian cell cultivation process. *Cytotechnology*. Oct 2010;62(5):413-22. doi:10.1007/s10616-010-9291-z

174. Schmidberger T, Posch C, Sasse A, Gülch C, Huber R. Progress toward forecasting product quality and quantity of mammalian cell culture processes by performance-based modeling. 2015;31(4):1119-1127. doi:https://doi.org/10.1002/btpr.2105
175. Charaniya S, Le H, Rangwala H, et al. Mining manufacturing data for discovery of high productivity process characteristics. *Journal of biotechnology*. 2010/06/01/ 2010;147(3):186-197. doi:https://doi.org/10.1016/j.jbiotec.2010.04.005
176. Poth M, Magill G, Filgertshofer A, Popp O, Großkopf T. Extensive evaluation of machine learning models and data preprocessings for Raman modeling in bioprocessing. 2022;53(9):1580-1591. doi:https://doi.org/10.1002/jrs.6402
177. Hassan SS, Farhan M, Mangayil R, Huttunen H, Aho T. Bioprocess data mining using regularized regression and random forests. *BMC Systems Biology*. 2013/08/12 2013;7(1):S5. doi:10.1186/1752-0509-7-S1-S5
178. Karimi Alavijeh M, Baker I, Lee YY, Gras SL. Digitally enabled approaches for the scale up of mammalian cell bioreactors. *Digital Chemical Engineering*. 2022/09/01/ 2022;4:100040. doi:https://doi.org/10.1016/j.dche.2022.100040
179. Park S-Y, Kim S-J, Park C-H, Kim J, Lee D-Y. Data-driven prediction models for forecasting multistep ahead profiles of mammalian cell culture toward bioprocess digital twins. *Biotechnology and bioengineering*. 2023;120:2494–2508. doi:https://doi.org/10.1002/bit.28405
180. Wong WC, Chee E, Li J, Wang XJM. Recurrent neural network-based model predictive control for continuous pharmaceutical manufacturing. 2018;6(11):242.
181. Iglesias JC, Mehta V, Venereo-Sanchez A, et al. Handling Massive Proportion of Missing Labels in Multivariate Long-Term Time Series Forecasting. IOP Publishing; 2021:012170.
182. Smiatek J, Clemens C, Herrera LM, et al. Generic and specific recurrent neural network models: Applications for large and small scale biopharmaceutical upstream processes. 2021;31:e00640.
183. Irfan K. *Carbon dioxide control in bioreactors and the application of principal component analysis to cell culture process data*. Newcastle University; 2017.
184. Tescione L, Lambropoulos J, Paranandi MR, Makagiansar H, Ryll T. Application of bioreactor design principles and multivariate analysis for

development of cell culture scale down models. *Biotechnology and bioengineering*. Jan 2015;112(1):84-97. doi:10.1002/bit.25330
185. Goudar CT, Piret JM, Konstantinov KB. Estimating cell specific oxygen uptake and carbon dioxide production rates for mammalian cells in perfusion culture. *Biotechnol Progress*. 2011;27(5):1347-1357.
186. Rathore AS, Mittal S, Pathak M, Mahalingam V. Chemometrics application in biotech processes: assessing comparability across processes and scales. *Journal of Chemical Technology & Biotechnology*. 2014;89(9):1311-1316. doi:https://doi.org/10.1002/jctb.4428
187. Goldrick S, Sandner V, Cheeks M, et al. Multivariate Data Analysis Methodology to Solve Data Challenges Related to Scale-Up Model Validation and Missing Data on a Micro-Bioreactor System. 2020;15(3):1800684. doi:https://doi.org/10.1002/biot.201800684
188. Nomikos P, MacGregor JF. Monitoring batch processes using multiway principal component analysis. 1994;40(8):1361-1375. doi:https://doi.org/10.1002/aic.690400809
189. Wold S, Kettaneh N, Fridén H, Holmberg A. Modelling and diagnostics of batch processes and analogous kinetic experiments. *Chemometrics and Intelligent Laboratory Systems*. 1998/12/14/ 1998;44(1):331-340. doi:https://doi.org/10.1016/S0169-7439(98)00162-2
190. Glassey J. Multivariate Data Analysis for Advancing the Interpretation of Bioprocess Measurement and Monitoring Data. In: Mandenius C-F, Titchener-Hooker NJ, eds. *Measurement, Monitoring, Modelling and Control of Bioprocesses*. Springer Berlin Heidelberg; 2013:167-191.
191. Mercier SM, Diepenbroek B, Dalm MCF, Wijffels RH, Streefland M. Multivariate data analysis as a PAT tool for early bioprocess development data. *Journal of biotechnology*. 2013/09/10/ 2013;167(3):262-270. doi:https://doi.org/10.1016/j.jbiotec.2013.07.006
192. Suarez-Zuluaga DA, Borchert D, Driessen NN, Bakker WAM, Thomassen YE. Accelerating bioprocess development by analysis of all available data: A USP case study. *Vaccine*. 2019/11/08/ 2019;37(47):7081-7089. doi:https://doi.org/10.1016/j.vaccine.2019.07.026
193. Kirdar AO, Conner JS, Baclaski J, Rathore AS. Application of multivariate analysis toward biotech processes: case study of a cell-culture unit operation. *Biotechnology progress*. Jan-Feb 2007;23(1):61-7. doi:10.1021/bp060377u

194. Tsang VL, Wang AX, Yusuf-Makagiansar H, Ryll T. Development of a scale down cell culture model using multivariate analysis as a qualification tool. *Biotechnology progress*. Jan-Feb 2014;30(1):152-60. doi:10.1002/btpr.1819
195. Rafferty C, Johnson K, O'Mahony J, Burgoyne B, Rea R, Balss KM. Analysis of chemometric models applied to Raman spectroscopy for monitoring key metabolites of cell culture. *Biotechnology progress*. Jul 2020;36(4):e2977. doi:10.1002/btpr.2977
196. Molina RS, Molina-Rodríguez MA, Rincón FM, Maldonado JD. Cardiac operative risk in Latin America: a comparison of machine learning models vs EuroSCORE-II. *Ann Thorac Surg*. 2022;113(1):92-99.
197. Rozemberczki B, Watson L, Bayer P, et al. The shapley value in machine learning. 2022;
198. Lundberg S, Lee S-I. A unified approach to interpreting model predictions. *Advances in neural information processing systems*. 2017;30
199. Yang R. Who dies from COVID-19? Post-hoc explanations of mortality prediction models using coalitional game theory, surrogate trees, and partial dependence plots. 2020:2020.06.07.20124933. doi:10.1101/2020.06.07.20124933 %J medRxiv
200. Stuible M, Gervais C, Lord-Dufour S, et al. Rapid, high-yield production of full-length SARS-CoV-2 spike ectodomain by transient gene expression in CHO cells. *Journal of biotechnology*. 2021;326:21-27. doi:10.1016/j.jbiotec.2020.12.005
201. Joubert S, Stuible M, Lord‐Dufour S, et al. A CHO stable pool production platform for rapid clinical development of trimeric SARS‐CoV‐2 spike subunit vaccine antigens. *Biotechnology and bioengineering*. 2023;120(7):1746-1761.
202. Stuible M, van Lier F, Croughan MS, Durocher Y. Beyond preclinical research: production of CHO-derived biotherapeutics for toxicology and early-phase trials by transient gene expression or stable pools. *Current Opinion in Chemical Engineering*. 2018;22:145-151. doi:https://doi.org/10.1016/j.coche.2018.09.010
203. Alavijeh MK, Baker I, Lee YY, Gras SL. Digitally enabled approaches for the scale up of mammalian cell bioreactors. *Digital Chemical Engineering*. 2022;4:100040.

204. Irfan K. *Carbon dioxide control in bioreactors and the application of principal component analysis to cell culture process data*. Dissertation. Newcastle University; 2017.
205. Tescione L, Lambropoulos J, Paranandi MR, Makagiansar H, Ryll T. Application of bioreactor design principles and multivariate analysis for development of cell culture scale down models. *Biotechnology Bioengineering*. 2015;112(1):84-97.
206. Goldrick S, Sandner V, Cheeks M, et al. Multivariate Data Analysis Methodology to Solve Data Challenges Related to Scale-Up Model Validation and Missing Data on a Micro-Bioreactor System. *Biotechnology journal*. 2020;15(3):1800684. doi:https://doi.org/10.1002/biot.201800684
207. Facco P, Zomer S, Rowland-Jones RC, et al. Using data analytics to accelerate biopharmaceutical process scale-up. *Biochemical Engineering Journal*. 2020;164:107791. doi:https://doi.org/10.1016/j.bej.2020.107791
208. Powers DN, Trunfio N, Velugula‐Yellela SR, Angart P, Faustino A, Agarabi C. Multivariate data analysis of growth medium trends affecting antibody glycosylation. *Biotechnology progress*. 2020;36(1):e2903.
209. Nomikos P, MacGregor JF. Monitoring batch processes using multiway principal component analysis. *AIChE Journal*. 1994;40(8):1361-1375. doi:https://doi.org/10.1002/aic.690400809
210. Wold S, Kettaneh N, Fridén H, Holmberg A. Modelling and diagnostics of batch processes and analogous kinetic experiments. *Chemometrics intelligent laboratory systems*. 1998;44(1-2):331-340.
211. Glassey J. Multivariate Data Analysis for Advancing the Interpretation of Bioprocess Measurement and Monitoring Data. *Advances in biochemical engineering/biotechnology*. 2013:167-191. doi:10.1007/10_2012_171
212. Mercier SM, Diepenbroek B, Dalm MC, Wijffels RH, Streefland M. Multivariate data analysis as a PAT tool for early bioprocess development data. *Journal of biotechnology*. 2013;167(3):262-270.
213. Roychoudhury P, O'Kennedy R, Faulkner J, McNeil B, M Harvey L. Implementing multivariate data analysis to monitor mammalian cell culture processes. *European Pharmaceutical Review*. 2013;18(3):15-20.
214. Kirdar AO, Conner JS, Baclaski J, Rathore AS. Application of multivariate analysis toward biotech processes: case study of a cell‐culture unit operation. *Biotechnology progress*. 2007;23(1):61-67.

215. Tsang VL, Wang AX, Yusuf‐Makagiansar H, Ryll T. Development of a scale down cell culture model using multivariate analysis as a qualification tool. *Biotechnology progress*. 2014;30(1):152-160.
216. Salim T, Chauhan G, Templeton N, Ling WLW. Using MVDA with stoichiometric balances to optimize amino acid concentrations in chemically defined CHO cell culture medium for improved culture performance. *Biotechnology Bioengineering*

2022;119(2):452-469.
217. Ding X, Liu J, Yang F, Cao J. Random radial basis function kernel-based support vector machine. *Journal of the Franklin Institute*. 2021;358(18):10121-10140.
doi:https://doi.org/10.1016/j.jfranklin.2021.10.005
218. Hassan SS, Farhan M, Mangayil R, Huttunen H, Aho T. Bioprocess data mining using regularized regression and random forests. *BMC systems biology*. 2013;7(1):1-7.
219. Rafferty C, Johnson K, O'Mahony J, Burgoyne B, Rea R, Balss KM. Analysis of chemometric models applied to Raman spectroscopy for monitoring key metabolites of cell culture. *Biotechnology progress*. 2020;36(4):e2977.
220. Poulain A, Perret S, Malenfant F, Mullick A, Massie B, Durocher Y. Rapid protein production from stable CHO cell pools using plasmid vector and the cumate gene-switch. *Journal of biotechnology*. 2017;255:16-27. doi:10.1016/j.jbiotec.2017.06.009
221. Kucheryavskiy S. mdatools – R package for chemometrics. *Chemometrics and Intelligent Laboratory Systems*. 2020;198:103937. doi:https://doi.org/10.1016/j.chemolab.2020.103937
222. Kuhn M, Wing J, Weston S, et al. Package 'caret'. 2020;223:7.
223. *Signal: Signal processing*. 2014. http://r-forge.r-project.org/projects/signal/
224. Davison AC, Hinkley DV. *Bootstrap methods and their application*. Cambridge university press; 1997.
225. Canty A, Ripley BJCsRfhcr-powpb. boot: Bootstrap R (S-Plus) Functions. R package version 1.3-18. 2016;
226. Greenwell B, Greenwell MB. Package 'fastshap'. 2020.
227. Lin J, Takagi M, Qu Y, Yoshida T. Possible strategy for on-line monitoring and control of hybridoma cell culture. *Biochemical engineering journal*. 2002;11(2-3):205-209.

228. Hippach M, Schwartz I, Pei J, Huynh J, Kawai Y, Zhu M. Fluctuations in dissolved oxygen concentration during a CHO cell culture process affects monoclonal antibody productivity and the sulfhydryl‐drug conjugation process. *Biotechnology progress*. 2018;34(6):1427-1437.
229. Tian J, He Q, Oliveira C, et al. Increased MSX level improves biological productivity and production stability in multiple recombinant GS CHO cell lines. *Eng Life Sci*. 2020;20(3-4):112-125.
230. Kuhn M. Futility analysis in the cross-validation of machine learning models. *arXiv preprint arXiv:14056974*. 2014;
231. Aman F, Rauf A, Ali R, Hussain J, Ahmed I. Balancing Complex Signals for Robust Predictive Modeling. *Sensors (Basel, Switzerland)*. Dec 18 2021;21(24)doi:10.3390/s21248465
232. Nohara Y, Matsumoto K, Soejima H, Nakashima N. Explanation of machine learning models using shapley additive explanation and application for real data in hospital. *Comput Methods Programs Biomed*. 2022;214:106584. doi:10.1016/j.cmpb.2021.106584
233. Yang R. Who dies from COVID-19? Post-hoc explanations of mortality prediction models using coalitional game theory, surrogate trees, and partial dependence plots. *MedRxiv*. 2020:2020.06. 07.20124933. doi:https://doi.org/10.1101/2020.06.07.20124933
234. Scapin D, Cisotto G, Gindullina E, Badia L. Shapley value as an aid to biomedical machine learning: a heart disease dataset analysis. IEEE; 2022:933-939.
235. Grilo AL, Mantalaris A. Apoptosis: A mammalian cell bioprocessing perspective. *Biotechnology advances*. 2019;37(3):459-475.
236. Downey BJ, Graham LJ, Breit JF, Glutting NK. A novel approach for using dielectric spectroscopy to predict viable cell volume (VCV) in early process development. *Biotechnology progress*. 2014;30(2):479-487. doi:https://doi.org/10.1002/btpr.1845
237. Pappenreiter M, Sissolak B, Sommeregger W, Striedner G. Oxygen Uptake Rate Soft-Sensing via Dynamic kLa Computation: Cell Volume and Metabolic Transition Prediction in Mammalian Bioprocesses. *Frontiers in Bioengineering Biotechnology Advances*. 2019;7:195.
238. Kuystermans D, Al-Rubeai M. Bioreactor systems for producing antibody from mammalian cells. *Antibody expression production*. 2011:25-52.

239. Krampe B, Al-Rubeai M. Cell death in mammalian cell culture: molecular mechanisms and cell line engineering strategies. *Cytotechnology*. 2010;62(3):175-88. doi:10.1007/s10616-010-9274-0
240. Anane E, Knudsen IM, Wilson GC. Scale-down cultivation in mammalian cell bioreactors—The effect of bioreactor mixing time on the response of CHO cells to dissolved oxygen gradients. *Biochemical Engineering Journal*. 2021/02/01/ 2021;166:107870. doi:https://doi.org/10.1016/j.bej.2020.107870
241. Zhang A, Tsang VL, Moore B, et al. Advanced process monitoring and feedback control to enhance cell culture process production and robustness. *Biotechnology Bioengineering* Dec 2015;112(12):2495-504. doi:10.1002/bit.25684
242. Opel CF, Li J, Amanullah A. Quantitative modeling of viable cell density, cell size, intracellular conductivity, and membrane capacitance in batch and fed-batch CHO processes using dielectric spectroscopy. *Biotechnology progress*. 2010;26(4):1187-1199. doi:https://doi.org/10.1002/btpr.425
243. Lee HW, Carvell J, Brorson K, Yoon S. Dielectric spectroscopy-based estimation of VCD in CHO cell culture. *Journal of Chemical Technology & Biotechnology*. 2015;90(2):273-282. doi:https://doi.org/10.1002/jctb.4522
244. Martínez-Monge I, Martínez C, Decker M, et al. Soft-sensors application for automated feeding control in high-throughput mammalian cell cultures. *Biotechnology bioengineering* Apr 2022;119(4):1077-1090. doi:10.1002/bit.28032
245. Templeton N, Dean J, Reddy P, Young JD. Peak antibody production is associated with increased oxidative metabolism in an industrially relevant fed-batch CHO cell culture. *Biotechnology Bioengineering*. 2013;110(7):2013-2024. doi:https://doi.org/10.1002/bit.24858
246. Martínez-Monge I, Roman R, Comas P, et al. New developments in online OUR monitoring and its application to animal cell cultures. *Applied microbiology biotechnology*. 2019;103(17):6903-6917.
247. Joubert S, Stuible M, Lord-Dufour S, et al. A CHO stable pool production platform for rapid clinical development of trimeric SARS-CoV-2 spike subunit vaccine antigens. *Biotechnology and bioengineering*. Jul 2023;120(7):1746-1761. doi:10.1002/bit.28387
248. Lemire L, Pham PL, Durocher Y, Henry O. Practical Considerations for the Scale-Up of Chinese Hamster Ovary (CHO) Cell Cultures. *Cell Culture Engineering and Technology*. Springer; 2021:367-400.

249. Garcia-Ochoa F, Gomez E. Bioreactor scale-up and oxygen transfer rate in microbial processes: an overview. *Biotechnology advances*. 2009;27(2):153-176.
250. Hoshan L, Jiang R, Moroney J, et al. Effective bioreactor pH control using only sparging gases. *Biotechnology Progress*. 2019;35(1):e2743.
251. Nie L, Gao D, Jiang H, et al. Development and Qualification of a Scale-Down Mammalian Cell Culture Model and Application in Design Space Development by Definitive Screening Design. *AAPS PharmSciTech*. 2019;20(6):1-10.
252. Möller J, Rodríguez TH, Müller J, et al. Model uncertainty-based evaluation of process strategies during scale-up of biopharmaceutical processes. *Computers & chemical engineering*. 2020;134:106693.
253. Tescione L, Lambropoulos J, Paranandi MR, Makagiansar H, Ryll T. Application of bioreactor design principles and multivariate analysis for development of cell culture scale down models. *Biotechnology Bioengineering*. 2015;112(1):84-97.
254. Mostafa SS, Gu XJBp. Strategies for improved dCO2 removal in large‐scale fed‐batch cultures. 2003;19(1):45-51.
255. Mitchell‐Logean C, Murhammer DW. Bioreactor headspace purging reduces dissolved carbon dioxide accumulation in insect cell cultures and enhances cell growth. *Biotechnology progress*. 1997;13(6):875-877.
256. Sieblist C, Hägeholz O, Aehle M, Jenzsch M, Pohlscheidt M, Lübbert A. Insights into large‐scale cell‐culture reactors: II. Gas‐phase mixing and CO2 stripping. *Biotechnology journal*. 2011;6(12):1547-1556.
257. Naciri M, Kuystermans D, Al-Rubeai M. Monitoring pH and dissolved oxygen in mammalian cell culture using optical sensors. *Cytotechnology*. 2008;57:245-250.
258. Heidemann R, Lütkemeyer D, Büntemeyer H, Lehmann J. Effects of dissolved oxygen levels and the role of extra-and intracellular amino acid concentrations upon the metabolism of mammalian cell lines during batch and continuous cultures. *Cytotechnology*. 1998;26(3):185-197.
259. Restelli V, Wang MD, Huzel N, Ethier M, Perreault H, Butler M. The effect of dissolved oxygen on the production and the glycosylation profile of recombinant human erythropoietin produced from CHO cells. *Biotechnology bioengineering*. 2006;94(3):481-494.
260. Möller J, Bhat K, Riecken K, Pörtner R, Zeng AP, Jandt U. Process‐induced cell cycle oscillations in CHO cultures: Online monitoring and

model‐based investigation. *Biotechnology bioengineering*. 2019;116(11):2931-2943.
261. Xu WJ, Lin Y, Mi CL, Pang JY, Wang TY. Progress in fed-batch culture for recombinant protein production in CHO cells. *Applied microbiology and biotechnology*. Feb 2023;107(4):1063-1075. doi:10.1007/s00253-022-12342-x
262. Xu S, Hoshan L, Jiang R, et al. A practical approach in bioreactor scale-up and process transfer using a combination of constant P/V and vvm as the criterion. *Biotechnology progress*. 2017;33(4):1146-1159. doi:https://doi.org/10.1002/btpr.2489
263. Xu J, Tang P, Yongky A, et al. Systematic development of temperature shift strategies for Chinese hamster ovary cells based on short duration cultures and kinetic modeling. Taylor & Francis; 2019:191-204.
264. Hsu WT, Aulakh RP, Traul DL, Yuk IH. Advanced microscale bioreactor system: a representative scale-down model for bench-top bioreactors. *Cytotechnology*. Dec 2012;64(6):667-78. doi:10.1007/s10616-012-9446-1
265. Handlogten MW, Zhu M, Ahuja SJBEJ. Intracellular response of CHO cells to oxidative stress and its influence on metabolism and antibody production. 2018;133:12-20.
266. Baez A, Shiloach J. Effect of elevated oxygen concentration on bacteria, yeasts, and cells propagated for production of biological compounds. *Microbial cell factories*. 2014;13:1-7.
267. Goudar CT, Piret JM, Konstantinov KB. Estimating cell specific oxygen uptake and carbon dioxide production rates for mammalian cells in perfusion culture. *Biotechnology Progress*. 2011;27(5):1347-1357.
268. Wiese AG, Pacifici RE, Davies KJ. Transient adaptation of oxidative stress in mammalian cells. *Archives of biochemistry and biophysics*. Apr 1 1995;318(1):231-40. doi:10.1006/abbi.1995.1225
269. Santa-Gonzalez GA, Gomez-Molina A, Arcos-Burgos M, Meyer JN, Camargo M. Distinctive adaptive response to repeated exposure to hydrogen peroxide associated with upregulation of DNA repair genes and cell cycle arrest. *Redox biology*. Oct 2016;9:124-133. doi:10.1016/j.redox.2016.07.004
270. Rish AJ, Drennen JK, Anderson CA. Metabolic trends of Chinese hamster ovary cells in biopharmaceutical production under batch and fed-batch conditions. *Biotechnology progress*. Jan 2022;38(1):e3220. doi:10.1002/btpr.3220

271. Huang Y-M, Hu W, Rustandi E, Chang K, Yusuf-Makagiansar H, Ryll T. Maximizing productivity of CHO cell-based fed-batch culture using chemically defined media conditions and typical manufacturing equipment. *Biotechnology progress*. 2010;26(5):1400-1410. doi:https://doi.org/10.1002/btpr.436
272. Velez‐Suberbie ML, Tarrant RD, Tait AS, Spencer DI, Bracewell DG. Impact of aeration strategy on CHO cell performance during antibody production. *Biotechnology progress*. 2013;29(1):116-126.
273. Grein TA, Loewe D, Dieken H, Weidner T, Salzig D, Czermak P. Aeration and shear stress are critical process parameters for the production of oncolytic measles virus. *Frontiers in bioengineering biotechnology*. 2019;7:78.
274. Xing Z, Lewis AM, Borys MC, Li ZJ. A carbon dioxide stripping model for mammalian cell culture in manufacturing scale bioreactors. *Biotechnology Bioengineering*. 2017;114(6):1184-1194.
275. Becker M, Junghans L, Teleki A, Bechmann J, Takors R. The less the better: how suppressed base addition boosts production of monoclonal antibodies with Chinese hamster ovary cells. *Frontiers in Bioengineering Biotechnology*. 2019;7:76.
276. Xiao W, Wang R-S, Handy DE, Loscalzo J. NAD (H) and NADP (H) redox couples and cellular energy metabolism. *Antioxidants redox signaling*. 2018;28(3):251-272.
277. Hosios AM, Vander Heiden MG. The redox requirements of proliferating mammalian cells. *Journal of Biological Chemistry*. 2018;293(20):7490-7498.
278. Pereira S, Kildegaard HF, Andersen MR. Impact of CHO Metabolism on Cell Growth and Protein Production: An Overview of Toxic and Inhibiting Metabolites and Nutrients. *Biotechnology journal*. 2018;13(3):1700499. doi:https://doi.org/10.1002/biot.201700499
279. Synoground BF, McGraw CE, Elliott KS, et al. Transient ammonia stress on Chinese hamster ovary (CHO) cells yield alterations to alanine metabolism and IgG glycosylation profiles. *Biotechnology journal*. 2021;16(7):2100098. doi:https://doi.org/10.1002/biot.202100098
280. Chen P, Harcum SW. Effects of elevated ammonium on glycosylation gene expression in CHO cells. *Metabolic Engineering*. 2006/03/01/ 2006;8(2):123-132. doi:https://doi.org/10.1016/j.ymben.2005.10.002

281. Yoo DW, Lee SM, Moon SY, Kim IS, Chang CL. Evaluation of conductivity‐based osmolality measurement in urine using the Sysmex UF5000. *Journal of clinical laboratory analysis*. 2021;35(1):e23586.
282. Mitsubayashi K, Toma K, Arakawa T. 12 - Cavitas bio/chemical sensors for Internet of Things in healthcare. In: Mitsubayashi K, Niwa O, Ueno Y, eds. *Chemical, Gas, and Biosensors for Internet of Things and Related Applications*. Elsevier; 2019:177-191.
283. Kim NS, Lee GM. Response of recombinant Chinese hamster ovary cells to hyperosmotic pressure: effect of Bcl-2 overexpression. *Journal of biotechnology*. 2002;95(3):237-248.
284. Han YK, Kim YG, Kim JY, Lee GM. Hyperosmotic stress induces autophagy and apoptosis in recombinant Chinese hamster ovary cell culture. *Biotechnology bioengineering*. 2010;105(6):1187-1192.
285. Kamachi Y, Omasa T. Development of hyper osmotic resistant CHO host cells for enhanced antibody production. *J Biosci Bioeng*. Apr 2018;125(4):470-478. doi:10.1016/j.jbiosc.2017.11.002
286. Weegman BP, Nash P, Carlson AL, et al. Nutrient regulation by continuous feeding removes limitations on cell yield in the large-scale expansion of mammalian cell spheroids. *PLoS One*. 2013;8(10):e76611.
287. Horvat J, Narat M, Spadiut O. The effect of amino acid supplementation in an industrial Chinese Hamster Ovary process. *Biotechnology progress*. Sep 2020;36(5):e3001. doi:10.1002/btpr.3001
288. Shen D, Kiehl TR, Khattak SF, et al. Transcriptomic responses to sodium chloride-induced osmotic stress: A study of industrial fed-batch CHO cell cultures. *Biotechnology progress*. 2010;26(4):1104-1115. doi:https://doi.org/10.1002/btpr.398
289. Wu MH, Dimopoulos G, Mantalaris A, Varley J. The effect of hyperosmotic pressure on antibody production and gene expression in the GS-NS0 cell line. *Biotechnology and applied biochemistry*. Aug 2004;40(Pt 1):41-6. doi:10.1042/ba20030170
290. Omasa T, Higashiyama K, Shioya S, Suga K. Effects of lactate concentration on hybridoma culture in lactate-controlled fed-batch operation. *Biotechnology and bioengineering*. Mar 5 1992;39(5):556-64. doi:10.1002/bit.260390511
291. Sheikholeslami Z, Jolicoeur M, Henry O. Probing the metabolism of an inducible mammalian expression system using extracellular isotopomer analysis. *Journal of biotechnology*. 2013;164(4):469-478.

292. Fan L, Frye CC, Racher AJ. The use of glutamine synthetase as a selection marker: recent advances in Chinese hamster ovary cell line generation processes. *Pharmaceutical Bioprocessing*. 2013;1(5):487-502.
293. Budge JD, Roobol J, Singh G, et al. A proline metabolism selection system and its application to the engineering of lipid biosynthesis in Chinese hamster ovary cells. *Metabolic Engineering Communications*. 2021;13:e00179.
294. Carinhas N, Duarte TM, Barreiro LC, Carrondo MJT, Alves PM, Teixeira AP. Metabolic signatures of GS-CHO cell clones associated with butyrate treatment and culture phase transition. *Biotechnology and bioengineering*. 2013;110(12):3244-3257. doi:https://doi.org/10.1002/bit.24983
295. Chevallier V, Andersen MR, Malphettes L. Oxidative stress‐alleviating strategies to improve recombinant protein production in CHO cells. *Biotechnology and bioengineering*. 2020;117(4):1172-1186.
296. Chevallier V, Schoof EM, Malphettes L, Andersen MR, Workman CT. Characterization of glutathione proteome in CHO cells and its relationship with productivity and cholesterol synthesis. *Biotechnology and bioengineering*. 2020;117(11):3448-3458.
297. Orellana CA, Marcellin E, Schulz BL, Nouwens AS, Gray PP, Nielsen LK. High-antibody-producing Chinese hamster ovary cells up-regulate intracellular protein transport and glutathione synthesis. *Journal of proteome research*. 2015;14(2):609-618.
298. Dean J, Reddy P. Metabolic analysis of antibody producing CHO cells in fed-batch production. *Biotechnology and bioengineering*. 2013;110(6):1735-1747. doi:https://doi.org/10.1002/bit.24826
299. Sellick CA, Croxford AS, Maqsood AR, et al. Metabolite profiling of CHO cells: Molecular reflections of bioprocessing effectiveness. *Biotechnology journal*. 2015;10(9):1434-1445. doi:https://doi.org/10.1002/biot.201400664
300. Sellick CA, Croxford AS, Maqsood AR, et al. Metabolite profiling of recombinant CHO cells: Designing tailored feeding regimes that enhance recombinant antibody production. *Biotechnology and bioengineering*. 2011;108(12):3025-3031. doi:https://doi.org/10.1002/bit.23269
301. Rico J, Nantel A, Pham PL, et al. Kinetic model of metabolism of monoclonal antibody producing CHO cells. *Current Metabolomics*. 2018;6(3):207-217.
302. Savizi ISP, Maghsoudi N, Motamedian E, Lewis NE, Shojaosadati SA. Valine feeding reduces ammonia production through rearrangement of

metabolic fluxes in central carbon metabolism of CHO cells. *Applied microbiology and biotechnology*. 2022;106(3):1113-1126.
303. Chitwood DG, Uy L, Fu W, Klaubert SR, Harcum SW, Saski CA. Dynamics of Amino Acid Metabolism, Gene Expression, and Circulomics in a Recombinant Chinese Hamster Ovary Cell Line Adapted to Moderate and High Levels of Extracellular Lactate. *Genes*. 2023;14(8):1576.
304. Kirsch BJ, Bennun SV, Mendez A, et al. Metabolic analysis of the asparagine and glutamine dynamics in an industrial Chinese hamster ovary fed-batch process. *Biotechnology and bioengineering*. 2022;119(3):807-819. doi:https://doi.org/10.1002/bit.27993
305. Carrillo-Cocom LM, Genel-Rey T, Araíz-Hernández D, et al. Amino acid consumption in naïve and recombinant CHO cell cultures: producers of a monoclonal antibody. *Cytotechnology*. Oct 2015;67(5):809-20. doi:10.1007/s10616-014-9720-5
306. Xing Z, Kenty B, Koyrakh I, Borys M, Pan S-H, Li ZJ. Optimizing amino acid composition of CHO cell culture media for a fusion protein production. *Process Biochemistry*. 2011/07/01/ 2011;46(7):1423-1429. doi:https://doi.org/10.1016/j.procbio.2011.03.014
307. Coronel J, Klausing S, Heinrich C, Noll T, Figueredo-Cardero A, Castilho LR. Valeric acid supplementation combined to mild hypothermia increases productivity in CHO cell cultivations. *Biochemical Engineering Journal*. 2016/10/15/ 2016;114:101-109. doi:https://doi.org/10.1016/j.bej.2016.06.031
308. Bergin A, Carvell J, Butler M. Applications of bio-capacitance to cell culture manufacturing. *Biotechnology Advances*. 2022/12/01/ 2022;61:108048. doi:https://doi.org/10.1016/j.biotechadv.2022.108048
309. Metze S, Blioch S, Matuszczyk J, et al. Multivariate data analysis of capacitance frequency scanning for online monitoring of viable cell concentrations in small-scale bioreactors. *Analytical Bioanalytical Chemistry*. 2020;412:2089-2102.
310. Reyes SJ, Keebler M, Schulte A, et al. Impact of Protein Production on Metabolic Activity of CHO Stable Cell Line Producing Palivizumab. Kuhner Accessed 2024-04-12, 2024. https://kuhner.com/wAssets/docs/download/Poster/Poster-NRC/TOM_Antibody_Poster-4_NRC-Kuhner.pdf

311. Reyes SJ, Durocher Y, Pham PL, Henry O. Modern Sensor Tools and Techniques for Monitoring Controlling Improving Cell Culture Processes. *Processes*. 2022;10(2)doi:10.3390/pr10020189
312. Luttmann R, Bracewell DG, Cornelissen G, et al. Soft sensors in bioprocessing: A status report and recommendations. *Biotechnology journal*. 2012;7(8):1040-1048. doi:https://doi.org/10.1002/biot.201100506
313. Sommeregger W, Sissolak B, Kandra K, von Stosch M, Mayer M, Striedner G. Quality by control: Towards model predictive control of mammalian cell culture bioprocesses. *Biotechnology journal*. 2017;12(7):1600546. doi:https://doi.org/10.1002/biot.201600546
314. Mandenius C-F, Gustavsson R. Mini-review: soft sensors as means for PAT in the manufacture of bio-therapeutics. *Journal of Chemical Technology & Biotechnology*. 2015;90(2):215-227. doi:https://doi.org/10.1002/jctb.4477
315. Randek J, Mandenius CF. On-line soft sensing in upstream bioprocessing. *Critical reviews in biotechnology*. Feb 2018;38(1):106-121. doi:10.1080/07388551.2017.1312271
316. Bayer B, Duerkop M, Pörtner R, Möller J. Comparison of mechanistic and hybrid modeling approaches for characterization of a CHO cultivation process: Requirements, pitfalls and solution paths. *Biotechnology journal*. 2023;18(1):2200381. doi:https://doi.org/10.1002/biot.202200381
317. Martínez VS, Dietmair S, Quek L-E, Hodson MP, Gray P, Nielsen LK. Flux balance analysis of CHO cells before and after a metabolic switch from lactate production to consumption. *Biotechnology and bioengineering*. 2013;110(2):660-666. doi:https://doi.org/10.1002/bit.24728
318. Tang P, Xu J, Louey A, et al. Kinetic modeling of Chinese hamster ovary cell culture: factors and principles. *Critical reviews in biotechnology*. 2020;40(2):265-281.
319. Kyriakopoulos S, Ang KS, Lakshmanan M, et al. Kinetic Modeling of Mammalian Cell Culture Bioprocessing: The Quest to Advance Biomanufacturing. *Biotechnology journal*. 2018;13(3):1700229. doi:https://doi.org/10.1002/biot.201700229
320. Mahanty B. Hybrid modeling in bioprocess dynamics: Structural variabilities, implementation strategies, and practical challenges. *Biotechnology and bioengineering*. Aug 2023;120(8):2072-2091. doi:10.1002/bit.28503

321. Schmidberger T, Posch C, Sasse A, Gülch C, Huber R. Progress toward forecasting product quality and quantity of mammalian cell culture processes by performance-based modeling. *Biotechnology progress*. 2015;31(4):1119-1127. doi:https://doi.org/10.1002/btpr.2105
322. Mondal PP, Galodha A, Verma VK, et al. Review on machine learning-based bioprocess optimization, monitoring, and control systems. *Bioresource Technology*. 2023/02/01/ 2023;370:128523. doi:https://doi.org/10.1016/j.biortech.2022.128523
323. Rathore AS, Nikita S, Thakur G, Mishra S. Artificial intelligence and machine learning applications in biopharmaceutical manufacturing. *Trends in Biotechnology*. 2023/04/01/ 2023;41(4):497-510. doi:https://doi.org/10.1016/j.tibtech.2022.08.007
324. Kotidis P, Kontoravdi C. Harnessing the potential of artificial neural networks for predicting protein glycosylation. *Metab Eng Commun*. Jun 2020;10:e00131. doi:10.1016/j.mec.2020.e00131
325. Baako T-MD, Kulkarni SK, McClendon JL, Harcum SW, Gilmore J. Machine Learning and Deep Learning Strategies for Chinese Hamster Ovary Cell Bioprocess Optimization. *Fermentation*. 2024;10(5):234.
326. Walsh I, Myint M, Nguyen-Khuong T, Ho YS, Ng SK, Lakshmanan M. Harnessing the potential of machine learning for advancing "quality by design" in biomanufacturing. *MAbs*. 2022;14(1):2013593.
327. Poth M, Magill G, Filgertshofer A, Popp O, Großkopf T. Extensive evaluation of machine learning models and data preprocessings for Raman modeling in bioprocessing. *Journal of Raman Spectroscopy*. 2022;53(9):1580-1591. doi:https://doi.org/10.1002/jrs.6402
328. Park S-Y, Kim S-J, Park C-H, Kim J, Lee D-Y. Data-driven prediction models for forecasting multistep ahead profiles of mammalian cell culture toward bioprocess digital twins. *Biotechnology and bioengineering*. 2023;120(9):2494-2508. doi:https://doi.org/10.1002/bit.28405
329. Wong WC, Chee E, Li J, Wang X. Recurrent neural network-based model predictive control for continuous pharmaceutical manufacturing. *Mathematics*. 2018;6(11):242.
330. Smiatek J, Clemens C, Herrera LM, et al. Generic and specific recurrent neural network models: Applications for large and small scale biopharmaceutical upstream processes. *Biotechnology Reports*. 2021;31:e00640.

331. Rogers AW, Song Z, Ramon FV, Jing K, Zhang D. Investigating 'greyness' of hybrid model for bioprocess predictive modelling. *Biochemical Engineering Journal*. 2023/01/01/ 2023;190:108761. doi:https://doi.org/10.1016/j.bej.2022.108761
332. Mowbray MR, Wu C, Rogers AW, Rio-Chanona EAD, Zhang D. A reinforcement learning-based hybrid modeling framework for bioprocess kinetics identification. *Biotechnology and bioengineering*. 2023;120(1):154-168. doi:https://doi.org/10.1002/bit.28262
333. Cui T, Bertalan T, Ndahiro N, et al. Data-driven and physics informed modelling of Chinese Hamster Ovary cell bioreactors. *Computers & Chemical Engineering*. 2024/01/18/ 2024:108594. doi:https://doi.org/10.1016/j.compchemeng.2024.108594
334. Pinto J, Mestre M, Ramos J, Costa RS, Striedner G, Oliveira R. A general deep hybrid model for bioreactor systems: Combining first principles with deep neural networks. *Computers & Chemical Engineering*. 2022/09/01/ 2022;165:107952.
doi:https://doi.org/10.1016/j.compchemeng.2022.107952
335. Agharafeie R, Oliveira R, Ramos JRC, Mendes JM. Application of hybrid neural models to bioprocesses: A systematic literature review. *Authorea Preprints*. 2023;
336. Maton M, Bogaerts P, Vande Wouwer A. Hybrid Dynamic Models of Bioprocesses Based on Elementary Flux Modes and Multilayer Perceptrons. *Processes*. 2022;10(10):2084.
337. Narayanan H, von Stosch M, Feidl F, Sokolov M, Morbidelli M, Butté A. Hybrid modeling for biopharmaceutical processes: advantages, opportunities, and implementation. *Frontiers in Chemical Engineering*. 2023;5:1157889.
338. Yatipanthalawa BS, Fitzsimons SEW, Horning T, Lee YY, Gras SL. Development and validation of a hybrid model for prediction of viable cell density, titer and cumulative glucose consumption in a mammalian cell culture system. *Computers & Chemical Engineering*. 2024:108648.
339. Solle D, Hitzmann B, Herwig C, et al. Between the Poles of Data-Driven and Mechanistic Modeling for Process Operation. *Chemie Ingenieur Technik*. 2017;89(5):542-561. doi:https://doi.org/10.1002/cite.201600175
340. Simutis R, Lübbert A. Hybrid approach to state estimation for bioprocess control. *Bioengineering*. 2017;4(1):21.

341. Iglesias Jr CF, Xu X, Mehta V, et al. Monitoring the Recombinant Adeno-Associated Virus Production using Extended Kalman Filter. *Processes*. 2022;10(11):2180.
342. Narayanan H, Behle L, Luna MF, et al. Hybrid‐ekf: Hybrid model coupled with extended Kalman filter for real‐time monitoring and control of mammalian cell culture. *Biotechnology and bioengineering*. 2020;117(9):2703-2714.
343. Tsopanoglou A, del Val IJ. Moving towards an era of hybrid modelling: advantages and challenges of coupling mechanistic and data-driven models for upstream pharmaceutical bioprocesses. *Current Opinion in Chemical Engineering*. 2021;32:100691.
344. Narayanan H, Sokolov M, Morbidelli M, Butté A. A new generation of predictive models: The added value of hybrid models for manufacturing processes of therapeutic proteins. *Biotechnology and bioengineering*. 2019;116(10):2540-2549.
345. Ohadi K, Legge RL, Budman HM. Development of a soft‐sensor based on multi‐wavelength fluorescence spectroscopy and a dynamic metabolic model for monitoring mammalian cell cultures. *Biotechnology and bioengineering*. 2015;112(1):197-208.
346. Polak J, Huang Z, Sokolov M, et al. An innovative hybrid modeling approach for simultaneous prediction of cell culture process dynamics and product quality. *Biotechnology journal*. 2024;19(3):2300473.
347. Iglesias Jr CF, Ristovski M, Bolic M, Cuperlovic-Culf M. rAAV Manufacturing: The Challenges of Soft Sensing during Upstream Processing. *Bioengineering*. 2023;10(2):229.
348. Reyes SJ, Lemire L, Molina RS, et al. Multivariate data analysis of process parameters affecting the growth and productivity of stable Chinese hamster ovary cell pools expressing SARS‐CoV‐2 spike protein as vaccine antigen in early process development. *Biotechnology progress*. 2024:e3467.
349. DiPietro R, Hager GD. Chapter 21 - Deep learning: RNNs and LSTM. In: Zhou SK, Rueckert D, Fichtinger G, eds. *Handbook of Medical Image Computing and Computer Assisted Intervention*. Academic Press; 2020:503-519.
350. Graves A, Graves A. *Supervised sequence labelling*. Springer; 2012.
351. Goodfellow I, Bengio Y, Courville A. *Deep learning*. MIT press; 2016.
352. McKinney W. Data structures for statistical computing in python. Austin, TX; 2010:51-56.

353. Harris CR, Millman KJ, Van Der Walt SJ, et al. Array programming with NumPy. *Nature*. 2020;585(7825):357-362.
354. Paszke A, Gross S, Massa F, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*. 2019;32
355. Jiang Z, Sharfstein ST. Sodium butyrate stimulates monoclonal antibody over-expression in CHO cells by improving gene accessibility. *Biotechnology and bioengineering*. May 1 2008;100(1):189-94. doi:10.1002/bit.21726
356. Park JH, Noh SM, Woo JR, Kim JW, Lee GM. Valeric acid induces cell cycle arrest at G1 phase in CHO cell cultures and improves recombinant antibody productivity. *Biotechnology journal*. Mar 2016;11(4):487-96. doi:10.1002/biot.201500327
357. An HJ, Froehlich JW, Lebrilla CB. Determination of glycosylation sites and site-specific heterogeneity in glycoproteins. *Current opinion in chemical biology*. Oct 2009;13(4):421-6. doi:10.1016/j.cbpa.2009.07.022
358. Edwards E, Livanos M, Krueger A, et al. Strategies to control therapeutic antibody glycosylation during bioprocessing: Synthesis and separation. *Biotechnology and bioengineering*. Jun 2022;119(6):1343-1358. doi:10.1002/bit.28066
359. Green A, Glassey J. Multivariate analysis of the effect of operating conditions on hybridoma cell metabolism and glycosylation of produced antibody. *Journal of Chemical Technology & Biotechnology*. 2015;90(2):303-313. doi:https://doi.org/10.1002/jctb.4481
360. Ivarsson M, Villiger TK, Morbidelli M, Soos M. Evaluating the impact of cell culture process parameters on monoclonal antibody N-glycosylation. *Journal of biotechnology*. 2014// 2014;188doi:10.1016/j.jbiotec.2014.08.026
361. López-Meza J, Araíz-Hernández D, Carrillo-Cocom LM, López-Pacheco F, Rocha-Pizaña Mdel R, Alvarez MM. Using simple models to describe the kinetics of growth, glucose consumption, and monoclonal antibody formation in naive and infliximab producer CHO cells. *Cytotechnology*. Aug 2016;68(4):1287-300. doi:10.1007/s10616-015-9889-2
362. Meuwly F, Papp F, Ruffieux PA, Bernard AR, Kadouri A, von Stockar U. Use of glucose consumption rate (GCR) as a tool to monitor and control animal cell production processes in packed-bed bioreactors. *Journal of*

biotechnology. 2006/03/09/ 2006;122(1):122-129. doi:https://doi.org/10.1016/j.jbiotec.2005.08.005

363. Kozma B, Hirsch E, Gergely S, Párta L, Pataki H, Salgó A. On-line prediction of the glucose concentration of CHO cell cultivations by NIR and Raman spectroscopy: Comparative scalability test with a shake flask model system. *Journal of Pharmaceutical and Biomedical Analysis*. 2017/10/25/ 2017;145:346-355. doi:https://doi.org/10.1016/j.jpba.2017.06.070

364. Lederle M, Tric M, Roth T, et al. Continuous optical in-line glucose monitoring and control in CHO cultures contributes to enhanced metabolic efficiency while maintaining darbepoetin alfa product quality. *Biotechnology journal*. Aug 2021;16(8):e2100088. doi:10.1002/biot.202100088

365. Rashedi M, Rafiei M, Demers M, et al. Machine learning-based model predictive controller design for cell culture processes. 2023;120(8):2144-2159. doi:https://doi.org/10.1002/bit.28486

366. Gibbons L, Maslanka F, Le N, et al. An assessment of the impact of Raman based glucose feedback control on CHO cell bioreactor process development. *Biotechnology progress*. 2023;39(5):e3371. doi:https://doi.org/10.1002/btpr.3371

367. Templeton N, Dean J, Reddy P, Young JD. Peak antibody production is associated with increased oxidative metabolism in an industrially relevant fed-batch CHO cell culture. *Biotechnology and bioengineering*. 2013;110(7):2013-2024. doi:https://doi.org/10.1002/bit.24858

368. Moore B, Sanford R, Zhang A. Case study: The characterization and implementation of dielectric spectroscopy (biocapacitance) for process control in a commercial GMP CHO manufacturing process. *Biotechnology progress*. May 2019;35(3):e2782. doi:10.1002/btpr.2782

369. Wang X, Zhou G, Liang L, et al. Deep learning-based image analysis for in situ microscopic imaging of cell culture process. 2024;129:107621.

370. Xu X, Farnós O, Paes BC, Nesdoly S, Kamen AA. Multivariate data analysis on multisensor measurement for inline process monitoring of adenovirus production in HEK293 cells. *Biotechnology Bioengineering*. 2024;

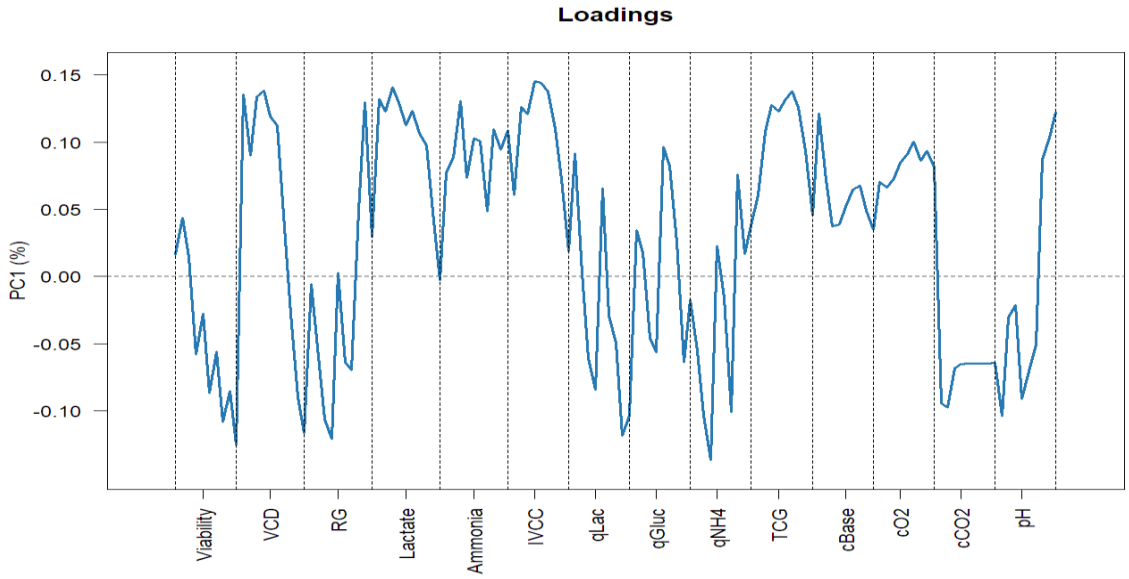
371. Lalonde M-E, Durocher Y. Therapeutic glycoprotein production in mammalian cells. *Journal of biotechnology*. 2017/06/10/ 2017;251:128-140. doi:https://doi.org/10.1016/j.jbiotec.2017.04.028

372. Zhu J. Mammalian cell protein expression for biopharmaceutical production. *Biotechnology Advances*. 2012/09/01/ 2012;30(5):1158-1170. doi:https://doi.org/10.1016/j.biotechadv.2011.08.022

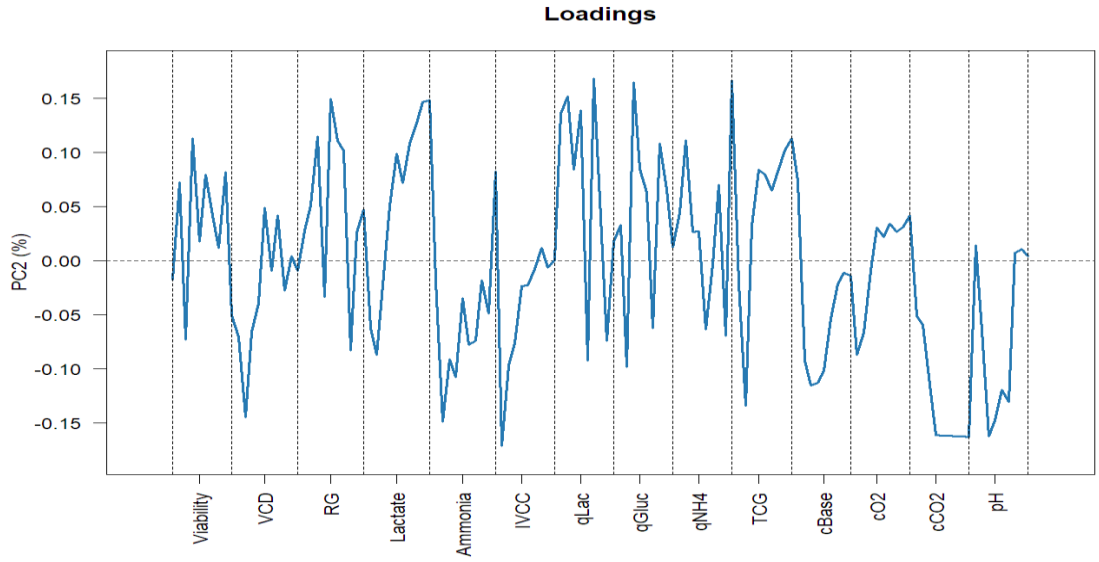
373. Lai T, Yang Y, Ng SKJP. Advances in mammalian cell line development technologies for recombinant protein production. 2013;6(5):579-603.
374. Goh JB, Ng SKJCRiB. Impact of host cell line choice on glycan profile. 2018;38(6):851-867.
375. Xu S, Hoshan L, Jiang R, et al. A practical approach in bioreactor scale-up and process transfer using a combination of constant P/V and vvm as the criterion. 2017;33(4):1146-1159. doi:https://doi.org/10.1002/btpr.2489
376. Amer M, Feng Y, Ramsey JD. Using CFD simulations and statistical analysis to correlate oxygen mass transfer coefficient to both geometrical parameters and operating conditions in a stirred-tank bioreactor. 2019;35(3):e2785. doi:https://doi.org/10.1002/btpr.2785
377. Green A, Glassey J. Multivariate analysis of the effect of operating conditions on hybridoma cell metabolism and glycosylation of produced antibody. 2015;90(2):303-313. doi:https://doi.org/10.1002/jctb.4481
378. Kumar J, Chauhan AS, Shah RL, Gupta JA, Rathore AS. Amino acid supplementation for enhancing recombinant protein production in E. coli. *Biotechnology and bioengineering*. Aug 2020;117(8):2420-2433. doi:10.1002/bit.27371
379. Powers DN, Wang Y, Fratz-Berilla EJ, et al. Real-time quantification and supplementation of bioreactor amino acids to prolong culture time and maintain antibody product quality. *Biotechnology progress*. Nov 2019;35(6):e2894. doi:10.1002/btpr.2894
380. Domján J, Pantea E, Gyürkés M, et al. Real-time amino acid and glucose monitoring system for the automatic control of nutrient feeding in CHO cell culture using Raman spectroscopy. 2022;17(5):2100395. doi:https://doi.org/10.1002/biot.202100395
381. Lecina M, Soley A, Gràcia J, et al. Application of on-line OUR measurements to detect actions points to improve baculovirus-insect cell cultures in bioreactors. *Journal of biotechnology*. Sep 18 2006;125(3):385-94. doi:10.1016/j.jbiotec.2006.03.014
382. Lecina M, Comas P, Martínez-Monge I, Cairó JJ. Monitoring Tools for the Development of High Cell Density Culture Strategies. In: Pörtner R, ed. *Cell Culture Engineering and Technology: In appreciation to Professor Mohamed Al-Rubeai*. Springer International Publishing; 2021:485-510.
383. Ihling N, Munkler LP, Paul R, et al. Non-invasive and time-resolved measurement of the respiration activity of Chinese hamster ovary cells

enables prediction of key culture parameters in shake flasks. 2022;17(8):2100677. doi:https://doi.org/10.1002/biot.202100677
384. Deshpande RR, Heinzle E. On-line oxygen uptake rate and culture viability measurement of animal cell culture using microplates with integrated oxygen sensors. *Biotechnology letters*. May 2004;26(9):763-7. doi:10.1023/b:bile.0000024101.57683.6d
385. Amribt Z, Dewasme L, Wouwer AV, Bogaerts P. Parameter Identification for State Estimation: Design of an Extended Kalman Filter for Hybridoma Cell Fed-Batch Cultures. *IFAC Proceedings Volumes*. 2014/01/01/ 2014;47(3):1170-1175. doi:https://doi.org/10.3182/20140824-6-ZA-1003.01185
386. Fernandes S, Richelle A, Amribt Z, Dewasme L, Bogaerts P, Wouwer AV. Extended and Unscented Kalman Filter design for hybridoma cell fed-batch and continuous cultures. *IFAC-PapersOnLine*. 2015/01/01/ 2015;48(8):1108-1113. doi:https://doi.org/10.1016/j.ifacol.2015.09.116
387. Dewasme L, Fernandes S, Amribt Z, Santos LO, Bogaerts P, Vande Wouwer A. State estimation and predictive control of fed-batch cultures of hybridoma cells. *Journal of Process Control*. 2015/06/01/ 2015;30:50-57. doi:https://doi.org/10.1016/j.jprocont.2014.12.006
388. Soons ZITA, Shi J, Stigter JD, van der Pol LA, van Straten G, van Boxtel AJB. Observer design and tuning for biomass growth and kLa using online and offline measurements. *Journal of Process Control*. 2008/08/01/ 2008;18(7):621-631. doi:https://doi.org/10.1016/j.jprocont.2007.12.008
389. Survyla A, Levisauskas D, Urniezius R, Simutis R. An oxygen-uptake-rate-based estimator of the specific growth rate in Escherichia coli BL21 strains cultivation processes. *Computational and Structural Biotechnology Journal*. 2021/01/01/ 2021;19:5856-5863. doi:https://doi.org/10.1016/j.csbj.2021.10.015

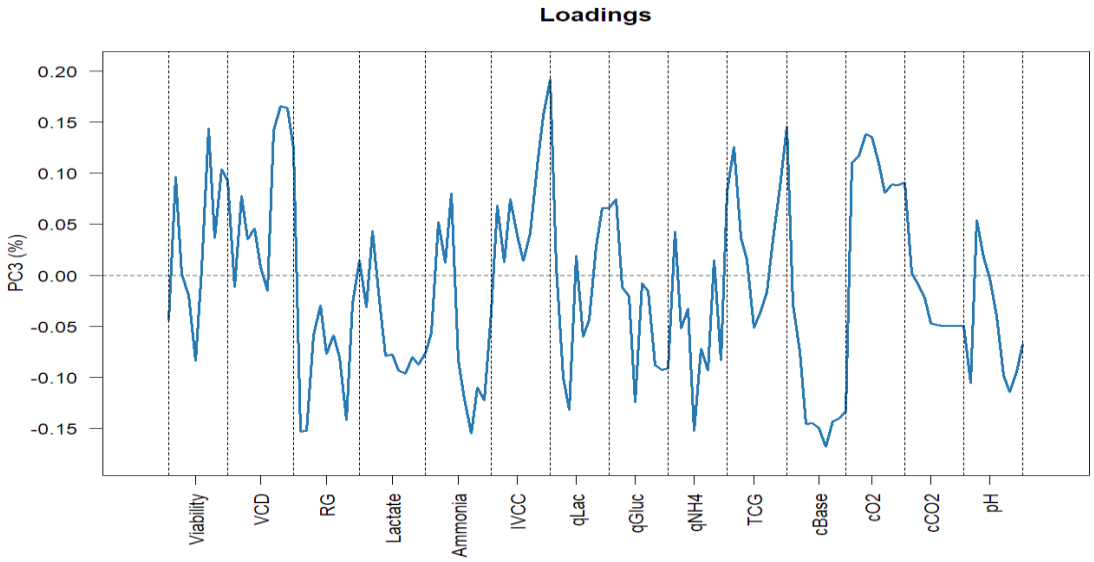
APPENDICES



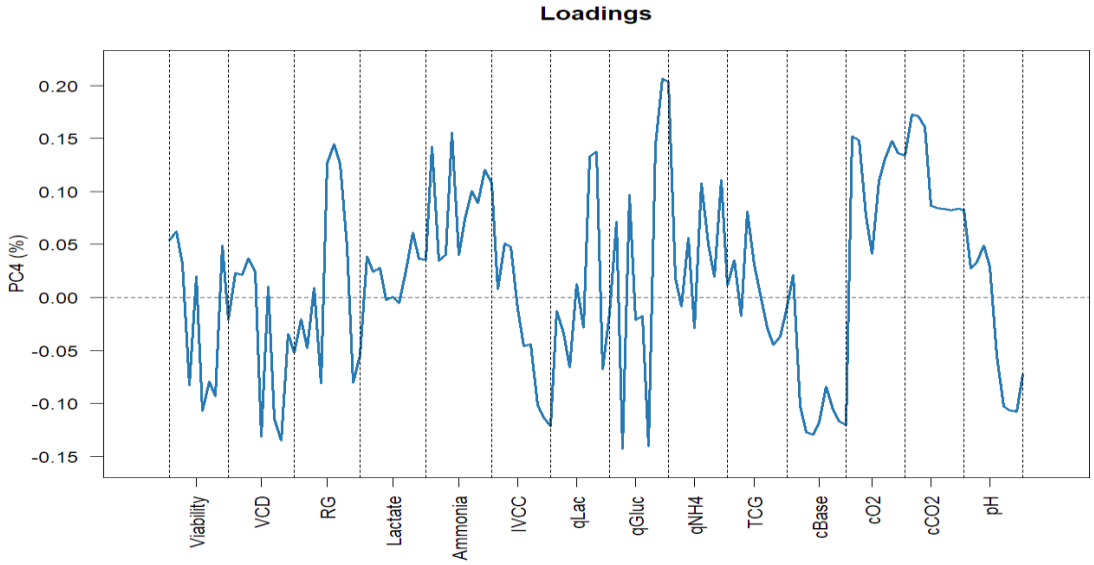
Supplement Figure 1. Multifors 0.75L Wu pool loadings for PC1.



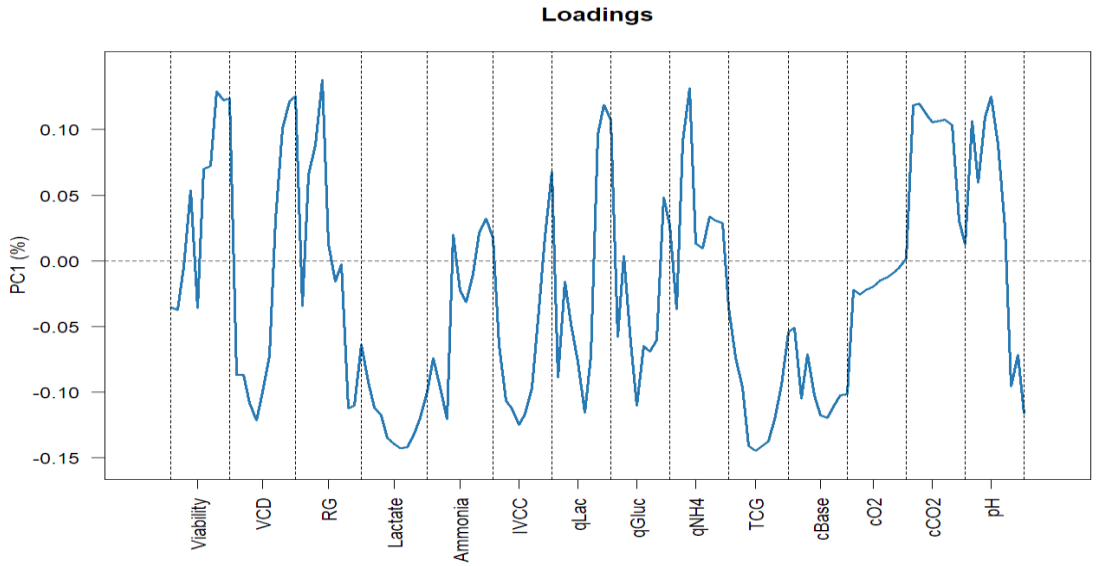
Supplement Figure 2. Multifors 0.75L Wu pool loadings for PC2.



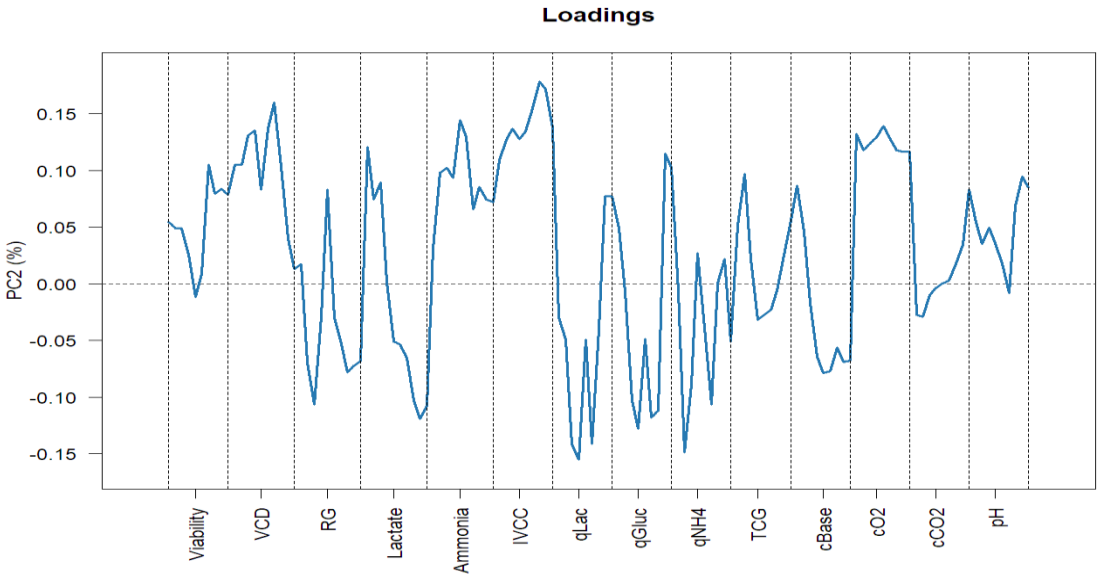
Supplement Figure 3. Multifors 0.75L Wu pool loadings for PC3.



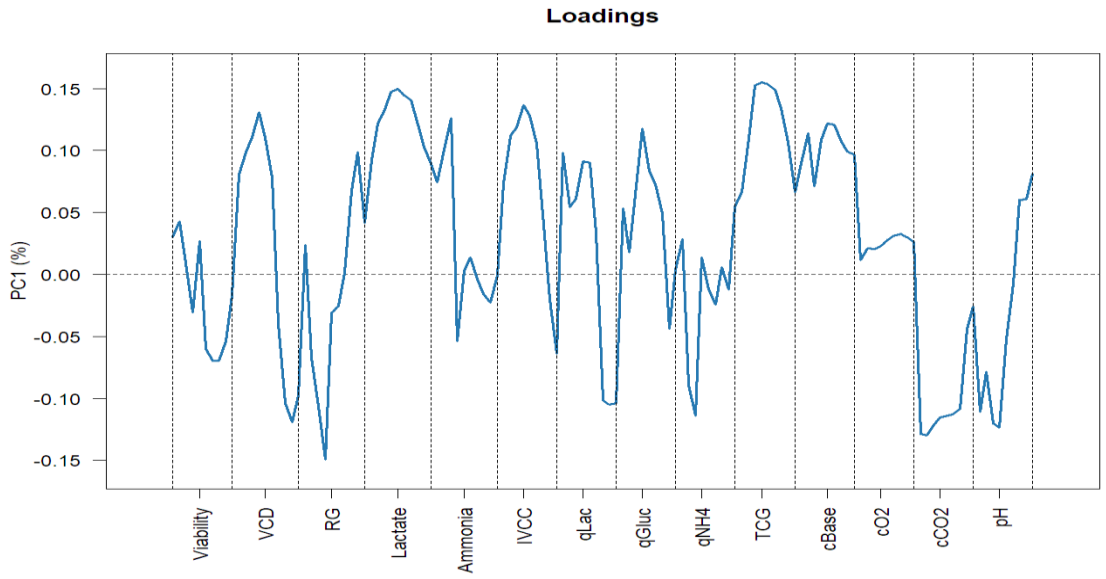
Supplement Figure 4. Multifors 0.75L Wu pool loadings for PC4



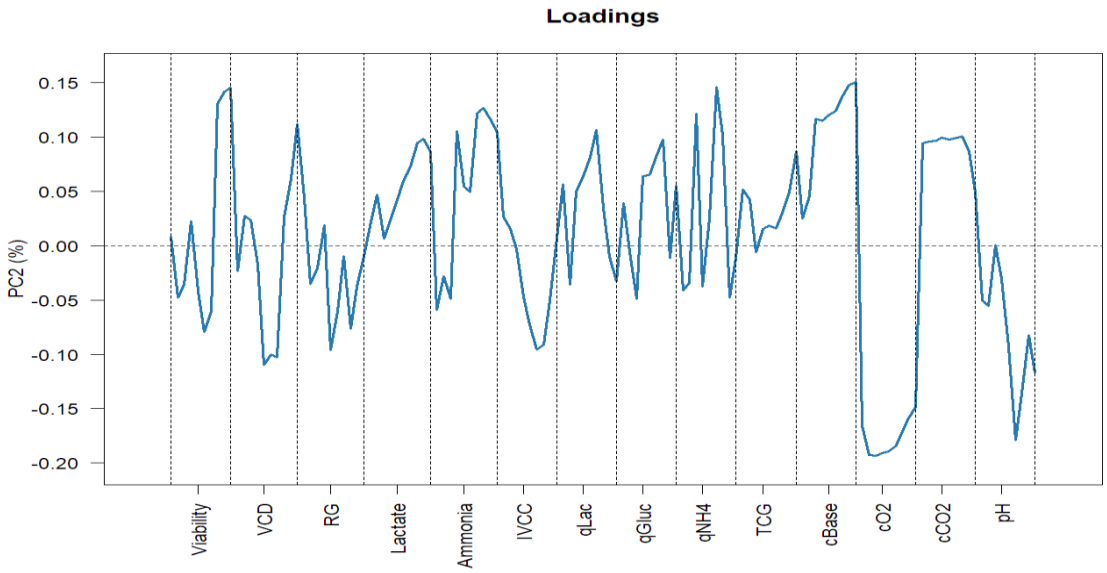
Supplement Figure 5. Multifors 0.75L and BioFlo 1L loadings for PC1.



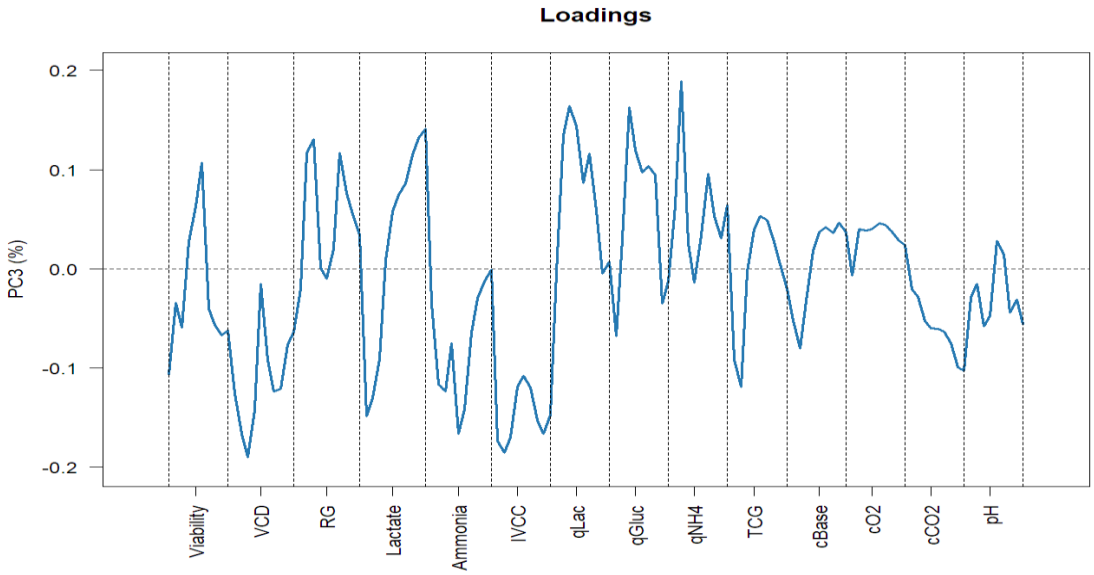
Supplement Figure 6. Multifors 0.75L and BioFlo 1L loadings for PC2.



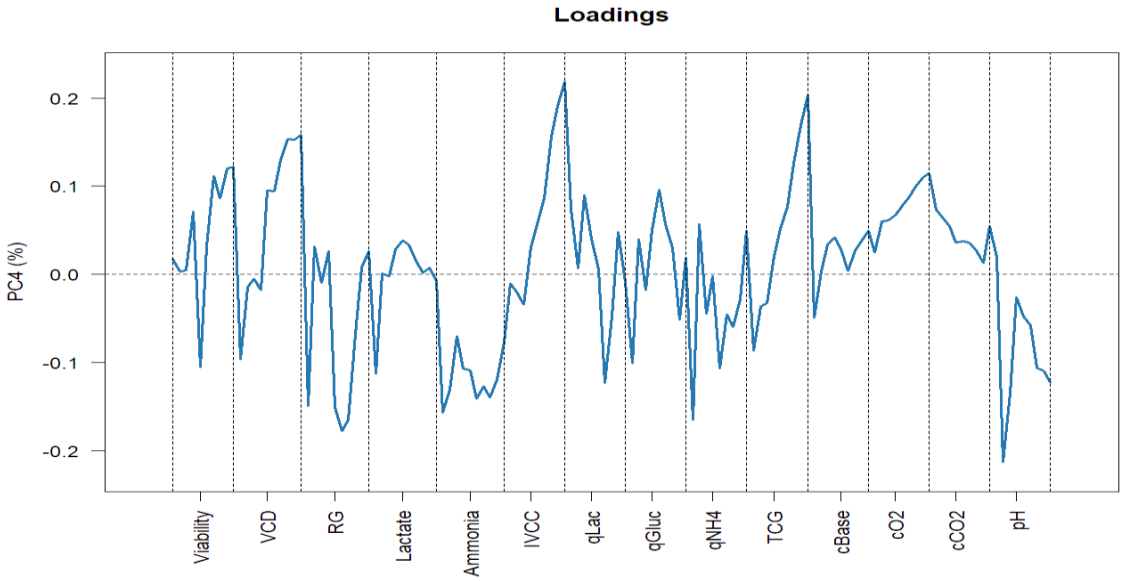
Supplement Figure 7. Multifors 0.75L, BioFlo 1L, and BioFlo 10L loadings for PC1.



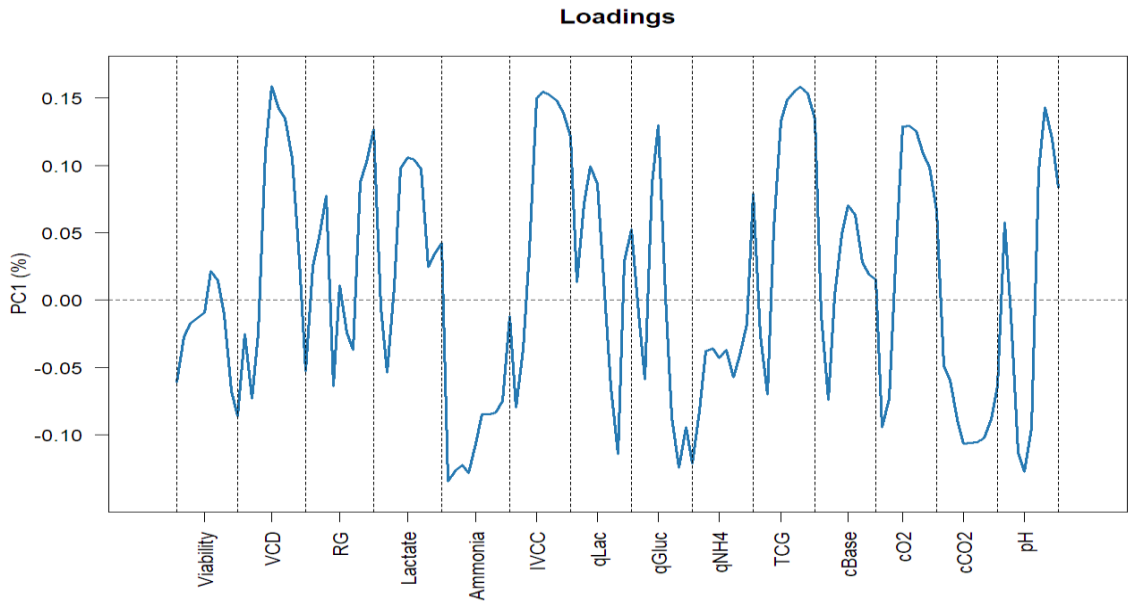
Supplement Figure 8. Multifors 0.75L, BioFlo 1L, and BioFlo 10L loadings for PC2.



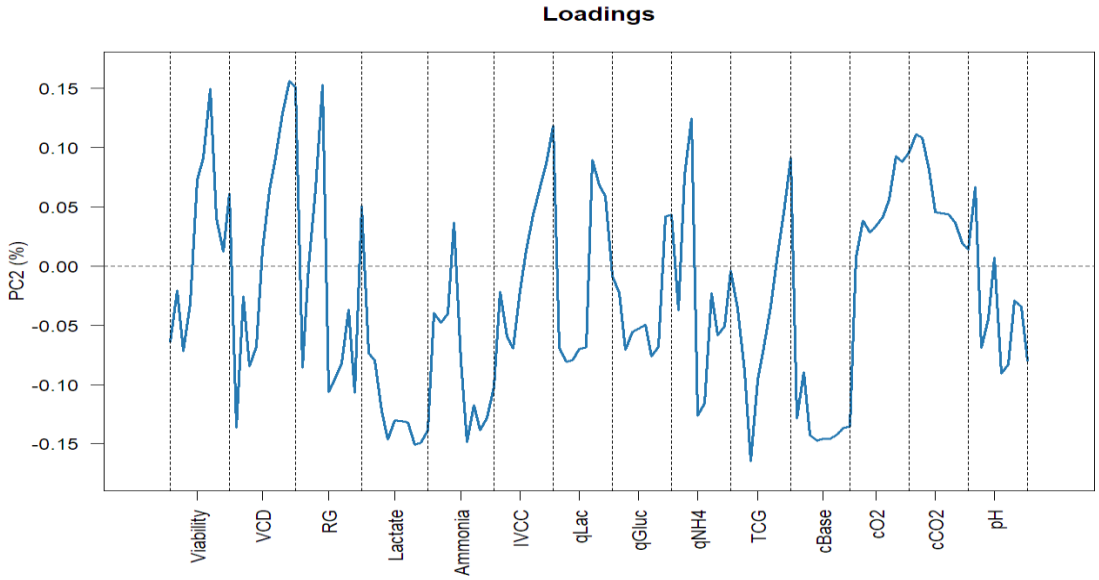
Supplement Figure 9. Multifors 0.75L, BioFlo 1L, and BioFlo 10L loadings for PC3.



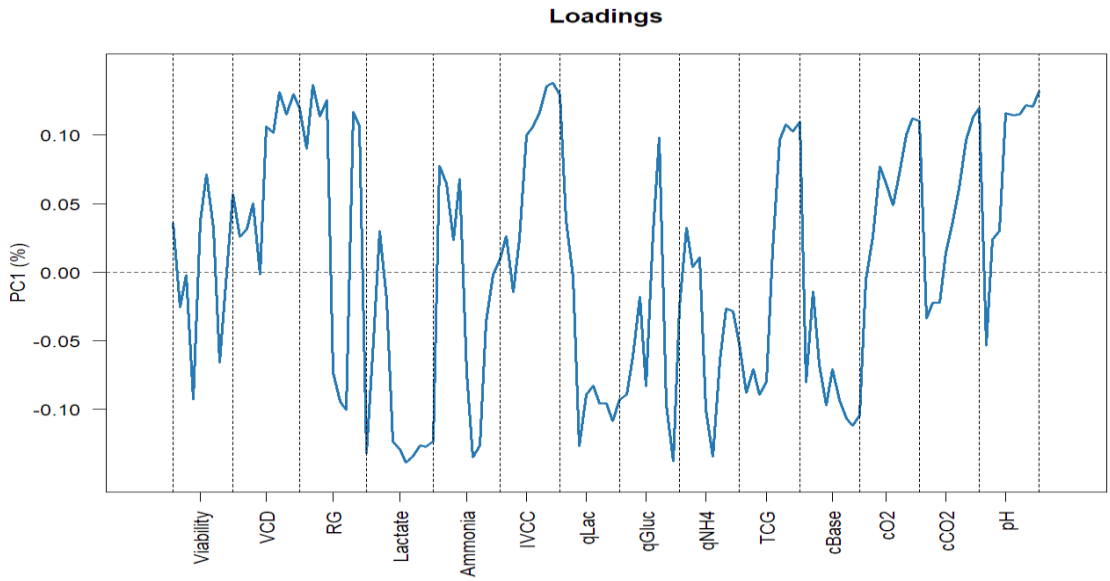
Supplement Figure 10. Multifors 0.75L, BioFlo 1L, and BioFlo 10L loadings for PC4.



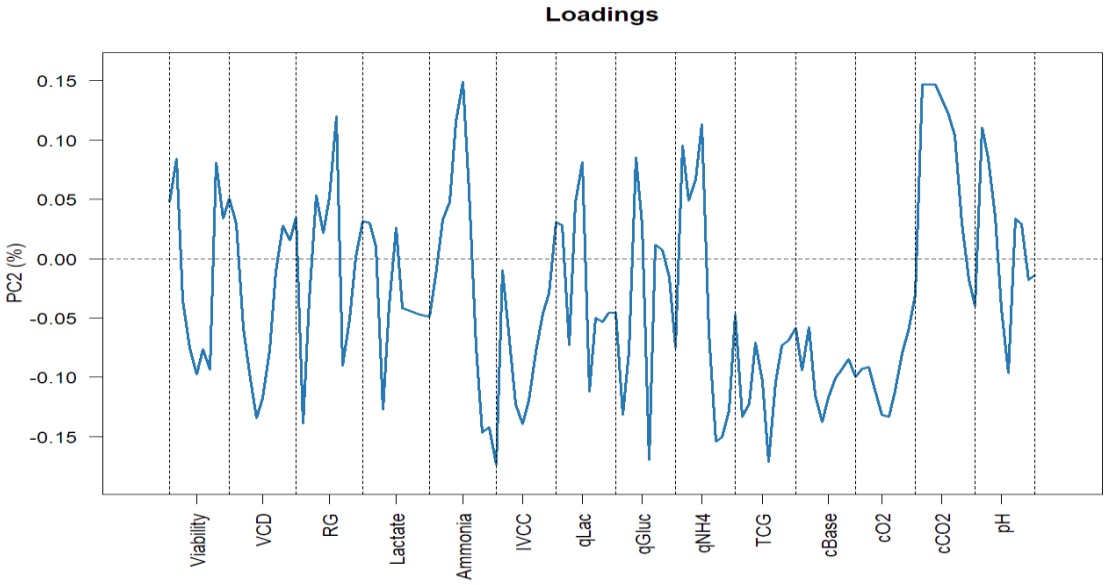
Supplement Figure 11. Wu, Beta, and WuTL pools loadings for PC1.



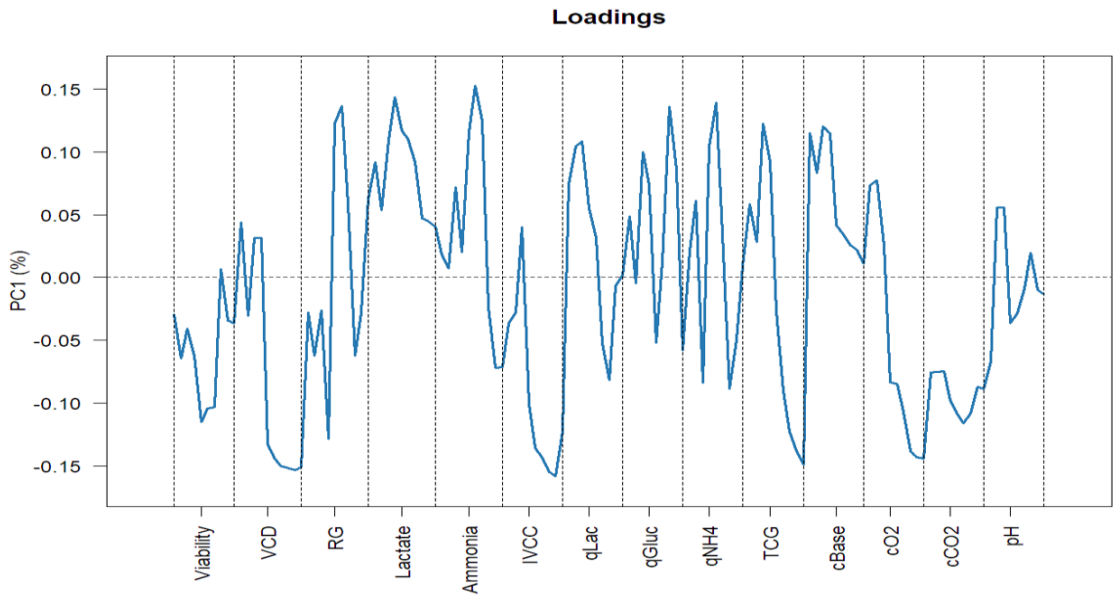
Supplement Figure 12. Wu, Beta, and WuTL pools loadings for PC2.



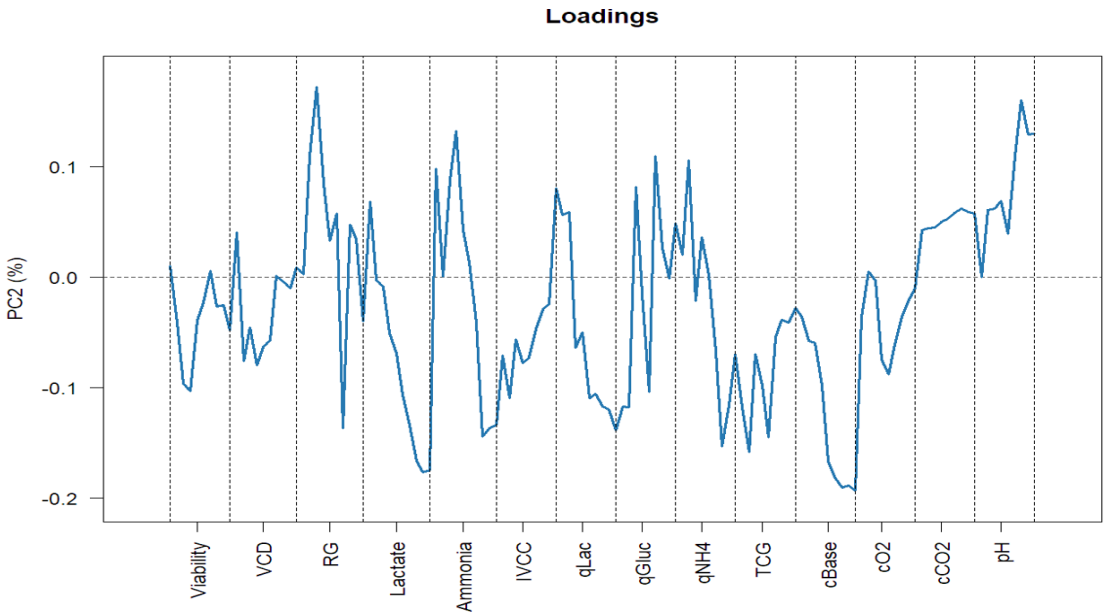
Supplement Figure 13. Delta pool loadings for PC1.



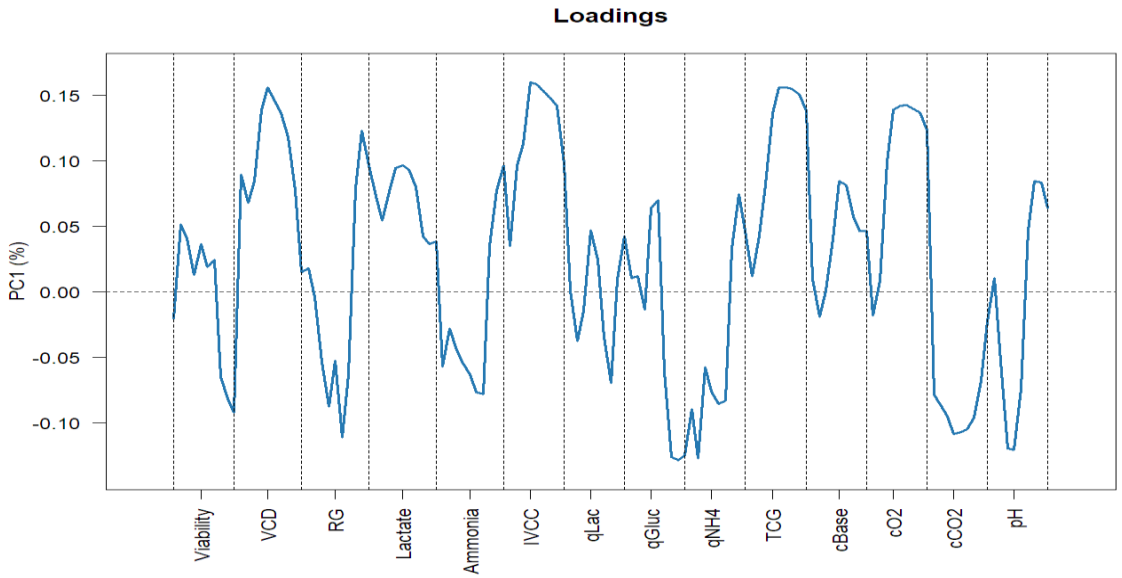
Supplement Figure 14. Delta pool loadings for PC2.



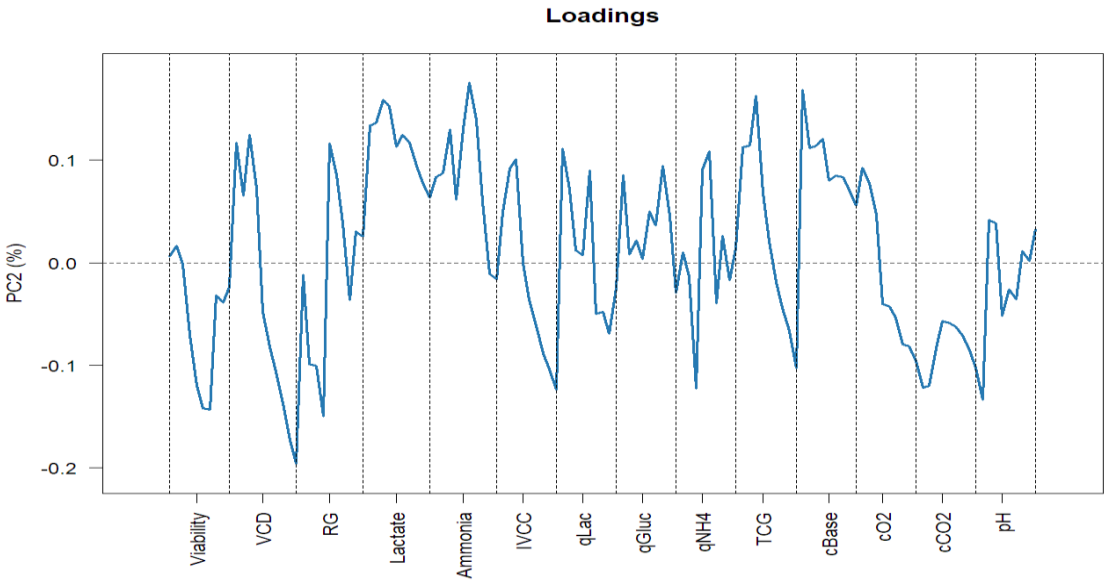
Supplement Figure 15. MSX study loadings for PC1.



Supplement Figure 16. MSX study loadings for PC2.



Supplement Figure 17. All pools loadings for PC1.



Supplement Figure 18. All pools loadings for PC2.

Table Supplement 1. Two-tailed T-test of cell passage's impact at equal seeding densities on various key outputs for Wuhan pool.

Wuhan pool	**Low Density**		**High Density**	
Variable	**P5 vs P8**	**P5 vs P11**	**P5 vs P8**	**P5 vs P11**
Base Volume	0.99	0.65	0.08	0.78
Endpoint Lactate	0.90	0.43	0.91	0.84
Total Oxygen Sparging	0.06	0.58	0.25	0.19
Endpoint Titer	0.07	0.40	0.07	0.72
Total Glucose Consumption	0.12	0.84	0.35	0.13
Endpoint IVCC	0.34	0.51	0.53	0.76
Max Lactate	0.43	0.55	0.94	0.56
Endpoint Ammonia	0.57	0.59	0.84	0.80

Table Supplement 2. Two-tailed T-test of cell passage's impact on different key variables for WuTL and Beta pool.

Values with asterisk * represent conditions in which statistical significance was found (p value < 0.05).

WuTL pool	**P5 vs P8**	**P5 vs P11**
Base Volume	0.19	0.50
Endpoint Lactate	0.07	0.06
Total Oxygen Sparging	0.18	0.51
Endpoint Titer	0.25	0.49
Total Glucose Consumption	0.01*	0.47
Endpoint IVCC	0.13	0.71
Max Lactate	0.07	0.12
Endpoint Ammonia	0.66	0.53
Beta pool	**P5 vs P8**	**P5 vs P11**
Base Volume	0.26	0.42
Endpoint Lactate	0.04*	0.47
Total Oxygen Sparging	0.12	0.06
Endpoint Titer	0.33	0.22
Total Glucose Consumption	0.14	0.18
Endpoint IVCC	0.05	0.02*
Max Lactate	0.10	0.10
Endpoint Ammonia	0.13	0.17

Table Supplement 3. Two-tailed T-test of MSX impact on different key variables for Delta, Beta, and WuTL pools.

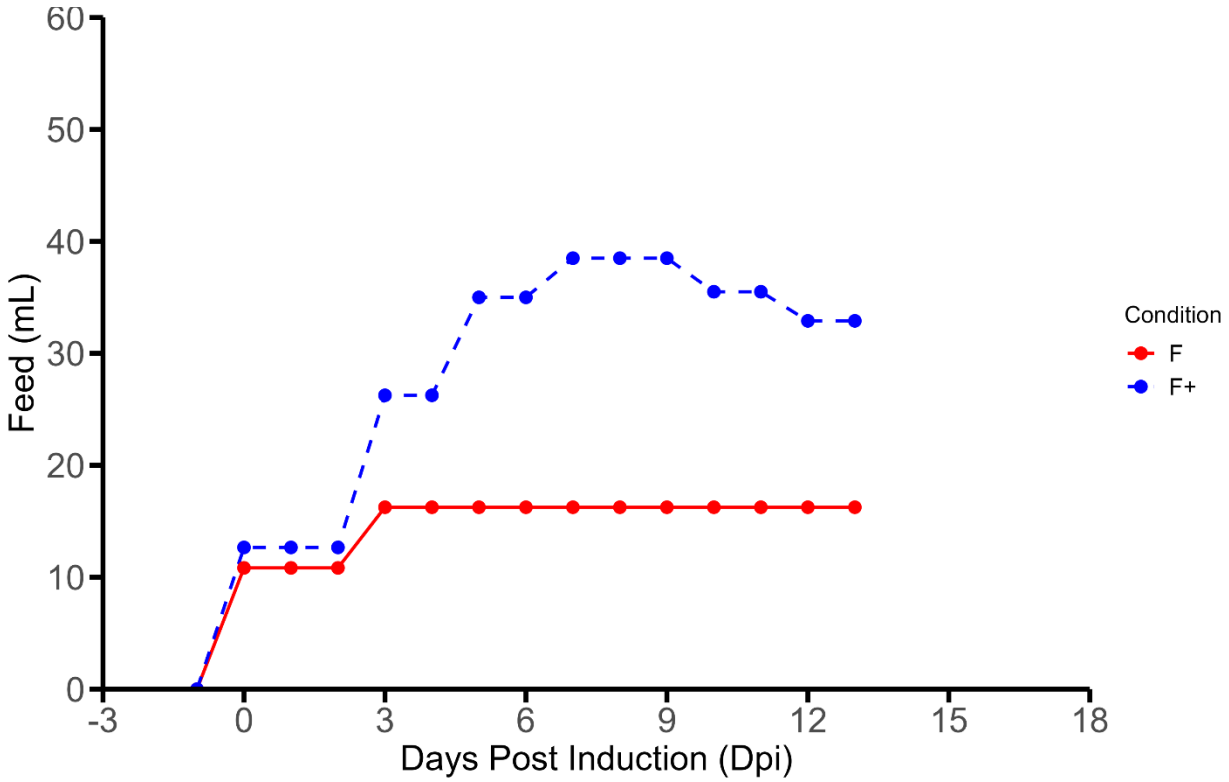
Values with asterisk * represent conditions in which statistical significance was found (p value < 0.05).

Pool	***Delta***	***Beta***	***WuTL***
Base Volume	0.07	0.89	0.37
Endpoint Lactate	0.15	0.93	0.61
Total Oxygen Sparging	0.69	0.74	0.70
Endpoint Titer	0.91	0.63	0.52
Total Glucose Consumption	0.31	0.27	0.46
Endpoint IVCC	0.88	0.64	0.57
Max Lactate	0.18	0.45	0.50
Endpoint Ammonia	0.47	0.77	0.01*

Table Supplement 4. Two-tailed T-test of MSX impact at different passage numbers on different key variables for Delta pool.

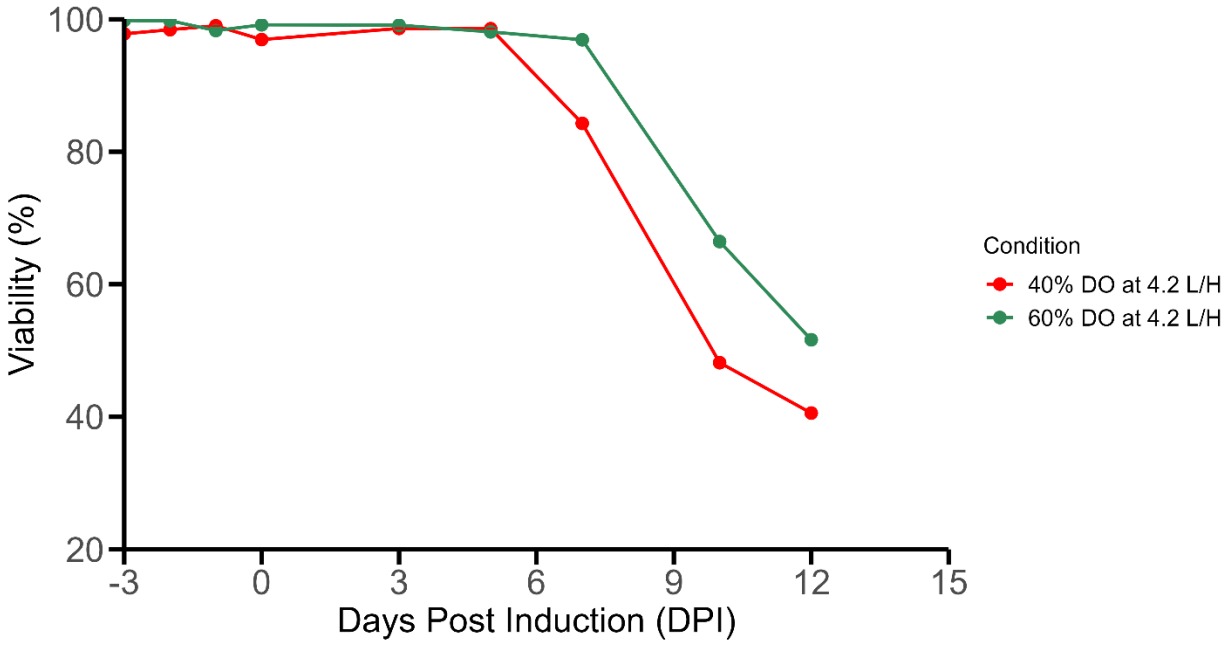
Values with asterisk * represent conditions in which statistical significance was found (p value < 0.05).

MSX level	**MSX+ vs MSX++**		
Delta	**P5**	**P8**	**P11**
Base Volume	0.38	0.78	0.04*
Endpoint Lactate	0.95	0.72	0.14
Total Oxygen Sparging	0.96	0.76	0.81
Endpoint Titer	0.56	0.57	0.79
Total Glucose Consumption	0.76	0.58	0.64
Endpoint IVCC	0.39	0.71	0.83
Max Lactate	0.99	0.66	0.21
Endpoint Ammonia	0.98	0.11	0.36



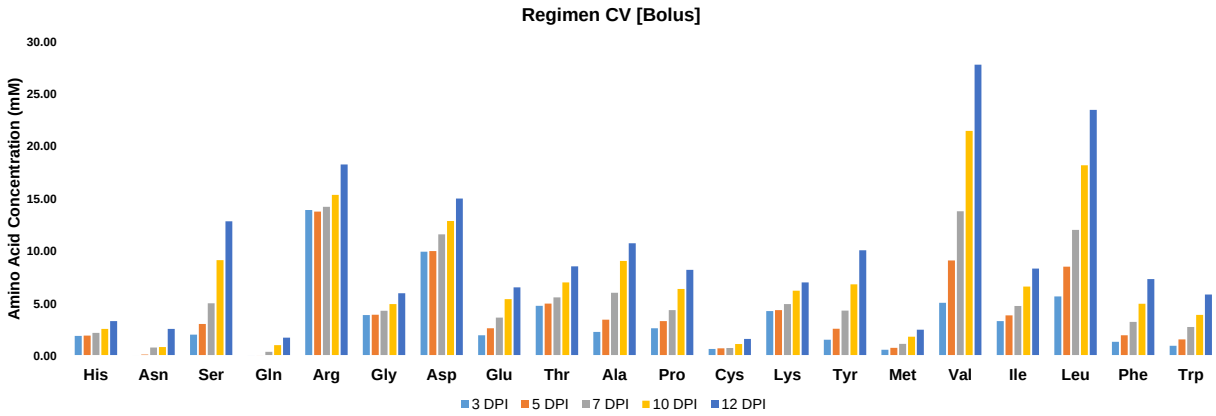
Supplement Figure 19. Daily feed profile of F and F+ regimens.

Fixed feed volume percentages (feed volume/Current medium volume) for F+ were designed to follow the increase and plateau of a cell culture run so as to mimic dynamic cellular kinetics. Conversely the fixed feed volume percentages (feed volume/Initial medium volume) for the F regimen were designed to be kept relatively constant across the production process. Feed additions are divided by the days between feeding events to gauge a representative feed per day plot.

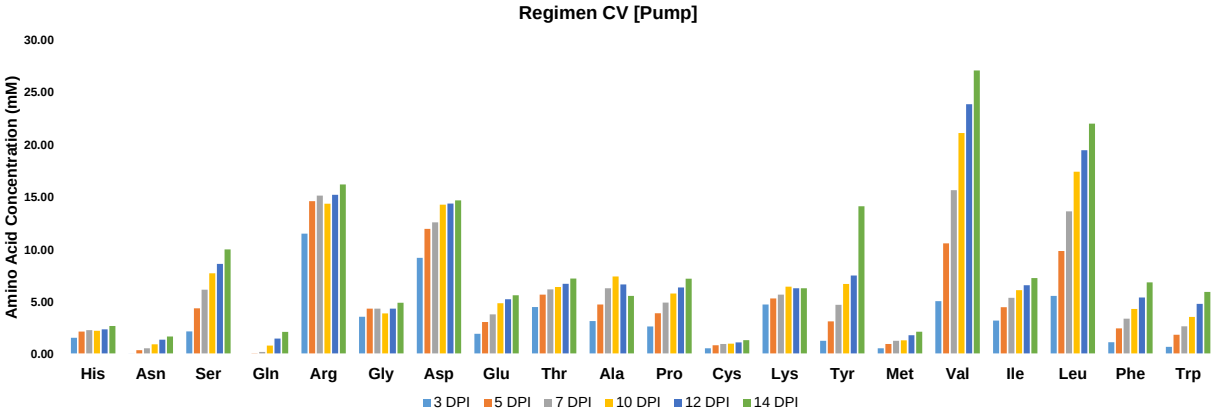


Supplement Figure 20. Viability profile of 40% vs 60% DO at air cap of 4.2 L/H. Increased DO level demonstrates an increase in viability outcomes.

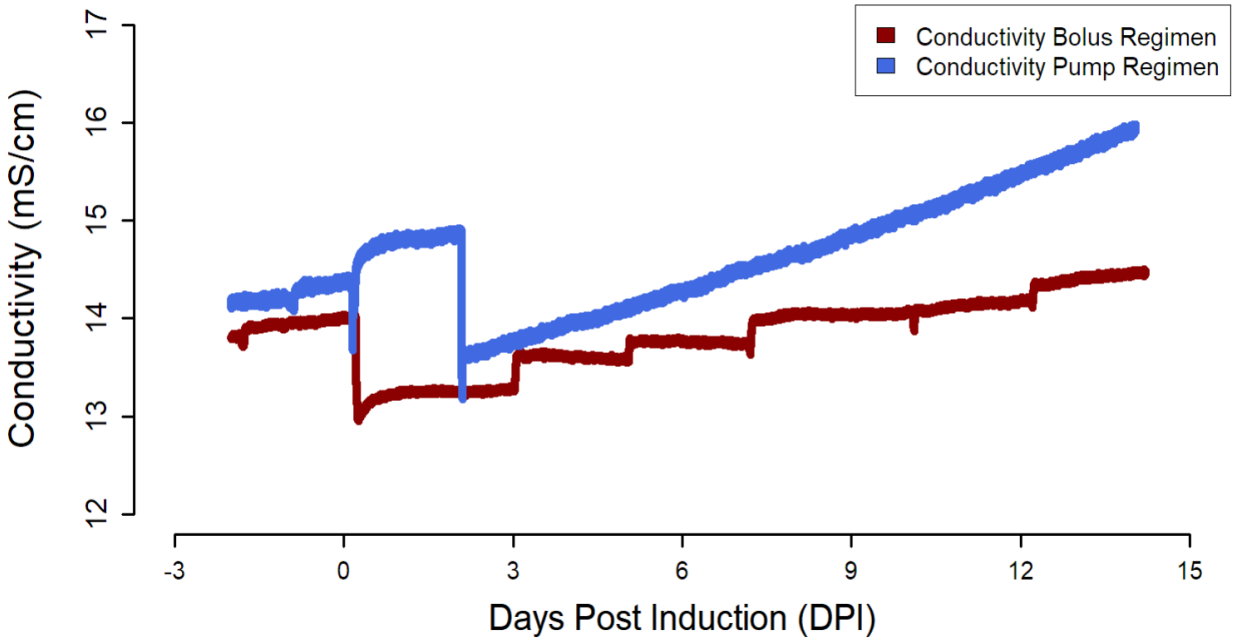
A



B

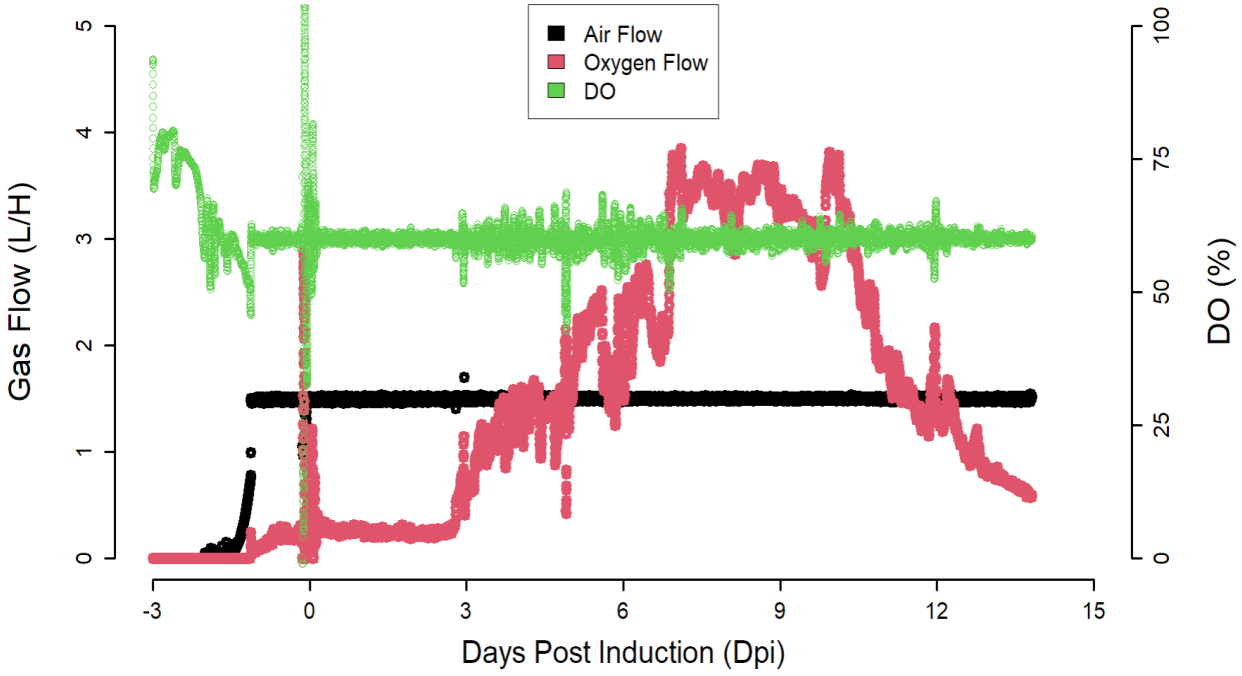


Supplement Figure 21. Amino acid profiles for regimen A) CV [Bolus] and B) CV [Pump]. Days 3, 5, 10, and 14 DPI are plotted. Measurement of amino acid concentrations correspond to sampling before feed additions.



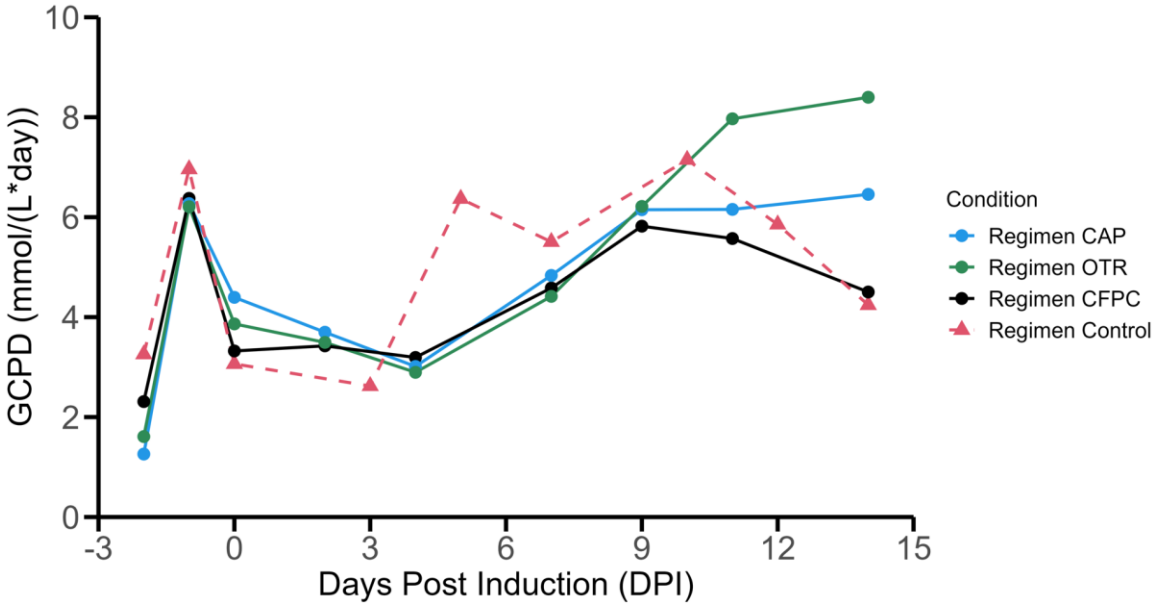
Supplement Figure 22. Conductivity measurements of bolus and slow pump feed additions.

Blue line represents pump feed addition while the maroon line represents bolus feed addition. Maroon measurements taken from control process that was temperature shifted at 0 dpi (as evidenced by the decrease in conductivity at this time point). Blue measurements taken from capacitance fed process that was temperature shifted at 2 dpi (as evidenced by the decrease in conductivity at this time point). For the bolus feed addition (red), each feeding event is accompanied by a step wise increase in conductivity measurements. Conversely, slow bolus additions demonstrate a constant linear increase in conductivity.



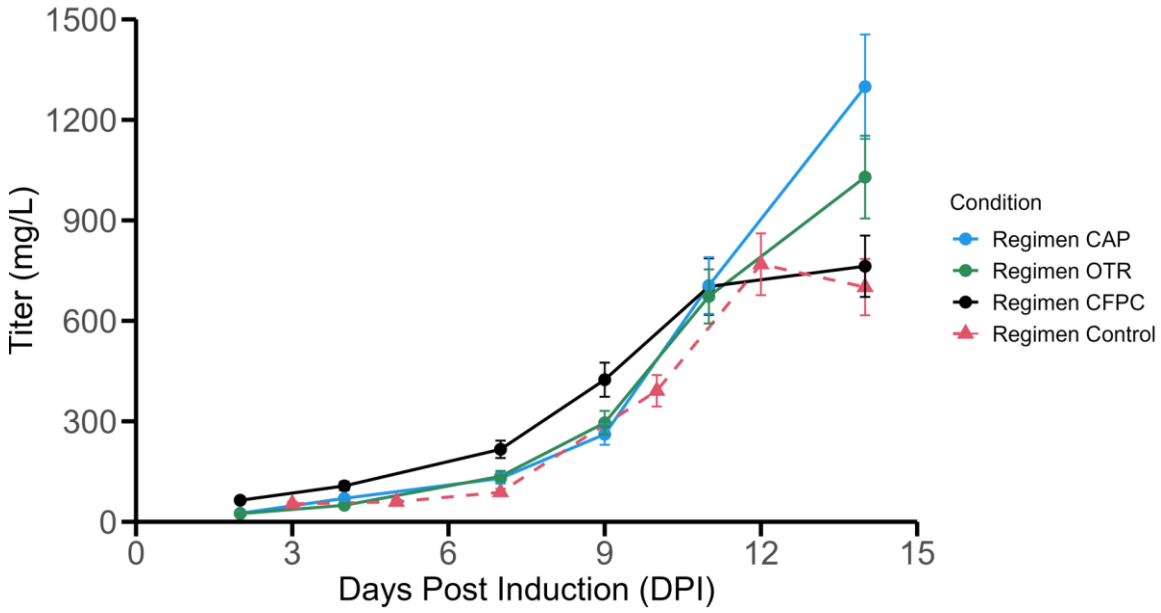
Supplement Figure 23. Gassing profile of a standard regimen culture fed in a bolus fashion.

Green represents the DO %. Black represents the air sparging indicative of the Air Cap. Red represents on demand oxygen sparging once Air Cap has been reached. Air Cap commonly observed to be saturated by -1 dpi and increased oxygen requirements observed to happen in the late production phase of the culture (6-12 dpi).



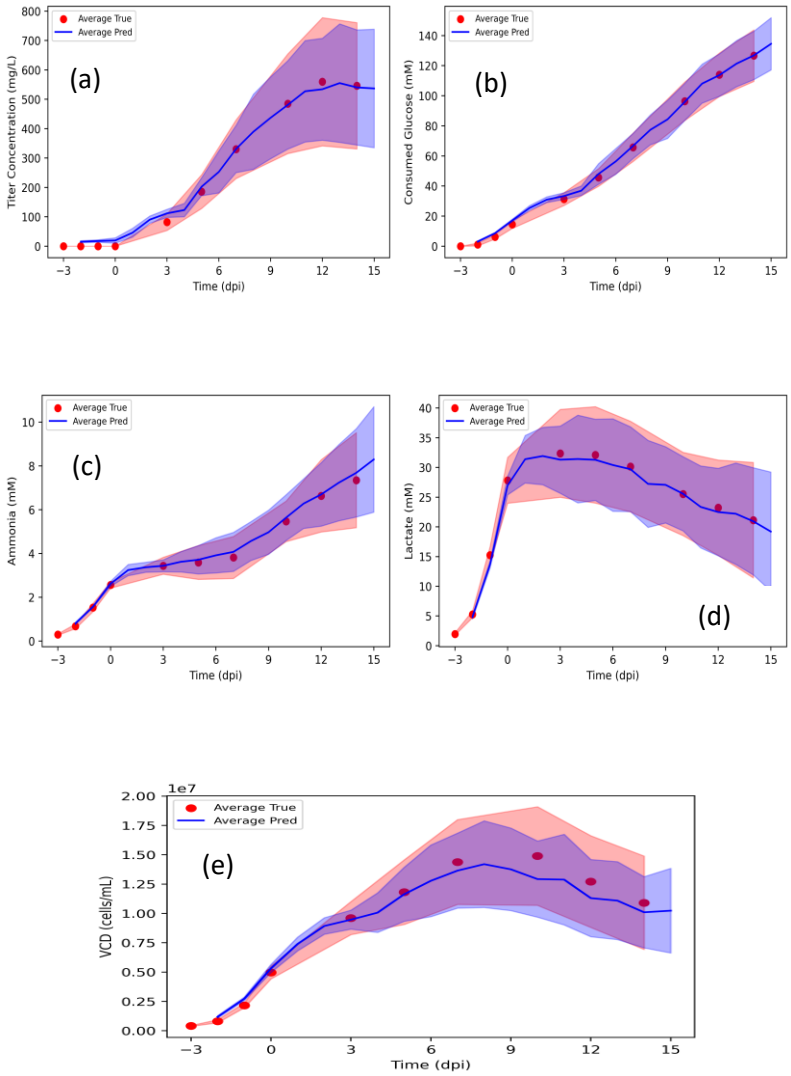
Supplement Figure 24. Glucose consumed per day profile (GCPD) of feeding strategies study.

A decrease in volumetric glucose consumption is observed to occur post induction (0 dpi). Subsequent increase glucose consumption occurs late in the production phase (6-14 dpi).



Supplement Figure 25. Titer profile of feeding regimens study.

Rapid increase in protein concentration is observed to happen late in the production phase (7 - 14 dpi) concomitant with increased oxygen and glucose requirements. Error bars represents the coefficient of variance associated to each variable.



Supplement Figure 26. Model results for trained features in the training dataset comprised of 17 cultures (8 Delta pool batches, 5 Beta pool batches, and 4 WuTL pool batches).

Blue continuous lines are every day averaged predictions while the shaded blue region represents the standard deviation of the training predictions. Red dots show averaged sampling day data while the red shaded region represents the standard deviation of the training dataset. (a) product titer, (b) cumulative glucose consumed, (c) ammonia, (d) lactate, (e) viable cell density (VCD).