Process Intensification of *Spodoptera frugiperda (Sf)* Cell Growth via Multi-Parallel Bioreactor System

by

Randy Stein

B.S., Chemical Engineering University of Pittsburgh (2012)

Submitted to the MIT Sloan School of Management and the Department of Chemical Engineering in Partial Fulfillment of the Requirements for the Degrees of

Master of Business Administration
and
Master of Science in Chemical Engineering

In conjunction with the Leaders for Global Operations Program at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2020

Signature of Author _____
MIT Sloan School of Management, Department of Chemical Engineering
May 8, 2020

Certified by _____
Kristala Prather, Thesis Supervisor
Arthur D. Little Professor, MIT Department of Chemical Engineering

Certified by _____
Roy Welsch, Thesis Supervisor
Professor of Statistics and Data Science, Sloan School of Management

Approved by _____
Maura Herson
Assistant Dean of MBA Program, MIT Sloan School of Management

Approved by _____
Patrick Doyle
Robert T. Haslam (1911) Professor of Chemical Engineering

This Page is Intentionally Blank

Process Intensification of *Spodoptera frugiperda (Sf)* Cell Growth via Multi-Parallel Bioreactor System

by
Randy Stein

Submitted to the MIT Sloan School of Management and the Department of Chemical Engineering on May 8, 2020 in Partial Fulfillment of the Requirements for the Degrees of Master of Business Administration and Master of Science in Chemical Engineering

Abstract

The objective of this project is to improve the yield of the fed-batch manufacturing process for the production of Flublok influenza vaccine, which was approved by the FDA in 2018. In short, *Spodoptera frugiperda* (SF+) insect cells are grown to a specific target cell density and then infected with baculovirus containing the gene of interest (GOI). For this particular process, the recombinant hemagglutinin (rHA) is extracted from the cell and used to produce the influenza vaccine.

Protein Sciences developed a fed-batch process which improved on the traditional batch process by feeding supplementary nutrients to boost cell growth. The Fed-Batch process doubled the target cell density at the time of infection which resulted in a two-fold increase in the final yield of rHA and a 30% reduction in cost of goods. This Fed-Batch process can be further optimized to increase rHA yield and reduce the cost of goods. It is important to note that simply increasing cell biomass is not enough; the cells must also be able to produce rHA at a similar specific productivity in order to increase the yield.

Exploratory process improvement experiments were performed on the ambr250 ® multi-parallel bioreactor system, with the goal of identifying the growth conditions for maximizing SF+ cell density. The conditions yielding the best results from these experiments were replicated in 3L glass bioreactors. Using data from these experiments, an optimized Fed-batch process can be developed. In addition, a statistical model was developed to relate key process parameters to SF+ cell density. This model can be used to quantitively ascertain how cell density is impacted by changing process parameters.

Thesis Supervisor: Kristala Prather
Arthur D. Little Professor, MIT Department of Chemical Engineering

Thesis Supervisor: Roy Welsch
Professor of Statistics and Data Science, Sloan School of Management

Acknowledgements

"Endings are never easy. I always build them up so much in my head they can't possibly live up to my expectations, and I just end up disappointed. I guess it's because we all want to believe that what we do is very important; that people hang onto our every word; that they care what we think. The truth is, you should consider yourself lucky if you even occasionally get to make someone, anyone, feel a little better. (Scrubs, My Finale)"

This thesis is dedicated to my family, especially my mother and father, who helped me through some incredibly tough moments over the past two years. Right there with me were my Speakeasy friends, Houston friends, Lake Jackson friends, Sloanies, and roommates. Together, all of you formed a support network whose support and lightheartedness were immeasurable to me.

For all their hard work, I would like to appreciate my advisors Roy Welsch, Kristala Prather, and Indresh Srivastava. I would like to include all of those in the LGO staff and faculty (especially Ted, Patty, Katja, and Anna) who spent numerous hours answering questions, resolving issues, and offering their support. Last, I would like to acknowledge the assistance that Mia from MIT Student Mental Health and Counseling gave me.

Thank you to everyone at Sanofi Protein Sciences and Genzyme who made this work possible. I would like to acknowledge specifically the MSAT group at Genzyme, especially Alissa, Woo, and Canghai. Thank you to the great group of folks in Meriden: Nikolai, Jamal, Kate, Sam, Marcus, Liz, Mireli, and others. My gratitude also to Regis, Jean-Michel, and Sanofi-Pasteur for sponsoring the internship.

I extend a final gratitude to Dana Cagle, Ryan Powell, and Johanna Graishe, for believing in me and writing the Letters of Recommendation that helped to get me in the program. I know you were not there to witness my work, but I hope that it lived up to your recommendations. I will certainly pay it forward. In addition, to all of the staff and operators at SPC: I always wanted to pass the lessons and knowledge I learned from you to classmates at Sloan.

To the LGO class of 2020, congratulations and thank you for all the love and support! I hope our friendship continues to grow throughout the years. It seems like just yesterday I was driving more than thirty hours from Houston with Dante, and now it is all over. A special shoutout to fellow chemical engineers Ketan and Aaron, and teaching assistants Colin Grambow, Samuel Winslow, Thejas Wesley, and Joy Zeng, without whom I could not have passed Course 10.

Finally, my Nonni and Nonno are always in my thoughts, who would have loved to see me finish a program like this. Also, to my dog Chip, who was a very good boy.

Table of Contents

Table of Figures

Acronyms and Abbreviations

BEVS – Baculovirus Expression Vector System

CDC – Centers for Disease Control

CFS – Complete Feed Solution

CO_2/pCO_2 – carbon dioxide/dissolved carbon dioxide

DNA – deoxyribonucleic acid

DO% - dissolved oxygen percentage

DOT – dissolved oxygen tension

FDA – Food and Drug Administration

GOI – gene of interest

k_La – volumetric mass transfer coefficient (expressed in h^{-1})

mM – millimolar, unit of concentration

mOsm/kg – milliosmoles per kilogram of water, unit of osmotic pressure

MOI – multiplicity of infection

N_2 – nitrogen

O_2/pO_2 – oxygen/dissolved oxygen

PID – proportional-integral-derivative

PSFM – Protein Sciences ® Formulary Medium

R&D- research and development

rHA – recombinant hemagglutinin

RMSE - root mean squared error

RPM – revolutions per minute

STR- stirred tank reactor

SF+ - *Spodoptera Frugiperda expresSf+* cell line

Sf9 – *Spodoptera Frugiperda 9* cell line

TCA – tricarboxylic acid

VCD – viable cell density

(v/v) – volume per volume (expressed as a %)

WHO – World Health Organization

1. Introduction

This work is intended to scratch the surface of machine learning and optimization as they apply to protein production using an insect cell line like *Spodoptera frugiperda*. This work blends an intense focus on biochemical engineering with business pragmatism. The biopharmaceutical field, in general, struggles with the problem of determining what is optimal and what is "good enough." When it comes to R&D, the lifeblood of any biopharma business, there exists a struggle between perfection and the ability to deliver a viable solution to the consumer. In the context of optimizing an existing drug product process, this thesis aims to provide a framework to address the following issue: how can a company utilize the brilliance of its R&D staff, create an effective team, take advantage of simple machine learning techniques, and optimize a process in a timely and cost-friendly manner? Specifically, this thesis focuses on optimizing the cell growth of a *Spodoptera frugiperda 9* (Sf9)-like cell line, dubbed SF+. The approach is straightforward: perform a series of exploratory experiments on the multi-parallel ambr-250 ® system, recreate the best conditions in three-liter Applikon ® glass bioreactors, and develop a rudimentary model to predict growth for the cell line. While the goal is to maximize protein production for this cell line, this approach focuses solely on cell growth. Toward the end of this work, the groundwork is laid to apply this approach to maximizing rHA protein production on an Sf cell line. Furthermore, commentary will also be provided on organizational generational issues for R&D labs; these considerations can be force multipliers to get the most out of a team of researchers. For simplicity throughout the rest of this work, the cell line is referred to as the SF+ line, the ambr-250 ® system as the ambr system, and the three-liter Applikon ® glass bioreactors as 3L glass bioreactors.

1.1 Thesis Outline

This thesis is split into an introduction and background, where the author provides information on the project, its motivation, and a description of the industry and technology. In the next section, literature review, articles, and studies relevant to the project are reviewed. Research methodology and results follow, giving an overview of the five ambr experiments, data collection methods and processes, and the confirmation experiments run in 3L glass reactors. Following this is a model generation and results section, which presents the results of two regression models created from the data collected during this work. Finally, this thesis culminates with discussing key results and conclusions, business impacts, and future recommendations.

2. Background

The background is intended to provide a brief overview of influenza, its history, the history of the flu vaccine, an overview of batch and fed/fed-batch recombinant vaccine production processes, a history of Sanofi, and the motivations behind this current work.

2.1 Impact of Influenza

For hundreds of years, influenza has plagued mankind with various outbreaks, epidemics, and pandemics that have cumulatively killed millions of people. In addition to these events, the seasonal flu kills several hundreds of thousands of people around the world. After the Spanish flu outbreak in 1918, a concentrated effort was made to develop a vaccine for the flu, and by the 1940s scientists developed the first inactivated flu vaccine approved for use in the United States.

The CDC estimates that in 2017-2018 influenza was responsible for more than 48 million infections, 959,000 hospitalizations, and 79,400 deaths in the United States. In addition, the median annual economic impact was estimated to be $2.47 million per county in the United States [1]. These costs include medical care, loss of productivity, and deaths. Flu vaccine producers are tasked to tackle this heavy health burden each year. In 2018, 155.3 million flu vaccine doses were delivered by producers, resulting in a flu vaccination coverage of only 37.1% in adults 18-64 and 57.9% among children [2].

Vaccines are developed and sold by private, for-profit pharmaceutical companies. As an example, the vaccines business segment of Sanofi Pasteur had a business operating income of $1.77 billion in 2016. This is compared to a business operating income of $8.83 billion for their pharmaceuticals segment [3]. Therefore, these companies have tremendous financial incentive to reduce their cost of goods and improve their vaccine yields. Major inactivated flu vaccine manufacturers in the United States include Sanofi Pasteur, Seqirus (CSL), and GlaxoSmithKline. The major market holders make significant investments in R&D as well as acquire new technologies to maintain a competitive advantage. In 2003, Medimmune developed a live (attenuated) Flumist vaccine, and it was acquired in 2007 by AstraZeneca for $15.6 billion. Finally, in 2013, Protein Sciences received approval to produce a recombinant flu vaccine, and it was acquired by Sanofi Pasteur in 2017 for $750 million. Optimization of the production of recombinant flu vaccine, which currently provides 25 million doses (15% of the US Market) will be the focus of this research paper[4].

2.2 The Influenza Virus

Before discussing the details of the recombinant virus vaccine, it is important to introduce some general background information on the influenza virus. There are four types of influenza viruses: A, B, C, and D. Generally, influenza A and B viruses are included in each year's influenza vaccine. Influenza C infections are not known to cause epidemics, and Influenza D viruses primarily affect cattle. Influenza A viruses are divided into subtypes based on two proteins on the surface of the virus: the hemagglutinin (H) and the neuraminidase (N) [5].

Hemagglutinin is a spike-shaped protein that extends from the surface of the virus. The name hemagglutinin comes from its ability to cause red blood cells to clump (agglutinate): Influenza viruses have many hemagglutinin molecules on their surface, so as they bind to red blood cells, many of them tend to coalesce [6]. The protein is composed of two primary forms of chains. The first chain targets specific sugar chains on cellular proteins, and, once bound, the second chain on hemagglutinin initiates the attack. A fusion peptide on the hemagglutinin protein locks the virus to the cell membrane, and the cell membrane subsequently becomes fused to the viral membrane. Viral RNA is allowed to enter the cell, beginning the infection process [6].

Neuraminidase is a protein that is composed of four identical subunits arranged in a square. It is typically attached to the virus surface through a long protein stock. The main function of neuraminidase occurs after the virus leaves an infected cell. It ensures that the virus does not remain on the cell surface by snipping the sugar chains to which the hemagglutinin is bound. There are 11 different neuraminidase subtypes, N1 through N11, defined by their interaction with antibodies. Each subtype is defeated by a similar type of antibody. Some of the subtypes promote infection in people, some in pigs, and some in birds. As the influenza virus

spreads and infects different organisms, these subtypes can mutate randomly into new combinations that are able to evade existing antibodies. This is one of the reasons that the influenza vaccine needs to be monitored and adjusted every year [7].

2.3 Protein Sciences

Protein Sciences is a biotech company established in 1983 based in Meriden, CT, which was acquired by Sanofi in 2017. Protein Sciences' mission is to save lives and improve health through the creation of innovative vaccines. Protein Sciences developed a novel recombinant influenza vaccine called Flublok ®. SF+ cells are grown to a specific target cell density and then infected with baculovirus containing the GOI [8]. The rHA is extracted from the cell and used to produce the influenza vaccine. Flublok ® was approved for use by the FDA in persons aged 18-49 in 2013, and then expanded to adults 18 and older as of October 2014 [9]. An overview of recombinant flu vaccines is provided in the next section.

2.4 Recombinant Flu Vaccines

Recombinant flu vaccines are produced using a method that does not require an egg-grown vaccine virus. Instead, for Flublok the gene sequence is obtained from a naturally occurring ("wild type") recommended flu vaccine virus. The genes encoding for the protein hemagglutinin are then cloned into a vector for the development of recombinant baculovirus. In particular, the insect cells that are used are that of *Spodoptera frugiperda,* or the Fall Armyworm. The baculovirus expression vector system (BEVS) is particularly suited for recombinant protein production for several reasons. Firstly, the eukaryotic insect cell that the baculovirus infects can perform higher order post-translational modification of proteins compared to yeast and bacterial cells. Next, baculoviruses are considered safe because they are unable to replicate in mammalian

cells. It is also particularly easy to culture insect cells to very high cell densities (> 1 x 10^7 cells/mL). Finally, two baculoviral genes, p10 and polyhedron, that are considered not essential for the replication of the baculovirus to the production of recombinant proteins, can be removed and replaced by a transgene of interest (e.g. a gene to express hemagglutinin).

For influenza, the first step is cloning of the hemagglutinin DNA sequence from wild type strains of the influenza virus. Various influenza strains are evaluated every season to match as close as possible to the strains that are circulating in the population. Influenza strains follow a standard World Health Organization (WHO) naming convention:

- The antigen type (A, B, C)
- The host of origin (swine, chicken)
- Geographical origin (Abu Dhabi, Texas)
- Strain Number (66, 87)
- Year of Isolation (57, 2010)
- For influenza A, the hemagglutinin and neuraminidase antigen are included in parentheses (H1N1)

The next step is transfecting SF+ cells with the extracellular baculovirus and the plasmid containing the gene of interest to obtain a recombinant virus, which is used to produce the target influenza antigen. Firstly, DNA from a baculovirus is linearized by cleaving at two specific sites on the baculovirus plasmid. Then, a plasmid containing the gene for the influenza hemagglutinin is recombined with the linearized baculovirus DNA to produce a recombined baculovirus plasmid. This plasmid is then transfected into SF+ cells using a (Sanofi-internal) liposome. Over the next few days, the viruses reproduce more recombinant viruses. It is important to note that

the recombinant baculoviruses are specialized to produce human influenza hemagglutinin protein [10]. The recombinant baculovirus is expanded through three passages termed as P1, P2, and P3. The recombinant baculovirus at P3 is frozen in liquid nitrogen and termed as a working virus bank (WVB). The working virus bank is expanded to a working virus stock and used to infect insect cells to produce the antigen of interest (hemagglutinin protein). After scale-up, process evaluation, and optimization, the technology is handed over to produce commercial batches. A summary of this vaccine production process is shown in Figure 2.1[11]. Flublok ® was originally produced using a batch production process, which is covered in the next section.

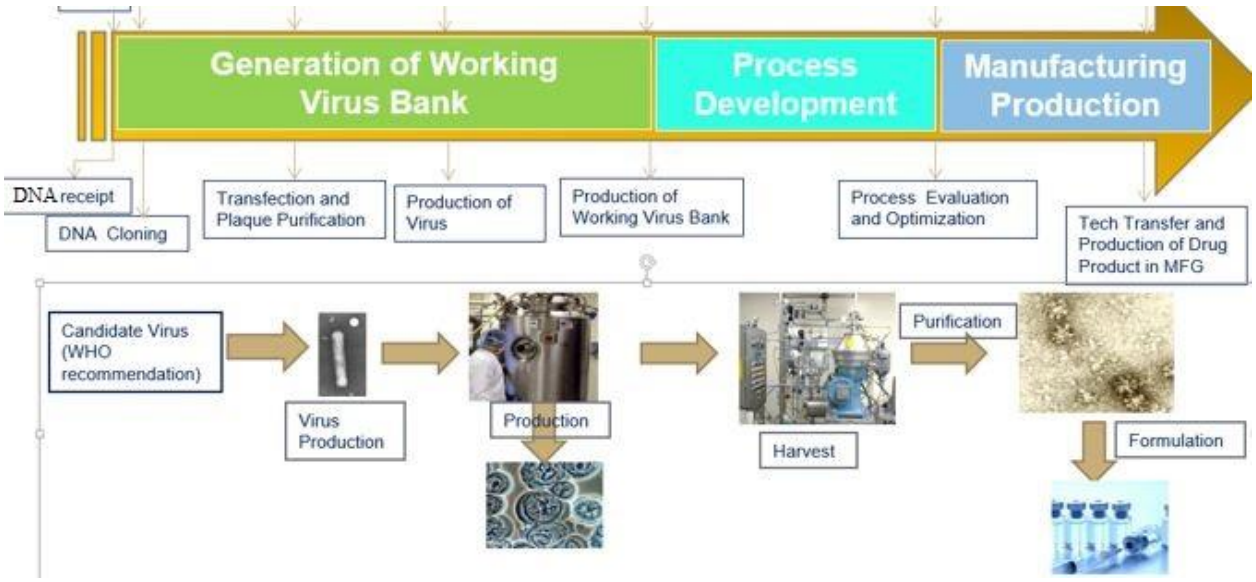

Figure 2.1 An Overview of Recombinant Vaccine Production Process

2.5 Batch Vaccine Production Process

In the traditional batch production process, passage 3 working virus banks are taken from a cryogenic tank and used to scale up to passage 4 and 5 to generate the working virus stock. In parallel, frozen cells are scaled up from smaller flask to larger STRs. Then, these cell at a target

cell density are infected with working virus stock at passage 5 to produce recombinant hemagglutinin protein. The infected cells are harvested at a target viability, and the protein of interest is extracted with detergent. The extracted material is depth filtered, and the target protein is purified into the commercial drug substance. The purification processes are proprietary to Sanofi and are not relevant to the scope of this research paper. There are several limitations to the traditional batch process. Firstly, several key nutrients are depleted and cannot be replenished in batch production, limiting cell growth beyond a certain density. Examples of these include glucose and glutamine. Next, growth can be impacted by accumulation of the byproducts ammonia and lactate, although the effect is less severe than it is for mammalian cells. Finally, in batch mode, oxygen depletion and carbon dioxide accumulation can also negatively affect cell growth and protein production.

2.6 Fed-Batch Vaccine Production Process

Because of the limitations of the batch process, the Sanofi team developed a more efficient process to increase the volumetric productivity of the target antigen using the fed-batch process. In 2018, the FDA approved a fed-batch process developed by Protein Sciences for influenza vaccine production. In short, the fed-batch process involves replenishing depleted nutrients to a growing cell culture to increase cell density and increase protein production. Figure 2.2 shows a simplified overview of the fed-batch process.

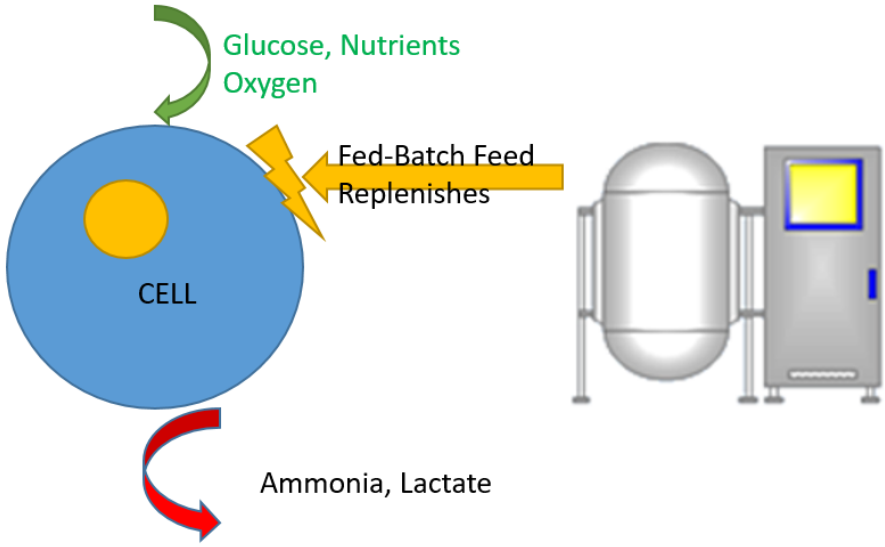

Figure 2.2 Simplified Overview of the Fed-Batch Process

To improve the yield for the vaccine antigen, the team at Sanofi developed a fed-batch process that involves doubling the cell density at the time of infection compared to a batch process by adding a single pulse of a feeding solution 24 hours before infection and then a second pulse at the time of infection. The feeding solution is a mixture of amino acids, vitamins, trace elements, yeastolate, and glucose. To improve the process efficiency, the Sanofi team streamlined the feeding strategy to only one feed at a specific target cell density 24 hours before infection. Cells were then infected at approximately double the cell density compared to the batch process, resulting in an approximately double volumetric rHA yield (mg/L) over the optimized batch process, with no detrimental effect on specific rHA productivity (mg/10^9 cells).

2.7 Project Motivation

Therein lies an opportunity to further improve on the fed-batch process to improve the yield of rHA and reduce the cost of goods. The thought process is that if a 2-3-fold higher cell

density at the time of infection is achieved, assuming the same specific productivity, the overall volumetric yield of rHA will be 2-3 times higher than the current yield. However, keeping the specific productivity constant at a higher cell density is not trivial, and this will be discussed in more detail in the future recommendations section. Nevertheless, establishing conditions that optimize SF+ cell growth and creating a model to predict cell growth lay the groundwork for future optimization efforts. Increasing the protein yield for the flu vaccine drug substance provides Sanofi flexibility to produce more vaccine dosages from the existing facility. First, they could simply create more drug substance to produce the vaccine to sell, as there is currently not enough vaccine production to satisfy demand. Next, producing more drug substance at a higher cell density could enable Sanofi to use its current footprint for producing more vaccine doses.

2.8 Project Overview

The goal of this project is to further optimize the Flublok ® fed-batch process by performing exploratory experiments on a Sartorius ambr multi-parallel bioreactor system. The aim of these experiments is to intensify the fed-batch process and find key variables of interest that can maximize SF+ cell growth. Some of the variables manipulated include feed percentage, dissolved oxygen setpoint, semi-continuous vs. bolus feeding, and glutamine/glucose supplementation. In some cases, shake-flask experiments were performed in parallel with the ambr experiments to validate small scale experiments. To the author's knowledge, there is no published work on *Sf* cells grown using the ambr bioreactor system to date.

Literature was reviewed to identify the best starting point for designing and running cell growth experiments on the ambr system. As the exploratory experiments progressed, select conditions of interest were used to design the subsequent experiments. In this way, a continuous

improvement mindset was incorporated, and things were not necessarily set in stone. After the five exploratory ambr experiments, confirmation runs in 3L glass reactors were performed using the most promising conditions from the ambr experiments.

During the exploratory experiments, amino acid, pH, and other nutrient data were also collected to determine if there were any key nutrients that are required for cell growth. Finally, data from all five ambr experiments were fed to a regression model to estimate peak SF+ cell growth. Data from the 3L glass confirmation runs were used to test the model.

3. Literature Review

In the first part of the literature review, reference works are summarized that cover the metabolism of SF+ as well as key variables relevant to the growth of SF+ cells in batch and fed-batch culture, such as dissolved carbon dioxide, semi-continuous feeding, and media formulation. Most of the references study the Sf9 cell line, which is similar to the SF+ line. However, there are likely subtle differences in metabolism between the cell lines which should be considered when reviewing this work. Results of previous studies were used to guide experiments to maximize cell growth on the ambr system. Because, to my knowledge, Sf9 cells have not been grown in this bioreactor system, one must compromise and use the results on a relative basis rather than an absolute basis. The learnings from the works covered within the literature review are used heuristically to design the experiments for cell growth.

When approaching the problem of maximizing cell density of the SF+ line, one looks for variables that can be manipulated to achieve the best possible results. In a biological system, these variables are numerous – they can range from engineering conditions like agitation speed, batch size, and oxygen addition rate to biological considerations like media composition, trace elements, or pH. As noted in the abstract, the focus of this work is to try to provide a practical approach to optimizing cell growth of an insect cell line. This practical approach requires solutions that are commercially viable within the context of a business – changes must be implementable within an acceptable timeframe (often less than two years), maintain stringent drug substance quality, and not be prohibitively costly. The sub-sections that follow highlight several variables that were considered in the ambr design of experiments. Additionally, some of these variables were examined in additional experiments that are discussed later in the results section.

3.1 Glucose

Glucose is the most important carbohydrate for insect cell growth. It is the most consumed nutrient and the main energy source for the cell [10]. Glucose is primarily provided to insect cells through the base medium and replenished in a fed-batch process using feed medium. Sf9 cells can also use other carbon sources like sucrose, maltose, and fructose. Because of the prevalent use of glucose in the bio-pharmaceutical world, its simplicity, and perceived disadvantage of using other carbon sources [10], these alternatives were not considered in this work. There is an active glycolytic process in insect cells: glucose is consumed through the parallel pentose phosphate pathway and the Citric Acid (TCA) cycle. Figure 3.1 and Figure 3.2 show summaries of both pathways [12] [13].

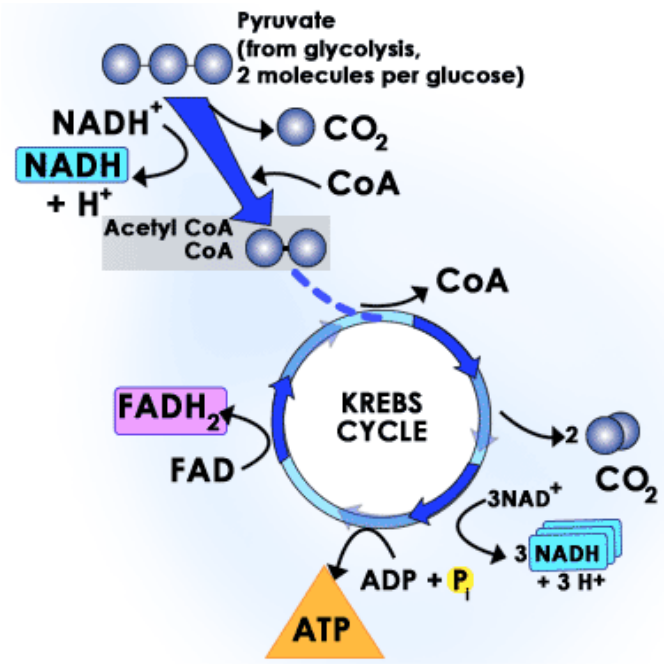

Figure 3.1 Summary of Citric Acid (TCA) Cycle

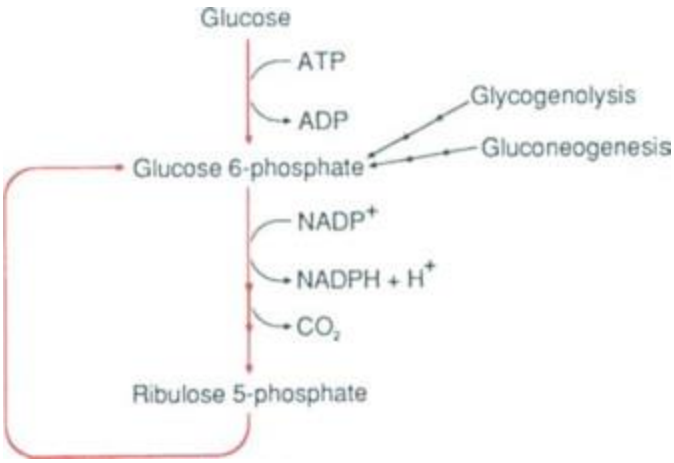

Figure 3.2 Summary of Pentose Phosphate Pathway

The first and primary strategy with glucose is to maintain at least a critical concentration in the bioreactor. Below this level, there are insufficient nutrients to sustain the culture, and biomass generation either stops or begins taking inefficient pathways. When glucose is exhausted, glutamine likely becomes the main energy supplier; alanine formation stops; and ammonia will start to form and accumulate [14]. Next, if glucose supply exceeds cellular requirements, overflow metabolism occurs through the TCA cycle, generating byproducts including alanine, lactate, and ammonia. Further, when glucose is in excess, the rate of consumption will increase [10]. This further signifies the importance of establishing a glucose concentration sweet spot: enough to supply the culture with energy, but not so much that it is wasted. There are three methods to replenish glucose in a bioreactor: bolus (batch) addition, semi-continuous addition, and continuous addition. Bolus and semi-continuous methods of feeding are considered in this work. Also, glucose can also be added within a feed or separately as a stand-alone component. Bolus addition is very simple for operators maintaining a biological process, but it introduces an osmolality shock and glucose concentration in the reactor. Semi-

continuous feeding strategy allows for much more precise glucose concentration control, keeping osmolality and glucose levels stable. However, in a highly regulated environment, semi-continuous feeding will require process control, pumps, flowmeters, and possibly round-the-clock shift coverage. Semi-continuous feeding must be justified through higher yield or better-quality product.

3.2 Glutamine and Glutamate

Glutamine is the most consumed amino acid in insect cell cultures, and it is rapidly depleted. Just like with glucose, ample glutamine must be made available in the medium to avoid the formation of toxic byproducts and metabolic wastes. As mentioned in the glucose subsection, in glucose limited cultures, glutamine will become a main energy supplier. When this happens, ammonia is formed as a byproduct via the glutaminase/glutamate dehydrogenase pathway, as shown in Figure 3.3 [15] .

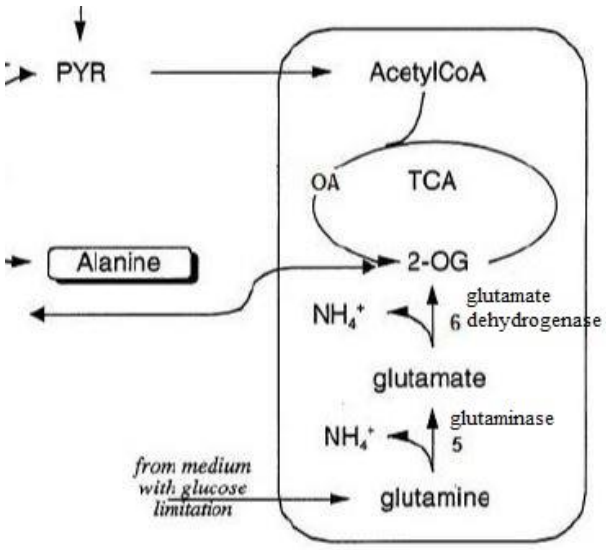

Figure 3.3 Glutaminase/Glutamate Dehydrogenase Pathway

It is likely that glutamine and glutamate act as osmolytes in intracellular fluid. Osmolytes help maintain the integrity of cells by affecting the viscosity and ionic strength of biological fluids. Osmolytes also directly impact the cell volume, and insect cells must achieve certain sizes before they can progress through the cell cycle [16]. Glutamate is a non-essential amino acid and is readily interchanged with glutamine. In the presence of glucose, through the glutamate-pyruvate transaminase pathway, glutamate produces alanine, which serves as an ammonia sink in the cell culture [17].

Equation 3.1 Pathway to Produce Alanine from Glutamate

$$Glutamate + Pyruvate \rightarrow Alanine + 2 - Oxoglutarate$$

Also, like glucose, the initial concentration of glutamine will affect its consumption rate, with higher concentrations leading to higher uptake rates. In insect cell media, glucose/glutamine concentration ratios are maintained commonly in the 2:1-7:1 range. Another important consideration of glutamine in production environments is its half-life. Glutamine is the only amino acid to degrade significantly in media [10]. While for lab-scale experiments it is easy to refresh and replenish media, large quantities of media may be stored for months at a time. Because the half-life of glutamine is roughly 25 days at 27 °C [18], the concentration of glutamine in aged media may be tangibly less than expected. This could have significant effects on cell growth profiles, protein production, or experimental results.

3.3 Asparagine, Serine, and other Amino Acids

Two other important amino acids considered as variables to examine were asparagine and serine. Asparagine is an amino acid that can also be used for energy production, and its metabolic uptake rate increases in glutamine limited cultures [19]. Asparagine can be used for both

energy production and biomass generation [10]. Serine is also used for the energy production and biomass generation. It can also be used to synthesize nucleic acids [10]. Cheng, citing Bruggert, et al. 2003[20], claims that cell division will stop once only serine in the media is depleted. However, in Bruggert's work, the Sf9 cells stopped dividing once all of arginine, tyrosine, and serine were depleted. Nevertheless, it is still an important amino acid for SF+ metabolism. While the impact of other amino acids was not examined as deeply as the ones above, some merit further examination including glycine, lysine, leucine, arginine, tyrosine, and cysteine.

3.4 Ammonia, Lactate, and Alanine

The three main metabolic byproducts to be concerned about with insect cell cultures are ammonia, lactate, and carbon dioxide. Alanine, while an amino acid, serves as a nitrogen sink for SF+ cultures and is the only amino acid produced in significant amounts by insect cells. When glucose supply exceeds requirements, overflow metabolism through the TCA cycle will occur, generating metabolic byproducts including alanine, lactate, and ammonia [21]. Lactate production is generally a signal of low protein yield and off-balance metabolism. Lactate production below 2 mM indicates a stress-free culture and shows that most glucose is completely oxidized [10]. Ammonia, as mentioned previously, is produced as a byproduct of glutamine consumption through the glutaminase/glutamate dehydrogenase pathway. This occurs when glucose is depleted. For the purposes of this study, it is relevant to maintain ammonia levels at a concentration that is not detrimental to cell growth. One study found that ammonia concentrations of up to 10 mM were not detrimental to insect cell growth [14]. Carbon dioxide is produced as a result of respiration. A gas control strategy must be employed to remove carbon dioxide from the headspace of bioreactor cultures to avoid saturation of the media. The Sanofi team found that that dissolved carbon dioxide levels greater than 100 mmHg inhibited cellular

growth [22]. Therefore, a gas sweep was employed to maintain a dissolved carbon dioxide concentration below this level.

3.5 Dissolved Oxygen and Oxygen Transport

Dissolved oxygen is essential for cell respiration. If cells receive too little oxygen, they will cease to divide, and the culture will die. Sf9 metabolism is affected under a dissolved oxygen tension (DOT) of 10%. Between 10-30% DOT, there are no significant differences between specific growth rate, cell concentration, nor amino acid consumption/production [23]. While the effects of excess dissolved oxygen have not been studied as extensively, one study found the presence of actively moving elongated Sf9 cells when cultures had excess oxygen. These elongated cells were resistant to baculovirus infection, so supplying excess oxygen could lead to lower protein yields in a BEVS [24]. Oxygen is typically provided to a reactor through a sparger, and agitation uniformly distributes the culture as well as aids in oxygen transport. Oxygen supply is one of the key limiting criteria for insect cell growth, so quantifying oxygen transport is important in the design and scale-up of the bioprocess. Criteria such as volumetric oxygen transfer coefficient, or k_La, are determined to evaluate how oxygen transport will scale from the benchtop to larger bioreactors. Factors such as vessel geometry, agitation, sparger geometry, media properties, and morphology of cells influence this oxygen transfer coefficient [25]. The k_La of the 250 mL ambr bioreactor was determined empirically as a function of the cell culture volume, agitation, and total gas sparge rate, as provided in Equation 3.2:

Equation 3.2 Empirical Formula to Determine kLa of ambr-250 ® Bioreactors

$$k_L a = V^{\alpha} \omega^{\beta} v^{\gamma}$$

V refers to volume in mL, ω refers to the agitation in rpm, v refers to the gas sparge rate in ccm, and [α β γ] is a vector of experimentally determined exponents. For 3 L Applikon reactors used later in the experiment, an initial k_La of 17 [1/h] was assumed. For these 3L bioreactors, Applikon Biotechnology in conjunction with Delft University of Technology reported Equation 3.3 which can be used to determine the volumetric oxygen transfer coefficient. Note that the root mean squared error (RMSE) for this equation was reported to be 38.6 h^{-1} for 3 L bioreactors, which may be an issue when estimating low k_La values.

Equation 3.3 Empirical Formula to Determine kLa of 3 L Applikon Bioreactor

$$k_L a = \alpha \left(\frac{P_g}{V_L}\right)^{\beta} * (v_{gs})^{\gamma}$$

where P_g is the gassed power input of the stirrer, V_L is the liquid volume of the reactor, v_{gs} is the superficial gas velocity, and [α β γ] is a vector of experimentally determined constants [26].

3.6 Bioreactor Foaming

Proteins, lipids, and other organic substances present in cell cultures act as natural surfactants, breaking surface tension and allowing air to form bubbles at the surface of the bioreactor. The foam amount and character depend on agitation, sparge rate, cell density, and media composition [27]. Foam is present in SF+ cultures, so care must be taken so that the culture does not overflow the bioreactor. Antifoam can be added to reduce foam in cultures, but because drug product is harvested with the cells in the BEVS for rHA production, the antifoam substance cannot be completely filtered out of the material collected. Extra precaution must be taken to determine the interference of antifoam on various unit operations during the downstream purification process. Besides overflow, foam can plug filters and gas lines, potentially depriving

the bioreactor of its overlay flow or blocking air outlets. Blocking air outlets is especially dangerous as it can lead to a sharp rise in bioreactor pressure. One simple way besides antifoam to minimize foam is to reduce the total gas flow to the reactor.

3.7 Common Media Compositions

Several serum-free feed media formulations have been used in Sf9 growth studies. Bédard and others have been instrumental in specifying compositions of feed media that have been used to optimize Sf9 growth and production. Some example media formulations will be listed in this section to give an idea of the composition of a typical Sf9 feed.

One of the earliest works that outlines a serum-free medium formulation for Sf9 cells is *Maximization of Recombinant Protein Yield in the Insect Cell/Baculovirus System by One-Time Addition of Nutrients to High-Density Batch Cultures* [28]. The paper outlines four concentrates: A (amino acids), B (yeastolate ultrafiltrate [29]), C (Lipid Emulsion per Maiorella[30] *et al.* 1988), and D (Vitamins, trace metals, and iron). Bédard further expands on this work in *Fed-batch Culture of Sf-9 Cells Supports 3 x 10^7 Cells per mL and Improves Baculovirus-expressed Recombinant Protein Yields.* In this work, four different feed media formulations are used to supplement a fed-batch culture of Sf9. The first of these feeds is the same formulation as the one described in Bédard *et al.* 1994. Table 3.1 summarizes the amino acid concentrations in each feed formulation, Table 3.2 summarizes the vitamin and trace metal concentration in each formulation, and Table 3.3 lists the (v/v)% of yeastolate and lipids in each formulation[28]. The concentrations were determined from concentrate information and methodology provided in the above references.

Table 3.1 Amino Acid Concentrations of Example Sf9 Fed-Batch Cultures

	Culture 1 "NCC1" (mM)	**Culture 2 "NCC2" (mM)**	**Culture 3 "NCC3" (mM)**	**Culture 4 "NCC4" (mM)**
Arginine	3.54	0.35	3.41	3.29
Asparagine	2.34	0.00	2.25	2.4
Aspartic Acid	2.32	0.00	2.23	2.32
Cysteine	0.38	1.32	0.37	2.69
Glutamic Acid	3.60	3.41	3.46	7.09
Glutamine	6.03	9.04	5.81	15.72
Glycine	7.64	0.09	7.35	7.72
Histidine	1.14	0.51	1.10	1.66
Isoleucine	5.04	0.00	4.86	5.04
Leucine	1.68	3.04	1.62	6.23
Lysine	3.02	1.49	2.90	4.54
Methionine	0.30	0.00	0.29	0.29
Phenylalanine	0.80	0.00	0.77	0.80
Proline	2.68	2.19	2.58	4.92
Serine	9.23	5.48	8.89	14.83
Threonine	1.30	1.36	1.25	2.79
Tryptophan	0.43	0.11	0.42	0.54
Tyrosine	0.00	0.95	0.93	1.65
Valine	0.75	1.32	0.72	2.71

Table 3.2 Vitamin and Trace Metal Concentrations of Example Sf9 Fed-Batch Cultures

	Culture 1 "NCC1" (µM)	**Culture 2 "NCC2" (µM)**	**Culture 3 "NCC3" (µM)**	**Culture 4 "NCC4" (µM)**
Thiamine * HCl * ½ H_2O	0.84	0.82	0.81	1.42
Riboflavin	0.75	0.73	0.72	1.28
D-Calcium Pantothenate	0.64	0.62	0.62	1.08
Pyridoxine HCl	6.86	6.70	6.60	11.6
Para-aminobenzoic acid	8.23	8.03	7.93	14.0
Nicotinic Acid	4.59	4.48	4.41	7.78
i-Inositol	7.83	7.64	7.54	13.3
Biotin	2.31	2.25	2.22	3.92
Choline Chloride	504	492	486	856
Cyanocobalamin	0.62	0.61	0.60	1.06
Folic Acid	0.64	0.62	0.62	1.08
Molybdic Acid, Ammonium Salt	0.12	0.12	0.12	0.21
Cobalt Chloride Hexahydrate	0.74	0.72	0.71	1.26
Cupric Chloride *2H_2O	0.41	0.40	0.40	0.70
Manganese Chloride *4H_2O	0.37	0.36	0.35	0.62
Zinc Chloride	1.04	1.01	1.00	1.76
Ferrous Sulfate *7H_2O	6.98	6.82	6.72	11.9
Aspartate	9.44	9.21	9.08	16.0

Table 3.3 Yeastolate and Lipid Concentrations of Example Sf9 Fed-Batch Cultures

	Culture 1 "NCC1" (v/v%)	**Culture 2 "NCC2" (v/v%)**	**Culture 3 "NCC3" (v/v%)**	**Culture 4 "NCC4" (v/v%)**
50x Yeastolate Ultrafiltrate	1.76	3.87	3.82	6.74
100x Lipid Emulsion	0.88	0.77	0.76	1.35

It is important to note that the concentrations given in these tables are concentrations supplemented via feed to the cell cultures, not the base concentrations in the original cell culture. The actual concentration is also dependent on residual concentrations of components in the depleted media as well as contributed components from yeastolate. Of these formulations, "NCC4" performed the best producing a maximum Sf9 cell density just shy of 30 million cells/mL.

3.8 Other Considerations

The above sections do not intend to be an exhaustive list of the factors important to SF+ cell growth and protein production. Other important factors include cell culture pH and shear stresses introduced by agitation. Typically, Sf9 cultures are maintained in the 6.0 – 7.0 range, and one study found no significant effects of pH on cell viability, growth, and protein expression in this range. However, pH rising above 7.0 due to overfeeding or ammonia accumulation should be avoided as it is likely to affect protein production level/quality of the product. These effects were not studied in detail in this work as it is focused primarily on increasing the cell growth rather than protein production.

Another important factor to consider is cell stress due to shear forces. SF+ cells are susceptible to damage due to hydrodynamic forces due to their relatively large size and lack of a

cell wall. Higher agitation speeds have been known to significantly decrease cell viability [31]. A very low concentration of a chemical like Pluronic F68 is added to mitigate shear stresses. An engineering factor like the P/V ratio, or power input, can be used to ensure that shear forces scale appropriately with the size of the bioreactor. At the conditions in the ambr bioreactors, the power input is likely in the range of 10-20 W/m^3 [32]. In the Applikon 3L bioreactors, power input is likely in the range of 15-25 W/m^3 [33]. Cell viability should not have been significantly affected in these ranges.

4. Ambr-250 ® Experiment Methodology

4.1 Cell Lines and Growth Media

SF+ cells are a proprietary cell line of Sanofi – Protein Sciences. They are maintained at 27 °C in 1 L shake flasks (Corning, NY) with 300 mL of Protein Sciences Formulary Medium (PSFM). SF+ cells are grown in serum-free media and produce higher yields of recombinant proteins when compared with its parent cell line Sf9 [22]. The SF+ cells were routinely sub-cultured every 2 or 3 days and maintained in a shaking incubator at 27 °C and 0% humidity.

4.2 Analytical Assays

Samples from bioreactors were collected at different time points. Cell density and size were determined using a Vi-Cell XR Analyzer (Beckman Coulter, Miami, FL). For cell densities greater than 1 x 10^7 cells/mL, cells were split to a ratio of 3:1, with three parts PSFM and one-part cell culture. The Vi-Cell analyzer was not attached to the ambr system, so some settling of cells was possible when moving the sample to be analyzed. Precautions were taken to minimize this variability. From the ambr system, samples were taken routinely and analyzed immediately using a BioProfile FLEX2 (Waltham, MA). The concentrations measured were glucose, lactate,

glutamine, glutamate, ammonia, sodium, potassium, and calcium. The pH, pCO_2, and pO_2 of the bioreactor were also measured. The ambr system constantly tracks pH, temperature, dissolved oxygen % (DO%), off-gas CO_2 and O_2 levels, and impeller speed in rpm. Other metabolites, calcium, glucose, glutamate, glutamine, iron, lactate, magnesium, phosphate, and sodium were also tested using a Cedex Bio HT Analyzer (Roche, Indianapolis, IN). Finally, for select samples, a full amino acid panel was provided from AminoAcids.com (St. Paul, MN). Cell samples for amino acids were centrifuged at 3000xg for five minutes, decanted, frozen at 80°, and shipped for analysis.

4.3 ambr Bioreactor System

The ambr system used in this experiment had 12 parallel bioreactor slots. Each single-use bioreactor vessel was 250 mL without baffles and included a double pitched-blade impeller, spot based DO sensor, disposable pH electrode, integrated gas and liquid inlet filters, sparge and overlay gassing, and an integrated condenser [34]. An example of the single-use bioreactor is given in Figure 4.1 [34].

Figure 4.1 Example of a Disposable 250 mL ambr Bioreactor

5. Ambr-250 ® Bioreactor Experiment Results and Discussion

Five exploratory experiments were performed on the ambr system. The goal of these experiments was to find conditions optimal for SF+ cell growth, collect data on SF+ growth in ambr bioreactors, and evaluate the effect of changing process conditions like feed percentage, semi-continuous feeding, and amino acid supplementation. Experiment 1 focused on establishing the base agitation and sparge conditions to use in the subsequent experiments. Experiment 2 evaluated the effect of changing feed percentage on SF+ cell growth, with each bioreactor

receiving only one feed during the experiment. Experiment 3 looked at both changing feed percentage and glutamine/glucose supplementation, with each bioreactor receiving one feed each day for five days. The fourth experiment investigated semi-continuous feeding. The fifth and final experiment assessed the effect of amino acid supplementation and varying dissolved oxygen level on cell growth.

5.1 Experiment 1 – Establishing SF+ ambr Gas and Agitation Strategy

The objective of this experiment was to determine the base agitation and gas sparging strategy for SF+ cells in the ambr bioreactors. Agitation and air sparge volumetric flow rate were varied, and peak VCD was recorded for each bioreactor. Then, a statistical software package was used to determine which agitation and air sparge rate would provide adequate oxygen transport and locally optimum growth environment for the cell culture. This experiment was insufficient to find the globally optimum conditions for working volume, agitation, air sparge flow rate, and gas overlay flow rate. Indeed, in the fifth experiment, gas control was changed in such a way that improved cell growth. The reason that more time was not spent on determining the best conditions for growth is that the ambr system used was only available for a limited period. Therefore, an adequate, rather than optimal, agitation and air sparger flow rate was deemed sufficient. Table 5.1 summarizes the conditions for each of the bioreactors in the experiment. A working volume of 180 mL and overlay flow rate of 5 ccm was used for each bioreactor. The air sparge volumetric flow refers to the initial flow of air to the bioreactor. As oxygen demand increased, oxygen was added to the bioreactor through the sparger up to a maximum of 15 ccm. This high limit was not varied among the reactors, except for Bioreactor 9 due to an oxygen deficiency. Bioreactors were seeded at approximately 1.0 x 10^6 million cells/mL. Bioreactors 11 and 12 were used to test the fed-batch process for Experiment 2.

Table 5.1 Summary of Conditions for Experiment 1

**Note – kLa values determined using linear regression, results may not be exact*

	200 RPM	350 RPM	500 RPM	0.5 ccm	1.5 ccm	2.5 ccm	*k_La (h^{-1})
Bioreactor 1	_	+	_	_	+	_	3.10
Bioreactor 2	+	_	_	_	+	_	1.81
Bioreactor 3	_	+	_	_	_	+	3.16
Bioreactor 4	_	_	+	_	_	+	4.45
Bioreactor 5	_	_	+	+	_	_	4.33
Bioreactor 6	_	+	_	+	_	_	3.04
Bioreactor 7	_	+	_	_	+	_	3.10
Bioreactor 8	_	_	+	_	+	_	4.39
Bioreactor 9	+	_	_	+	_	_	1.75
Bioreactor 10	+	_	_	_	_	+	1.87

5.1.1. Experiment 1 Viable Cell Density Plots

Figure 9.1 shows the effect of increasing the initial air sparge flow on the growth profiles of bioreactors 2, 9, and 10. All three bioreactors had working volumes of 180 mL and an agitation rate of 200 RPM. In experiments 2-5, cell viability plots will also be shown, but since there were no abnormalities and this was an experiment solely to establish base conditions, they were not included.

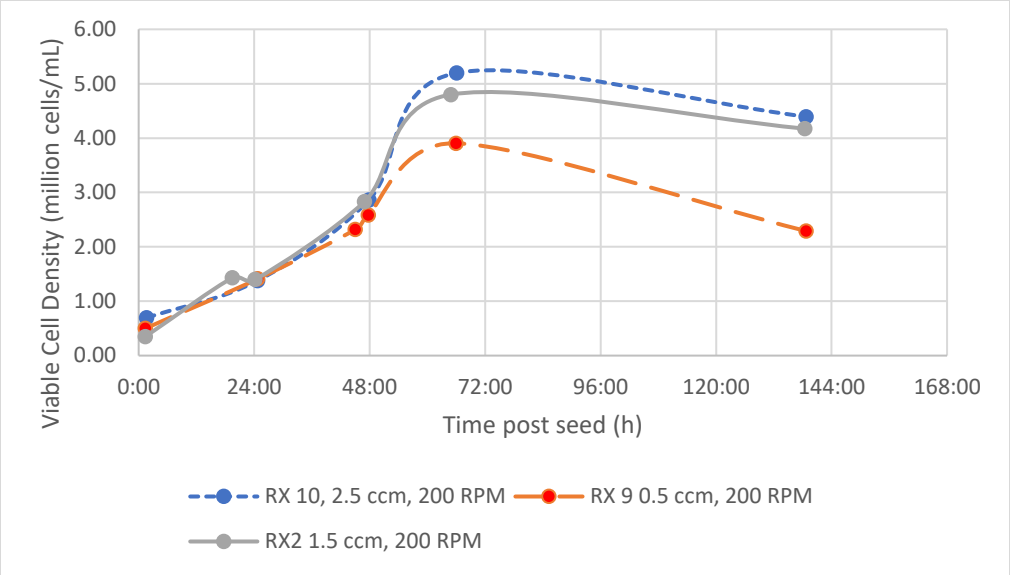

Figure 5.1 Effect of Increasing Air Sparge Flow Rate on Cell Growth Profiles

As one can see, the reactor with the lowest ccm of air sparge flow had a markedly worse cell growth. The same is true of reactors at the other agitation setpoints of 350 and 500 RPM, although the difference was most profound in the reactors with the lowest agitation rate. Figure 5.2 shows how important agitation was for growth of the SF+ cells in the ambr bioreactors. At the lowest air sparge rate, the reactors with agitation rates of 350 RPM to 500 RPM performed similarly while the reactor with a 200 RPM agitation rate performed demonstrably worse.

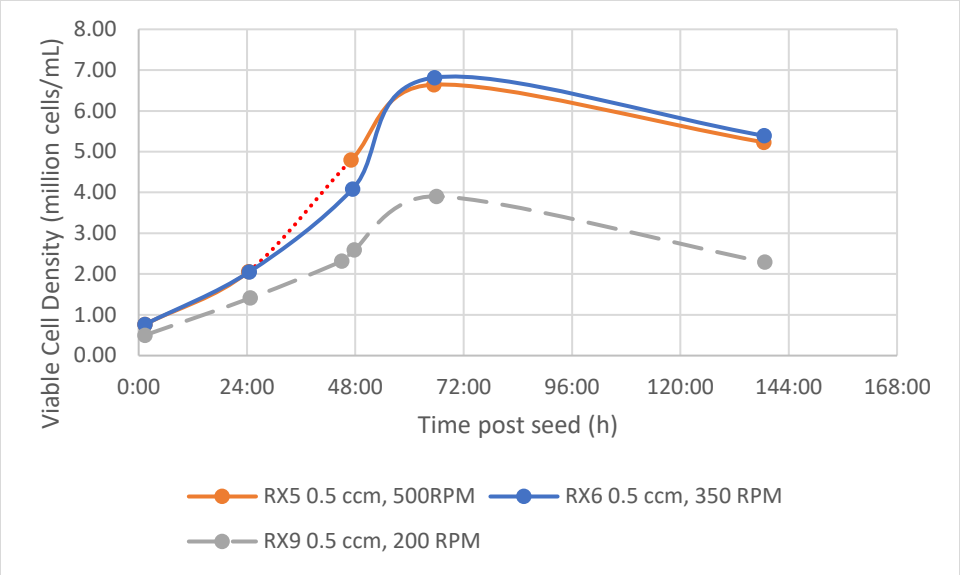

Figure 5.2 Effect of Agitation Rate on SF+ Cell Growth, 0.5 ccm Air Sparge Rate

Next, results at the highest agitation rate, 500 RPM, at varying gas sparge rates show that as agitation rate increases, there are diminishing returns to increasing the gas sparge rate. Further, while increasing the gas sparge rate will have a positive effect on oxygen transport, it increases the total overall gas flow rate to the bioreactor. A higher total gas flow rate leads to more foam, which can cause plugging problems, especially in the ambr bioreactors. It is a good principle to have as high a percentage of total gas flow rate be oxygen flow as possible without inducing oxidative stress. Elevated oxygen concentrations in the inlet gas have been known to stress cells at certain positions inside the bioreactor [35]. Figure 5.3 shows the cell growth patterns of bioreactors 4,5, and 8 at 500 RPM and varying air sparge flow rates.

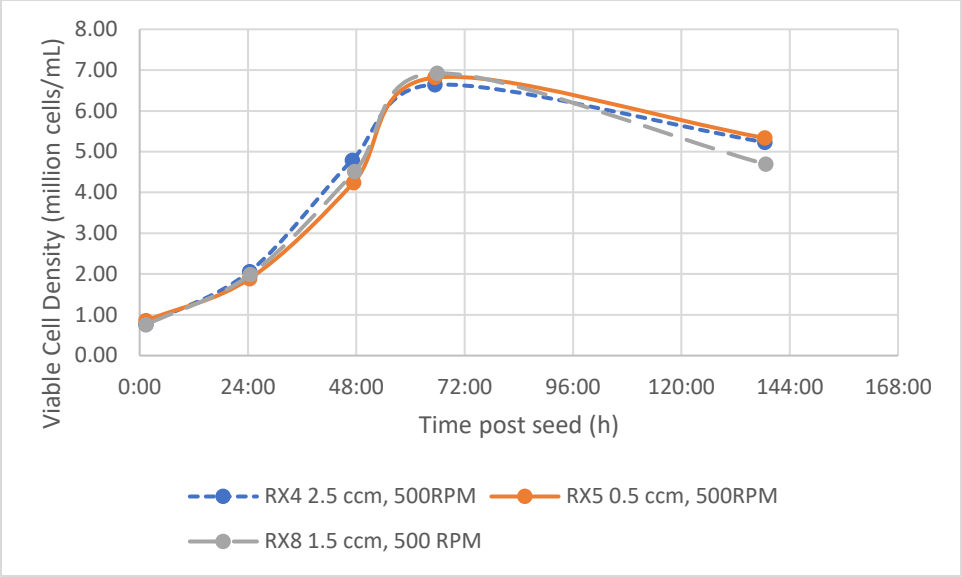

Figure 5.3 Effect of Air Sparge Rate on SF+ Cell Growth, 500 RPM Agitation Rate

While at 300 RPM and 500 RPM the ambr system was able to adequately track the dissolved oxygen setpoint, the reactors at 200 RPM had trouble maintaining the oxygen level in the bioreactor. This caused the oxygen flow rate to sharply ramp up to its 15 ccm limit and was likely the cause of the lower peak viable cell density as compared to the reactors with a higher agitation setpoint. Figure 5.4 contains the same reactors as Figure 5.1.

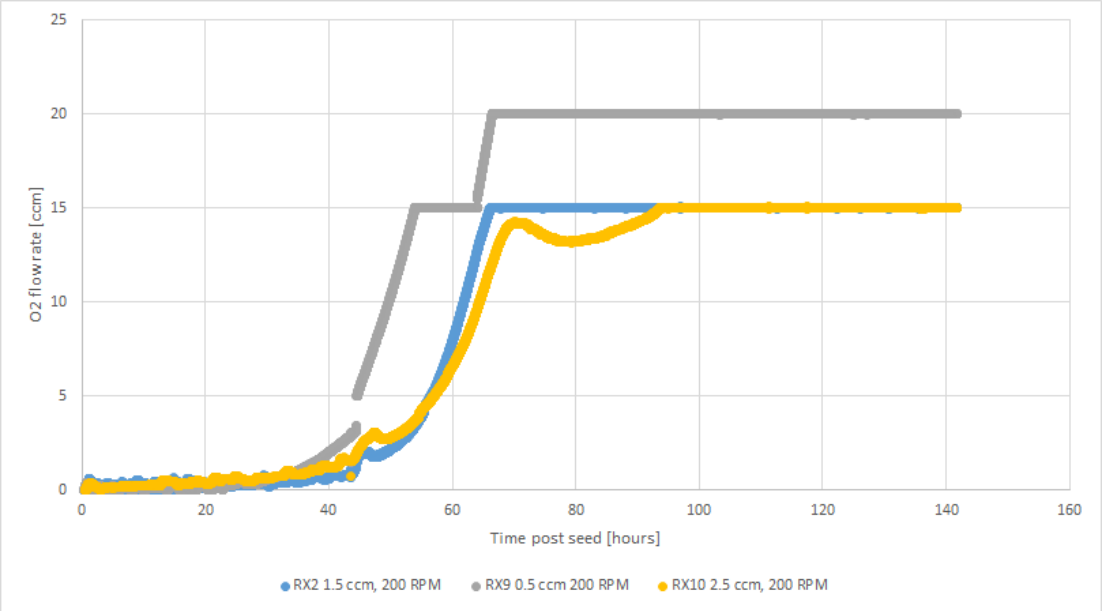

Figure 5.4 Oxygen Flow Profile of Bioreactors with 200 RPM Agitation Setpoint

5.1.2. Experiment 1 Glucose Consumption

Because the glucose metabolism of the bioreactors was linear and similar, one could estimate a specific glucose consumption rate for SF+ cells and compare with other literature values. Figure 5.5 shows the glucose concentration versus time for three reactors in the experiment.

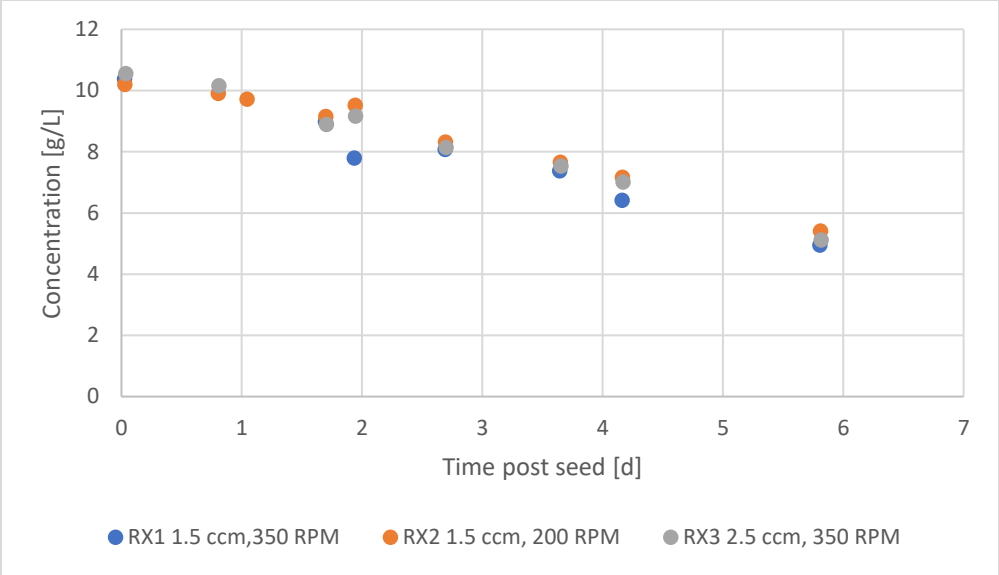

Figure 5.5 Glucose Concentration vs. Time for Three Experiment 1 Bioreactors

From this plot, one could fit the data to a linear model and estimate the specific glucose consumption rate to be 4.5 x 10^{-17} mol/cell*s. This is within the same power of ten as reported literature values of Sf9 glucose consumption. One study reported 2.5-3.2 x 10^{-17} mol/cell*s for Sf9 cells [36]. Potential reasons for the slight difference are subtle variations in the SF+ line, measurement noise, or the presence of serum in the literature Sf9 culture.

5.1.3. Experiment 1 Glutamine Consumption

The glutamine depleted relatively quickly for each of the bioreactors and stabilized thereafter. While later experiments show that glutamine becomes depleted completely after three days, the batch process without feed demonstrated that there was enough glutamine to sustain the cultures until another factor stopped their growth.

5.1.4. Experiment 1 Discussion

The viable cell density data from this experiment was used in conjunction with a statistical software package to determine the base agitation rate (500 RPM), working volume (170 mL), and air sparge flow rate (2.5 ccm) for the subsequent experiments. Oxygen control was established at 0 ccm to start, with an increase up to a high limit of 15 ccm, in order to maintain setpoint. An air overlay of 5 ccm was set to efficiently strip carbon dioxide. Again, these conditions were likely not the global optimum, but they did provide an adequate environment for SF+ cell growth. It was promising that the batch growth profiles of SF+ cells from this experiment were very similar to the growth profiles of SF+ cells in shake flasks (data not included), and thus further optimization work could be continued.

5.2 Experiment 2 Varying Single Feed Percentage

The goal of the second experiment was to advance from the batch bioreactor process to a single feed fed-batch process. A bolus of complete feed solution (CFS), proprietary to Sanofi, was added the second day after seeding the reactors. This is similar to the SF+ cell growth process for manufacturing. Six conditions were run in duplicate to determine the effect of feed percentage on cell growth. Two conditions used an agitation rate of 750 RPM to compare to the base conditions adopted from Experiment 1. Two conditions split their feed into two days (4% v/v each day) to compare with a reactor that had only a single feed of 8% (v/v). The high limit for the single feed percentage was set at 12% (v/v) - based on literature review and historical experience, a single feed would not be greater magnitude than that. The conditions for Experiment 2 are given in Table 5.2.

Table 5.2 Summary of Conditions for Experiment 2

	Agitation Rate (rpm)	Feed Percentage #1 (v/v)	Feed Percentage #2 (v/v)
Bioreactor 1	750	8%	N/A
Bioreactor 2	500	8%	N/A
Bioreactor 3	500	2%	N/A
Bioreactor 4	500	4%	4%
Bioreactor 5	500	12%	N/A
Bioreactor 6	500	6%	N/A
Bioreactor 7	500	6%	N/A
Bioreactor 8	500	12%	N/A
Bioreactor 9	500	8%	N/A
Bioreactor 10	750	8%	N/A
Bioreactor 11	500	4%	4%
Bioreactor 12	500	2%	N/A

5.2.1. Experiment 2 Cell Growth and Viability Plots

Figure 5.6 shows the viable cell density over time for each of the reactors at 500 RPM with a single feed (Bioreactors 2,3,5,6,8,9 and 12). An additional bioreactor from Experiment 5, with a 16% feed, was also included in these data. The black line indicates the approximate time of the feed.

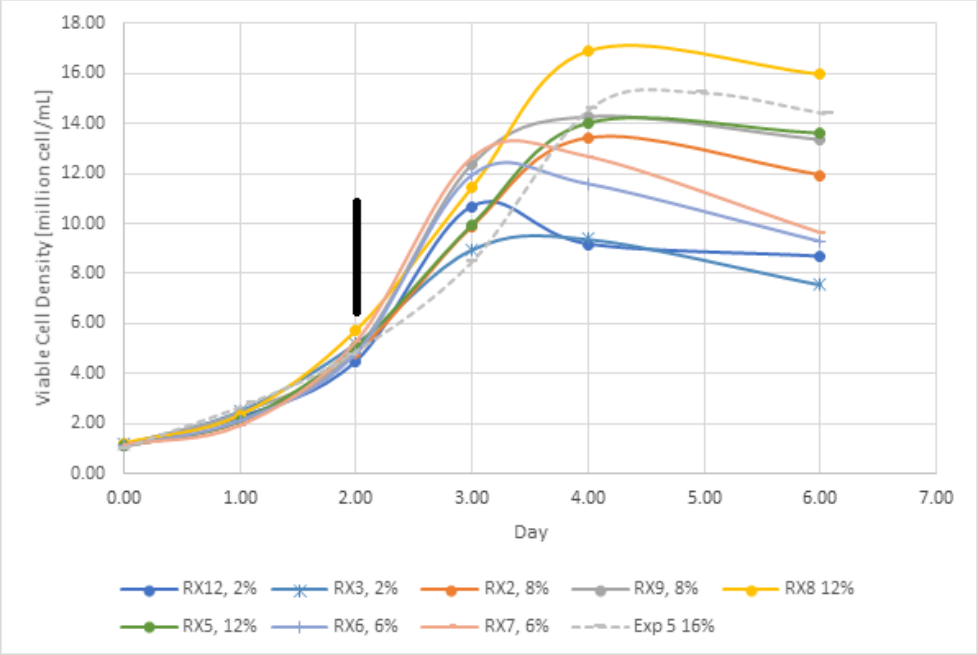

Figure 5.6 Viable Cell Density Plot, Experiment 2, Varying Single Feed Percentage

Peak viable cell density increases with increasing feed percentage, with diminishing returns after 12% feed. Figure 5.7 better illustrates this trend.

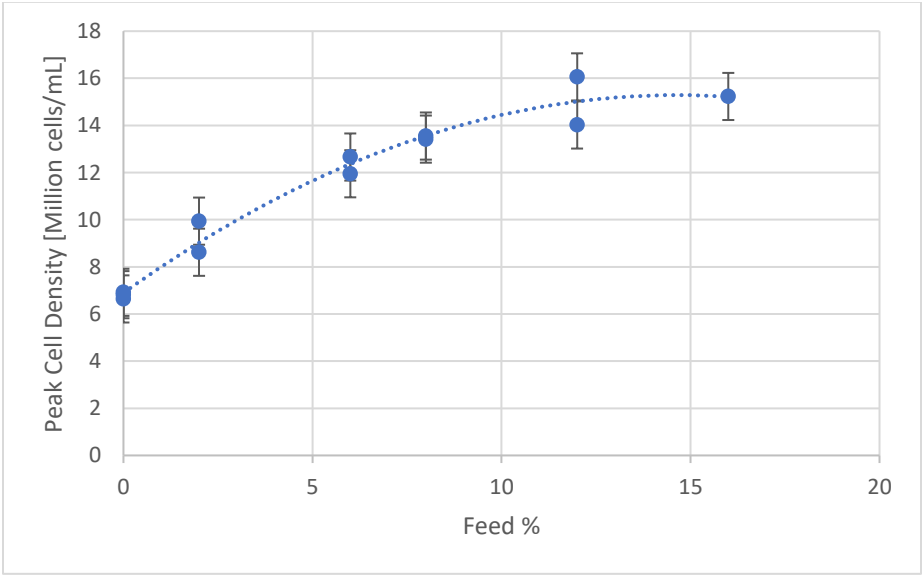

Figure 5.7 Peak Cell Density for Single Feed Bioreactors

Figure 5.8 shows the viability plot for each of the reactors from Figure 5.6. Most of the bioreactors can maintain a viability above 90% with only a single feed until day 5 post-seed, apart from the reactors with 2% v/v feed. This suggests that a single feed at 4% or above is enough to sustain a cell culture for up to five days.

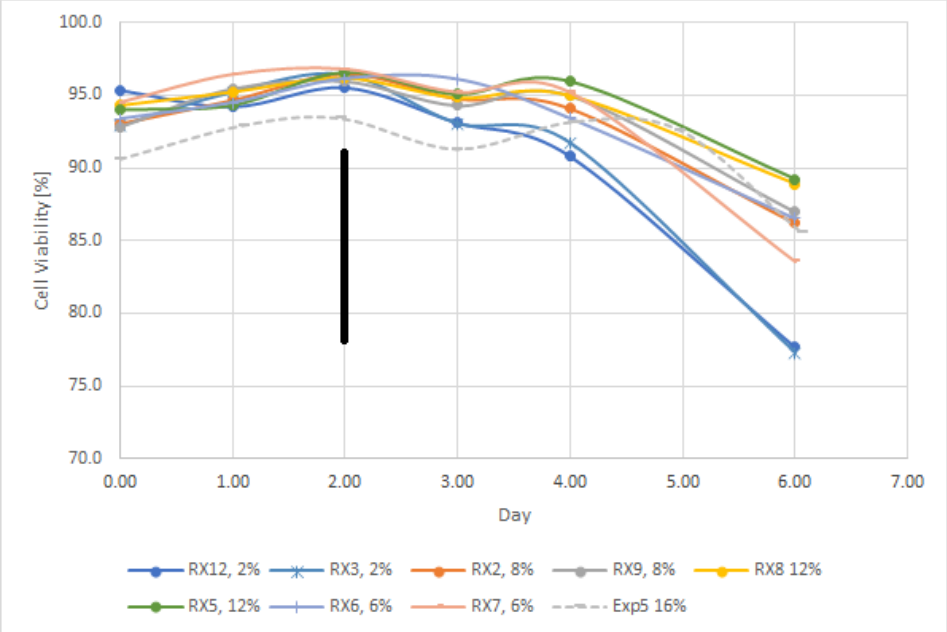

Figure 5.8 Cell Viability Plot, Experiment 2, Varying Single Feed Percentage

5.2.2. Experiment 2 Glucose Consumption

As in Experiment 1, a specific glucose consumption rate was calculated by linearly fitting the glucose consumption rates of specific bioreactors versus average cell density. This is shown in Figure 5.9.

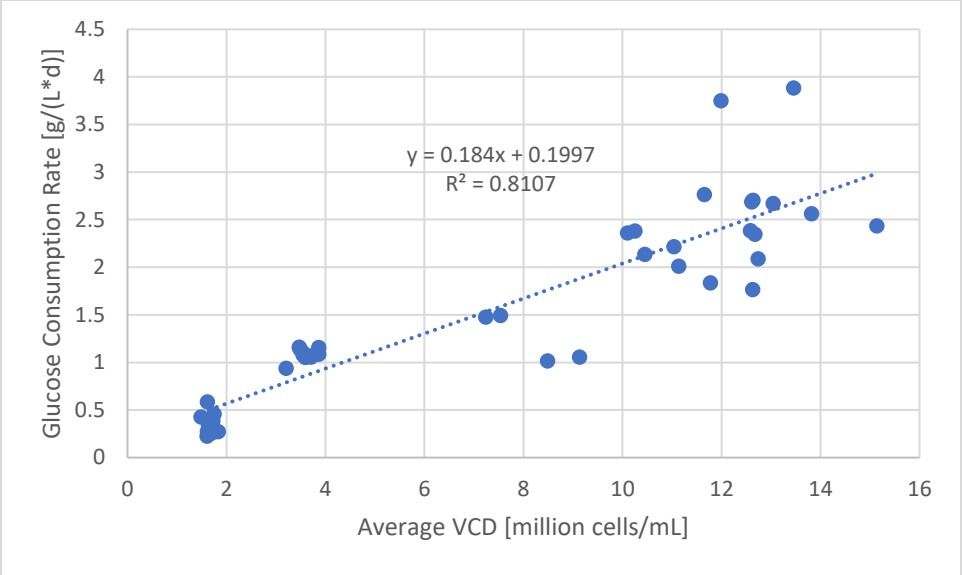

Figure 5.9 Glucose Consumption Rate vs Average VCD, Experiment 2

The resulting specific glucose consumption rate found was 1.2 x 10^{-17} mol/cell*s. While this is a very crude estimation, once again it is promising that it is within the same power of ten as rates quoted in literature for Sf9 cells. It was evident from this experiment that glucose consumption did in fact scale with amount added in the feed, as cited in the literature review. Further discussion on glucose consumption rate for all the experiments will be covered at the end of this chapter.

5.2.3. Experiment 2 Glutamine, Glutamate, Iron, and Phosphate

Glutamate concentration remained relatively stable in the cultures until it was fed, with a bump in concentration proportionate to the amount of feed added. Glutamine was consumed rapidly by day three in all cases except for the reactors with 12% feed. After day three, glutamine concentration slowly rose in all single-feed bioreactors, suggesting that SF+ cells can synthesize glutamine themselves [19]. Figure 5.10 is a side-by-side comparison of the glutamine and

glutamate concentrations versus time for several bioreactors from Experiment 2. The units on the y-axis are omitted for proprietary reasons.

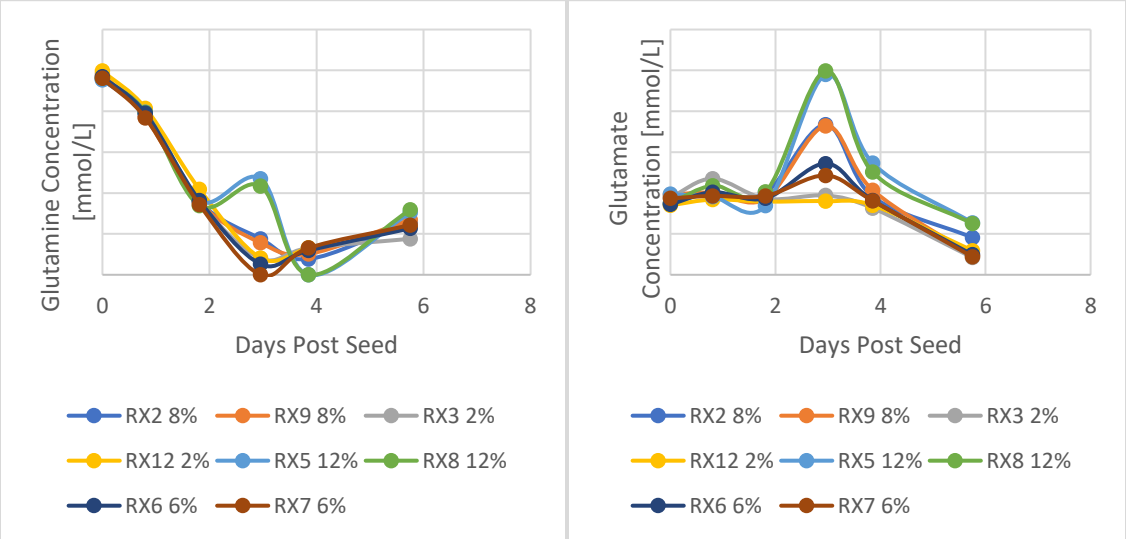

Figure 5.10 Glutamine and Glutamate Concentrations vs. Time, Experiment 2

Extracellular phosphate levels were also rapidly depleted in the culture. Phosphate was depleted completely in the 12% feed reactors. Levels started to rise gradually at the end, possibly due to cell death. Because phosphate was depleted, a side experiment was done in shake flasks to investigate any benefit to supplementing phosphate. After supplementing a small amount of phosphate to prevent depletion, there was no benefit to cell growth. These results are included in Appendix A. Iron concentration in the bioreactor followed a similar pattern to glutamate, with a rise proportional to the feed percentage. These are shown in Figure 5.11.

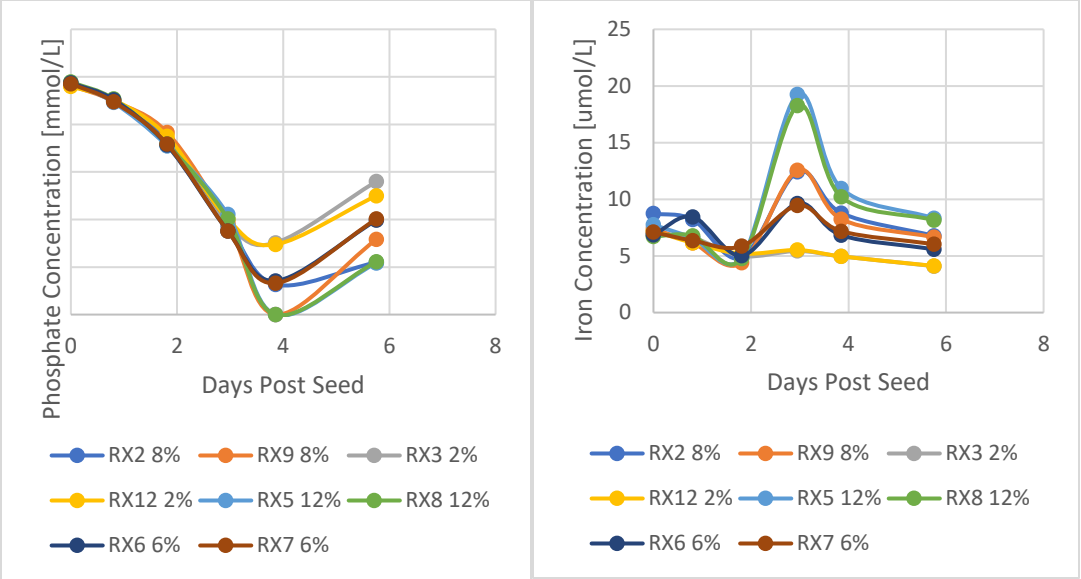

Figure 5.11 Phosphate and Iron Concentrations vs. Time, Experiment 2

5.2.4. Experiment 2 Osmolality

One thing that was measured in this experiment was the ability of the bioreactor medium to absorb the shock of a bolus addition of CFS. Figure 5.12 shows the osmolality versus time for the single feed bioreactors. Because of the osmolality rise caused by the 12% v/v feed, this was established as the high-limit basis for the multiple bolus experiment.

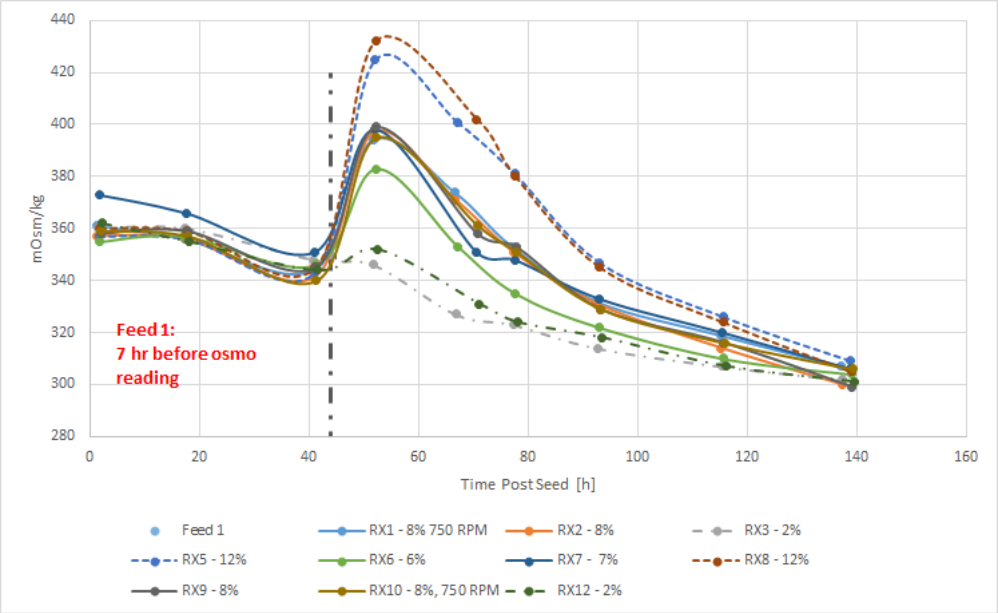

Figure 5.12 Osmolality vs. Time, Experiment 2

5.2.5. Experiment 2 Discussion

The single feed experiment demonstrated that peak viable cell density increased with increasing feed percent. An upper limit of 12% for a single bolus was established based on osmolality rise after feed. There was no appreciable rise in ammonia concentration in the bioreactors during this experiment, and there was no significant accumulation of lactate for any of the cultures. The reactors at 750 RPM died more quickly than the 500 RPM reactors, possibly due to increased shear forces. For this reason, these results were not included in this section, and the agitation rate for the rest of the ambr experiments was maintained at 500 RPM. There was also no apparent cell growth benefit to feeding two feeds of 4% v/v (Bioreactors 4 and 11) as opposed to a single feed of 8% v/v (Bioreactors 2 and 9). The next ambr experiment evaluated the effect of multiple feed additions (up to five total, one feed addition each day).

5.3 Experiment 3 Varying Multiple Feed Percentage

After evaluating the effects of a single feed, the effects of varying the bolus percentage (v/v) of multiple feeds was investigated. To isolate the impact of other components in the media, glucose was maintained above a minimum level to avoid depletion for all bioreactors. While it is likely that the level of glucose in the bioreactor could be further optimized to reduce the generation of byproducts, it was felt that other avenues of investigation would be more worthwhile for the limited remaining time using the ambr system. The twelve bioreactors were split into three groups and fed at three different levels, once a day, for five days: 4%, 8%, and 12%. In addition, two bioreactors of each group were supplemented with glutamine to avoid depletion. Finally, each bioreactor was run in duplicate. Each reactor had an agitation rate of 500 RPM, air sparge rate of 2.5 ccm, air overlay of 5 ccm, and was seeded at 1.0×10^6 million cells/mL. The experimental design of Experiment 3 is listed in Table 5.3.

Table 5.3 Summary of Conditions for Experiment 3

	Glutamine Supplementation?	Daily Feed Percentage (v/v)
Bioreactor 1	No	12%
Bioreactor 2	Yes	8%
Bioreactor 3	Yes	4%
Bioreactor 4	No	8%
Bioreactor 5	Yes	12%
Bioreactor 6	No	4%
Bioreactor 7	No	8%
Bioreactor 8	Yes	8%
Bioreactor 9	No	4%
Bioreactor 10	Yes	4%
Bioreactor 11	Yes	12%
Bioreactor 12	No	12%

5.3.1. Experiment 3 Cell Growth and Viability Plots

Unlike the experiment with single feeds, the bioreactor with the highest volume percent feed in Experiment 3, 12% v/v, performed the worst. The bioreactors that reached the highest peak cell density were the 4% v/v-fed bioreactors. Each insect cell culture followed a similar growth pattern to the single-feed reactors, with most of the cultures showing signs of death 120 hours post-seed. Figure 5.13 shows the viable cell density plot for the non-glutamine supplemented reactors of Experiment 3, and Figure 5.14 shows the cell viability plot for these same reactors. The black lines indicate the approximate time of feed. Note, due to over-feeding, there were only four 8% v/v additions, and there were only three 12% v/v additions.

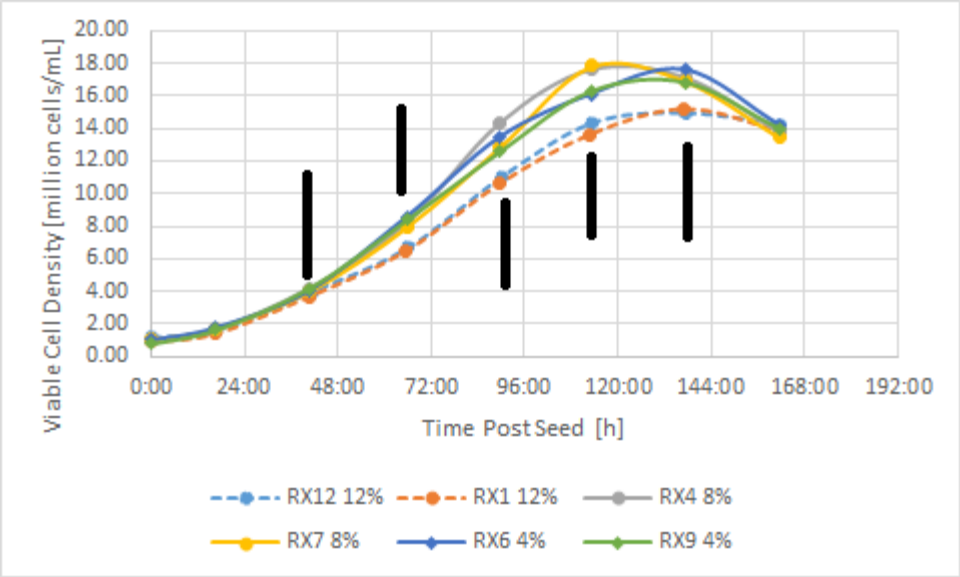

Figure 5.13 Viable Cell Density Plot, Experiment 3, Varying Multiple Feed Percentage

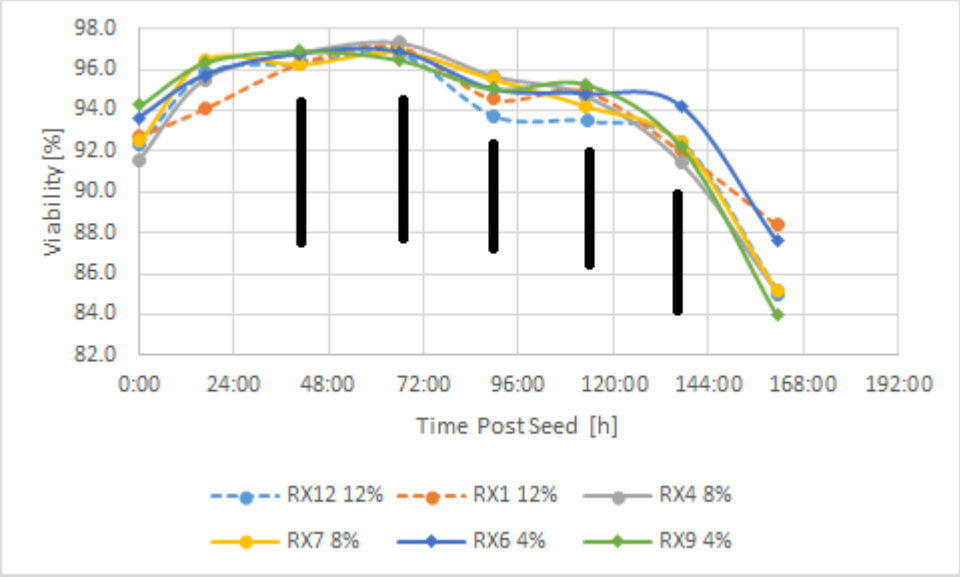

Figure 5.14 Cell Viability Plot, Experiment 3, Varying Multiple Feed Percentage

5.3.2. Experiment 3 Glutamine, Glutamate, Iron, Phosphate, and Ammonia

The most obvious outcome from multiple feeds was over-feeding of nutrients. As mentioned before, the 8% reactors received four feeds instead of five, and the 12% reactors received only three. Still, this was enough to sustain glutamine, glutamate, and iron levels in the reactor. In fact, both iron and glutamate accumulated in the 8% and 12% reactors. Figure 5.15 shows the glutamine concentration trend over time for the 4%, 8%, and 12% reactors. None of these bioreactors received glutamine supplementation.

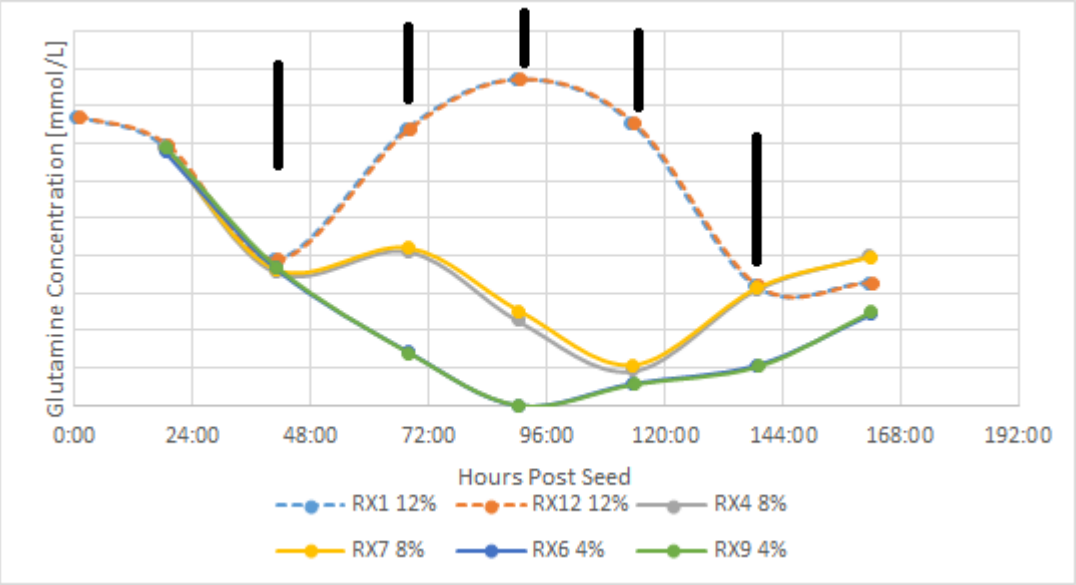

Figure 5.15 Glutamine Concentration vs. Time, Experiment 3

The 12%-fed reactors never dipped below the level that required glutamine supplementation, so the reactors that were planned to be supplemented ended up being additional duplicates for the 12% condition. Glutamine levels in the 4%-fed reactors depleted just before 4 days post seed, while the 8%-fed reactors did not have glutamine depleted for roughly another day.

Figure 5.16 shows how iron accumulated in the bioreactors, significantly so in the 12%-fed bioreactors. Glutamate also followed a similar pattern, so it was not included in this section. It is available for viewing in Appendix A.

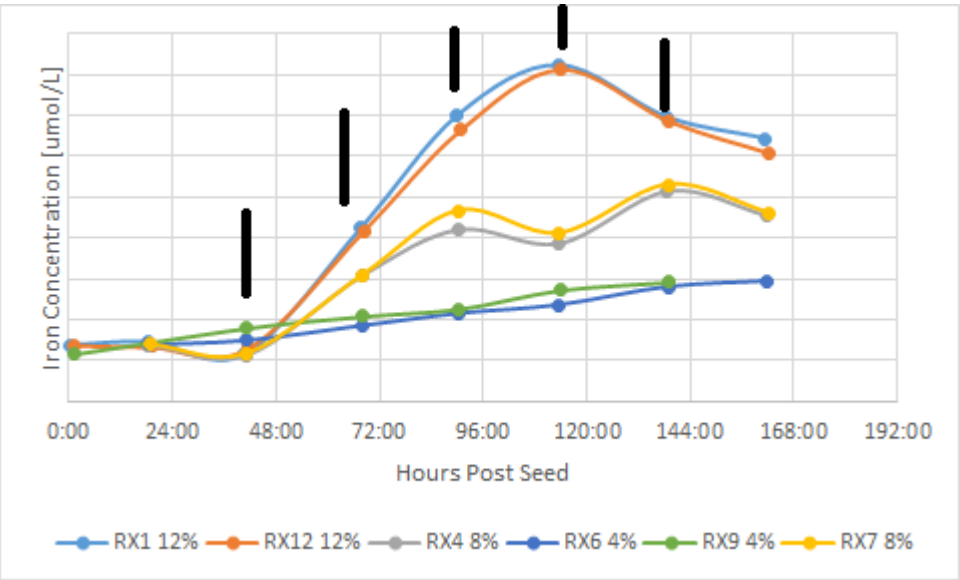

Figure 5.16 Iron Concentration vs Time, Experiment 3

Interestingly, phosphate was completely depleted in all reactors, with no gradual rise as exhibited in the single feed reactors. Phosphate is not replenished by the CFS. It is possible that the accompanied rise in pH, which is discussed in the next subsection, resulted in lower levels of extracellular phosphate. Nevertheless, the 4% reactors still exhibited viable cell growth after phosphate depletion, which suggests that it is not a limiting factor. As mentioned before, a side experiment of phosphate supplementation was performed in shake flasks to confirm this, and the results are available in Appendix A.

Also, unlike the previous experiment, there was significant accumulation of ammonia, especially in the 12%-fed bioreactors. Ammonia levels rose to over 25 mM in these cultures, which is enough to have a toxic effect on the insect cells [14]. Figure 5.17 shows the ammonia concentration over time for the non-glutamine supplemented bioreactors.

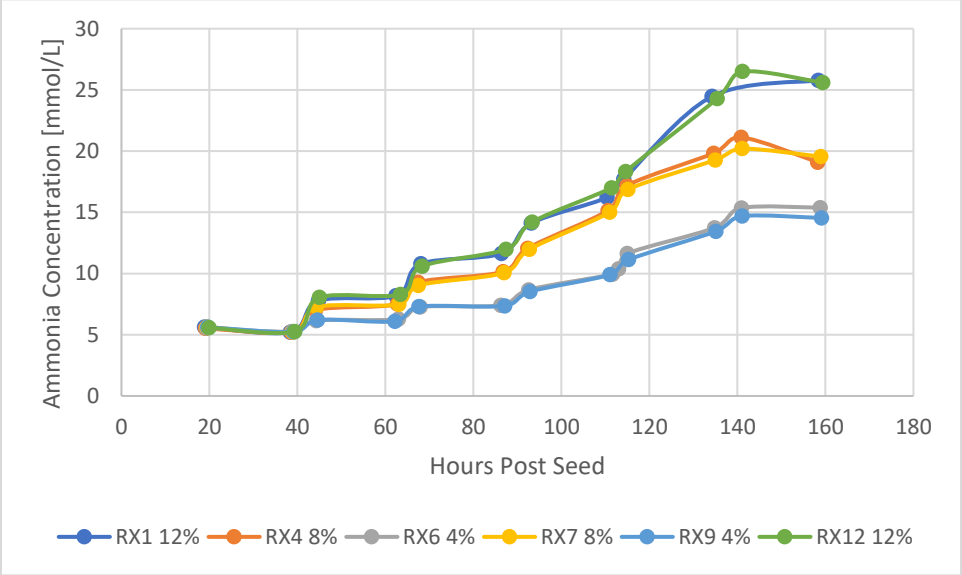

Figure 5.17 Ammonia Concentration vs. Time, Experiment 3

5.3.3. Experiment 3 pH and Osmolality

Experiment 3 revealed a problem with feeding the bioreactors multiple times with CFS: there was a precipitous rise in pH after the third feed. High pH is not conducive to significant protein production. By 120 hours post seed, all the bioreactors had risen above pH 6.7, whereas the pH is normally between 6 and 6.4. Feeding multiple times also had a negative effect on osmolality. Figure 5.18 shows the osmolality plot for Experiment 3.

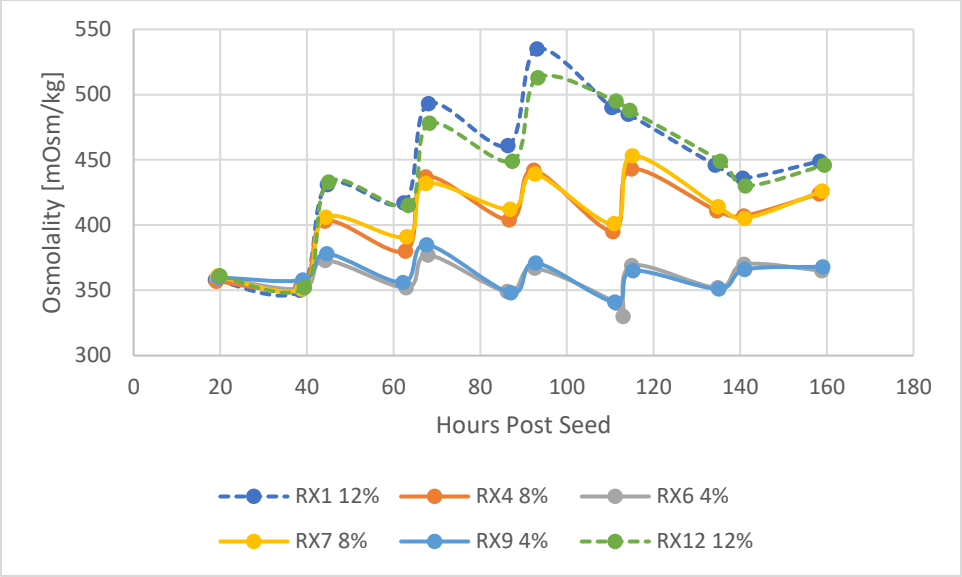

Figure 5.18 Osmolality vs. Time, Experiment 3

The osmolality of the 12%-fed reactors rose to well over 500 mOsm/kg, which is unacceptable. An osmolality at that level are sure to affect protein production, if not create a stressful environment for the cells. The 4%-fed reactors, on the contrary, were able to maintain an osmolality below 400 mOsm/kg.

5.3.4. Experiment 3 Discussion

The results from Experiment 3 show that multiple feeds did indeed almost double the viable cell density of the cultures. However, multiple feeds, especially after three feeds, introduced new problems that would hinder protein production. Multiple feeds increase the pH to problematic alkaline levels. pH control is an option; however, it was somewhat cumbersome to set up on the ambr system, so it was not explored in these experiments. Further, feeds with a volume percentage above 6% raise the osmolality to a concentration that would be stressful for cells. From an optimization standpoint, the 4%-fed reactors reached the highest peak cell density,

which is in stark contrast to the single-feed reactors. Supplementing glutamine did not seem to have a statistically significant effect on cell growth. One interesting result is that glutamate levels remain stable when glutamine is supplemented, but the insect cells produce glutamine from glutamate when it is depleted, as shown in Figure 5.19.

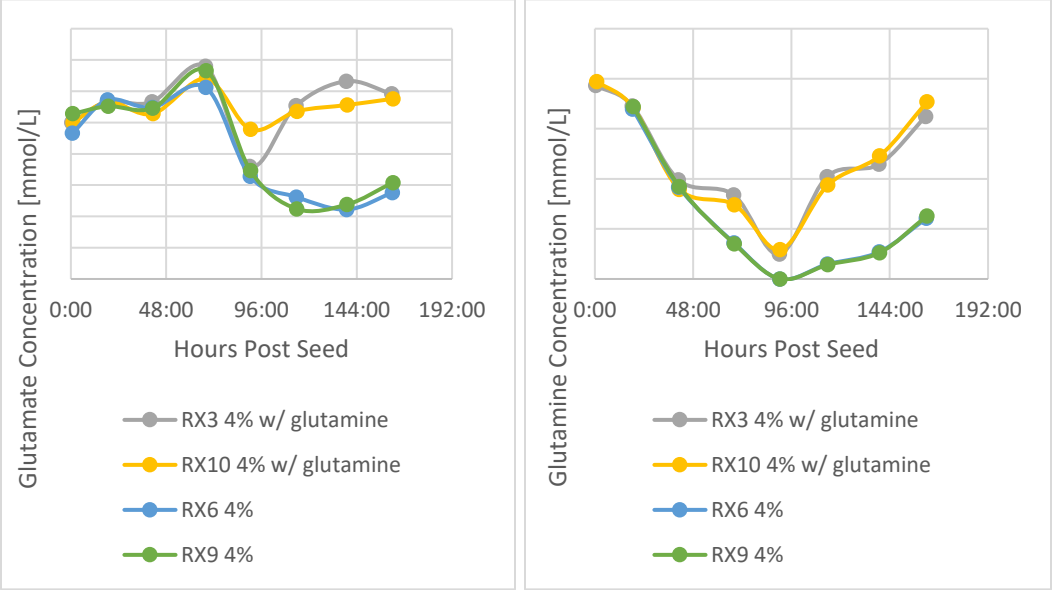

Figure 5.19 Glutamate and Glutamine Plots, 4% Feed Bioreactors

If one views the results in a vacuum, multiple feeds is indeed the way to maximize SF+ cell density. However, when one considers the overall goal of maximizing protein production, without pH control, multiple feeds create unsuitable conditions to produce drug substance. The focus of this work is to maximize SF+ cell growth, but one must appreciate the dualism that maximizing cell growth may affect specific protein productivity so much that overall yield is reduced.

5.4 Experiment 4 Semi-Continuous Feeding

The goal of the fourth experiment was to determine if there were any significant beneficial effects from replenishing feed in a semi-continuous mode rather than in a full bolus addition. The ambr system has a limit to how slowly it can pump per hour. Therefore, some of the semi-continuous feeding methods were accomplished by feeding "mini-boluses" every hour. For example, if the feeding schedule were 8% over eight hours, one would feed 1% at the start of every hour. Table 5.4 shows the experiment design for Experiment 4. Bioreactors 7 and 12 were used to investigate the effect of lowering the dissolved oxygen setpoint, which will be discussed in more detail in the next subsection.

Table 5.4 Experiment 4 Design Conditions

	Daily Feed Percentage (v/v)	**Feeding Duration**
Bioreactor 1	6%	6 hours
Bioreactor 2	8%	8 hours, every other day
Bioreactor 3	6%	6 hours
Bioreactor 4	6%	10 hours
Bioreactor 5	8%	8 hours
Bioreactor 6	4%	6 hours
Bioreactor 8	10%	6 hours
Bioreactor 9	6%	2 hours
Bioreactor 10	10%	10 hours
Bioreactor 11	12%	6 hours, every other day

5.4.1. Experiment 4 Discussion

Cell growth and viability charts from Experiment 4 showed no significant improvement to cell growth with semi-continuous feeding. For brevity, these graphs are not included in the main body of this work, but they are included in Appendix A. Semi-continuous feeding has a few benefits – it eliminates the spike in osmolality after feeding, and it allows one to specifically control the rate at which glucose and other nutrients are fed to the bioreactor. However, with semi-continuous feeding the pH increase issue was still present. Additionally, the feed temperature had to be carefully controlled to avoid the degradation of glutamine. From an engineering and manufacturing standpoint, semi-continuous feeding introduces variability into the process. It requires equipment such as pumps, increased maintenance costs, and potential round-the-clock shift coverage. Because the focus of this thesis is on both operations and maximizing cell growth, semi-continuous feeding was not investigated further. In this engineer's opinion, the benefits do not outweigh the extra variability and costs.

5.5 Experiment 5 Dissolved Oxygen Setpoint

The results of the Experiment 4 bioreactors with lower dissolved oxygen setpoints were promising, so investigating the dissolved oxygen setpoint became a major focus of the last experiment. Other changes explored were amino acid supplementation and pH adjustments, but these resulted in only negligible cell growth improvements. Thus, the focus of this section is primarily on the results of the bioreactors with lower dissolved oxygen setpoints. Three bioreactors were tested at two different dissolved oxygen setpoints. The summary of these reactors is listed in Table 5.5. Because there was an attempt to obtain as much information as possible in this final ambr experiment, the design of the experiment was somewhat subpar. In

hindsight, a control bioreactor with the original dissolved oxygen setpoint would have been included. The bioreactors in this experiment should also have been duplicated. Still, one could compare results on a relative basis to cultures from both Experiment 3 and 4. Finally, cascade control was used to control the dissolved oxygen setpoint for this experiment. A primary dissolved oxygen controller served as the setpoint for a secondary sparge oxygen percent controller. The secondary sparge oxygen percent controller, starting at an initial setpoint of 20% oxygen, increased its setpoint as oxygen needs of the bioreactor increased. The controller manipulated an actuator that increased the percentage of oxygen flow to the sparger while keeping the total gas flow constant. That way, less total gas was delivered to the bioreactor, reducing the amount of foam created. It is possible that this change in gas control impacted cell growth, so when comparing the results of bioreactors from this experiment to previous experiments, it is important to keep that in mind.

Table 5.5 Reduced DO% Setpoint Experiment Conditions

	Experiment	**Dissolved Oxygen Setpoint**	**Feeding Method**
Bioreactor 1	Experiment 5	Setpoint A	8% daily
Bioreactor 4	Experiment 5	Setpoint B	4% daily
Bioreactor 5	Experiment 5	Setpoint A	4% daily
Bioreactor 12	Experiment 5	Setpoint C	6% daily
Bioreactor 7	Experiment 4	Setpoint C	6% daily
Bioreactor 12	Experiment 4	Setpoint A	6% daily

For proprietary reasons, the exact dissolved setpoints are not disclosed, but they are in the range of 10%-75% DO. Setpoint C is greater than Setpoint A which is greater than Setpoint B.

5.5.1. Experiment 5 Dissolved Oxygen Cell Growth and Viability

Reducing the dissolved oxygen setpoint proved to improve peak viable cell density by 10-15%, especially for the 4%-fed reactors. Figure 5.20 and Figure 5.21 show the viable cell density and viability plots for bioreactors relevant to the dissolved oxygen experiment. Black lines indicate the approximate time of feeding.

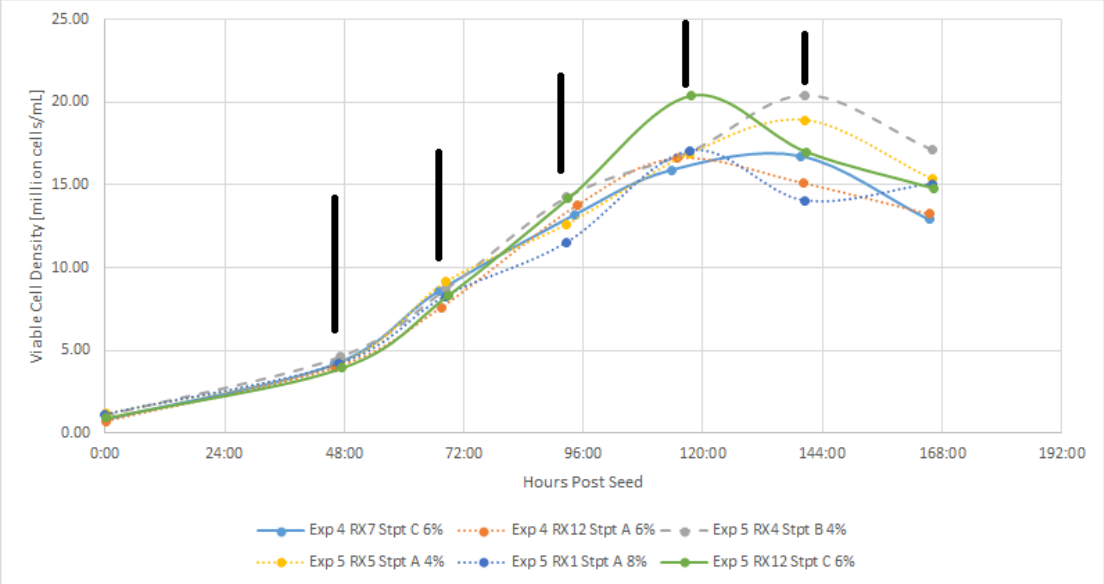

Figure 5.20 Viable Cell Density Plot for Reduced DO% Experiment

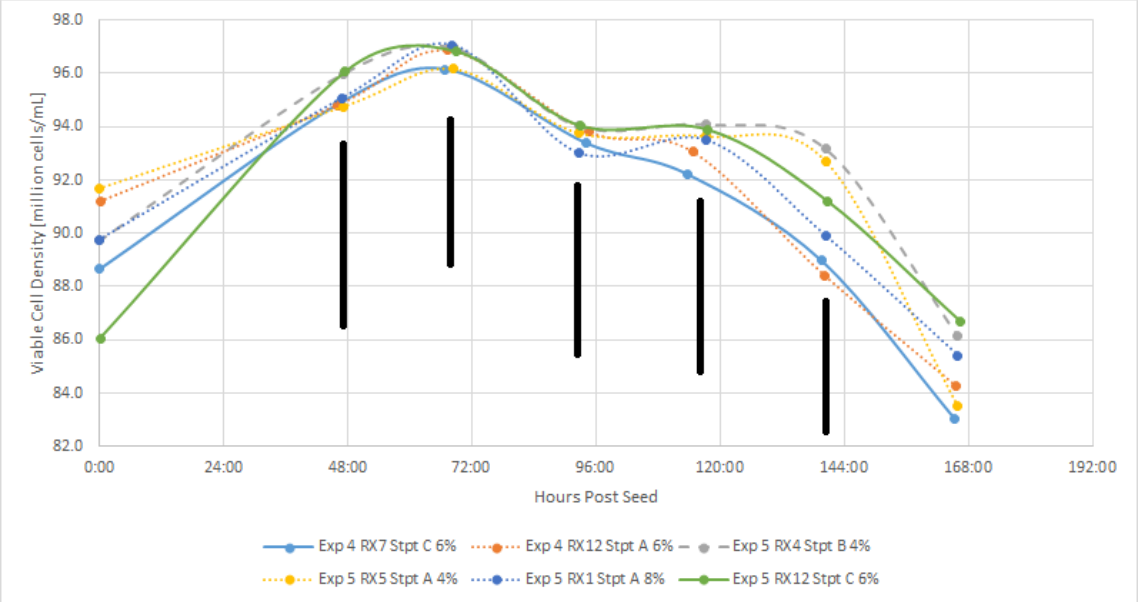

Figure 5.21 Viability Plot for Reduced DO% Experiment

While the cell growth of the reactor with setpoint C matched that of the bioreactors with lower dissolved oxygen setpoints, there is an obvious advantage of using less oxygen to support the cell culture. Because the bioreactors with lower DO% setpoints performed as well as the SF+ cell cultures, DO% was investigated in the subsequent experiments in 3L Applikon bioreactors.

5.6 Scale-up to 3L Bioreactor Confirmation Runs Preview

Growth of over 20 million cells/mL was found to be possible on the ambr system, which is similar to the cell density of Sf9 cultures supported in shake flasks [37]. From the results of the experiments conducted on the ambr system, several key factors were identified to maximize SF+ cell growth. First, lowering the dissolved oxygen setpoint had a positive effect on cell growth. Next, multiple feeds significantly boosted the growth of the cells, although more than three feeds had a negative impact on pH and osmolality. Semi-continuous feeding and glutamine supplementation were not found to have a profound effect on cell growth. The next action taken

was to reproduce the best conditions in 3L glass bioreactors, to see how well the ambr system scales to larger bioreactors. The best conditions from the five experiments are listed in Table 5.6.

Table 5.6 Summary of Top 5 ambr Experiment Results

Condition	**Oxygen Setpoint**	**Cell Density [million cells/mL]**	**Best Doubling Time [hr]**
12% feed every other day (three total feeds)	A	20.99	23.2
4% daily	B	20.41	22.2
6% daily with amino acid supplementation	A	20.38	19.7
4% daily with pH adjustment	B	20.06	22.0
4% daily	A	18.90	19.2

5.7 ambr Experiments: Glucose and Glutamine Consumption

The final section is intended to show the power of collecting and tracking data on amino acid and nutrient consumption, even in a lab setting. Ideally cells used to seed bioreactors are as identical as possible, but because cell cultures are "split" every few days for maintenance, cells that have been split only a few times will be slightly different than cells that have been split 30 to 40 times. Other issues can arise during research that may be difficult to spot like instrument failure or sparger blocking. These issues may cause results that one may misconstrue as being a legitimate experimental result. As an example, for Experiments 2 to 5 the initial glucose consumption rate (before any supplementary feed was added) and glutamine consumption rate were recorded for each bioreactor. The only variable that differed among the bioreactors was dissolved oxygen setpoint. Figure 5.22 shows a boxplot of pre-feed glucose consumption rate and Figure 5.23 shows a boxplot of the pre-feed glutamine consumption rate. These boxplots compare the rates between bioreactors with DO% setpoint A and DO% setpoint C. A t-test rejects the alternative hypothesis that there is a difference between the mean initial glucose consumption rates of bioreactors with setpoint A versus setpoint C (*p-value* 0.5167). However, for glutamine consumption rates, a t-test suggests there is a difference between the mean initial glutamine consumption rates of bioreactors with setpoint A versus setpoint C (*p-value* 2.632e-09). In short, collecting and analyzing nutrient consumption data can be a useful tool for troubleshooting or understanding changes in a cell line's metabolism as experimental conditions change.

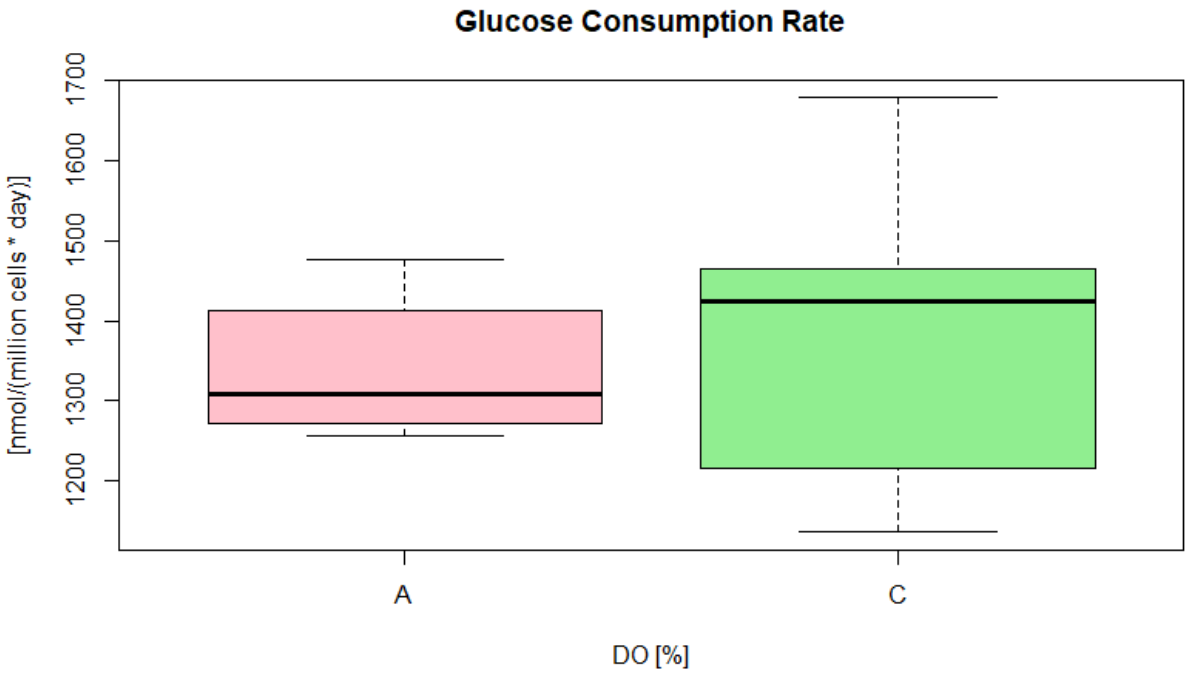

Figure 5.22 Pre-feed Glucose Consumption Rate DO% A vs. C

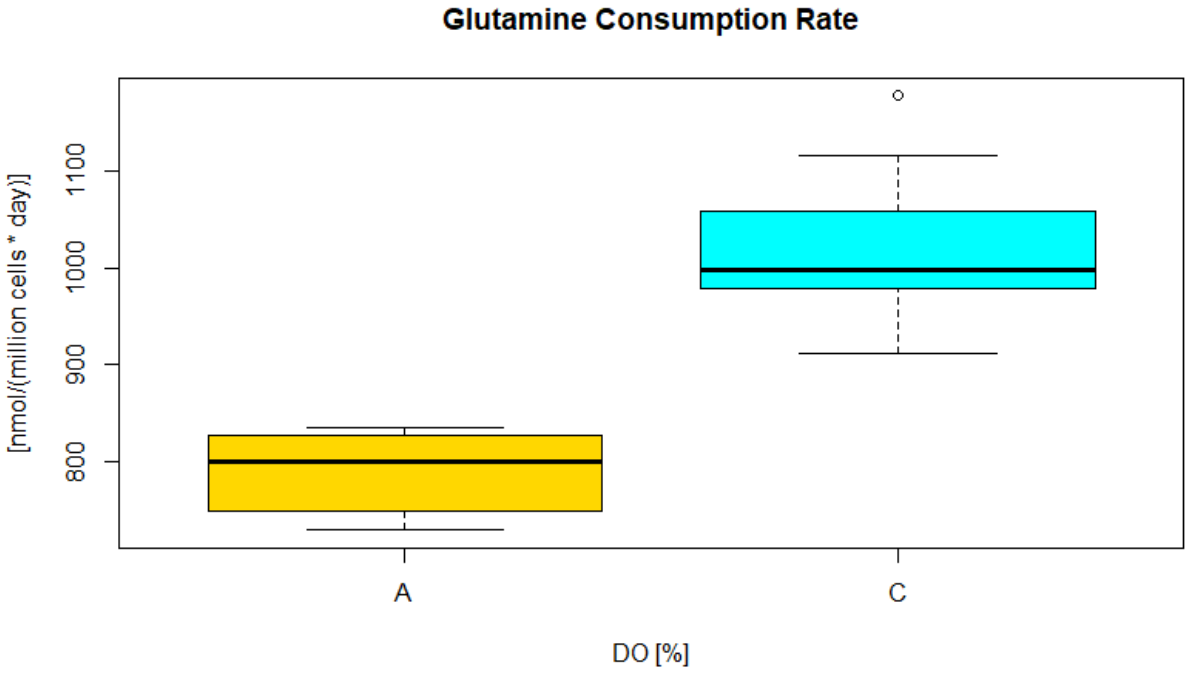

Figure 5.23 Pre-feed Glutamine Consumption Rate DO% A vs. C

6. 3L Confirmation Runs Methodology

Confirmation runs of the best conditions were carried out on benchtop 3L bioreactors at Protein Sciences in Meriden, CT. The cell line is the same line used as the ambr experiments. The Vi-Cell analysis is also the same, but there was no access to a Cedex Bio HT nor a FLEX2 machine. Glucose, lactate, and ammonia concentration, osmolality, dissolved CO_2, and pH could be measured immediately using a BioProfile FLEX analyzer.

6.1 Bioreactors and Equipment

3L glass bioreactors were used in this study. Agitation was provided by two scoping marine impellers. The lower impeller was placed 2 cm above the sparger, and the top impeller was placed just below the gas-liquid surface. Culture parameters (dissolved oxygen, temperature, agitation, and gas flow rates) were controlled using an ACC BioStation loaded with Rockwell Factory Talk software. All setpoints were maintained using PID controllers within the Rockwell PLC. Gas flow rates were measured by mass flow controllers, temperature was measured by a Weed Instruments temperature probe, and dissolved oxygen was measured with a Mettler Toledo polarographic probe. An overlay flow was controlled using a rotameter, with enough air added to the headspace of the reactor to strip carbon dioxide. Cell transfer, cell seed, and CFS feeding were performed aseptically using a welder/fuser (G.E. Healthcare Life Sciences). Figure 6.1 shows an example benchtop 3L bioreactor [38].

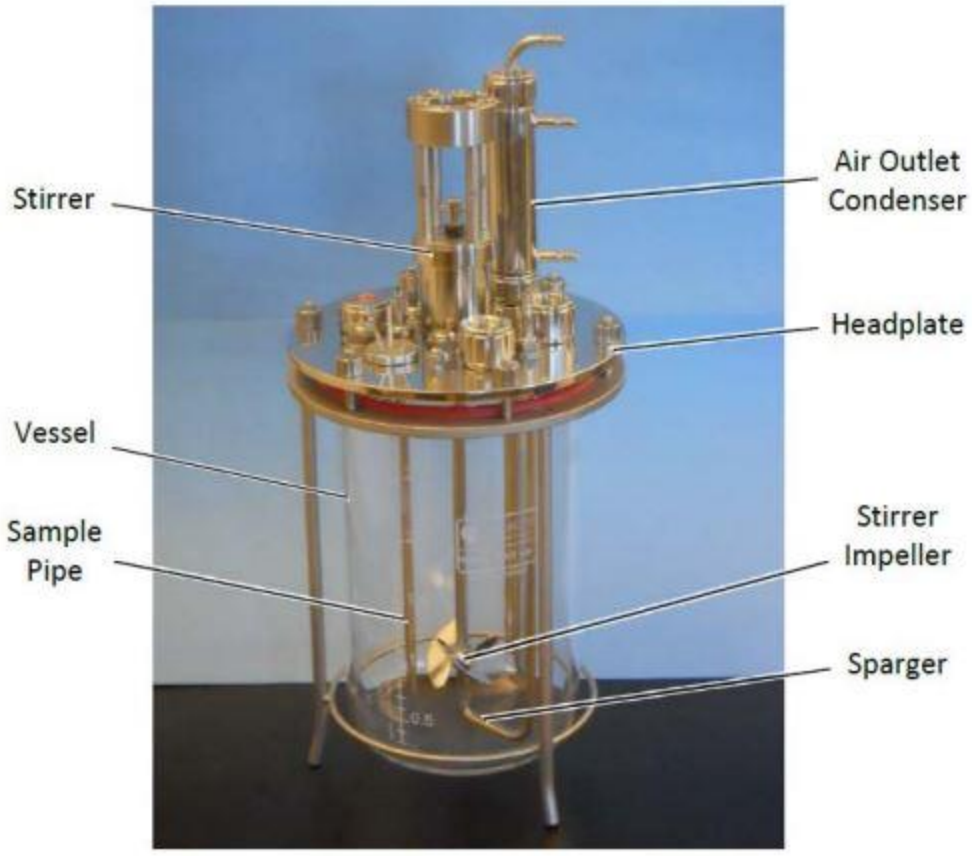

Figure 6.1 Applikon 3L Benchtop Glass Bioreactor

7. 3L Confirmation Runs Results and Discussion

Table 6.6 shows the top five viable cell densities achieved on the ambr system. There were several conditions to choose from to reproduce. Due to limited time and reactor availability, one condition was chosen to be run in duplicate. Then, the feed percentage and dissolved oxygen were manipulated to compare the results from the ambr experiments to the results in the 3L bioreactors. Table 7.1 summarizes the conditions of the confirmation run experiments.

Table 7.1 3L Confirmation Run Conditions

	Oxygen Setpoint	Feed Method (v/v%)
Confirmation Run 1	A	4% Daily
Confirmation Run 2	A	4% Daily
Confirmation Run 3	A	4% Daily
Confirmation Run 4	B	4% Daily
Confirmation Run 5	A	6% Daily

7.1 Three-Liter Confirmation Runs Cell Viability and Density Plots

Overall, the transition from the ambr bioreactor process to the three-liter bioreactors went smoothly. The peak viable cell density for the 4%-fed reactor at oxygen setpoint A was 18.90 million cells/mL. The average peak viable cell density for 4% three-liter Applikon experiments was 23.39 million cells/mL, a 24% improvement. Cell growth in three-liter benchtop bioreactors likely improved due to better oxygen transport. Again, for reference, the k_La in the larger bioreactors was around 17 h^{-1} as opposed to 5 in the smaller ambr bioreactors. Figure 7.1 compares the viable cell density for the 4% bioreactors at dissolved oxygen setpoint B. Figure 7.2 shows the viability plots for the same reactors.

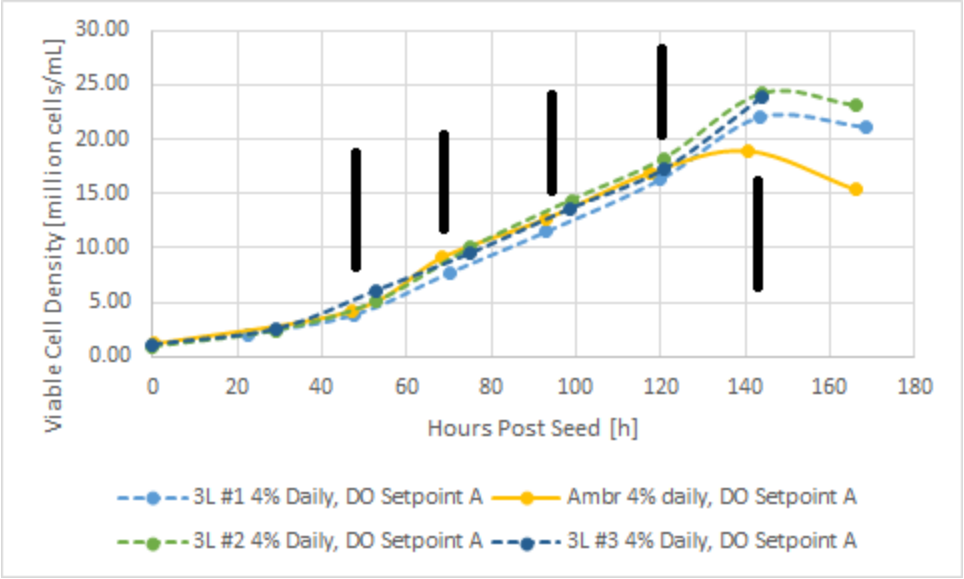

Figure 7.1 Viable Cell Density Plot, Confirmation Runs, 4% Fed Confirmation Runs vs. Ambr

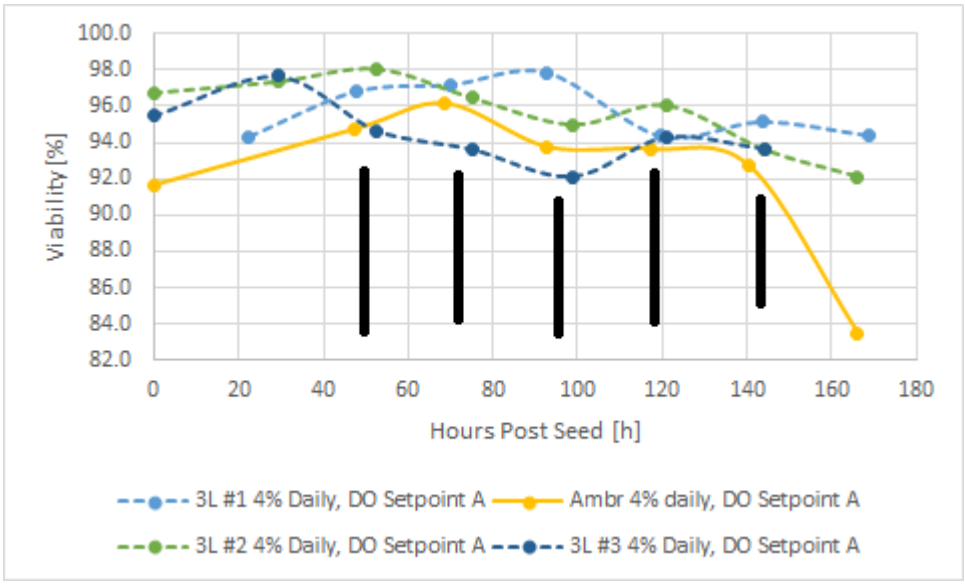

Figure 7.2 Viability Plot, 4% Fed Confirmation Runs vs. Ambr

Figure 7.1 and Figure 7.2 demonstrate the better cell growth performance of SF+ cells in the 3L bioreactors as opposed to the ambr system. The confirmation run cultures reached higher

peak cell densities and died less quickly after the final feed. The process established on the ambr system transitioned nicely to the benchtop scale. This shows the promise of using an ambr system in conjunction with larger bioreactors to save time when scaling up a bioprocess. This will be discussed in more detail in future sections.

The next two figures compare a 3L bioreactor that was fed 4% daily for five feeds and a 3L bioreactor that was fed 6% daily for five feeds. This was to see that, as with the ambr experiments, the 4% daily fed cultures out-performed the higher-fed cultures. As Figure 7.3 and Figure 7.4 show, peak total cell density was indeed higher for the 4% fed reactor, and there was no significant difference in viability between the two bioreactors.

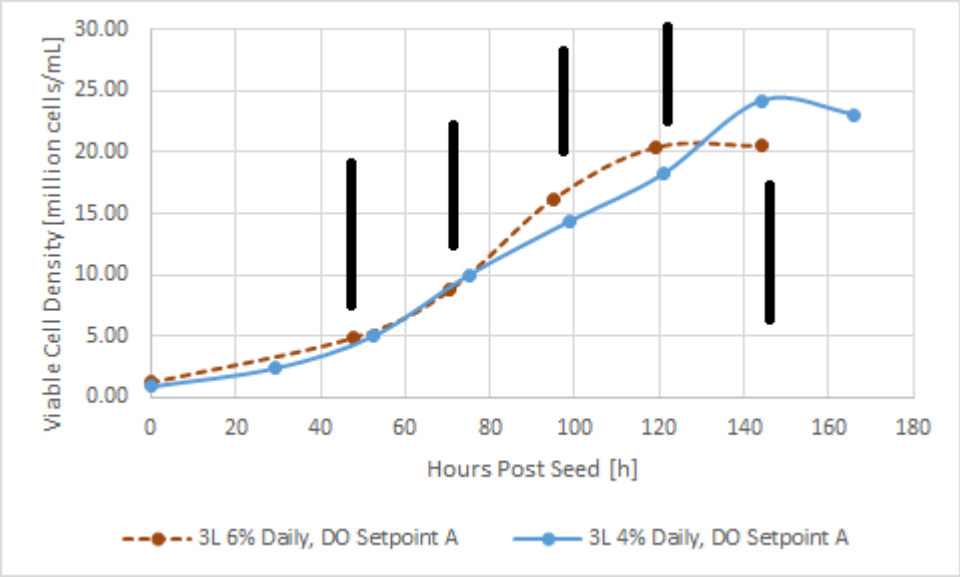

Figure 7.3 Viable Cell Density Plot, Confirmation Runs, 4% Daily Feed vs. 6% Daily Feed

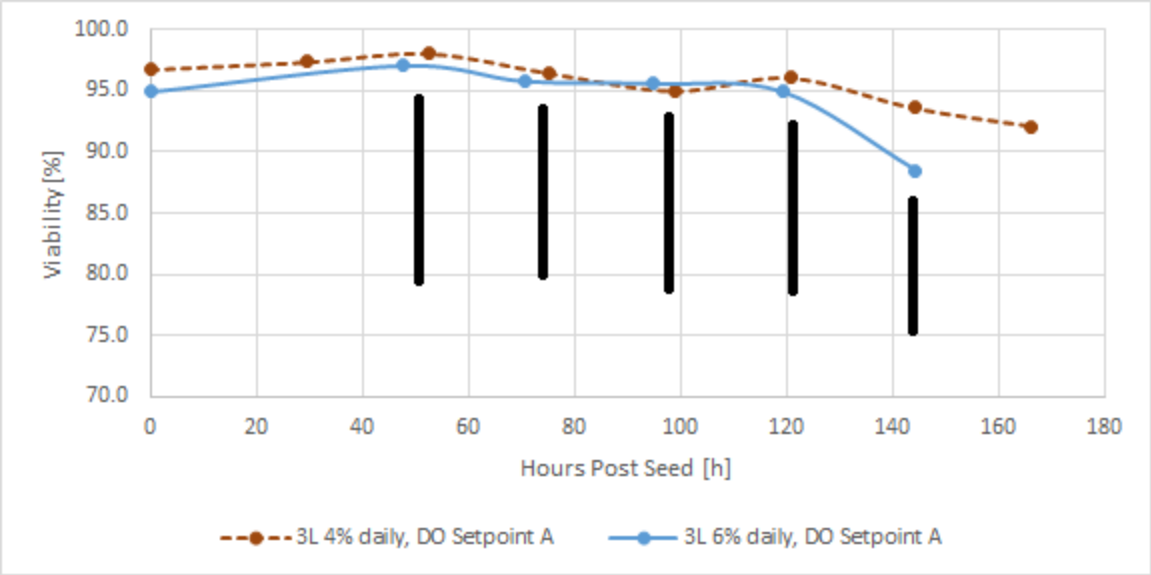

Figure 7.4 Viability Plot, Confirmation Runs, 4% Daily Feed vs. 6% Daily Feed

The final viable cell density and viability plots, Figure 7.5 and Figure 7.6 compare the reactors with two different oxygen setpoints. These setpoints are the same ones tested in Experiment 5 on the ambr bioreactors. To recall, Setpoint B is lower than Setpoint A, in the range of 10-75 % dissolved oxygen. Both are lower than Setpoint C, which was the default dissolved oxygen setpoint and not tested on the 3L bioreactors.

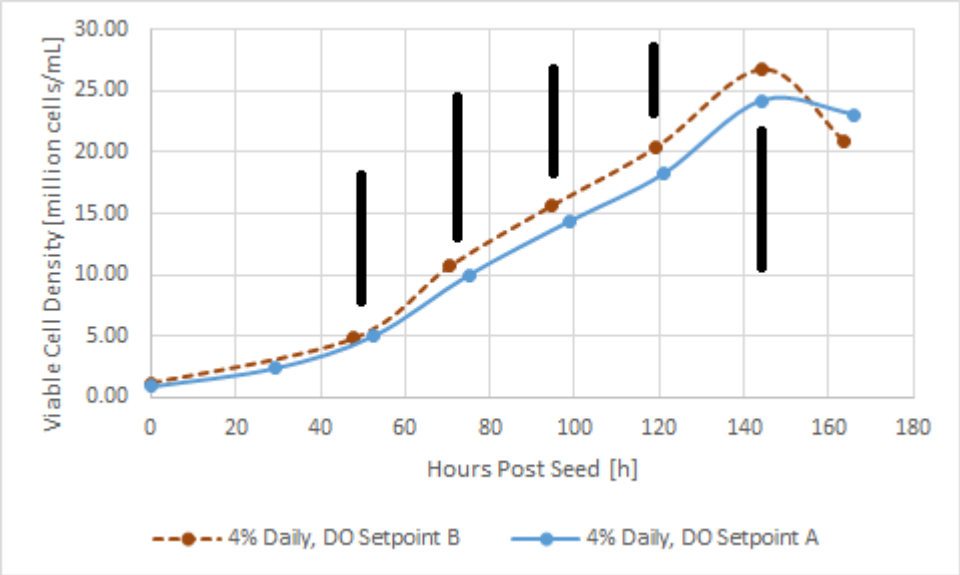

Figure 7.5 Viable Cell Density Plot, Confirmation Runs, DO Setpoint A vs. Setpoint B

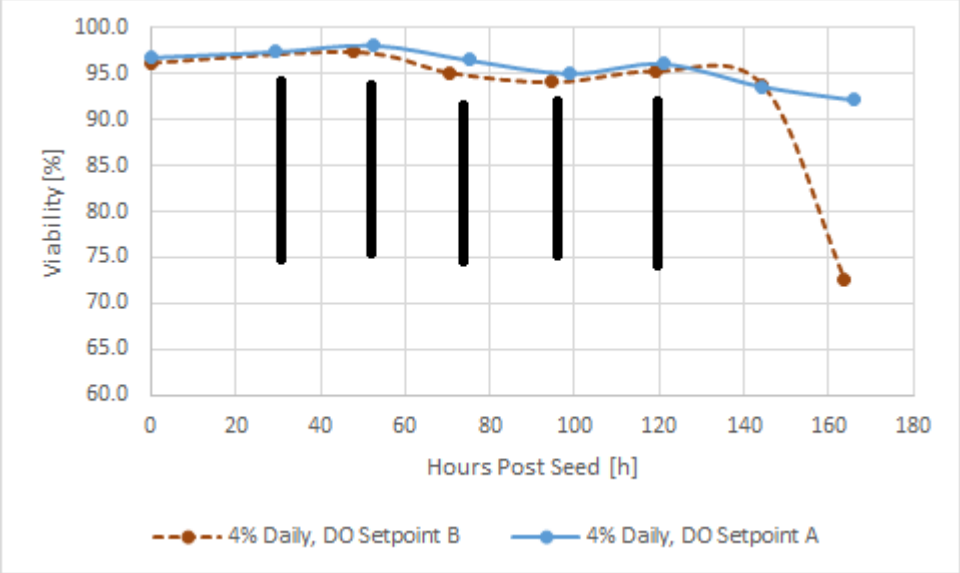

Figure 7.6 Viability Plot, Confirmation Runs, DO Setpoint A vs. Setpoint B

Here, lowering the dissolved oxygen setpoint from Setpoint A to Setpoint B improved the peak viable cell density by nearly 12%. However, it is obvious from the viability chart that the cells died off much more quickly when the dissolved oxygen setpoint was lower.

7.2 3L Confirmation Runs Discussion

Results from the 3L confirmation runs demonstrate that the ambr system does indeed scale very well to benchtop reactors. For the SF+ or Sf9 cell line, one can expect a 15-25% relative cell growth improvement from the ambr system to the benchtop from better oxygen transport conditions alone. It is difficult to say if a local optimum was achieved with these experiments, for this base and feed media. It's extremely unlikely to be a global optimum for SF+ cells, considering Sf9 cultures of up to 50 million cells/mL have been reported [39]. This highlights just how many variables can be manipulated to change cell growth performance – media, oxygen transport, temperature, pH, etc. Some of these variables were not tested due to time constraints. With twelve parallel bioreactors, the ambr system can enable researchers to get much more creative with their experimental design and expand the horizon of the variables tested.

With the current base medium (PSFM) and feed medium (CFS), it appears that the two variables that had the most significant influence were daily feed percentage (v/v %) and dissolved oxygen (%). It is likely that even higher cell densities would be achieved with improvements to media. However, in a bioprocess that has already been approved by the FDA, changes to media can be both timely and costly. These considerations are discussed in more detail in later sections, where we apply the learnings from this experiment to real-world manufacturing technology scale-up organizations. These organizations must be swift in bringing

new drug products to market, while ensuring that the drug product is of the utmost quality, the bioprocess is safe, and the process is optimized as much as possible. The ambr system can be a powerful tool in achieving these objectives and doing it right on the first pass.

Table 7.2 compares results from the ambr system to the results of the confirmation runs. In short, the best reactors were the ones with the lowest dissolved oxygen setpoint and a 4% daily feed.

Table 7.2 Comparison of ambr Results to Confirmation Runs

	Oxygen Setpoint	**Feed Method (v/v%)**	**Viable Cell Density (million cells/mL)**
Confirmation Run 1	A	4% Daily	22.04
Confirmation Run 2	A	4% Daily	24.19
Confirmation Run 3	A	4% Daily	23.93
Confirmation Run 4	B	4% Daily	26.83
Ambr Run 1	C	4% Daily	16.82
Ambr Run 2	C	4% Daily	17.58
Ambr Run 3	A	4% Daily	18.90
Ambr Run 4	B	4% Daily	20.41

In the next chapter, results from both experiments are used in conjunction with relatively straightforward regression techniques to attempt to predict cell density. Hopefully, the section will reveal the potential of combining machine learning with good experimental design to achieve desired results more quickly and with less hands-on effort.

8. Development of Viable Cell Density Predictive Model

A powerful aspect of the ambr bioreactor system is the ability to generate a large dataset which can be used to predict a dependent variable of choice, such as peak viable cell density, protein yield, or peak doubling time. When doing experiments in larger bioreactors, it can take months, tying up multiple scientists and reactor spaces. With the ambr system, one can collect twelve sets of data at once in one to two weeks, with only one to two scientists required to run the system. Assuming an average number of six 3L bioreactors, to complete the same number of runs would require up to four weeks and roughly four to five workers (enough to split cell cultures, set up reactors, maintain reactors, and clean). While there are many linear regression methods, lasso regression and random forest regression were chosen to create models for this work. Lasso regression was chosen because in general, it produces simple, sparse models. Random forest regression was chosen because for most datasets, it produces a highly accurate model, and it also gives estimates of what variables are important in the estimation. In the next few sections, the results of these models are presented.

8.1 Variable Selection

One of the first things to do when creating a regression model is to determine which variables will be explanatory. In manufacturing, one may be given a dataset and not know which variables, or features, to include. But in this case, it is easy to start with the variables that were manipulated in the experiments, like dissolved oxygen percentage and feed percentage. The advantage to this is that there is some scientific and engineering basis in the model parameters, but the disadvantage is that there might be other lurking variables that have a strong effect on the response variable. That is why experimental design and controlling for confounding variables is

so important. While precautions were taken for all the experiments of this work, some things were inevitably hard to control. Examples include the viable cell density at time of first feed, the seed cell density, and minor physical differences between reactor vessels. However, tolerance to variability is a good thing in a model, especially when it will be applied to a manufacturing process. At that large of a scale, it is almost impossible to control every source of variability. The continuous variables that were selected as independent variables for the model were DO%, Feed 1 % (v/v), Feed 2 % (v/v), Feed 3 % (v/v), Feed 4% (v/v), and k_La (h^{-1}). The oxygen transport coefficient k_La was incorporated to allow some engineering comparison between the ambr bioreactors and the 3L bioreactors. Additionally, it helped reduce the number of variables by combining agitation rate and sparge volumetric flow rate. Binary variables that were included were glutamine supplementation (true/false), glucose supplementation (true/false), semi-continuous feeding method (true/false), and amino acid supplementation (true/false). The dependent variable was maximum viable cell density (million cells/mL).

8.2 Linear Regression Considerations

In multiple linear regression, standardizing variables is a necessary practice to bring all variables to the same scale. The method of z-scoring, or standardizing a variable, is straightforward:

Equation 8.1 Method of Standardizing Sample Variables

$$z = \frac{x - \bar{x}}{s}$$

Here, $\bar{x}$ is the sample mean and s is the sample deviation. Because only three levels of dissolved oxygen were tested, an assumed mean and standard deviation were used to standardize these variables. The same was also done for the feed percentage variables. The sample mean and

standard deviation were used to normalize the k_La variable and viable cell density. Another condition to consider in linear regression analysis is homoskedasticity. Homoskedasticity refers to a state in which the variance of the residual, or error term, in a regression model is constant. If the variance of the error term is not homoscedastic, the model may be poorly defined. Finally, multiple regression deals with models that are linear in the features. These features do not necessarily have to be linear themselves, they could be the square of an experimental variable, the $\log_{10}$ of the variable, or the cosine of the variable. For the purposes of this thesis, the requirements for the regression methods that follow are assumed to be met. The performance of the models generated is likely to hold under the conditions which the data was collected, i.e. small bioreactors 250 mL to 3 L in volume. Critique of the models developed will continue in the discussion section of this chapter.

8.3 Lasso Regression Model

The first regression model created was a Lasso regression model. In the subsections that follow, a brief background, a pseudocode algorithm, and model results are given.

8.3.1. Lasso Regression Background

Lasso (least absolute shrinkage and selection operator) regression is a sparse model method – a model with high predictive capability that assumes that many model parameters are equal to zero. The method looks for weighted predictors that solve the following equation[40]:

Equation 8.2 Optimization Problem for Lasso Regression

$$\min_{\theta}(Y - A\theta)^T(Y - A\theta) + \alpha \sum_{i=1}^{n} |\theta_i|$$

Here, θ are the weights determined by solving the above, α is the penalty factor, Y is a vector of the dependent variable, and A is a matrix of predictor variables. The penalty is equivalent to the sum of the magnitude of the coefficients [40]. Lasso regression is a simple technique to reduce complexity and prevent over-fitting. Sparse models are also useful for interpretability, which is critical when scaling up a process, because one wants to find the factors that have a significant effect on the dependent variable of interest. Lasso also is tolerant of multicollinearity, or a state where predictor variables are linearly related, or codependent [41]. In this case, the predictor variables are not codependent because they are controlled variables of the experiments.

8.3.2. Lasso Regression Algorithm

The Lasso regression for this work was performed in MATLAB. The *lasso* command returns fitted least-squares regression coefficients for the linear models of the predictor matrix A and the response Y [42]. It is best practice to use a training, test, and validation set for regression techniques, and a 70-15-15% split is a good standard to use. In this work, there were not enough datasets to justify splitting into three datasets. Therefore, the data was split 60-40% into a training and test set. The model was trained using the training set and the minimum error was selected with the validation set. In the lasso command, a 7-fold cross validation was used to set the penalty. The pseudocode for the lasso regression method is given on the next page.

Algorithm: Use Lasso regression to predict viable cell density

Input: Predictor data and viable cell density results

Output: Best-fit parameters

Model performance metrics

%% Data Manipulation

Input → Load Data

Split Data into training, test, and validation sets

Normalize Data

%% **Model Training**

[Lasso Parameters, Lasso Fit] = lasso (X_train, Y_train,'CV',k)

y_hat_lasso = x_test*Lasso Parameters

for each y_hat_lasso

Compute error between predicted and actual values

end

Select minimum error

Output → Select Best Lasso Parameters based on minimum error

%% Model Metrics

Output → Compute RMSE of test and training set

8.3.3. Lasso Regression Model Results

Figure 8.1 shows a plot of the actual viable cell density of the training and test sets versus the predicted viable cell density. The reported RMSE for the training set was 4.6 million cells/mL and 5.00 million cells/mL for the test set. Figure 8.2 shows the residual plots for both the training and test sets. A residual is the difference between the actual value and the predicted value.

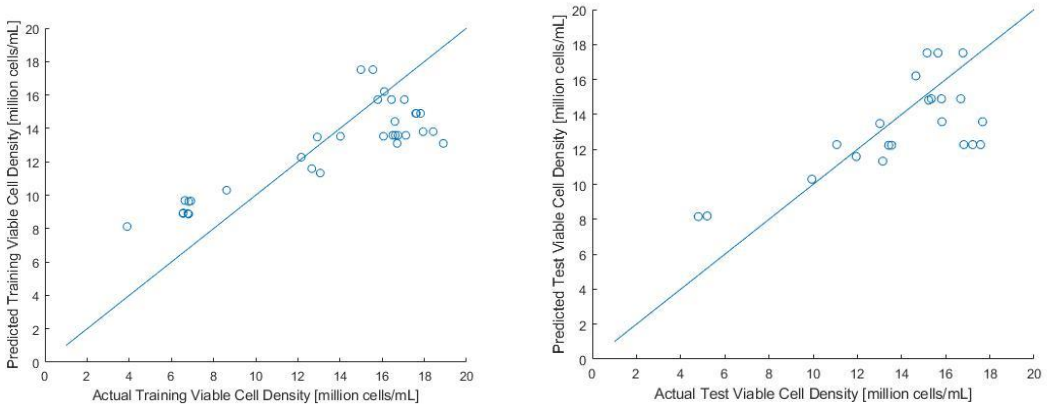

Figure 8.1 Actual vs. Predicted Viable Cell Density, Lasso Regression Model

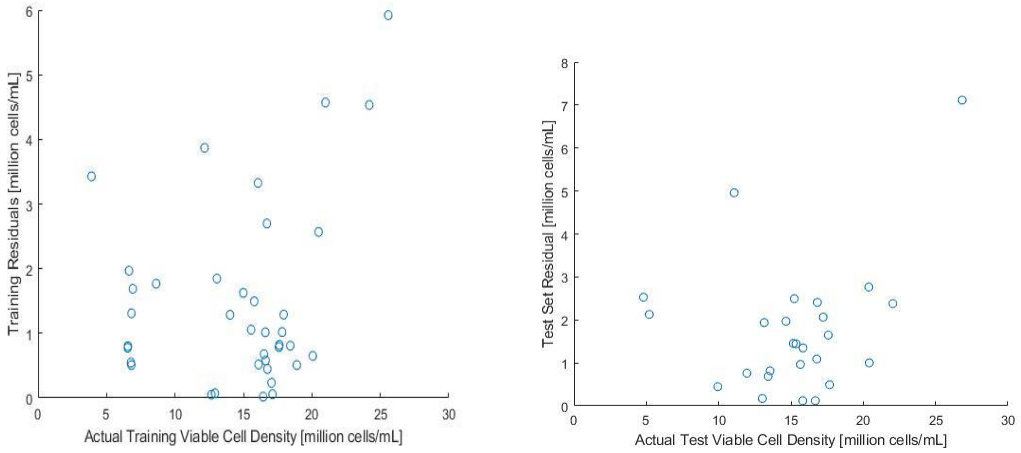

Figure 8.2 Residual Plots for Lasso Regression Model

The model generated from the Lasso regression algorithm tends to overestimate viable cell density at the low end and underestimate the viable cell density at the high end. This could be because there were not enough experiments that produced results at the low and high end of the dependent variable. Further, an RMSE that is roughly 20-40% of the predicted variable of interest is not an outstanding result. Despite the poor result, one can evaluate the best parameters selected by the algorithm. The predictors selected by the model were dissolved oxygen, feed 1 percentage, feed 3 percentage, and k_La. The parameters for all other predictors were zero, and therefore were ignored by the model. Biologically, the parameters selected by the model are reasonable. Feed 1 and 3 percentage are important, and dissolved oxygen and k_La were shown to be important from the results presented in the previous sections. Furthermore, the coefficient for dissolved oxygen percentage was negative, meaning that a lower DO% led to a higher maximum viable cell density. This is what was found from the ambr and three-liter experiments. Lastly, all the binary predictor variables were ignored from the model. This was also the same conclusion from the ambr experiments. While the results of the model are not perfect, a reasonable model was obtained. That model is given by Equation 8.3.

Equation 8.3 Viable Cell Density Prediction from Lasso Regression Model

$$\widetilde{VCD}\ \left[\frac{million\ cells}{mL}\right] = A * \left(\frac{DO\% - \bar{x}}{s}\right) + B * \left(\frac{F1\% - \bar{x}}{s}\right) + C * \left(\frac{F3\% - \bar{x}}{s}\right) + D * \left(\widetilde{k_L a}\right)$$

8.4 Random Forest Regression Model

The next regression model used was a random forest model. This is different than other regularization methods such as lasso, ridge regression, and elastic net. Random forest models utilize decision trees, and the background of the method will be covered in the next subsection.

8.4.1. Random Forest Regression Background

Random forest regression models are based on decision trees. At each node in the decision tree, a separation is made based on a criterion such that the resulting groups are as different from each other as possible. Figure 8.3 shows a simple example of a decision tree [43].

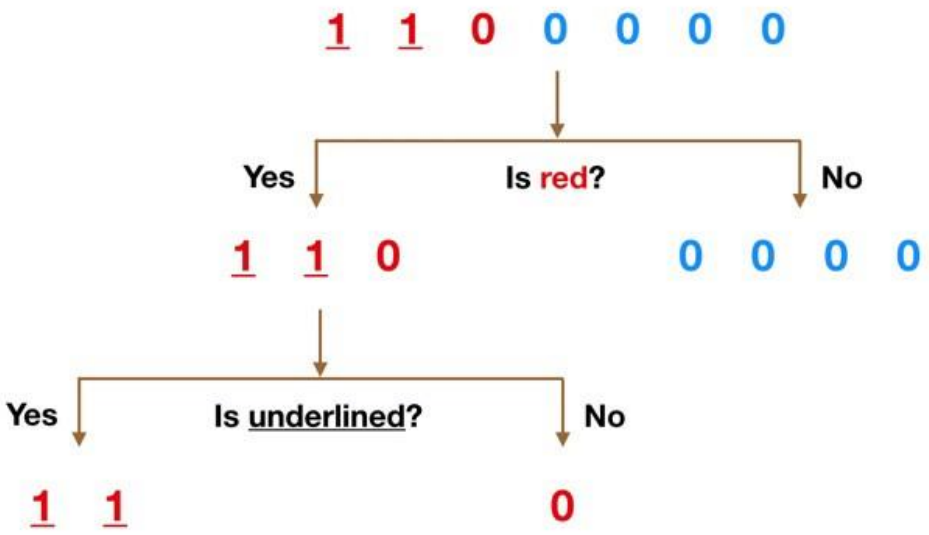

Figure 8.3 Simple Example of a Decision Tree

Random forest is comprised of many individual decision trees that function as an ensemble. Each individual tree in the random forest comes up with a prediction, and the classification with the most votes becomes the model's prediction [43]. Bagging is a technique for reducing the variance of a prediction function, and it works well for methods such as decision

trees. Random forest methods build a large collection of de-correlated trees and averages them. The idea is to average as many noisy but unbiased models as possible to reduce variance [44]. Two methods of evaluating random forest models are out-of-bag samples and variable importance plots. Out-of-bag error is the mean prediction error on each sample using only trees that did not have that sample in their bootstrap sample. Variable importance plots measure the prediction strength of each variable [44].

8.4.2. Random Forest Regression Algorithm

Random forest algorithms are accomplished in MATLAB using the *Treebagger* command. The features OOBPrediction and OOBPredictorImportance store information on the out-of-bag observations for each tree and out-of-bag estimates of feature importance, respectively [45]. An ensemble of fifty trees was chosen for this algorithm, which should be an acceptable threshold for this dataset, beyond which there is no performance improvement[46]. Pseudocode for the random forest regression model is provided on the next page.

Algorithm: Use Random Forest regression to predict viable cell density

Input: Predictor data and viable cell density results

Output: Model performance metrics

%% Data Manipulation

Input → Load Data

Split Data into test and training sets

%% **Model Training**

```
Tree Model =
TreeBagger (50,Training Data,'MaxVCD', 'Method','regression',…
 'OOBPrediction','On','OOBPredictorImportance','On')
```

%% Model Metrics

Tree Output = predict (Tree Model, Test Data);

Output → Compute RMSE for Training and Test Data

8.4.3. Random Forest Regression Model Results

Because of overfitting to the training set, plots of training data are not included in the results. Figure 8.4 shows a plot of actual test set viable cell density versus predicted test set viable cell density.

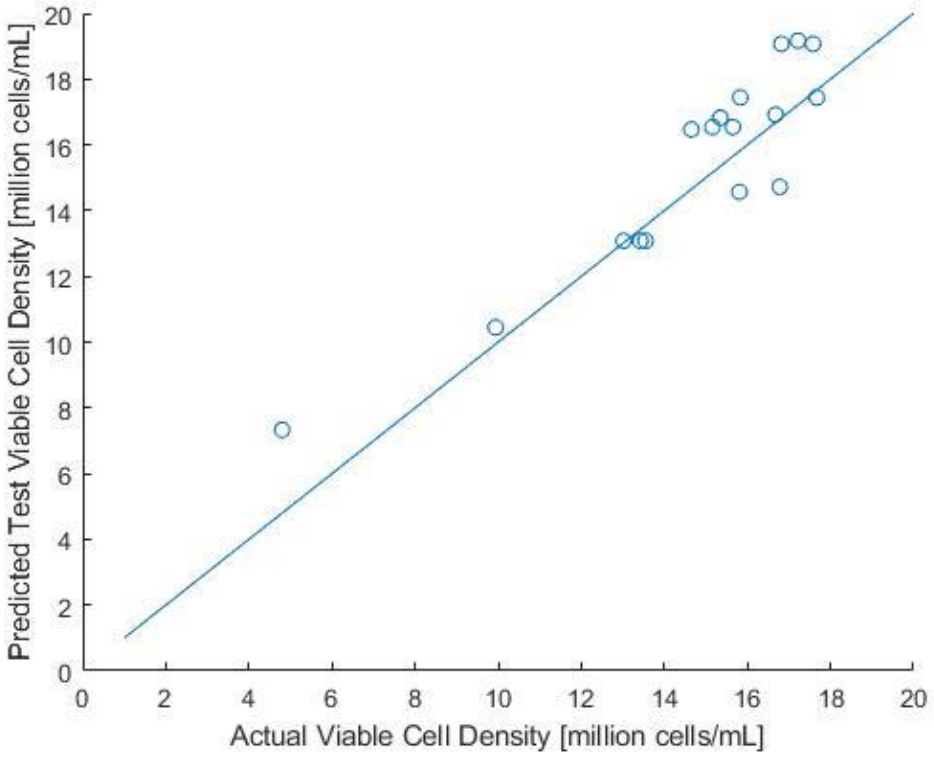

Figure 8.4 Actual vs. Predicted Viable Cell Density, Random Forest Regression Model

The RMSE of the test set was reported as 2.19 million cells/mL, which improves upon the lasso regression RMSE of the test set by nearly 3 million cells/mL. In this case, it appears random forest regression is a better method to predict peak viable cell density, but it would be wise to confirm this with additional experiments.

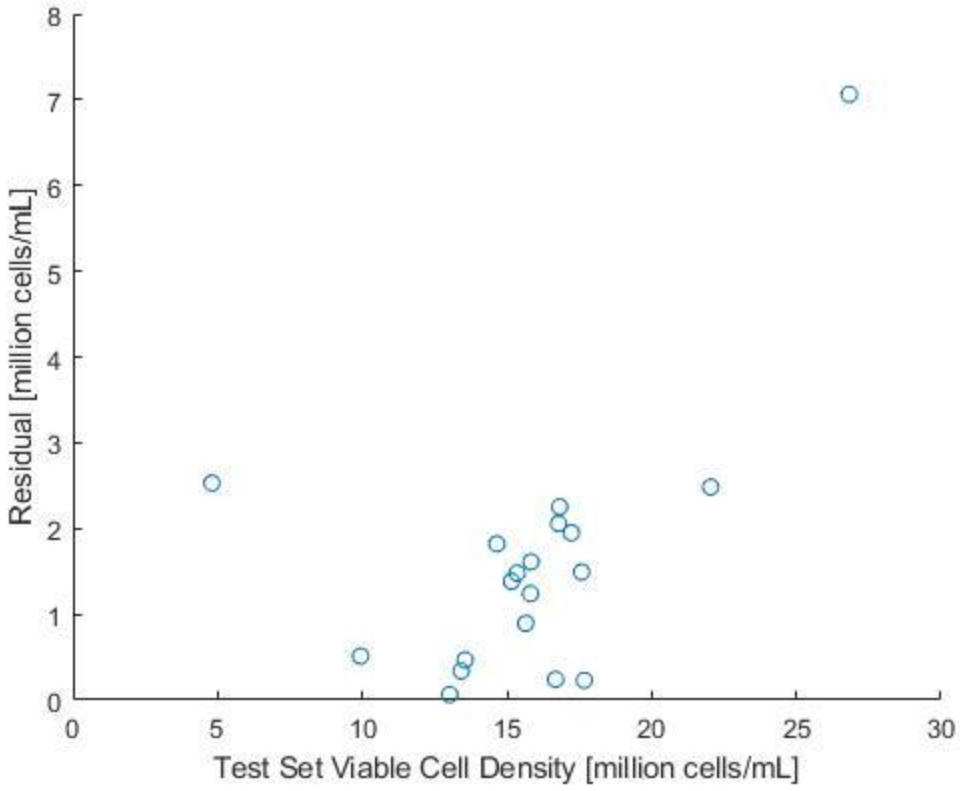

Figure 8.5 Residual Plot for Random Forest Regression Model

However, looking at the residual plot, the model does appear to fall apart at the high end of the viable cell density range. There were very few bioreactors that resulted in a viable cell density at the high end of the viable cell density range, which could have led to this result. It is possible that the performance of the model would have improved had more conditions produced a viable cell density at the high end of the range. Figure 8.6 shows the out-of-bag error plot for the random forest regression model, and Figure 8.7 shows the feature importance plot.

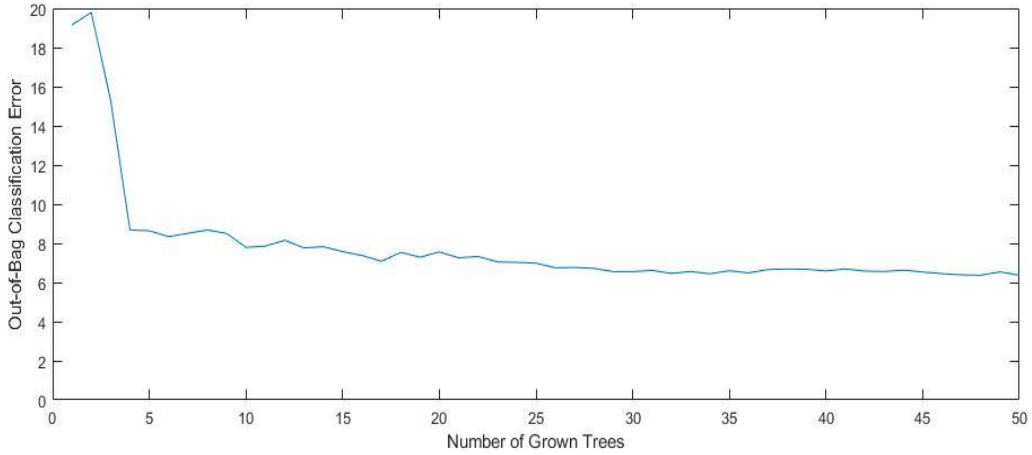

Figure 8.6 Out-of-Bag Error Plot for Random Forest Regression Model

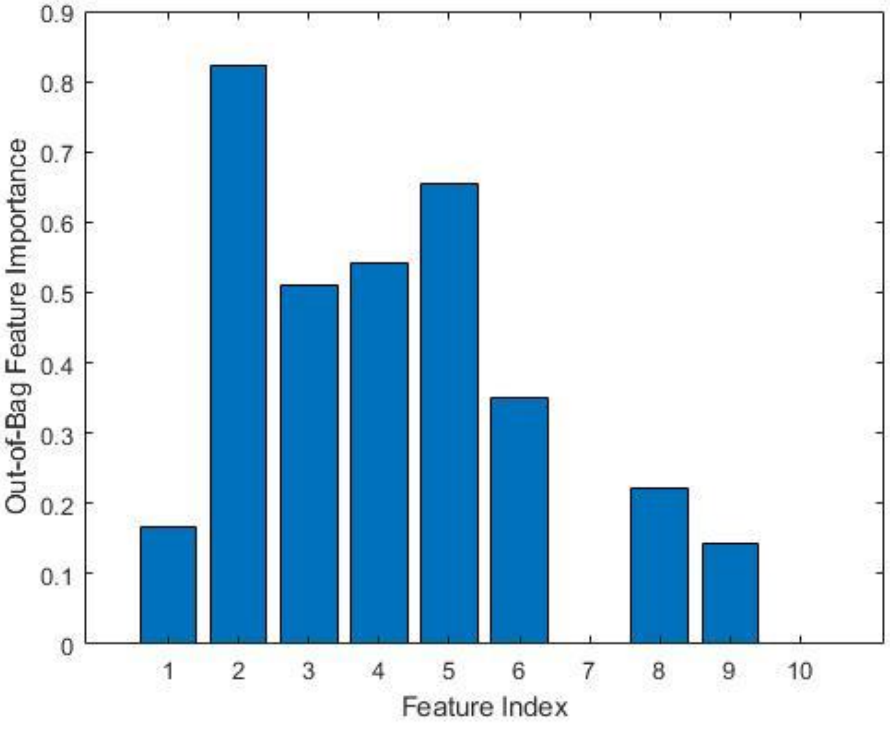

Figure 8.7 Feature Importance Plot for Random Forest Regression Model

From these plots, one can see that model performance levels off around four to five grown trees. Interestingly, the random forest model keeps all but two features, and the feature importance plot suggests that the model performs better without the dissolved oxygen feature. On a first principles basis, this does not make sense. From the results of the bioreactor experiments, dissolved oxygen is one of the most important features. This highlights the importance of having different disciplines on a scale-up team. This will be discussed further in the final chapters of this work. In the next subsection, the results of the predictive models are summarized.

8.5 Viable Cell Density Predictive Model Discussion

From the datasets collected from the ambr experiments and confirmation runs, two models were developed: one based on lasso regression and one based on random forest regression. Both models performed well in predicting viable cell density, as evaluated by RMSE on test datasets. However, the random forest regression model improved on the RMSE of the test set by more than 100%, suggesting it is a better algorithm to employ when predicting peak viable cell density. While the results of the predictive models are mixed, the experiments show that the ambr system can be used to develop a reasonable predictive model. Further, only two algorithms were tested; there are many more algorithms that would be applicable, such as elastic net or ridge regression. As creating a model was an ancillary goal of this work, experimental conditions were not chosen to facilitate the creation of a great model. Better planning and experimental design would certainly improve the model's performance.

9. Future Work

9.1 Applicability to Maximization of rHA Production via SF+ Cells

The methodology applied in this work can be further applied to maximization of protein production. The problem becomes much more difficult because there are many more variables to consider with the infection process. Some of these include cell density at time of infection, viability at which to harvest the cells, multiplicity of infection (MOI), and time after infection until first feed. One also must balance the potential decrease in specific productivity (mg protein/billion cells) with the increase in cell density (million cells/L). Specific productivity tends to decrease when infecting at cell densities above 10 million cells/mL. The ambr system can be used to accelerate the process development timeline, but precautions must be taken when using viral cultures. Timely (and potentially costly) decontamination procedures must be performed when switching back to cell growth only or a different cell line entirely. That is why for a BEVS process it is recommended to have dedicated biological safety cabinets for the viral and non-viral cultures. To date, there are not any published works that report rHA production using Sf9 cells at high density (here defined as infection at greater than 7 million cells/mL). However, some research articles to consider for guidance are *Insect Cell Culture for Industrial Production of Recombinant Proteins*[17], *Enhanced Growth of Sf-9 Cells to a Maximum Density of 5.2 x 10^7 Cells per mL and Production of β-Galactosidase at High Cell Density by Fed Batch Culture*[39] and *Fed-batch Culture of Sf-9 Cells Supports 3 x 10^7 cells per mL and Improves Baculovirus-expressed Recombinant Protein Yields*[37]. As for the development of a statistical model, a similar approach could be used for protein production, with the dependent variable being total volumetric productivity (mg protein per L) instead of peak viable cell density.

9.2 Thoughts on Medium Optimization

To reiterate the point stated in the conclusion of Elias *et al.*[39], "in order to successfully exploit the potential of the fed batch method not only the composition of the nutrient feed but also the feeding regime have to be carefully designed." The author of this present work could not concur more, and based on these experiments, it is suggested that the feed medium be defined prior to optimizing the feeding regimen. The idea is that a specific cell density or protein yield is targeted with a specific medium, and then the key variable (cell density or protein yield) is incrementally improved through process improvement. If both feed medium and process design are optimized at the same time, there may be too many variables to home in on the optimal formulation and process. Like a chemical process, the synthesis of a certain chemical is typically determined first and then scaled up to production level. The advancement of computing and data analysis has unlocked the potential for different methods to optimize a process. Having a chemical engineer, biologist, and data analyst collaborating could add to each other's strengths and expedite the scaleup process. Some ideas include using ^{15}N-labeled proteins to track which amino acids are observed in rHA protein[20]. Another idea is to use machine learning techniques outlined in this work or even a genetic algorithm[50]. In summary, both the pre-infection and post-infection feed media should be defined before process optimization begins.

9.3 Ambr-250 ® System Replacing Traditional Benchtop Setup

The final section of this document is a commentary on the impact of upgrading a conventional benchtop bioprocess setup to the ambr system. For this hypothetical process scaleup, we assume 48 batches are required, taking 2 weeks of preparation, setup, growth, and cleanup per batch. The ambr system requires only 1.5 weeks per batch. A traditional benchtop setup might have 12 benchtop bioreactors. We also assume that we have a three-person lab team

to run the process. Two people are enough to run the ambr process, so we will measure the labor required in man-hours. A comparison of the two processes is given in Table 9.1.

Table 9.1 Comparison of Traditional Benchtop Bioreactor to ambr System

	Ambr-250 ® system[51]	**Traditional Benchtop Bioreactors**[52]
Base Capital and Maintenance Costs	$700,000	$200,000
Consumable Cost (per 12 batches)	$5,000	Negligible
Total Cost for Experiment	$720,000	$200,000
Weeks to Complete Experiment	6	8
Man-hours Required for 48 batches	480	960

In short, there is a significant upfront capital cost and recurring consumable cost for the ambr system, so experiments run on this system must be designed carefully and selected based on priority. Exploratory experiments, as done in this work, are not recommended. However, the testing, sampling, and data-collecting features of the ambr system vastly surpass the abilities of a benchtop bioreactor. In addition, the ambr system improves the efficiency of lab resources, allowing scientists to be moved to other experiments or to analyze data.

10. Final Recommendations and Conclusions

10.1 Business Impact

The business implications of improving the rHA protein yield by infecting at high cell density are numerous. If you increase output, you can produce more material with the same infrastructure footprint. Further, as mentioned in the background, there is still not enough flu vaccine supply to meet demand. With the COVID-19 crisis, it is likely that flu vaccine demand will increase, as people will be advised to get an influenza vaccine to avoid coinfection with both influenza and coronavirus.

Assuming roughly 25 million doses of Flublok per year[4], and approximately 1 million doses per batch, it takes roughly 25 batches per year to produce a season's supply of Flublok vaccine. If yield is improved by 15%, the same number of doses can be produced in 22 batches, lowering the variable costs of the plant, and reducing the cost of goods sold. Alternatively, a 15% yield increase could produce 3,750,000 additional flu vaccine doses per year. At an average price of $45 per dose[47], this would generate an additional $170 million in revenue per year. Sanofi does not separate financial statements by business segment, but assuming a rough cost of goods of 40% of revenue (based on GlaxoSmithKline's financial statements)[48], this would generate an additional $100 million in profits per year.

Rather than increasing yield at an existing plant, Sanofi could also choose to build a new facility to produce more vaccine doses. Adjusting for inflation and allowing for a 50% engineering cost overrun, one can assume that a new facility would cost $280 million in 2020[49]. If 25 million doses could be produced in this new facility, it would generate a cash flow of $1.125 billion in revenue and $675 million in profits at full capacity. Assuming a new facility

would require five years to be licensed for production, $280 million at a discount rate of 8% is roughly $410 million in present-day dollars. Assuming the plant would sell 25% of full capacity in year one of operation, 50% of full capacity in year two, and 75% of full capacity in year three, it would break-even after three years of operation (eight years after project kick-off). While Sanofi has limited capital and would have to select a new site carefully, it would still be a winning strategy.

10.2 Highlighted Results and Conclusions

Process optimization of SF+ cultures obtained over 20 million cells/mL in the ambr system. The process to grow SF+ cells in the ambr system was established. This process was replicated in 3L Applikon reactors and obtained a cell density of over 25 million cells/mL. The increase in peak cell density was likely the result of improved oxygen transport in the larger bioreactors. This demonstrates that the ambr bioreactor system is an appropriate system for quickly scaling up Sf9-based bioprocesses. Some of the specific finds of the ambr experiments are that gas control (overlay, sparger, and oxygen) has a significant impact on cell growth. Lowering the dissolved oxygen setpoint improved the maximum peak cell density of SF+ cells by 10-15%. It was determined that a single feed was enough to reach the target cell density for the next phase of the scaleup, which is infection at high cell density and protein production.

Both a lasso regression and random forest regression model were developed using the data from both the ambr and the benchtop bioreactor experiments. The lasso regression model had an RMSE of 5 million cells/mL, which is not a great result. The random forest regression model had an RMSE of 2.19 million cells/mL on the test set, which is a significant improvement. Although the lasso model was not a terrific predictor of maximum viable cell density, the

variables retained by the model were scientifically explicable: DO%, Feed 1% (v/v), Feed 3% (v/v), and k_La.

Appendix A – Supplemental Experiment Results

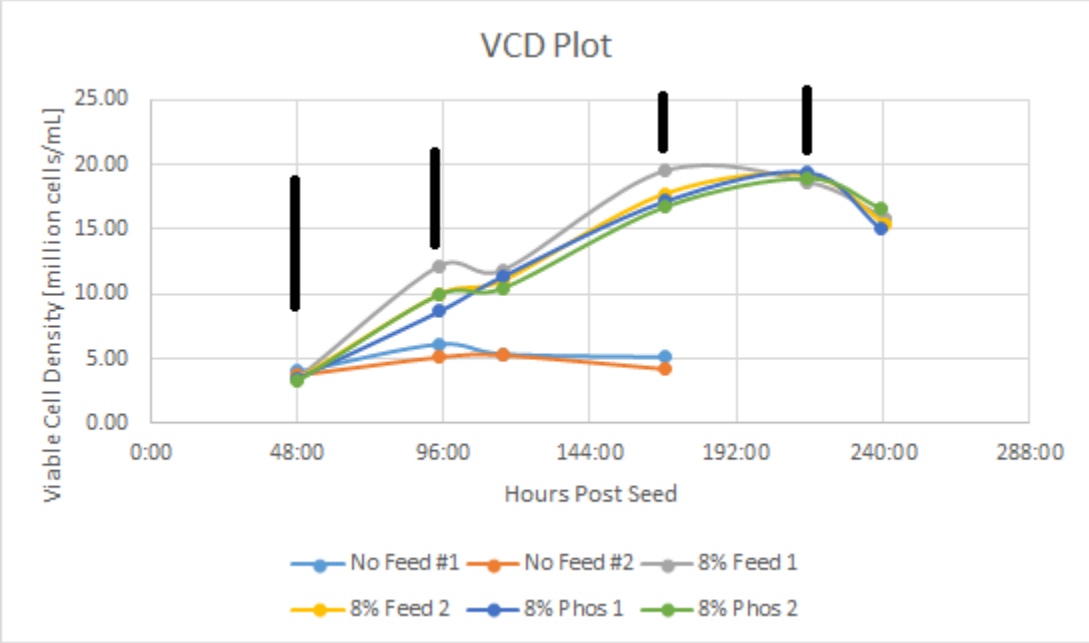

Figure A.1 – Viable Cell Density Plot for Phosphate Supplementation Shake Flask Experiment

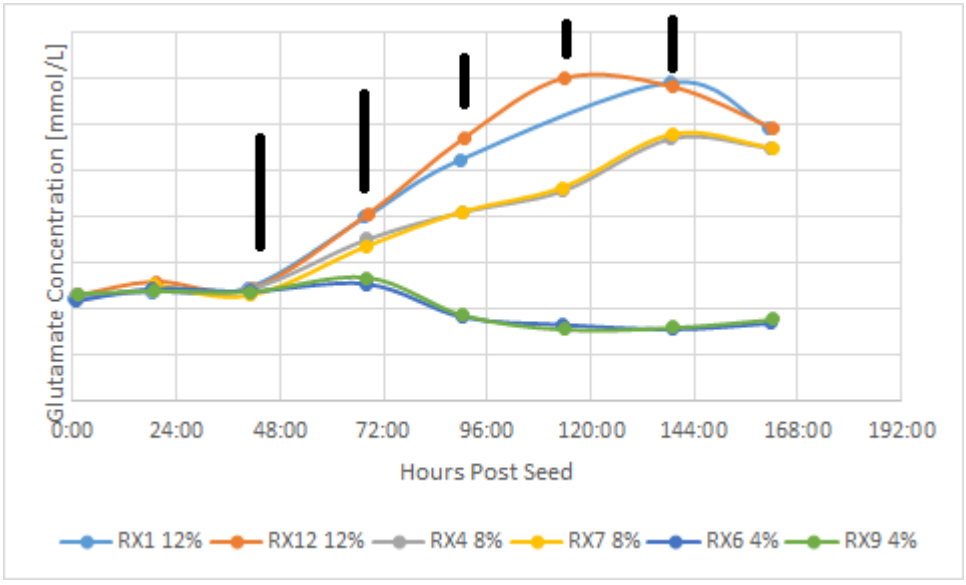

Figure A.2 – Glutamate Concentration Plot, Experiment 3

**Black lines indicate approximate feed times*
**12% v/v reactors received 3 feeds, 8% v/v reactors received 4 feeds, and 4% reactors received 5 feeds*

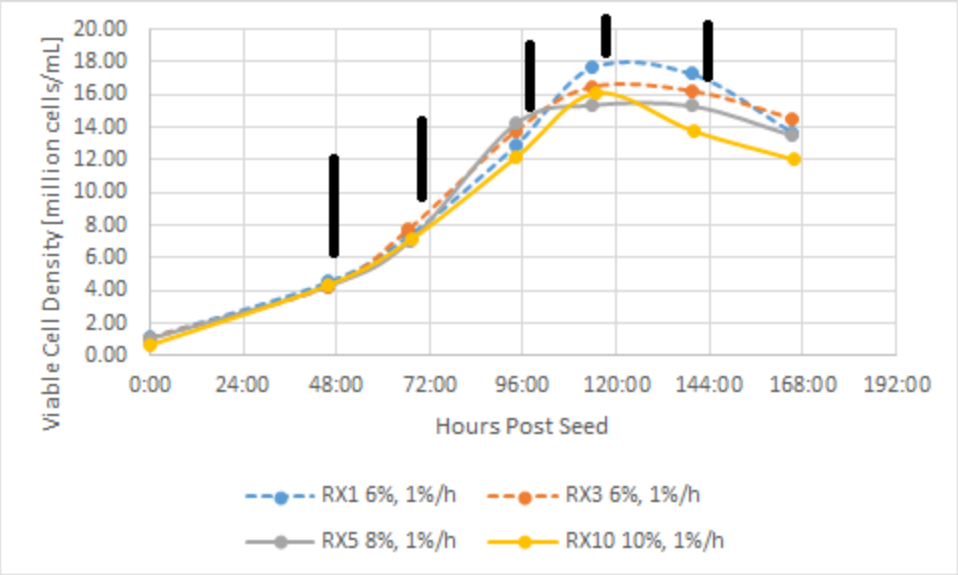

Figure A.3 Viable Cell Density Plot, Experiment 4, 1% v/v per hour, Varying Total % v/v

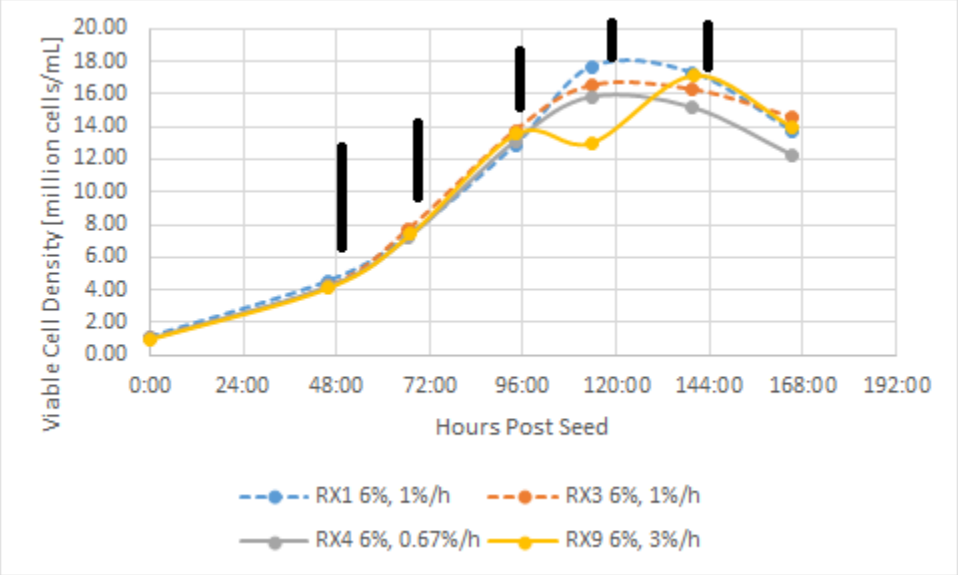

Figure A.4 Viable Cell Density Plot, Experiment 4, 6% Total v/v, Varying v/v % per hour

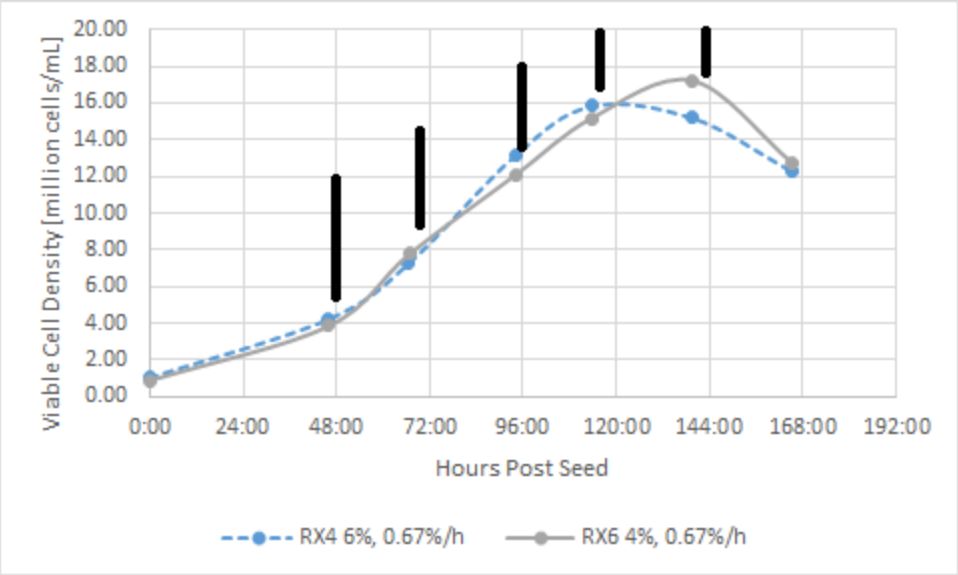

Figure A.5 Viable Cell Density Plot, Experiment 4, 0.67% v/v per hour, Varying Total % v/v

11. Bibliography

1. Mao, L., Yang, Y., Qiu, Y. & Yang, Y. Annual economic impacts of seasonal influenza on US counties: Spatial heterogeneity and patterns. *Int J Health Geogr* **11**, 16 (2012).
2. Estimated Influenza Illnesses, Medical visits, Hospitalizations, and Deaths in the United States — 2017–2018 influenza season | CDC. https://www.cdc.gov/flu/about/burden/2017-2018.htm (2019).
3. Fourth quarter and full year 2018 results - Sanofi. https://www.sanofi.com/investors/financial-results-and-events/financial-results/resultats-t4-2018.
4. Sanofi Ships First Flu Vaccines for 2018-2019 Season. http://www.news.sanofi.us/2018-07-30-Sanofi-Ships-First-Flu-Vaccines-for-2018-2019-Season (2018).
5. CDC. Types of Influenza Viruses. *Centers for Disease Control and Prevention* https://www.cdc.gov/flu/about/viruses/types.htm (2019).
6. PDB101: Molecule of the Month: Hemagglutinin. *RCSB: PDB-101* http://pdb101.rcsb.org/motm/76.
7. PDB101: Molecule of the Month: Influenza Neuraminidase. *RCSB: PDB-101* http://pdb101.rcsb.org/motm/113.
8. Recombinant Influenza (Flu) Vaccine | CDC. https://www.cdc.gov/flu/prevent/qa_flublok-vaccine.htm (2019).
9. Sanofi completes the acquisition of Protein Sciences. http://www.news.sanofi.us/press-releases?item=137157.

10. Cheng, Y.-L. Investigation of Sf-9 Cell Metabolism Before and After Baculovirus Infection Using Biovolume: a Case for the Improvement of Adeno-Associated Viral Vector Production. in (2009).
11. Meghrous, J. Recombinant Vaccine Production Process.
12. SparkNotes: The Citric Acid Cycle: The Reactions of the Citric Acid Cycle. https://www.sparknotes.com/biology/cellrespiration/citricacidcycle/section2/.
13. Bhagavan, N. V. Pentose Phosphate Pathway - an overview | ScienceDirect Topics. https://www.sciencedirect.com/topics/medicine-and-dentistry/pentose-phosphate-pathway.
14. Bédard, C., Tom, R. & Kamen, A. Growth, Nutrient Consumption, and End-Product Accumulation in Sf-9 and BTI-EAA Insect Cell Cultures: Insights into Growth Limitation and Metabolism. *Biotechnology Progress* **9**, 615–624 (1993).
15. Drews, M. *et al.* Pathways of glutamine metabolism in Spodoptera frugiperda (Sf9) insect cells: evidence for the presence of the nitrogen assimilation system, and a metabolic switch by 1H/15N NMR. *Journal of Biotechnology* **78**, 23–37 (2000).
16. Bussolati, O., Uggeri, J., Belletti, S., Dall'Asta, V. & Gazzola, G. C. The stimulation of Na,K,Cl cotransport and of system A for neutral amino acid transport is a mechanism for cell volume increase during the cell cycle. *FASEB J.* **10**, 920–926 (1996).
17. Ikonomou, L., Schneider, Y.-J. & Agathos, S. N. Insect cell culture for industrial production of recombinant proteins. *Appl Microbiol Biotechnol* **62**, 1–20 (2003).
18. Ferrance, J. P., Goel, A. & Ataai, M. M. Utilization of glucose and amino acids in insect cell cultures: Quantifying the metabolic flows within the primary pathways and medium development. *Biotechnology and Bioengineering* **42**, 697–707 (1993).

19. Öhman, L., Ljunggren, J. & Häggström, L. Induction of a metabolic switch in insect cells by substrate-limited fed batch cultures. *Appl Microbiol Biotechnol* **43**, 1006–1013 (1995).
20. Brüggert, M., Rehm, T., Shanker, S., Georgescu, J. & Holak, T. A. A novel medium for expression of proteins selectively labeled with 15N-amino acids in Spodoptera frugiperda (Sf9) insect cells. *J. Biomol. NMR* **25**, 335–348 (2003).
21. Doverskog, M., Ljunggren, J., Öhman, L. & Häggström, L. Physiology of cultured animal cells. *Journal of Biotechnology* **59**, 103–115 (1997).
22. Meghrous, J. *et al.* Dissolved carbon dioxide determines the productivity of a recombinant hemagglutinin component of an influenza vaccine produced by insect cells. *BIOTECHNOL. BIOENG* **112**, 2267 (20151101).
23. Palomares, L. A. & Ramirez, O. T. The effect of dissolved oxygen tension and the utility of oxygen uptake rate in insect cell culture. *Cytotechnology* **22**, 225–237 (1996).
24. Hidalgo, D., Paz, E., Palomares, L. A. & Ramirez, O. T. Real-time imaging reveals unique heterogeneous population features in insect cell cultures. *Journal of Biotechnology* **259**, 56–62 (2017).
25. Zhou, Y. *et al.* Effects of Agitation, Aeration and Temperature on Production of a Novel Glycoprotein GP-1 by Streptomyces kanasenisi ZX01 and Scale-Up Based on Volumetric Oxygen Transfer Coefficient. *Molecules* **23**, (2018).
26. De Jong, T. F., Fernandes, C., D'Ambruoso, A. & Koechlein, D. Scalability in Laboratory Bioreactors Based on Constant Volumetric Mass Transfer. https://www.applikon-biotechnology.com/files/applikon-poster-scalability-in-lab-bioreactors-kla.pdf.

27. Zhang, S., Handa-Corrigan, A. & Spier, R. E. Foaming and media surfactant effects on the cultivation of animal cells in stirred and sparged bioreactors. *Journal of Biotechnology* **25**, 289–306 (1992).
28. Maximization of recombinant protein yield in the insect cell/baculovirus system by one-time addition of nutrients to high-density batch cultures | SpringerLink. https://link.springer.com/article/10.1007/BF00762387.
29. Gibco™ Yeastolate Ultrafiltrate (50X) 100mL Dissociation Reagents | Fisher Scientific. https://www.fishersci.no/shop/products/gibco-yeastolate-ultrafiltrate-50x/11558906.
30. Maiorella, B., Inlow, D., Shauger, A. & Harano, D. Large-Scale Insect Cell-Culture for Recombinant Protein Production. *Bio/Technology* **6**, 1406–1410 (1988).
31. Huhtala, A., Linko, P. & Mutharasan, R. Protein response of insect cells to bioreactor environmental stresses. *Journal of Biotechnology* **118**, 278–289 (2005).
32. Xu, P. *et al.* Characterization of TAP Ambr 250 disposable bioreactors, as a reliable scale-down model for biologics process development: Characterization and Application of Ambr 250 in Clone Selection and Process Development. *Biotechnol Progress* **33**, 478–489 (2017).
33. Zanghi, A. *et al.* Increased efficiency and product quality with the UniVessel® Single Use bioreactor for CHO fed-batch cultures. 10.
34. ambr® 250 high throughput.
35. Baez, A. & Shiloach, J. Effect of elevated oxygen concentration on bacteria, yeasts, and cells propagated for production of biological compounds. *Microb Cell Fact* **13**, (2014).
36. Vajrala, S. G. & Murhammer, D. W. Effect of CO2 on uninfected Sf-9 cell growth and metabolism. *Biotechnology Progress* **32**, 465–469 (2016).

37. Bédard, C., Perret, S. & Kamen, A. A. Fed-batch culture of Sf-9 cells supports 3 3 10e7 cells per ml and improves baculovirus-expressed recombinant protein yields. *Biotechnology Letters* **19**, 629–632 (1997).
38. Buford, J., McMillan, J., Doshi, H. & Bryans, D. M. SOP: Applikon ez-Control Bioreactor Controller Operation. 21.
39. Elias, C. B., Zeiser, A., Bédard, C. & Kamen, A. A. Enhanced growth of sf-9 cells to a maximum density of 5.2 × 107 cells per mL and production of β-galactosidase at high cell density by fed batch culture. *Biotechnology and Bioengineering* **68**, 381–388 (2000).
40. Braatz, R. D. & Anthony, B. W. Process Data Analytics & ML: Linear-in-the-Parameter Models.
41. Wolfe, A. Regulation Techniques for Multicollinearity: Lasso, Ridge, and Elastic Nets. 23.
42. Lasso or elastic net regularization for linear models - MATLAB lasso. https://www.mathworks.com/help/stats/lasso.html.
43. Yiu, T. Understanding Random Forest. *Medium* https://towardsdatascience.com/understanding-random-forest-58381e0602d2 (2019).
44. Hastie, T., Tibshirani, R. & Friedman, J. Random Forests. in *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (eds. Hastie, T., Tibshirani, R. & Friedman, J.) 587–604 (Springer, 2009). doi:10.1007/978-0-387-84858-7_15.
45. Create bag of decision trees - MATLAB. https://www.mathworks.com/help/stats/treebagger.html.
46. Oshiro, T., Perez, P. & Baranauskas, J. How Many Trees in a Random Forest? in vol. 7376 (2012).
47. Flublok Prices, Coupons & Savings Tips. *GoodRx* https://www.goodrx.com/flublok.

48. financial-statements-2019.pdf.

49. Sanofi flu vaccine plant to triple US production. *Reuters* (2007).

50. Marteijn, R. C. L. *et al.* Optimization of a feed medium for fed-batch culture of insect cells using a genetic algorithm. *Biotechnol. Bioeng.* **81**, 269–278 (2003).

51. Ahn, W. ambr Rough Cost Estimates. (2020).

52. *Ace Glass Incorporated Reactors Catalog Bench & Kilo Class*. https://www.aceglass.com/html/3dissues/Reactors/offline/download.pdf (2014).