PROCESS MODELING AND OPTIMIZATION OF BIOPHARMACEUTICAL

MANUFACTURING

BY

OU YANG

A dissertation submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Chemical and Biochemical Engineering

Written under the direction of

Marianthi Ierapetritou

And approved by

_____

_____

_____

_____

New Brunswick, New Jersey

May 2021

ABSTRACT OF THE DISSERTATION

PROCESS MODELINGAND AND OPTIMIZATION OF BIOPHARMACEUTICAL

MANUFACTURING

by OU YANG

Dissertation Director:

Marianthi Ierapetritou

Biopharmaceutical manufacturing mainly produces large molecule-based products and is becoming increasingly important in the last few years, leading to high market demand. Two approaches are investigated in this dissertation to satisfy the high market demand, including the selection of high-performance process operating mode and improvement of operating strategies. Continuous processing shows a significant benefit in increasing productivity, reducing the footprint, and cost-effectiveness over conventional operations. Quantitative evaluation of the overall process performance is critical to decide the preferred mode of operation (batch or continuous). At the same time, enhancing the understanding of conventional unit operations is also essential to improve productivity while maintaining the product quality for current manufacturing setups.

In this dissertation, different simulation methods, including flowsheet modeling and mathematical modeling, are explored to accurately predict and evaluate large molecule-based drug production. Flowsheet modeling is implemented to design and construct fully integrated frameworks for continuous monoclonal antibody and gene therapy drug productions and compare their performance with the batch processes using techno-economic analysis. Scenario studies are used to evaluate process cost-effectiveness under varied production scales and upstream/downstream parameters. Under the guidance of quality by design (QbD), mechanistic and surrogate models are

built to capture the nonlinear bioprocess dynamics between operating conditions and output variables, aiming to improve productivity while maintaining critical quality attributes of the products. Furthermore, feasibility analysis has been used to determine the design space for protein production based on the requirement of commercial product quality attributes. This dissertation provides a framework covering process design and decision-making in biopharmaceutical manufacturing, paving the way for cost-effective manufacturing and high-quality biological drug production.

Acknowledgements

Looking back to the past four years, I have been encouraged by many people. I would like to take a moment to thank them.

Foremost, I wish to express my deepest and sincerest gratitude to my advisor Prof. Marianthi Ierapetritou. Thank you for your patience that helped me get through the challenges and explore interesting projects of my research. Your wisdom, passion and support have made my Ph.D. study productive. Thanks for your trust and encouragement to make me always improve myself. I can't express how grateful I am to be one of your students. You not only teach me to be a good researcher but also to be a kind, brave and confident person.

I would like to thank my committee members, Prof. Shishir Chundawat, Prof. Rohit Ramachandran and Dr. Tao for their valuable and constructive feedback on my work. I appreciate all the discussions and comments. It is my privilege to have a chance to work with all of you. I would also like to thank my financial support from FDA and Eli Lilly and Company.

It was my great pleasure to work with everyone from the Ierapetritou group, Parham, Lisia, Sebastian, Zilong, Nirupa, Abhay, Atharv, Pooja, Yingjie, Yue, Siddharth, Praneeth, Shu, Chaoying, Jayanth, Michael, Huayu, Robert, Dare. I would like to thank Lisia and Zilong for introducing me to the group and bring me to this wonderful journey. I would like to thank Parham for helping me with the understanding of bioreactor modeling when I just joined the group. Thanks, Atharv, for always standing by my side, supporting me, and encouraging me. To Pooja, thanks for being my friend, my roommate. I really enjoy the time that we work together and help each other. I would like to thank Prof. Chundawat's research group for providing me experimental data for this work, especially thank Viki and Aron. From Rutgers to Delaware, I am very fortunate to have all of you that bring me memorable times in my life.

I would like to thank my previous advisors, Prof. Veiglia and Prof. Cheng, who provided me chances to work in your lab during my undergraduate study. For the first time, bring me the world of research and further inspired me to pursue a Ph.D. study.

Moreover, I would also like to acknowledge my family, my parents, and my grandparents for all your endless love and for your unlimited support. I would not be able to complete this work without your encouragement. I hope one day I can make you proud. I want to express my special thanks to my friends, Qiushi, Mengxi, Yi, thanks for your trust and choose me to be one of your friends and thanks for your companion for the past years and in life long.

Table of Contents

List of Figures

Chapter 4.3

List of Tables

Chapter 5.4

1 Introduction

1.1 General overview of biopharmaceutical manufacturing

Biopharmaceutical manufacturing focuses on the production of large molecule-based products in heterogeneous mixtures, which can be used to treat cancer, inflammatory, and microbiological diseases [1]. Major biological drugs include monoclonal antibodies, vaccines, recombinant hormones/proteins, cells and gene therapies. Monoclonal antibodies are one of the most famous biological products, which takes the highest percent (37%) among all the drugs on the market (the year 2016), including drugs in development [2]. It has rapid routes to a clinical proof -of concept, low cost, low risk and high yield. On the other hand, gene therapy process is just recommended and approved worldwide in 2017 which has ability to make site-specific modification to human genomes and target on therapeutic treatment [3]. In year 2020, FDA provides guidance for industry on design of long term follow up studies of human gene therapy products which indicating novel gene therapy development is promising in future applications [4] .

In recent years, the global biopharma market increased rapidly from $4.4 billion in 1990 to $275 billion in 2018 [5]. This brings an increasing demand for biologic-based drugs that drives the need for manufacturing efficiency and effectiveness [6]. To deal with these challenges, many companies are improving the manufacturing processes by transitioning from batch to continuous operation mode [7] and employing smart manufacturing systems [1].

Continuous biopharmaceutical process reduces the equipment footprints compared to the corresponding batch or semi-batch process and offers rapid capacity adjustment and process scaling up [8]. In addition, continuous process improves product quality, reduces bioburden risks and increases volumetric productivity of the process. It also standardizes the biological production by building multiproduct facility which can produce both stable and nonstable proteins[8]. All the above advantages of continuous biopharmaceutical processing result in process intensification. It

has been reported that continuous operations have been used in small molecule drug production, and there exists FDA approved products such as Vertex's ORKAMBI^{TM} and Prezista [9].

Smart manufacturing system has capabilities to capture productivities, product qualities, product and process sustainability as well as has agilities to response to changes of the process[1]. This requires the digitization of each unit operations, data collections, analysis, system control and supply chain management. Especially in these years, digital twin idea has been proposed [10] which can guide and control the physical plant by running a virtual plant on the side. Process modeling plays an important role of these advances, which can assist biopharmaceutical manufacturing in product development, process prediction, decision making, and risk analysis.

In this dissertation, process modeling methods have been used to resolve the current challenges in biopharmaceutical manufacturing by improving the process cost efficiency as well as productivity. From the first approach, flowsheet modeling has been used to design and evaluate batch and continuous processes for monoclonal antibody and gene therapy production, demonstrating the advantages of continuous operations. Another approach focuses on improving current conventional unit operation, upstream fed-batch production bioreactor using both mechanistic and surrogate models.

1.2   Product quality in biopharmaceutical manufacturing

To fulfill the FDA regulations and obtain safe products, biopharmaceutical operations should be strictly controlled and operated under a sterilized process environment, and the products have to satisfy the critical quality attributes (CQAs). CQAs are product-specific properties that have been defined as "a physical, chemical, biological, or microbiological property to characteristic that should be within an appropriate limit, range or distribution to ensure the desired product quality" in ICH Pharmaceutical development Q8 guideline [11]. Since the deviation of the drug properties would cause risk such as life threatening, loss of efficacy or effect on pharmacokinetics (PK),

pharmacodynamics (PD), drug manufactures have to maintain a certain operating range during the drug production.

As mentioned in the previous section, mAbs are the most popular biological products used to treat cancers and micro-bio diseases. These proteins are generally produced from Chinese hamster ovary (CHO) cells in bioreactors and purified by a series of filters and chromatography. Examples of the quality attributes for monoclonal antibody productions include size-related variants, charge-related variants, oxidation-related variants, Fc glycosylation and structural variants [12]. In recent years, researchers have been investigating ways to increase the protein production by adjusting operating conditions and culture media formulation to meet the high drug demands. In the meantime, product qualities such as antibody-dependent cellular cytotoxicity (ADCC) and complement-dependent cellular cytotoxicity (CDC) are also critical to be maintained. It was found that the presence of a terminal galactose of the Fc region during N-linked glycosylation process increases the intensity of the CDC response. The reduction of fucosylation would improve the effect of ADCC activities [13]. More than 70% of the approved mAbs belong to IgG1 isotype [14]. N-linked glycosylation is a critical step to formulate the final IgG product and glycan fractions representing the level of afucosylation, galactosylation, high mannose, and sialylation are quality attributes that significantly affect protein folding, binding, and further controls product safety, efficacy, and potency [14-17].

Quality by design (QbD) proposed by FDA assists process development ensures the quality of the products through the manufacturing processing [11]. The basis of QbD is to link the critical process parameters with CQAs and further define a design space for process operations. In recent years, mathematical modeling has been applied to QbD for biopharmaceuticals process understanding and development [18, 19]. In this dissertation, another aim is to simulate upstream production bioreactor and correlate their critical process parameters such as temperature and pH to protein glycosylation in mAbs production. We developed a framework to achieve predictive process

modeling with different modeling approaches and further find design space for high productivity and qualities.

1.3   Outline of the dissertation

This dissertation focuses on the improvement of biological drugs' production by using process modeling methods and is organized as follows. Chapter 2 reviews the application of process system engineering tools in biopharmaceutical process development and targets on integrated process design and upstream bioreactor modeling. Following this, analysis methods that are for plant decision making and design space investigations are introduced. Chapter 3 describes approaches to reduce costs of the biopharmaceutical manufacturing by converting batch process to continuous using flowsheet modeling. The techno-economic analysis is used to evaluate the performance of continuous operations. Two types of the biological products, the production of monoclonal antibodies and recombinant adeno-associated virus are evaluated in case studies. Chapter 4 focuses on the improvement of upstream bioreactor operation by capturing the cell growth, protein production, and product qualities under different temperatures and pH using both mechanistic and surrogate models. Process system engineering tools such as feasibility analysis are used to investigate design space for upstream operations to maintain productivity and product quality. Chapter 5 further explored the development of bioreactor predive model with the support of experimental data in Herceptin production. A realistic model is built and captured the temperature effect during the cell culture and can be potentially integrated with PAT and control system to achieve an on-line soft sensor. Chapter 6 discusses conclusion of this dissertation and provides the potential future work.

2   Literature Review

In this chapter, works on modeling and analysis methods in biopharmaceutical manufacturing has been reviewed. In section 2.1, modeling approaches in biopharmaceutical manufacturing have been reviewed, including modeling for the integrated processes from upstream to downstream operation and modeling for upstream production bioreactor. Modeling methods such as mechanistic model, data-driven model, as well as different simulators that are available in the market are reviewed. Section 2.2, described analysis methods in process decision making and design space defining.

2.1   Modeling approaches in biopharmaceutical manufacturing

2.1.1   Integrated process modeling

In biopharmaceutical manufacturing, unit-to-unit connections are very important. Upstream output such as throughput and product qualities significantly affect the choice of downstream unit operations and equipment size. Since downstream processing consists of multiple types of equipment, optimizing operating strategies for each unit while bridging the connections from one unit operation to another will improve the efficiency of the operations. Thus, it is important to integrate all the unit operations and predict final product quality or productivity from the integrated process. Integrated process modeling usually combines with the analysis method to achieve process design and optimization. Many of the process integrations are focusing on continuous biopharmaceutical processing and evaluate process cost effectiveness[20-23], understand residence time distribution[24] and implement process optimization[25].

To build an integrated model, one approach is to use a mechanistic model and data-driven model to capture each of the unit operation and then integrate them together. For example, Sencar et al. [24] used mechanistic model and empirical model to achieve unit by unit modeling and integrate them to a RTD model. This model supports the tracking of material through different unit operations as well as effects on process disturbance. Gomis-Fons et al. modeled chromatography

steps and integrated them with the control system to achieve real-time control and optimize the operations [26]. These models usually build based on the platform such as python, matlab, Java, etc., which provide flexibility on model construction. However, these models require a high level of programming skills and a detailed understanding of each unit operation. Another approach is to use simulators to capture and compare integrated processes and track the mass balance and adjust process schedules. Simulators such as Aspen, SuperPro designer, Biosolve, gProms etc. are user-friendly platforms that also contain inbuilt analysis methods for process evaluations. Table 2.1 1 briefly provides the examples and platforms that have been used in integrated process modeling. In section 2.2, the detailed modeling, analysis and decision-making methods will be discussed.

Table 2.1 1 Integrated process modeling

Integrated Process			
Categories	**Methods**	**Platforms**	**Comments**
Residence time distribution	Probability distribution function for each unit operations	Python	Correlate input material operating conditions, design parameters with outlet profile. Easy to update.
Activity tracking and decision making	Discrete Event Simulation [27]	Extend Sim, Simul8,	Discrete/dynamic system, track activity, scheduling and resource utilization
Material tracking and decision making	Mechanistic/Empirical model [28]	SuperPro Designer, Biosolve,	Track material balance and optimize cost-effectiveness. Process debottlenecking, capacity planning
Process risk assessment	Implement process model with Monte Carlo analysis [29]	MATLAB	Evaluate parameter sensitivity, impurity purification and product quality. Hard to apply to computationally expensive model
Overall process optimization	Integrate flowsheet model with optimization solvers [30]	SuperPro Designer-VB-Matlab	Optimize environment impact and cost-effectiveness by adjusting 4 operating parameters

2.1.2   Bioreactor modeling

Bioreactor is one of the most important unit operations in the upstream process closely relative to the overall plant production and product quality. The outputs of the production bioreactor affect the downstream purification process. Thus, the understanding, predicting and control of the bioreactor operation is critical to the improvement of the overall process manufacturing.

In upstream operations, researchers and companies try to adjust media, cell line and operating conditions to achieve target productivities and CQAs. Temperature, pH, agitation, aeration rate, amino acid addition, and pressure change are all critical factors that need to be controlled. With the change of the parameters above, osmolality, dissolved oxygen or carbon dioxide, share force, mixing efficiency, metabolite, and nutrient concentrations would be changed, further affecting cell intracellular metabolism and final biologics drug production. Thus, to understand the correlations of critical process parameters with critical quality attributes, different methods including mechanistic, data-driven, and hybrid modeling have been used.

For mechanistic modeling, kinetic modeling has capability to show dynamic changes of the product titer and CQAs under various operating conditions. Kinetic modeling can be built based on the understanding of process structure including cell growth algorithms, biochemical reactions or adsorption mechanisms. Due to a large number of parameters in the model, parameter estimation and process validation are required. To reduce the number of parameters, kinetic modeling is commonly coupled with empirical modeling to build semi-empirical models. These models can be used for reduction of dimensionality of the problem by combining some parameters into a single parameter. Using lumped equations such as Monod-based or Michaelis–Menten models, cell growth and biochemical reactions inside cells can be well captured [31]. These models have used to simulate the temperature, pH effect to the cell growth and productivity [32, 33]. Xu et al.[32] used kinetic model to capture the effect of temperature shift on CHO cell culture performance in

fed-batch bioreactor. By obtaining parameters under different temperatures from first 8 days' culture, the model predicted the viable cell density, glucose, glutamine, glutamate, lactate concentration and titer for extended duration up to 14 days with different temperatures and initial cell density. CQAs such as glycosylation profile can be captured to interpolate different temperatures and pH as well. Sou et al. [34] successfully used kinetic model to simulate cell culture, nucleotide, nucleotide sugar metabolic and N-linked glycosylation under two different temperatures. Villiger et al. [35] built pH, manganese, and ammonia concentration effect to cell culture and glycosylation process in their model and described the dynamic profile of glycan fractions. Mechanistic model can also be used to assist the media selection and optimize feeding strategies. Kotidis et al. [36] investigated the glycosylation precursor feeding including galactose and uridine effects on cell growth, protein productivity and quality. Radhakrishnan et al.[37] applied kinetic model capturing the time-dependent media supplement effects on the glycosylation quantitatively. With the development of continuous processing, kinetic modeling has also been used to simulate both cell culture and protein glycosylation processes in perfusion bioreactor [38, 39]. However kinetic model contains many kinetic pathways, which requires a significant amount of data for parameter estimation.

To deal with this problem, stoichiometric model has been widely used to capture biochemical components, intermediate molecular and intracellular reactions. By calculating flux of each components, stoichiometric model is able to predict specific metabolic rate which has been used in media selection and feed media adjustment[40]. In addition, the model based on genome scale can be used for cell culture performance and glycoprofiles prediction from mutant cell lines[41]. Dynamic process can also be modeled by capturing the fluxes at different time intervals [42]. To further improve the dynamic simulation, stoichiometric model is coupled with kinetic model and track the intracellular metabolism with the extracellular environments and kinetic changes[43, 44].

As for agitation and dissolved oxygen effects, CFD simulation have been used to investigate the physical environment of the bioreactor and simulate its heterogeneities. To improve the modeling accuracy, CFD simulation has been coupled with PBM model to capture the liquid gas interactions as well as gas aggregation and breakage[45] It has also been integrated with kinetic model [46, 47] to understand physical and chemical stimuli to cell growth and protein production. However, mechanistic model is usually computational expensive which is not convenient to be used for process optimization and control.

Data-driven model has been applied to reduce the computational time while capturing the critical operating parameters and quality attributes at different scales. Selvarasu et al. [48] processed data with principal component analysis (PCA) and partial least square regression (PLSR) to identify the key nutrient components' effect on the cell viability and productivity which can be used for media selection. Sokolov et al. [49] used sequential multivariate tools to predict titer, aggregation, low molecular weight components, and glycan groups in the product under deep-well plates (DWP) scale, AMBR scale, Lab scale and pilot scale. The work shows the capability of the multivariate analysis in scale-up and decision making for biopharmaceutical manufacturing. Usually, multivariate analysis requires a large amount of data for accurate prediction. To deal with this problem, recently, Zurcher et al.[50] used multivariate analysis data tools to capture dynamic protein glycosylation profiles using extracellular data with small datasets. However, data-driven model still has weakness with poor extrapolation capabilities.

Hybrid model integrated data-driven model and mechanistic model and demonstrated its potentials in cell growth and metabolism simulation. Narayanan et al. [51] developed system of equations based on mass balance and integrated the equations with ANN to capture the cell density, metabolite concentration and productivity. Results showed that hybrid model improved the prediction accuracy and interpolation/extrapolation capabilities. In addition, hybrid model can also be used to predict bioreactor performance under different culture conditions [52]. Stosch, M. et al.

[18] used artificial neural network to correlate temperature, pH and feed rate with biomass and protein production rate and integrated this correlation in the kinetic model. By integrated data-driven model with flux balance analysis, pH and temperature affect to intracellular metabolism can also be investigated[53]. Table 2.1 2 shows examples of capabilities and methods for process modeling which can be potentially used in model building. However, it needs to note that, although process modeling has capabilities to capture all the above operating conditions and critical quality attributes, none of the modeling work incorporates all the process information within a single model.

Table 2.1 2 Upstream bioreactor modeling examples

Upstream Process			
Categories	Methods	Platforms	Comments
Bioreactor fluid dynamics, system heterogeneity	CFD simulation [45] CFD +PBM simulation [54] CFD+ kinetics model [46]	Ansys Fluent, COMSOL Multiphysics	Support to understand operations such as agitation, aeration, nutrients feeding. Guide process scale-up. Computationally expensive. Can reduce the computational time by using a compartment model, hard to be validated.
Cell growth, nutrients, and Metabolism. Product quality (protein glycosylation)	Kinetic model [32, 34, 55]	MATLAB, gPROMS, Visual Basic for Applications	Capture and predict the dynamic profile of the cell culture. Correlate CPPs and CQAs. Require a large amount of data for parameter estimations.
	Stoichiometric methods [56]	MATLAB, OptFlux etc.	Deal with a large amount of mechanistic reaction, genome-scale simulation. Need to integrate with the kinetic model to capture the dynamic profiles
	Multivariate tools [50]	MATLAB	Require a large amount of data. Represent input-output correlations. Do not capture the mechanistic correlations.
Media formulation	Multivariate analysis MFA [40, 48]	MATLAB	Identify nutrient correlations, improve

			productivity and cell viability
Product impurities	Regression model and Multivariate analysis [49]	MATLAB	Capture predict titer, aggregation, low molecular weight components, and glycan groups

The pros and cons of the kinetic model, flux balanced model, and statistical data-driven model are summarized and compared in Table 2.1 3. Table 2.1 2 and Table 2.1 3 can be used together to suggest the best modeling method for different bioreactor modeling projects.

Table 2.1 3 Comparison of computational methods in bioreactor modeling

Method	Kinetic model	Stoichiometric model	Stats model
Brief intro	Use kinetic parameters to represent culture conditions and product	Use stoichiometric relations and flux to represent reactions	Use stats methods such as PCA or PLSR to find the relation between operation conditions with product
Pro	1. Shows dynamic change of the model 2. Product prediction	1. Less experimental data are required 2. No parameter estimation needed 3. Help to understand the operating processes	1. Straightforward, less end-user expertise required 2. Commercial software available, software such as MATLAB also have packages for those 3. Handle multiple variables
Con	1. Large amounts of experimental data are required for parameter estimation, validation 2. Some parameters cannot be obtained are from literature data thus cause the inaccuracy	1. Need to solve underestimated problems 2. The method itself is hard to predict the dynamic changes of the system 3. Only consider steady state conditions	1. Large amount of data is required to find correlations 2. Accuracy can achieve up to 70%

2.2   Biopharmaceutical process design and operations

2.2.1   Decision making (Economic analysis)

In silico integrated process modelling is usually used for early stage process development and decision making and process optimization. Using economic analysis, batch, hybrid and continuous processing are compared and the economic benefits for varies cost categories can be observed. This method can be used as decision support tool to assist operating modes selection [21, 23, 28]. A more advanced decision support tool which considers, cost, environmental impact, product quality and supply robustness is also developed for production pattern selections between multiple-product small scale platform and single-product large scale platform [57]. Decision making includes not only the integrated process but also single unit operation. As describe in section 2.1, modeling of single unit operation would support the decision making on process operation and real time control by building the correlations of process input and output. In this section, we will focus on the decision making from techno-economic analysis which assists the selection of unit operation, operating mode and multi-product plant development. Methods such as deterministic analysis and stochastic cost analysis are introduced in details.

Simulation software

This section focuses on the different simulation software that can be used for economic analysis in continuous biopharmaceutical processing. The three simulation software packages that have been commonly used for cost analysis are Bio-Solve (Biopharm Services, UK), SuperPro Designer (Intelligen, Scotch Plains, NJ) and Aspen Batch Process Developer (Aspen Technology, Burlington, MA). **Table 1** presents the software used in selected recent published papers on economic analysis of biopharmaceutical manufacturing, with focus on continuous processing. It can be noted that Bio-Solve software is more commonly used in the recent years especially for continuous biopharmaceutical manufacturing whereas SuperPro Designer and Aspen have not yet been applied for published continuous processing simulation.

Table 1 Simulation packages used in recent publications in cost analysis of biopharmaceutical production

Paper	About Author	Content	Software
[58] Lim et al. 2006	UCL[1]	Fed-batch vs Perfusion	DES[2]
[59] Pollock et al. 2013	UCL, Pfizer R&D global biologics	Fed-batch vs perfusion	DES
[60] Pollock et al. 2013	UCL, Pfizer R&D global biologics	Design Semi-continuous affinity for clinical and commercial manufacture	DES
[61] Liu et al. 2013	UCL	Cost-effective design (Batch)	MINLP[3]
[62] Hammerschmidt et al. 2014	UNRLSV[4], ACIB[5], Novartis Pharma AG	Batch vs continuous vs hybrid process (precipitation)	BioSolve[6]
[63] Xenopoulos 2015	Biopharm Process Solutions R&D	Integrated batch vs continuous (multi-column)	BioSolve
[64]Walther et al 2015	BioRealization, Sanofi, R&D	Integrated batch vs continuous (multi-column)	BioSolve
[65] Li and Venkatasubramanian 2016	Columbia University	Integrated batch process, focus on downstream	SuperPro Designer
[66] Klutz et al. 2016	Invite GmbH, TU Dortmund University	Batch vs continuous vs hybrid integrated process	No simulation packages
[67]Bunnak et al. 2016	UCL, University of Manchester	Fed-batch and perfusion	BioSolve
[68] Torres-Acosta et al. 2016	UCL, Centro De Biotecnologia-FEMSA	Batch vs batch with ATPE downstream	BioSolve
[69] Xu et al. 2016	Merck	Fed-batch vs perfusion vs concentrated fed-batch (media cost)	BioSolve
[25] Liu et al. 2016	UCL	Integrated batch cost-effective optimization	MINLP

[70] Pollock et al. 2017	UCL, Pfizer Biotherapeutic Pharmaceutical Science	Batch vs continuous vs hybrid integrated process	DES
[71] Arnold et al. 2018	MedImmune	Batch vs continuous integrated process	BioSolve & experiment
[28] Pleitt et al. 2019	Thermo Fisher Scientific; The University of Queensland	Batch vs continuous focus on downstream	SuperPro Designer and Biosolve
[72] Hummel et al. 2019	Pall Life Sciences	Batch, Continuous single-use vs stainless steel focus on downstream	Biosolve
[73] Cataldo et al. 2019	ACIB, UNRLSV, A Sandoz Company	Batch vs continuous vs hybrid focus on downstream	Biosolve
[74] Gupta et al. 2020	Indian Institute of Technology Delhi	Batch vs continuous integrated process	SuperPro Designer

[1] UCL: University College London
[2] DES: Discrete-event simulation engine, extend with MySQL database
[3] MINLP: Mixed integer nonlinear programming, Constrained programming
[4] UNRLSV: University of natural Resources and life sciences Vienna
[5] ACIB: Austria Centre for industrial biotechnology
[6] BioSolve: BioSolve (Biopharm Services Ltd)

Bio-Solve is an Excel-based process and cost modeling software (Biopharm Services, UK). It can be applied to calculate cost of goods (COG) with a breakdown of cost drivers in both batch and continuous process based on a scalable process model, as the scale can be varied from lab scale to manufacturing scale. It can also be used for cash flow analysis. The software can perform multiple process comparisons and can be used to identify the most cost-effective technologies [75]. The software contains yearly updated built-in cost data library, including equipment, consumables and solution components. The software allows users to customize cost data based on models [62, 63]. However, this software is not able to capture the dynamic nature of the process. The latest BioSolve software shows ability to map batch, continuous and perfusion process into a Gantt chart [75].

SuperPro Designer (Intelligen, Scotch Plains, NJ) is a flowsheet driven simulator and can be used to integrate both process and business modeling. With this software, material and energy balances can be automatically calculated based on the mathematical model built by the user [65]. Thus, equipment sizing, debottlenecking, cost, and economic evaluation can be performed both in batch and continuous process. Combined with SchedulePro, scheduling in batch process can be achieved in both SuperPro and SchedulePro Designer. However, the simulation model can only be specified with fixed resource, time delay and failure events in bioprocess. In the continuous mode, there is limited equipment that exists in the current library without scheduling option. Cost analysis and project economic evaluation can be done by SuperPro Designer. The software can show the basic cost analysis including direct and indirect costs, and also do the profitability and cash flow analysis. Sensitivity analysis, Monte Carlo analysis can also be done with SuperPro model but the data need be transferred to excel and distribution of parameters should be provided in Excel spreadsheet. On the Excel spreadsheet, the add-ons packages need to be used, for example Crystal Ball (Oracle).

Similar to SuperPro Designer, Aspen Batch Process Developer (Aspen Technology, Burlington, MA) is a recipe-driven simulator that can also be used to simulate biopharmaceutical processes with economic analysis. The software can be used to obtain material balances, cycle-times calculation, perform scale-up, alternative process evaluation and process debottlenecking. The software can also simulate single or multi-product campaigns and track the in-plant emission such as CO_2 emission for environmental purpose. However, for Aspen Batch Process Developer, the simulation is limited to building batch processes. In addition to Aspen Batch, Aspen Technology also developed Aspen Chromatography that is used to simulate batch and continuous chromatographic separation processes that are applied in pharmaceutical and biotechnology. One of the advantages of Aspen Process is Aspen Chromatography and Aspen Batch process developer can be combined with Aspen Plus Dynamic so that the process can not only be analyzed under

steady state but also evaluated under dynamic process for both batch and continuous processes built in the software.

To understand the dynamic behavior of the process including process failure, delay, and rescheduling, a simulator such as ExtendSim should be used [59, 70]. By using ExtendSim, discrete event simulation can be applied to both discrete and continuous biopharmaceutical simulation. ExtendSim is a simulator that can be used for resource management, a mass balance analysis, in-process testing and costing analysis. A detailed discussion and its application can be found in Paige Ashouri's thesis [76]. The thesis also provides the discussion of previous research that applied the software to biopharmaceutical manufacturing [27, 76].

Instead of using simulation packages, some publications performed economic analysis and optimization of biopharmaceutical production by using programming software such as GAMS. Liu et al. [61] applied mixed-integer nonlinear programming (MINLP) to provide sizing strategies, process debottlenecking and decision making of facility configuration in order to build cost-effective chromatography. In 2016, Liu et al [25] applied stochastic mixed integer linear programming (MILP) model to optimize both upstream and downstream equipment sizing, and their working cycle determinations. Combined with chance-constrained programming (CCP) techniques, the most cost-effective decision was made with the consideration of upstream titer and chromatography resin yield. However, these methods have been only applied on mAbs production in batch mode.

Economic analysis method

The economic analysis method includes deterministic cost calculation, sensitivity cost analysis, and stochastic analysis. Since many of the cost calculations of biopharmaceutical manufacturing are the same as those of chemical plants including labor cost, maintenance cost, and even profitability cost, their cost calculations are not going to be covered in detail in this review. This section only presents the calculation of biopharmaceutical manufacturing cost.

For capital cost calculation, there are three main items: working capital, start-up and validation cost, and direct fixed capital (DFC). In biopharmaceutical plants, working capital is 10-20% of the DFC, while start-up and validation are typically 20-30% of DFC. For single-use system, the processing flexibility increases and validation, start-up and commercialization times become shorter [77]. The calculation of capital cost is similar to the calculation done for chemical plants using available scaling relations [77] as in equation (2.2.1).

$$C_2 = C_1 \cdot (\frac{size_2}{size_1})^a \tag{2.2.1}$$

Where C_2 is the target cost of the equipment; C_1 is the reference cost; $size_1$ and $size_2$ are sizes of reference equipment and target equipment; a is a scale-up factor and ranges between 0.5 and 1. In biopharmaceutical calculation, two equipment have been used [61]:

$$C^{bior} = C_{ref}^{bior} \cdot (\frac{V^{bior}}{V_{ref}^{bior}})^{SUF^{bior}} \tag{1.2.2}$$

$$C^{Col} = C_{ref}^{Col} \cdot (\frac{D^{Col}}{D_{ref}^{Col}})^{SUF^{Col}} \tag{2.2.3}$$

V is the bioreactor volume; *D* is chromatography diameter; *SUF* is a scale-up factor.

Cost of goods per gram (COG/g) combines direct and indirect costs of the process and is widely used in analyzing the impact of investment and operating decisions. The COG/g analysis can be applied in both single steps comparison and integrated manufacturing process.

$$COG = \frac{annual\ direct\ cost + annual\ indirect\ cost}{annual\ product\ output} \tag{2.2.4}$$

In equation COG/g equation, material cost and consumables cost belong to direct cost. In biopharmaceutical process, material cost includes media, buffer and cleaning cost (for example,

waste for injection cost) [77]. Liu et al. [61] built cost calculation model for batch biopharmaceutical facilities and provided the equations to calculate media cost:

$$C^{media} = \theta \cdot N^{batch} \cdot Pc^{media} \cdot \alpha \cdot V^{bior} \qquad (2.2.5)$$

θ is media over-fill allowance, and N^{batch} is number of batches of bioreactor, while Pc^{media} is the price of media. α is bioreactor working volume ratio, and V^{bior} represents the volume of the bioreactor.

Consumables are the items that need to be replaced during processing, for example, filters and membranes in filtration, diafiltration, and resins in chromatography. In some studies, the resin cost is also considered as a material cost. In most of the cases, it is considered as a consumable cost. When calculating the cost of consumables, lifetime and reusable cycles need to be considered. For example, in Liu et al. [61] study, the resin cost in chromatography column was calculated as:

$$C^{resin} = \sum \frac{A \cdot Pc_s^{resin} \cdot N^{batch} \cdot N_s^{cyc} \cdot TotV_s^{col}}{L} \qquad (2.2.2)$$

A is an over-packing factor for resin, and Pc_s^{resin} is the resin price. N^{batch} is number of batches, while N_s^{cyc} is number of cycles per batch. $TotV_s^{col}$ is total column volume, and L is the resin lifetime, while S represents the different chromatography steps.

Sensitivity and Monte Carlo analysis can also be applied in economic evaluation. Sensitivity analysis can help with decision making regarding testing solutions, identifying variables, finding the optimal solutions and evaluating the risk [78]. In economic analysis, we can observe the most sensitive parameters relative to the cost and help to design the most cost-effective process under the existing conditions.

Monte Carlo analysis can help predict the system response to certain events with specific probabilities. By defining the parameters' minimum, maximum and probabilistic distributions, this analysis helps to understand the real manufacturing behavior with two or more key parameters'

variation. It can also help to analysis uncertainty and failure rates exist in real manufacturing process.

Since the output of analysis is the cost, the sensitivity and Monte Carlo analysis in economic analysis of biopharmaceutical production can be calculated by Visual Basic Analysis or Crystal Ball (Oracle). In sensitivity analysis, the variation of sensitivity analysis can be determined from literature resources or experimental data and define best, base and worst case scenarios[79]. As for Monte Carlo analysis, random values need to be generated based on the distribution defined by user: uniform, normal and triangular distribution are usually used. In bioprocess with known best, worst, and base case scenarios, triangular distribution is normally applied [59, 79].

Deterministic cost analysis

Deterministic cost analysis mainly includes calculation of investment capital cost, COG/g and cost in contributions in different categories, including material, depreciation, and utilities cost. By comparison of categories, the cost bottleneck of each operation can be found [59, 67, 69, 80].

In both upstream and downstream processing, capital cost, material cost, consumable and labor cost are important categories of the overall cost and thus have the highest weight compared to other cost categories. Continuous processing results in smaller footprint by using smaller perfusion bioreactors with high cell concentration, smaller size of chromatography columns and elimination of hold tanks between operations [8]. In upstream, understanding and optimizing labor cost and material cost is very important since labor cost constitutes a large portion of the overall perfusion COGs/g. Vermasvuori and Hurme [80] showed that among different cost categories, approximately 30 % of the cost is from labor cost for stirred-tank perfusion and hollow fiber perfusion bioreactors. Regarding the material cost, there are great differences between batch and continuous processing. In the continuous perfusion process, media is continuously fed into and collected out of the bioreactor, and in batch or fed-batch process, media is maintained in the bioreactor. Pollock et al.[59] showed that material costs take 38% of total cost in continuous processing and only take 24

% in batch processing. The specific cell line used and the media composition greatly affect product titer and productivity. Under the assumption of keeping the same production rate, selecting a cost-effective media is an important consideration [8]. Xu et al. [69] compared the productivity, media cost among fed-batch, concentrated fed-batch and perfusion bioreactor under same cell line and media conditions. The result showed that media cost for perfusion process (2.29±0.28 g/L/day) was lower than that in fed-batch (on day 14, titer 6.8±0.2 g/L) and concentrated fed-batch process (on day 18, titer 27.5 g/L). This comparison was based on an initial assessment, and media optimization can also be explored in a future study [69]. Klutz et al. [66] discussed the comparison between fed-batch operation in fully integrated batch process and perfusion operation in fully integrated continuous process at 200 kg/yr manufacturing scale. The result showed the COG/g from perfusion upstream process is 1.2-3.2 higher than fed-batch process due to higher media cost.

In commonly used batch biopharmaceutical manufacturing, up to 80 % of total manufacturing cost is due to downstream processes [81]. More than 70 % of this downstream cost is dominated by chromatography and the material cost [82]. In addition, the decision of resin and filter reuse plays important role in downstream material or consumables cost. Especially in protein A chromatography, the resin cost is higher than other chromatography options and accounts for a large portion of the overall raw material cost [83]. From previous literatures, the cost of protein A resin is 10,000-15,000 $/L and AEX resin cost is 1500 $/L [60]. Thus, many researches focus on the study of protein A chromatography in order to reduce the resin use in the process or find another substitute equipment. Torres-Acosta et al. [68] compared the economic cost between chromatographic purification and aqueous two-phase extraction (ATPS) of uricase, an intracellular product enzyme, purification in batch mode. The deterministic analysis shows that the chromatographic option has higher capital cost, consumables cost (where resin cost takes a high percentage of consumables cost), and lower raw material cost compared to the ATPS method. The ATPS process can also be applied in continuous mode [84] however no economic analysis has been

performed yet. Pollock et al. [60] evaluated the periodic counter current (PCC) semi-continuous chromatography including column design, operation, and economic analysis. They applied economic analysis to compare PCC chromatography and conventional batch process. The results showed that at the stage of PoC (proof-of-concept) 4 kg/batch scale model, the PCC method reduced resin use by 50 % leading to reduction of batch manufacturing direct costs by 31 %. The three columns system resulted in buffer usage reduction by 39 % in protein A chromatography operation and reduced overall buffer usage by 12 %. However, the investigators pointed out that these benefits become less significant as the production scale increases.

Different from upstream and downstream processing, the analysis of integrated processing not only focuses on overall cost but also analyzes different cost categories in order to evaluate the contribution of each unit operations to the whole process. Cash flow and net present value are also analyzed for future decision making. Walther et al. [64] compared integrated continuous biomanufacturing (ICB) with conventional batch process in mAbs and nonemAbs production. The publication showed integrated continuous process reduced operation cost by 21 % and capital cost by 47 % in mAbs production and 80 % and 72 % cost for non-mAb production respectively. The breakdown cost analysis showed the ICB process has higher upstream filters and media cost but lower downstream resins cost. In every unit operations, the capital cost is lower compared to conventional process. NPV results provided information that the ICB platform can obtain benefits immediately after the manufacturing start. Arnold et al. [71] compared integrated continuous antibody production with batch processing using experimental and computational results. By keeping the same annual production rate, the continuous processing has 15 % lower operating cost in COG/g and the capital cost in continuous processing reduces 50 % comparing to that in batch.

Sensitivity and stochastic cost analysis

Sensitivity analysis can be applied to analyze the effects of process parameters in order to determine the most important parameters in terms of cost. Bunnak et al. [67] compared fed-batch and

perfusion operation for mAbs production using sensitivity analysis to identify the most sensitive parameters in each system. The results indicated that titer, bioreactor working volume, and perfusion rate are three crucial process parameters to overall COGs/g. It also indicated that increasing the pooling duration in the process can decrease the overall COGs/g. Longer pooling duration brings lower frequency of both upstream and downstream processes but larger downstream equipment needs to be applied as a tradeoff. Another highlight of this work is that the paper applied life-cycle assessment and used sensitivity analysis to investigate water consumption, solid waste generation and energy requirement of the process. Table 2.2 1 summarizes the main variables of perfusion bioreactor, multicolumn chromatography and integrated process that have been used in sensitivity analysis.

Most commonly, sensitivity analysis is applied using deterministic methods, which involves the evaluation of cost at a few points of independent variables in a certain range. The method is used to analyze protein titer and production capacity effects on upstream, downstream and integrated processing cost in order to compare the flexibility of various operating modes. Bunnak et al. [67] found as the annual production rate increases from 28 kg/year to 1000 kg/year, the capital cost contribution to overall COGs/g decreases in both batch and perfusion process and consumables and materials cost increases. Pollock et al. [59] analyzed production scales and product titers effects on overall cost with the use of different types of bioreactors. The results showed that as the production scales increases from 100 kg/yr to 1000 kg/yr, and titer increase from 2 g/L to 10 g/L, the advantages of perfusion bioreactor becomes less significant. The paper also analyzed the working conditions of bioreactor such as viable cell density. It shows that as long as the viable cell density in perfusion bioreactor can be maintained 3 folds of that in fed-batch bioreactor, the ATF perfusion process will save cost in any scale.

Hammerschmidt et al. [62] compared the integrated hybrid and fully continuous precipitation process with traditional batch chromatography process. Three operation modes were tested under

different scales, including development and fully commercial production scale. The result showed that as the scale increases, the highest cost of batch process shifted from consumable cost to capital cost. For continuous and hybrid process (batch process upstream with continuous precipitate downstream), the highest cost shifted from labor cost to material costs. In all scales, hybrid process has the lowest cost comparing to other operation modes. The study showed that with an increase in titer capacity from 2 g/L to 10 g/L, the cost of downstream cost contributions increases from 48 % to 72 % of the overall cost. A similar analysis is also applied to integrated analysis with multicolumn chromatography downstream process. Pollock et al. [70] showed the comparison among batch, hybrid and continuous integrated process. The work showed the trend of media and buffer costs, single-use component cost, chromatography resin and QC/QA cost as the operation increase from Phase I development scale to commercial scale. In addition to comparing break down cost of continuous and batch process, the author also analyzed the impact of manufacturing scales with company sizes in COGs/g of different operation modes. The work provided the analysis of operational feasibility and evaluated the process environmental effect. It also applied sensitivity analysis on decision making by calculating an overall aggregated score that is a multi-attribute decision-making methodology to represent environmental economical and operational robustness of the system changed by the importance of economic benefits.

Stochastic cost analysis such as Monte Carlo simulation can be applied to evaluate variability and robustness of the process. Pollock et al. [59] showed that under 500 kg/year annual production rate, and 5 g/L titer condition, the amount of annual output of ATF perfusion process was higher comparing to batch process, however, it had lower robustness because of high standard deviations. ATF perfusion process has lower COG/g than fed-batch, even at the worst-case situation. Walther et al. [64] applied Monte Carlo analysis to estimate the overall NPV of continuous facility with the possibility of product or technology transfer success with varying product demand.

Furthermore, the sensitivity analysis and Monte Carlo analysis can be applied to evaluate the effect of parameter optimization. Torres-Acosta et al. [79] applied economic modelling tool to evaluate royalactin production with aqueous two phase systems. Through sensitivity analysis, the upstream titer was found as one of the key parameters affecting the cost of production. By evaluating the optimized process, Monte Carlo analysis shows the distribution of COG/g and decide the most feasible option to investigate.

These two analyses also play an important role in process selection and decision making. Torres-Acosta et al.[85] designed four types of ATPS processes for tetracycline purification and then optimize the process by sensitivity analysis. For process decision making, Monte Carlo analysis presented downstream process yield and upstream titers' effect on COG/g distribution in all the scenarios. As the result, ATPS formed by cholinium chloride/K_3PO_4 are selected.

Table 2.2 1 Variables for sensitivity analysis

Operation	Variables
Perfusion Bioreactor	Perfusion runtime, pooling duration, perfusion rate, bioreactor working volume, protein titer, failure rate [67]
Multicolumn chromatography	Binding capacity, chromatography yield, number of cycles, protein titer, A column lifespan, resin, filter capacity, material prices, number of facility staff. [60, 65]
Integrated process	Protein titer, failure rate, Company sizes, manufacturing scales (preclinical scale, proof-of-concept scale (phase I, phase II), phase III, commercial scale) [23, 64, 66]

2.2.2 Design space of biopharmaceutical processes

Design space is defined in the ICH Q8 guidance document as " the multidimensional combination of interaction of input variables and process parameters that have been demonstrated to provide assurance of quality" [86, 87]. These input variables can be process parameters such as temperature, pH, flow rate, feeding strategies, pressure but also can be materials, size of the equipment. The product quality is specific to different products and processes as described in Chapter 2. Design space provides acceptable operating ranges as well as the edge of process failure, which allows the

operational flexibility and ensures there are no significant quality or regulatory problems during the operation.

The general approach for defining a design space including 1) define parameters based on prior knowledge and risk assessment 2) use experiments to determine critical parameters 3) finding operating ranges by experimental or modeling approaches 4) Validate the design space in manufacturing operations [87]. In biopharma, current method to obtain the design space is to use design of experiment (DoE) supporting with statistical analysis and risk analysis tools [88]. For DoE, central composite design [89] and Plackett-Burman [90] design have been used to investigate the operating range and provide information for the regression model building. Response surface modeling [91] or multi-variate analysis[92] are powerful statistical analysis methods that are used for predictive model building. Risk analysis such as Failure mode and effects analysis (FMEA) [93, 94] and Monte Carlo analysis [91] are applied to ensure the robustness of the design space with respect to lot to lot differences. However, due to the poor extrapolation of the statistical model, the DoE has to be wide enough to capture process success to find a reliable process boundary. At the same time, with the multi-dimensional variables, finding a design space that satisfies all the process parameters would need a lot of materials and time.

In recent years, the mechanistic model that provides better process understanding and predictability has been developed. By testing a wide range of parameters and adding process perturbations, the design space can be well defined [95]. A contour plot can be generated by testing the whole operating range using mechanistic model. System engineering tools such as sensitivity analysis and process optimization methods can also apply to both upstream and downstream operations. Global optimization methods have been used to optimize the operating parameters. For example, Pirrung, S. et al. [96] built hybrid model for 3 chromatography system and applied process optimization simultaneously to maximize production yield while maintaining 99.9% minimum purities. Baur, D. et al. [97] modeled twin-column CaptureSMB optimized process by capacity utilization and

productivity with constant yield and purity. To define a design space, sensitivity analysis has been used to investigate the feeding strategies. Kiparissides, A. et al. [98] applied global sensitivity analysis to identify a feeding regime to maximize mAb production. Kotidis, P. et al. [99] used constrained global sensitivity analysis explored >8000 feeding strategies in silico and identified design space to meet required antibody titer and glycan fractions. Although these models reduced the materials and time that needed through experimental screening, they are still computationally expensive.

3 End to End Batch and Continuous Biopharmaceutical Manufacturing Design

3.1 Background

As discussed in Chapter 1, continuous manufacturing reduces the equipment footprints compared to the corresponding batch or semi-batch process, and continuous operating mode offers rapid capacity adjustment and process scaling up. The high demand of the biopharma market brings the development of continuous manufacturing as a future direction.

In this Chapter, biopharmaceutical manufacturing efforts toward process intensification are presented and evaluated. Flowsheet modeling is used for plant decision-making, including new process design, existing process optimization, and novel process evaluation. Detailed integrated fed-batch (with fed-batch upstream and batch downstream) and continuous processes are developed to produce mAbs and gene therapy products using Superpro designer (Intelligen, Scotch Plains, NJ). The analysis considers material balances, process scheduling, and equipment occupancy charts to enable a comprehensive evaluation of the two alternative production routes. Deterministic cost including different cost categories is utilized to evaluate the improvement of continuous processing over fed-batch in base case scenario. Different scenarios are built using sensitivity analysis to examine the effect of important operating parameters and manufacturing scales on overall continuous process performance.

3.2 Modeling method

As mentioned in the Chapter 2, three simulation software packages have been commonly used for cost analysis: SuperPro Designer (Intelligen, Scotch Plains, NJ), Bio-Solve (Biopharm Services, UK), and Aspen Batch Process Developer (Aspen Technology, Burlington, MA). This process design and analysis in this study is based on SuperPro Designer.

SuperPro Designer is a recipe-driven simulation tool that can be used to both batch and continuous integrated biopharmaceutical manufacturing process modeling. By adding a sequence of tasks in each unit operations, material and energy balances can be calculated and tracked through the whole process. Users have an option to choose either the rating or design mode in their process simulation. Under rating mode, the design parameters of equipment can be set, and different products or different production rates can be tested in the model for facility fit. Under the design mode, the maximum equipment dimensions can be set, and the simulator can automatically calculate the design parameters for users. Process debottlenecking for example reducing the starting time between two-unit operations can also be achieved in the simulation tool. The software provides equipment occupancy chart, gantt chart, material consumable chart labor required chart and throughput analysis to track and analyze detail of process operation. SuperPro designer has the capacity for business modeling, such as economic analysis, cost of goods analysis and essential parameters can be adjusted within the model or transferred to Excel for further analysis.

Since Aspen Batch Process developer is specific on batch process and continuous chromatography and Bio-Solve is specific on economic analysis. SuperPro designer provides flexibility on process design, facility analysis and process scheduling on both batch, continuous and semi-continuous processing which is the best fit for both process design and economic analysis. Thus, the model in this study is built in SuperPro designer simulator, which contains detailed material balance, cleaning methods and equipment scheduling. Economic analysis and throughput analysis are also performed in the Superpro designer for process evaluation.

3.3   Economic analysis

Total capital investment and operating cost are two major costs in economic analysis. Total capital investment which includes direct working capital, startup and validation costs, and direct fixed capital cost, is calculated based on equipment expenditure. Direct fixed capital contains equipment purchase cost, installation, piping, instrumentation, insulation, electrical, buildings, auxiliary

facilities, engineering, construction cost as well as contractor's fee and contingency. Each cost is estimated by multiplying a certain coefficient to equipment purchase cost as listed in Table 3.3 1 Economic analysis capital cost estimation following the work of Harrison et al [77].

Table 3.3 1 Economic analysis capital cost estimation

Cost item	Multiplier
Piping	0.35
Instrumentation	0.35
Insulation	0.03
Electrical facilities	0.15
Buildings	0.45
Yard improvement	0.15
Auxiliary	0.5
Engineering	0.25
Construction	0.35
Contractor's fee	0.05
Contingency	0.10
Startup and validation	20%

The cost of equipment is evaluated based on literature and vendor resources. Working capital represents the assets that must be available to ensure the facility processing and includes 1 month of labor, raw material cost, utilities and waste treatment cost. Scaling rate is applied to estimate capital cost as shown in Eq. 3.3.1.

$$C_2 = C_1 \cdot (\frac{size_2}{size_1})^a \qquad\qquad 3.3.1$$

The operating cost contains material, consumables, utilities, labor-depended, quality control (QC) and quality assurance (QA) cost and facility dependent cost (including the cost related to use of facility: equipment maintenance, depreciation of the facility and miscellaneous cost). Material and consumables costs are referenced from literature and online vendor resources. QC/QA cost is estimated as 15% of total labor cost. The total operating cost of the product ($/g) is also named as cost of goods per gram (COG/g) and it is usually used to calculate operating cost per unit of

production. In this paper, we redefine the COG/g as shown in Equation (3.3.2) and consider both total capital investment and operating cost.

$$COG = \frac{\frac{Total\ capital\ investment}{Years\ of\ operation}(\frac{\$}{yr}) + Annual\ operating\ cost(\frac{\$}{yr})}{Annual\ production\ rate\ (\frac{g}{yr})} \qquad (3.3.2)$$

3.4 Case study 1 mAb production

Monoclonal antibody (mAbs) are used for the treatment of cancers, autoimmune diseases, inflammatory, infectious, and microbial diseases[102]. mAbs are widely used with four out of the top five therapeutic proteins being mAbs and among all the therapeutic proteins, mAbs count for almost half of the selling amounts [103]. It is also predicted that the mAbs market size will be worth $138.6 Billion by the year 2024 [102]. As a highly demanded biological product, the market and development of mAbs are rapidly grown in recent years – four new products are approved every year and 78 have been approved from 2017 [104].

In this case study, fed-batch and continuous mAb production lines are designed to evaluate benefits of continuous operations. Different analysis approaches including deterministic cost analysis, sensitivity analysis are used to discover the possibilities and challenges of continuous applications in future biopharmaceutical manufacturing.

3.4.1 Process description

Biopharmaceutical manufacturing for mAbs production contains the following steps: inoculation, cell cultural and protein production, clarification, primary capture, polishing, and final formulation. **Error! Reference source not found.** and **Error! Reference source not found.** provide process op eration flowsheets for fed-batch and continuous processes respectively. Inoculation process is applied to cell culture scale up by using a series of cell culture passages [105] such as test tube, T

flasks, shake flasks or seed cell culture bioreactor which are shown in Error! Reference source not f ound.. The purpose of this step is to grow the cells to certain density and viability before transferring them to production bioreactor. From literature, inoculation cell density ranges between 0.2-1x10^6 cells/ml [106-110]. Some publications show that the density can reach as high as 2-3 x10^6 cells/ml [111]. In this paper, 1.5x10^6 cells/ml is considered as the peak cell density in inoculation bioreactor. Fed-batch and continuous processes share the similar inoculation bioreactor. Extracellular protein such as mAbs is traditionally produced in fed-batch production bioreactor under temperature 36.5-37 ℃, with dissolved oxygen 30-50%. Production media, antifoam are added at the beginning of the operation with adjusted dissolved oxygen concentration, and pH. Then cells from inoculation are transferred to production bioreactor. When the cell concentration reaches 0.7 x10^6 cells/ml, nutrients, such as glucose, Hy-Soy are added. Nutrients, base, process air and antifoams are added based on the culture requirements. In this way, the bioreactor is actually operated as fed-batch bioreactor with additional nutrient addition. In the fed-batch model, the cell culture can last for 14 days with the titer of 1.5 g/L. The cell density is assumed 5.5x10^6 cells/ml at the end of the cell culture and productivity is calculated as 0.107 g/L/day with cell specific productivity 37 pg/cell/day. After the fermentation, material from bioreactor is transferred to centrifuge, where biomass and most impurities are removed. Before transferring the product to the primary capture step for further purification, several filtration steps need to be considered to remove the remaining cells in the solution and protect protein A column and resin from contamination and clogging. In industrial manufacturing, purification through protein A column is one of the most important steps in downstream processing, which removes most of the impurities such as host cell protein (HCP), DNA, media component and virus particles [112]. The primary capture operation contains the following procedures:

1) Pre-sanitization and pre-equilibrium: using acidic buffer, alkaline buffer, salt buffer and water to wash and activate resin;

2) Equilibrium: salt buffer sodium chloride and sodium phosphate are applied to adjust pH and prepare for protein loading. This step prevents the protein aggregation and deactivation during loading process;

3) Loading: protein binds to the resin with little amount of impurities during this step, and most of the impurities, water, and unused media is removed in flow through;

4) Washing: salt buffer sodium chloride and sodium phosphate are used to remove unbonded impurities and residuals;

5) Elution: protein product with minor impurities are recovered from resin and transported to the next step by using acid salt;

6) Regeneration and final wash: acid and alkaline buffer are used to remove protein and impurities left inside of column.

During the primary capture, the column usually runs for several cycles to finish the purification of one batch. This is done since a large amount of protein is produced from the bioreactor, and the capacity of protein A column is limited by resin binding capacity (30 g/L). Due to the protein loss during loading, washing and elution steps, the protein yield of this unit operation is assumed 95%. By adding acid, virus inactivation of elution proceeds in blending tank for 1 hour to prevent virus growth. Alkaline buffer is then added to neutralize the solution pH and followed by diafiltration to prepare for polishing step. Anion exchange chromatography (AEX) working under flow-through mode is used for polishing step to remove the residual impurities, such as DNA, and final formulation is realized by ultrafiltration and diafiltration. In the model, ultrafiltration and diafiltration are operated in the same equipment, and protein concentration is adjusted during these operations. In fed-batch process, pooling tank needs to be used between each unit operations to maintain the process sequencing and free up expensive or time-consuming unit operations [113]. Steaming in place (SIP) and cleaning in place (CIP) are also applied and scheduled before and after equipment usage, including seed bioreactor, production bioreactor, storage tanks, virus inactivation

tanks, diafiltration, and ultrafiltration. It is need to point out that Figure 3.4 1 and Figure 3.4 2 only show the main operations in the two processes operations. Pooling tanks, filters are not included in these figures but are considered in the simulation.
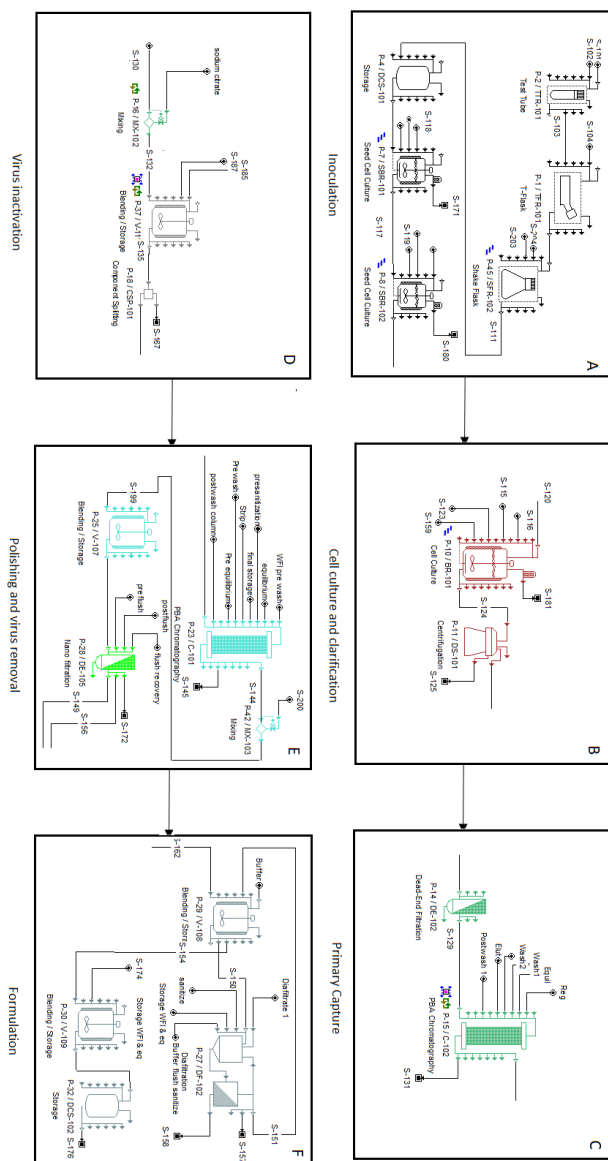

Figure 3.4 1 Fed-batch process flowsheet for monoclonal antibody production

In perfusion bioreactor, high cell densities can be achieved using continuous fresh nutrient fed and continuous product, by-product, and waste removed from bioreactor. Under this mode of operation,

the percentage of cell viability, product quality, and volumetric productivity can be improved compared to fed-batch process. It is reported that cell density of perfusion bioreactor can reach 20-60x10^6 cells/ml and most cell density of perfusion reactions are within this range[59, 66, 109-111, 114-116]. Walther et al. [64] achieved density as high as 100x10^6 cells/ml with an annual production rate of 1236 kg/yr. The perfusion bioreactor viable cell density chosen for this paper is 60x10^6 cells/ml with volumetric productivity of 1.68 g/L/day and cell specific productivity 28 pg/cell/day. Chotteau et al. [111] found that the product quality varies during the early stage of the bioreactor culture and productive stage. The product harvests should be started a few days after the system reaching cell density target. However, the time should be balanced against the overall productivity in manufacturing process. In this, 4 days are considered as the setup time before the cell cultural reaching steady state. However, the protein lost during this period is not considered. 0.02 um Tangential Flow Filtration (TFF) is connected to perfusion bioreactor which is used to separate cells from liquid product and return cells to fermentation. Cell bleeding which is used to maintain cell density in perfusion bioreactor is not included in this system and the cell density is assumed constant. The fermentation lasts for 24 days even though some papers use up to 60 days [59, 64, 110]. Filter clogging, and fouling may occur during perfusion so maintaining proper operation can reduce the failure of bioreactors. Similar to batch processes, filters are also added before the periodic countercurrent protein A column (PCC) for primary capture. Blending storage tank is designed to distribute the product to different columns. Four columns work in parallel which mostly increase the resin usage and reduce the protein loss and buffer consumption. The resin binding capacity for PCC process is 60 g/L. A pooling tank is installed right after the protein A column to collect the protein before transferring the product to virus inactivation. Two virus inactivation tanks work alternatively i.e. one tank is working under acidic inactivation while the other is cleaning. Two static mixers and pH adjusters are applied before and after the inactivation tanks. Similarly, two AEX columns also work alternatively to achieve continuous polishing. Before the polishing step, countercurrent staged defiltration is applied for continuous buffer exchange with

5 diafiltration volumes which replace nearly all of the buffer and salt by sodium phosphate [117]. Nanofiltration removes virus in the product. Protein is further concentrated by using staged single pass tangential flow filtration (SPTFF) and buffers are exchanged by using countercurrent staged diafiltration. The lifetime of membrane in filtration is assumed to be 10 cycles. Casey et al. [118] applied high concentration SPTFF cycles testing and approved that the ultrafiltration can complete 10 cycles successfully and concentrate IgG from 45 g/L to 200 g/L. In this paper, the continuous system only requires protein concentrate from 11 g/L to 30 g/L. Thus applying 10 cycles lifetime might result in an over estimation of the consumable cost. Since all the processes operate in a continuous mode under lower flow rate, no large pooling tanks are needed between unit operations. The most important operating parameters are summarized in Table 3.4 1. The maximum operating time of the two processes is 330 days/yr.
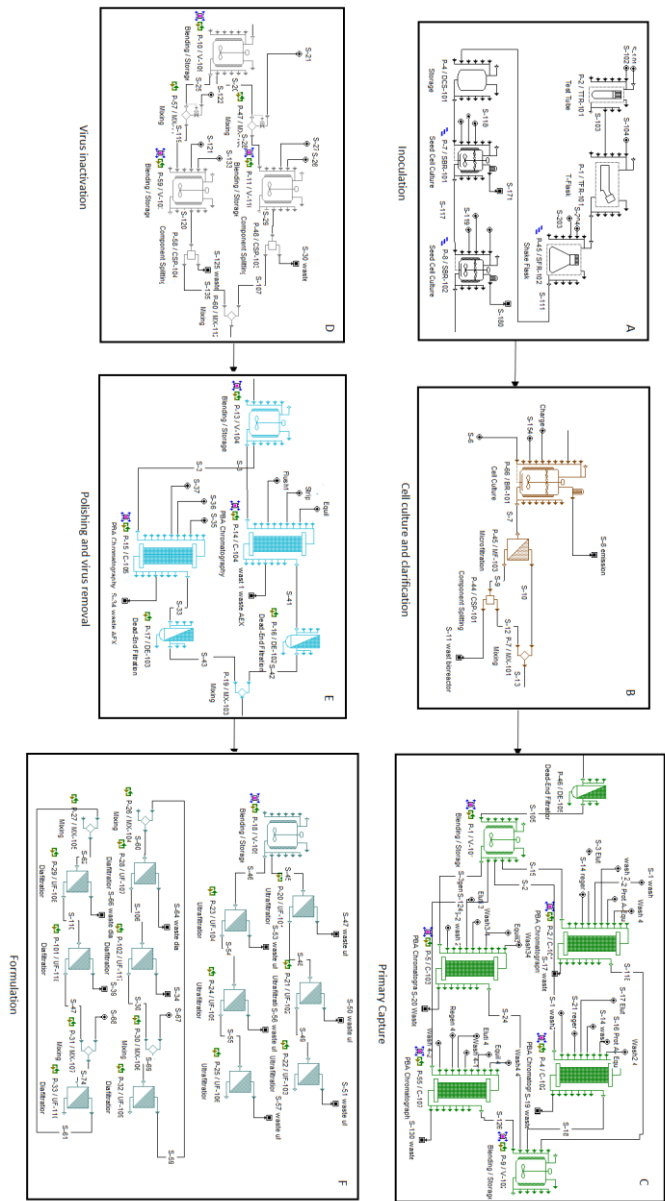
Figure 3.4 2 Continuous process flowsheet for monoclonal antibody production

Table 3.4 1 Operating conditions

	Fed-batch Processing	Continuous Processing
Inoculation cell density (seed)	1.5×10^6 cells/ml	1.5×10^6 cells/ml

Bioreactor cell density (production)	5.5×10^6 cells/ml	6×10^7 cells/ml
Max bioreactor volume set	15000 L	3000 L
Perfusion rate		1 vvd
Downstream yield	75%	75%
Cell cultural time	14 days	24 days
The annual number of batch	48	9
Annual production rate	618 kg/yr	624 kg/yr

In downstream process, the yield of each unit operation is shown in Table 3.4 2. The yield for both batch and continuous downstream process is 75%.

Table 3.4 2 Downstream unit operations yield

	Clarification	Primary capture	Virus inactivation	Diafiltration	Polishing	Virus removal	Final filtration
Batch	99%	96%	95%	96%	95%	98%	95%
Continuous	99%	98%	95%	91%	95%	98%	96%

Equipment occupancy charts are shown in Figure 3.4 3 and Figure 3.4 4. One equipment is chosen to represent each unit operation to simplify the results' presentation. For example, in Figure 3.4 3, the inoculation bioreactor is shown to represent inoculation process while the inoculation flasks, and tubes are not shown in the chart. In fed-batch process, the operation starts right after the previous operation finished. In continuous process, all the major steps of operations start almost at the same time, except the inoculation and CIP and SIP processes. The annual production rate can

be reached by using multiple equipment in fed-batch process. It is found that, process bottleneck of fed-batch process is in upstream, thus, two seeds bioreactors and four production bioreactors work in parallel and each of those bioreactors start a new batch to reduce the time between consecutive batches. In this case, the number of equipment utilization between fed-batch and continuous processes is different. Scheduling is also applied to adjust the order of each procedure in different unit operations. For example, in both fed-batch and continuous processes, product transferring in task is manually linked to the product transferring out task from previous operation. In continuous process, the operation needs to be performed continuously under the same time period as the previous operation. CIP and SIP are manually adjusted so that maximum numbers of operations can be fitted to the minimum amount of CIP and SIP units.
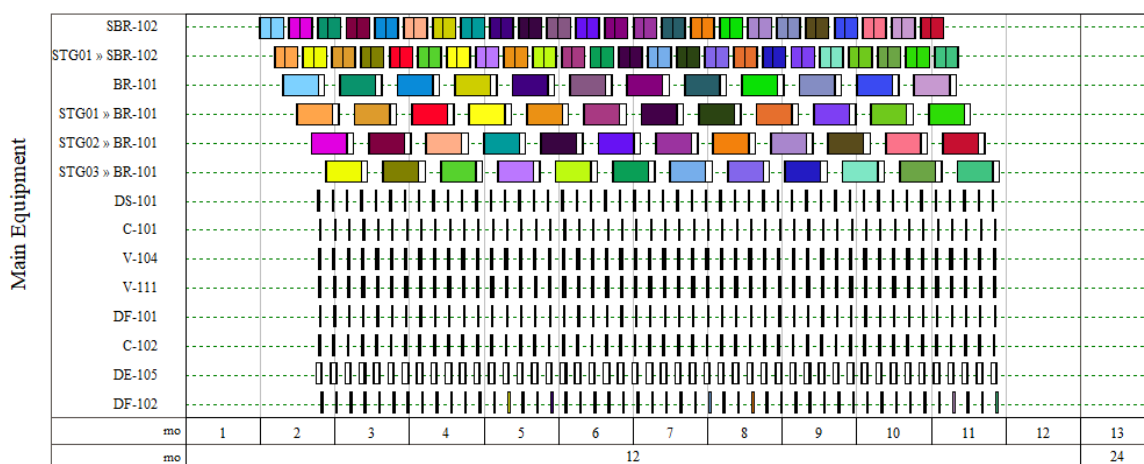
Figure 3.4 3 Equipment occupancy chart for the fed-batch process (SBR: inoculation bioreactor, BR: production bioreactor, DS: centrifuge, C-101: protein A column in primary capture, V104 and V111: virus inactivation tank, DF-101: diafiltration, C-102: AEX column, DE: virus removal, DF-102: diafiltration and ultrafiltration)
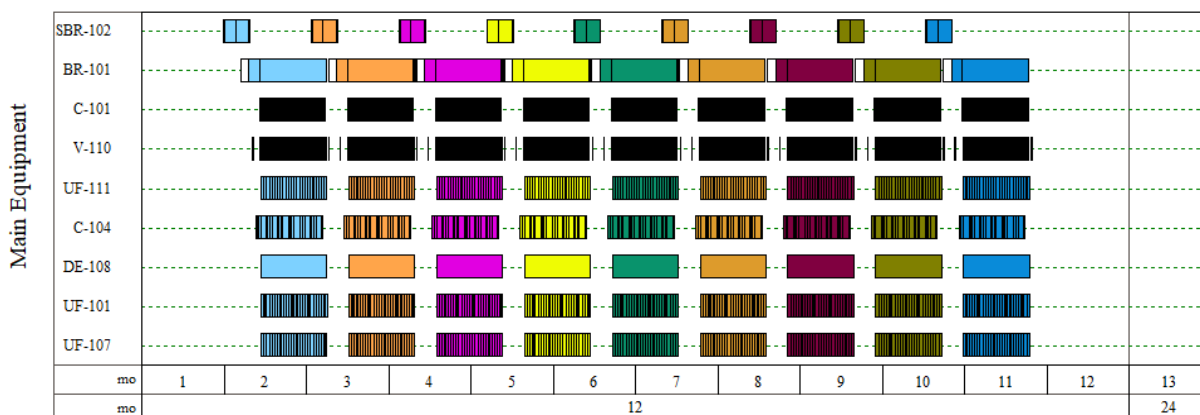
Figure 3.4 4 Equipment occupancy chart for the continuous process (SBR: inoculation bioreactor, BR: production bioreactor, DS: centrifuge, C-101: one protein A column in primary capture, V110: virus inactivation tank, UF-111: diafiltration, C-104: AEX column, DE: virus removal, UF-101: ultrafiltration, UF-107: diafiltration)

As for economic analysis, the most important prices and parameters that can be used in economic analysis are listed in the Table 3.4 3. Regarding the selling price, the prices of monoclonal antibodies that are approved by the FDA from the year 1997 to 2016 are shown in Figure 3.4 5 [119]. It can be noticed that most of the prices are below 100 $/mg. 20 $/mg is assumed as selling price in base case scenario which represents the average selling prices of drugs below 100$/mg.
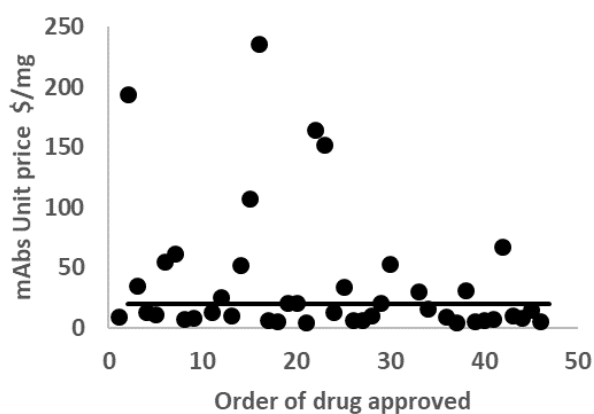
Figure 3.4 5 Monoclonal antibody price approved from 1997 to 2016 (The figure is plotted based on literature data) [119]

Table 3.4 3 Unit prices

Equipment	Ref. Size	Ref. Cost	Ref. Year	Index	Reference
Bioreactor	2000 L	$716,024*	2016	0.6	Liu et al.[1]
Centrifuge	40 L/min	$760,000*	2015	0.25	Klutz et al.[2]
Filtration skid (applied to all dead end filter)	1 m^2	$8057*	2015	0.58	Klutz et al.[2]
Xcell ATF 10	1 item	$180,000	2013		Pollock et al.[3]
Chromatography	D=1m	$198896*	2016	0.8	Liu et al.[1]
PH and concentration adjustment	1 item	$14039	2015		Kultz et al.[2]
SPTFF	0.12 m^2	$5428	2018	0.8	Calculated from Pall Company website

*calculated cost with currency conversion

Consumables cost	Cost	Reference	Assumptions
Clarification membrane (applied to process after cell clarification)	776 $/m^2	Klutz et al.[2]	Batch: replace every cycle Continuous: replace every day
Xcell ATF 10 consumable cost	$16,300	Pollock et al.[3]	
Protein A Resin cost (batch)	30g/L, 7484* $/L 100 cycles	Liu et al.[1]	
Protein A resin cost (cont.)	60g/L, 18120 $/L		300 cycle number is assumed
UF membrane	981 $/m^2	Li,Y. and Venkatasubramanian, V[4]	Batch: replace every 25 cycles Continuous: replace every 10 cycles
AEX resin cost	819* $/L 100 cycles	Liu et al.[1]	
Sterile filtration membrane	4137 * $/L	Klutz et al.[2]	Batch: replace every cycle Continuous: replace every 60 hrs

Disposable cost	Cost	Reference	
Membrane and filter	2 $/m^2	Assumed	Klutz et al.[2] Mentioned the membrane and filter disposable costs is 10 Eruo/module by assumption
Waste water	0.05 $/kg	Klutz et al.[2]	
Plastics	9.3 $/kg	Klutz et al. [2]	
Bioreactor disposal	0.32-0.35$/kg		

Sensitivity analysis is also used to evaluate the effect of different parameters considered in simulating the continuous process. Parameters considered are the upstream titers, downstream yields and manufacturing scales to the overall process economic benefits. In fed-batch and continuous processes, most facilities are interconnected and the operating time of one procedure is scheduled depending on another process. Due to the complexity of the process, different scenarios are designed for process evaluation. For upstream titer analysis, facility size is kept constant while titer varies from 1.5 g/L to 5.5 g/L. The downstream yield in the base case scenario is 75%. Different scenarios with yield of 80%, 85%, and 90% are analyzed for fed-batch and continuous process. For plant capacity analysis, the facility size varies from 50 g/L to 1200 g/L and all the other operating conditions are kept constant.

3.4.2   Cost comparison: fed-batch and continuous processes

This section provides the comparison between fed-batch and continuous processes based on the base case scenario where the production rate is adjusted to 620 kg/yr and selling price of mAb is assumed $20/mg based on the analysis in Figure 3.4 5. The analysis includes capital investment cost and operating cost with cost breakdowns for both operating modes. The result shows that the overall capital cost in fed-batch process is $165 million, which is 3 times that of continuous process ($53 million). The real plant cost can exceed the cost calculated due to the levels of details considered in the process, for example, buffer preparation tanks and transfer panels that are not built in the model. The operating cost of fed-batch and continuous processes are $61 million/yr and $32 million/yr, respectively. The overall cost of fed-batch mode is $99/g and that of the continuous mode is $61/g. Under the assumption of 15 years' project with 3 years' facilities construction and with 7% interest, the net present value difference between fed-batch and continuous process is $720 million.
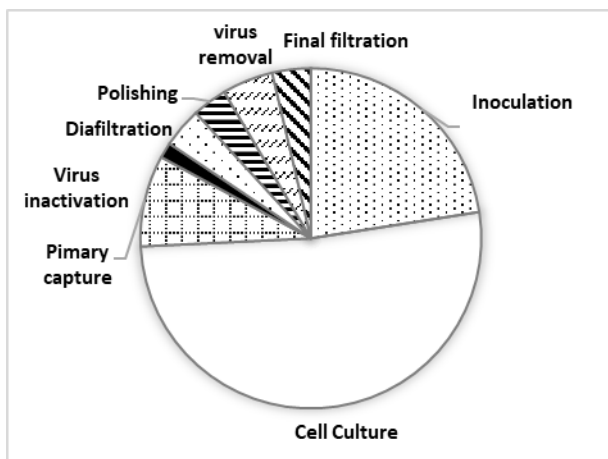
Figure 3.4 6 Cost of each unit operation in fed-batch process
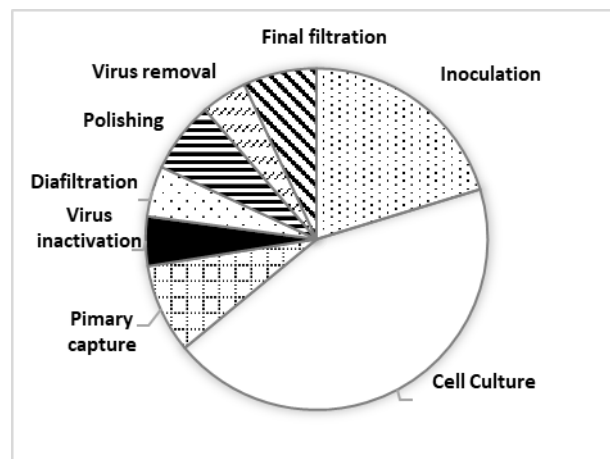
Figure 3.4 7 Cost of each unit operation in continuous process

Figure 3.4 6 and Figure 3.4 7 show the cost of different unit operations in fed-batch and continuous processing, respectively. The upstream cell culture unit operation contributes as high as 51% of the overall batch process and inoculation contributes 22%, which are mainly due to the high capital cost of the two unit operations. In order to reduce the time between consecutive batches and achieve 620 kg/yr annual production rate, 2x2 inoculation bioreactors (2 different bioreactors and 2 pieces each) are used in inoculation operation and 4 production bioreactors are used in cell culture. Labors are also scheduled to each of the bioreactor. With multiple inoculation bioreactors and 15000L bioreactors, the facility dependent and labor cost result in 74% cost for the inoculation operation and 63% for the cell culture. Comparing to fed-batch process, continuous process reduces the number of bioreactors to two inoculation bioreactors and 1 perfusion production bioreactor. In addition, the size of production bioreactor also decreases from 15000L to 2800L. Thus, the facility dependent cost and labor cost in continuous process significantly decreases in comparison to fed-batch. As a result, the upstream cost in continuous process is reduced to one half of original cost. However, upstream unit operations still occupy the highest portion of overall cost. Material costs are another major cost in upstream unit operation in both processes which take 31% of the cell culture cost in fed-batch process and 75% of that in continuous. This is because of low volumetric productivity (0.107 g/L/day) and low culture medium conversion rate in fed-batch bioreactor. In

continuous process, the volumetric productivity increases to 1.68 g/L/day. However, due to the usage of same culture medium, it assumes that there is no significant increasing of medium conversion rate in perfusion process. From the simulation result, the material cost in perfusion process does not account to significant savings to the overall cost.

Continuous processing not only has benefit in upstream operations but also in downstream operations. Figure 3.4 8 compares the COG/g of different unit operations in fed-batch and continuous processing. Continuous process reduces the primary capture by two folds, which accounts for the highest portion of the downstream cost in both operating modes. In the primary capture step, the size of protein A column reduces from 286L to 4 columns of 3.14L which decreases the capital cost and facility dependent cost. The PCC system has internal connection between the protein A columns, which improves the protein yield and resin utilization during this step. Thus, material cost from equilibrium, washing, elution and regeneration buffers is also reduced. In addition, the PCC step uses more expensive resin but provides higher resin binding capacity and life cycles, which compensates the high unit price. Above all, most savings are achieved from the cost reduction of protein A resin which is only 28.4% of that in fed-batch process.
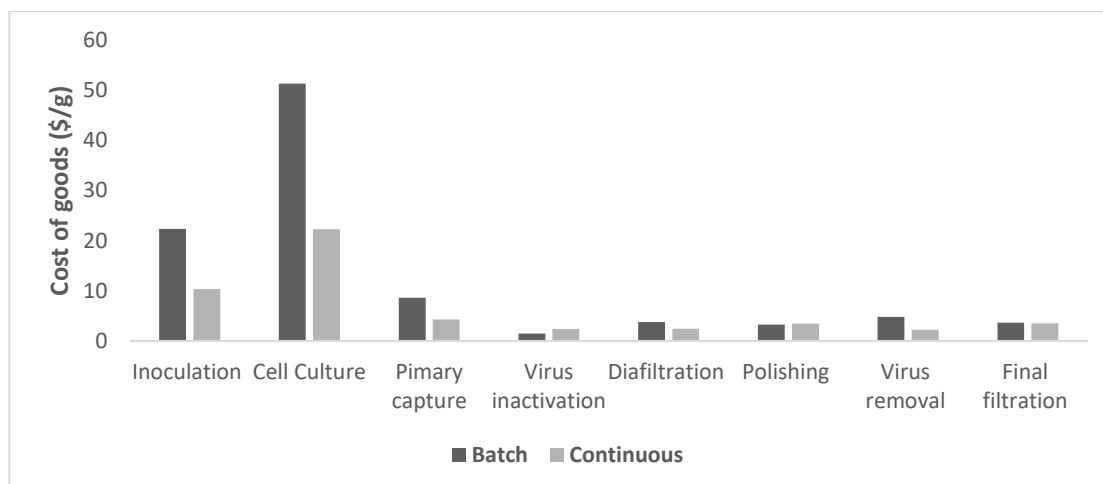
Figure 3.4 8 COG of fed-batch and continuous upstream downstream unit operation

However, not all the unit operations in continuous process contribute to cost savings over fed-batch process. Virus inactivation and polishing operations have higher COG/g than batch process as shown in Figure 3.4 8. In virus inactivation process, two tanks work alternatively in a way that as one tank is under cleaning, the other one is working under low pH virus inactivation. Similarly, two AEX columns are used in polishing step. In these two steps, there is no significant improvement of process operations. Although the size of equipment decreases in both unit operations, labor cost increases due to requirement of continuous operations. There are also no significant savings of final filtration step in continuous process. The final filtration contains not only the ultrafiltration and diafitration steps but also sterilize filter and final packaging, so that additional labors are assigned in this process. Thus, labor cost of filtration in continuous process is higher than that of batch. However, material cost in this operation is reduced which compensates the labor cost, thus, no significant savings can be observed. Figure 3.4 9 shows the cost comparison between fed-batch and continuous processes for all different cost categories. The main benefits of continuous process are due to capital cost investment and facility dependent cost, which are caused by the smaller footprint of all the unit operations through the process. In addition, large storage tanks are not used in continuous processing [71] that results on lower operating and capital cost. Consumables cost is also reduced by the resin cost savings in primary capture operation as

explained in the previous paragraph. Even though continuous process shows savings on material and labor cost, the savings are not as apparent as that of facility dependent cost and consumables cost. In fed-batch process, each equipment is assigned one or two units of labors and is assumed that labor is needed only at the time of operation. In the continuous process, considering the automations used in the system, one or two labors are assigned to each unit operation that contains many equipment. However, the continuous process is running continuously, which means multiple labor units have to be occupied from start to the end of the process with shifts. Material cost is mainly due to upstream cell culture process. Perfusion method provides a capability of upstream operation with wilder utilization of cell lines and continuously adding culture medium. Thus, improving the conversion or utilization of medium by using different cell lines or medium will reduce the material cost per gram of product[109, 120, 121].
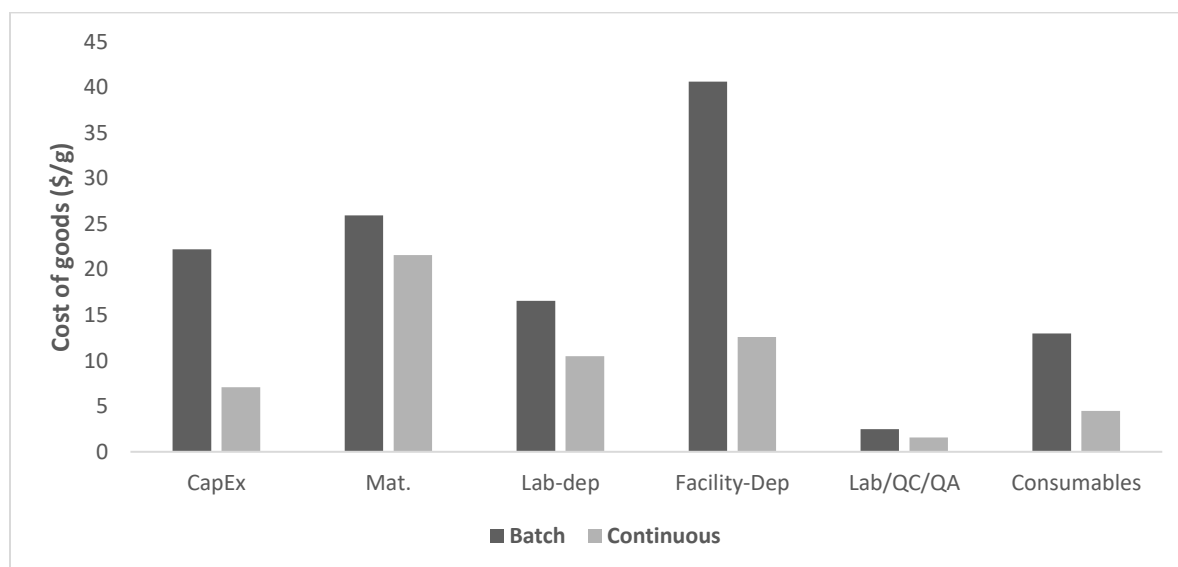

Figure 3.4 9 Cost of goods analysis between fed-batch and continuous

3.4.3 Sensitivity analysis

The parameter fluctuations and variations are very common in biopharmaceutical manufacturing. For example, upstream titers are varied by different cell lines and media used. Consumptions, such as resin and membranes, their life cycles, membrane capabilities, yields and costs are changed with

market perturbation and types of product produced. Thus, uncertainty and sensitivity analysis can be used to investigate the impact of variable parameters inputs on the final result [58]. In this section the analysis is focused on the uncertainty and sensitivity analysis of plant capacity, upstream titer and downstream yield on the overall economic analysis.

3.4.3.1 Throughput analysis

The throughput analysis is used to evaluate the plant capacity effect on production cost in fed-batch and continuous processes as shown in Figure 3.4 10, and the capacity ranges from 50 kg/yr to 1200 kg/yr. It shows that as the plant capacity increases, the unit operating cost ($/g) of product decreases both in fed-batch and continuous processes. In both operating modes, the labor dependent and QC/QA cost decrease significantly as plant capacity increases. This is mainly due to labor dependent cost, which is scale independent. With annual production rate increasing, the averaged labor dependent cost per unit of protein production decreases. For example, the labor dependent cost in continuous process ($6.51 million) remains constant in all plant capacities. However, as the throughput increases from 50 to 1200 kg/yr, the unit labor dependent cost decreases from $130.14/g to $5.50/g. Furthermore, facility dependent cost which depends on labor cost and capital cost decreases. It also shows that, consumables, QC/QA costs decreases. There is material cost savings which is not obvious, and the main reason is that the material cost is dominated by cell culture medium which is independent to the plant capacity.

Comparing to the fed-batch process, the benefit of the continuous process becomes less significant as plant capacity increases. As explained in the previous paragraph, the cost savings are mainly from facility dependent cost and labor cost. The operating costs of different unit operations that are dominated by these two types of costs will be reduced the most. From Figure 3.4 11, upstream cost savings are more significant in fed-batch process and downstream cost savings are more significant in continuous process. However, among the total cost, upstream cost takes up a higher portion than

downstream. Thus, the cost saving from fed-batch process is more significant as the plant capacity
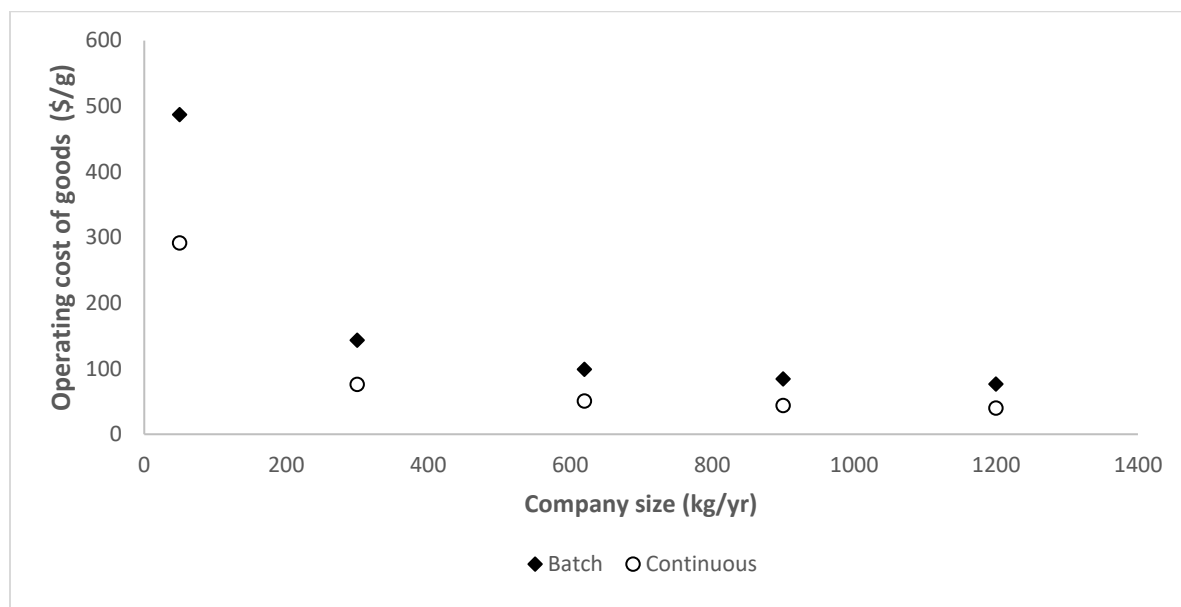
increases.

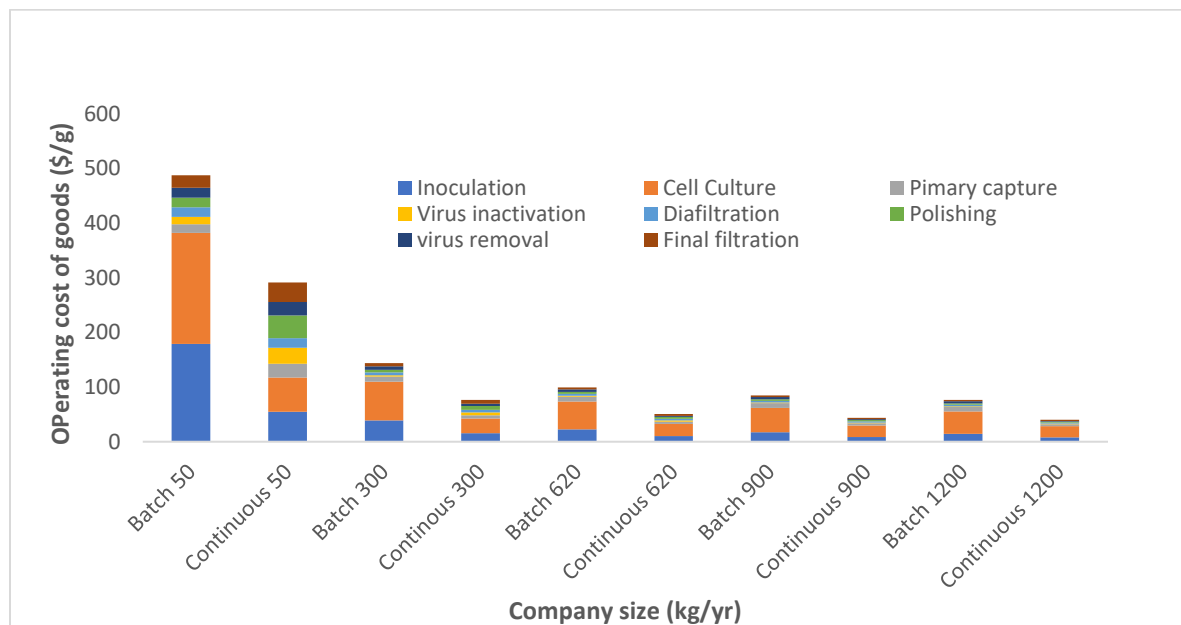Figure 3.4 10 Manufacturing scales change vs. COG/g change

Figure 3.4 11 A comparison between fed-batch and continuous cost on unit operations among the different plant capacity

3.4.3.2   *Upstream titer analysis*

Titers represent the concentration of protein product harvested from production bioreactor. Development of upstream process, improvement of seed bioreactor, and the choices of cultural media and cell lines increase the titers in bioreactors which will improve the upstream productivity, and this will affect downstream operating cycles, times and consumables used. Due to the number of interconnections and scheduling among the equipment in fed-batch and continuous system, increasing the upstream titer will cause process rescheduling and system conflicts. Thus, multiple scenarios with different titers are designed for the upstream titer analysis.

In Figure 3.4 12, the simulation results for fed-batch and continuous processing are compared under different upstream titers in different scenarios: 1.5 g/L (scenario 1), 2.5 g/L (scenario 2) and 5.5 g/L (scenario 3). It shows that, with increasing titer, the overall operating cost per unit product ($/g) decreases in both operating modes. With titer ranging from 1.5 g/L to 5.5 g/L, the operating cost ($/g) of fed-batch process is reduced by 58%, and the continuous process is reduced by 50%. This is due to the increasing annual production rates of the two operation modes, shown in Table 3.4 4. Under the same upstream operating conditions, the increase of upstream productivity shifts the bottleneck of the process from upstream to downstream. Since facility sizing remains constant through all scenarios, to purify more protein, more cycles need to be used in major purification operations. From scenario 1 to scenario 3, the fed-batch upstream productivity increases from 17.1 kg/batch to 62.7 kg/batch, and the number of protein A column cycle increase from 2 cycles to 7 cycles and the unit operating time increases from 0.62 days to 1.83 days. To prevent product overflow, hold vessels that are used between unit operations, and virus inactivation tanks need to be operated in cycles. Thus, the overall annual batch numbers decrease from 48 cycles to 45 cycles. Similarly, in continuous process, with titer increasing, primary capture step is also the bottleneck. Since upstream production rate increases from 109 kg/batch to 398 kg/batch, the PCC cycle numbers increase from 127 cycles to 468 cycles, which causes the operating time to increase from

24 days to 73 days. Thus, the overall working cycles reduce from 9 batches to 3 batches. For this case, continuous processing has less annual protein production rate, so cost savings are not as significant as for the fed-batch process.

Table 3.4 4 Annual production rate of fed-batch and continuous process with different upstream titers

Upstream titer	Fed-batch production rate (kg/yr)	Continuous production rate (kg/L)
1.5 g/L	618.16	620.00
2.5 g/L	987.00	776.98
5.5 g/L	1227.00	856.58

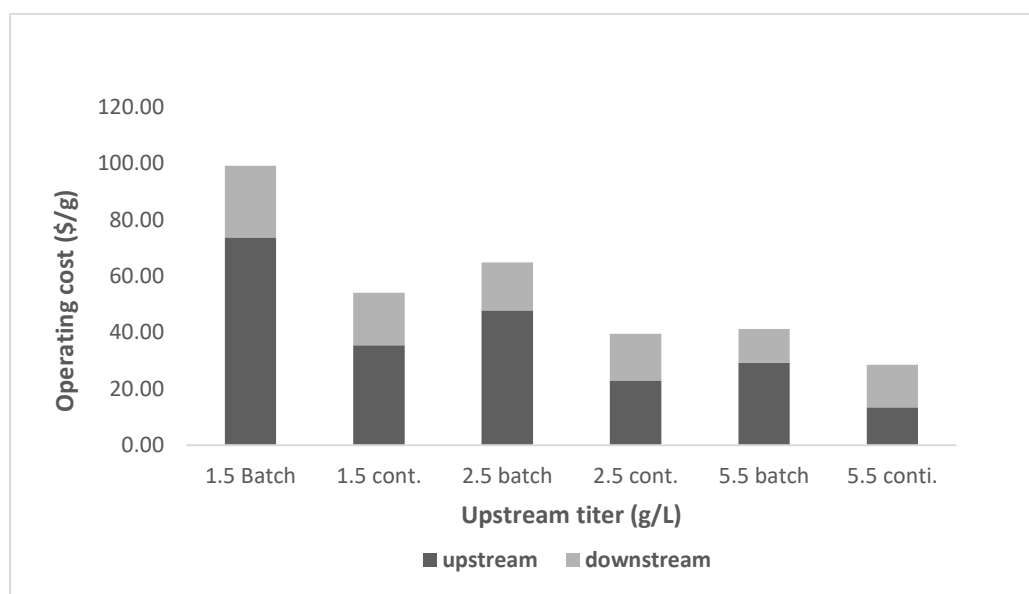Figure 3.4 12 A comparison between fed-batch and continuous operating cost with different upstream titer

Even though cost savings in fed-batch process becomes more significant as titer increases, the operating cost per unit of fed-batch product is still higher than that of continuous. This is mainly due to facility-dependent cost and consumables cost of continuous operations being 3 times lesser

than that of fed-batch. It is also interesting to show that as titer increases, the cost ratio between upstream and downstream increases in fed-batch operations but decreases in continuous operations. It indicates that, continuous downstream unit operations are more sensitive to the titer in upstream and as titer increases, the process bottleneck will shift from upstream to downstream. To improve the productivity and flexibility of continuous processing, multiple downstream facilities should run in parallel or larger facilities need to be used. Since continuous primary capture uses 4 columns of 3.14 L, either adding another set of PCC operation or increasing the column sizes is cost effective. Thus, the operating cost of continuous process can be further reduced. Hammerschmidt et al. [62] considered fed-batch plus the chromatography system and perfusion plus the precipitation system and compared the cost of goods for different upstream titer. The results showed that for the fed-batch system, as titer increased, the ratio between upstream and downstream decreased which means the cost shifts from upstream to downstream. However, this result is opposite to what we obtained. It needs to be highlighted that, the facility size in Hammerschmidt et al. analysis is varied with titer while it is kept constant in this paper. By varying facility sizes, increasing the upstream titer also shifts operating cost from upstream to downstream in fed-batch and continuous processes.

3.4.3.3   Downstream yield

Different operating scenarios are designed to analyze downstream yield effect on operating cost in both fed-batch and continuous processes. Downstream yield in base case scenario is 75% and downstream yield variations range from 70% to 80% while the facility sizes are kept constant. In fed-batch and continuous processes, different downstream unit operations have different yield. In order to decrease the overall yield, among all the downstream unit operations, highest operation yields are decreased. Similarly, in order to increase the yield, lowest operation yields are adjusted and increased.

As downstream yield increases, the overall operating cost decreases, shown in **Figure S3.** When yield increases, the annual production rate increases. Thus, the utilization of equipment increases

and the unit operating cost increases. However, when the operating yield reaches 80%, the size of holding tank in fed-batch process is not large enough to maintain the operating volume. Here, the number of operating cycles is increased and the total operating time per batch also increases. Thus, the cost savings in operating cost from downstream yield 75% to 80% is not as much as that from 70% to 75%. In continuous process, cost savings can be observed with an increase in downstream yield. However, there are no obvious changes or cost shifts between upstream and downstream. The cost savings of continuous process decreases is not compared to fed-batch process, which shows that the continuous process is not as sensitive as fed-batch.

3.4.4   Conclusion and future work

This case study highlights efforts in biopharmaceutical manufacturing towards process intensification highlighting the advantages of alternative process units and continuous operating mode. It provides a comprehensive comparison of traditional fed-batch process as it compares with intensified continuous operation in biopharmaceutical manufacturing by using economic metrics and sensitivity analysis. From base case scenario, with 620 kg/yr annual production rate, continuous process, with small footprint and high automations, reduces capital cost, facility dependent cost and labor cost significantly compared to fed-batch process. Consumables and material costs of continuous process are also reduced, mainly because the resin utilization in continuous process is increased, and with small footprint, the buffer usage for downstream cleaning and primary capture is also reduced. From sensitivity analysis, scenarios are designed to compare fed-batch and continuous processes with different operating conditions, including, various plant capacities, upstream titers and downstream yield. Results show that, as plant capacity increases (from 50 kg/yr to 1200 kg/yr), operating cost in both fed-batch and continuous processes decreases and cost in continuous process is lower than that of fed-batch process in all ranges. Similarly, as titer increases or downstream yield increases, the operating costs also decreases. Increasing upstream titers shifts the process bottleneck from upstream to downstream, especially primary capture step, which causes

longer downstream operating time. In continuous process, adding another downstream line or increasing size of primary capture column can resolve bottleneck cost-effectively.

Several assumptions are made in this model to simplify the process and calculations. In upstream process, same inoculation operation is used in both fed-batch and continuous processing. In recent years, researchers have been working on inoculation operation by using N-1 perfusion inoculation bioreactor that has high-density seeding, followed by fed-batch bioreactor. The method may increase the inoculation cell density, reduce the production bioreactor fermentation time and improve the production capacity and product quality [106, 122]. In addition, most of the cleaning recipes in continuous process are modified and similar actions are done in the fed-batch process. Total operating times per batch in continuous process and cycle numbers are calculated based on unit operation running time but not optimized. Failure rate and contaminations are not considered in the production bioreactor section in this work. Since perfusion bioreactor contains external filters and cells are recycled back to the fermenter, the contamination chance in perfusion process should be higher than that of fed-batch process. Filter clogging, and fouling can also cause process failure. However, when process failure happens, the whole batch of product needs to be disposed in the fed-batch processing, however in continuous process, only the product affected by the failure should be discarded.

Instead of economic benefits, continuous biopharmaceutical processing shows advantages in the area of production system standardization, process flexibility, scaling, product quality, and it is available to use single-use technology. The development of continuous process is also under a good regulatory condition. In 2011, US Food and Drug Administration's strategic plans[123] pointed out that the continuous manufacture of biologicals is encouraged[124] and the product quality can be improved by using Quality by Design (QbD). FDA has also confirmed that the continuous process offers advantages in development and manufacturing[125]. Continuous biopharmaceutical processing allows rapid capacity adjustment by adjusting the number of parallel production lines[8].

The duration of production can also be changed based on the requirement of certain product. Since the continuous processing reduces the total time that product stays inside of production line, especially in the bioreactor, both stable and non-stale protein product can also be produced in multiproduct continuous facilities so that implement process standardization. More important, the decreased residence time in bioreactor and elimination of intermediate hold steps improve the product quality by reducing the risk of product degradation and post secretion enzymatic modification. With the uniformed microenvironment control, the continuous processing also minimizes the product variability[126]. In addition, scales in continuous processing are equivalence for pilot, clinical and commercial production, which increases the speeding of scale up and minimizes the technology transfer. However, challenges are also come along with those benefits. The integrated fully continuous processing requires high degree of automation with minimal operator involvement. While different vendors share different control system which increases integration difficulty[8].

For future work, the continuous process needs to be optimized, with respect to the batch time, number of operating cycles, process lines and facility sizes, with consideration of process failure rate. Single-use methods also need to be involved in the analysis. In single-use system, cleaning such as CIP and SIP do not need to be used which reduces buffer and energy consumption. However, the single-use equipment needs to be replaced and disposed. In addition, it also has the limitation of facility sizes, for example, production bioreactor is limited under 2000 L. Thus, including all the above considerations, a more realistic and cost-effective integrated continuous process model can be built.

3.5   Case study 2 rAAV production

It has been reported that, till the year 2019, there were 22 gene therapy products approved for human diseases worldwide [127]. From year the 2018 to 2021, US FDA approved 19 gene/cell therapy products [128]. The global gene therapy market is estimated to grow from UDS 3.8 billion

in 2019 to USD 12.0 billion by 2024 [129]. As an emerging biological product, gene therapy has capability to make site-specific modifications of human genomes and can be used in the treatment of monogenic disease, cancer, cardiovascular diseases, infectious, degenerative arthritis, ocular diseases [127]. The idea of gene therapy is to change original human genes to target genes which includes gene insertions, deletions, target modifications, amplifications and translocations. However, in order to successfully deliver the genes to target cells, finding a molecular carrier that can efficiently release genes without immunes response is critical. Viral vectors are the most important microcarrier with approved efficacy [3]. Recombinant adeno-associated virus (rAAV) is one of the most widely used viral vectors and three approved drugs Glybera (Alipogene Tiparvovec), Luxturna (Voretigene neparvovec-rzyl) and Zolgensma (Onasemnogene Abeparvovec-xior) are produced based on rAAV [127]. From the 2017 report, among 483 gene therapy vectors in development, 103 are AAV based [130]. Although many clinical trials using gene therapy process have been performed, building a large-scale manufacturing process is still a major challenge at the current stage[131].

In this section, flowsheet modeling will be used to design an rAAV manufacturing process for both batch and continuous operation and show the modeling capabilities in cutting-edge product design. Economic analysis is used to obtain capital and operating costs for each process and scenario study is going to apply to evaluate plant capabilities under different scales.

3.5.1 Process description

Like mAbs production, rAAV production also includes upstream and downstream, where upstream mainly focus on rAAV production and downstream is used for impurities removal. For upstream, there are three ways to replicate and pack the rAAV with the target genes 1) mammalian cell with helper virus, 2) mammalian cells with helper free transfection 3) insect cell with baculovirus, as shown in Figure 3.5 1. In general, rAAV vectors plasmid that is containing transgene needs to be co-transfected with regulatory (rep) and structural capsid (cap) genes to the packaging cell line

(mammalian cell). During this step, helper virus will be added to maintain replication of the transgenes. However, this method usually brings contaminations by helper virus which will further complicate the downstream purification. Another approach is to encode the transgenes, rep and cap to baculovirus and transfer them to insect cells such as Sf9 cell line. Although the baculovirus method can potentially scale up to a large manufacturing process with high productivity, it still has a long clinical time and brings baculovirus contamination. Compared to helper virus and baculovirus, helper free transfection (triple transfection) is promising to be used because it has capabilities to achieve fast transfection, produce different rAAV stereotypes while does not add extra contaminations to downstream. For this process, three plasmids that contain a transgene, rap,cap gene and adenovirus helper gene respectively with ratio 1:1.5:2 are used for co-transfection [131].

rAAV is an intracellular product thus needs to be lysed after the cell culture. One option is to harvest the cells first and then lyse them. Another approach is to lyse the cells directly. During the cell culture, part of the rAAV is secreted to the culture media. Thus, even though cells are lysed after the harvest, it still has to be combined with the culture media before the further purification. In our case study, the cells with media are directly removed from the bioreactor to the next unit operation for cell lysis and product clarification. After the cell lysis, benzenes are added to remove host DNA and RNA. Three filtration steps are then used to remove cell debris, host cell proteins (HCP), large impurities, and further concentrate the product. To prepare for the chromatography operation, product buffer is exchanged using diafiltration. The rest impurities are majorly removed by two chromatography, affinity chromatography and anion exchange chromatography. Finally, ultrafiltration and diafiltration are used for the final formulation. It needs to mention that, ultracentrifugation has been widely used in the rAAV gene therapy development in laboratory. By consulting with our industrial collaborator, this method is hard to be used in large-scale manufacturing processes. Thus, ultracentrifugation is avoided to be used in this design.
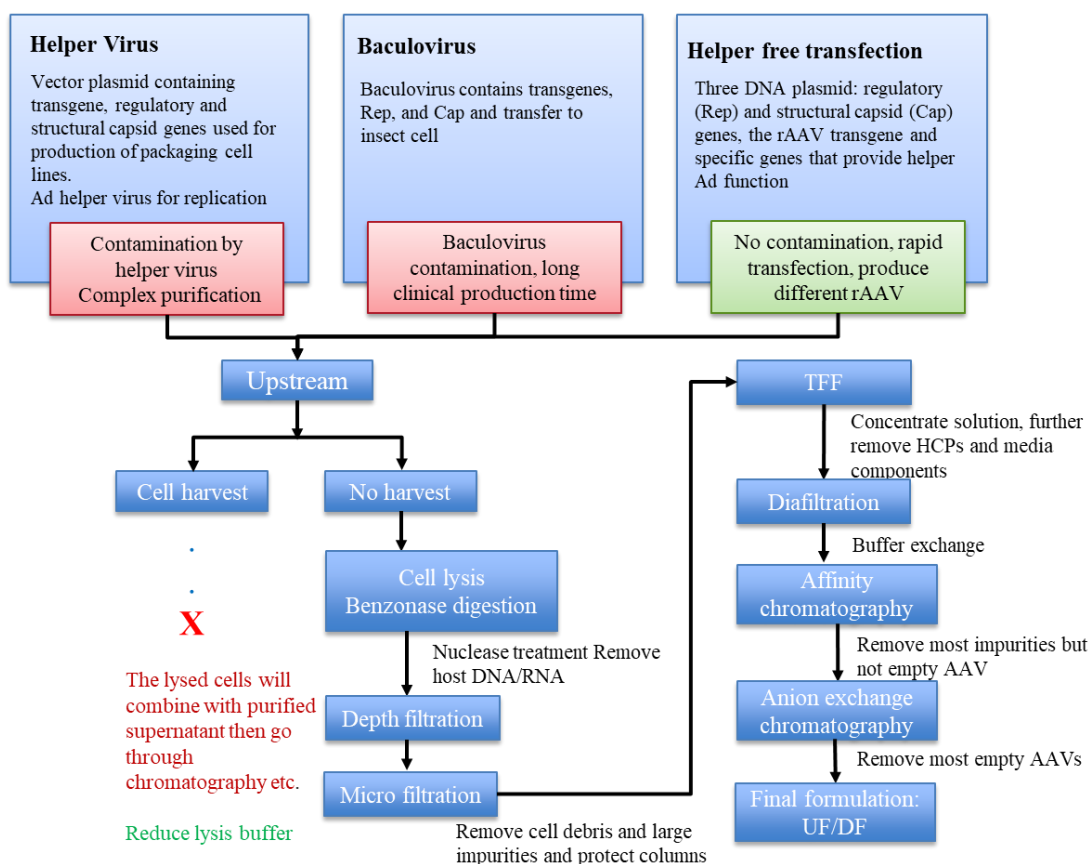
Figure 3.5 1 Selection of rAAV production lines

3.5.2 Batch process design and optimization

Based on the description in the section 3.5.1, the flowsheet model is designed on SuperPro Designer as shown in Figure 3.5 2. SuperPro Designer can be used for process design, scheduling and economic analysis which has been introduced in Chapter 2. Table 3.5-1 summarized detailed specifications for major unit operations. The base case scenario of the batch process has three 250L single-use batch bioreactors with working volume 200L. There is one downstream line with sum yield 43%. The size of each unit operation is commercially available. In this scenario, the final formulation of the rAAV product is 5.8×10^{15} vg/L and the annual production rate is 742 L/year. The total batch number is 138 batch/year.

Figure 3.5 2 Single-use batch process for rAAV production

Table 3.5 1 Critical parameters for batch process design

Bioreactor Parameters			Affinity Capture		
Viral titer	3.61×10^{14}	vg/L	Load ratio	10^{16}	vg/L
Empty partilces	9.58×10^{14}	particle/L	Column Dia	14	cm
Bioreactor scale	250	L	Column Length	15	cm

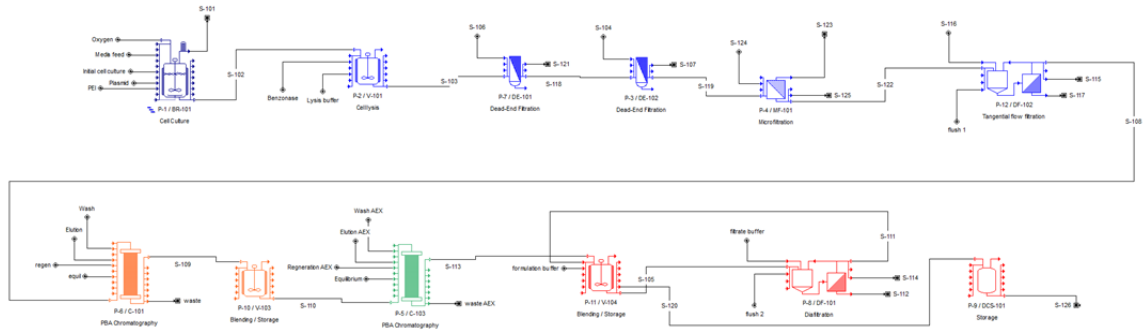| Cell Lysis			Polishing column		
Tween Twinty	0.005	w/v	Load ratio	7.8×10^{15}	vg/L
			Column Dia	30	cm
			Column Length	25	cm

Filtration			UF/DF		
Depth filter 1 flux	195.7	LMH	UF Permeate Flux	50	LMH
Filter Area applied	1.1	m2	DF Permeate Flux	20	LMH
Depth filter 2 flux	200	LMH			
Filter Area applied	1.1	m2			
Micro filter flux	200	LMH			
Filter Area applied	0.55	m2			

Depth filter (x2)	Micro filter	UF/DF	Affinity Capture	Polishing column	UF/DF
0.95x0.95	0.9	0.9	0.95	0.7	0.95

To understand the cost-effectiveness of the integrated system, economic analysis is applied to the simulation. For this designed plant, the total capital cost is 8.09 million USD with equipment cost 0.621 million USD while the operating cost is 48.62 million USD. This number consistent with cost estimation of adenovirus which is another vector that has been used in gene therapy [132]. In

the adenovirus case study, the total cost of operating cost with capital cost is around 0.3 million USD/batch for single-use process with a 200L culture scale. In our case study, operating and capital costs in total is 0.36 million USD/batch with the same production scale. In terms of the operating cost composition, 88% of the cost is from material cost. Benzonase cost takes the highest present which is around 51% of the material cost. Labor cost is 7% and consumables cost is 3% from Figure 3.5 3. As for consumables cost, single-used culture bags and mixing bags take 62% of the cost.
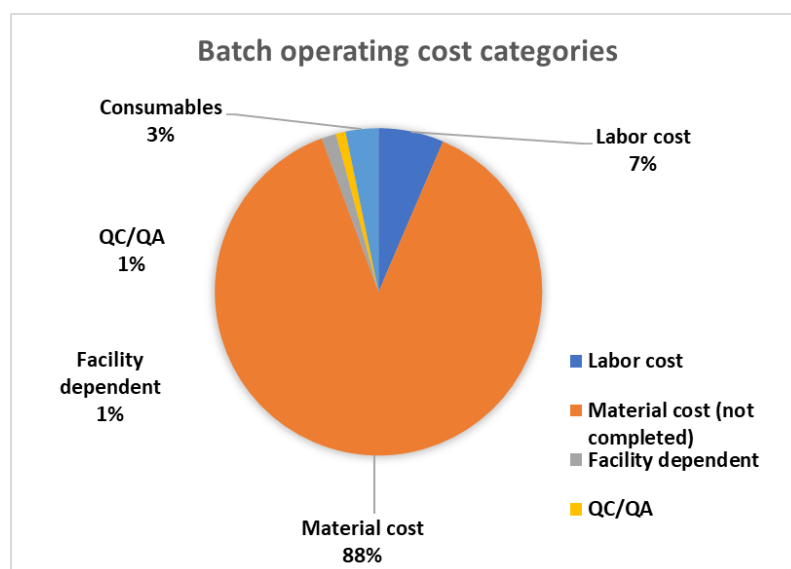
Figure 3.5 3 Batch operating cost categories breakdown for base case scenario

From the simulation, the equipment occupancy chart can be generated, as shown in Figure 3.5 4. The figure indicates the bottleneck for this process majorly from upstream production bioreactor because downstream unit operations are not sufficiently used. Then here comes to the next questions about how to optimize plant to ensure the maximum productivity while maintaining the cost efficiency.
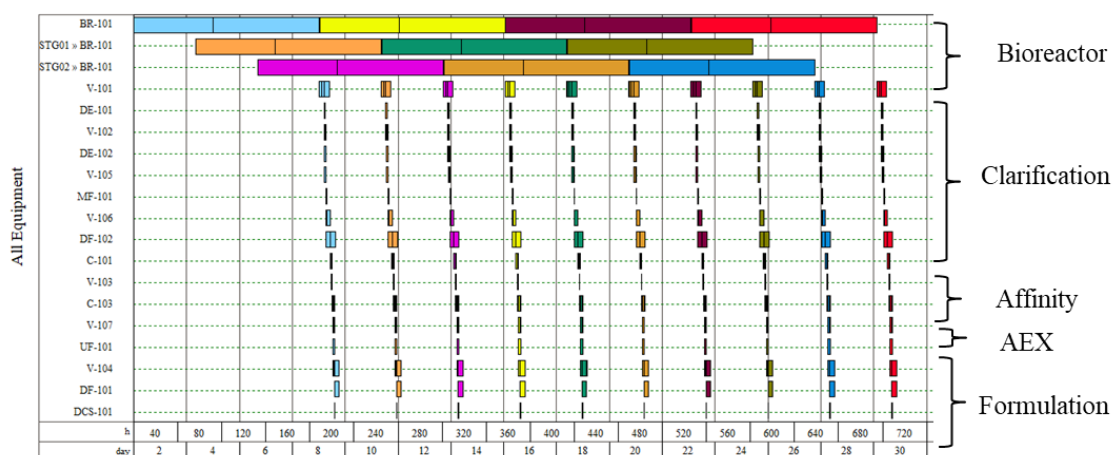
Figure 3.5 4 Equipment occupancy chart of batch rAAV production for base case scenario

Since upstream bioreactor operation is the bottleneck of this process, the staggered bioreactors are added. In this case, with multiple bioreactors working at the same time, downstream lagging time is reduced, and the overall productivity is increased. To investigate the cost change as the number of upstream staggered bioreactor increases, capital cost and operating cost for unit volume of the production are calculated. Note here, the upstream titer and final formulation are not changed. From Figure 3.5 5, as the upstream bioreactor number increases, the unit operating, and capital cost first reduced and then increased. It represents that as the overall productivity increases, the unit cost is reduced. Once the system reaches maximum plant capacity, the overall productivity will stop increasing. Further, increase the upstream bioreactor number will increase the labor cost, facility cost and reduce the bioreactor operating efficiency, thus increases the overall costs.
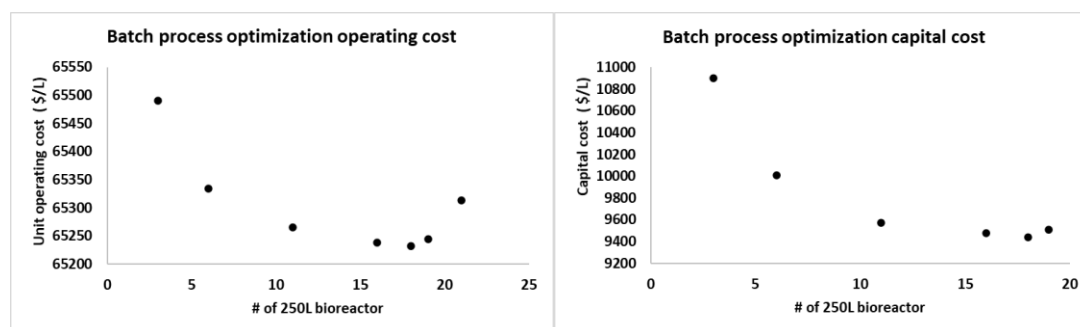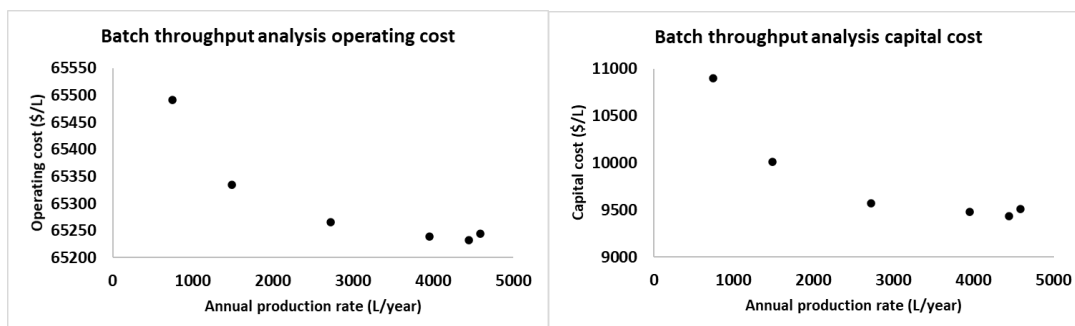
Figure 3.5 5 Batch process optimization with upstream bioreactor numbers

As a result, the most cost-effective process happens when there are seventeen 250 L bioreactors and one downstream line. Figure 3.5 6 shows the updated equipment occupancy chart with the optimum operating capacity. It also finds that using eighteen 250 L bioreactors can reach maximum productivity. Finally, the total plant capacity is 4588 L/year with formulation 5.8×10^{15} vg/L.
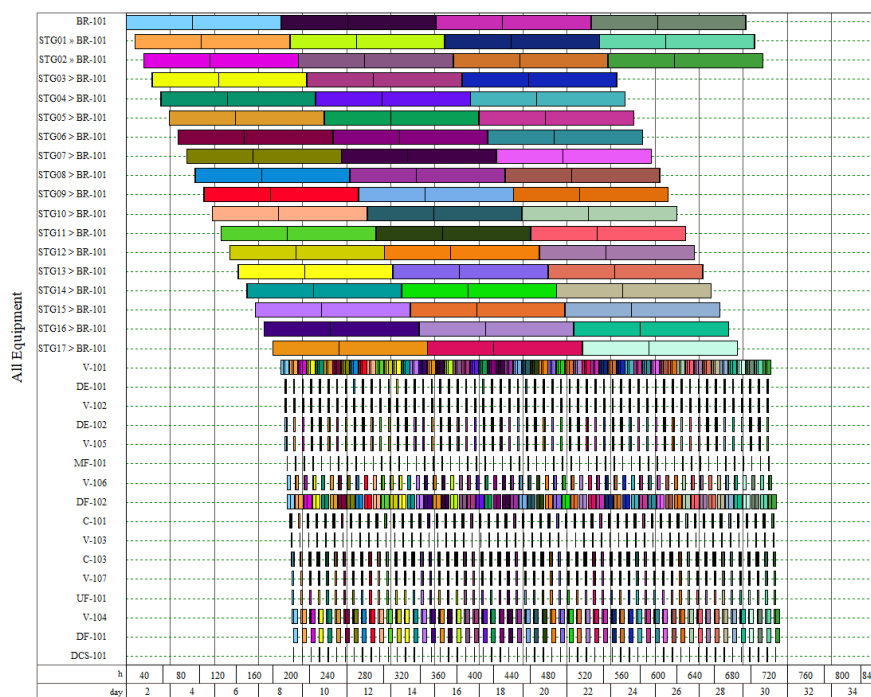
Figure 3.5 6 Equipment occupancy chart with optimized batch process

3.5.3 Cost comparison between batch and continuous operation

To further improve the process cost efficiency of the rAAV production, a continuous process that can reach the maximum operating throughput (4588 L/year) is designed. Currently, both continuous

upstream bioreactor and downstream chromatography operations have been used in clinical research, for example, wave bioreactor, stir bioreactor with acoustic filter, and continuous chromatography operations [131, 133, 134]. However, there is no manufacturing scale continuous gene therapy process designed in the literature. Thus, based on the experience from continuous mAbs manufacturing operations, we designed a flowsheet for continuous rAAV production.

The continuous plant has the same unit operations as that of the batch. Since the continuous process for gene therapy has not been well developed, the following assumptions are made: 1) most of continuous downstream operating parameters especially for filters and chromatography are developed based on continuous monoclonal antibody production. 2) The process yield is assumed the same as that of batch 3) chromatography resin capacities and cycles are assumed the same as that of batch. 4) all the equipment is assumed single-used and the bags will be replaced every batch. Figure 3.5 7 shows the flowsheet simulated platform for continuous rAAV production. Specifically, a perfusion bioreactor is operated continuously for four days in perfusion mode. The product titer is assumed the same as that of the batch process. Following the bioreactor, two continuous tubular blending tanks are connected in series for cell lysis and benzonase digestion. Then three continuous filtration, single-pass tangential flow filtration and diafiltration are used for clarification, concentration, and buffer exchange. A 4-column periodic counter-current (PCC) continuous affinity chromatography followed by 4-column PCC anion exchange chromatography is applied. Finally, counter-current ultrafiltration and diafiltration are designed for the final formulation. As a result, four 250 L plus two 100 L single-use bioreactors and one continuous downstream line are used in the design.
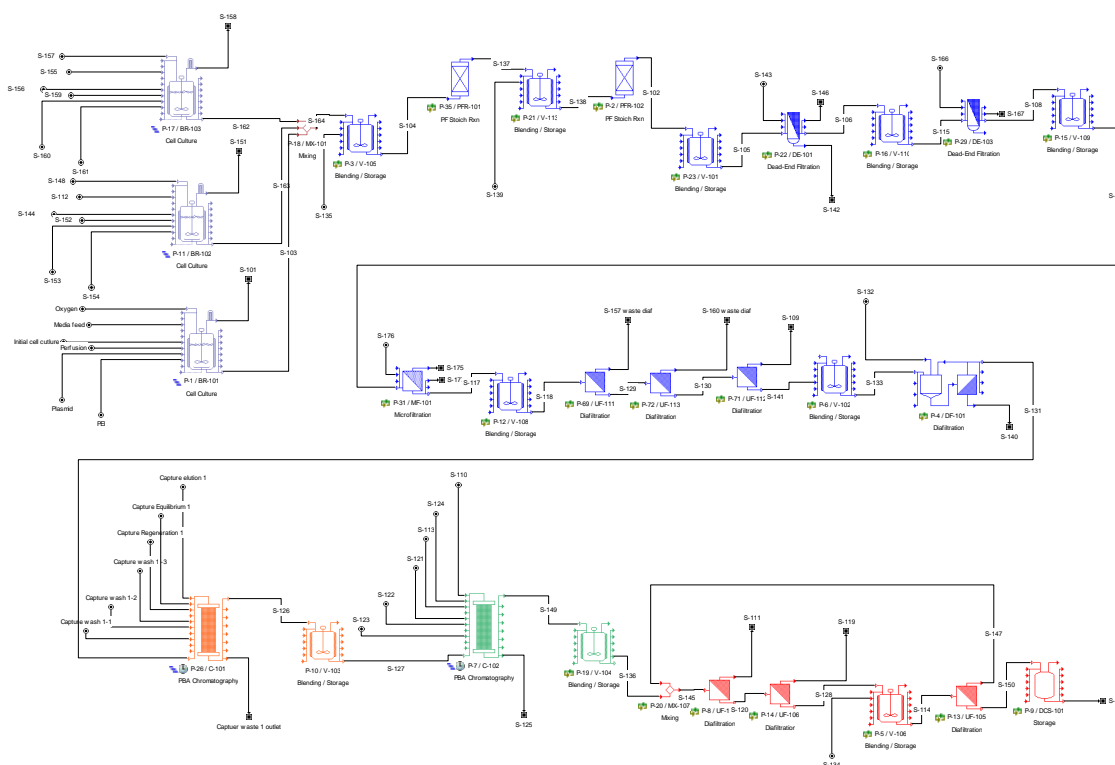
Figure 3.5 7 Single-use continuous flowsheet model for rAAV production

To compare batch and continuous process, economic analysis is applied to both batch and continuous operations on their maximum scale, that 4927 L/year for continuous and 4588 L/year for batch operation. The total capital cost for continuous operation is 0.04 million USD/L and 0.3 million USD/L for that of batch. The total equipment cost of batch process is 16.441 million USD which is around 20 times higher than that of continuous. The operating costs for continuous and batch processes are 0.044 million USD/L and 0.080 million USD/L, respectively. Figure 3.5 8 shows operating cost categories for continuous operations. Material cost takes the highest percent of the overall cost of continuous operation, while consumables cost is the second highest cost. However, by comparing the cost from different unit operations, it is interesting to see that continuous operating saves the cost mostly from bioreactor and clarification steps. For continuous affinity chromatography, most of the costs are from resin cost and labor cost. It shows that continuous operation does not reduce the size of the column significantly – 4x1.48L for continuous

and 2.31L for batch. Then more resins are used during the continuous operation. For polishing and final formulation, the labor cost of the continuous process is much higher than that of batch.

Figure 3.5 8 Operating cost for batch vs continuous production.

3.5.4 Analysis of fed-batch and continuous under different scales

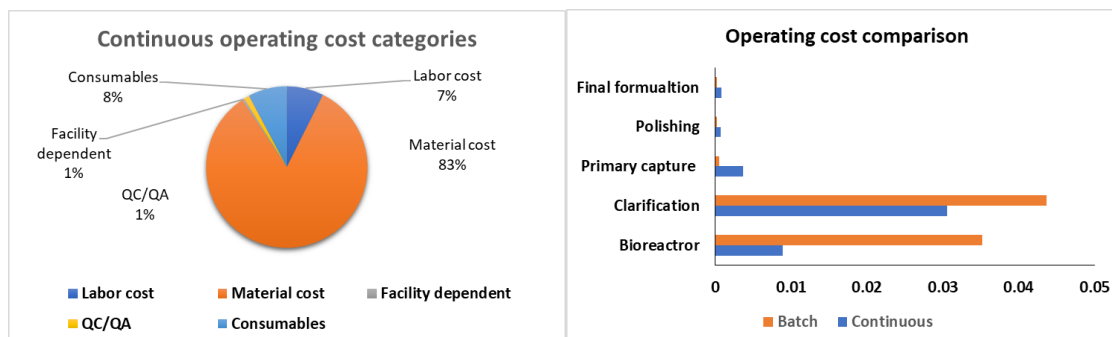Throughput analysis is applied to investigate the cost variations of batch and continuous operations under different scales, as shown in Figure 3.5-9. In this comparison, downstream equipment sizes and numbers are kept constant. The size of the upstream bioreactor kept the same, but the number of bioreactors used is reduced or increased aiming to adjust the annual throughput. The annual throughput is changed from 700 L/year to 4000 L/year. Remind from Figure 3.5 9, increasing the throughput of the batch process will reduce the operating cost. However, compared to continuous operation, this reduction is not obvious. The continuous operating cost shows a going down trend as the annual production rate is reduced but going down periodically. This happened when the number of upstream operating bioreactors changed. As it is known from the continuous operation, adding another piece of equipment increases the capital cost and labor cost. If the throughput is changed slightly, the average facility-dependent cost and labor cost will be increased significantly. This is the reason why there are variations in the continuous operating cost analysis. Even though continuous operations cannot work efficiently all the time along with the changes of the scale, the operating cost still shows benefits or similar to that of batch in large scale. The graph also indicates continuous operations are not always good than that of batch process. Continuous operating is good for large-scale operating, and batch process is good for small-scale operation for rAAV production.

With the development of continuous gene therapy techniques, the upstream and process efficiency of the downstream will improve. The cost-effectiveness of continuous operation will be further improved.
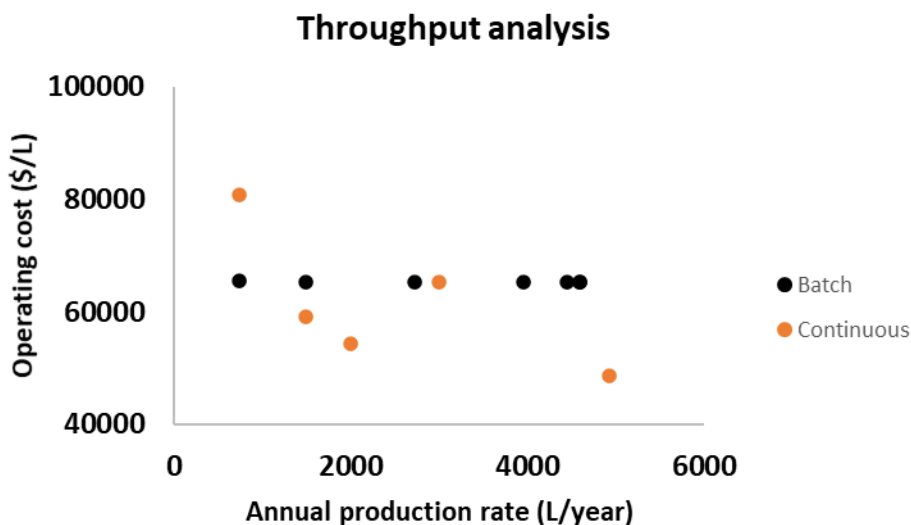
Figure 3.5 9 Throughput analysis batch vs continuous

3.5.5   Conclusion and Future work

In this case study, two large-scale rAAV production processes are designed for batch and continuous mode. Using economic analysis, the batch process is optimized by adjusting the number of upstream production bioreactors. The optimized batch process is compared with continuous operation from different cost categories and different production scales. As a result, continuous operation saves both capital cost and operating cost compared to that of batch on a large scale, and these advantages are reduced as the manufacturing scales are reduced. This work guides the design of large-scale gene therapy process manufacturing as more and more gene therapy drugs are approved in recent years. For the future work, different operating parameters such as titer, or downstream yield can be investigated for both batch and continuous operating modes.

4   A framework of fed-batch bioreactor modeling and design space identification

4.1   Background

During monoclonal antibody production from CHO cells, one of the most critical product quality attributes is relative to protein glycosylation. Glycosylation is a post-translational protein modification process where the oligosaccharide covalently binds to the protein backbone [16]. There are two types of glycosylation O-linked and N-linked glycosylation. They use different sugar residuals and link to different amino acid on protein where O-linked glycosylation binds to hydroxyl group of serine or threonine and N-linked glycosylation links to amide nitrogen of asparagine. [135] Usually, N-linked glycosylation widely exists in mAb production. The glycosylation process starts from endoplasmic reticulum that oligosaccharides first bind to protein peptide and then transfer to Golgi apparatus for further modification.

In Golgi apparatus, there are three major compartments: Cis-Golgi, Medial Golgi and Trans-Golgi During the glycosylation process, glycans are moved from Cis-Golgi to Trans-Golgi. Different enzymes/glycosyltransferases exist in different compartment to add or remove sugars from protein glycans. For example, in Cis-Golgi, the mannosidases (ManI and ManII) are used to remove mannose; in medial Golgi, N-acetylglucosaminyl transferase I, II, III (GnTI, GnTII, GnTIII) and fucosyltransferase FucT are able to add glucose or fucose; in trans Golgi, galactose and Sialic acid can be added to glycans by galactosyltransferase (GalT) and sialyltransferase (SiaT). The sugar sources from nucleotide sugar donor are synthesized in cytoplasm and nuclei. Thus, the glycosylation process is majorly affected by glycan concentration, nucleotide sugar donor concentration and glycosyltransferases activities. These three components are closely related to metabolic cycles and cell culture conditions.

In bioreactor culture system, glycosylation process can be affected by both chemical and physical stimuli for the same cell line [14, 136, 137]. These stimuli can change the cell metabolism such as the glycolytic pathway, purine/pyrimidine metabolism, TCA cycles, which influence the synthesis and functions of nucleotide donors, glycosyltransferases and glycosidase, and further affect the biocatalytic reactions of glycosylation [138]. The limitation of glucose, glutamine [139], galactose [140], manganese [37, 141], uridine, ManNAc, sialic acid [14] in culture media and feed additions result the reduction of protein productivity, glycosylation site occupancy, UDP-Gal, UDP-GlcNAc, Neu5Ac synthesis. It should be mentioned that different cell lines perform differently under the same conditions, leading to different glycan fractions compositions. pH can dramatically affect the cell growth and glycosylation process though changed ammonia level during the cell culture. It has been showed that pH and ammonia concentration cause the delocalization and kinetic rate change of glycosyltransferases including ManII, GnTI, GalT, and SiaT [14]. Usually the pH range tests from 6.2-8.5, and with different cell lines, the effect from pH and ammonia concentration will also be different. Therefore, the degree and direction of pH effect to N-linked glycosylation cannot be drawn [142]. Temperature has also been found affecting intracellular enzyme concentrations and activities that increases or reduces protein productivity and glycosylation which depending on different cell lines and biological products [137, 143-148]. To maintain product quality, understanding and controlling the operating parameters of the glycosylation process is critical.

Kinetic modeling has been widely used to capture the relations between the operating conditions with the dynamic profile of glycan fractions as mentioned in Chapter 2. Here research works that have been focusing on glycosylation kinetic modeling is provided with more details. In terms of feeding strategies, Radhakrishnan et al. [37] used kinetic model to quantify the change of glycan fractions under dynamic addition of media supplements (including MnCl2 and EDTA). Kotidis et al. [36] investigated the glycosylation precursor feeding, including galactose and uridine effects on cell growth, protein productivity and quality. The authors also applied dynamic optimization to

maximize the concentration of galactosylated mAb fractions. Luo et al.[149] modified the kinetic model by adding delta function to capture the amino acid and copper effect on the monoclonal antibody productivity and glycosylation. As for process parameters, Villiger et al. [35] captured pH, manganese, and ammonia concentration and predicted their effect on cell culture and glycosylation process. Sou et al. [34] successfully used their kinetic model to simulate cell culture, nucleotide, nucleotide sugar metabolic, and N-linked glycosylation under two different temperatures. However, the kinetic models usually contain a large number of kinetic pathways, which can be a challenge in parameter estimation. Many of the above models simulate the process under different conditions separately as for example to capture the effects of temperature. In addition, the mechanistic model is usually computationally expensive, which can be prohibitive for the evaluation of design space.

To deal with this challenge, statistical analysis tools have been used to capture the influence of multiple independent factors to the system outputs while reducing the computational cost. Thus, it has been used in media formulation and building process-product correlations to characterize the biological system [6]. Sokolov et al. [49] used sequential multivariate tools with partial least squares regression (PLSR) to predict titer, aggregation, low molecular weight components, and glycan groups in the product under different process scales. They illustrate the capability of multivariate analysis in scale-up and decision making for biopharmaceutical manufacturing. Zurcher et al. [50] also used PLSR to capture protein glycosylation profiles using process variables and extracellular variables (127 variables). The authors compared the performance of using just process variables to that with the addition of extracellular variables and showed that using only the process variables leads to higher error in predicting afucosyaltion, fucosylation, and galactosylation.

In this chapter we incorporate temperature and pH factors into the mechanistic kinetic model and capture the dynamic change of the viable cell density, glucose concentration, mAb concentration as well as the glycan fractions. With the results that is obtained from the mechanistic model, further,

we applied dynamic kriging to capture the dynamic trend of these outputs under different operating conditions. It is found that dynamic kriging is able to capture the dynamic behavior of output variabilities under the change of process operating parameters with high accuracy. Furthermore, a surrogate-based adaptive sampling approach has been used to determine the feasible operating region (design space) for protein production and glycosylation process based on the requirement of commercial product quality attributes.

4.2   Modeling and analysis method

4.2.1   Kinetic modeling for cell growth and glycosylation process

The cell system is very complicated with metabolic cycles, TCA cycles and each of these cycles contains many reactions. Considering all the reactions will increase the equation development and parameter estimation difficulties. Thus, the unstructured culture model that treats cells as a black box is used to simulate cell growth. Typically, this method contains algebraic differential equations from mass balance and empirical kinetic equations including Monod kinetics, mass action kinetics and Michaelis-Menten kinetics.

First, mass balance is developed to capture mass balance of each components in the bioreactor as shown from equation 4.2.1 to 4.2.5.

Volume: $\dfrac{dV}{dt} = F_{in} - F_{out}$ (4.2.1)

Viable cell: $\dfrac{d(V[X_v])}{dt} = F_{in}[X_{v0}] + \mu V[X_v] - \mu_d V[X_v] - F_{out}[X_v]$ (4.2.2)

Dead cell: $\dfrac{d(V[X_d])}{dt} = F_{in}[X_{d0}] + \mu_d V[X_v] - F_{out}[X_d]$ (4.2.3)

mAb production: $\dfrac{dV[mAb]}{dt} = F_{in}[mAb_0] + q_{mAb}V[X_v] - F_{out}[mAb]$ (4.2. 4)

Metabolites concentration: (4.2. 5)

$\dfrac{d(V[metabolite])}{dt} = F_{in}[metabolite]_{in} - F_{out}[metabolite] + q_{metabolite}VX_v$

Where μ represents cell growth rate; μ_d is cell death rate; F is flow rate, $[metabolite]$ represents metabolite concentrations which can also represent nutrients such as glucose; the protein/metabolites production or consumption rate indicates as q in this equation.

In this study, the out model is developed based on the dynamic model that is developed by Val et al. [150]. In order to consider the effects of temperature and pH in kinetic model, it is critical to understand the mechanism behind it. It has been found that reducing the temperature would reduce GnTII, GalTI, GalTIII and FucT expression level, leading to the reduction of the intracellular NSD concentration, mAb productivity and further reduction of the mature glycoforms in the final product[148]. Thus, the model is developed based on adding linear regression terms to capture the temperature effect on cell growth kinetics and intracellular enzyme expression. Controlling pH would affect the osmolality and impact the cell metabolism. pH shift also usually affects protein productivity[35]. Thus, pH is only linked to the rate relative to protein production. The modified kinetic model is based on the Val et al. [150]. For brevity in presentation the entire model is shown in the appendix and this section concentrates on the modifications.

The growth rate μ is represented by equation (4.2.6).

$$\mu = (\frac{\mu_{max}}{T} + a)(\frac{C_{Glc}}{K_{Glc} + C_{Glc}} - \frac{[X_v]}{\alpha_x}) \qquad (4.2.\ 6)$$

where μ_{max} is the maximum growth rate, T is temperature, α_x is cellular carrying capacity, K_{Glc} is Monod constant for glucose, and a is regression constant.

Equation (4.2.7) represents death rate μ_d.

$$\mu_d = (\frac{\mu_d^{max}}{T} + b)\frac{(\frac{K_{d\mu}}{T} + c)}{(\frac{K_{d\mu}}{T} + c) + \mu} \qquad (4.2.\ 7)$$

where μ_d^{max} is maximum death rate, $K_{d\mu}$ is inverse specific death rate. b and c are regression constants.

Equation (4.2.8) show the glucose consumption rate q_{Glc}.

$$q_{\text{Glc}} = -\frac{1}{\left(\frac{Y_{XGlc}}{T} + d\right)[C_{Glc}] + \left(\frac{K_{Glc}}{T} + f\right)}[C_{Glc}] \tag{4.2. 8}$$

where Y_{XGlc} is yield coefficient of glucose; C_{Glc} is glucose concentration; d and f are regression constant.

The mAb production rate can be represented by equation (4.2.9).

$$q_{mAb} = \left(\frac{Y_{\frac{mAb}{Glc}}}{T} + e\right)q_{Glc}e^{-\frac{1}{2}\left(\frac{pH_{shift} - pH_{opt}}{w}\right)^2} \tag{4.2.9}$$

where $Y_{mAb/Glc}$ is yield coefficient of mAb production from glucose consumption; e is regression constant; pH_{opt} is optimal culture pH; w is pH dependent mAb productivity constant; pH_{shift} is the pH control during the cell culture.

In terms of the glycosylation process, the Golgi apparatus is considered as plug flow reactor (PFR) where along the axial direction, different distributions of enzymes are existed. Enzyme concentration is also linearly correlated to temperature. The mass balance is shown in equation (4.2.10). It assumes that there is no axial dispersion within the compartment through the PFR and protein transfer maintains a linear velocity. The Golgi diameters are constant, and no mass transfer limitation affects the glycosylation reactions.

$$\frac{\partial [G_m]}{\partial t} = -V_1 \frac{\partial [G_m]}{\partial z} + \sum_n^{Enzyme} v_{m,n} r_n \tag{4.2.10}$$

where $[G_m]$ represents glycan (m) concentration, V_1 represents the linear velocity of glycans through the Golgi apparatus, z is the length of Golgi, r_n is the kinetic rate for enzyme reaction n, $v_{m,n}$ is the reaction coefficient of glycan m that catalyzed by enzyme n.

It should be mentioned that q_p can be obtained from the cell culture model and used as an input condition to calculate the linear velocity of protein transferring in golgi apparatus as shown in the equation (4.2.11)

$$(V_1)(Vol_{Golgi})[Man9]_{z=0} = \frac{q_p}{MW_{mAb}}\left(\frac{2umol_{Glyc}}{umol_{mAb}}\right) \tag{4.2.11}$$

where $Vol_{Golgi} = 25 \ \mu m^3$ is the volume of Golgi Apparatus[150, 151], $[Man9]_{z=0} = 55\frac{\mu mol}{L_{Golgi}}$ represents the Man9 concentration at the entrance of Golgi. MW_{mAb} is average molecular weight which is 150 kDa. The reactions with the product and substrates related inhibitions can be simulated by Michaelis-Menten kinetics, sequential-order Bi-Bi kinetics, and random-order Bi-Bi kinetics as shown in equation (4.2.12) to equation (4.2.16) [151].

$$r_n = \frac{k_{f,n}[E_n][G_m]}{k_{d,\frac{m}{n}}\left(1 + \frac{[G_m]}{k_{d,\frac{m}{n}}} + \frac{[G_{m-1}]}{k_{d,m-1/n}}\right)} \tag{4.2.12}$$

$$r_n = \frac{k_{f,n}[E_n][NS_k][G_m]}{K_{d,m/n}K_{d,k/n}\left(1 + \frac{[NS_k]}{K_{d,k/n}} + \frac{[NS_k]}{K_{d,k/n}}\frac{[G_m]}{K_{d,m/n}} + \frac{[NS_k]}{K_{d,k/n}}\sum_{z=1}^{N.C.}\frac{[G_z]}{K_{d,z/n}} + \frac{B_k}{K_{i,k/n}}\frac{[G_{m+1}]}{k_{d,(m+1)/n}} + \frac{[B_k]}{K_{d,k/n}}\right)} \tag{4.2.13}$$

$$r_n = \frac{k_{f,n}[E_n][NS_k][G_m]}{K_{d,\frac{m}{n}}K_{d,\frac{k}{n}}\left(1 + \frac{[NS_k]}{K_{d,\frac{k}{n}}} + \frac{[G_m]}{K_{d,\frac{m}{n}}} + \sum_{z=1}^{N.C.}\frac{[G_z]}{K_{d,\frac{z}{n}}} + \frac{[NS_k]}{K_{d,\frac{k}{n}}}\frac{[G_m]}{K_{d,\frac{m}{n}}} + \frac{[NS_k]}{K_{d,\frac{k}{n}}}\sum_{z=1}^{N.C.}\frac{[G_z]}{K_{d,\frac{z}{n}}} + \frac{B_k}{K_{i,\frac{k}{n}}}\frac{[G_{m+1}]}{k_{d,\frac{(m+1)}{n}}} + \frac{[G_{m+1}]}{k_{d,\frac{(m+1)}{n}}} + \frac{[B_k]}{K_{d,k/n}}\right)} \tag{4.2.14}$$

$$[En] = [E_{n,max}] \times e^{\left(\frac{(z-z_{n,max})}{2w_n}\right)^2} \tag{4.2.15}$$

$$[E_{n,max}] = a_n \times T + b_n \tag{4.2.16}$$

$k_{f,n}$ is the rate-limiting turnover rate for enzyme n. $[NS_k]$ is nucleotide sugar concentration. $K_{d,\frac{m}{n}}$ is the dissociation constant of the acceptor enzyme complex and $K_{d,\frac{k}{n}}$ is the dissociation constant of the donor enzyme complex. $K_{d,\frac{z}{n}}$ is the dissociation constant of the competitor enzyme complex, $[En]$ is enzyme concentration.

The mechanistic model is built on MATLAB. The systems of ordinary differential equation are solved in MATLAB (ODE45) and the partial differential equations from intracellular structured model is first discretized along the axial using finite difference methods and solved by ODE solvers in MATLAB.

4.2.2 Kriging and dynamic kriging

Kriging or gaussian modeling is an interpolation method that uses sum of spatial weighted distance of observed function values at nearby sample points to predict new points[152, 153]. It also provides the mean squared error for the prediction[154]. Equation (7) shows the general equation of kriging model to predict $\hat{f}(x^i)$.

$$\hat{f}(x^i) = \beta f(x^i) + \varepsilon(x^i) \tag{7}$$

The first part of the equation represents a known regression model that defines the global trend of the data $f(x^i)$, β is unknown parameter. In this work, constant and linear regression models are tested and the one that gives the least mean squared error is selected. The second part is a residual term $\varepsilon(x^i)$ that indicates the error at location x^i which is usually normally distributed with zero mean and variance σ^2. Covariance function between $\varepsilon(x^i)$ and $\varepsilon(x^j)$ is shown in equation (8).

$$Cov(\varepsilon(x^i), \varepsilon(x^j)) \tag{8}$$

$$= \sigma^2 R(x^i, x^j)$$

where R is the correlation function and exponential, linear, squared exponential models are commonly used in kriging surrogates[155], shown in equation (9)-(11).

$$R(\theta, x^i, x^j) = \exp\left(-\sum_{h=1}^{d} \theta_h |x_h^i - x_h^j|^{\theta_{d+1}}\right) \quad 0 < \theta_{d+1} < 2 \tag{9}$$

$$R(\theta, x^i, x^j) = \exp\left(-\sum_{h=1}^{d} \theta_h |x_h^i - x_h^j|^2\right) \tag{10}$$

$$R(\theta, x^i, x^j) = \max(0, 1 - \sum_{h=1}^{d} \theta_h |x_h^i - x_h^j|) \tag{11}$$

θ_h is unknown parameter. Parameters θ_h, β and σ^2 can be predicted by using maximum likelihood.The detailed derivation of kriging can be found in papers Bhosekar and Ierapetritou[155]. In this work, the kriging

models are built using DACE toolbox [156] and different correlation functions are tested and selected for model training.

Dynamic kriging is a modification of kriging model as shown in equation (12).

$$\hat{f}(x_k^i) = \beta f\left(x_k^i, \hat{f}(x_{k-1}^i)\right) + \varepsilon(x^i, \hat{f}(x_{k-1}^i)) \tag{12}$$

The dynamic system is first discretized into different time points k, and the kriging model is used as an autoregressive model that collects the predicted results $\hat{f}(x_{k-1}^i)$ from the previous time point (k-1) and combines with the state variables or control input x_k^i to estimate the future time point $\hat{f}(x_k^i)$[157]. This means that the kriging algorithm is iteratively called and the database that is used for prediction is dynamically updated. The method maximizes the amount of information about the dynamic response for the prediction[153].

4.2.3 Feasibility analysis

To find the feasible region of bioreactor operation, the process operations need to satisfy the productivity (product titer) and product quality (different glycan fractions) constraints as shown in equation (13).

$$g_i(x) \leq 0 \quad j \in J \tag{13}$$

where $g_i(x)$ represent different constraints as mentioned above, x includes the uncertain variables including temperature and pH and operating parameters including the total operating time, initial conditions of cell density, glucose and mAb concentrations. In this study, it assumes that the design parameters and control parameters are constant. To satisfy all the constraints, feasibility function is defined in equation (14). If $\varphi < 0$, θ represents the feasible region. When $\varphi = 0$, the defined condition is right at the boundary of feasible region.

$$\varphi(x) = \max_{j \in J} g_i(x) \qquad (14)$$

Initial sample points can be generated by space filing and a feasible region can be obtained by calculating feasibility function. To further simplify the model, kriging is used to determine the feasible region. Adaptive sampling method is used to improve the accuracy of the feasible boundary. By maximizing the modified EI function, shown in equation (15), the new sample points that close to feasible region and unexplored region are generated to update the kriging model.

$$EI_{feas}(x) = \hat{s}(x)\phi\left(-\frac{\hat{y}(x)}{\hat{s}(x)}\right) = \hat{s}(x)\frac{1}{\sqrt{2\pi}}e^{-0.5\left(\frac{\hat{y}(x)^2}{\hat{s}(x)^2}\right)} \qquad (15)$$

Standard error $\hat{s}(x)$ can be obtained from kriging prediction at location x, and $\hat{y}(x)$ is the predicted value. The detailed explanation of the modified EI function can be found in paper [158]. To test the performance of the feasibility analysis, CF%, CIF% and NC% values are calculated and represented in equation (16) to (18).

$$CF\% = \frac{CF}{CF + ICIF} \times 100 \qquad (16)$$

$$CIF\% = \frac{CIF}{CIF + ICF} \times 100 \qquad (17)$$

$$NC\% = \frac{ICF}{CF + ICF} \times 100 \qquad (18)$$

CF% represents the percentage of the feasible region that is successfully explored. CIF% represents the percentage of the infeasible region that is successfully explored. NC% represents the percentage of the feasible region that is overestimated. CF is feasible region that is correctly defined by kriging model. CIF is infeasible region that is correctly defined by kriging model. ICF is an infeasible region but defined as feasible region by kriging model. ICIF is feasible region but defined as infeasible region by kriging model.

4.3   Case study

To clearly demonstrate the use of mechanistic modeling and dynamic kriging, a case study is presented in this section. First, mechanistic model that captures temperature and pH effects on cell culture and glycosylation process is developed in section 4.3.1 and the results are compared with literature data. Based on the simulation built, a design of experiment (DoE) is used to generate designs for dynamic kriging model building and results are shown in section 4.3.2. Finally, section 4.3.3 describes how to implement feasibility analysis to the dynamic kriging model to obtain the design space for protein production.

4.3.1   Mechanistic modeling experiment

A kinetic model is built to capture the viable cell, glucose, and mAb concentrations under different temperatures and pH. It assumes the cells are first cultured under 37°C pH=7, and then shifted to a different temperature and pH on day 5. Three runs are used to test the temperature effect. For each of the run, pH was set to 7 and temperature is shifted to 33.5°C, 35°C and 36.5°C respectively. Similarly, three runs are used to test the pH effect with pH shifting to 6.8, 6.9 and 7, respectively. Conditions of the cell culture follow the following assumptions. Glucose is added on days 5, 8 and 12 during the cell culture. A feed with nutrients is also provided to the system on days 2, 5, 8, 12 to maintain cell growth. Due to lack of experimental data, all the fittings are based on the regression of the kinetic rate constants under different temperatures from literature [16, 21].

The viable cell density and glucose concentration under different temperatures (constant pH=7) are shown in Figure 4.3 1.a and Figure 4.3 1.b. Due to the addition of glucose and nutrients, sharp peaks can be observed. The figures indicate that viable cell concentration is reduced with the temperature shifting down, which also causes the reduction of glucose consumption rate. Similar trend has been observed in studies[136, 159]. pH effect to viable cell and glucose concentrations are not considered in this work. It has been reported in the literature that no significant effects have been

observed with a certain range of pH change. For example, Trummer et al. [145] showed that specific cell growth rate is not affected when pH shifted from 7.10 to 6.9 on day 5. Figure 1.c shows that the protein titer increases with the reduction of the temperature. Since the viable cell concentration is reduced, the specific protein production rate (qp) increases as the temperature reduces which is consistent with what is reported in the literature[32, 148]. Figure 2 also shows that the reduction of pH reduces the protein titer. This observation has also been reported in Villiger et al[160].
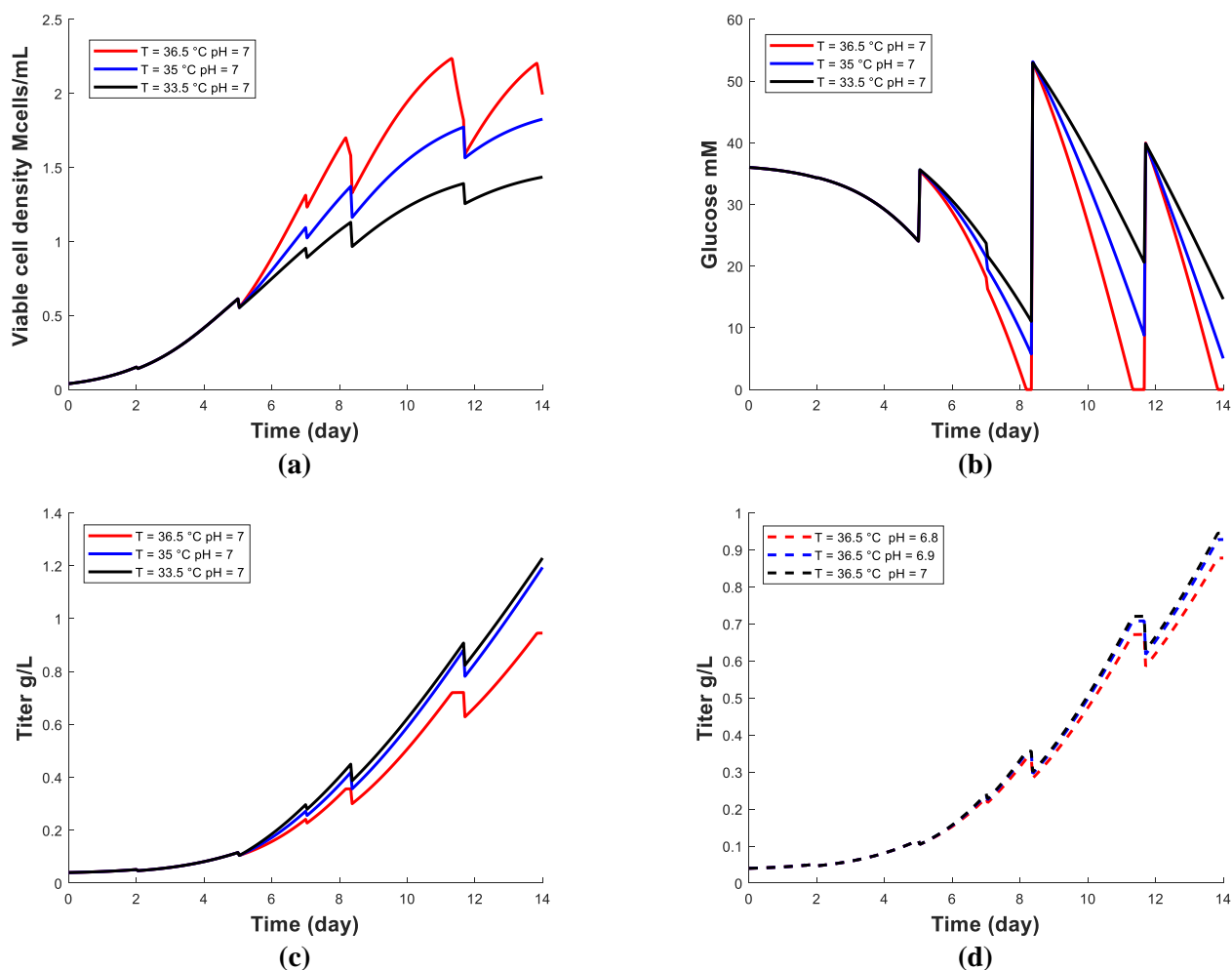

Figure 4.3 1 Dynamic change in (a) viable cell density, (b) glucose concentration and (c) mAb concentration under different temperatures. (d) Dynamic change in mAb concentration under different pH values.

In a previous study, it was shown that the protein specific production rate and glycan fractions are affected by different temperatures in the CHO cell culture. In general, for mAbs, the shifting of

temperature usually causes an increase in protein production rate, while decreasing the temperature usually causes a reduction in glycosylation processes, such that the galactosylation, fucosylation and sialyation are reduced [148] [136].

The pH effects on the protein productivity and glycosylation vary from cell line to cell line. It is shown in Table 4.3 1 that increasing and decreasing the pH can both cause a decrease in protein productivity. Changing the protein productivity has a great effect on protein glycosylation. Jiang et al. [36] reported that the galactosylation increases as q_p decreases, regardless of the shift in pH. However, inconsistency is still observed in the trend of Man5 and G2FS1, as shown in Table 4.3 1

Table 4.3 1 pH effect on titer and glycan fraction production from CHO cell

Protein	pH range	qp	Glycan fraction	Ref
mAb	↓ pH 7.15-6.70.	↓	Reduced: G0F, G0, Man5 Increased: G1F, G2F, G2FS1	[35]
mAb	↓ pH 7.2 -6.9	-	Reduced: G0	[146]
mAb	↓ pH 6.9-6.7	↓	Increased: G1F+G2F, Man5, galactosylation	[161]
	↑ pH 6.9-7.3	↓	Decrease: Sialylation	

In general, a higher protein production rate reduces the residence time of protein in the Golgi apparatus. Protein has less contact time with the enzymes which reduces glycosylation in the final product. Thus, in the model, the production rate is used to connect the cell culture process and protein glycosylation. By incorporating temperature and pH effects to protein productivity, glycosylation process can be then captured. On the other hand, temperature also affects enzyme expression. The mRNA expressions of the glycosyltransferase including GnTII, β-GalT and FucT are significantly lower when temperature decreases.[148] The parameters for enzyme concentration under different temperature are obtained from the literature[34]. It also needs to note that the reduction of temperature reduces the consumption of glucose which further reduces nucleotide

sugar including UDP-Glc, UDP-Gal and UDP-GlcNAc synthesis thus reduced the processed glycan structure. Figure 4.3 2 shows the change of glycan fractions under different temperature shift. Decreasing the temperature increases the afucosylation and reduces galactosylation and fucosylation which is consistent with that in the literature.
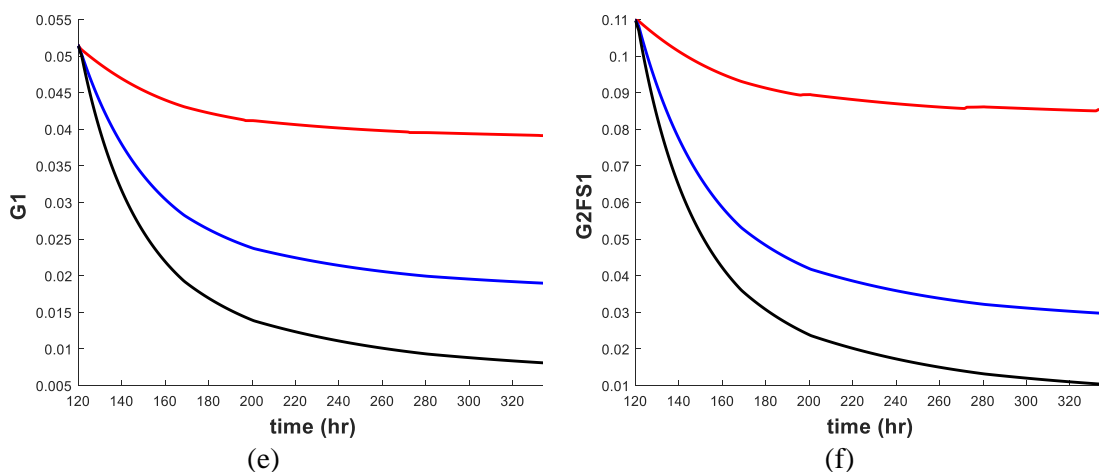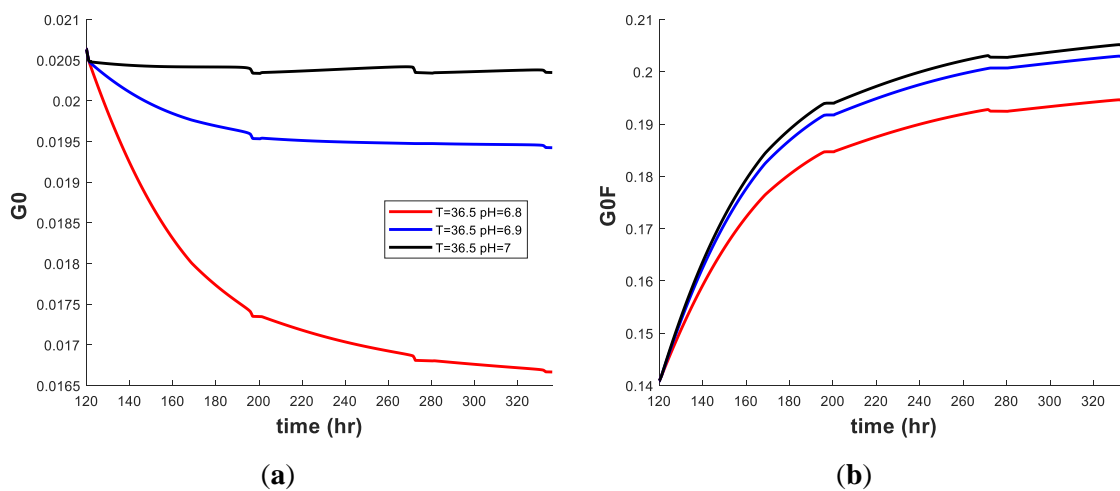

(**a**)

(**b**)

(c)

(d)

(e)

(f)
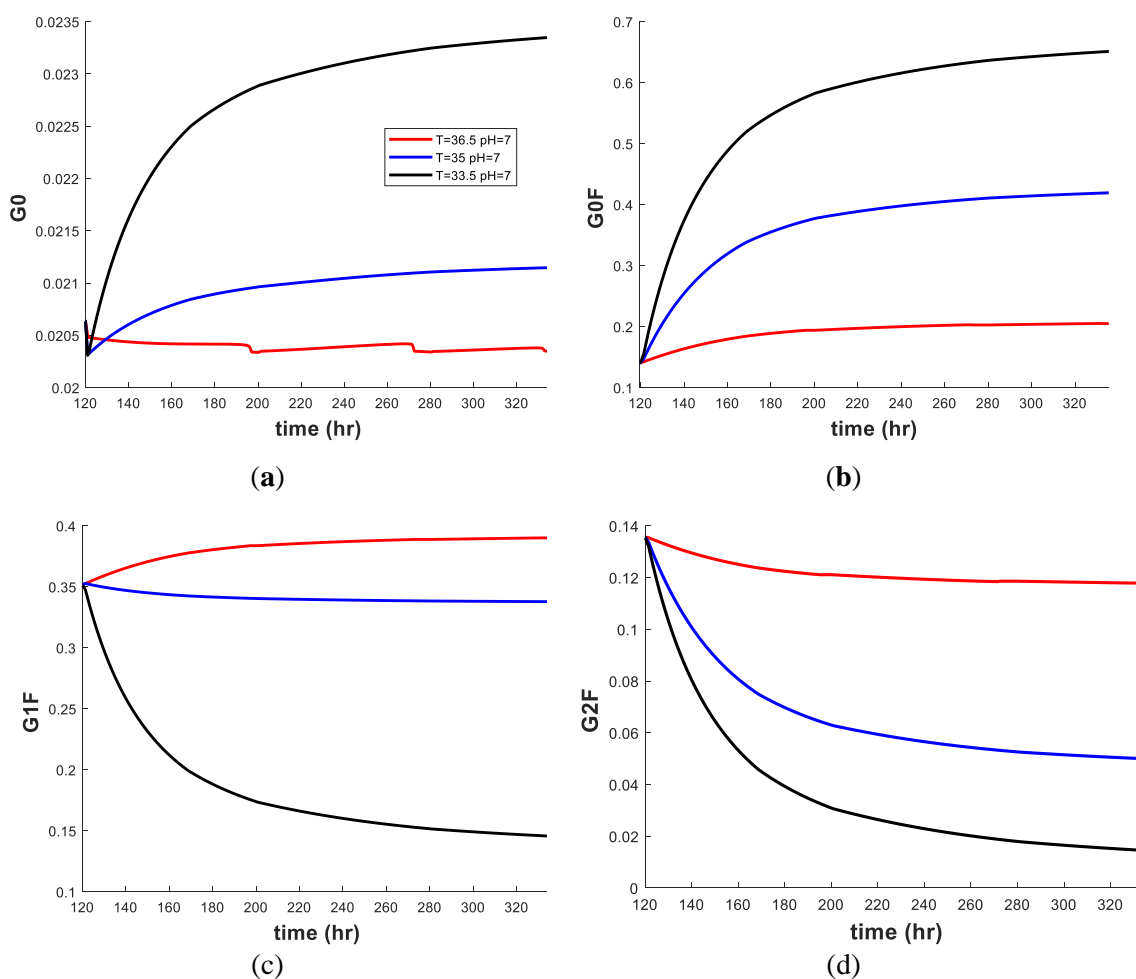
Figure 4.3 2 Glycan fractions under different temperature (a) G0, (b) G0F, (c) G1F, (d) G2F, (e) G1, (f) G2FS1

Figure 4.3 3 shows the change of glycan fractions under different pH. Decreasing the pH reduces the afucosylation but increases galactosylation and fucosylation which consists with the results shown in the Table 4.3 1.
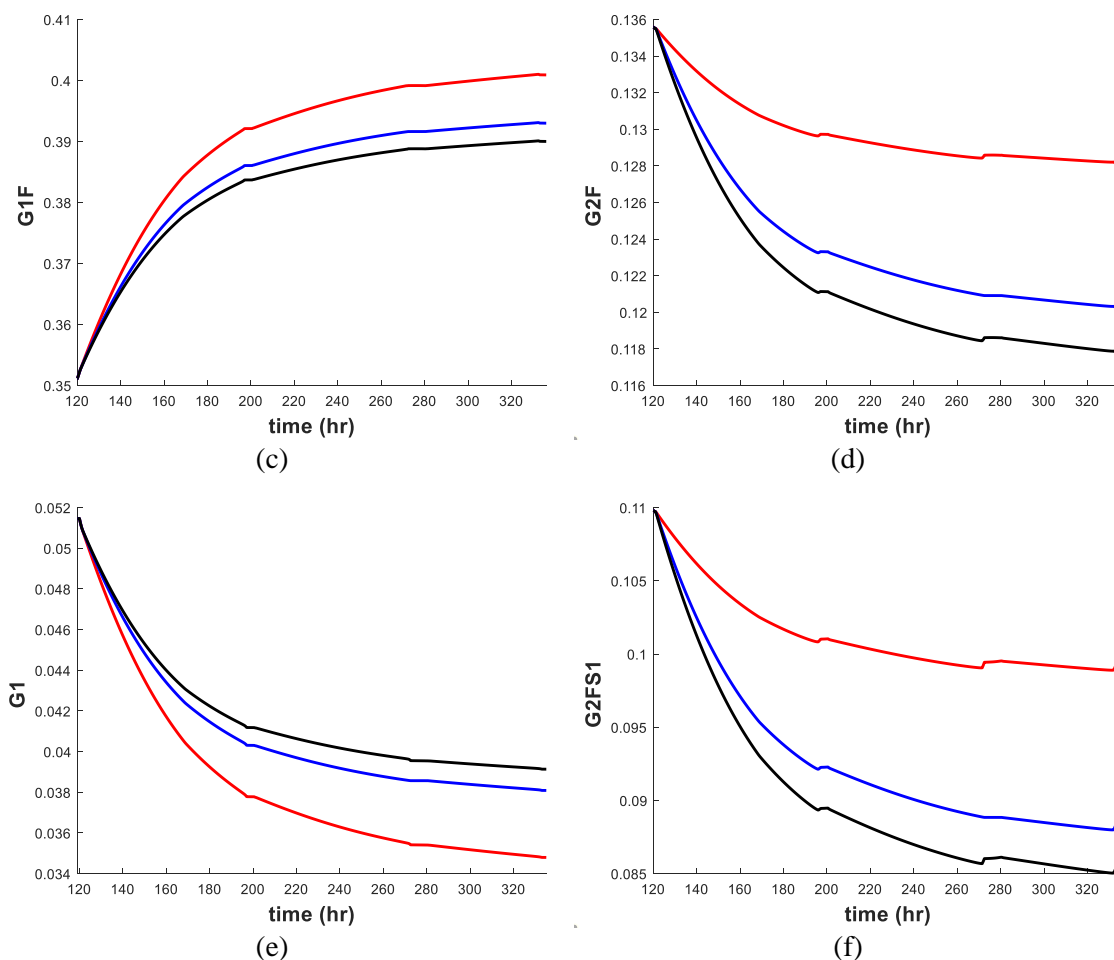

(**a**)

(**b**)

Figure 4.3 3 Glycan fractions under different pH (a) G0, (b) G0F, (c) G1F, (d) G2F, (e) G1, (f) G2FS1.

All the trend of glycan fractions under different temperature and pH are summarized in Table 4.3 2. Comparing the results from Table 4.3 1 as well as the literature that focused on the effect of temperature shift, it indicates that all the predictions from the developed mechanistic model are consistent with the results from the literature. From the prediction, shifting down pH would reduce the titer and increase the fucosylation and galactosylation, and shifting temperature down would increase the product titer and reduce the galactosylation.

Table 4.3 2 Model predicted glycan fractions under different Temperature and pH

| Conditions	Titer	G0	G0F	G1/G2	G1F/G2F	G2F1S
pH shifted down	↓	↓	↓	↓	↑	↑
Temp Shifted down	↑	↑	↑	↓	↓	↓

4.3.2 Kriging model fitting results

Regular Kriging vs. Dynamic Kriging

The kinetic model results are used as training sets to build the regular kriging and dynamic kriging models. A design of experiment (DoE) based on two level full factorial design is used to obtain data from the kinetic model. Five conditions with temperature and pH shifts are used: 1) T = 33.5 °C, pH = 7; 2) T = 33.5 °C, pH = 6.8; 3) T = 35 °C, pH = 6.9; 4) T = 36.5 °C, pH = 6.8; 5) T = 36.5 °C, pH = 7. The kriging model reduces the computational complexity, provides an easier means of model fitting, and has capabilities in process optimization. The purpose of this section is to test the capabilities of dynamic kriging in the prediction of product concentration and quality with a small amount of data.
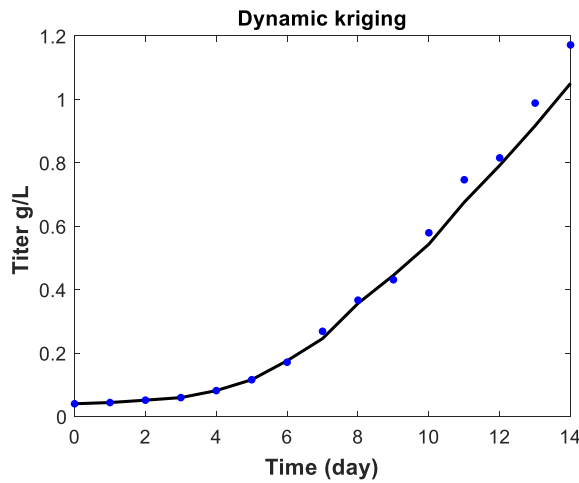
For regular kriging, to predict viable cell concentration, glucose and protein titer, input parameters including time, temperature, pH and glucose addition at different time points are used. For dynamic kriging, the input parameters start with the same input parameters as for regular kriging, and the input datasets are dynamically updated based on the predictions from the previous time points. A random culture condition (T = 34.5 °C, pH = 6.85) is selected to test the fitting performance. Figure 4.3 4 a,b,c show the good prediction of dynamic kriging results for viable cell, glucose and mAbs. In general, dynamic kriging provides higher prediction accuracy than regular kriging, especially in the prediction of glucose concentration, which has a more complicated trend, as shown in Figure 4.3 4 b,d. This is mainly because dynamic kriging considers more sample points and correlations during the model prediction. By considering the output from previous time points, the prediction of the next time point is very sensitive to system change.
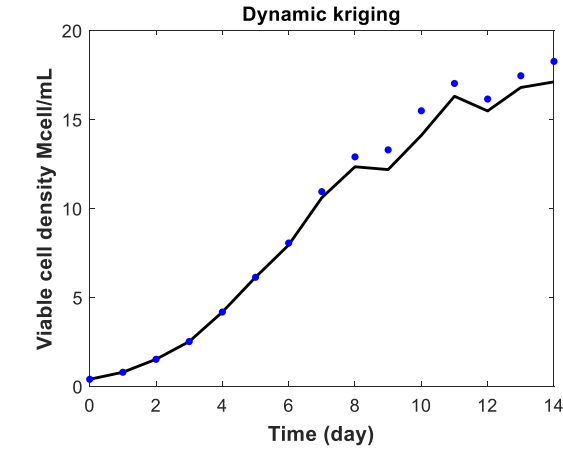
(**a**)
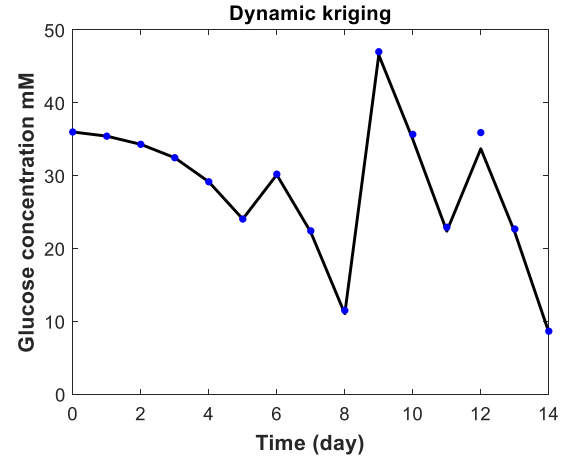
(**b**)

(**c**)
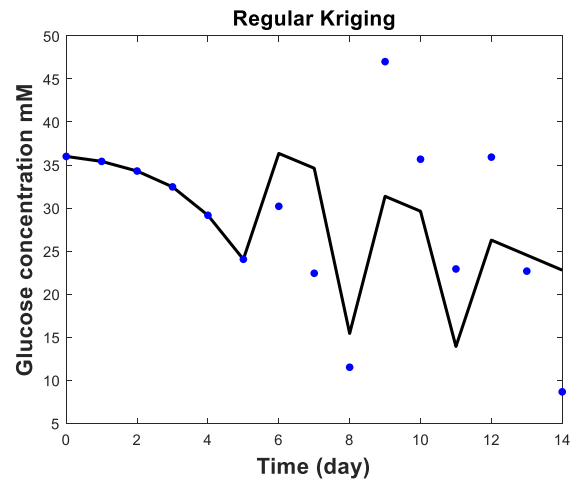
(**d**)

Figure 4.3 4 Prediction of viable cell (a), glucose (b), and mAbs (c) from dynamic kriging. Prediction of glucose concentration (d) from regular kriging

To predict the glycan fractions, the operating parameters, cell culture data and glycan fractions (at $t-1$ time points) are used as the input datasets that train the dynamic kriging model. During the prediction, the input parameters together with viable cell concentration, protein titers and Man5 glycan fractions at ($t-1$) time points are used as the input to predict the Man5 glycan fractions at t time point by dynamic kriging. The same idea is applied to other glycan fractions.

Theoretically, the prediction can start from day 1, since the mechanistic model provides the glycan fractions at an early stage. Thus, data from day 1 from the mechanistic model can be used as training datasets for data-driven model training, as shown in Figure 4.3 5. In the experimental work, however, the glycan fractions' data for the first 4 days are hard to obtain. This is mainly due to the low product concentration in the solution, which results in a low intensity during the glycan fraction measurements. The predictions from day 1 and day 5 are all tested in this study, and the mean relative squared errors (MRSE) [38] are compared in Table 4.3 3. This shows that dynamic kriging is able to predict the glycan fractions with high accuracy (all the MRSE <10%). As shown in Figure 4.3 5, dynamic kriging provides good prediction in the early stage, and the error becomes large at the end of the cell culture. It is found that the higher errors occur in the prediction of G1F, G2F and G2FS1, which could be due to the limited training dataset used in this case study.
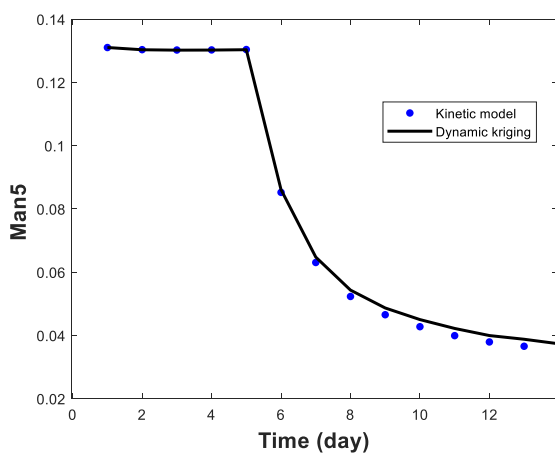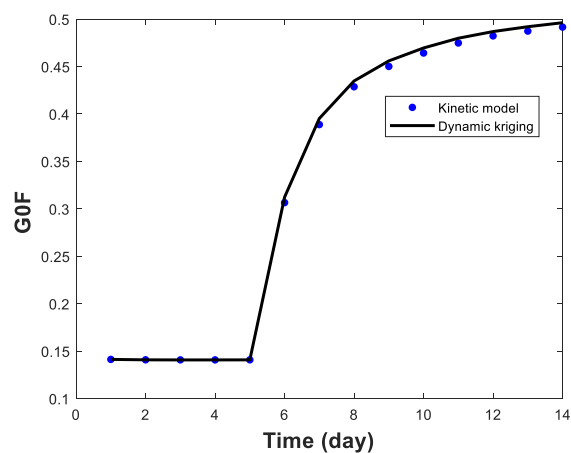
(**a**)   (**b**)
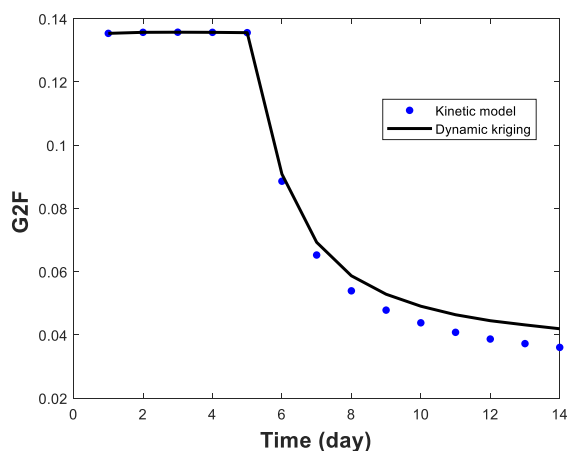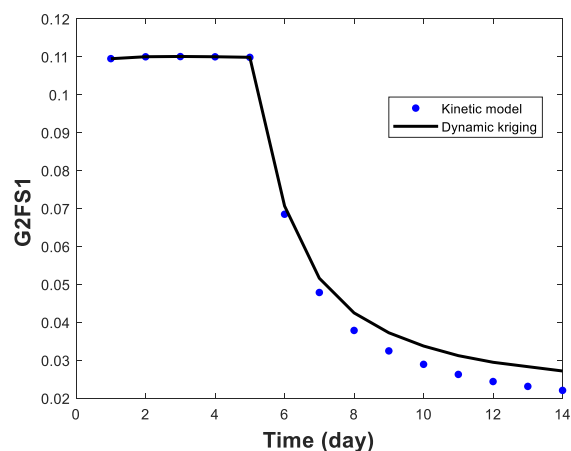
(**c**)   (**d**)

Figure 4.3 5 Prediction of different glycan fractions from dynamic kriging: (a) Man5, (b) G0F, (c) G2F, (d) G2FS1.

Table 4.3 3 Mean relative squared error (MRSE) of dynamic kriging prediction

Glycan fractions	Man5	G0	G1	G1F	G2F	G2FS1
Day 5	0.029	0.004	0.027	0.088	0.073	0.088
Day 1	0.018	0.004	0.017	0.072	0.044	0.0510

4.3.2.1   Prediction of Temperature Effect from Dynamic Kriging Model

Different operating conditions are also used to further test the prediction of dynamic kriging. Since 33.5, 35 and 36.5 °C are used as training data sets, different temperatures T = 34, 34.5 and 36 °C are tested at pH 7. With the temperature increasing, G0 and G0F are reduced, while G1, G1F, G2F and G2FS1 are increased. From Figure 4.3 6, all the results are matched to those obtained by the mechanistic model.
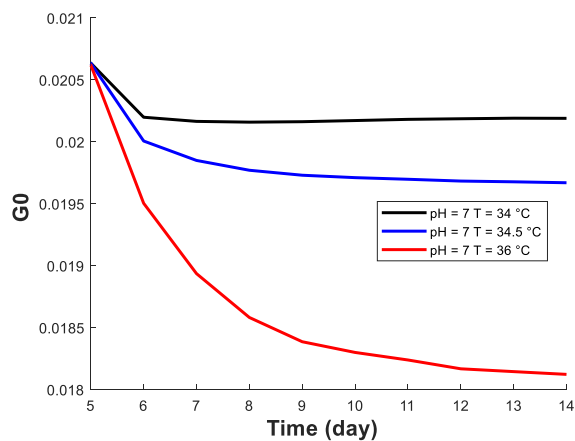
(**a**)
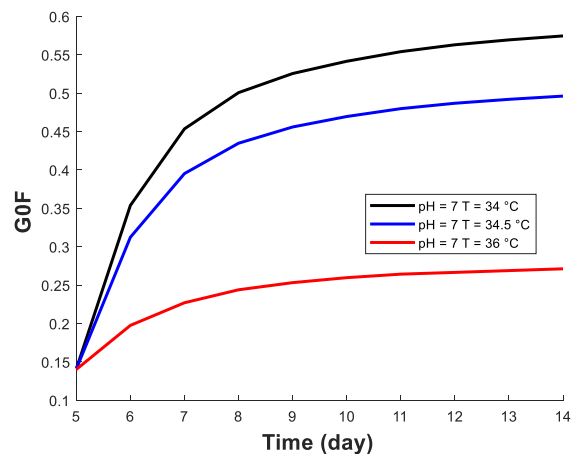
(**b**)
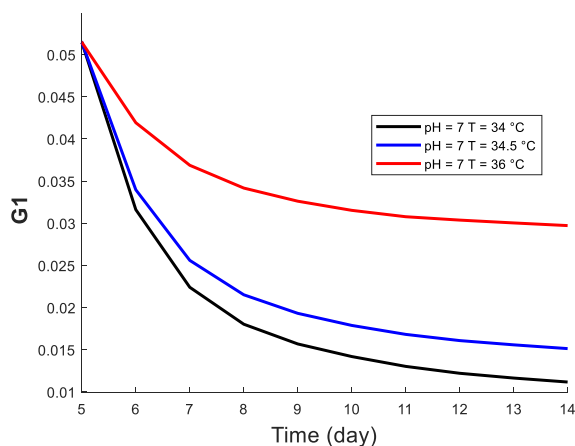
(**c**)
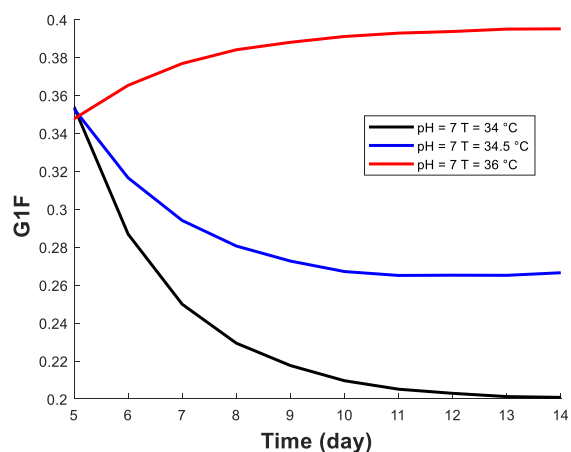
(**d**)
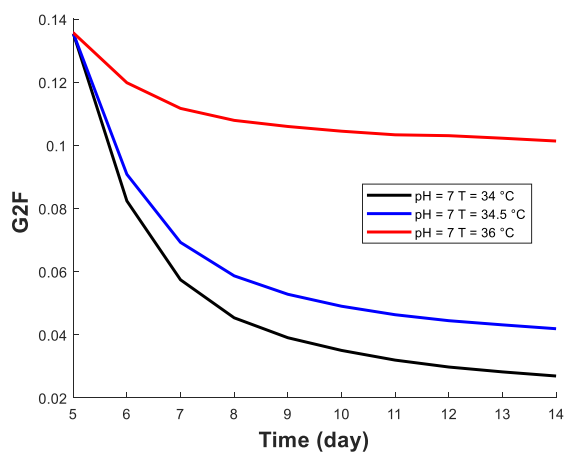
(**e**)
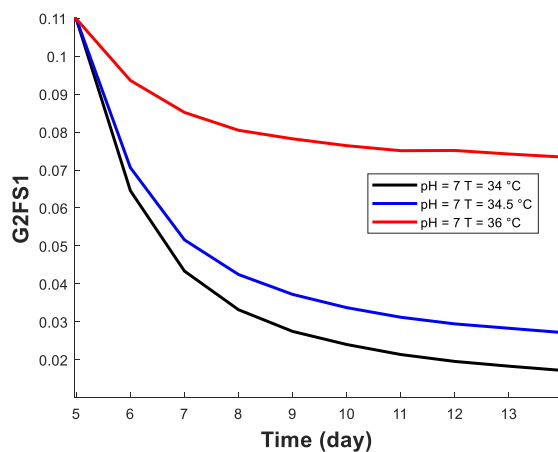
(**f**)

Figure 4.3 6 The prediction of glycan fractions under different temperatures from dynamic kriging: (a) G0, (b) G0F, (c) G1, (d) G1F, (e) G2F, (f) G2FS1.

4.3.3 Feasibility analysis results

As mentioned in the previous section, the glycan fractions are critical quality attributes used to evaluate the biological product's performance and for the testing of biosimilars. For this case study, three quality attributes (high mannose, afucosyaltion and galactosylation of Herceptin (trastuzumab)) are tested. According to the FDA Briefing Document Oncologic Drugs Advisory Committee Meeting, the requirements for the glycan profile of a product biosimilar to Herceptin are obtained [162]. To satisfy the quality attributes, the total afucosylation needs to be maintained within the range of 2% to 14.5%, galactosylation needs to stay within 20% to 70%, and high mannose needs to be below 8%. Within the operating range, the end points of high mannose, afucosyaltion and galactosylation prediction at day 14, under different temperatures and pH values, are shown in Figure 4.3 7. This shows that afucosylation ranges from 2.5% to 6%, that is, within the requirement for the afucosyaltion ratio. However, the galacosylation changes from 5% to 47%, which means that in this range of conditions there is a risk of being lower than the required range of values. Thus, a constraint needs to be considered to ensure that GI can reach at least 20% in the feasibility analysis. Similarly, high mannose reaches up to 9%, which also needs to be considered as a constraint in the feasibility analysis.

(**a**)   (**b**)

(**c**)
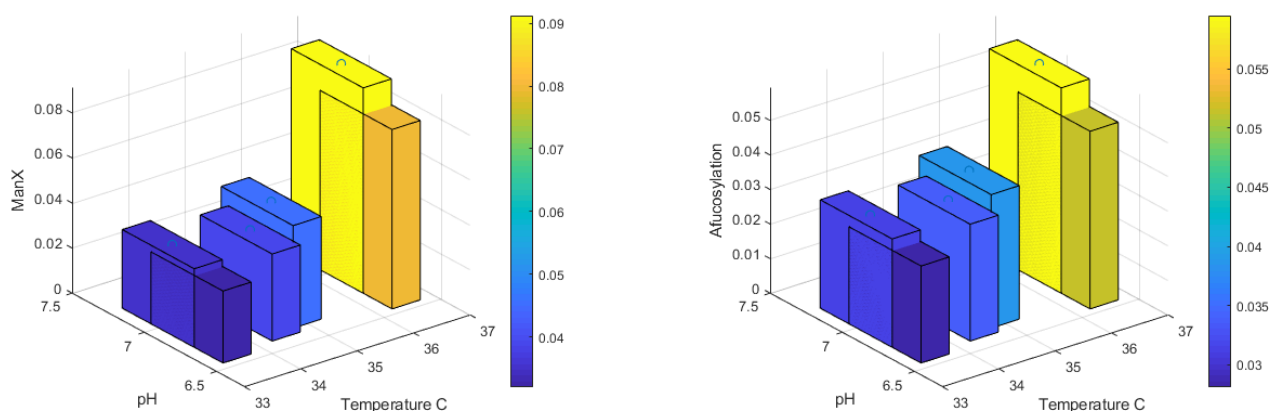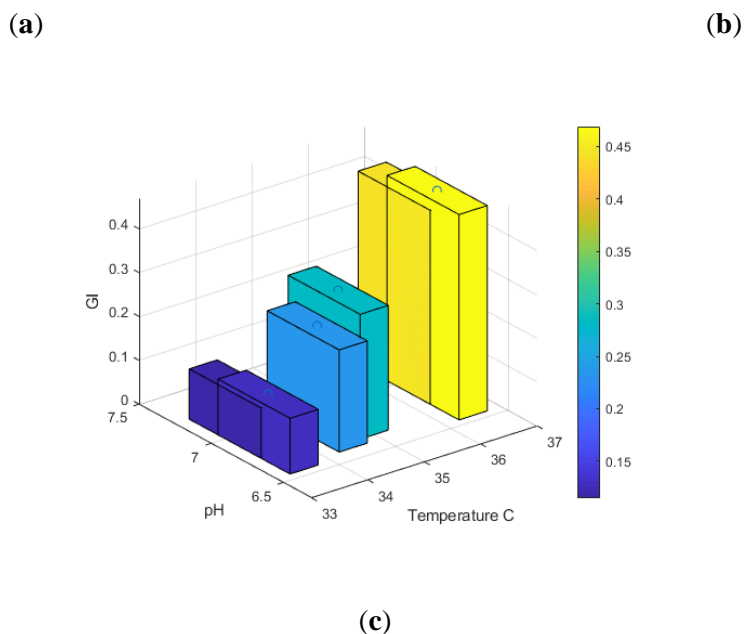
Figure 4.3 7 (a) High mannose, (b) afucosylation, (c) GI fractions within the defined operating ranges. The color variation represents glycan fractions.

From Figure 4.3 1 c,d, we see that temperature and pH have effects on the protein titer. Thus, in order to maintain high productivity, the protein concentration is set as more than 1 g/L. This section aims to maintain the afucosyaltion and galactosylation within the required range, and these components can be predicted by dynamic kriging. One approach is to develop the dynamic kriging that predicts the glycan index directly based on the culture conditions, viable cell, and mAb concentrations. The predictions of the glycan index are shown in Figure 4.3 8. Table 4.3 4 shows the MRSE values. The results show that dynamic kriging provides a good prediction of all the glycan indices (MRSE < 2%). The second approach used has a good prediction performance. It should be noticed that although only the end point glycan fractions are important for quality assurance, understanding the dynamic profile of different glycan fractions is important in order to enable better quality control during production.
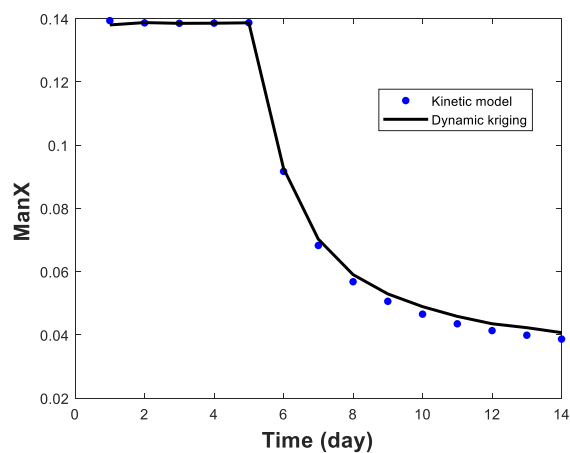
(**a**)
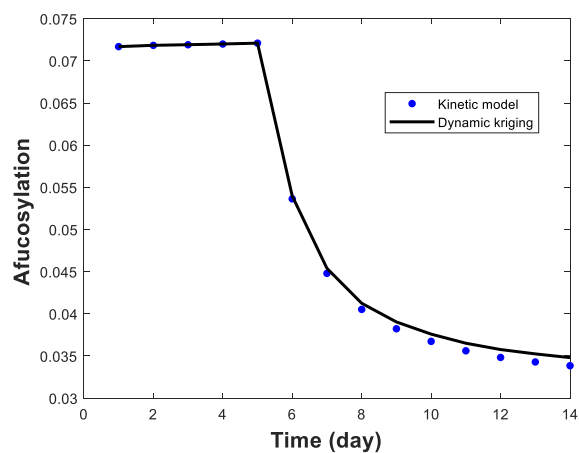
(**b**)
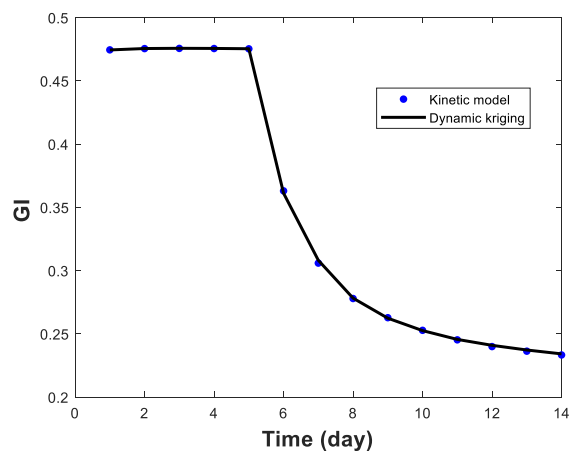
(**c**)
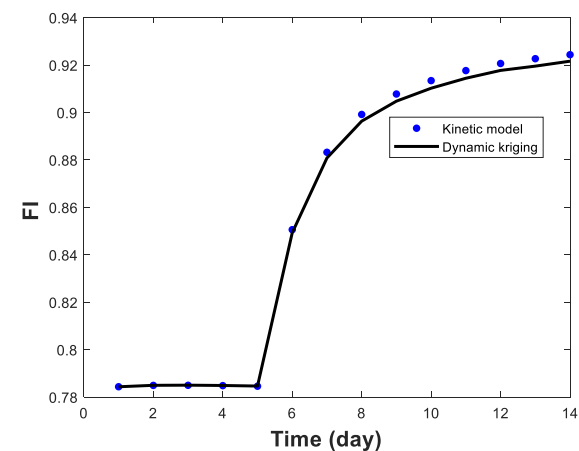
(**d**)

Figure 4.3 8 Glycosylation index predicted by dynamic kriging: (a) high mannose, (b) afucosylation, (c) GI, (d) FI.

Table 4.3 4 Means relative squared error (MRSE) of dynamic kriging prediction.

Glycan Index	Afucoyslation	ManX	GI	FI
MRSE	0.0121	0.0186	0.0025	0.0026

In total, 25 initial sample points were obtained from the kriging model and used to obtain the initial

feasible region. Adaptive sampling is used to improve the accuracy of the feasible region. Figure

4.3 9 demonstrates the results of the feasibility analysis. The dark blue line shows the predicted feasible region's boundary (feasibility function = 0), which indicates the operating region that satisfies all the constraint requirements. The circle points are infill points, which are added to improve the accuracy of the feasible region. Finally, the three performance measures' distributions are CF = 0.997, CIF = 0.999 and NC = 0.001, which indicates that the feasible region has been accurately determined.
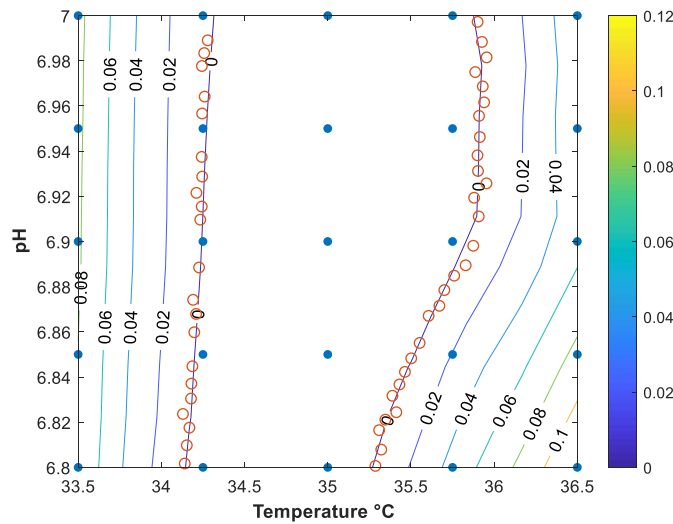
Figure 4.3 9 Feasible operating region obtained from adaptive sampling. Contour plot for feasibility function values under different pH values and temperatures. Zero line represents the feasible region's boundary. Initial sampling points are shown as blue dots and red circles are adaptive sampling points.

4.4 Conclusion

In this work, we developed a kinetic model to successfully capture the effects of temperature and pH on protein production and the mAb's product quality attributes. By using mechanistic relations, the kinetic model improves the understanding of the effects of physical stimuli on the intracellular reactions. The mechanistic model is further simpli-fied using dynamic kriging, which is able to provide dynamic predictions with high accuracy. This step reduces the computational costs and improves the efficiency and accuracy of model fitting. Surrogate-based feasibility analysis is used to determine the design space for bioreactor modeling. The adaptive sampling method is used to

determine the design space for process operations efficiently. The developed methods can be used to predict titer and protein qualities under different conditions. They handle multiple con-straints and provide the design space boundaries for efficient bioreactor operation. Addi-tional work is needed to implement the proposed framework, using more experimental data for CHO bioreactor operations. In the experimental cell culture system, variabilities and noises may exist. Data cleaning and preprocessing are needed to improve prediction accuracy.

5 Predictive modeling of cell culture and protein glycosylation processes

5.1 Background

In Chapter 4, a framework has been developed in capturing the dynamic trend of cell culture and protein production by both mechanistic and data driven model and further defined of design space for bioreactor operations. In this section of the dissertation, we would like to apply the mechanistic model in a real case study with experimental support to further test our model capabilities.

In this chapter, a simplified single-cell model is developed to dynamically simulate the CHO cell culture process in fed-batch bioreactors. The model contains unstructured cell culture model which is coupled to structured glycosylation model with reduced number of parameters. Cell culture model simulates cell growth and death, nutrients consumption and metabolites' concentrations. The glycosylation model predicts major glycoprotein concentrations varying with time including intermediates glycosylated proteins and final product protein quality attributes. Experimental data are provided from our collaborators and used for model parameter estimation and model validation. Temperature and pH are changed to investigate their effect on product compositions. In this study, cell was first cultured under temperature 35 ℃ and pH 7.1. Starting from day 4, 4 conditions are tested including 1) keep the temperature and pH constant 37 ℃ 2) shift temperature down, from 37 ℃ to 35 ℃, pH kept constant 3) shift the temperature up, from 37 ℃ to 39 ℃, pH kept constant 4) shift the pH down, from 7.1 to 6.7, the temperature kept constant. Using experimental data, the model is further developed by including temperature effect in the mechanistic model. The intracellular parameters that estimated by the model are used to interpret the experimental phenomena and guide the future experiments. At the beginning of this chapter, I would like to first acknowledge Prof. Chundawat's research group for providing us cell culture experimental data.

5.2   Modeling method

To capture temperature and pH effect on cell growth and glycosylation process and guide process optimization, first principle kinetic modeling needs to be developed. A kinetic model is built based on the simplified cell kinetic metabolism. As mentioned in Chapter 4, both structured and unstructured models are used to construct the model to predict extracellular components and intracellular reactions. In this section, the model is developed based on the character of the CHO-K1 cell line.

In the fed-batch system, media feed and glucose addition are generally used to maintain cell activities. Especially for small scale cell culture, the reduction of the volume due to sampling is also needed to capture. Equation 5.2.1 captures the mass balance of volume where F_{in} and F_{out} represents the volumetric flow in and flow out of the system. In the similar way, mass balance is applied to simulate viable cell density, dead cell, mAb production and metabolites concentration and represented by equation 5.2.1 to 5.2.5.

Volume: $\frac{dV}{dt} = F_{in} - F_{out}$ $\qquad\qquad$ (5.2.1)

Viable cell: $\frac{d(V[X_v])}{dt} = F_{in}[X_{v0}] + \mu V[X_v] - \mu_d V[X_v] - F_{out}[X_v]$ $\qquad$ (5.2.2)

Dead cell: $\frac{d(V[X_d])}{dt} = F_{in}[X_{d0}] + \mu_d V[X_v] - F_{out}[X_d]$ $\qquad$ (5.2.3)

mAb production: $\frac{dV[mAb]}{dt} = F_{in}[mAb_0] + q_{mAb}V[X_v] - F_{out}[mAb]$ $\qquad$ (5.2.4)

Metabolites concentration: $\qquad\qquad$ (5.2.5)

$$\frac{d(V[metabolite])}{dt} = F_{in}[metabolite]_{in} - F_{out}[metabolite] + q_{metabolite}VX_v$$

The cell growth rate μ, death rate μ_d are lumped parameters which are represented by Monod equations based on their limited substrates (i) and inhibited metabolites (j).

Cell growth rate: $\mu = \mu_{max} \prod_i^n \frac{C_i}{K_{Ci}+C_i} \prod_j^m \frac{K_{Cj}}{K_{Cj}+C_i}$ $\qquad$ (3.2.1-6)

μ_{max} represents the maximum specific growth rate and K_{Ci} and K_{Cj} are Monod constant for components i and j. From the literature research, it has been found that for CHO cell line, major nutrient includes glucose is used in glycolysis and form adenosine trophophyte (ATP) to support cell activities [163]. And during the cell growth, the production of ammonia negatively impacts productivity, and quality. One explanation is that ammonium increases the electrochemical gradient and impacts the intracellular pH, which further inhibits enzymatic activity. Thus, in this model, glucose and ammonia are the two major components to consider. Although the death rate is also affected by growth limited components such as ammonia [35], as ammonia is one of the factors of the cell growth rate, a more lumped equation is developed in equation (5.2.7) [150].

$$\mu_d = \mu_d^{max} \frac{K_D}{K_D + \mu} \tag{5.2.7}$$

where μ_d^{max} is the maximum death rate and K_{DCj} is Monod constant. The specific production rate of protein can be represented by nutrient related terms (Y_{mAb/C_i}) and non-nutrient related term m_{mAb} as shown in equation (5.2.8).

$$q_{mAb} = Y_{mAb/C_i} q_{ci} + m_{mAb} \tag{5.2.8}$$

Y_{x/C_i} is the mAb yield coefficient from component C_i which is glucose in this equation. Two ways can be used to capture metabolites and nutrients consumption. The first way is similar to mAb production that considers the nutrients uptake as growth and non-growth related terms [34, 164-166]. Another way is to consider the nutrient concentration and specific yield of consumption based on cell growth [150]. In this system, the carbon source glucose, byproduct ammonia concentration can be modeled by equation (3.2.1-10).

$$q_{ci} = \frac{\mu}{Y_{\mu/C_i}} + m_{c_i} \tag{5.2.9}$$

Y_{XCi} is the yield coefficient of cell from component C_i and Y_{CjCi} is the yield coefficient of C_j from C_i consumption. Through the cell culture, glucose is consumed thus $Y_{\mu/C_{glc}}$ and $m_{c_{glc}}$ are negative

numbers. To relate temperature parameter with temperature, Arrhenius and polynomial equations are used as shown in the Table 5.2 1 below.

Table 5.2 1 Parameters with temperature change for cell culture

Parameters	Equation
$Y_{\frac{u}{c_{amm}}}, Y_{\frac{u}{c_{glc}}}, Y_{\frac{mAb}{c_{glc}}}$	$P_m = A_m e^{-\frac{E_{a_m}}{RT}}$
$\mu_{max}, K_{Ci}, K_{Cj}, \mu_d^{max}, m_{mAb}, m_{c_i}$	$P_l = BT^2 + CT + D$

It shows that most of the yield coefficient terms can be represented by Arrhenius equation and all of them are changed monotonically as the temperature increases. Monod constants and growth independent terms are captured by polynomial terms. These trends are dependent from cell line to cell line.

The majority of the glycosylation process happens in Golgi Apparatus. To capture the dynamic pathway of glycosylation process in Golgi system, a single PFR model has been developed. These models are based on cisternal maturation model that consider the transport of cargo proteins through the Golgi complex [167] similar to what we have introduced in Chapter 4. The model is shown in (5.2.10).

$$\frac{\partial[G_m]}{\partial t} = -V_1 \frac{\partial[G_m]}{\partial z} + \sum_{n}^{Enzyme} v_{m,n} r_n \qquad (5.2.10)$$

The equation assumes no axial dispersion within the compartment through the PFR, and protein transfer maintains a linear velocity. The Golgi diameters are constants and there is no mass transfer limitation affecting the glycosylation reactions. $[G_m]$ represents glycan (m) concentration and r_n represents kinetic rate for enzyme reaction n. $v_{m,n}$ is the reaction coefficient of glycan m that catalyzed by enzyme n. V_1 represents the linear velocity that protein glycan transfers through the Golgi apparatus. Z is the length of Golgi. Three types of reactions exist in the Golgi apparatus for different types of enzymes. In these reactions, one enzyme will have multiple substrates, thus there

are competitive and product inhibition reactions. The equations 3.2.1-11 to 3.2.1-13 are the simplified enzyme kinetics. Equation (5.2.11) shows the Michelis Menten equation with competitive and product inhibition for mannose removal by enzyme ManI and ManII. The Equation (5.2.12) shows the sequential-order Bi-Bi equation for N-acelylglucosamine and galactose addition by GnTI, GnTII and GalT. In this reaction, enzyme first bind with nucleotide sugar precursor and then react with different types of glycan. Equation (5.2.13) is the random-order Bi-Bi equation for fructose, sialic acid addition by FucT and SiaT. Unlike the sequential order reaction, the enzyme can first react with either glycan or nucleotide sugar precursor and then the other one.

$$r_n = \frac{k_{f,n}[E_n][G_m]}{k_{d,\frac{m}{n}}(1 + \frac{[G_m]}{k_{d,\frac{m}{n}}} + \frac{[G_{m-1}]}{k_{d,m-1/n}})} \tag{5.2.11}$$

$$r_n = \frac{k_{f,n}[E_n][NS_k][G_m]}{K_{d,m/n}K_{d,k/n}(1 + \frac{[NS_k]}{K_{d,k/n}} + \frac{[NS_k]}{K_{d,k/n}}\frac{[G_m]}{K_{d,m/n}} + \frac{[NS_k]}{K_{d,k/n}}\sum_{z=1}^{N.C.}\frac{[G_z]}{K_{d,z/n}} + \frac{B_k}{K_{i,k/n}}\frac{[Glyc_{i+1}]}{k_{d,(i+1)/n}} + \frac{[B_k]}{K_{i,k/n}})} \tag{5.2.12}$$

$$r_j = \frac{k_{f,n}[E_n][NS_k][G_m]}{K_{d,\frac{m}{n}}K_{d,\frac{k}{n}}(1 + \frac{[NS_k]}{K_{d,\frac{k}{n}}} + \frac{[G_m]}{K_{d,\frac{m}{n}}} + \sum_{z=1}^{N.C.}\frac{[G_z]}{K_{d,\frac{z}{n}}} + \frac{[NS_k]}{K_{d,\frac{k}{n}}}\frac{[G_m]}{K_{d,\frac{i}{n}}} + \frac{[NS_k]}{K_{d,\frac{k}{n}}}\sum_{z=1}^{N.C.}\frac{[G_z]}{K_{d,\frac{z}{n}}} + \frac{B_k}{K_{i,\frac{k}{n}}}\frac{[G_{m+1}]}{k_{d,\frac{(i+1)}{n}}} + \frac{[G_{m+1}]}{k_{d,\frac{(i+1)}{n}}} + \frac{[B_k]}{K_{i,k/n}})} \tag{5.2.13}$$

$k_{f,n}$ is the rate-limiting turnover rate for enzyme n. $[NS_k]$ is nucleotide sugar concentration. $K_{d,m/n}$ is the dissociation constant of the acceptor enzyme complex and $K_{d,k/n}$ is the dissociation constant of the donor enzyme complex. $K_{d,\frac{z}{n}}$ is the dissociation constant of competitor enzyme complex.

In the PFR system, the enzyme is distributed along the axial direction and can be represented by the following equation:

$$[E_n](z) = E_{n,max}\exp\left(-\frac{1}{2}\left(\frac{z - z_{n,max}}{w_n}\right)^2\right) \tag{5.2.14}$$

$E_{n,max}$ represents the peak concentration of enzyme, and w_n is the width of the distribution. $z_{n,max}$ represents the position along the Golgi where the enzyme reaches highest concentration. Similar to cell culture model, dissociation rate is also correlated with temperature in different format as shown in Table 5.2 2.

Table 5.2 2 Parameter with temperature change for glycosylation

Parameters	Equation
$K_{d,ManII}$, $K_{d,FucT_{G0}}$, $UDP-Gal$	$P_m = A_m e^{-\frac{E_{a_m}}{RT}}$
The rest of the dissociation rate and nucleotide surgar concentration	$P_l = B_l T^2 + C_l T + D_l$

It has been found that only dissociations rate for ManII and FucT$_{G0}$, UDP-Gal concentration satisfy the Arrhenius equation. All other components have polynomial trends as temperature increases. To simplify the model, it assumes enzyme concentration kept constant, however it is not true. This means the change of the rate constant under the temperature shift shows a lumped effect of both enzyme activities and concentrations.

The model is developed in MATLAB R2018b, the unstructured cell culture model is solved by ODE45. The structured glycosylation model is the system of partial differential equations (PDEs) and is first decomposed along the axial coordinate of Golgi apparatus by finite difference method. The resulted ODEs are solved by ODE15s

5.3   Parameter estimation

In the mechanistic model, parameters including maximum growth rate, death rate, yield coefficient, dissociation constant need to be estimated to capture the dynamic trend for the cell density, productivity and glycan fractions. The general way to describe the kinetic model can be represented in equation (5.3.1).

$$y_{cal} = f(\theta; x) \tag{5.3.1}$$

where x and y_{cal} represent the input; θ is the parameter estimated. To obtain the parameters, least-squares minimization is used to obtain the least difference between the calculated results and experimental results. The objective function is shown in equation (5.3.2).

$$F(\theta) = \sum_{i}^{n} \sum_{j}^{m} \sum_{t}^{T} \left(y_{obs_i,j,t} - y_{sim_{i,j,t}} \right)^2 \tag{5.3.2}$$

$y_{obs_i,j,t}$ and $y_{cal_{i,j,t}}$ are the jth components in ith experimental runs under different conditions at time t. Due to different components are under different magnitudes and will devote various weights to the parameter estimation when adding them up during the parameter estimation. Each of the component is standardized using equation (5.3.3). The obtained model output is also processed with the same treatment as shown in equation (5.3.4).

$$y_{obsi,j,t} = \frac{y_{obsi,j,t,original} - \overline{y_{obsi,j,original}}}{\sigma_{obsi,j,original}} \tag{5.3.3}$$

$$y_{sim_{i,j},t} = \frac{y_{sim_{i,j},t,model} - \overline{y_{obsi,j,original}}}{\sigma_{obsi,j,original}} \tag{5.3.4}$$

Where $y_{obsi,j,t,original}$ represent the data that are obtained directly from experiment. $\overline{y_{obsi,j,original}}$ are the average value of the data from day 0 to day 14. The standard deviation of original data from experiment is $\sigma_{obsi,j,original}$. $y_{sim_{i,j},t,model}$ represents the outputs that solved from ODEs. In this way, all the data are transferred to the data format with zero mean and standard deviation equals 1. But the shape of the data and their trends will not be changed. In this way, the contributions from each component are weighted. The estimation starts with initial points that are obtained from literature and then simulated annealing is applied for global optimization. The idea for simulated annealing is to screen multiple possible local minima within the large search space.

The output will be used as the initial points for local optimization where 'fminsearchcon' and 'fmincon, sqp' function in MATLAB 2018b are used in sequence.

The parameter estimation starts from cell culture model. Single-cell productivity is thus obtained and used as input to the glycosylation model. There are 11 parameters estimated in cell culture model including growth and death rate, Monod constants, growth-related and nongrowth-related parameters. 14 parameters are estimated for glycosylation model. It needs to notice that only enzyme dissociation rate and nucleotide sugar concentrations in glycosylation model are estimated and enzyme concentrations in Golgi apparatus are obtained from literatures [150, 165, 168].

On one hand, although researches in the literature investigate the temperature and pH effect to glycosylation process, few of them corporate temperature in their model. On the other hand, physical stimuli to cell growth, protein production, and product quality are independent from cell line to cell line and product to product. Thus, the first step of the model development is to understand how model parameters are changed with temperature and pH. In this study, parameters are first evaluated independently from run to run and correlate the change with temperature using the Arrhenius equation or regression model. Since cells are first cultured under 37 °C, pH =7.1 for the first four days and then shifted to a different pH and temperature except for the control run. All the bioreactor runs are fitted for the first four days together with control runs, in order to improve the generalization and accuracy of the parameter estimation. For those runs with temperature shift, a new set of parameters are obtained. Once the parameters are obtained, the regression models are used to capture the trend.
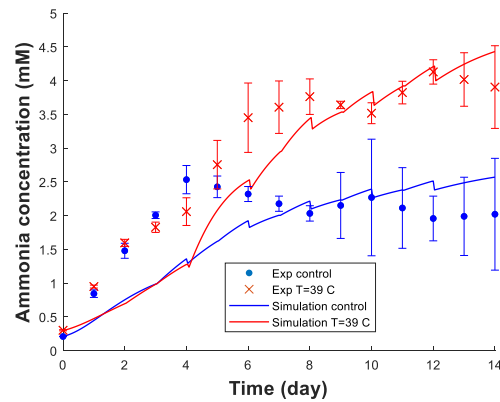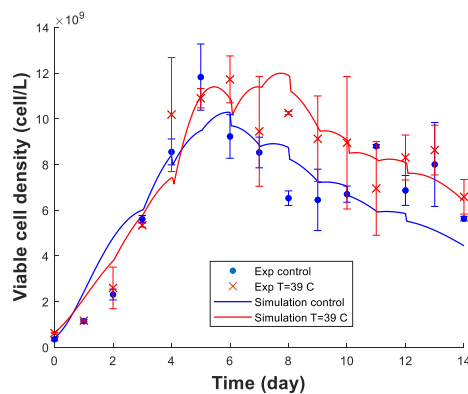
5.4   Results

5.4.1   Cell culture and protein production

Model parameters are estimated under four culture conditions. In general, the prediction of viable cell density, glucose concentration, ammonia concentration and protein titers are all in good agreement with experimental results shown in the Figure 5.4 1 to Figure 5.4 3.

For cell growth, exponential growth phase, stationary phase and death phase exist in all bioreactor runs. With pH shifting down, cells have a slightly (1-2 days) longer exponential growth phase than control phase while their maximum cell concentrations and dynamic trend do not have a significant change. This result consistent with that in the literature which use the same cell line[53]. The reduction of the temperature slightly increases the cell concentration which also consistent with that found in the literature[137]. However, it is hard to conclude if temperature shifts up affect the cell growth. Reading from the Figure 5.4-1, viable cell concentration with temperature shift up is slightly higher than that of the control. However, their initial culture concentrations are around two times higher as well. (0.6 M cell/ml for T=39 ℃ , and 0.3 M cell/ml for T =37 ℃). Nolan et al. [43] reported an increasing integrated viable cell density when increases the initial concentration from 0.9 M cell/mL to 1.8 M cell/ml. The results indicated that increasing, decreasing temperature and pH increase the titer as well as the cell productivity. Especially with the reduction of the temperature, cell productivity is doubled compared to the control run in the cell death phase. Similar results are found in the literature for CHO-K1 cell [169]. One explanation is that the mild hypothermia condition reduces the death rate and nutrient consumption rates during the cell culture. Last but not least, ammonia concentration is also tested in this study. The reduction of temperature reduces the production of ammonia. Increasing the temperature and decreasing the pH increase the ammonia concentration. Decreasing pH increases the ammonia concentration for around 4 times.

In terms of the model, the general trend of each component that under different temperature and pH shift are captured by kinetic model. However, viable cell density under temperature shift down

is poorly captured, especially for the exponential growth phase. This is mainly due to the high initial ammonia concentration at the beginning of the cell culture (around 1-3-fold higher) however the viable cell density does not show a significant decrease for the first few days. For the current developed equation, only glucose and ammonia concentrations are considered as factors to cell growth. The high concentration of the ammonia reduces the growth rate of the cell growth in the model which cause the deviations of the prediction. To deal with this problem, one suggestion is to include more factors, such as asparagine, lactate and glutamine in the prediction of the cell growth rate [43, 163]. Another possible reason is that ammonia concentration does not have a significant effect to cell growth when it is below to a certain level[170]. A biological replication of the temperature and shifts down is suggested to further confirm the critical factors that affect temperature change.
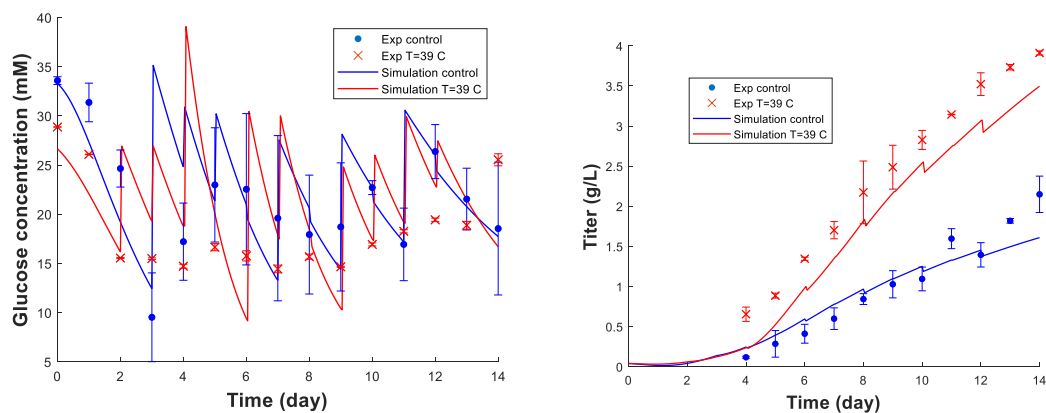
Figure 5.4 1 experimental data and simulation fitting results for cell culture components under control and temperature shift up.
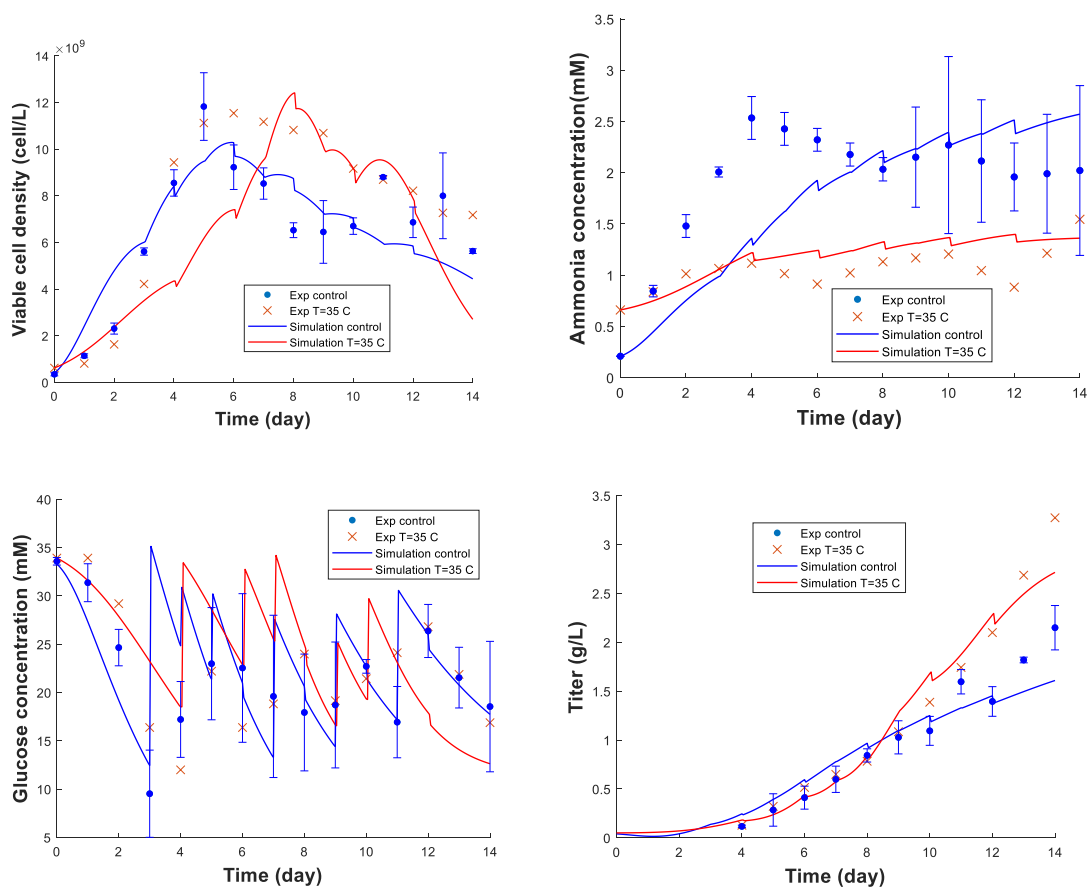
Figure 5.4 2 experimental data and simulation fitting results for cell culture components under control and temperature shift down.
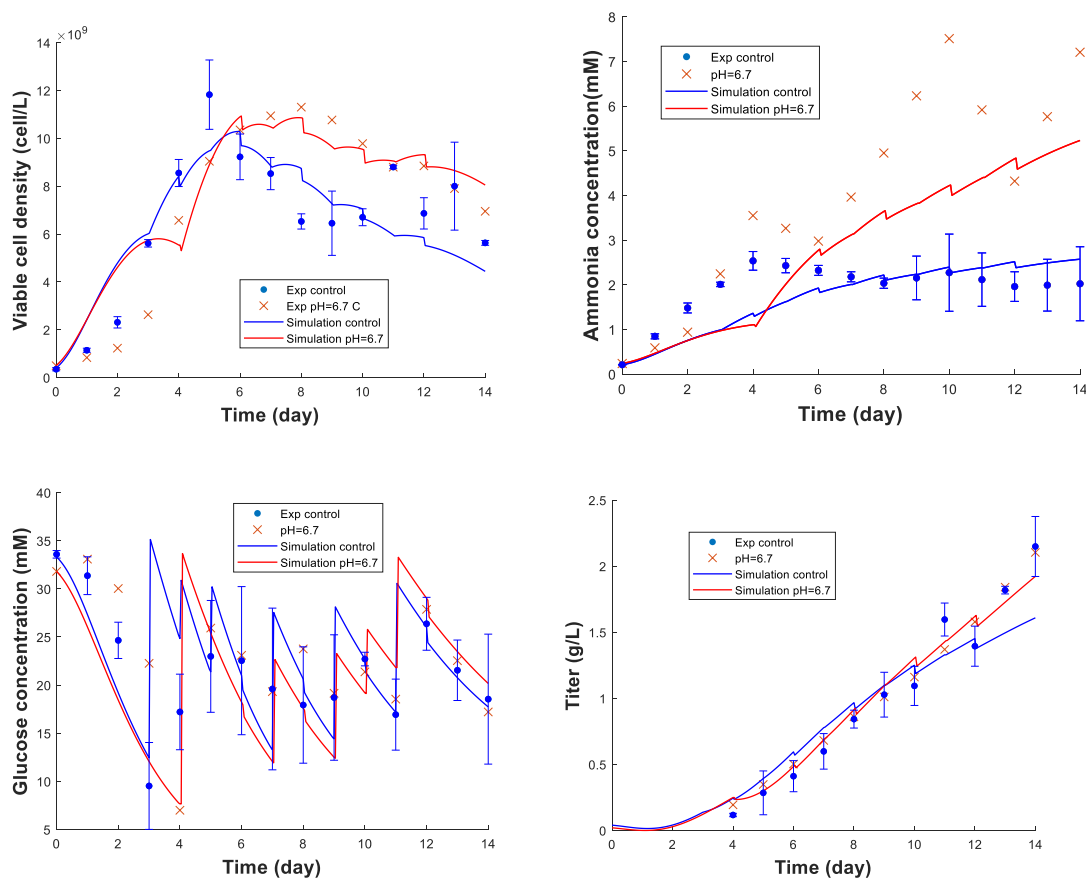
Figure 5.4 3 experimental data and simulation fitting results for cell culture components under control and temperature pH shift down.

5.4.2 Golgi N-linked Glycosylation

5.4.2.1 Glycan fractions

Experimental data shows that G0, G0F, G1F, G2F, Man5 and G1 are the major glycan fractions in the final products which represent the level of glycosylation by calculating the galactosylation, high mannose, and afucosylation index. To simulate reactions in the Golgi apparatus, the biochemical reaction pathways are simplified into the following branches in Figure 5.4 4 based on the important glycans that can be measured from experiments. ManI, ManII, GnTII, GalT, and FucT are enzymes. UDP-GlcNAc, GDP-Fuc, and UDP-Gal are major nucleotide sugars.
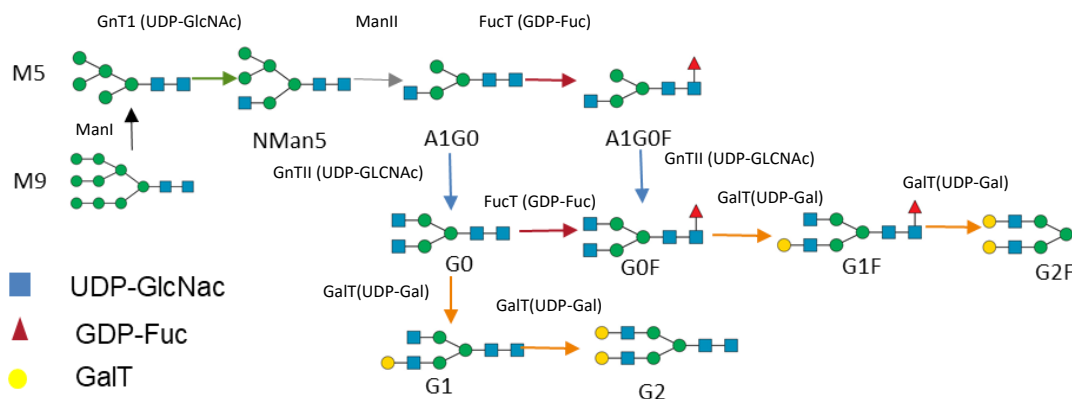
Figure 5.4 4 Simplified glycosylation pathway

Glycan fractions can only be measured starting from day 4 and day 6 onward due to the limited protein concentrations at the beginning of the cell culture experiment. Glycan fraction fittings for control, temperature and pH shifts are shown in Figure 5.4 5 to Figure 5.4 8. It indicates that during the cell culture, there is an increasing trend of Man5, G0F and G0 for all the runs independent from the temperature and pH shift. G1F and G2F have decreasing trends. G1 keeps constant for most of the time with an insignificant decreasing trend. This means that, from the stationary phase to the cell death phase, the level of glycosylation is reduced, especially for fucosylation. One possible explanation is that with the cell productivity increased during the cell death phase, protein does not have long enough time to stay inside the Golgi apparatus which reduces the time for protein post-translation. Another explanation is that the high mAb productivity consumes more energies insides the cell and reduces the enzyme expressions for N-linked glycosylation[148]. Mechanistic model can capture this trend as shown in the blue line for all the bioreactor runs. However, for the run with temperature shift down, as shown in Figure 5.4 5, the prediction is not accurate from day 6 to day 8. This is mainly because of the inaccurate model prediction for cell concentration from day 5 to day 7. It is also found that there is a high prediction error that exists on the fitting of pH shift.

With the replications of the control runs and temperature shift up runs, errors are added to the figure and R^2 values are calculated to evaluate the goodness of the fitting. All the fittings have $R^2 > 0.96$, which indicates the fittings are good.

$R^2 = 0.997$

$R^2 = 0.998$

$R^2 = 0.993$

$R^2 = 0.997$

$R^2 = 0.996$

$R^2 = 0.995$
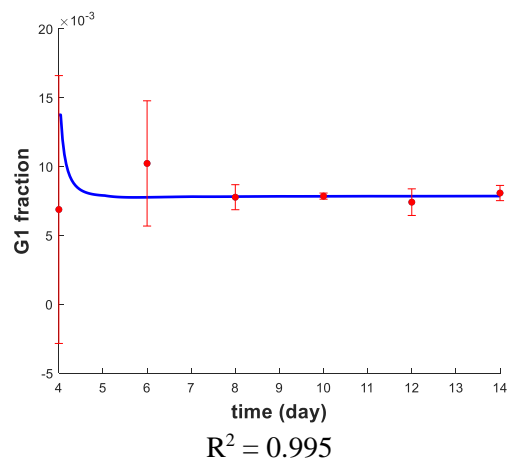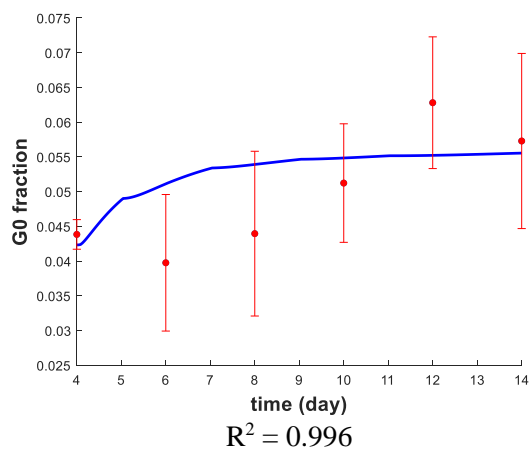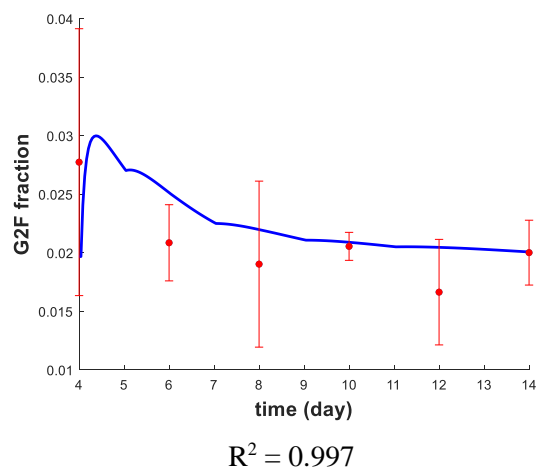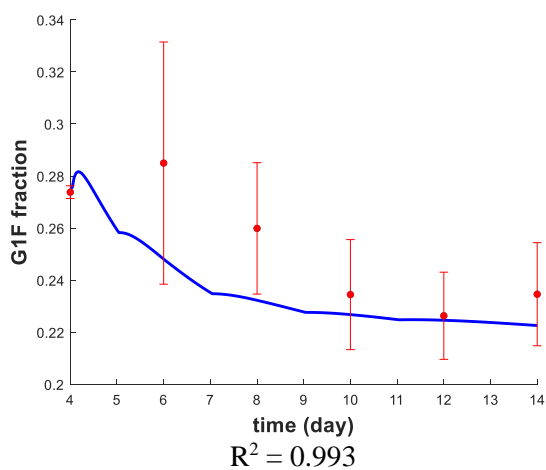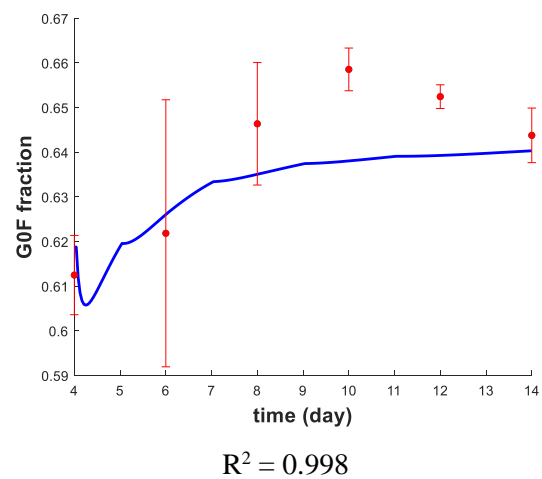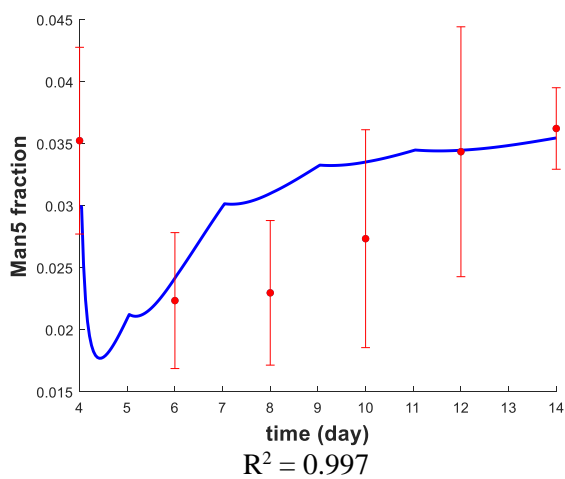
Figure 5.4 5 Glycosylation experimental data and simulation fitting results for control run

$R^2 = 0.996$

$R^2 = 0.966$

$R^2 = 0.975$

$R^2 = 0.996$

$R^2 = 0.999$

$R^2 = 0.997$
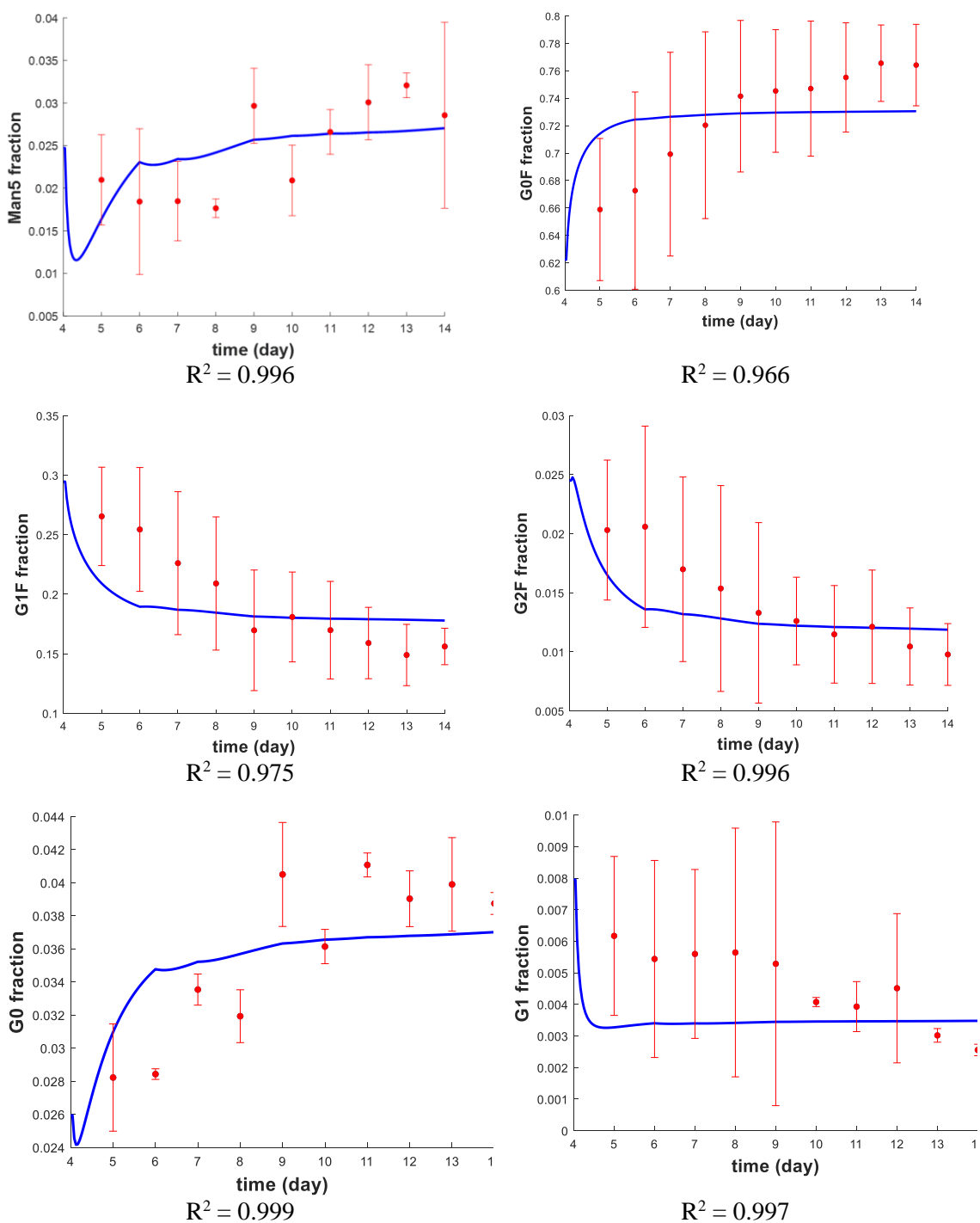
Figure 5.4 6 Glycosylation experimental data and simulation fitting results for temperature shift up
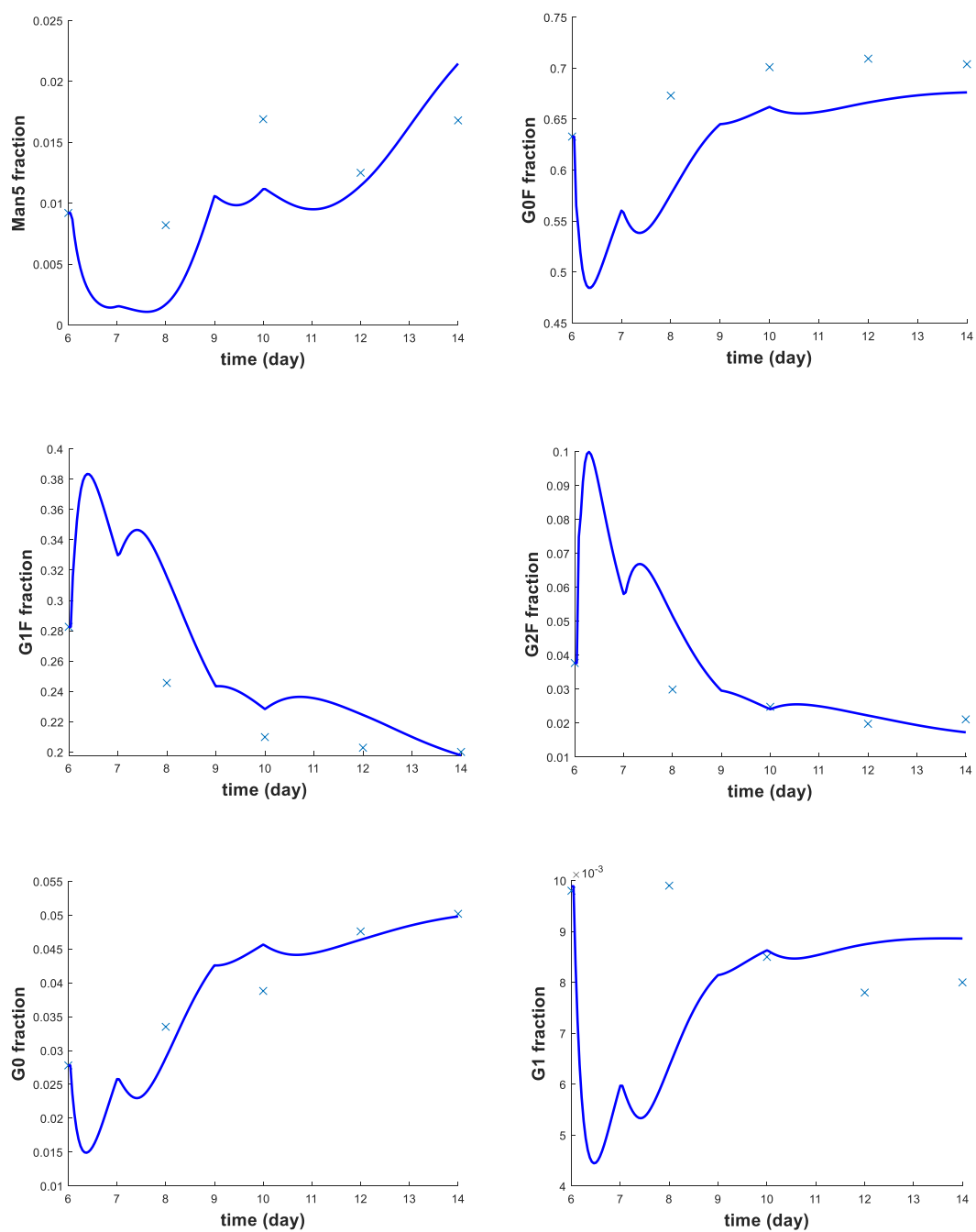
Figure 5.4 7 Glycosylation experimental data and simulation fitting results for temperature shift down
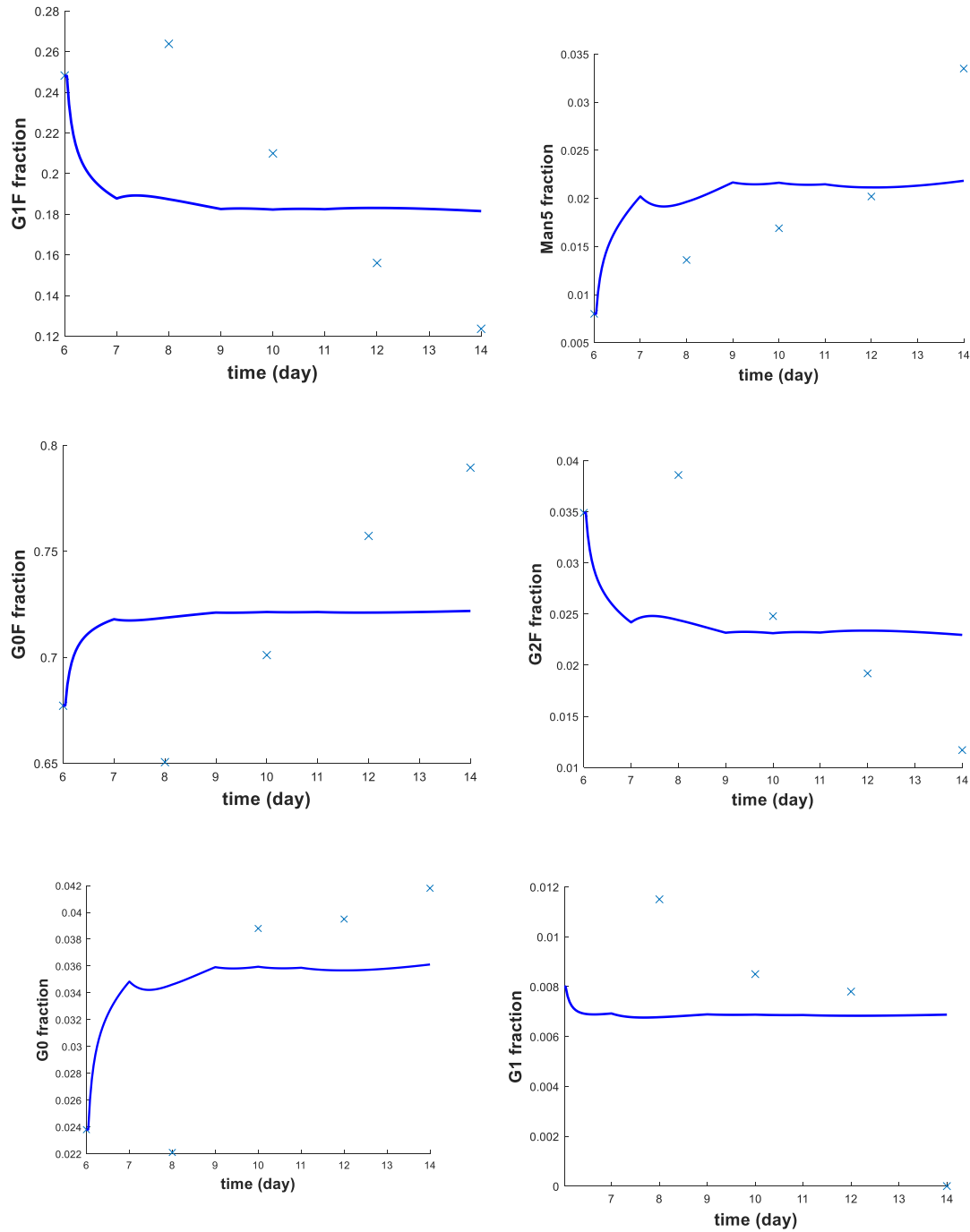
Figure 5.4 8 Glycosylation experimental data and simulation fitting results for pH shift down

However, the parameter fitting for another two runs cannot be evaluated using R^2 due to the lack of replications. The goodness of parameter fitting is further evaluated by calculating the root mean squared error (RMSE). For all the fittings under different conditions, RMSE are less than 0.03.

Model prediction vs experimental measured is plotted in Figure 5.4 9 indicating the good agreements between mechanistic model and experimental data for control and temperature shifts.

$T = 37\ °C,\ pH = 7.1$
RMSE=0.0138

$T = 39\ °C,\ pH = 7.1$
RMSE = 0.0278

$T = 35\ °C,\ pH = 7.1$
RMSE = 0.0258

$T = 37\ °C,\ pH = 6.7$
RMSE = 0.0274
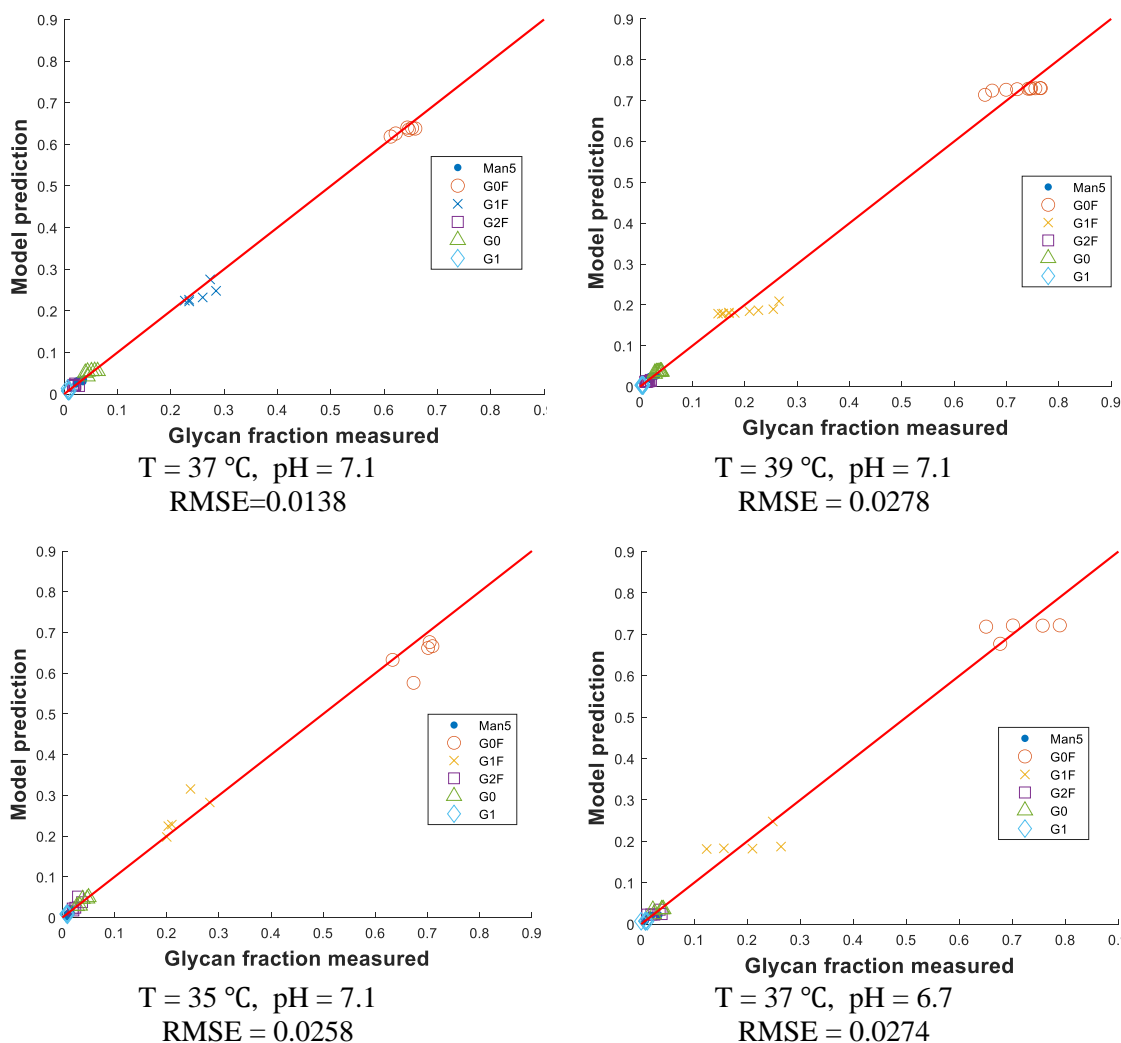
Figure 5.4 9 Comparison of experimentally measured glycan fractions to mechanistic model output

RMSE is relative to the scale of the data and is only powerful when comparing the fittings for the same sets of data by using different fitting methods. To test the model accuracy statistically, relative root means square error (RRMSE) is calculated [171]. Considering the measurement and biological replication errors from the experiment, the mean relative standard deviation (MRSD) of the experimental data are also obtained for the runs with replications. As shown from the table, for those fittings with large RRMSE, their experimental MRSD is also significant.

Table 5.4 1 RRMSE and MRSD calculation for control run and temperature shift up run

Condition	T = 37 ℃, pH = 7.1		T = 39 ℃, pH = 7.1	
Glycans	RRMSE	Exp MRSD	RRMSE	Exp MRSD
Man5	15.9 %	23.6 %	18.2 %	20.5 %
G0	14.3 %	18.4 %	9.6 %	4.6 %
G0F	1.8 %	1.74 %	4.4 %	7.14 %
G1	37.2 %	36.7 %	36.5 %	39.1 %
G1F	7.8 %	8.6 %	17.5 %	21.0 %
G2F	20.3 %	23.4 %	21.2 %	39.3 %

As it is known from R^2 calculation, control and temperature shift up runs show good accuracy of model fittings. Based on the MRSD calculations, a metric is obtained to evaluate the goodness of fittings for all the components, which is summarized in Table 5.4 2.

Table 5.4 2 RRMSE calculation and evaluation for temperature shift down and pH shift down run

Glycans	T = 35 ℃, pH = 7.1	T = 37 ℃, pH = 6.7	Acceptable range
Man5	34.8 %	33.9 %	< 30 %
G0	9.4 %	19.6 %	<20 %
G0F	7.6 %	6.5 %	<10 %
G1	19.2 %	53.4 %	<40 %
G1F	14.8 %	23.1 %	<20 %
G2F	37.2 %	32.3 %	<40 %

From the result, most of the glycan fittings for temperature shift down can be well captured, except for Man5. Although Man5 and G1 cannot be well captured when it has pH shift up, most of the components are within the acceptable range. From the cell culture data, it is found that ammonia concentration of bioreactor 4 is higher than that of other runs and is around 3-fold higher than that of control run in the cell stationary and death phase. High ammonia concentration would increase the intracellular pH, which will further reduce the enzyme activities. However, in the model ammonia effect has not been captured.

5.4.2.2   Glycan Indexes

In order to further understand the temperature and pH effect to protein quality attributes, glycan indexes are calculated and compared as shown in Figure 5.4 10 (experimental results) and Figure 5.4 11 (simulation results) which can indicate the relative relations of dynamic change of glycan indexes. The red line provides the upper and lower bound of the required glycan index for Herceptin

production [162]. It is obvious to see that afucosylation and high mannose are all within the range of the required target quality attributes. In addition, shifting down the temperature or reducing pH reduce the high mannose fraction. Adjusting temperature and reducing pH reduce the fraction of afucosylation. This reflects that temperature and pH shift increase the fucosylation. However, the level of galactosylation is below the lower bound. It shows that changing temperature and reducing pH all decrease the level of galactosylation. We received preliminary pH shifts up data from our collaborators and the results indicate that increasing pH would increase the level of glycosylation, which is interesting to investigate further.
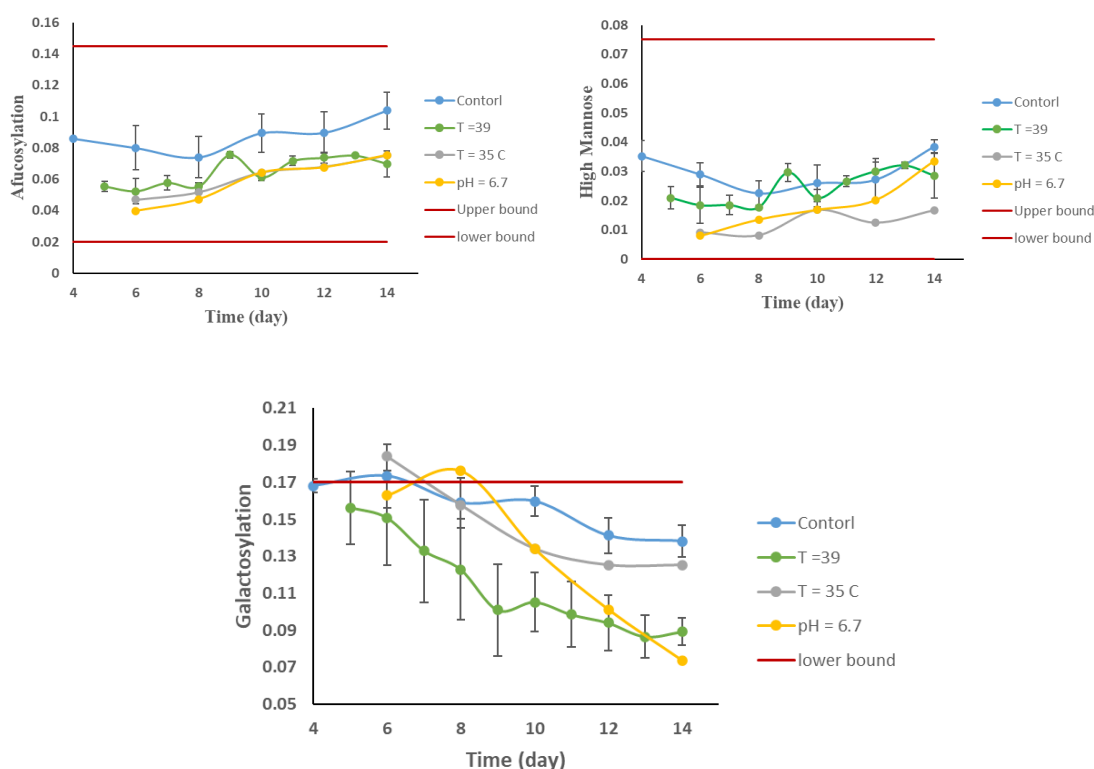
Figure 5.4 10 Comparison of experimental data for glycan Index under different conditions

Similar to Figure 5.4 10, the simulation results also correctly captured the temperature and pH effects on product quality attributes.
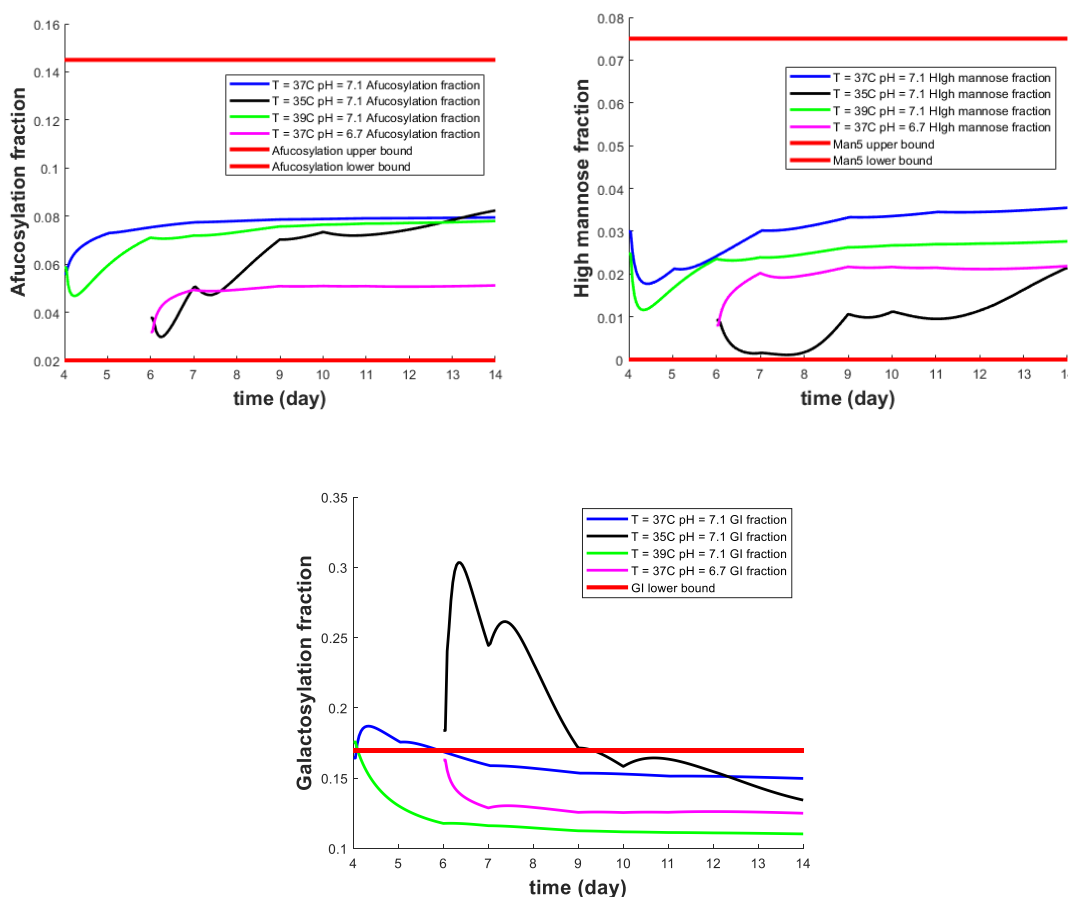
Figure 5.4 11 Comparison of simulation data for glycan Index under different conditions

One benefit of the mechanistic model is to predict intracellular components under the change of physical or chemical stimuli, interprets the effect of the glycosylation process and further guides the improvement of product quality. The normalized dissociation rates and nucleotide sugar concentrations estimated from mechanistic model are shown in the Table 5.4 3 below. Recap from equations 5.2.11 to 5.2.13, increasing the dissociation rate will decrease the enzymatic reaction rate. Reduction of the nucleotide sugars also reduces the production rate. It has found that reducing the temperature increases the dissociation rate of the galacosylation and reduces the formation of UDP-Gal., thus small amount of G1F and G2F are formed. When the temperature increases, the dissociation rate for ManII increased, which reduces the amount of glycan for later reactions. With the increase of the dissociation rate, the galactosylation of the product thus decreased. It also finds

that, with pH increasing, many of the enzyme dissociation rates increased. This might cause by the increase of ammonia concentration in the extracellular media.

Table 5.4 3 Kinetic constant simulated from mechanistic modeling

Conditions	$k_{d,\text{manII}}$	$k_{d,\text{GalT (G0)}}$	$k_{d,\text{GalT(G0F)}}$	$k_{d,\text{GalT(G1F)}}$	UDP-GlcNAc	UDP-Gal
T = 35 ℃	0.55	1.11	1.50	1.50	0.53	0.50
T = 37 ℃	1.00	1.00	1.00	1.00	1.00	1.00
T = 39 ℃	1.50	1.24	1.35	1.28	0.71	1.48

As currently none of the bioreactor runs discussed in this study provide product quality that satisfies the FDA required quality attributes. Further researches need to focus on the improvement of the galactosylation of the protein. For example, investigate the effect of ammonia, uridine, and Mn^+ additions during the cell culture. After that, since temperature shift increases the productivity of the protein production significantly, to further improve the level of protein galactosylation with reduced temperature, one approach is to increase the formation of UDP-Gal. It has been found that adding galactose and uridine possibly increases the UDP-Gal and further increases the galactosylation [36, 172].

5.5 Conclusion

In this study, a mechanistic kinetic model has been successfully developed to capture the temperature effect to fed-batch bioreactor cell growth, protein production and glycosylation. It shows that increasing and decreasing temperature during the cell culture decreases the protein galactosylation for different reasons. The increase of the temperature does not affect the formation of UDP-Gal but increases the dissociation rate of ManII, GalT, which further reduces the formation of posttranslational glycosylation process. Thus, to improve galactosylation, the enzyme expression and activities need to be investigated by maintaining ammonia concentrations within a certain level. However, reduction of the temperature reduces nucleotide sugar concentration as well as the dissociation rate of GalT. Increasing the nucleotide sugar by adding galactose or uridine during the cell culture is suggested.

For future work, replications need to be provided for temperature and pH shifting down, which can further improve the model fitting with the consideration of biological run uncertainties. DoE needs to be finished to test whether increasing pH has positive effects on increasing the level of galactosylation of the protein, which will help achieve the required quality attributes for our target product. Based on the replicated data, the most fitting should be reconducted for temperature and pH shifting down. Aspartate, lactate, and glutamate can be tested for each run and evaluate their effect on cell growth. If it is necessary, considering their effect on cell growth rate in the model would further improve the model accuracy

6   Summary and future work

In this dissertation, process modeling and system analysis methods have been used to improve the biopharmaceutical manufacturing process, specifically for monoclonal antibodies (mAbs) production. Two approaches are introduced, including plant decision-making and single unit operation optimization. Plant decision-making with new process design, existing process optimization, and novel process evaluation are achieved by flowsheet modeling. End-to-end continuous integrated processes are developed for different biologic drug production and compared by techno-economic analysis, which reveals the benefits of continuous operations. By applying deterministic and sensitivity cost analysis, different operating modes under multiple scales and process parameters are evaluated. Two case studies demonstrate the benefit of continuous manufacturing on a large-scale manufacturing operation. However, for small-scale production, the benefits vary from case to case.

Single unit operation mainly focuses on the upstream production bioreactor as it affects not only overall process productivity but also drug efficacy and potency. A framework is built to improve process understanding as well as finding an optimum design space to satisfy required productivity and quality. Mechanistic modeling is first used to correlate process parameters with cell growth and critical product quality attributes. Based on the mechanistic model experiment, a surrogate model is developed. The surrogate model requires a small amount of data but can provide a high accuracy on product quality prediction. Feasibility analysis is then implemented to the model for operating design space investigation. With the idea of the framework, the mechanistic model is further developed

based on experimental data from our collaborators and guides experimental product improvement.

Based on the research in this dissertation, a few recommendations on future research are listed below.

- For integrated flowsheet development, a more detailed continuous manfaucturing process with auxiliary equipment, buffer preparation tanks as well as online PAT system for continuous manufacturing can be further designed. Unlike batch process, the online sensing system would be critical to ensure the product quality and reduce the production waste caused by a failure in the continuous process.

- Continuous process can be further optimized by flowsheet model. As provided in Chapter 3, case study two, the batch process can be optimized based on the number of upstream bioreactors selected. The continuous process can also be optimized by the selection of multiple production lines or different upstream downstream arrangement.

- Although in recent years, different researches have been investigating components such as ammonia, galactose, uridine's effects on protein glycosylation but there is no research specifically on Herceptin production. Since different cell-lines will have different reactions to chemical and physical stimuli, it deserves to investigate their effect to the quality attributes improvement. Furthermore, as it is mentioned in the dissertation, increasing pH might increase galactosylation, thus it needs to be tested and confirmed by experimental work.

- With the multiple effects discovered, the model can be further developed to capture more components. To handle multiple parameters within the system, the hybrid

model is promising to use, which contains mechanistic model part to capture cell growth and glycosylation under controlled conditions and use a surrogate model to capture the effect of the parameters on the rate constants.

- Although the framework is built for upstream bioreactor modeling, the same framework can be applied to other unit operations such as chromatography and filtration. With the individual unit getting built, the integrated model that captures critical process parameters and critical quality attributes can be obtained. Thus, sensitivity analysis and feasibility analysis can be applied to these parameters to investigate design space for the overall process.

Acknowledgement of previous publications

Several chapters of this thesis have been published or being prepared for publications. The following publications are acknowledged.

Parts of Chapter 2 are being prepared under the citations

- Yang O., Qadan, M., Ierapetritou, M., "Economic analysis of Batch and Continuous Biopharmaceutical Antibody Production: A review." Journal of Pharmaceutical Innovation (2019).

- Chen, Y., Yang, O., Sampat, C., Bhalode, P., Ramachandran, R., Ierapetritouet, M., "Digital Twins in Pharmaceutical and Biopharmaceutical Manufacturing: A Review." Processes (2020).

- Chopda, V., Gyorgypal, A., Yang, O., Singh, R., Ramachandran, R., Zhang, H., Tsilomelekis, G., Chundawat, S., Ierapetritou, M. Recent Advances in Integrated Process Analytical Techniques, Modeling and Control Strategies to Enable Continuous Biomanufacturing of Monoclonal Antibodies.

Parts of Chapter 3 are being prepared under the citations

- Yang O., Prabhu, S., Ierapetritou, M., "A Comparison Between Batch and Continuous Monoclonal Antibody Production and Economic Analysis." Industrial & Engineering Chemistry Research (2019).

- Yang O., Tao, Y., Qadan, M., Ierapetritou, M., "Process Design and Comparison for Batch and Continuous Recombinant Adeno-Associated Virus."

Chapter 4 is published under the citations

- Yang, O., Ierapetritou, M., "mAb Production Modeling and Design Space Evaluation Including Glycosylation Process." Processes (2021)

Chapter 5 is being prepared under the citations

- Yang, O., Chopda, V., Gyorgypal, A., Singh, R., Chundawat, S., Ramachandran, R., Zhang, H., Tsilomelekis, G., Ierapetritou, M. "Investigate Mechanistic of Temperature and pH Effects on N-linked Glycosylation in Herceptin Biosimilar Production."

Bibliography

1. Lin-Gibson, S. and V. Srinivasan, *Recent Industrial Roadmaps to Enable Smart Manufacturing of Biopharmaceuticals.* IEEE Transactions on Automation Science and Engineering, 2019: p. 1-8.

2. Mullin, R., *Contract services for biologics evolve.* Chemical & Engineering News, 2017. **95**(33): p. 34-35.

3. Gonçalves, G.A.R. and R.d.M.A. Paiva, *Gene therapy: advances, challenges and perspectives.* Einstein (Sao Paulo, Brazil), 2017. **15**(3): p. 369-375.

4. *Long term follow-up after administration of human gene therapy products* F.a.D.A. U.S. Departmetn of Health and Human Services, Center for Biologics evaluation and reserach Editor. 2020.

5. Ronald A. Rader, E.S.L. *Biopharma Market: An Inside Look*. 2018; Available from: https://www.pharmamanufacturing.com/articles/2018/biopharma-market-an-inside-look/#:~:text=The%20biopharma%20sector%20has%20seen,of%20the%20total%20pharmaceutical%20market.

6. Narayanan, H., et al., *Bioprocessing in the Digital Age: The Role of Process Models.* Biotechnol J, 2020. **15**(1): p. e1900172.

7. Fisher, A.C., et al., *Advancing pharmaceutical quality: An overview of science and research in the U.S. FDA's Office of Pharmaceutical Quality.* Int J Pharm, 2016. **515**(1-2): p. 390-402.

8. Konstantinov, K.B. and C.L. Cooney, *White paper on continuous bioprocessing. May 20–21, 2014 continuous manufacturing symposium.* Journal of Pharmaceutical Sciences, 2015. **104**(3): p. 813-820.

9. Lee, S.l. *Present and Future for Continuuous Manufacturing FDA Perspective* in *3rd FDA/PQRI Conference on Advancing Product Quality* 2017.

10. Badr, S. and H. Sugiyama, *A PSE perspective for the efficient production of monoclonal antibodies: integration of process, cell, and product design aspects.* Current Opinion in Chemical Engineering, 2020. **27**: p. 121-128.

11. *ICH Harmonised tripartite guideline Pharmaceutical Development Q8 (R2)* I. Process, Editor. 2009.

12. Alt, N., et al., *Determination of critical quality attributes for monoclonal antibodies using quality by design principles.* Biologicals, 2016. **44**(5): p. 291-305.

13. Kiyoshi, M., et al., *Assessing the Heterogeneity of the Fc-Glycan of a Therapeutic Antibody Using an engineered FcγReceptor IIIa-Immobilized Column.* Scientific Reports, 2018. **8**(1): p. 3955.

14. del Val, I.J., C. Kontoravdi, and J.M. Nagy, *Towards the implementation of quality by design to the production of therapeutic monoclonal antibodies with desired glycosylation patterns.* Biotechnol Prog, 2010. **26**(6): p. 1505-27.

15. Ehret, J., et al., *Impact of cell culture media additives on IgG glycosylation produced in Chinese hamster ovary cells.* Biotechnol Bioeng, 2019. **116**(4): p. 816-830.

16. Lodish H, B.A., Zipursky SL, et al, *Protein Glycosylation in the ER and Golgi Complex*, in *Molecular Cell Biology*. 2000, New York: W. H. Freeman: New York: W. H. Freeman.

17. Higel, F., et al., *N-glycosylation heterogeneity and the influence on structure, function and pharmacokinetics of monoclonal antibodies and Fc fusion proteins.* Eur J Pharm Biopharm, 2016. **100**: p. 94-100.

18. von Stosch, M., J.-M. Hamelink, and R. Oliveira, *Hybrid modeling as a QbD/PAT tool in process development: an industrial E. coli case study.* Bioprocess and biosystems engineering, 2016. **39**(5): p. 773-784.

19. Phillip Elliott, S.B., Jingxu Bi, Hu Zhang, *Quality by design for biopharmaceuticals a historical review and guide for implementation.* Pharmaceutical Bioprocessing 2013. **1**(1).

20. Walther, J., et al., *The business impact of an integrated continuous biomanufacturing platform for recombinant protein production.* Journal of Biotechnology, 2015. **213**: p. 3-12.

21. Arnold, L., et al., *Implementation of Fully Integrated Continuous Antibody Processing: Effects on Productivity and COGm.* Biotechnology Journal, 2019. **14**(2): p. 1800061.

22. Yang, O., S. Prabhu, and M. Ierapetritou, *Comparison between Batch and Continuous Monoclonal Antibody Production and Economic Analysis.* Industrial & Engineering Chemistry Research, 2019. **58**(15): p. 5851-5863.

23. Pollock, J., et al., *Integrated Continuous Bioprocessing: Economic, Operational, and Environmental Feasibility for Clinical and Commercial Antibody Manufacture.* Biotechnology Progress, 2017. **33**(4): p. 854-866.

24. Sencar, J., N. Hammerschmidt, and A. Jungbauer, *Modeling the Residence Time Distribution of Integrated Continuous Bioprocesses.* Biotechnol J, 2020: p. e2000008.

25. Liu, S., S.S. Farid, and L.G. Papageorgiou, *Integrated Optimization of Upstream and Downstream Processing in Biopharmaceutical Manufacturing under Uncertainty: A Chance Constrained Programming Approach.* Industrial & Engineering Chemistry Research, 2016. **55**(16): p. 4599-4612.

26. Gomis-Fons, J., et al., *Model-based design and control of a small-scale integrated continuous end-to-end mAb platform.* Biotechnology Progress, 2020. **n/a**(n/a): p. e2995.

27. Sachidananda, M., et al., *Discrete Event Simulation Modelling for Dynamic Decision Making in Biopharmaceutical Manufacturing.* Procedia CIRP, 2016. **49**: p. 39-44.

28. Pleitt, K., et al., *Evaluation of process simulation as a decisional tool for biopharmaceutical contract development and manufacturing organizations.* Biochemical Engineering Journal, 2019. **150**: p. 107252.

29. Zahel, T., et al., *Integrated Process Modeling-A Process Validation Life Cycle Companion.* Bioengineering (Basel), 2017. **4**(4).

30. Sandu Taraş, A.W., *Simulation and Multi-Objective Optimization of Bioprocesses with Matlab and SuperPro Designer using a Client-Server Interface.* Chemical Engineering Transactions, 2011(25): p. 207-212.

31. Tang, P., et al., *Kinetic modeling of Chinese hamster ovary cell culture: factors and principles.* Crit Rev Biotechnol, 2020. **40**(2): p. 265-281.

32. Xu, J., et al., *Systematic development of temperature shift strategies for Chinese hamster ovary cells based on short duration cultures and kinetic modeling.* MAbs, 2019. **11**(1): p. 191-204.

33. Hogiri, T., et al., *Optimization of a pH-shift control strategy for producing monoclonal antibodies in Chinese hamster ovary cell cultures using a pH-dependent dynamic model.* J Biosci Bioeng, 2018. **125**(2): p. 245-250.

34. Sou, S.N., et al., *Model-based investigation of intracellular processes determining antibody Fc-glycosylation under mild hypothermia.* Biotechnol Bioeng, 2017. **114**(7): p. 1570-1582.

35. **32**(5): p. 1135-1148.

36. Kotidis, P., et al., *Model-based optimization of antibody galactosylation in CHO cell culture.* 2019. **116**(7): p. 1612-1626.

37. Radhakrishnan, D., A.S. Robinson, and B.A. Ogunnaike, *Controlling the Glycosylation Profile in mAbs Using Time-Dependent Media Supplementation.* Antibodies (Basel, Switzerland), 2017. **7**(1): p. 1.

38. Karst, D.J., et al., *Modulation and modeling of monoclonal antibody N-linked glycosylation in mammalian cell perfusion reactors.* Biotechnology and Bioengineering, 2017. **114**(9): p. 1978-1990.

39. Shirahata, H., et al., *Dynamic modelling, simulation and economic evaluation of two CHO cell-based production modes towards developing biopharmaceutical manufacturing processes.* Chemical Engineering Research and Design, 2019. **150**: p. 218-233.

40. Xing, Z., et al., *Optimizing amino acid composition of CHO cell culture media for a fusion protein production.* Process Biochemistry, 2011. **46**(7): p. 1423-1429.

41. Spahn, P.N., et al., *A Markov chain model for N-linked protein glycosylation – towards a low-parameter tool for model-driven glycoengineering.* Metabolic Engineering, 2016. **33**: p. 52-66.

42. Hutter, S., et al., *Glycosylation flux analysis reveals dynamic changes of intracellular glycosylation flux distribution in Chinese hamster ovary fed-batch cultures.* Metabolic Engineering, 2017. **43**: p. 9-20.

43. Nolan, R.P. and K. Lee, *Dynamic model of CHO cell metabolism.* Metabolic Engineering, 2011. **13**(1): p. 108-124.

44. Bayrak, E.S., et al., *Computational Modeling of Fed-Batch Cell Culture Bioreactor: Hybrid Agent-Based Approach.* IFAC-PapersOnLine, 2015. **48**(8): p. 1252-1257.

45. Li, X., et al., *Development of a Computational Fluid Dynamics Model for Scaling-up Ambr Bioreactors.* Biotechnology and Bioprocess Engineering, 2018. **23**(6): p. 710-725.

46. Farzan, P. and M.G. Ierapetritou, *A Framework for the Development of Integrated and Computationally Feasible Models of Large-Scale Mammalian Cell Bioreactors.* 2018. **6**(7): p. 82.

47. Menshutina, N.V., et al., *Modelling of hollow fiber membrane bioreactor for mammalian cell cultivation using computational hydrodynamics.* Bioprocess and Biosystems Engineering, 2020. **43**(3): p. 549-567.

48. Selvarasu, S., et al., *Combined data preprocessing and multivariate statistical analysis characterizes fed-batch culture of mouse hybridoma cells for rational medium design.* Journal of Biotechnology, 2010. **150**(1): p. 94-100.

49. Sokolov, M., et al., *Sequential Multivariate Cell Culture Modeling at Multiple Scales Supports Systematic Shaping of a Monoclonal Antibody Toward a Quality Target.* Biotechnology Journal, 2018. **13**(4): p. 1700461.

50. Zürcher, P., et al., *Cell culture process metabolomics together with multivariate data analysis tools opens new routes for bioprocess development and glycosylation prediction.* Biotechnology Progress, 2020. **n/a**(n/a): p. e3012.

51. Narayanan, H., et al., *A new generation of predictive models: The added value of hybrid models for manufacturing processes of therapeutic proteins.* Biotechnol Bioeng, 2019. **116**(10): p. 2540-2549.

52. Psichogios, D.C. and L.H. Ungar, *A hybrid neural network-first principles approach to process modeling.* AIChE Journal, 1992. **38**(10): p. 1499-1511.

53. Zalai, D., et al., *Combining mechanistic and data-driven approaches to gain process knowledge on the control of the metabolic shift to lactate uptake in a fed-batch CHO process.* Biotechnology Progress, 2015. **31**(6): p. 1657-1668.

54. Borys, B.S., et al., *Scale-up of embryonic stem cell aggregate stirred suspension bioreactor culture enabled by computational fluid dynamics modeling.* Biochemical Engineering Journal, 2018. **133**: p. 157-167.

55. Agarwal, N., et al., *Kinetic modeling as a tool to understand the influence of cell culture process parameters on the glycation of monoclonal antibody biotherapeutics.* Biotechnol Prog, 2019. **35**(5): p. e2865.

56. Sha, S., et al., *Mechanistic modeling and applications for CHO cell culture development and production.* Current Opinion in Chemical Engineering, 2018. **22**: p. 54-61.

57. Shirahata, H., M. Hirao, and H. Sugiyama, *Multiobjective decision-support tools for the choice between single-use and multi-use technologies in sterile filling of biopharmaceuticals.* Computers & Chemical Engineering, 2019. **122**: p. 114-128.

58. Lim, A.C., et al., *A computer-aided approach to compare the production economics of fed-batch and perfusion culture under uncertainty.* Biotechnol Bioeng, 2006. **93**(4): p. 687-97.

59. Pollock, J., S.V. Ho, and S.S. Farid, *Fed-batch and perfusion culture processes: Economic, environmental, and operational feasibility under uncertainty.* Biotechnology and Bioengineering, 2013. **110**(1): p. 206-219.

60. Pollock, J., et al., *Optimising the design and operation of semi-continuous affinity chromatography for clinical and commercial manufacture.* Journal of Chromatography A, 2013. **1284**: p. 17-27.

61. Liu, S., et al., *Designing cost-effective biopharmaceutical facilities using mixed-integer optimization.* Biotechnol Prog, 2013. **29**(6): p. 1472-83.

62. Hammerschmidt, N., et al., *Economics of recombinant antibody production processes at various scales: Industry-standard compared to continuous precipitation.* Biotechnol J, 2014. **9**(6): p. 766-775.

63. Xenopoulos, A., *A new, integrated, continuous purification process template for monoclonal antibodies: Process modeling and cost of goods studies.* J Biotechnol, 2015. **213**: p. 42-53.

64. Walther, J., et al., *The business impact of an integrated continuous biomanufacturing platform for recombinant protein production.* J Biotechnol, 2015. **213**: p. 3-12.

65. Li, Y.F. and V. Venkatasubramanian, *Integrating Design of Experiments and Principal Component Analysis to Reduce Downstream Cost of Goods in Monoclonal Antibody Production.* Journal of Pharmaceutical Innovation, 2016. **11**(4): p. 352-361.

66. Klutz, S., et al., *Cost evaluation of antibody production processes in different operation modes.* Chemical Engineering Science, 2016. **141**: p. 63-74.

67. Bunnak, P., et al., *Life-cycle and cost of goods assessment of fed-batch and perfusion-based manufacturing processes for mAbs.* Biotechnol Prog, 2016. **32**(5): p. 1324-1335.

68. Torres-Acosta, M.A., et al., *Economic analysis of uricase production under uncertainty: Contrast of chromatographic purification and aqueous two-phase extraction (with and without PEG recycle).* Biotechnol Prog, 2016. **32**(1): p. 126-33.

69. Xu, S., et al., *Bioreactor productivity and media cost comparison for different intensified cell culture processes.* Biotechnol Prog, 2016.

70. Pollock, J., et al., *Integrated continuous bioprocessing: Economic, operational, and environmental feasibility for clinical and commercial antibody manufacture.* Biotechnol Prog, 2017.

71. Arnold, L., et al., *Implementation of Fully Integrated Continuous Antibody Processing: Effects on Productivity and COGm.* Biotechnol J, 2018.

72. Hummel, J., et al., *Modeling the Downstream Processing of Monoclonal Antibodies Reveals Cost Advantages for Continuous Methods for a Broad Range of Manufacturing Scales.* Biotechnology Journal, 2019. **14**(2): p. 1700665.

73. Cataldo, A.L., et al., *Economics and ecology: Modelling of continuous primary recovery and capture scenarios for recombinant antibody production.* Journal of Biotechnology, 2020. **308**: p. 87-95.

74. Gupta, P., et al., *Economic assessment of continuous processing for manufacturing of biotherapeutics.* Biotechnol Prog, 2020: p. e3108.

75. *The biopharmaceutical industry's leading process analysis and economic modelling package*. BioSolve Process 7 2016 [cited 2017 8/9]; Available from: https://biopharmservices.com/wp-content/uploads/2016/02/Process-2pp-A4-2016-2-FINAL.pdf.

76. Ashouri, P., *A dynamic simulation framework for biopharmaceutical capacity management*, in *Department of Biochemical Engineering*. 2011, University College London. p. 259.

77. Petrides, D., *Bioprocess design and economics*, in *Bioseparations Science and Engineering* P.W.T. Roger G. Harrison, Scott R. Rudge and Demetri P. Petrides, Editor. 2015.

78. Pannell, D.J., *Sensitivity analysis: strategies, methods, concepts, examples.* Agric Econ, 1997. **16**: p. 139-152.

79. Torres-Acosta, M.A., et al., *Economic analysis of Royalactin production under uncertainty: Evaluating the effect of parameter optimization.* Biotechnol Prog, 2015. **31**(3): p. 744-9.

80. Vermasvuori, R. and M. Hurme, *Economic comparison of diagnostic antibody production in perfusion stirred tank and in hollow fiber bioreactor processes.* Biotechnol Prog, 2011. **27**(6): p. 1588-98.

81. Roque, A.C.A., C.R. Lowe, and M.Â. Taipa, *Antibodies and genetically engineered related molecules: production and purification.* Biotechnol prog, 2004. **20**(3): p. 639-654.

82. Azevedo, A.M., et al., *Chromatography-free recovery of biopharmaceuticals through aqueous two-phase processing.* Trends in Biotechnology, 2009. **27**(4): p. 240-247.

83. Farid, S.S., *Economic drivers and trade-offs in antibody purification processes.* BioPharm International, 2009. **2009**(7).

84. Espitia-Saloma, E., et al., *An integrated practical implementation of continuous aqueous two-phase systems for the recovery of human IgG: From the microdevice to a multistage bench-scale mixer-settler device.* Biotechnol J, 2016. **11**(5): p. 708-16.

85. Torres-Acosta, M.A., et al., *Economic evaluation of the primary recovery of tetracycline with traditional and novel aqueous two-phase systems.* Sep Purif Technol, 2018. **203**: p. 178-184.

86. *Guidance for Industry Q8 Pharmaceutical Development* U.S.D.o.H.a.H.S.F.a.D.A.C.f.D.E.a.R.C.f.B.E.a. Reserach, Editor. 2006.

87. Chatterjee, S. *Design Space considerations* 2012; Available from: https://www.fda.gov/media/85336/download.

88. Eon-Duval, A., et al., *Application of the quality by design approach to the drug substance manufacturing process of an Fc fusion protein: Towards a global multi-step design space.* Journal of Pharmaceutical Sciences, 2012. **101**(10): p. 3604-3618.

89. Abu-Absi, S.F., et al., *Defining process design space for monoclonal antibody cell culture.* Biotechnology and Bioengineering, 2010. **106**(6): p. 894-905.

90. Bhatia, H., et al., *A design space exploration for control of Critical Quality Attributes of mAb.* International Journal of Pharmaceutics, 2016. **512**(1): p. 242-252.

91. Kim, T.K., et al., *Applying the Quality by Design to Robust Optimization and Design Space Define for Erythropoietin Cell Culture Process.* Bulletin of the Korean Chemical Society, 2019. **40**(10): p. 1002-1012.

92. Mercier, S.M., et al., *Multivariate PAT solutions for biopharmaceutical cultivation: current progress and limitations.* Trends in Biotechnology, 2014. **32**(6): p. 329-336.

93.   Nagashima, H., et al., *Application of a Quality by Design Approach to the Cell Culture Process of Monoclonal Antibody Production, Resulting in the Establishment of a Design Space.* Journal of Pharmaceutical Sciences, 2013. **102**(12): p. 4274-4283.

94.   Jiang, C., et al., *Defining process design space for a hydrophobic interaction chromatography (HIC) purification step: Application of quality by design (QbD) principles.* 2010. **107**(6): p. 985-997.

95.   Vogg, S., T. Müller-Späth, and M. Morbidelli, *Design space and robustness analysis of batch and counter-current frontal chromatography processes for the removal of antibody aggregates.* Journal of Chromatography A, 2020. **1619**: p. 460943.

96.   Pirrung, S.M., et al., *Optimization of biopharmaceutical downstream processes supported by mechanistic models and artificial neural networks.* 2017. **33**(3): p. 696-707.

97.   Baur, D., et al., *Optimal model-based design of the twin-column CaptureSMB process improves capacity utilization and productivity in protein A affinity capture.* Biotechnology Journal, 2016. **11**(1): p. 135-145.

98.   Kiparissides, A., E.N. Pistikopoulos, and A. Mantalaris, *On the model-based optimization of secreting mammalian cell (GS-NS0) cultures.* Biotechnology and Bioengineering, 2015. **112**(3): p. 536-548.

99.   Kotidis, P., et al., *Constrained global sensitivity analysis for bioprocess design space identification.* Computers & Chemical Engineering, 2019. **125**: p. 558-568.

100.   Jagschies, G., et al., *Biopharmaceutical Processing: Development, Design, and Implementation of Manufacturing Processes*. 2018: Elsevier.

101.   Gosling, L. *Bioprocess simulation: taking the error out of trial-and-error.* [cited 2018 Sep 11]; Available from: http://www.iptonline.com/articles/public/chemsimllc.pdf.

102.   Newswire, P.R., *Monoclonal Antibodies (mAbs) Market Analysis By Source (Chimeric, Murine, Humanized, Human), By Type of Production, By Indication (Cancer, Autoimmune, Inflammatory, Infectious, Microbial, Viral Diseases), By End-use (Hospitals , Research, Academic Institutes, Clinics, Diagnostic Laboratories) And Segment Forecasts, 2018- 2024.* 2017, Y: Grand View Research, Market Research Report

103.   Dimitrov, D.S., *Therapeutic Proteins*, in *Therapeutic Proteins: Methods and Protocols*, V. Voynov and J.A. Caravella, Editors. 2012, Humana Press: Totowa, NJ. p. 1-26.

104.   *Monoclonal Antibodies Approved by the EMA and FDA for Therapeutic use (status 2017)* 2017 May 18, 2017 [cited 2018 July 31st]; Available from: http://www.actip.org/products/monoclonal-antibodies-approved-by-the-ema-and-fda-for-therapeutic-use/.

105.   Li, F., et al., *Cell culture processes for monoclonal antibody production.* mAbs, 2010. **2**(5): p. 466-477.

106.   Bielser, J.M., et al., *Perfusion mammalian cell culture for recombinant protein manufacturing - A critical review.* Biotechnol Adv, 2018. **36**(4): p. 1328-1340.

107.   Goudar, C.T., et al., *Logistic equations effectively model Mammalian cell batch and fed-batch kinetics by logically constraining the fit.* Biotechnol Prog, 2005. **21**(4): p. 1109-18.

108.   Goudar, C.T., *Computer programs for modeling mammalian cell batch and fed-batch cultures using logistic equations.* Cytotechnology, 2012. **64**(4): p. 465-475.

109.   Xu, S., et al., *Bioreactor productivity and media cost comparison for different intensified cell culture processes.* Biotechnol Prog, 2017. **33**(4): p. 867-878.

110.   Warikoo, V., et al., *Integrated continuous production of recombinant therapeutic proteins.* Biotechnol Bioeng, 2012. **109**(12): p. 3018-29.

111.   Chotteau, V., Y. Zhang, and M.F. Clincke, *Very High Cell Density in Perfusion of CHO Cells by ATF, TFF, Wave Bioreactor, and/or CellTank Technologies - Impact of Cell Density and Applications.* Continuous Processing in Pharmaceutical Manufacturing, 2015: p. 339-356.

112. Gronemeyer, P., R. Ditz, and J. Strube, *Trends in upstream and downstream process development for antibody manufacturing.* Bioengineering, 2014. **1**(4): p. 188-212.

113. Karimi, I.A. and G.V. Reklaitis, *Optimal Selection of Intermediate Storage Tank Capacity in a Periodic Batch Semicontinuous Process.* Aiche Journal, 1983. **29**(4): p. 588-596.

114. Pollock, J., et al., *Integrated continuous bioprocessing: Economic, operational, and environmental feasibility for clinical and commercial antibody manufacture.* Biotechnol Prog, 2017. **33**(4): p. 854-866.

115. Klutz, S., et al., *Developing the biofacility of the future based on continuous processing and single-use technology.* Journal of Biotechnology, 2015. **213**: p. 120-130.

116. Walther, J., et al., *Perfusion Cell Culture Decreases Process and Product Heterogeneity in a Head-to-Head Comparison With Fed-Batch.* Biotechnol J, 2018: p. e1700733.

117. Zydney, A.L., *Continuous downstream processing for high value biological products: A Review.* Biotechnology and Bioengineering, 2016. **113**(3): p. 465-475.

118. Casey, C., et al., *Protein concentration with single-pass tangential flow filtration (SPTFF).* Journal of Membrane Science, 2011. **384**(1-2): p. 82-88.

119. Hernandez, I., et al., *Pricing of Monoclonal Antibody Therapies: Higher If Used for Cancer?* American Journal of Managed Care, 2018. **24**(2): p. 109-+.

120. Rajendra, Y., R.B. Peery, and G.C. Barnard, *Generation of stable Chinese hamster ovary pools yielding antibody titers of up to 7.6 g/L using the piggyBac transposon system.* Biotechnol Prog, 2016. **32**(5): p. 1301-1307.

121. Huang, Y.-M., et al., *Maximizing productivity of CHO cell-based fed-batch culture using chemically defined media conditions and typical manufacturing equipment.* Biotechnology Progress, 2010. **26**(5): p. 1400-1410.

122. Yang, W.C., et al., *Perfusion seed cultures improve biopharmaceutical fed-batch production capacity and product quality.* Biotechnology Progress, 2014. **30**(3): p. 616-625.

123. FDA. *Advancing Regulatory Science at FDA: A Strategic Plan, August 2011.* [cited 2018 Dec 9th]; Available from: https://www.fda.gov/downloads/ScienceResearch/SpecialTopics/RegulatoryScience/UCM268225.pdf.

124. Farid, S.S., B. Thompson, and A. Davidson, *Continuous bioprocessing: The real thing this time? 10th Annual bioProcessUK Conference, December 3-4, 2013, London, UK.* Mabs, 2014. **6**(6): p. 1357-1361.

125. Vogg, S., T. Müller-Späth, and M. Morbidelli, *Current status and future challenges in continuous biochromatography.* Current Opinion in Chemical Engineering, 2018. **22**: p. 138-144.

126. Zydney, A.L., *Perspectives on integrated continuous bioprocessing - opportunities and challenges.* Current Opinion in Chemical Engineering, 2015. **10**: p. 8-13.

127. Ma, C.-C., et al., *The approved gene therapy drugs worldwide: from 1998 to 2019.* Biotechnology Advances, 2020. **40**: p. 107502.

128. FDA. *Approved Cellular and Gene Therapy Products.* 2021 [cited 2021 3/17]; Available from: https://www.fda.gov/vaccines-blood-biologics/cellular-gene-therapy-products/approved-cellular-and-gene-therapy-products.

129. *Gene Therapy Market by Vectors [Non-viral (Oligonucleotides), Viral (Retroviral (Gammaretroviral, Lentiviral)), Adeno-associated], Indication (Cancer, Neurological Diseases), Delivery Method (In Vivo, Ex Vivo), Region - Global Forecast to 2024.* 2018 [cited 2021 3/17]; Available from: https://www.marketsandmarkets.com/Market-Reports/gene-therapy-market-122857962.html?gclid=Cj0KCQjw0caCBhCIARIsAGAfuMzdbr5WyZB9ibIkLu-nQxsfQ5p58KeZzRfnDMv8TQlJmcTCb5i4V00aAt6lEALw_wcB.

130. Smith, D.C., *AAV Vector Manufacturing – Challenges & Opportunities in the Manufacturing of AAV Vectors Used in the Delivery of Gene Therapy Treatments.* Drug Development and Delivery 2017.

131. Grieger, J.C., S.M. Soltys, and R.J. Samulski, *Production of Recombinant Adeno-associated Virus Vectors Using Suspension HEK293 Cells and Continuous Harvest of Vector From the Culture Media for GMP FIX and FLT1 Clinical Vector.* Molecular therapy : the journal of the American Society of Gene Therapy, 2016. **24**(2): p. 287-297.

132. Cytiva. *Process economic simulation for scalable production of adenovirus*. 2020 [cited 2021 3/18]; Available from: https://cdn.cytivalifesciences.com/dmm3bwsv3/AssetStream.aspx?mediaformatid=10061&destinationid=10016&assetid=27022.

133. Henry, O., et al., *Insights into adenoviral vector production kinetics in acoustic filter-based perfusion cultures.* Biotechnol Bioeng, 2004. **86**(7): p. 765-74.

134. Silva, R.J.S., et al., *Continuous Chromatography Purification of Virus-Based Biopharmaceuticals: A Shortcut Design Method.* Methods Mol Biol, 2020. **2095**: p. 367-384.

135. Lodish H, B.A., Zipursky SL, et al., *Molecular Cell Biology*. Protein Glycosylation in the ER and Golgi Complex. 2000, New York: W. H. Freeman.

136. Aghamohseni, H., et al., *A semi-empirical glycosylation model of a camelid monoclonal antibody under hypothermia cell culture conditions.* Journal of Industrial Microbiology & Biotechnology, 2017. **44**(7): p. 1005-1020.

137. Kim, S.-M., K.-H. Chang, and D.J. Oh, *Effect of Environmental Parameters on Glycosylation of Recombinant Immunoglobulin G Produced from Recombinant CHO Cells.* Biotechnology and Bioprocess Engineering, 2018. **23**(4): p. 456-464.

138. Sumit, M., et al., *Dissecting N-Glycosylation Dynamics in Chinese Hamster Ovary Cells Fed-batch Cultures using Time Course Omics Analyses.* iScience, 2019. **12**: p. 102-120.

139. Chee Furng Wong, D., et al., *Impact of dynamic online fed-batch strategies on metabolism, productivity and N-glycosylation quality in CHO cell cultures.* Biotechnol Bioeng, 2005. **89**(2): p. 164-77.

140. Raju, T.S. and R.E. Jordan, *Galactosylation variations in marketed therapeutic antibodies.* MAbs, 2012. **4**(3): p. 385-91.

141. Surve, T. and M. Gadgil, *Manganese increases high mannose glycoform on monoclonal antibody expressed in CHO when glucose is absent or limiting: Implications for use of alternate sugars.* Biotechnol Prog, 2015. **31**(2): p. 460-7.

142. Ivarsson, M., et al., *Evaluating the impact of cell culture process parameters on monoclonal antibody N-glycosylation.* J Biotechnol, 2014. **188**: p. 88-96.

143. Clark, K.J., F.W. Chaplin, and S.W. Harcum, *Temperature effects on product-quality-related enzymes in batch CHO cell cultures producing recombinant tPA.* Biotechnol Prog, 2004. **20**(6): p. 1888-92.

144. Kaufmann, H., et al., *Influence of low temperature on productivity, proteome and protein phosphorylation of CHO cells.* Biotechnol Bioeng, 1999. **63**(5): p. 573-82.

145. Trummer, E., et al., *Process parameter shifting: Part II. Biphasic cultivation-A tool for enhancing the volumetric productivity of batch processes using Epo-Fc expressing CHO cells.* Biotechnol Bioeng, 2006. **94**(6): p. 1045-52.

146. Agarabi, C.D., et al., *Bioreactor process parameter screening utilizing a Plackett-Burman design for a model monoclonal antibody.* J Pharm Sci, 2015. **104**(6): p. 1919-1928.

147. Yoon, S.K., J.Y. Song, and G.M. Lee, *Effect of low culture temperature on specific productivity, transcription level, and heterogeneity of erythropoietin in Chinese hamster ovary cells.* Biotechnol Bioeng, 2003. **82**(3): p. 289-98.

148. Sou, S.N., et al., *How does mild hypothermia affect monoclonal antibody glycosylation?* Biotechnol Bioeng, 2015. **112**(6): p. 1165-76.
149. Luo, Y., et al., *Modeling the Effect of Amino Acids and Copper on Monoclonal Antibody Productivity and Glycosylation: A Modular Approach.* Biotechnol J, 2020: p. e2000261.
150. Jimenez Del Val, I., Y. Fan, and D. Weilguny, *Dynamics of immature mAb glycoform secretion during CHO cell culture: An integrated modelling framework.* Biotechnol J, 2016. **11**(5): p. 610-23.
151. Jimenez del Val, I., J.M. Nagy, and C. Kontoravdi, *A dynamic mathematical model for monoclonal antibody N-linked glycosylation and nucleotide sugar donor transport within a maturing Golgi apparatus.* Biotechnol Prog, 2011. **27**(6): p. 1730-43.
152. Davis, E. and M. Ierapetritou, *A kriging method for the solution of nonlinear programs with black-box functions.* 2007. **53**(8): p. 2001-2012.
153. Boukouvala, F., F.J. Muzzio, and M.G. Ierapetritou, *Dynamic Data-Driven Modeling of Pharmaceutical Processes.* Industrial & Engineering Chemistry Research, 2011. **50**(11): p. 6743-6754.
154. Sacks, J., et al., *Design and analysis of computer experiments.* 1989: p. 409-423.
155. Bhosekar, A. and M. Ierapetritou, *Advances in surrogate based modeling, feasibility analysis, and optimization: A review.* Computers & Chemical Engineering, 2018. **108**: p. 250-267.
156. Lophaven, S.N., H.B. Nielsen, and J. Søndergaard, *DACE: a Matlab kriging toolbox.* Vol. 2. 2002: Citeseer.
157. Hernandez, A.F. and M.A. Grover, *Stochastic dynamic predictions using Gaussian process models for nanoparticle synthesis.* Computers & Chemical Engineering, 2010. **34**(12): p. 1953-1961.
158. Boukouvala, F. and M.G. Ierapetritou, *Derivative-free optimization for expensive constrained problems using a novel expected improvement objective function.* 2014. **60**(7): p. 2462-2474.
159. Vergara, M., et al., *Differential effect of culture temperature and specific growth rate on CHO cell behavior in chemostat culture.* PLoS One, 2014. **9**(4): p. e93865.
160. Villiger, T.K., et al., *Controlling the time evolution of mAb N-linked glycosylation, Part I: Microbioreactor experiments.* Biotechnol Prog, 2016. **32**(5): p. 1123-1134.
161. Jiang, R., H. Chen, and S. Xu, *pH excursions impact CHO cell culture performance and antibody N-linked glycosylation.* Bioprocess Biosyst Eng, 2018. **41**(12): p. 1731-1741.
162. FDA, *FDA Briefing Document Oncologic Drugs Advisory Committee Meeting.* 2017.
163. Pereira, S., H.F. Kildegaard, and M.R. Andersen, *Impact of CHO Metabolism on Cell Growth and Protein Production: An Overview of Toxic and Inhibiting Metabolites and Nutrients.* Biotechnol J, 2018. **13**(3): p. e1700499.
164. Villiger, T.K., et al., *Controlling the time evolution of mAb N-linked glycosylation - Part II: Model-based predictions.* Biotechnology Progress, 2016. **32**(5): p. 1135-1148.
165. Karst, D.J., et al., *Modulation and modeling of monoclonal antibody N-linked glycosylation in mammalian cell perfusion reactors.* Biotechnol Bioeng, 2017. **114**(9): p. 1978-1990.
166. Kontoravdi, C. and I.J. del Val, *Computational tools for predicting and controlling the glycosylation of biopharmaceuticals.* Current Opinion in Chemical Engineering, 2018. **22**: p. 89-97.
167. Glick, B.S., T. Elston, and G. Oster, *A cisternal maturation mechanism can explain the asymmetry of the Golgi stack.* FEBS Lett, 1997. **414**(2): p. 177-81.
168. Villiger, T.K., et al., *Controlling the time evolution of mAb N-linked glycosylation - Part II: Model-based predictions.* Biotechnol Prog, 2016. **32**(5): p. 1135-1148.

169. Nam, J.H., et al., *The effects of culture conditions on the glycosylation of secreted human placental alkaline phosphatase produced in Chinese hamster ovary cells.* Biotechnol Bioeng, 2008. **100**(6): p. 1178-92.
170. Xing, Z., et al., *Modeling kinetics of a large-scale fed-batch CHO cell culture by Markov chain Monte Carlo method.* 2010. **26**(1): p. 208-219.
171. Despotovic, M., et al., *Evaluation of empirical models for predicting monthly mean horizontal diffuse solar radiation.* 2016. **56**: p. 246-260.
172. Gramer, M.J., et al., *Modulation of antibody galactosylation through feeding of uridine, manganese chloride, and galactose.* Biotechnol Bioeng, 2011. **108**(7): p. 1591-602.