

**APPLICATIONS OF PROCESS MODELING AND OPTIMIZATION IN
BIOPHARMACEUTICAL MANUFACTURING**

by

Chaoying Ding

A dissertation submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Chemical Engineering

Summer 2024

© 2024 Chaoying Ding
All Rights Reserved

**APPLICATIONS OF PROCESS MODELING AND OPTIMIZATION IN
BIOPHARMACEUTICAL MANUFACTURING**

by

Chaoying Ding

Approved: _____
Millicent O. Sullivan, Ph.D.
Chair of the Department of Chemical and Biomolecular Engineering

Approved: _____
Levi Thompson, Ph.D.
Dean of the College of Engineering

Approved: _____
Louis F. Rossi, Ph.D.
Vice Provost for Graduate and Professional Education and
Dean of the Graduate College

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed:

Marianthi G. Ierapetritou, Ph.D.
Professor in charge of dissertation

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed:

Abraham M. Lenhoff, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed:

Kevin V. Solomon, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed:

Ou Yang, Ph.D.
Member of dissertation committee

ACKNOWLEDGMENTS

First and foremost, I would like to extend my heartfelt gratitude to my advisor, Prof. Marianthi Ierapetritou, for her guidance and support throughout the past five years of my Ph.D. study. Having her as my mentor has been one of the best decisions that I have made during my academic journey. Thanks for her patience, enthusiasm, kindness, and encouragement, which not only helped me get through the challenges in research but also built up my confidence along the way. I am extremely proud and grateful to be one of her students.

I am also deeply thankful to my thesis committee members, Dr. Abraham Lenhoff, Dr. Kevin Solomon, and Dr. Ou Yang, for their valuable suggestions and feedback. I want to express a special thanks to Dr. Lenhoff for his assistance in enhancing my understanding of chromatography and for his review of my first paper at UD.

The completion of my research projects would not have been possible without financial support, including the U.S. Food and Drug Administration (FDA) via grant number U01FD007695A, National Institute for Innovation in Manufacturing Biopharmaceuticals (NIIMBL) via grant number 70NANB17H002, GlaxoSmithKline via grant number RCA3000038468, and CSL Behring LLC via grant number 142680.

It is my great pleasure to be part of the Ierapetritou group and to collaborate with everyone. I would like to thank all former and current members of the group that I interacted with: Abhay, Ou, Atharv, Pooja, Yingjie, Katerina, Borja, Anjana, Yuqiu, Huayu, Dare, Jayanth, Yuqing, Katie, Ching-Mei, Dat, Arnav, Nikola, Shivam, Zhifei,

Jing, Margherita, Daniel, Eileen, and Omkar. In particular, I would like to thank Ou for her efforts in expanding my understanding of the biopharmaceutical field and assisting me in developing the necessary coding and software skills. I would also like to thank Yingjie for introducing me to the group and providing numerous valuable suggestions and insights for my research.

I am also grateful for the collaboration with my industrial partners, especially Christopher Gerberich, Hiren Ardeshta, Christopher Gillespie, Matthew Kujawa, Michael Bartkovsky, and Maen Qadan, whose valuable suggestions have enriched my work and provided diverse perspectives.

During my Ph.D. study, I had the opportunity to do an internship at GSK Biopharm Drug Substance Process Development (BDSD). I would like to extend my gratitude to my mentor Christopher Gerberich for teaching me practical chromatography skills, providing valuable industrial perspectives, and guiding me through challenges encountered during this internship. I am also thankful for my colleagues at BSDS, particularly Andre, Steve, Zheyuan, Xiaodan, Yu, Kevin, as well as our vendor collaborators Wei and Aaron from Purolite.

Finally, I would like to express my special thanks to my family for their unwavering support and understanding throughout the journey. Due to the pandemic, we haven't seen each other for five years. Without their encouragement and constant care, I would not have been able to reach this milestone. I also want to thank my friends for their support and encouragement during times when I felt depressed and doubted myself.

TABLE OF CONTENTS

LIST OF TABLES	x
LIST OF FIGURES	xiii
ABSTRACT	xix

Chapter

1	INTRODUCTION: APPLICATION OF MODELING IN ADVANCED BIOPHARMACEUTICAL MANUFACTURING PROCESSES	1
1.1	Digital Twin.....	2
1.1.1	General Concepts and Framework of Digital Twin	2
1.1.2	Development of Digital Twin in Biopharmaceutical Processes....	5
1.2	Modeling and System Analysis of Biopharmaceutical Processes	8
1.2.1	Mechanistic Modeling	10
1.2.2	Surrogate Modeling	14
1.2.3	Hybrid Modeling	16
1.2.4	Flowsheet Modeling	19
1.3	Challenges in Modeling Biopharmaceutical Processes	21
1.3.1	Development of Appropriate Models for Complex Biopharmaceutical Processes	21
1.3.2	Computational Complexity in System Analysis Applications	22
1.3.3	Proof-of-Concept for New Process Design	23
1.4	Research Goals	23
1.4.1	Design Space Identification for ProA.....	24
1.4.2	Machine Learning-based Optimization for IEX.....	25
1.4.3	Hybrid Model Development for HIC	26
1.4.4	Proof-of-Concept Study for Continuous mAb Production	27
2	DESIGN SPACE IDENTIFICATION OF CONTINUOUS PROTEIN A CHROMATOGRAPHY.....	28

2.1	Introduction	28
2.2	Process Description and Process Model.....	31
2.2.1	Operating Principles of CaptureSMB.....	31
2.2.2	Process Model	32
2.2.3	Evaluation of Process Performance.....	35
2.3	Methodology.....	36
2.3.1	Feasibility Analysis	36
2.3.2	Surrogate Model - Kriging	37
2.3.3	Adaptive Sampling	39
2.3.4	Surrogate Model Accuracy.....	40
2.3.5	Proposed Methodology.....	41
2.4	Results and Discussion	44
2.4.1	Model Validation.....	44
2.4.2	Original Design Space	45
2.4.3	Effect of t_{RR} on Design Space	49
2.4.4	Effect of Constraints on Design Space	53
2.4.4.1	Yield Constraint.....	53
2.4.4.2	Capacity Utilization Constraint	55
2.4.4.3	Productivity Constraint.....	57
2.4.4.4	Simultaneously Changing Constraints	61
2.4.5	Effect of Design Variables on Design Space	63
2.5	Conclusions	65
3	MACHINE LEARNING-BASED OPTIMIZATION FOR ION EXCHANGE CHROMATOGRAPHY	67
3.1	Introduction	67
3.2	Materials and Methods	70
3.2.1	Case Study	70
3.2.2	Mechanistic Model	72
3.2.2.1	General Rate Model.....	73
3.2.2.2	Steric Mass Action	74
3.2.3	Problem Statement.....	75
3.2.4	Optimization Framework.....	77

3.3	Results and Discussion	81
3.3.1	Comparison of Optimization Results	81
3.3.1.1	Optimization Results using ga.....	81
3.3.1.2	Optimization Results using ML-based Framework.....	83
3.3.2	Effect of the Peak Cutting Criteria	90
3.3.3	Design Space Analysis	91
3.4	Conclusions	96
4	HYBRID MODEL DEVELOPMENT FOR HYDROPHOBIC INTERACTION CHROMATOGRAPHY	98
4.1	Introduction	98
4.2	Material and Methods.....	101
4.2.1	Experimental Setup	101
4.2.2	Mechanistic Model	103
4.2.2.1	Kinetic-Dispersive Model (KDM)	103
4.2.2.2	HIC Isotherm	104
4.2.3	Hybrid Model	105
4.2.4	Determination of Model Parameters	107
4.2.4.1	Column Characteristics	107
4.2.4.2	Parameter Estimation.....	107
4.2.5	Process Optimization.....	111
4.3	Results and Discussion	112
4.3.1	Parameter Estimation using Mechanistic Model.....	112
4.3.2	Hybrid Model Development and Validation	115
4.3.2.1	Parameter Estimation.....	115
4.3.2.2	Extrapolation Capability.....	126
4.3.3	Process Optimization.....	129
4.4	Conclusions	130

5	FLWSHEET MODELING FOR INTEGRATED CONTINUOUS BIOPHARMACEUTICAL PROCESS	133
5.1	Introduction	133
5.2	Methods	136
5.2.1	Process Description	136
5.2.2	Process Scheduling	141
5.2.3	Process Economics	144
5.2.4	Simulation Software	148
5.2.5	Ecology Impacts	149
5.2.6	Scenario Analysis	150
5.2.7	Process Debottlenecking	150
5.2.8	Overall Methodology.....	151
5.3	Results and Discussion	152
5.3.1	Economic Analysis	152
5.3.2	Ecological Analysis	156
5.3.3	Scenario Analysis	157
5.3.3.1	Throughput Analysis	157
5.3.3.2	Upstream Titer Analysis.....	160
5.3.4	Process Debottlenecking	163
5.3.5	Alternative Design – Membrane Chromatography	166
5.4	Conclusions	167
6	CONCLUSIONS AND PERSPECTIVES OF FUTURE WORK	171
6.1	Conclusions	171
6.2	Future Work.....	173
	REFERENCES	175
Appendix		
A	ACKNOWLEDGEMENT OF PUBLICATIONS AND COPYRIGHT PERMISSIONS	195

LIST OF TABLES

Table 2.1:	A list of used model parameters.	34
Table 2.2:	Regression models and correlation functions with $d_j=x_j-x_i$ and θ_j	38
Table 2.3:	Summary of active constraints when changing process variables.	47
Table 2.4:	Accuracy metrics at different fixed process variables in Figure 2.8.	48
Table 2.5:	Comparison of the extreme values of feasible regions at different t_{RR} . ..	50
Table 2.6:	Comparison of max Q_{IC} at various t_{RR}	51
Table 2.7:	Summary of active constraints when changing t_{RR}	52
Table 2.8:	The extreme values of feasible regions at different Y constraints.	53
Table 2.9:	Comparison of the extreme values of feasible regions at different CU constraints.....	55
Table 2.10:	The extreme values of feasible regions at different P constraints.	58
Table 2.11:	Summary of design space changes with individual constraint increases.	60
Table 2.12:	Summary of design space changes with individual constraint decreases.....	60
Table 2.13:	Comparison of the extreme values of feasible regions when increasing three constraints.....	61
Table 2.14:	Comparison of the extreme values of feasible regions at different column lengths.	63
Table 3.1:	Parameters related to column geometry, mass transport, and binding behavior of the proteins ($i \in \{RNase, cyt, lyz\}$) applied in the mechanistic model.	71

Table 3.2:	Lower (lb) and upper bounds (ub) of decision variables in optimization case study.	77
Table 3.3:	Optimized operating conditions and objectives using ga and ML-based framework.	82
Table 3.4:	Comparison of obtained yield and purity using multi-step and single linear elution strategies at the optimal operating conditions.	86
Table 3.5:	Comparison of yield, purity, and area under the curve of Ribonuclease at various slope values.....	87
Table 3.5:	Comparison of the optimized results at different percentage thresholds.....	91
Table 3.6:	Comparison of feasibility function values for the sampled points in Figure 3.10a.....	95
Table 3.7:	Comparison of the sign of feasibility function values for the sampled points in Figure 3.10b.....	95
Table 4.1:	HIC chromatographic sequence used in calibration and validation experiments.	102
Table 4.2:	Experimental operating conditions used for model calibration and validation.	102
Table 4.3:	The operating sequence used for the extrapolation test and optimization case study.	110
Table 4.4:	Resin and column packing parameters of the KDM in HIC process. ...	113
Table 4.5:	Calibration NSSE using the Hybrid_MCL model for parameter estimation (1 hidden layer and tanh as the activation function).....	116
Table 4.6:	Comparison of NSSE after each run.	119
Table 4.7:	Comparison of average NSSE after each run.....	119
Table 4.8:	Comparison of calibration and validation error under different NN structures without the regularization strategy incorporated. All tested NN has one hidden layer.	122

Table 4.9:	Calibrated and validation error under different regularization rates using the hybrid model structure (sigmoid as activation function, 1 hidden layer with 2 nodes).	123
Table 4.10:	Comparison of SSE between mechanistic and best hybrid model.	123
Table 4.11:	Comparison of calibrated NSSE using different optimization algorithms.....	124
Table 4.12:	Summarized operating conditions for extrapolation test under same loading conditions (35 g/L loading capacity and 0.2% aggregate content). Exp: experimental conditions for model calibration and validation.	126
Table 4.13:	Optimized operating conditions and objectives using mechanistic and hybrid model.....	129
Table 5.1:	Details of the continuous platform for mAb production.	139
Table 5.2:	Capital cost estimation in the economic analysis.....	146
Table 5.3:	Equipment purchase cost estimation. The correlations were built with data from discussions with our industry partners, SuperPro Designer software, and various publications [150, 261, 269].	147
Table 5.4:	Consumable cost estimation. The price is based on discussion with our industrial partners and SuperPro Designer software.	148
Table 5.5:	Consumable unit masses. The data is from Pollock’s thesis [269] and adjusted based on our assumptions.	149
Table 5.6:	Summary of bioreactor scales and upstream titers considered for the base case and scenario analysis.	150
Table 5.7:	E-factors for process water and consumable usage in the upstream, downstream, and whole process.....	157
Table 5.8:	Summary of the capacity utilization in each chromatography.	164
Table 5.9:	Emerging of process bottlenecks with the increase of titer and perfusion rate.....	164
Table 5.10:	Unit operating COGs at critical titer and perfusion rate.	165

LIST OF FIGURES

Figure 1.1:	Schematic illustration of different components of DT.....	3
Figure 1.2:	A fully integrated DT framework for biopharmaceutical manufacturing.....	7
Figure 1.3:	General overview of the research goals.	24
Figure 2.1:	Schematic diagram of a CaptureSMB system for continuous capture.	32
Figure 2.2:	A schematic figure for the model performance metrics.	41
Figure 2.3:	Framework of Kriging-based feasibility analysis with adaptive sampling.	42
Figure 2.4:	Comparison of 2D feasibility plots of CaptureSMB process at different iteration numbers for adaptive sampling.	43
Figure 2.5:	Comparison of EI function plot at different iteration numbers when $Q_{IC} = 1.5 \text{ mL min}^{-1}$	43
Figure 2.6:	a) fitted breakthrough curves (line) at different flowrates with experimental data (dot) at the feed concentration of 1.2 mg mL^{-1} ; b) fitted breakthrough curves at different feed concentrations with experimental data. The flowrate is 1.0 mL min^{-1} for the runs with feeding concentrations of 1.2 and 4.6 mg mL^{-1} , whereas the flowrate is 1.5 mL min^{-1} for the experiment with a feeding concentration of 3.2 mg mL^{-1}	44
Figure 2.7:	a) 3D contour plot showing the whole input space; b) 3D contour plot only showing the feasible region.....	45
Figure 2.8:	2D projections of the 3D feasible region at different values of the process variables for the CaptureSMB process.	47
Figure 2.9:	2D contour plots at different fixed t_{IC} for the CaptureSMB process.	49
Figure 2.10:	Comparison of 3D feasible regions at different t_{RR}	49

Figure 2.11: Comparison of the 2D contour plots at different t_{RR} under varying fixed process variables for the CaptureSMB process.....	52
Figure 2.12: Comparison of the 2D contour plots under different Y constraints at varying fixed process variables for the CaptureSMB process.	54
Figure 2.13: Comparison of the 2D contour plots under different CU constraints at varying fixed process variables for the CaptureSMB process.	56
Figure 2.14: Comparison of the 2D contour plots under different P constraints at varying fixed process variables for the CaptureSMB process.	59
Figure 2.15: Comparison of 2D contour plots for simultaneously increasing three constraints in the CaptureSMB process. Base: $Y \geq 0.8$, $CU \geq 0.85$, $P \geq 15$; Alternative: $Y \geq 0.9$, $CU \geq 0.9$, $P \geq 17$	62
Figure 2.16: Comparison of the 2D contour plots under different column length at varying fixed process variables for the CaptureSMB process.	64
Figure 3.1: Operating protocols for the CEX considered in the modeling. Elu_1 , Elu_2 , and Elu_3 correspond to the multistep elution strategies. m_1 and m_2 represent the slopes for the first two linear elution phases; Δt_1 and Δt_2 denote the time intervals corresponding to those elution phases; c_{init0} refers to the initial salt concentration at the beginning of the elution process.	72
Figure 3.2: Proposed machine-learning (ML) based optimization framework.....	80
Figure 3.3: The full chromatograms at the optimized operating conditions obtained from ga (a) and the ML-based framework (b), with enlarged plots displayed in (c) and (d), respectively.....	83
Figure 3.4: The new chromatograms generated after the removal of the first linear elution step at the optimal operating conditions obtained from ga (a) and the ML-based framework (b).....	85
Figure 3.5: Different simulated chromatograms at different selected slope values for the first linear elution step.	86
Figure 3.6: Comparison of the obtained chromatogram at the optimal solution for the multi-step and single step case study.....	87

Figure 3.7:	Comparison of the obtained optimal chromatograms under different purity and yield requirements: (a) Purity requirement of 90% and yield requirement of 80%; (b) Purity requirement of 95% and yield requirement of 65%.....	89
Figure 3.8:	Comparison of the chromatograms under the optimized conditions at various percentage thresholds. The blue, red and green lines denote the column outlet concentration of lysozyme, cytochrome, and ribonuclease.....	91
Figure 3.9:	2D contour plots of the design space projected from high dimensional space at the optimal solutions. The red line represents the feasible boundary identified using the GPR model, and the red square denotes the optimal solution obtained from the ML-based framework.....	93
Figure 3.10:	Selected 2D contour plots of the design space to illustrate the framework. The white dash line is the feasible boundary with the feasibility function value of 0, while the red solid line indicates the salt constraint. The red square and circles are optimal solution and sampled points, respectively.....	94
Figure 4.1:	The overall schematic of parameter estimation using hybrid model.....	109
Figure 4.2:	Comparison of the calculated bind-and-elute curves with the experimental dataset under different operating conditions during parameter estimation. All the experiments were conducted using a flow rate of 300 cm/h. Mech: mechanistic model; Exp: Experiment....	114
Figure 4.3:	a) Comparison of the predicted bind-and-elute curves with the experimental dataset for model validation; Correlation between measured and predicted concentration for monomer (b) and dimer (c). Mech: mechanistic model; Exp: Experiment.	115
Figure 4.4:	Illustration of neural network structure and parameters in hybrid model.	116
Figure 4.5:	Comparison of simulated elution curves and experimental results using Hybrid_MCL model.	117
Figure 4.6:	Comparison of simulated and experimental elution curves using Hybrid_Desorption model.	118
Figure 4.7:	Comparison of predicted and experimental elution curves using Hybrid_Desorption model.	119

Figure 4.8: The NNSE for the calibration and validation dataset under different regularization rates.	120
Figure 4.9: Comparison of simulated and experimental elution curves after the incorporation of regularization strategy ($\lambda = 2.3E - 4$).	121
Figure 4.10: Comparison of NSSE for each experiment using mechanistic and hybrid models.	124
Figure 4.11: Comparison of the simulated bind-and-elute curves with the experimental dataset under different operating conditions during parameter estimation. All the experiments were conducted using a flow rate of 300 cm/h. Hybrid: Hybrid_Desorption model; Exp: Experiment.	125
Figure 4.12: a) Comparison of the predicted bind-and-elute curves with the experimental dataset for model validation; Correlation between measured and predicted concentration for monomer (b) and dimer (c). Hybrid: Hybrid_Desorption model; Exp: Experiment.	126
Figure 4.13: The comparison of predicted elution curves using hybrid and mechanistic models under four extrapolation tests.	128
Figure 4.14: The comparison of extrapolation test 1 with unmodified dimer concentration.	128
Figure 4.15: The predicted elution curves under optimized operating conditions obtained from mechanistic (a) and hybrid models (b).	130
Figure 5.1: Process flow diagram (PFD) for continuous mAb production process.	139
Figure 5.2: Main equipment occupancy chart for the continuous mAb production process. Legend: DSBS-101, seed bioreactor; Disposable Bior, production bioreactor; C-101, one protein A column in primary capture; Virus inactivation, plug flow reactor; C-104, one AEX column; C-106, one CEX column; DE-102–UF104: SPTFF.	142

Figure 5.3:	The schematic representation of buffer preparation step in primary capture. Red lines: loading step from the surge tank SDLB-101 between upstream and Protein A chromatography; orange lines: washing steps from washing buffer hold tank SDLB-114; turquoise lines: elution steps from elution buffer hold tank SDLB-109; green lines: regeneration steps from regeneration buffer hold tank SDLB-111; blue lines: equilibration steps from the washing buffer hold tank SDLB-114 (because the equilibration buffer and washing buffer are the same). DE-110, DE-107, DE-106: sterile filtration; SDLB-113: washing/equilibration buffer preparation tank; SDLB-110: elution buffer preparation tank; SDLB-112: regeneration buffer preparation tank; SDLB-133: surge tank after Protein A chromatography to cyclically collect the eluate.	143
Figure 5.4:	Scheduling of buffer preparation in primary capture. W: first and second washing steps; L: loading; Elu: elute; R: regeneration; E: equilibrate; H: hold; T: buffer transfer into the buffer hold tank. Red: operation in the first column (C-102); purple: operation in the second column (C-101); orange: operation in the third column (C-103). C: chromatography columns; SDLB-114: washing/equilibration buffer hold tanks; SDLB-109: elution buffer hold tanks; SDLB-111: regeneration buffer hold tanks.....	144
Figure 5.5:	Production expenditure breakdown with capital expenditure (CapEX) and operating expenditure (OpEX).	145
Figure 5.6:	The overall framework for cost and ecology evaluation of continuous mAb production. The green box represents the repetitive procedures for an alternative design.	151
Figure 5.7:	Capital cost (a) and operating expenditure (b) breakdown in different sections.	154
Figure 5.8:	Operating cost of goods (COGs) breakdown in different sections (a) and cost categories (b). VI: virus inactivation; VR: virus removal; FF: final formulation.....	154
Figure 5.9:	Cost breakdown with different sections.	155
Figure 5.10:	Change of total operating COGs (a), media consumption per batch (b), COGs breakdown at varying cost categories (c) and sections (d) with respect to different bioreactor scales.	158

Figure 5.11: Change of the ratio to the normal value of different ecology indicators (a) and comparison of the percentage of process water between upstream and downstream (b) with respect to bioreactor scales. 160

Figure 5.12: Change of total operating COGs (a), COGs breakdown at varying cost categories (b) and sections (c) versus the titer; Change of operating COGs in upstream, downstream, and the cost ratio (d)..... 161

Figure 5.13: Change of the ratio to the normal value of different ecology indicators (a) and comparison of the percentage of process water between upstream and downstream (b) with respect to different titers. 162

Figure 5.14: Comparison of unit operating COGs under different cost categories (a) and sections (b) at critical titer and perfusion rate. VI: virus inactivation; VR: virus removal; FF: final formulation. 165

Figure 5.15: Economical and ecological comparisons of the base case and hybrid case scenarios. VI: virus inactivation; VR: virus removal; FF: final formulation. 167

ABSTRACT

With the increasing global market demand, prohibitive drug cost, and biosimilars competition, the biopharmaceutical industry is under pressure to speed up the development and manufacturing of biological products. To increase biologics production while maintaining product quality, the industry attempts to 1) improve the processes through intensification and optimization, and 2) explore new process designs. With this shift comes increasing complexity leading to a large number of process variables to be monitored, controlled, and optimized. Moreover, the Quality-by-Design (QbD) initiative introduced by FDA to take product quality into account during the process design, is driving the biopharma industry to acquire deeper process insights. With the increasing demand for process understanding and the pressure in cost reduction, the industry is turning to *in silico* solutions to assist with such a transition. Given the industrial trend, this dissertation aims to apply process systems engineering (PSE) tools to empower the *in silico* development of biopharmaceutical manufacturing processes.

The first part of the thesis, as presented in Chapters 2-4, focuses on the applications of PSE tools on individual unit operations, specifically various chromatography units, to support process characterization and optimization. In Chapter 2, a surrogate-based feasibility analysis method is proposed to identify the design space of continuous Protein A chromatography, aiming to strike a balance between computational complexity and model prediction accuracy. Machine learning-based optimization framework is introduced in Chapter 3 to address the nonconvex and

nonlinear constrained optimization challenges encountered in biopharmaceutical separation. This framework is applied to a case study involving the separation of a ternary protein mixture using ion-exchange chromatography. The focus of Chapter 4 is to develop hybrid models with enhanced model predictability to describe the unclear and complex binding behavior within the hydrophobic interaction chromatography. The aim is to reduce the investment effort required for developing mechanistic model while extracting the missing relationships that cannot be captured by the mechanistic model.

Considering the trend in transitioning from batch to continuous processes, the second part of the thesis focuses on conducting proof-of-concept study to evaluate the feasibility of transitioning from batch to continuous biomanufacturing mode through the establishment of *in silico* platform, as presented in Chapter 5.

The applications of PSE tools in the biopharmaceutical processes, specifically through process modeling and system analysis, have demonstrated significant potential in enhancing process understanding and facilitating process development / improvement. The methodologies and frameworks proposed in this dissertation can provide further insights into the cost-effective development and manufacturing of high-quality biological drugs, promoting further investigation and implementation of *in silico* technology in biopharmaceutical industry.

Chapter 1

INTRODUCTION: APPLICATION OF MODELING IN ADVANCED BIOPHARMACEUTICAL MANUFACTURING PROCESSES

Tremendous growth has been witnessed in the biopharmaceutical market over the past few years, as evidenced by the increasing biological drug approvals by the EU and the US regulatory agencies [1-4]. Driven by patient demand and competition in the biosimilar market, the biopharmaceutical industry is under pressure to speed up the development and manufacturing of biological products [5]. However, the development of resource-intensive bioprocesses is invariably accompanied by high manufacturing costs, which makes biotherapeutics inaccessible to a significant part of the population [6]. To improve the efficiency of development and manufacturing of biologics while adhering to quality and regulatory requirements, the biopharmaceutical industry attempts to 1) improve the processes through intensification and optimization, and 2) explore new process designs. For instance, there is a transition from traditional fed-batch bioreactors to more complicated perfusion processes [7, 8] and from single-

Sections of Chapter 1 are adapted from the following journal articles and book chapters written by the author of this dissertation:

Ding, C., Yang, O., Ierapetritou, M., Towards digital twin for biopharmaceutical processes: concept and progress. In: Pörtner, R. (eds) Biopharmaceutical Manufacturing. Cell Engineering, vol 11. Springer.

Malinov, N., Raudenbush, K., Ding, C., Reddy, J., Ierapetritou, M., End-to-end process flowsheet modeling for biopharmaceutical production: current state and future potential. *Current Opinion in Chemical Engineering* [Under Review].

column batch chromatography to multi-column continuous operation [6, 9, 10]. However, this transition significantly increases the complexity of bioprocessing, necessitating the monitoring of additional process variables to ensure both reliable operations and product quality. Moreover, the Quality-by-Design (QbD) initiative introduced by FDA to take product quality into account during the process design, is driving the biopharma industry to acquire deeper insights into the connection between bioprocess and bioproduct [11]. With the increasing demand for computational capabilities and process understanding, the industry is turning to *in silico* solutions to assist with such transition by constructing a digital platform, referred to as a digital twin (DT), to evaluate the updated process designs and operational schemes before making final decisions [12, 13].

1.1 Digital Twin

1.1.1 General Concepts and Framework of Digital Twin

The concept of the digital twin (DT) can be traced back to 2002 when the “Conceptual Ideal for Product Lifecycle Management (PLM)” was first proposed by Dr. Grieves and presented to the industry at the University of Michigan [14-16]. Since the inception of DT, this terminology has progressively evolved over time, but the basic idea of DT has barely changed. The most widely accepted interpretation comes from Glaesegen and Stargel, that a DT is a combination of multi-physics, multiscale, and probabilistic simulation for a complicated process with the utilization of physical data, sensors, and models to reflect the behavior of its corresponding physical twin [13, 17]. A complete DT is composed of three key elements: a physical component that currently exists or will exist in the physical world (“physical twin”), a virtual/digital part that

platform is developed for generating DT to supply historical and real-time process information and to run data analytics and optimization [19]. Nonetheless, from the perspective of practical application, simplified or partial DT is more favorable for the industry because it is much easier to realize in the actual production process. In particular, a digital shadow is one of the simplified DT concepts in which there is only a one-way flow from physical to digital components but no channel for reverse data communication [13, 21].

The digital twin is attracting increasingly more attention recently due to the rapid development of the Internet of things (IoT), Industry 4.0 technologies, advanced data analytics, and artificial intelligence [22, 23]. A complete DT could provide a comprehensive and holistic understanding of the process by constructing a network of dependencies between real-time data and their underlying meta information. Specifically, the establishment of DT provides an efficient way for achieving the cyber-physical integration of manufacturing practices, which could help obtain all accessible process information across different scales and sites and enable real-time information communication with the physical process control system [24, 25]. By virtue of the digital object like process analysis and optimization, *in silico* process model could be developed and updated in sync with data from the physical component to provide guidance for the actual procedures and further improve the overall efficiency and mitigate the risks in operation and maintenance. In general, the applications of DTs have demonstrated their capabilities in facilitating remote sensing, real-time monitoring, data acquisition, information exchange, process visualization, and process knowledge extraction [13, 26-29]. Integrated DTs have been applied in various industries including

aerospace, semiconductor, energy production, smart product manufacturing, automotive transportation, and healthcare [17, 19, 23].

1.1.2 Development of Digital Twin in Biopharmaceutical Processes

A typical platform for biopharmaceutical processes can be divided into upstream (USPs) and downstream processes (DSPs), where USPs are mainly used for protein production and DSPs are for protein purification [30]. USPs involve inoculation and seed cultivation to scale up cell lines, as well as cell culture to produce proteins of interest. DSPs in biologic production consist of various unit operations, including product recovery, primary capture, virus inactivation, polishing, virus removal, and ultrafiltration/diafiltration (UFDF) steps [31, 32]. Product recovery aims to separate the biomass and most impurities from the media through the combination of centrifugation and depth filtration [33]. Primary capture is performed to remove most of the process-related impurities like culture media components, host cell proteins, and DNA. For primary capture, Protein A affinity chromatography is a standard method for most all monoclonal antibodies (mAbs) and related products because of its high selectivity and efficiency [34, 35]. Virus inactivation and removal steps are crucial for ensuring virus clearance, thereby meeting the regulatory requirements [36, 37]. The polishing step aims to eliminate product-related impurities that have similar properties to the target protein, such as fragments, aggregates, charge variants, residual process-related impurities, as well as viruses [38]. Commonly used chromatographic techniques during this step include ion-exchange chromatography (IEX), hydrophobic interaction chromatography (HIC), and mixed-mode chromatography (MMC). UFDF is conducted to obtain the therapeutic protein at the desired concentration and exchange buffer [39].

Among these unit operations, primary capture and polishing step are the most significant operations to achieve protein purification [40].

The implementation of DT in biopharmaceutical manufacturing could help accelerate the process development and gain thorough understanding of the process while reducing the efforts of performing time and resource intensive experiments [13, 41]. The realization of DT would allow for real-time monitoring, prediction of crucial quality attributes, identification of design space, and process optimization to provide real-time feedback and guide the operating procedures [1, 3]. Moreover, with the development of simulation platforms and the fast-growing biological drugs market, biopharmaceutical companies nowadays have been working toward applying digital twin concepts during process development and commercial manufacturing. The advanced computational hardware expands the capabilities of modeling which allows faster and more reliable model building.

To accelerate process scale-up and reduce experiments required, Pfizer built a two-phase bioreactor computational fluid dynamic (CFD) simulation to predict the transport and fluid dynamics inside of bioreactor from small to large scale. Their GPU workstation allows parallel computations which can complete massive simulations within a limited time and rule out nonideal physical conditions [42]. Biopharmaceutical companies intend to collaborate with digital transformation companies to build their digital twin system. For example, GlaxoSmithKline (GSK) cooperates with Siemens and ATOS to operationalize digital twins by building a real-time vaccine manufacturing simulation platform. The project aims to reduce the development times and optimize the quality of their vaccine product [43, 44]. In addition, BioNTech also worked with Siemens to digitalize their facility to produce COVID-19 vaccine and reduce the

transformation timeline from one year to five months by applying automation and digitalization technologies [45]. Sanofi announced a partnership with Dassault Systems in order to use their 3DEXPERIENCE platform to visualize and optimize operation activities. The collaboration covers the end-to-end data management system, virtual operation, process development scale-up and production industrialization [46]. From the above examples, it is clear that digital twin platforms contain key components including a robust modeling platform, data acquisition and centralized data management system. Control strategies should also be included for virtual plant and physical plant "communication" [47].

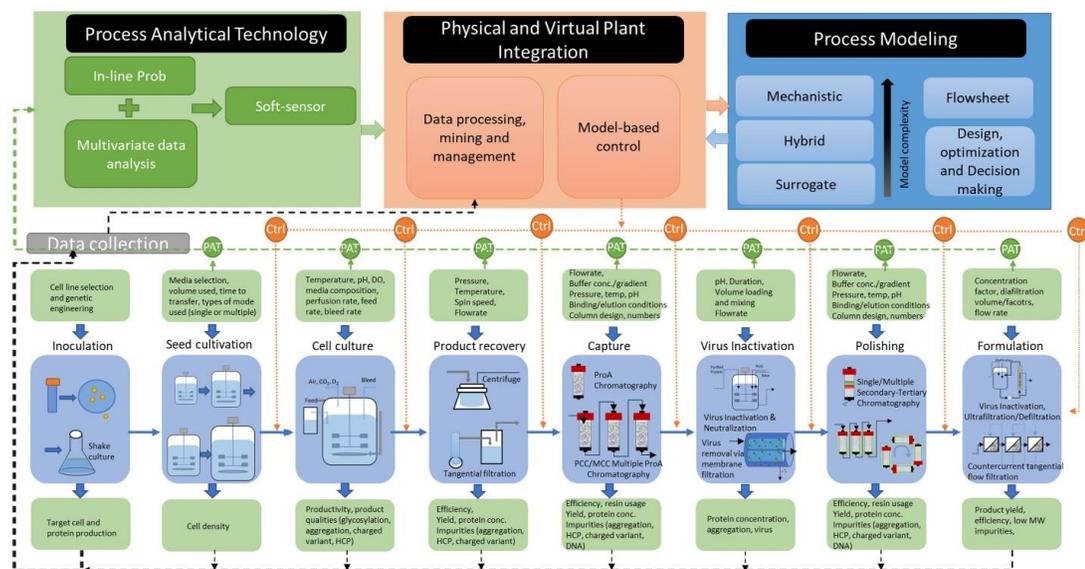


Figure 1.2: A fully integrated DT framework for biopharmaceutical manufacturing.

In summary, to realize a digital twin for existing physical plants, a complete virtual plant can be built, and a physical-virtual plant communication needs to be

established, as depicted in Figure 1.2. The digital twin platform includes the components of process analytical technologies, process modeling, and physical-virtual plant integration. The framework also highlights the methodologies and lists detailed inputs and outputs of each unit operation that can be measured and optimized. Among all, process modeling acts as a foundation, as the different modeling and analysis methods are critical to ensure that the data are appropriately analyzed, and the processes are accurately represented *in silico*.

1.2 Modeling and System Analysis of Biopharmaceutical Processes

To realize DT in biopharmaceutical manufacturing, it is essential to develop a comprehensive virtual representation of the physical plant, wherein process modeling plays a critical role. Process modeling, which transforms process knowledge into mathematical representations, could aid in reducing the experiment effort and gaining a deeper understanding of the process behavior. Based on the types of unit operations and modeling objectives, different modeling approaches such as mechanistic, surrogate, and hybrid modeling are commonly employed. Mechanistic modeling, established on the basis of governing phenomena, offers an in-depth understanding and insights into underlying mechanisms. However, its accuracy heavily relies on the current understanding and knowledge of the processes. A surrogate model serves as a computationally inexpensive representation of a complex system or process, commonly constructed using statistical techniques, machine learning (ML) algorithms, or other mathematical methods trained on data from the original system. By leveraging the existing process knowledge and information derived from the collected data, hybrid modeling combines mechanistic models with surrogate models to improve accuracy and capture the complex dynamics of the processes. After the development of unit operation

models, these models can be integrated together to construct the flowsheet model, representing the integrated processes [48, 49].

Following the development of models to virtually simulate physical processes, the further applications of process systems engineering (PSE) tools allow for in-depth system analysis and facilitate their application to real processes for analysis and improvement. Typically, the PSE tools include sensitivity analysis, feasibility analysis, economic analysis, sustainability analysis, and optimization techniques. Sensitivity analysis primarily serves as a tool for risk assessment by identifying the critical process parameters (CPPs) that influence the critical quality attributes (CQAs). It examines how variability in the model inputs results in the variations in model outputs [50]. Especially in complex processes with numerous parameters, sensitivity analysis plays a vital role in reducing the dimensionality of the overall parameter space, which further helps lower computational complexity or experimental efforts in subsequent analyses [51]. Feasibility analysis can be used to characterize the design space (feasible region) of a given process through feasibility function [52]. Identification of design space, determined by the ranges in which a given process must operate to ensure product quality, can help comprehend the main effects of process variables and their interactions on product quality, facilitating a deeper understanding of the process and establishing effective control strategies [53, 54]. Economic analysis is mainly used to assess the economic feasibility of industrial processes for decision-making and has been extensively applied across various engineering domains [55, 56]. This analysis involves evaluating process efficiency through process models and analyzing factors such as cost of goods (COGs), equipment cost, capital and operating expenditures, profitability, and cash flow [57]. Sustainability analysis is employed to estimate the potential

environmental impacts of products or processes, utilizing methodologies such as life-cycle assessment [58, 59]. Process optimization is a valuable tool in identifying the optimal operating conditions to achieve objectives such as maximizing yield or productivity while maintaining product quality. This type of analysis, conducted through process modeling, can significantly reduce required experimental efforts and costs [60].

In this section, current progress on process modeling and their applications in system identification and analysis of biopharmaceutical processes are reviewed based on each type of modeling strategy.

1.2.1 Mechanistic Modeling

In biopharmaceutical manufacturing, protein-based, or mRNA-based products are produced using a series of production steps including a bioreactor followed by separation, purification steps and final formulation steps. For bioreactor modeling, mechanistic models can correlate operating conditions such as dissolved oxygen (DO), pH, temperature, feeding strategy, media compositions to cell activities, metabolite concentrations, system homogeneity, and fluid dynamics. The models can predict and optimize process productivity, and product quality [61]. Inside of bioreactor, there is a complex system with multiphase transport phenomena, and dynamically changed cell populations characterized by complicated bioreactions [62]. These phenomena can be captured by the structured or unstructured kinetic model, flux balance analysis model, and CFD simulation. The kinetic model can be majorly used to predict and optimize titer and product quality by adjusting the temperature downshift and feeding strategies. Feeding strategies include feeding frequency, amount [63] as well as additional components [64, 65]. Apart from titer and quality, population balance model is used

[66] to predict the host cell protein (HCP) accumulation during the cell culture. However, kinetic models cannot include all metabolic reactions to avoid computational expensive calculations. In stoichiometric model (such as flux balance analysis and metabolic network analysis), the main assumption is considering that the intracellular reaction is under a pseudo-steady state and net flux is zero [67]. Using this model fluxes can be estimated for metabolic cycles inside the cell to understand metabolic bottlenecks under different operating conditions [68-70]. To capture the cell dynamics, a dynamic metabolic flux analysis is developed, which discretizes the time domain to multiple intervals and calculates the flux change during each of these intervals [71, 72]. Another approach is to integrate the stoichiometric model with the kinetic model [73]. As mentioned in the previous section, CFD simulation can be used to capture gas-liquid transfer and mixing heterogeneity of the bioreactor. CFD simulation can be integrated with kinetic modeling methods to predict titer under different operating conditions by using oxygen addition, nutrient heterogeneity, and agitation rate.

Among the downstream operations, column chromatography has played a dominant role in biopharmaceutical protein purification with high selectivity over the past two decades [74, 75]. Generally, in the production of mAb-related products, Protein A chromatography is applied during the primary capture step to remove process-related impurities [76, 77], while different modes of chromatography are utilized in the polishing step to further remove product-related and other residual impurities [78-80]. Current research on chromatography modeling has focused on the understanding of solute transport in mobile phase (inter- and intraparticle) and adsorption behavior in stationary phase [74]. With different assumptions and simplifications, there are different mechanistic models to describe mass transport of the moving phase inside the column,

including equilibrium dispersive model, lumped kinetic model, and general rate model (GRM) [81-83]. Among them, general rate model is the most comprehensive one that accounts for convection and axial dispersion in the interstitial column volume, film mass transfer of solute from the interstitial to the intraparticle pore space, pore and/or surface diffusion within the resin particle, and protein adsorption [84, 85]. The transport model is coupled with a kinetic model to describe the adsorption phenomena that drive the chromatographic separation. There are different mathematical forms of kinetic expressions in the literature [74, 86], and among these, the Langmuir model is a simple and widely used model to illustrate the intraparticle binding mechanism. Nonetheless, since the Langmuir model assumes independent binding sites and cannot account for the dependence of salt on the adsorption, the steric mass action (SMA) model, which incorporates the steric hindrance effects, is introduced for ion-exchange chromatography [86, 87]. After the mechanistic model building, parameter estimation and model validation are required to avoid overfitting and ensure the reliability and generality of the model. These developed mechanistic models can further contribute to the prediction of chromatographic behaviors and performance metrics like purity, yield and productivity, identification of the design space and process optimization [78, 88-91]. Bhojar et al. [92] developed a robust mechanistic model capable of predicting full Protein A chromatograms. This model holds the potential for predicting chromatographic behaviors in larger columns under varying flow rates, as well as different elution schemes and buffer conditions. Cebulla et al. [93] utilized mechanistic model-based optimization technique to separate von Willebrand factor fragments and human serum albumin. This approach resulted in higher product purity and reduced buffer consumption, compared to the originally conducted experiments.

Additionally, in order to capture the heterogeneity of internal flow in the radial direction of the column, computational fluid dynamics modeling is also applied to simulate the dynamic fluid flow [94, 95], but the application of CFD is limited due to the expensive computational cost, high knowledge requirement of the internal column geometry, and difficulty in experimental validation. It is worth noting that two dimensional general rate model (GRM2D) was recently developed by the CADET (Chromatography Analysis and Design Toolkit), which is an open-source package under the supervision of Dr. Eric von Lieres [96]. The newly built GRM2D model is adapted from GRM by introducing a radial coordinate to consider the inhomogeneity transport along the radial dimension [97].

Membrane-based unit operations like virus filtration and ultrafiltration are also indispensable in the biopharmaceutical downstream process [98, 99]. Virus filters are widely employed to ensure the safety of biotherapeutic proteins, and size-exclusion is normally considered the dominant mechanism for virus clearance [100]. The blocking model theory with varying blocking indexes combined with cake filtration model could accurately characterize the virus filtration behaviors such as the membrane fouling and filtration flux under different buffer conditions. In addition to the size-exclusion mechanism, there exist other mechanisms like adsorption described by the Langmuir and Freundlich theories [36, 37, 100-103]. Ultrafiltration modules are typically used to adjust the concentration of the target protein component and the buffer media composition [38]. Reliable transport-phenomena-based models have been developed to represent the ultrafiltration process, for which it is modeled via the combination of Stagnant Film Model for concentration polarization, Osmotic Pressure Model and Boundary Layer Model for the mass transfer through the membrane [41, 104, 105].

Binabaji et al. [106, 107] developed a mechanistic model to describe the behavior of ultrafiltration under highly concentrated antibody solution and validated the model through experimental efforts. This model, capturing the intermolecular interactions and back-filtration phenomenon at high concentration, could help increase the filtrate flux and maximum achievable concentration by evaluating the effect of membrane modules and operating protocols via simulation.

1.2.2 Surrogate Modeling

Mechanistic models suffer from the need for high computation power, and their accuracy highly relies on the obtained process knowledge. The surrogate model, also known as data-driven or black-box model, is an efficient approach to simulate the process with limited understanding and high complexity. With large amounts of data involved, the computationally cheap surrogate model in a lower dimensional space can be obtained to represent the original complicated process. Surrogate-based model has been applied both in the upstream and downstream operations in biopharmaceutical processes [108, 109].

In upstream, data-driven model has been used for cell-line selection [110], media optimization [111], process prediction, optimization, and scale up [112, 113]. Multivariate analysis has the capabilities of handling large amounts of data and information. Sokolov et al. [114] used principal component analysis (PCA) to find correlations between pH, temperature shift, and media composition with quality attributes including aggregate, fragments, charge variants, and glycans. The authors also used partial least square regression (PLS) model coupled with genetic algorithm to predict product titer and other quality attributes. Similarly Green and Glassey [115] use PLS and multi-way PLS to predict titer and quality attributes. The results show that

online measurements such as pH and DO could not provide good prediction on quality attributes. Combined with amino acids data, the quality attributes could be predicted, which is not limited to lab scale but also manufacturing scale [116]. The predictability of the model can potentially be used as a control chart to identify the deviations during the operation [112]. One limitation of the PLS model is that it can only capture the linear correlations. To deal with this issue, modelers can choose machine learning methods such as neural networks (NNs), random forests, extreme gradient boosting, support vector machines, Gaussian processes, etc. Alavijeh et al. [112] reviewed the advantages and disadvantages of each of the methods from the requirement of data source, model efficiency and robustness. Bashokouh et al. [111] used NN with multilayer feed-forward structure to perform a predictive model of titer and then apply the optimization algorithm to find optimum amount of fetal bovine serum in media, temperature and incubation time for maximum productivity.

The different applications of data-driven model have also been demonstrated in downstream unit operations like column chromatography [117, 118]. Wang et al. [119] applied NN modeling in protein chromatography for the estimation of adsorption and mass transfer parameters by mapping the *in silico* experimental chromatograms to the corresponding model parameters. Li et al. [120] developed two different types of surrogate models for efficient optimization of the simulated moving bed chromatography and compared the performance of original full-order and reduced-order models in terms of the computational time and accuracy. Additionally, Sachio et al. [121, 122] applied machine learning-based approaches to enable quick and efficient identification of the design space for Protein A chromatography, which is useful in the quantification of process flexibility, assessment of different operating points, and resin

screening. Moreover, Bayesian optimization was employed for inverse fitting parameters in ion-exchange and hydrophobic interaction chromatography operations, providing low computational demand and further facilitating rapid process development [117].

Besides the application in chromatographic processes, ANNs and regression models are trained based on the CFD simulation data of the membrane spacers under a wide range of potential commercial designs and the resulting surrogate model could accurately predict pressure loss and mass transfer coefficients [123]. In order to capture the complicated fouling dynamics, Zhang et al. [124] utilized the data collected from a pilot-scale ultrafiltration membrane system to obtain three different data-driven models (linear regression, NNs, and random forest), compared the performance of three modeling techniques, and optimized the backwash sequence timing.

1.2.3 Hybrid Modeling

Mechanistic and surrogate models could both be used to simulate the process, but they have different requirements for the process understanding and data. Mechanistic model requires *a priori* basis of knowledge about the process, and it is challenging to postulate appropriate assumptions and develop a precise model, but the model can be highly generalizable with physical or empirical interpretation of the process. Surrogate model is determined exclusively from data, so the quantity and quality of data significantly affect the model performance. Due to the lack of first-principle knowledge, the computationally efficient surrogate model has poor interpretability and generalizability. Thus, hybrid modeling is introduced to combine the advantages and alleviate the drawbacks of the two models. The missing knowledge in the mechanistic model can be represented by a surrogate model, constructing a hybrid

model. It was observed that this strategy has the potential to significantly improve the model predictive performance [24, 125, 126].

The hybrid model of bioreactor modeling shows higher accuracy, extrapolation and interpolation capabilities than both the mechanistic and data-driven models [127]. Narayanan et al. [128] used kinetic models to build mass balance equations. Then the authors applied BlackBox-PLS2 and stepwise-PLS2 to predict rate constant from operating parameters and metabolite concentrations. This method integrates kinetic models and data-driven models. Similarly, Stosch et al. [129] also used hybrid model to analyze the effects of temperature, pH, cell concentration, and feed rate on productivity. Another approach is to have kinetic models and data-driven models connected in series. Kotidis and Kontoravdi [130] developed HyGlycoM model to predict glycan concentrations under different feeding strategies. The model comprises two kinetic models for cell metabolism and nucleotide sugar donors (NSDs) and uses the output – NSDs concentration as inputs of the NN model to estimate the glycosylation process. Instead of using machine learning methods, the model-based design of experiment was applied to optimize media composition and feeding strategies. In this method, mechanistic model built in MATLAB is integrated with DesignExpert (AspenTech) a software for design of experiment and basic statistical analysis [131]. In this case, the design of experiments can run in silico which will significantly reduce the resource consumption for lab experiments and at the same time accelerate process development. Data driven model can be also integrated with stoichiometric model to understand metabolic shifts during the cell culture [132], and predict amino acid concentration [133], or quality attributes [134].

In addition to its application on the upstream bioreactors, hybrid modeling strategies are also applied in the chromatography and ultrafiltration procedures [135]. Narayanan et al. [136, 137] developed different hybrid models with varying degrees of process knowledge incorporated and evaluated their performance in terms of the interpolation and extrapolation capabilities, prediction accuracy and application into continuous chromatography system. Specifically, lumped kinetic model (LKM) is selected for the mass transport because the fluid dynamics inside the column are well described, while NN is introduced to represent the missing knowledge in the adsorption dynamics. The developed hybrid model outperformed the mechanistic model in terms of prediction accuracy and robustness. To investigate the Protein A resin lifetime, Gaussian Processes were first applied to correlate the aging parameters with cleaning and loading conditions, followed by integrating the GP to the lumped kinetic model to construct the hybrid model [138]. The developed model could give insights into the aging mechanism, which could further provide guidance about how to extend the resin lifetime and decrease the cost of goods through process analysis and optimization. Tang et al. [139] combined physics-informed neural networks (PINN) with LKM for fitting and predicting experimental breakthrough curves. However, this approach necessitates large datasets, and the model performance needs to be greatly improved.

With respect to ultrafiltration, Krippel et al. [140] established a hybrid model to predict the flux evolution and duration of ultrafiltration process under different operating conditions like various proteins, membrane types, input parameters and filtration modes. The prediction and robustness of this hybrid model are superior to the mechanistic film theory model, which could be further integrated into the virtual plant of digital twin. Thiess et al. [141] combined a regression model with a physicochemical

model to characterize and quantify the effect of module geometries of ultrafiltration cassettes on mass transfer and pressure drop.

1.2.4 Flowsheet Modeling

Besides individually modeling the unit operations, flowsheet models could be built by integrating these single unit procedures. The resulting flowsheet models could help capture process dynamics under process variation, and further identify design space and optimum operating conditions through system analysis and optimization [142]. Flowsheet model is an essential part of a complete DT, but it is considerably challenging to develop a flowsheet that can fully represent the physical plant due to the high complexity of the integrated process and the intense requirement of computational power resources. Consequently, current research on flowsheet development either focuses on coupling the downstream unit operations, or building the integrated process model in the commercial simulation software [143].

Several attempts have been observed in building the integrated process model for downstream operations. Coupling with the control system, the mechanistic models of different chromatographic systems are integrated to construct a flowsheet model to realize real-time decision making and automatically optimize the operations conditions [144]. In order to save computational time, data-driven and hybrid models are employed to build the flowsheet mode as well. Given the upstream product and antigen-binding fragment manufacturing process, Liu et al. [145] developed an integrated data-driven model for multiscale optimization to find the optimum chromatography decisions for purification like the number of chromatography columns and their sizes, the number of cycles per batch, and the operational flow velocities. Zahel et al. [146] established an integrated process model for three chromatographic steps with regression approaches

and then incorporated Monte Carlo simulation into the model, which can be used to estimate process capabilities, examine the product quality under process variations, and provide risk-based decision. Hybrid model is also developed by Pirrung et al. [147] to optimize the chromatography decisions in terms of the chromatography modes, operating sequence and conditions. Moreover, an integrated residence time distribution (RTD) model for an integrated continuous downstream process is developed to describe the inside mass flow, accelerate the start-up phase, and track the propagation of disturbances [148]. Recently, Rischawy et al. [149] first developed an integrated mechanistic model for the entire biopharma downstream process. This model incorporates capture, pH virus inactivation, depth filtration, adjustment, anion exchange chromatography, pH adjustment, cation exchange chromatography, virus filtration and ultrafiltration/diafiltration steps. The developed model has been calibrated on lab scale, validated at a manufacturing scale of 12,000 L, and tested to accurately respond to real input parameter variability in manufacturing scale, including protein concentration, volume, size variant composition, ion concentration, and pH.

On the other hand, flowsheet model for the biopharmaceutical process has been constructed in different process simulators, such as SuperPro Designer and BioSolve [150-152]. Although these modeling simulations are based on mass balance calculations and cannot provide detailed analyses of interconnections between different units, the simplified flowsheet models can help develop process design, evaluate various processing scenarios and achieve decision-making from the economic and environmental aspects, especially at the early-stage development. Petrides et al. [143] built an integrated flowsheet model for a batch biopharmaceutical process of mAbs in SuperPro Designer for process scheduling, debottlenecking and economic analysis.

Recently, a transition from batch to continuous biopharmaceutical operations has been noticed [98, 153, 154]. It is significant to assess the feasibility and viability of this transition and make decisions with the help of the flowsheet model. A lot of efforts have been made for the comprehensive comparison between different operating modes (such as batch, continuous and hybrid) and scenarios in terms of economic and ecological assessment [150, 155-158].

1.3 Challenges in Modeling Biopharmaceutical Processes

The fast development of digital twin capabilities and the successful attempts in engineering applications such as aerospace have drawn significant attention from the biopharmaceutical industry and the associated policymakers. Although building blocks of digital twin (e.g., models and PAT tools) and policy guidelines (e.g., United States' Executive Order on Advancing Biotechnology and Biomanufacturing Innovation [159], China's Five-Year Plan for Development of Smart Manufacturing [160], and European Union's Digitizing European Industry Initiative [161]) are in place, several challenges and gaps still exist, inhibiting the adoption of the technology in biopharmaceutical industry.

1.3.1 Development of Appropriate Models for Complex Biopharmaceutical Processes

In order to have an accurate virtual representation of digital twin in biopharmaceutical industry, having appropriate models to describe the physical biopharmaceutical manufacturing processes is critical for digital twin implementation. In current literature, even though various modeling methods have been developed and libraries of models are being incorporated into simulation software, their suitability for complex industrial development processes needs to be closely examined. For instance,

hydrophobic interaction chromatography is one of the widely used techniques in downstream polishing steps, but the binding mechanism inside is quite complex, which poses a challenge to the development of appropriate models that accurately describe its behavior. In addition, the developed models are highly dependent on the appropriate parameterization using data. In the early pharmaceutical development stage, the available data may be biased as they are not collected for model development, interfering with the evaluation and validation of the developed models. Such constraint in data quantity and data quality often leads to a reduced capability in using the models for process prediction and analysis.

1.3.2 Computational Complexity in System Analysis Applications

To keep pace with the advancements of artificial intelligence (AI) and digital twins and align with the quality by design initiative, the biopharma industry has attempted to embrace advanced model-based approach to expedite process development. Mechanistic models, rooted in prior process knowledge, have been employed for process characterization and optimization. Nonetheless, as these models evolve to consider precise representations of process dynamics, the computational complexity also increases, hindering the implementation of these models for system analysis like feasibility analysis and optimization. For instance, the design space for continuous frontal chromatography process [162] and twin-column countercurrent solvent gradient purification (MCSGP) [163] were identified based on mechanistic model by sampling the full operating space, but this approach would result in a tradeoff between model accuracy and computational cost. Increasing the sample size may result in improved accuracy of the identified design space, but it would increase the sampling cost simultaneously.

1.3.3 Proof-of-Concept for New Process Design

The commercial-scale production of biopharmaceuticals is currently performed in batches in which each unit operation is run sequentially, requiring human intervention to handle the product inflow and outflow between different unit operations [98, 164]. In addition to the large amount of workforce required, there are some other vital shortcomings associated with batch manufacturing, including low productivity and batch-to-batch variability in product quality [89, 165]. With the increasing global market demand, prohibitive drug cost, and biosimilar competition [166-168], batch operation has become inefficient and unable to meet those demands. Thus, a more cost-effective and flexible manufacturing platform is desired, such as continuous manufacturing. Conducting proof-of-concept (POC) or what-if study with physical implementation requires significant costs due to the investments in equipment, labor, capital, operating costs, etc. Thus, it is both necessary and advantageous to turn to *in silico* simulations to perform such feasibility studies by establishing end-to-end flowsheet models. Under the consideration of computational power limitations, it must be noted that for a fully integrated digital twin to be functional, the different unit operations models of various degrees of complexity need to be properly connected, enabling efficient information flow from upstream to downstream.

1.4 Research Goals

Considering the aforementioned challenges, the research in this dissertation aims to apply various process systems engineering tools to empower the *in silico* development of biopharmaceutical manufacturing processes. The objectives include improving model suitability/robustness, reducing computational effort, and supporting process assessments. For applications of PSE tools in supporting development, the

research specifically focuses on a key unit operation of downstream processing: chromatography, as illustrated in Figure 1.3. Chapters 2-4 introduce different modeling and system analysis methods incorporating machine learning (ML) algorithms to support process characterization and optimization of Protein A chromatography (ProA) in primary capture, as well as ion-exchange chromatography (IEX) and hydrophobic interaction chromatography (HIC) in polishing steps. Considering the trend in transitioning from batch to continuous processes, the research also targets to establish an end-to-end continuous platform for biomanufacturing, acting as an *in silico* proof-of-concept (POC) study for new processes in Chapter 5.

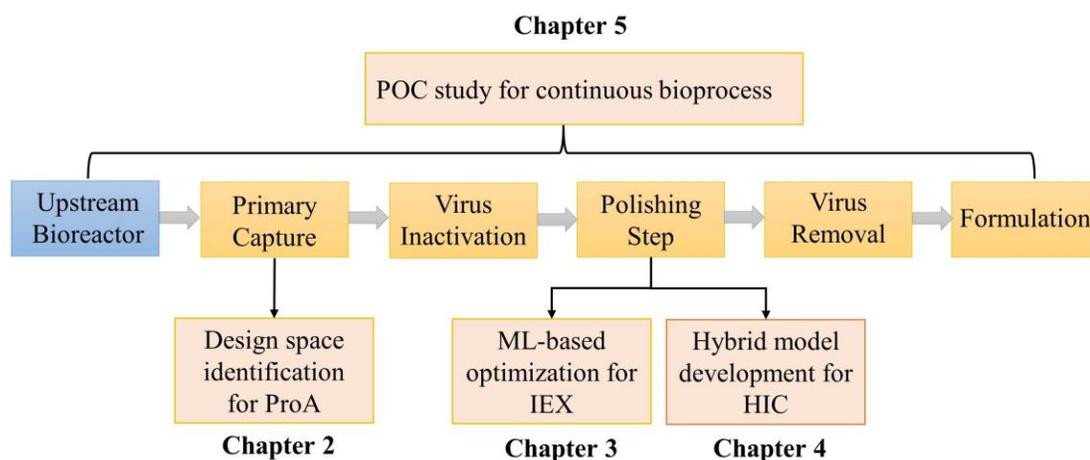


Figure 1.3: General overview of the research goals.

1.4.1 Design Space Identification for ProA

ProA is a standard procedure for initial capture to remove most of the process-related impurities, but it suffers from high resin cost. To address this issue, multi-column periodic chromatography has attracted wide attention because continuous

chromatography can obtain higher productivity and resin capacity utilization, and consume less of the buffer. Model-based approach is a valuable strategy to help determine the design space of the continuous capture process, facilitating process understanding and improving product quality.

To balance the computational complexity and model predictions, Chapter 2 aims to utilize a novel approach to identify the design space of continuous chromatography. Specifically, surrogate-based feasibility analysis with adaptive sampling is proposed and applied to establish the design space of the twin-column semi-continuous capture (CaptureSMB) process. The effects of process variables (including interconnected loading time, interconnected flow rate, and batch flow rate) on the design space are comprehensively examined based on an active set strategy. Besides, essential factors like recovery-regeneration time and constraints of column performance parameters (yield, productivity, and capacity utilization) are thoroughly investigated. The impact of design variables such as column length is also studied.

1.4.2 Machine Learning-based Optimization for IEX

Ion-exchange chromatography is an essential but complicated step in the biopharmaceutical downstream process, with multiple factors affecting the separation efficiency. Model-based optimization can help expedite process developments with limited time and resource investments. However, employing mechanistic models directly for optimization often encounters challenges due to high computational complexity. Additionally, finding optimal solutions in highly nonlinear and nonconvex processes can be difficult.

Therefore, Chapter 3 aims to apply a machine learning-based optimization framework to address the abovementioned challenges. Specifically, Gaussian Process

Regression (GPR) models are utilized as substitutes for the mechanistic model in calculating the constraints and objective function. To further reduce the required sampling, feasibility and optimization stages are incorporated into the framework, with a penalty introduced at each stage into the expected improvement function to guide the search process. This ML-based framework is applied to a case study to separate a ternary protein mixture using IEX. The effects of peak cutting criteria on the optimal results were examined, followed by a detailed analysis of design space.

1.4.3 Hybrid Model Development for HIC

HIC is often employed as a polishing step to remove aggregates for the purification of therapeutic proteins in the biopharmaceutical industry. To accelerate the process development and save the costs of performing time- and resource-intensive experiments, advanced model-based process design and optimization are necessary. Due to the unclear adsorption mechanism of the salt-dependent interaction between the protein and resin, the development of an accurate mechanistic model to describe the complex HIC behavior is challenging.

Therefore, Chapter 4 focuses on utilizing a hybrid modeling strategy to develop accurate models that represent the HIC process. To reduce the development effort of isotherm equations and extract missing information from the available data, a hybrid model is constructed by combining a simple and well-known multi-component Langmuir isotherm with a neural network. The accuracy of the developed hybrid model is compared with that of mechanistic model. Additionally, an extrapolation capability test was conducted to assess the generalizability of the hybrid model. Process optimization is also carried out to find the optimal operating conditions under product quality constraints using the developed hybrid model.

1.4.4 Proof-of-Concept Study for Continuous mAb Production

The most widely used method for biological production of mAbs is operated in batch operation mode. However, due to the fast expansion of market demand, continuous biomanufacturing emerges as a promising alternative to current batch operation as it offers benefits in terms of improved productivity, product quality, and reduced footprint.

Chapter 5 aims to conduct a proof-of-concept study by constructing a fully integrated *in silico* end-to-end continuous platform to evaluate the feasibility of transitioning from batch to continuous biomanufacturing processes. During the process design phase, innovative technologies like intensified seed expansion and continuous high cell density perfusion operations, single-pass tangential flow filtration, and single-use technologies are integrated, alongside media and buffer preparation steps, are incorporated. After the development of flowsheet model, economic and environmental analyses are conducted to assess the economic efficiency and environmental footprints. Moreover, scenario analysis is performed to assess the impacts of bioreactor scale and upstream titers on the process economics as well as on the environmental footprint. Process bottlenecks are identified and eliminated, and the integration of membrane chromatography is also examined.

Chapter 2

DESIGN SPACE IDENTIFICATION OF CONTINUOUS PROTEIN A CHROMATOGRAPHY

2.1 Introduction

The treatment of cancer and autoimmune diseases based on monoclonal antibodies (mAbs) has been identified as one of the most successful strategies [6, 169, 170]. Currently, the most widely used method for biological production of mAbs is batch operation mode in which each unit operation is operated in sequence, and the outflow of the previous unit is collected in a holding tank before transferring to the next unit [13, 98, 164, 171]. However, due to the rapid growth of market demand, continuous bioprocessing becomes a potential alternative for the production of mAbs by providing higher productivity and yield and smaller footprints [57, 99, 150]. Significant progress has been achieved in the upstream process like cell culture, which shifts the pressure of achieving continuous biopharmaceuticals to the downstream operations, especially for primary capture [31], which has not however received as much attention in the literature. However, capture is sometimes combined with upstream processing and harvest in

Chapter 2 is adapted from the following journal article written by the author of this dissertation:

Ding, C., Ierapetritou, M., A novel framework of surrogate-based feasibility analysis for establishing design space of twin-column continuous chromatography. *International Journal of Pharmaceutics*, 2021, 609: 121161.

determining strategies for process integration, which puts extra pressure on delivering high productivity in the primary capture step.

Protein A affinity chromatography is a standard procedure for the initial capture of almost all mAb products to remove most of the process-related impurities like culture media components, host cell proteins, and DNA [34]. To address the issue of high resin cost, continuous capture process, like multi-column periodic counter-current chromatography (PCC), has been developed, because continuous chromatography can obtain higher productivity and resin capacity utilization, and consume less of the buffer [9, 172, 173]. Several continuous capture chromatographic systems are available with different operation modes and column numbers, such as twin-column CaptureSMB [76, 174], 3-column and 4-column PCC [175-177], BioSMB process [178] combining PCC with disposal technology, and sequential multi-column chromatography (SMCC) [179].

In an attempt to acquire thorough process understanding of continuous biomanufacturing, FDA introduced the Quality by Design (QbD) initiative to provide regulatory flexibility and manufacturing efficiency to produce biotherapeutics with higher product quality [180-182]. The process understanding acquired from QbD initiative is beneficial for the identification of the design space, which is determined by the ranges in which a given process must operate to ensure product quality [53, 54]. Design space analysis can help comprehend the main effects of process variables and their interactions on product quality, facilitating the improvement of continuous bioprocess. Besides the experimental examination of the design space based on the design of experiments (DOE) [183], modeling can be utilized to lower the experimental work and delve into the effects of critical process parameters (CPPs) on the critical quality attributes (CQAs) [74]. Morbidelli's group has identified the design space of

continuous frontal chromatography process [162] and twin-column countercurrent solvent gradient purification (MCSGP) [163] based on the mechanistic models and pointed out the existing tradeoffs between model accuracy and computational complexity.

In this work, to reduce the computational burden introduced by the mechanistic models, surrogate-based feasibility analysis with an adaptive sampling approach [184] is proposed to approximate the original models and predict the design space. The introduction of surrogate model results in a reduced computational cost, but the accuracy of the surrogate model is highly dependent on the quality and quantity of the sampling set [108]. To deal with this issue, an adaptive sampling algorithm is utilized because this algorithm can help search towards the less explored regions and the boundary of the feasible region [185]. To validate the applicability of the proposed methodology, twin-column CaptureSMB is used as a case study although the approach is general to be used for any potential configuration. In summary, mechanistic models to describe the continuous capture process are first constructed, followed by establishing the surrogate models built on the relationship between the inputs (process variables including interconnected loading time t_{IC} , interconnected flowrate Q_{IC} and batch flowrate Q_B) and the outputs (the maximum value among productivity, yield, and capacity utilization constraints) acquired from the mechanistic models. After determining the design space of each case study, the impacts of each process variable on the design space are thoroughly investigated based on the active set strategy. In addition, crucial factors that affect the design space, including recovery-regeneration time and constraints of column performance like yield, productivity, and capacity

utilization, are also comprehensively examined. The influence of one of the most important design variables, the column length, is also investigated.

2.2 Process Description and Process Model

2.2.1 Operating Principles of CaptureSMB

A full cycle of the twin-column cyclic chromatography (CaptureSMB) process in steady-state is illustrated in Figure 2.1. Each cycle comprises two “switches” with an interconnected (IC) and a batch (B) step. During the IC phase (Step 1), two columns are connected in series and column 1 is overloaded beyond its dynamic binding capacity with the interconnected flowrate Q_{IC} and loading time t_{IC} . The breakthrough from the first column is captured by the second column. Once the first column is fully loaded, the two columns are washed in sequence to flush the unbound material into the second column using a wash buffer. During the batch phase (Step 2), the two columns are disconnected. Column 1 undergoes the recovery and regeneration step (R-R, including washing, elution, clean-in-place, and re-equilibration) while loading continues using column 2 with the feeding flowrate Q_B and loading time t_B . To ensure that this system operates in continuous mode, the R-R time (t_{RR}) is set to be equal to the batch loading time (t_B). At the end of the switch, the two columns switch their functions to perform similar steps (Steps 3 and 4) to complete one cycle. In summary, based on the operating principles described, the three process variables - interconnected loading time t_{IC} , interconnected flowrate Q_{IC} , and batch flowrate Q_B - are the ones defining the process design space and are considered for further system analysis and process optimization.

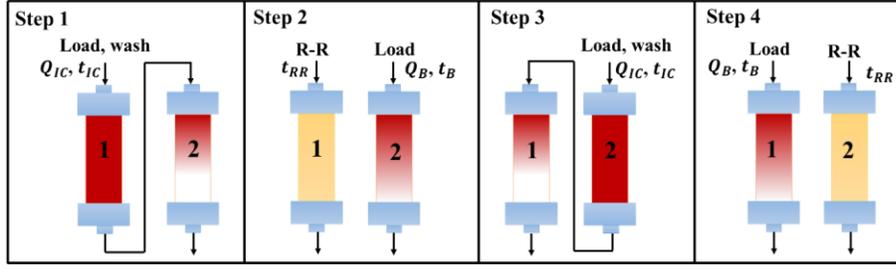


Figure 2.1: Schematic diagram of a CaptureSMB system for continuous capture.

2.2.2 Process Model

Lumped kinetic model (LKM) with Langmuir isotherm [81] is employed to describe the mass transport and kinetic adsorption inside the column, with the assumption of isothermal adsorption and radial homogeneity. Although the model used for this process is an approximate model, the main purpose of this work is to illustrate the application of the novel framework in identifying the design space of continuous chromatography. More detailed and accurate models like general rate model [186] or shrinking core model [136, 187] will be considered in our future work. The mass balance for the mobile phase and stationary phase of LKM are illustrated in Equations 2.1 and 2.2, respectively.

$$\frac{\partial c}{\partial t} = -u \frac{\partial c}{\partial z} + D_a \frac{\partial^2 c}{\partial z^2} - \frac{1-\varepsilon_c}{\varepsilon_c} \frac{\partial q}{\partial t} \quad 2.1$$

$$\frac{\partial q}{\partial t} = k_m (q^* - q) \quad 2.2$$

where c and q are the protein concentrations in the mobile and stationary phases, respectively; t and z denote the time and axial coordinate; ε_c is the total porosity; $u = Q/(A_{col}\varepsilon_c)$ is the superficial velocity; D_a is the apparent axial dispersion coefficient; k_m is the lumped mass transfer coefficient; and q^* is the equilibrium solid phase concentration of the protein.

Equations 2.3–2.4 shows the Danckwerts’ boundary conditions employed in the process [84]. At the column inlet, it is presumed that the change in axial concentration relates to the difference between the feed and the column inlet concentration. At the column outlet, there is no axial concentration gradient. The column is assumed to be initially free of proteins in both the mobile and stationary phase, as described in Equations 2.5–2.6. In addition, the apparent axial dispersion coefficient D_a can be estimated based on the reduced van Deemter relation as shown in Equations 2.7.

$$u c_{in}(t) = u c(t, 0) - D_a \frac{\partial c}{\partial z}(t, 0) \quad 2.3$$

$$\frac{\partial c}{\partial z}(t, L) = 0 \quad 2.4$$

$$c(0, z) = 0 \quad 2.5$$

$$q(0, z) = 0 \quad 2.6$$

$$D_a = A \frac{d_p}{2} u \quad 2.7$$

where A is the van Deemter coefficient and d_p is the average resin particle diameter, respectively. Note that $c_{in}(t)$ is time-dependent protein inlet concentration because of the different procedures (breakthrough uptake, load, elution, etc.) in the column.

For the lumped mass transfer coefficient k_m , an empirical correlation is employed in Equation 2.8 [188]. The adsorption mechanism is described using Langmuir isotherm as displayed in Equation 2.9.

$$k_m = k_{max} \left[S_1 + (1 - S_1) \left(1 - \frac{q}{q_{sat}} \right)^{S_2} \right] \quad 2.8$$

$$q^* = \frac{H \cdot c}{1 + \frac{H \cdot c}{q_{sat}}} \quad 2.9$$

where k_{max} is the maximum lumped mass transfer coefficient; q_{sat} is the saturation capacity of the resin; S_1 and S_2 are the saturation dependent kinetic constant and order; H is the Henry coefficient.

In terms of parameter estimation, van Deemter coefficients A and B are estimated by linearly fitting the reduced plate height against the mobile phase velocity. Mass transfer (k_{max} , S_1 , S_2) and adsorption parameters (q_{sat} and H) are obtained by inversely fitting the batch breakthrough curves (BTCs) under different feeding concentrations and flowrates. To solve the coupled partial differential equations, those equations were first discretized into 100 points along the axial coordinate using a first-order central finite differences method, followed by using the solver *ode15s* in *MATLAB* to resolve the obtained system of ordinary differential equations.

In this work, all the parameters to simulate the loading step of the continuous capture process are directly obtained from the literature [81] and listed in Table 2.1. It should be noted that the elution process is not considered in this work and the column is assumed to be empty after the recovery and regeneration step. Although this is an approximation of the entire process, the work presented here can be extended to capture those additional steps that will be considered in our future work.

Table 2.1: A list of used model parameters.

Parameters	Units	Value
A	-	35.13
ε	-	0.368
k_{max}	min^{-1}	0.1800
S_1	-	0.6245
S_2	-	2.071
q_{sat}	mg mL^{-1}	92.2816
H	-	246.8

Table 2.1 continued.

d_p	μm	44
γ	bar min cm ⁻²	0.01020

2.2.3 Evaluation of Process Performance

Productivity, resin capacity utilization, purity, and yield are four important metrics to evaluate the column performance of continuous chromatography. It is assumed that the purity requirements can be fulfilled under the current operating conditions and will not be considered when determining the design space [81]. Thus, only three criteria - yield, productivity, and capacity utilization - are examined to identify the design space. The yield (Y) of the continuous capture process is defined by the amount of protein obtained relative to the theoretical maximum amount obtainable, which can be calculated from Equation 2.10. The productivity (P) is equivalent to the amount of target protein produced per unit time and unit resin volume, which is given in Equation 2.11. The resin capacity utilization (CU) can be defined as the amount of target protein that can be produced per unit resin volume and cycle normalized with the equilibrium binding capacity at the feed concentration, as shown in Equation 2.12.

$$Y(t_{IC}, Q_{IC}, Q_B) = 1 - \frac{2(Q_{IC} \int_0^{t_{IC}} c_{out}(t) dt + Q_B \int_0^{t_B} c_{out}(t) dt)}{n c_{Feed} (t_{IC} Q_{IC} + t_B Q_B)} \quad 2.10$$

$$P(t_{IC}, Q_{IC}, Q_B) = Y \frac{c_{Feed} (t_{IC} Q_{IC} + t_B Q_B)}{n V_{Col} (t_{IC} + t_B)} \quad 2.11$$

$$CU(t_{IC}, Q_{IC}, Q_B) = Y \frac{c_{Feed} (t_{IC} Q_{IC} + t_B Q_B)}{V_{Col} (1 - \epsilon_c) q_{Feed}^*} \quad 2.12$$

where c_{Feed} is the feeding protein concentration; V_{Col} is the volume of one column; n represents the number of columns; c_{out} denotes the column outlet concentration, which is obtained by solving the coupled PDE equations in Equations 2.1–2.2.

2.3 Methodology

2.3.1 Feasibility Analysis

Feasibility can be mathematically quantified by the definition of feasibility function $\psi(d, \theta)$ given in Equation 2.13. The feasibility function is utilized to illustrate whether a given process can satisfy all constraints f_j by merely modifying control variable z at fixed uncertain parameters θ .

$$\psi(d, \theta) = \min_z \max_{j \in J} \{f_j(d, z, \theta)\} \quad 2.13$$

where d are the design variables like equipment size, z represents the control variables that vary on the range $Z = \{z: z^L \leq z \leq z^U\}$, θ denotes the uncertain parameters $\theta \in T = \{\theta: \theta^L \leq \theta \leq \theta^U\}$, $f_j(d, z, \theta), j \in J$ corresponds to the set of constraints restricting the feasible space. If $\psi(d, \theta) > 0$, it suggests that the process design is infeasible because of the violation of some constraints whereas $\psi(d, \theta) \leq 0$ means that the given process is feasible and $\psi(d, \theta) = 0$ represents the boundary of the feasible region.

In this article, no control variables are considered in the case study, which leads to the simplified feasibility function as shown in Equation 2.14.

$$\psi(d, \theta) = \max_{j \in J} \{f_j(d, \theta)\} \quad 2.14$$

In the twin-column continuous capture process, since we are interested in evaluating the design space of feasible operation, the process variables θ include the interconnected loading time t_{IC} , interconnected flowrate Q_{IC} and batch flowrate Q_B ,

while the design variable d represents column length L . The design space of continuous PCC process is restricted by yield (Y), capacity utilization (CU), and productivity (P) constraints. Specifically, Y, CU and P are set to be higher than or equal to 80%, 85%, and 15 mg mL⁻¹ h⁻¹ in the case study, respectively.

2.3.2 Surrogate Model - Kriging

The surrogate model is employed to replace the original mechanistic models of continuous chromatography procedures to reduce the computational expense. Specifically, the surrogate model is trained based on the inputs (process variables like t_{IC} , Q_{IC} , and Q_B) and the outputs (feasibility function values calculated from mechanistic model).

Kriging is a commonly used surrogate model that was first developed by mining engineer Danie Krige. *Ordinary kriging* is a spatial interpolation method relying solely on the point observations of the target variable. In specific, this approach makes predictions of un-visited points by utilizing the sum of the Euclidean distance of the observed function values at nearby sampling locations [189]. It assumes that the variation is random and only related to spatial distance with a constant mean. An alternate to *ordinary kriging* is *regression kriging* [190, 191], which is based on a hybrid interpolation technique combining point observations and regression of target variables to estimate the un-sampled locations. In this approach, regression is first used based on auxiliary information, followed by employing simple kriging with a known mean of 0 to interpolate the residuals from the established regression model. *Regression kriging* was chosen in our calculations because this method can produce better predictions [190, 192]. The general form of a *regression kriging* model can be formulated in Equations 2.15–2.16.

$$\hat{f}(x) = \mathbf{f}(x)^T \boldsymbol{\beta} + \varepsilon(x) \quad 2.15$$

$$\mathbf{f}(x)^T \boldsymbol{\beta} \equiv \beta_1 f_1(x) + \dots + \beta_m f_m(x) \quad 2.16$$

where $\mathbf{f}(x)^T \boldsymbol{\beta}$ is a linear regression model constituted by m known basis functions $f_j(x)$ ($j = 1, 2, \dots, m$) that define the global trend of mean prediction at location x and m unknown parameters β_j ($j = 1, 2, \dots, m$); $\varepsilon(x)$ is a residual term at x that is commonly distributed with zero mean and covariance σ^2 , as shown in Equation 2.17.

$$\text{Cov}(\varepsilon(x^i), \varepsilon(x^j)) = \sigma^2 R(x^i, x^j) \quad 2.17$$

where R represents the correlation model.

Compared with other surrogated models, the benefit of the Kriging model lies in that it can provide the estimated variance of the prediction [193], which can be useful for adaptive sampling. There are lots of regression and correlation models that can be employed to fit the Kriging model. In this work, three regression models and three correlation models shown in Table 2.2 are tested for the model training because these models exhibit good performance in our previous work [185]. The regression coefficients $\boldsymbol{\beta}$ are estimated from least-square estimator and the hyper-parameter (θ_j) is predicted using maximum likelihood. More detailed introduction and derivation about kriging can be found in this review [108]. In this paper, Kriging model is constructed through the DACE toolbox [194] built-in *MATLAB*. In our simulation, the combination of quadratic regression & exponential correlation gives us the best model performance.

Table 2.2: Regression models and correlation functions with $d_j = x_j - x_i$ and θ_j .

Regression model	$\mathbf{f}(x)$	Correlation function	$R_j(\theta, d_j)$
Constant	a	Exponential	$\exp(-\theta_j d_j)$
Linear	$a_i x_i + b$	Gauss	$\exp(-\theta_j d_j^2)$
Quadratic	$a_i x_i^2 + b_{ij} x_i x_j + c_i x_i + d$	Linear	$\max\{0, 1 - \theta_j d_j \}$

2.3.3 Adaptive Sampling

Adaptive sampling can help improve the accuracy of the surrogate model and reduce sampling cost simultaneously. By maximizing a modified expected improvement (EI) function [195] in Equation 2.18, an adaptive sampling strategy is in favor of searching towards the areas with high uncertainty and close to the boundary.

$$\max EI_{fea}(x) = s\phi\left(-\frac{y}{x}\right) = s \cdot \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{0.5y^2}{s^2}\right) \quad 2.18$$

where $EI_{fea}(x)$ is the modified expected improvement (EI) function value at x ; y and s represent the surrogate model predictor and corresponding standard error (\sqrt{MSE}), respectively; $\phi(\cdot)$ denotes the normal probability distribution function.

The principle behind this maximization problem can be explained based on the partial derivatives:

$$\frac{\partial EI_{feas}}{\partial s} = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{0.5y^2}{s^2}\right) \left(1 + \frac{y^2}{s^2}\right) > 0 \quad 2.19$$

$$\frac{\partial EI_{feas}}{\partial y} = -\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{0.5y^2}{s^2}\right) \frac{y}{s} \quad 2.20$$

It is found that the derivative of EI_{feas} with respect to s in Equation 2.19 is always larger than 0, suggesting that EI_{feas} monotonically increases as the prediction error s increases. Consequently, maximizing EI function can help sample new points in the unexplored region. From Equation 2.20, it should be noted that $\partial EI_{feas}/\partial y$ is negatively related to the surrogate predictor y . When $y < 0$, this derivative is positive, implying that EI_{feas} increases with the increase of y . Maximizing EI_{feas} will increase y and force it to approach zero. On the contrary, if $y > 0$, this derivative is negative, and maximization will decrease y and push it to get close to zero. Thus, Equation 2.20 suggests that maximizing EI function favors searching towards the feasible region

boundary. That is how adaptive sampling works to improve the model accuracy without exhaustively sampling the whole space.

To maximize this modified EI function, a local optimization solver *fmincon* is utilized. The performance of *fmincon* depends on the initial guess, so Latin Hypercube DOE is employed to generate 1000 different points, followed by evaluating the modified EI function values at those points. The point with the largest value is chosen as the initial guess of optimization. Although 1000 points are selected, it is still very computationally efficient because the optimization is based on the generated surrogate model, rather than the computationally expensive original model. It was found that only 0.49 seconds are required to perform 1000 calculations and 0.062 seconds for the optimization based on the EI function in *MATLAB R2019b*. All the simulations were conducted on Intel® Xeno® E-2274G CPU with 32.0 GB RAM.

2.3.4 Surrogate Model Accuracy

Since surrogate model is an approximation of the original model, we want to evaluate the accuracy of the feasibility analysis. In this article, three metrics [185, 195], namely percentage of Correct Feasible region (CF%), percentage of Correct Infeasible region (CIF%), and percentage of Not Conservative feasible region (NC%), are used as performance measures. To be specific, CF% means the percentage of feasible regions in the original function which has been correctly identified by the surrogate model; CIF% represents the percentage of infeasible regions in the original function that has been properly identified by the surrogate model; NC% illustrates the percentage of overestimated feasible region by the surrogate model. Accordingly, if CF% and CIF% are close to 100% and NC% is close to 0, the surrogate model can approximate the original function with very high accuracy.

The relationship between those metrics is shown in Figure 2.2. The rectangle area is the entire range of two-dimensional uncertain parameters; the blue and green circles illustrate the feasible region of the original function and surrogate model, respectively. With the aim of calculating the three metrics, the entire space is split into four zones: CF (Correct Feasible region in orange); CIF (Correct InFeasible region in white); ICF (InCorrect Feasible region in green); ICIF (InCorrect InFeasible region in blue). Thus, the three accuracy measures can be calculated in Equations 2.21-2.23.

$$CF\% = \frac{CF}{CF+ICIF} 100 \quad 2.21$$

$$CIF\% = \frac{CIF}{CIF+ICF} 100 \quad 2.22$$

$$NC\% = \frac{ICF}{ICF+CF} 100 \quad 2.23$$

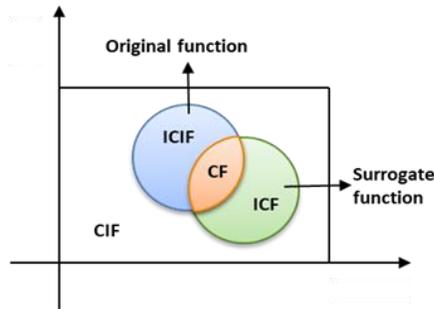


Figure 2.2: A schematic figure for the model performance metrics.

2.3.5 Proposed Methodology

The basic idea of the proposed strategy is shown in Figure 2.3. Firstly, initial sampling points are selected based on a space-filling design of experiment (DOE) to construct the initial Kriging model. Specifically, a rectangular grid sampling plan is chosen for the initial DOE to cover the input space uniformly. Subsequently, new

sampling points are added to iteratively update the Kriging model based on the adaptive sampling strategy. The adaptive sampling will stop if the number of iterations exceeds the user-defined maximum iterations. In this work, 800 is chosen as the maximum number of iterations because the feasible regions will not change as the number of iterations continues to increase, as shown in Figure 2.4. In addition, a plot of EI function with iterations is also provided in Figure 2.5 as quantitative evidence to show that the estimated feasible region has converged. The final Kriging model is used to perform feasibility analysis to predict the feasible region.

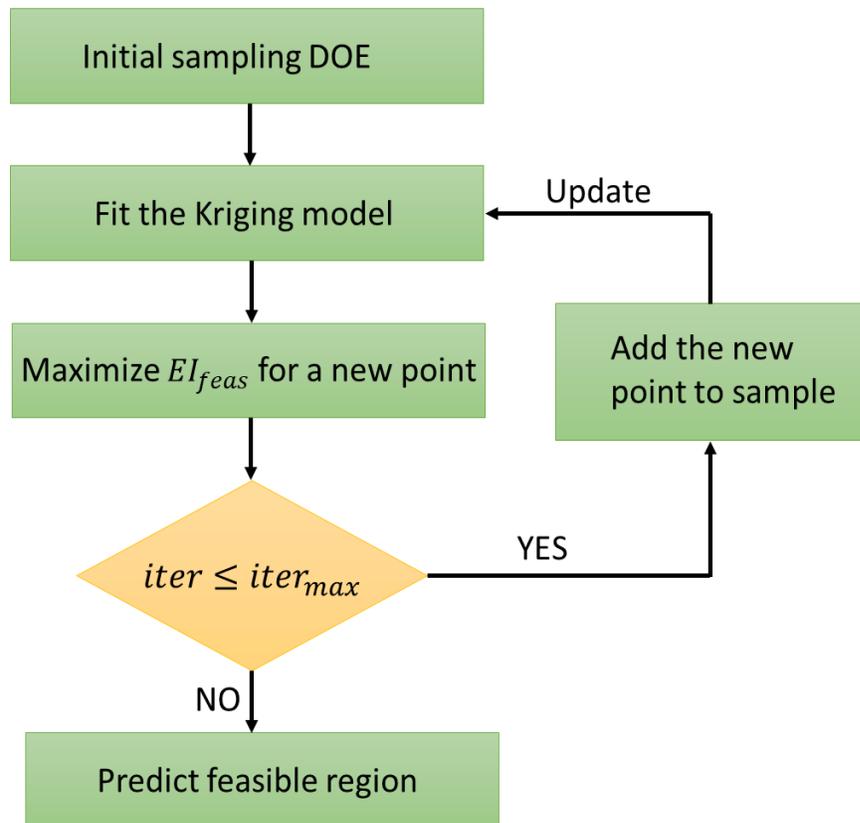


Figure 2.3: Framework of Kriging-based feasibility analysis with adaptive sampling.

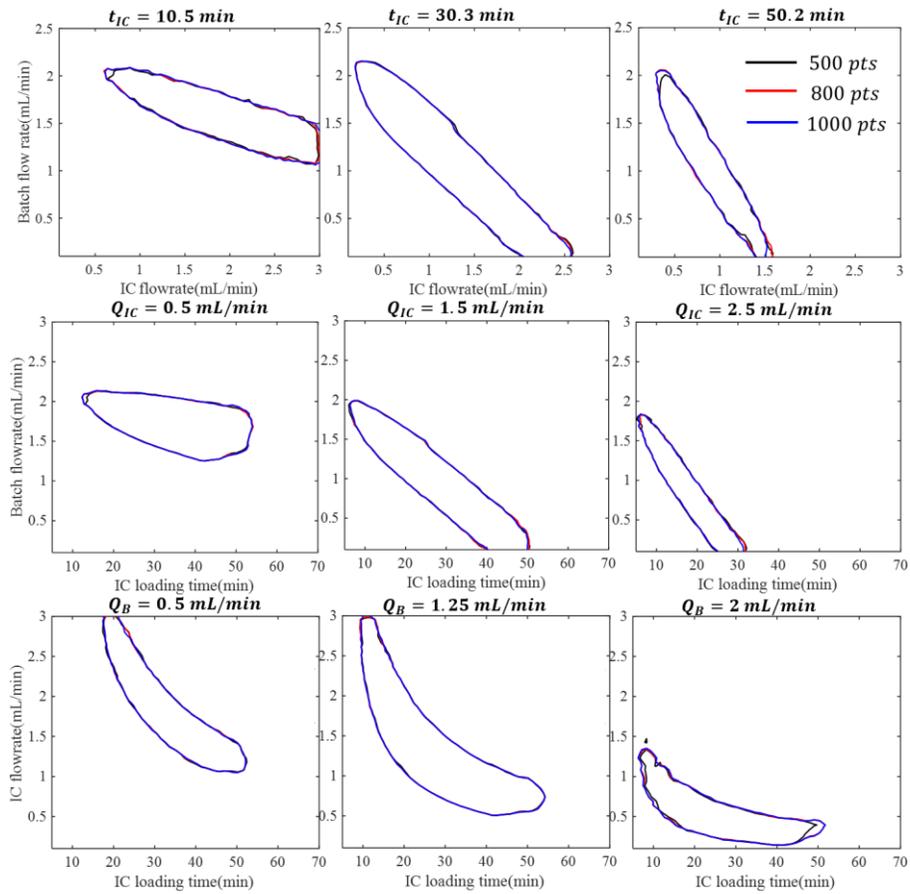


Figure 2.4: Comparison of 2D feasibility plots of CaptureSMB process at different iteration numbers for adaptive sampling.

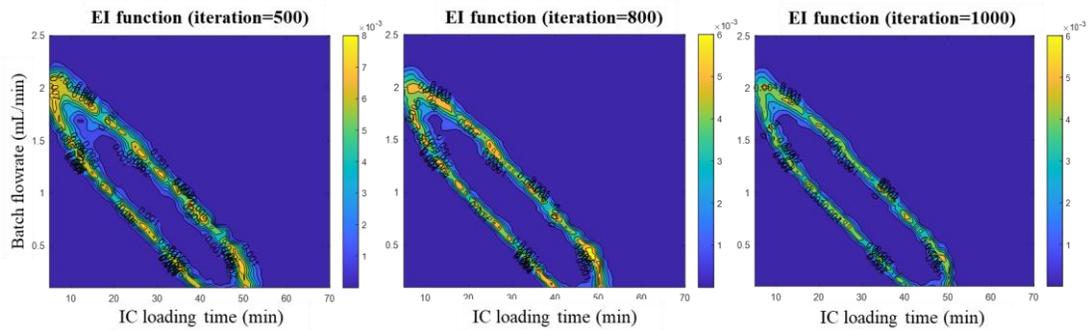


Figure 2.5: Comparison of EI function plot at different iteration numbers when $Q_{IC} = 1.5 \text{ mL min}^{-1}$.

2.4 Results and Discussion

2.4.1 Model Validation

The model parameters are directly obtained from the literature [81], but the breakthrough curves are still tailored to validate our developed model. The experimental data are extracted from [81] and the breakthrough curves under different feed concentrations and flowrates are fitted as shown in Figure 2.6. The root-mean-square error (RMSE) of each simulation ranges from 3.47% to 11.30% with an average of 5.77%. This developed process model is applicable for simulating the basic trend of the elution behavior, although there are some deviations existing between the fitted curves and the experimental data.

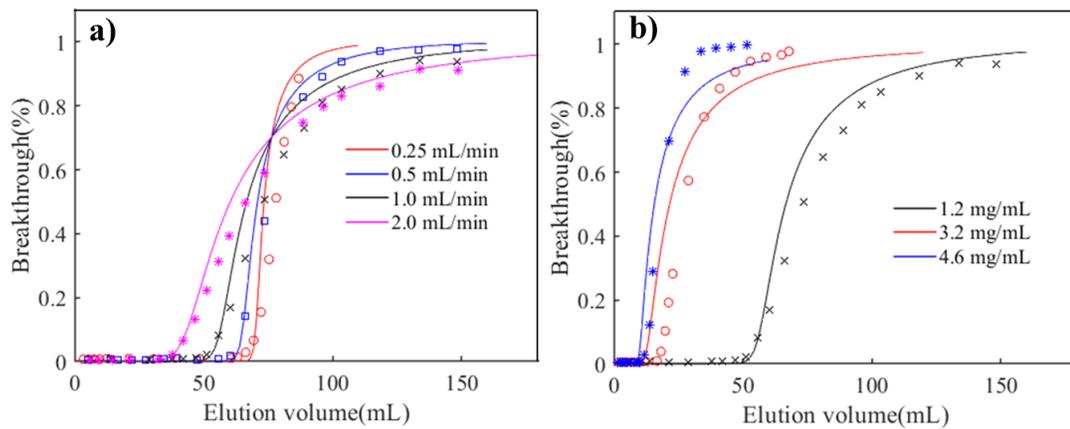


Figure 2.6: a) fitted breakthrough curves (line) at different flowrates with experimental data (dot) at the feed concentration of 1.2 mg mL^{-1} ; b) fitted breakthrough curves at different feed concentrations with experimental data. The flowrate is 1.0 mL min^{-1} for the runs with feeding concentrations of 1.2 and 4.6 mg mL^{-1} , whereas the flowrate is 1.5 mL min^{-1} for the experiment with a feeding concentration of 3.2 mg mL^{-1} .

2.4.2 Original Design Space

A total of 60^3 sampling points were used to construct the three-dimensional (3D) design space, as shown in Figure 2.7a. The mechanistic model requires nearly 72 hours to finish 60^3 calculations, but it is much more computationally efficient using the proposed framework. The time based on the surrogate model comprises two parts: the time to run the surrogate-based feasibility algorithm to build the kriging, and the time to run 60^3 calculations of the final kriging model. Specifically, the first part (which includes time of 1000 calculations on the original simulations and time of optimizing the EI function) takes roughly 1.7 hours, while the second part only requires 0.2 hours. The space with the feasibility function value less than or equal to 0 corresponds to the design space as described in Section 2.3.1. To clearly observe the design space, Figure 2.7b illustrates only the feasible region of the CaptureSMB process, which is located within the plotted circles.

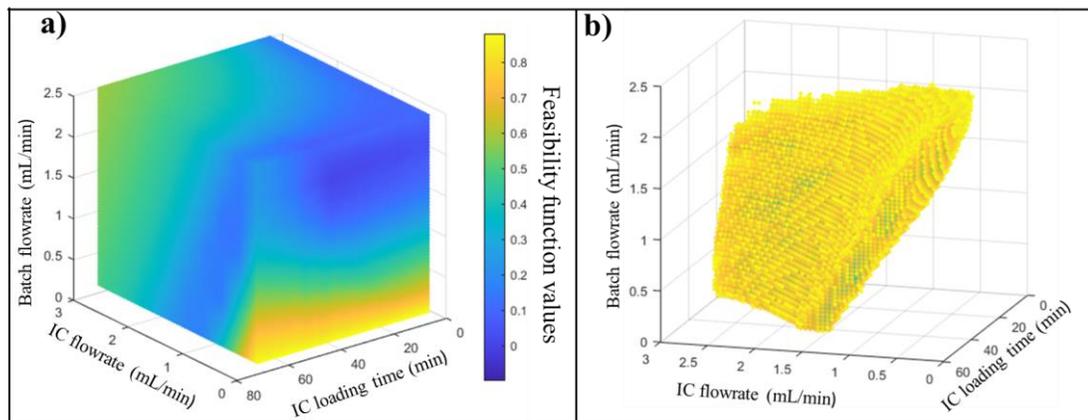


Figure 2.7: a) 3D contour plot showing the whole input space; b) 3D contour plot only showing the feasible region.

In order to visualize the change of feasible region with respect to each process variable, it is convenient to project the 3D feasible region onto 2D subspaces. 2D contour plots of feasibility analysis at nine different fixed input parameters are shown in Figure 2.8. The blue line of each subplot is the feasible region boundary, and the area within the blue represents the feasible region. When 3D feasible region is projected at different fixed interconnected loading time in Figure 2.8a–c, it should be noticed that the feasible region shifts towards the bottom left corner, i.e., smaller interconnected flowrate Q_{IC} and batch flowrate Q_B , with the increase of t_{IC} . In addition, the area of the feasible region that satisfies the column performance constraints first increases and then decreases, indicating that appropriate process variables should be selected in order to maintain a larger design space. The same trend can be observed with the increase of Q_{IC} (Figure 2.8d–f) or Q_B (Figure 2.8g–i) in the CaptureSMB case study. In order to analyze the reasons for this change, an active set strategy is employed. More specifically, 11 data points are selected at the feasible boundary when t_{IC} is set at 30.3 min, followed by calculating each constraint function value of those data points at t_{IC} of 10.5 and 50.2 min. If this calculated constraint function value is larger than zero, this constraint can be regarded as active (binding), meaning that this constraint plays an important role in initiating the change of the feasible region. It is found that productivity and yield constraints are binding with the increase of t_{IC} , while productivity and capacity utilization constraints are active with the decrease of t_{IC} . The same conclusion can be achieved when we analyze the change of the feasible region with respect to the other two process variables (Q_{IC} or Q_B). The active constraints introducing the changes of feasible regions are summarized in Table 2.3. It is necessary to evaluate the accuracy of the surrogate model, so the corresponding performance metrics of each subset plot in

Figure 2.8 are summarized in Table 2.4. It can be observed that all the CF% and CIF% values are higher than 97% (close to 100%) and NC% is smaller than 2.66% (close to 0), indicating that the developed surrogate model can predict the feasible region with high accuracy.

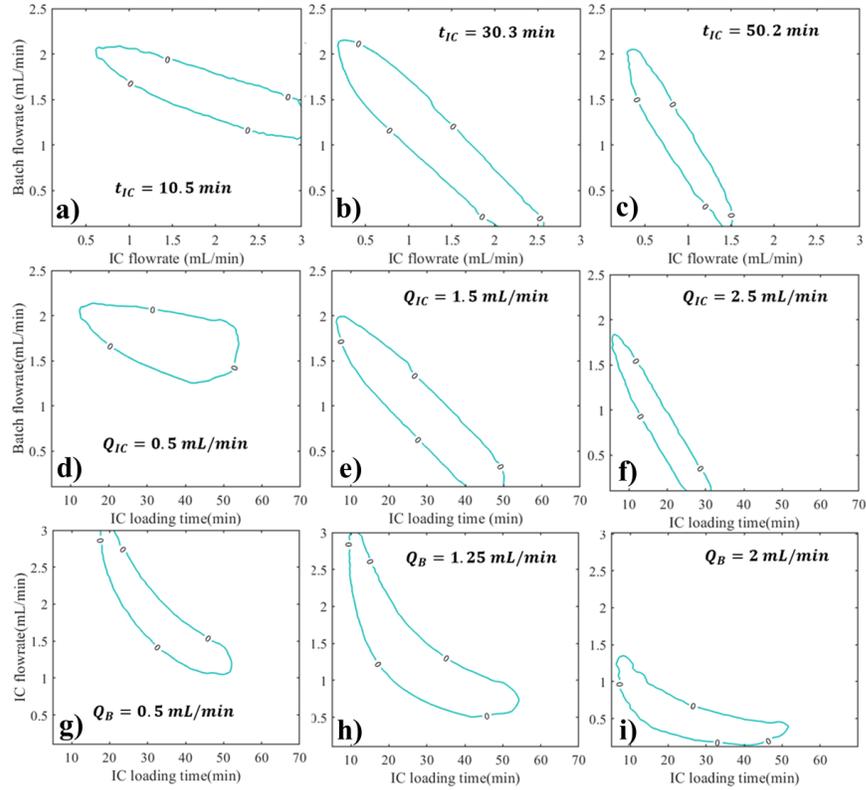


Figure 2.8: 2D projections of the 3D feasible region at different values of the process variables for the CaptureSMB process.

Table 2.3: Summary of active constraints when changing process variables.

Process variables (t_{IC}, Q_{IC}, Q_B)	Shift	P constraint	CU constraint	Y constraint
↓	↗	✗	✗	
↑	↙	✗		✗

Table 2.4: Accuracy metrics at different fixed process variables in Figure 2.8.

2D feasible region	Fixed value	CF%	CIF%	NC%
Fixed t_{IC} (min)	10.5	97.52	99.68	2.11
	30.3	99.04	99.43	2.29
	50.2	96.80	99.74	2.66
Fixed Q_{IC} (mg/mL)	0.5	99.26	99.74	1.47
	1.5	97.76	99.67	1.70
	2.5	97.61	99.81	2.53
Fixed Q_B (mg/mL)	0.5	97.73	99.78	1.82
	1.25	98.01	99.62	1.99
	2	97.06	99.44	2.10

An interesting phenomenon is observed at large t_{IC} as presented in Figure 2.9. When t_{IC} is smaller than 45.8 min (Figure 2.9a), the feasible region shifts towards the bottom left corner with the increase of t_{IC} to satisfy the yield constraint, which conforms with the above analysis. Nonetheless, when t_{IC} is close to the maximum feasible value, a different trend is recognized, i.e., the feasible region shrinks inwards as t_{IC} increases. Based on the active set strategy, Y and P constraints are active on the right feasible boundary, while only P constraint is binding on the left feasible boundary. The active Y constraint on the right boundary makes it move to the bottom left corner to prevent product loss, demonstrating the same trend at small t_{IC} . However, at large t_{IC} , the change of active constraints on the left feasible boundary from Y and P constraint to P constraint contributes to the opposite shrinking trend. Specifically, the left boundary shifts to the right to increase flowrates, thus increasing the produced sample load inside the column and further improving productivity.

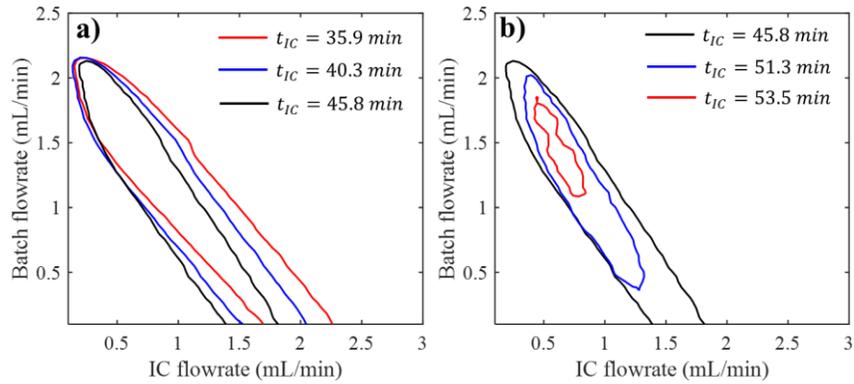


Figure 2.9: 2D contour plots at different fixed t_{IC} for the CaptureSMB process.

2.4.3 Effect of t_{RR} on Design Space

Recovery and regeneration time t_{RR} is an important factor with a great impact on the design space. In the CaptureSMB process, t_{RR} dictates the batch loading duration t_B , further influencing other process variables and column performance. Thus, the effect of t_{RR} on the design space of the CaptureSMB process is investigated and the comparison of 3D feasible regions is illustrated in Figure 2.10. The evident difference of 3D feasible regions can be discerned with the change of t_{RR} .

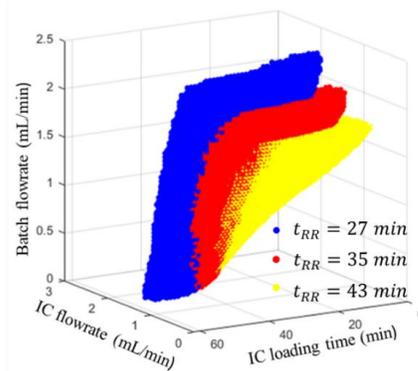


Figure 2.10: Comparison of 3D feasible regions at different t_{RR} .

To explore this difference systematically, the extreme values of three feasible regions are summarized in Table 2.5. From this table, it can be noticed that the decrease of t_{RR} increases the area of feasible regions by 19.1%, while the increase of t_{RR} decreases by 18.5% of the original feasible area. The reduction of t_{RR} leads to an increase in the extreme values of Q_B and t_{IC} because the decrease of t_{RR} will reduce the preload within the column and activate the CU constraint. To satisfy the CU constraint, the loaded amount of product inside the column should be increased by increasing Q_B and t_{IC} . On the contrary, the growth of t_{RR} will trigger the Y constraint, thereby decreasing the extreme values of Q_B and t_{IC} to avoid product loss. maximum Q_{IC} is unchanged at different t_{RR} , possibly due to the upper bound (UB) of Q_{IC} .

Table 2.5: Comparison of the extreme values of feasible regions at different t_{RR} .

Variable	Feasible bound	t_{RR} (min)		
		43	35	27
Q_{IC} (mL/min)	max	3	3	3
	min	0.1 (↓)	0.15	0.25 (↑)
Q_B (mL/min)	max	1.8 (↓)	2.13	2.5 (↑)
	min	0.1	0.1	0.1
t_{IC} (min)	max	45.7 (↓)	53.5	62.3 (↑)
	min	5 (↓)	6.1	9.4 (↑)
Size		18.5% ↓		19.1% ↑

To examine the change of Q_{IC} concerning various t_{RR} , the UB of Q_{IC} is increased to 4 mL/min to run the simulation and the corresponding result is displayed in Table 2.6. Surprisingly, maximum Q_{IC} is positively correlated with the change in t_{RR} , exhibiting an opposite trend than Q_B and t_{IC} . Theoretically, with the increase of

flowrate, the mass transfer resistance becomes larger, resulting in broadening of the breakthrough curves and faster breakthrough. Another hypothesis for the decrease of the dynamic binding capacity at large flowrates might be because part of the protein does not diffuse further inside the particles, making the adsorption site unable to effectively capture the protein. However, it should be noted that this assumption cannot be validated by the lumped kinetic model employed to characterize the process. Thus, continuously increasing Q_{IC} does not necessarily increase the dynamic loading of the column and might have adverse effects on capacity utilization, explaining why maximum Q_{IC} decreases when t_{RR} decreases. With the rise of t_{RR} , maximum Q_{IC} increases because the growth of t_{RR} improves load conditions inside the chromatographic column, compensating for the unfavorable effects launched by the large flowrates.

To further investigate the changes caused by t_{RR} , Figure 2.11 shows the 2D contour plots at different t_{RR} under varying fixed process variables. It can be obviously seen that the feasible region shifts towards the bottom left corner as t_{RR} increases and moves to the top right corner with the decrease of t_{RR} . By employing the active set strategy, P and Y constraints dominate with the growth of t_{RR} , while P and CU constraints are effective as t_{RR} decreases, as summarized in Table 2.7, which is in accordance with the analysis for the change in the feasible regions under different process variables.

Table 2.6: Comparison of max Q_{IC} at various t_{RR} .

Variable	UB	Feasible bound	$t_{RR} = 43 \text{ min}$	$t_{RR} = 35 \text{ min}$	$t_{RR} = 27 \text{ min}$
Q_{IC} (mL/min)	4	max	3.93 (↑)	3.41	3.21 (↓)

Table 2.7: Summary of active constraints when changing t_{RR} .

$t_{RR}(\text{min})$	Shift	P constraint	CU constraint	Y constraint
↓	↗	✘	✘	
↑	↙	✘		✘

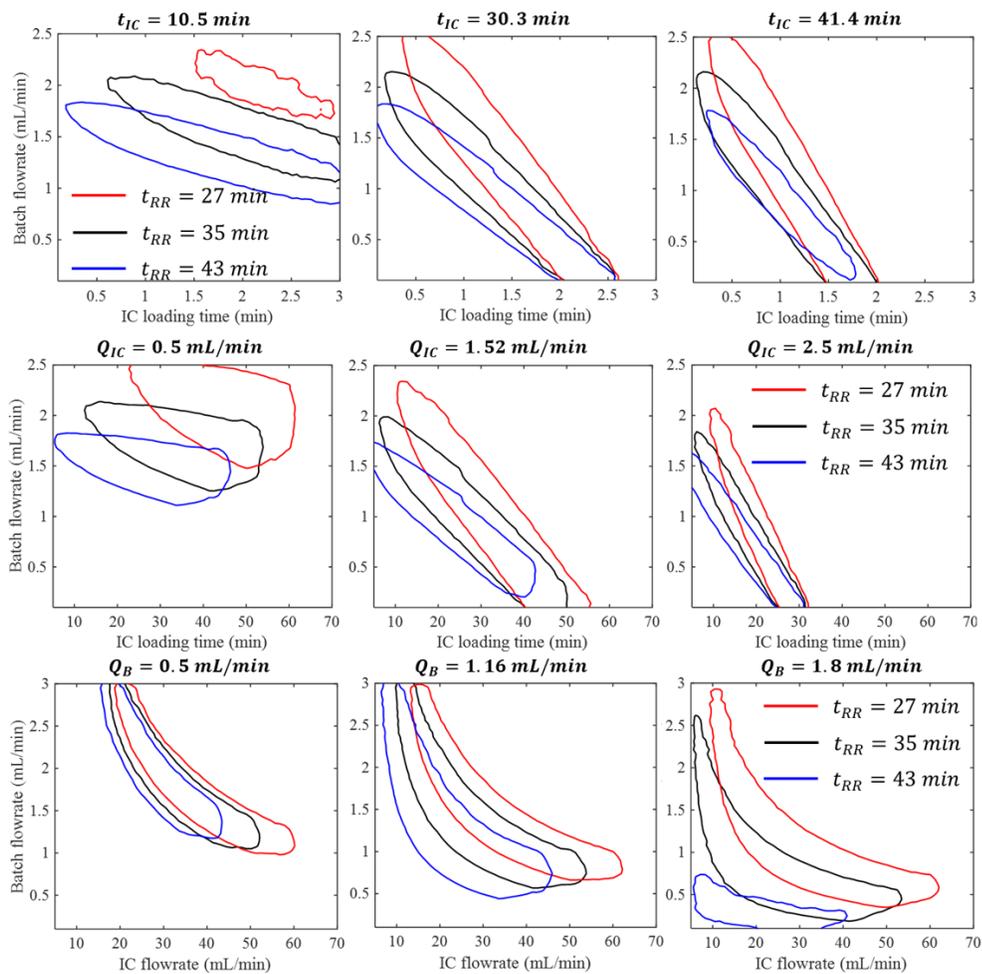


Figure 2.11: Comparison of the 2D contour plots at different t_{RR} under varying fixed process variables for the CaptureSMB process.

2.4.4 Effect of Constraints on Design Space

Constraints can directly affect the design space of the CaptureSMB process, so the impacts of different constraints on the feasible regions are studied. We first investigate the effect of each constraint on the design space individually, following by examining the impact of changing the three constraints simultaneously.

2.4.4.1 Yield Constraint

To delve into the effect of yield constraint, the extreme values of three feasible regions at different Y constraints (0.7, 0.8, and 0.9, respectively) are listed in Table 2.8. When Y constraint increases from 0.8 to 0.9, the feasible region is reduced by 53.2%, while the feasible region increases by 70.8% with Y constraint decreasing from 0.8 to 0.7. From Table 2.8, it can be found that the change of Y constraint has a significant impact on the highest Q_B but little effect on maximum Q_{IC} because Q_B is closely related to the product loss. In the continuous CaptureSMB process, the breakthrough from the 1st column can be caught by the 2nd column, contributing to the little impact of Y constraint on the highest Q_{IC} . maximum t_{IC} keeps constant because this value is determined by productivity constraint. Besides, the change of Y constraint constitutes some impacts on the minimum values of process variables.

Table 2.8: The extreme values of feasible regions at different Y constraints.

Variable	Feasible bound	$Y \geq 0.7$	$Y \geq 0.8$	$Y \geq 0.9$
Q_{IC} (mL/min)	max	3	3	2.95
	min	0.1	0.15	0.25
Q_B (mL/min)	max	2.5	2.13	1.81
	min	0.1	0.1	0.1
t_{IC} (min)	max	53.5	53.5	53.5
	min	5	6.1	8.31
Size		70.8% ↑	--	53.2% ↓

2D feasibility plots under different constraint are shown in Figure 2.12. The feasible regions with larger Y constraint values should lie in those with smaller Y constraint values, conforming to the 2D contour plots in Figure 2.12. As Y constraint increases from 0.8 to 0.9, the feasible region shrinks towards the bottom left corner to prevent more product loss, while the feasible region expands towards the top right corner when the yield constraint decreases from 0.8 to 0.7. It should be noteworthy that the left feasible boundary remains almost unmoved, and only the right boundary shifts with the change of Y constraint, indicating that only the right feasible boundary is governed by the Y constraint.

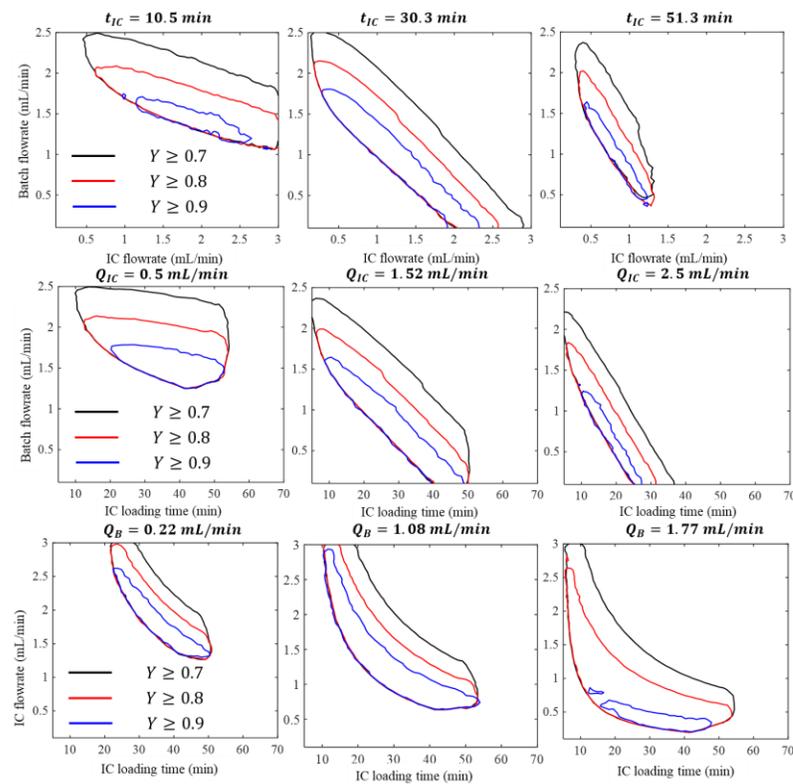


Figure 2.12: Comparison of the 2D contour plots under different Y constraints at varying fixed process variables for the CaptureSMB process.

2.4.4.2 Capacity Utilization Constraint

The effect of CU constraint on the design space is studied, and the comparison of the extreme values at different CU constraints (0.75, 0.85, and 0.95) is shown in Table 2.9. When CU constraint increases from 0.85 to 0.95, the feasible region is reduced by 83.4%. The highest Q_{IC} is substantially reduced from 3 to 1.38 mL/min because large Q_{IC} will adversely affect the dynamic binding capacity, as we analyzed in Section 2.4.3. There is no change of maximum t_{IC} observed because this value is determined by P constraint, but a significant increase of minimum t_{IC} can be observed as large t_{IC} is beneficial for CU. The lowest Q_{IC} and Q_B are both slightly increased to increase the sample load inside the column. As CU constraint decreases from 0.85 to 0.75, the feasible region increases by 48%. The highest Q_{IC} stays unchanged with the decrease of CU constraint due to the upper limit of Q_{IC} , while the minimum feasible extreme values of t_{IC} , Q_{IC} and Q_B are determined by the corresponding lower bounds.

Table 2.9: Comparison of the extreme values of feasible regions at different CU constraints.

Variable	Feasible bound	$CU \geq 0.75$	$CU \geq 0.85$	$CU \geq 0.95$
Q_{IC} (mL/min)	max	3	3	1.38
	min	0.1	0.15	0.4
Q_B (mL/min)	max	2.13	2.13	2.01
	min	0.1	0.1	0.43
t_{IC} (min)	max	53.5	53.5	53.5
	min	5	6.1	25.9
Size		48% ↑	--	83.4% ↓

To clearly visualize the effect of CU constraints on feasible regions, Figure 2.13 displays 9 sets of 2D feasibility plots of different CU constraints at various fixed process

variables. No matter CU increases from 0.75 to 0.85 or from 0.85 to 0.95, the feasible region of $CU \geq 0.85$ or $CU \geq 0.95$ is still within the feasible range of $CU \geq 0.75$ or $CU \geq 0.85$.

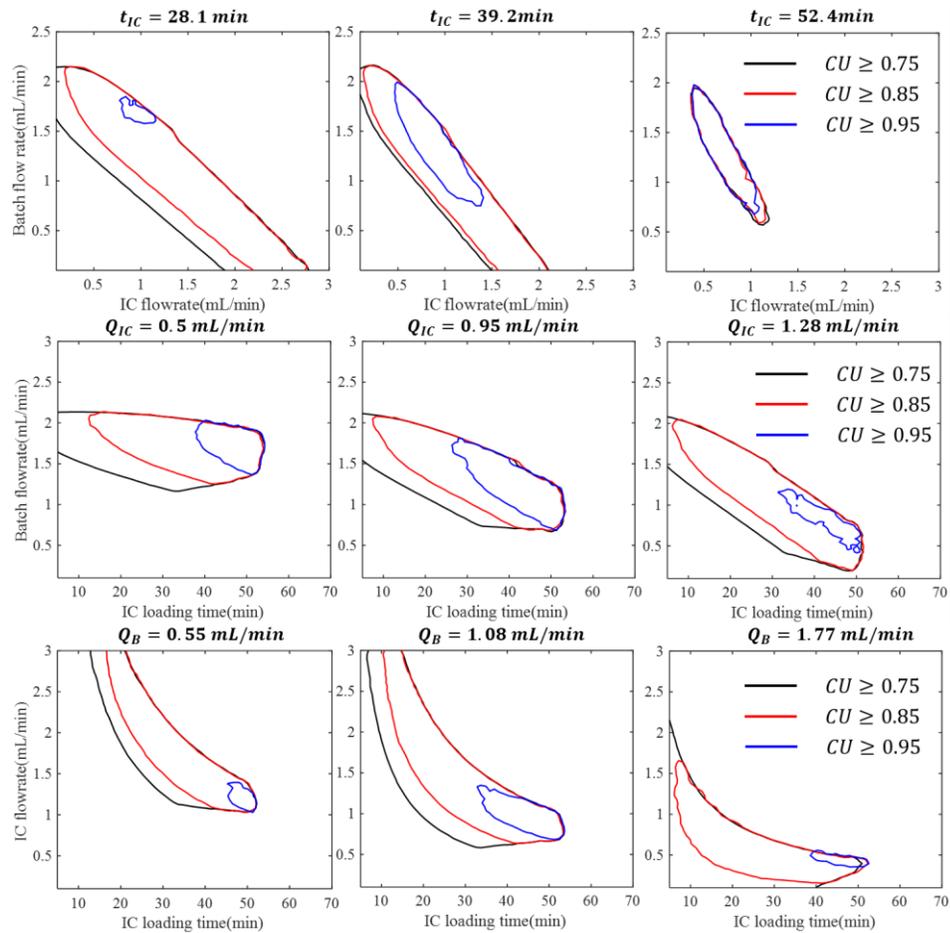


Figure 2.13: Comparison of the 2D contour plots under different CU constraints at varying fixed process variables for the CaptureSMB process.

If the CU constraint increases, the feasible region shrinks towards the top right corner to increase the sample load, while the feasible region expands towards the bottom

left corner when the CU constraint decreases. It can be also observed that the right feasible boundary remains unmovable, and only the left feasible boundary moves, implying that CU constraint is active on the left feasible boundary. Given that the long duration of interconnected loading is advantageous for CU, the 2D feasible regions at fixed Q_B or Q_{IC} are located at large t_{IC} . Compared with that at smaller t_{IC} , the discrepancy of feasible regions under different CU constraints is neglectable at large fixed t_{IC} , possibly because large t_{IC} will make the column fully loaded.

2.4.4.3 Productivity Constraint

Table 2.10 reveals the feasible bounds at different P constraints (10, 15, and 20 mg mL⁻¹ h⁻¹). With the increase of P constraint from 15 to 20 mg mL⁻¹ h⁻¹, the feasible region decreases by 51.2%, while the area increases by 31.2% when P constraint decreases from 15 to 10 mg mL⁻¹ h⁻¹. It can be noted that maximum t_{IC} decreases significantly with the increase of P constraint and vice versa, because short time is good for productivity. Based on previous analysis, CU and Y constraints are responsible for maximum Q_{IC} and Q_B , respectively, so the two values remain constant when P constraint is changed. Minimum Q_{IC} and Q_B both increase slightly when P constraint rises but remain unchanged as P constraint decreases owing to their lower bounds.

In order to further study the changes of feasible regions introduced by P constraint, the comparison of 2D feasibility plots under different fixed process variables are presented in Figure 2.14. No matter P constraint increases from 10 to 15 or from 15 to 20 mg mL⁻¹ h⁻¹, the changing trend is the same, i.e., the feasible regions almost overlap at small t_{IC} but the difference between regions becomes larger with the increase of t_{IC} . The occurrence of this phenomenon indicates that lower t_{IC} benefits productivity, which also explains why the 2D feasible regions at fixed Q_B or Q_{IC} are

located at small t_{IC} values when P constraint is increased. In addition, it can be found that the right feasible boundary hardly changes with P constraint increasing because Y constraint dominates this boundary. However, the left boundary shrinks towards the top right corner at large t_{IC} values, indicating that the left boundary is determined by P constraint under this situation, in line with the analysis of the opposite trend of Figure 2.9. Regarding the feasible regions at fixed Q_{IC} , the change of feasible regions when changing P constraints decreases as Q_{IC} increases. At first, a significant difference is identified between the feasible regions under different P constraints, but the difference becomes smaller and smaller with the increase of Q_{IC} . There are two possible explanations for this occurrence. One is that the feasible range of t_{IC} is lower at fixed large Q_{IC} compared with small Q_{IC} as t_{IC} is an extremely important factor in deciding productivity. Another possible reason is that larger Q_{IC} can produce more products during the same time period, resulting in the little change of feasible regions. However, this phenomenon cannot be observed at fixed Q_B possibly because the possible range of t_{IC} at different Q_B is very similar.

Table 2.10: The extreme values of feasible regions at different P constraints.

Variable	Feasible bound	$P \geq 10$	$P \geq 15$	$P \geq 20$
Q_{IC} (mL/min)	max	3	3	3
	min	0.1	0.15	0.3
Q_B (mL/min)	max	2.13	2.13	2.13
	min	0.1	0.1	0.18
t_{IC} (min)	max	70	53.5	30.34
	min	6.1	6.1	6.1
Size		31.2% ↑	--	51.2% ↓

To visualize the effects of each constraint on the design space in CaptureSMB process more clearly, Table 2.11 and Table 2.12 summarize the changes in the feasible region initialized by increasing or decreasing constraints, respectively.

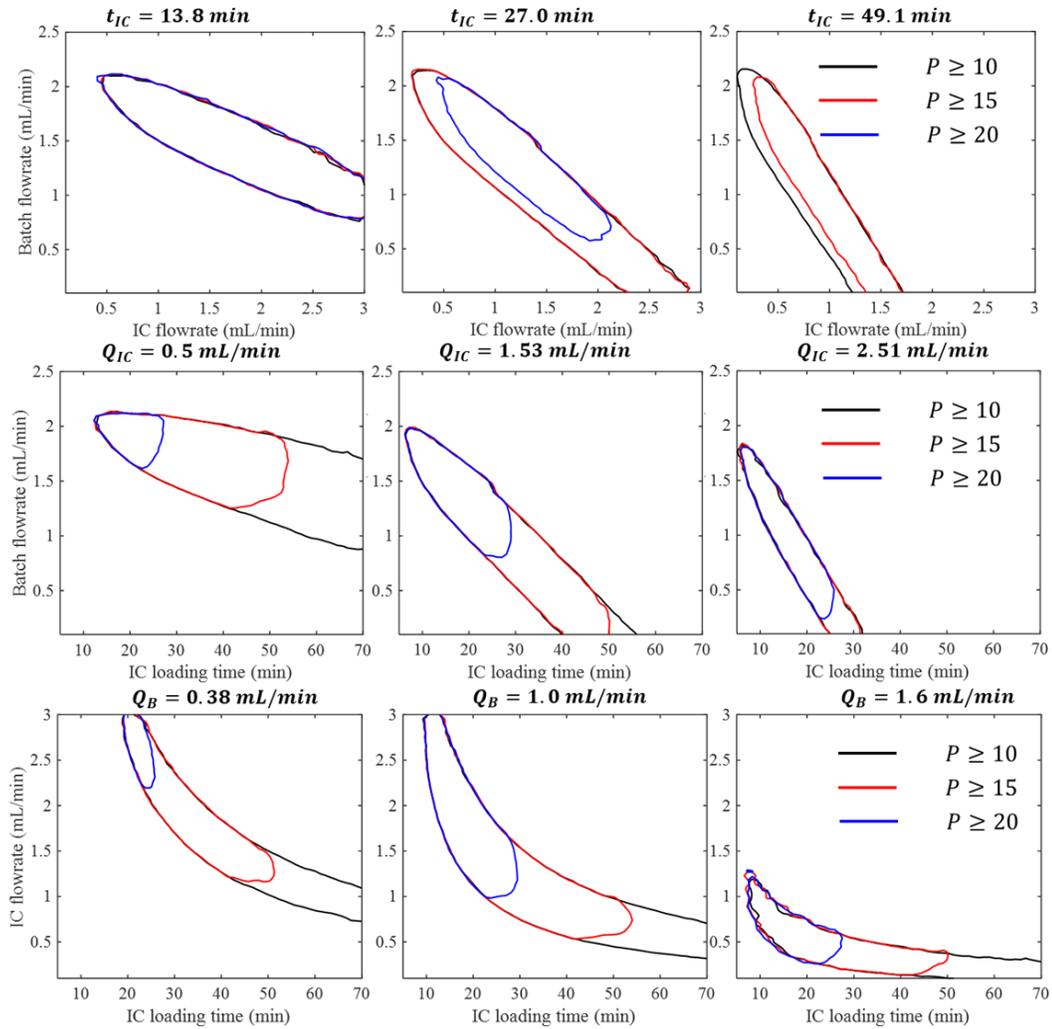


Figure 2.14: Comparison of the 2D contour plots under different P constraints at varying fixed process variables for the CaptureSMB process.

Table 2.11: Summary of design space changes with individual constraint increases.

Constraints	Size (↓)	Shifting	Feasible bounds	Remark
Y (0.8→0.9)	53.2%	Bottom left corner (only right feasible boundary)	Greatly decrease the max Q_B (2.13→1.81)	Smaller process variables benefit Y
CU (0.85→0.95)	83.4%	Top right corner (only left feasible boundary)	Significantly increase the min t_{IC} (6.1→25.9); Significantly decrease the max Q_{IC} (3→1.38)	Larger process variables benefits CU, especially larger t_{IC}
P (15→20)	51.2%	Shrink towards small t_{IC} ; Top right corner at large fixed t_{IC} (only left feasible boundary)	Greatly decrease the max t_{IC} (53.5→30.3)	Lower t_{IC} benefits P; Larger flowrates benefit P at fixed t_{IC}

Table 2.12: Summary of design space changes with individual constraint decreases.

Constraints	Size (↑)	Shifting	Feasible bounds
Y (0.8→0.7)	70.8%	Top right corner (only right feasible boundary)	Increase the max Q_B (2.13→2.5)
CU (0.85→0.75)	48%	Bottom left corner (only left feasible boundary)	No change of max Q_{IC} possibly because of the upper bound of Q_{IC} ; Little decrease of min t_{IC} (6.1→5) due to the lower bound of t_{IC}
P (15→10)	31.2%	Expand towards large t_{IC} ; Bottom left corner at large fixed t_{IC} (only left feasible boundary)	Significantly increase the max t_{IC} (53.5→70)

2.4.4.4 Simultaneously Changing Constraints

After examining the influence of each constraint on the design space of the continuous CaptureSMB process, the impact of changing all three constraints concurrently is investigated. The three constraints (Y, CU, and P) are increased from 0.8 to 0.9, 0.85 to 0.9, and 15 to 17 mg mL⁻¹ h⁻¹, respectively, as denoted by ‘Alternative’. Table 2.13 lists the change of extreme values under different situations and the corresponding binding constraints for this change calculated from the active set strategy, which is in agreement with the above analysis of increasing each constraint. Besides, when increasing three constraints simultaneously, the feasible area decreases by 83.3%.

Table 2.13: Comparison of the extreme values of feasible regions when increasing three constraints.

Constraints	Variable	Feasible bound	Base case	Alternative with different constraints	Active constraints
Y (0.8 → 0.9)	Q_{IC} (mL/min)	max	3	2.23 (↓)	CU, Y
		min	0.15	0.39 (↑)	P, CU
CU (0.85 → 0.9)	Q_B (mL/min)	max	2.13	1.77 (↓)	Y
		min	0.1	0.1	Lower bound
P (15 → 17)	t_{IC} (min)	max	53.5	42.46 (↓)	P
		min	6.1	16.02 (↑)	CU
Size				83.3% (↓)	

Figure 2.15 displays the difference of 2D contour plots when increasing three constraints simultaneously. The increase in Y constraint forces the right feasible boundary to move toward the bottom-left corner, and CU constraint is active for shifting

the left boundary toward the top-right corner. Besides, lower t_{IC} favors P and higher t_{IC} profits CU, so the shifted feasible region is located at the middle of t_{IC} due to the tradeoff between P and CU.

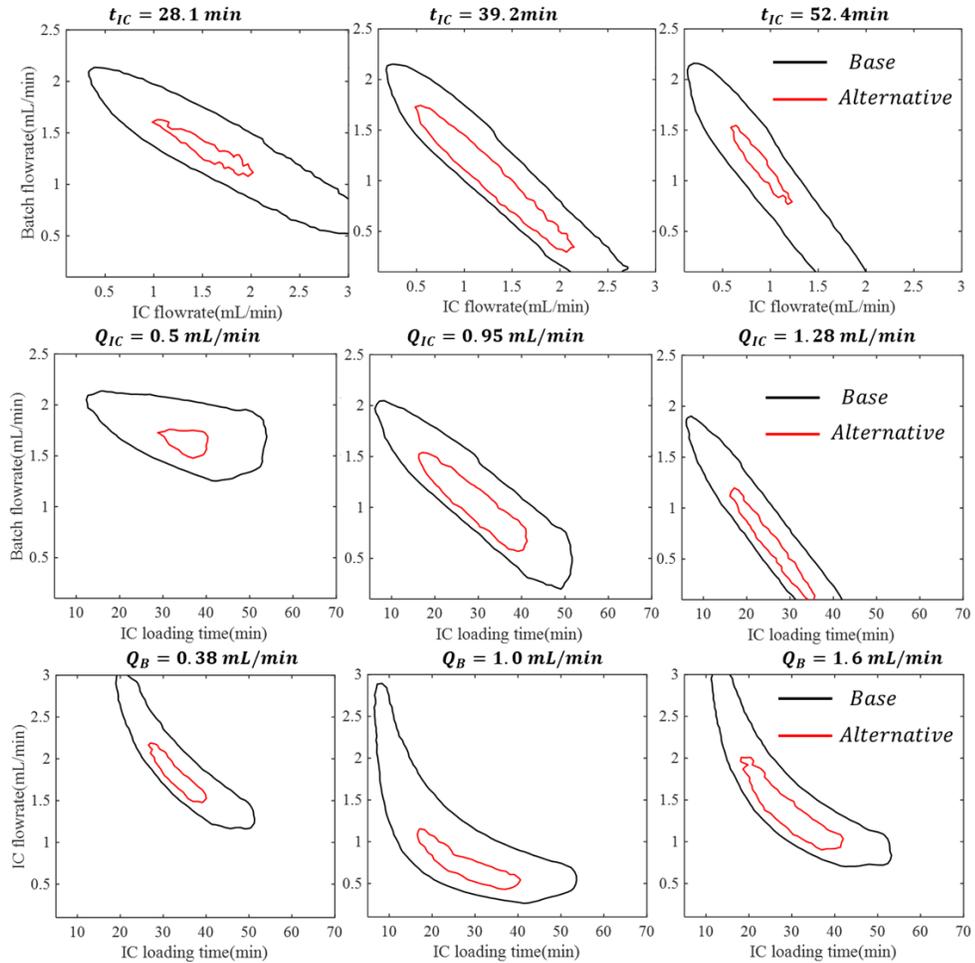


Figure 2.15: Comparison of 2D contour plots for simultaneously increasing three constraints in the CaptureSMB process. Base: $Y \geq 0.8$, $CU \geq 0.85$, $P \geq 15$; Alternative: $Y \geq 0.9$, $CU \geq 0.9$, $P \geq 17$.

2.4.5 Effect of Design Variables on Design Space

After studying the effects of t_{RR} and constraints on the feasible regions, we also explore the column length of the CaptureSMB system to see how this affects the design space. As seen from Table 2.14, when the column length is decreased from 10 cm to 5 cm, the area of design space drastically decreases to 24.6%. Specifically, no change is observed in t_{IC} , but maximum Q_{IC} and Q_B substantially decrease due to Y constraint, which is reasonable because the flowrates should decrease to prevent product loss as the column becomes short.

Table 2.14: Comparison of the extreme values of feasible regions at different column lengths.

Variable	Feasible bound	$L = 10\text{ cm}$	$L = 5\text{ cm}$
Q_{IC} (mL/min)	max	3	1.82
	min	0.15	0.1
Q_B (mL/min)	max	2.13	1.08
	min	0.1	0.1
t_{IC} (min)	max	53.5	53.5
	min	6.1	6.1
Size		--	75.4% ↓

From Figure 2.16, it can be found that the feasible region shifts towards the bottom-left corner, possibly due to yield constraint. An active set strategy is employed to help evaluate the reasons for this change. It is found that yield constraint is always active, which confirms the analysis for this change, and P constraint is binding at large t_{IC} . In addition, CU constraint is active at large Q_{IC} because large Q_{IC} will lead to increased mass transfer resistance and negatively affect the sample load inside the column, as described in Section 2.4.3. It is worth noting that when we study the same

shifting trend of feasible regions caused by the process variables, only Y and P constraints are binding, but all three constraints are binding when the column length decreases.

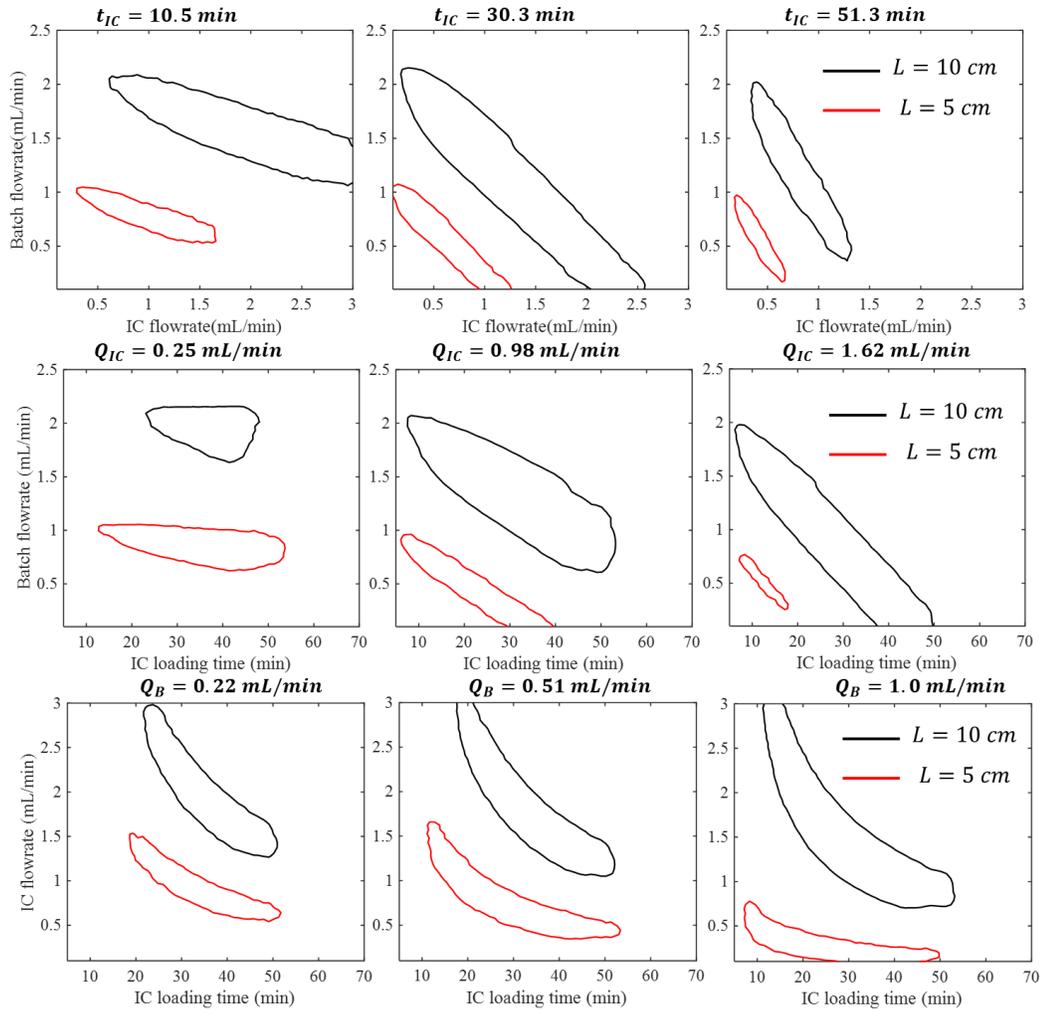


Figure 2.16: Comparison of the 2D contour plots under different column length at varying fixed process variables for the CaptureSMB process.

2.5 Conclusions

This chapter proposed a systematic approach to identify the design space of continuous chromatography. In an effort to balance computational complexity and model predictions, surrogate-based feasibility analysis with adaptive sampling is applied to establish the design space of twin-column CaptureSMB. The effects of process variables, including interconnected loading time, interconnected flowrate, and batch flowrate, are inspected in the design space. It was found that productivity and yield constraints are active with the increase of process variables while productivity and capacity utilization constraints are binding with the decrease of process variables.

The effects of each constraint on the design space are also investigated, followed by changing three constraints simultaneously. With the increase of yield requirement, the left feasible boundary almost keeps unchanged and only the right boundary shifts towards the bottom-left corner. When CU constraint increases, the right feasible boundary remains the same, and only the left feasible boundary moves to the top right corner. The interconnected loading time t_{IC} has significant impacts on the P constraint and the feasible region shifts to lower t_{IC} as P constraint increases. More detailed analyses about the effects of constraints on the design space are directed to the corresponding sections. Finally, the impact of the column length was thoroughly studied. It is found that yield constraint is always active, forcing the feasible regions to shift towards the bottom left corner (or left) to avoid product loss, but productivity and capacity utilization constraints are only binding at large t_{IC} and large Q_{IC} .

In addition to the findings above, there are some other aspects that can be considered using the proposed framework and will be explored in future publications. For example, lumped kinetic model is used to train the surrogate model, but this model does not consider particle diffusion inside the resin. Thus, more detailed mechanistic

models like the general rate model or shrinking core model will be used to investigate the design space of the continuous capture process. Moreover, the developed process model in this work does not include the simulation of elution process, which might have some impacts on the calculated yield and productivity. More comprehensive studies will be performed to simulate the elution process and investigate the purity under the multi-component systems. Furthermore, continuous chromatography systems with different column numbers will be also considered and compared. Additionally, the influence of modeling parameters caused by different resins on the design space will be examined.

Chapter 3

MACHINE LEARNING-BASED OPTIMIZATION FOR ION EXCHANGE CHROMATOGRAPHY

3.1 Introduction

Ion-exchange chromatography (IEX) serves as a crucial separation technique in the biopharmaceutical downstream process, particularly for the polishing steps [57, 155]. IEX is primarily employed for the separation of charge variants through either stepwise or gradient elution methods [78]. This central cut process is complicated, with multiple factors affecting its efficiency, including pH, buffer concentration, resin selection, elution strategies, and product pooling [82]. Solely relying on experimental approaches to characterize and optimize this multivariate process can be challenging, potentially leading to increased time and resource demands [196, 197]. To accelerate process development and reduce the associated costs of conducting time- and resource-intensive experiments, Process Systems Engineering (PSE) tools, such as model-based optimization, could be implemented to solve the multiparametric problem, which could help ensure a robust and optimal separation process [149, 198].

Chapter 3 is adapted from the following journal article written by the author of this dissertation:

Ding, C., Ierapetritou, M., Machine learning-based optimization of a multi-step ion exchange chromatography for ternary protein separation. *Computers & Chemical Engineering*, 2024, 184:108642.

In response to the rapid advancements of artificial intelligence (AI), Industry 4.0, and Digital Twins (DTs), and in alignment with the Quality by Design (QbD) initiative, there is a growing demand for gaining a more profound understanding of the underlying processes [12, 13, 126]. Mechanistic models, rooted in both physical laws and process knowledge, have been extensively applied for process characterization and optimization [199]. In the chromatographic procedures, mechanistic modeling primarily focuses on describing the solute transport within the inter- and intra-particle mobile phase and the adsorption process occurring in the stationary phase [74]. The general rate model is widely recognized as the most comprehensive one for representing mass transport due to its capacity to account for convection, axial dispersion, film mass transfer, and pore diffusion. In the context of the adsorption within IEX systems, the steric mass action (SMA) model is notably adopted and outstands in capturing salt-dependent effects [78]. For example, to leverage the potential of mechanistic modeling, Cebulla et al. [93] applied model-based approach to optimize the IEX process for isolating a fragment from human serum albumin. The resulting optimal operating conditions were shown to yield higher purity, reduced processing time, and decreased buffer consumption in comparison to the original experimental procedures.

However, as these mechanistic models evolve to incorporate more precise representations of process dynamics, the corresponding computational complexity also increases [189]. Consequently, it can become inefficient and time-consuming to solve such complex problems with traditional optimization approaches. To address this challenge, machine learning (ML)-based strategies have emerged as a promising alternative to balance the computational burden and model accuracy [137, 200, 201]. In our previously published work [89], surrogate-based feasibility analysis was proposed

in order to identify the design space of twin-column continuous chromatography. Feasibility analysis could help find design space that represents the operational ranges ensuring product quality. To enhance the efficiency of solving the computationally expensive simulations involved in feasibility analysis, a surrogate model was utilized to represent the mechanistic model and this approach yielded a substantial 36-fold reduction in computational time. Our work was based on the direct design approach for the identification of the design space, but it is noteworthy that the inverse design, also referred to as the inverse problem, has also gained popularity in the field of natural sciences and engineering [202]. Although this framework could be directly extended to address the constrained optimization problems, it is crucial to recognize that in that case, the feasibility analysis and the objective function would go through different training phases, potentially leading to an increased sampling budget. If using the separate training strategies, the sampling effort might be directed toward the feasible region with large objective values for the feasibility analysis or the infeasible region with small objective values for the optimization part. To address this challenge, we have introduced a surrogate-based optimization methodology coupled with feasibility analysis [60] to effectively address optimization challenges within the constraints of operational and quality requirements in IEX processes.

In this work, the ML-based optimization framework was applied to a case study focused on optimizing a multicomponent separation process involving ribonuclease, cytochrome and lysozyme. This case study, adapted and modified from a previous publication [90], aimed to maximize productivity while adhering to operational and quality constraints. To achieve this, Gaussian process regression models (GPR) models were constructed to approximate both the objective function and the constraints by using

data generated from the mechanistic model. To reduce the sampling budget, the optimization framework was divided into feasibility and optimization stages, each incorporating an adaptive sampling strategy. The optimal operating conditions, constraints, objective values, and required simulation time obtained through our ML-based approach were compared to those obtained by directly applying traditional optimization algorithms, such as genetic algorithms (ga) using the mechanistic model. Moreover, the effects of peak cutting criteria on the optimization results and active constraint were investigated, and a detailed analysis of the process' design space at the optimal solution was performed.

3.2 Materials and Methods

3.2.1 Case Study

The investigation into the separation of a protein mixture, encompassing ribonuclease (RNase), cytochrome (cyt), and lysozyme (lyz), is pursued as a case study, as it is a prototype and benchmark example employed in the academic modeling [84, 90, 91, 198, 203-205]. The separation process is achieved through the use of cation-exchange columns (CEX) packed with SP Sepharose FF beads, with sodium chloride playing a crucial role in modifying the protein affinities. When modeling the isolation process, the central component (cyt) is the protein of interest, with the other two components considered as impurities. The parameters associated with column geometry, mass transport and binding behaviors of the three proteins for modeling are obtained from He et al. [90], which are listed in Table 3.1.

Table 3.1: Parameters related to column geometry, mass transport, and binding behavior of the proteins ($i \in \{\text{RNase, cyt, lyz}\}$) applied in the mechanistic model.

Catalog	Symbol	Description	Value	Unit
Column geometry	L	Column length	1.4E-2	m
	d	Column diameter	1E-2	m
	r_p	Particle radius	4.50E-5	m
	ε_c	Column porosity	0.37	
	ε_p	Particle porosity	0.75	
Mass transport	$D_{a,x}$	Axial dispersion	5.75E-8	m^2s^{-1}
	$k_{f,i}$	Film mass transfer	6.90E-6	ms^{-1}
	$D_{p,i}$	Pore diffusion	6.07E-11	m^2s^{-1}
Isotherm	Λ	Ionic capacity	1200	$mol\ m^{-3}$
	$k_{a,i}$	Adsorption coefficients	[7.70, 1.59, 35.5]	s^{-1}
	$k_{d,i}$	Desorption coefficients	[1000, 1000, 1000]	s^{-1}
	ν_i	Characteristic charges	[3.70, 5.29, 4.70]	
	σ_i	Steric factors	[10.0, 10.6, 11.83]	
	c_{ref}	Reference salt concentration in the particle liquid phase	Maximum inlet salt concentration	$mol\ m^{-3}$
	q_{ref}	Reference salt concentration in the particle solid phase	1200 (Equal to ionic capacity Λ)	$mol\ m^{-3}$

The operation protocol for the CEX column is also taken from [90], as shown in Figure 3.1. Specifically, the column is first equilibrated with the protein-free running buffer with an initial salt concentration of $50\ mol\ m^{-3}$. Subsequently, the column is loaded for a duration of 10 seconds at a constant ionic strength of $50\ mol\ m^{-3}$, with lysozyme, cytochrome, and ribonuclease each at concentrations of $1\ mol\ m^{-3}$. Following the loading step, a 50-second washing phase is carried out at the same ionic strength. In the elution stage, multi-step gradient elution strategies are implemented, encompassing two linear gradient elutions with different time intervals ($\Delta t_1, \Delta t_2$) and

various slopes (m_1, m_2), along with one isocratic step at a constant salt concentration ($c_{init0} + m_1\Delta t_1 + m_2\Delta t_2$). The final step, which would typically involve the recovery and regeneration phase, is omitted during the modeling process and thus not depicted in Figure 3.1.

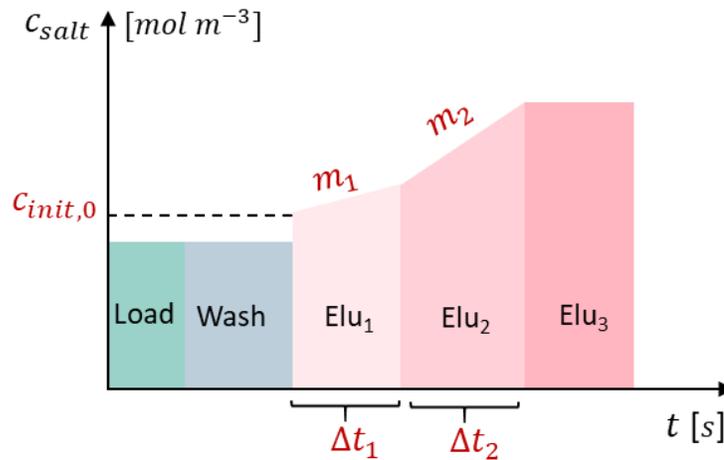


Figure 3.1: Operating protocols for the CEX considered in the modeling. Elu₁, Elu₂, and Elu₃ correspond to the multistep elution strategies. m_1 and m_2 represent the slopes for the first two linear elution phases; Δt_1 and Δt_2 denote the time intervals corresponding to those elution phases; c_{init0} refers to the initial salt concentration at the beginning of the elution process.

3.2.2 Mechanistic Model

The mechanistic model should be formulated to describe the mass transport of the mobile phase and the adsorption phenomena of the stationary phase. The General Rate Model is selected for its ability to accommodate various levels of mass transfer resistance. As for the binding behavior inside IEX process, the Steric Mass Action

isotherm is employed as it accounts for the salt concentration, molecule charges and steric hindrance.

3.2.2.1 General Rate Model

The general rate model was applied in the exact form as implemented within the Chromatography Analysis and Design Toolkit (CADET) [96], as illustrated in Equations 3.1-3.2.

$$\frac{\partial c_i}{\partial t} = -u \frac{\partial c_i}{\partial z} + D_{ax} \frac{\partial^2 c_i}{\partial z^2} - \frac{1-\varepsilon_c}{\varepsilon_c} \frac{3}{r_p} k_{f,i} (c_i - c_{p,i}(\cdot, r_p)) \quad 3.1$$

$$\frac{\partial c_{p,i}}{\partial t} = D_{p,i} \left(\frac{\partial^2 c_{p,i}}{\partial r^2} + \frac{2}{r} \frac{\partial c_{p,i}}{\partial r} \right) - \frac{1-\varepsilon_p}{\varepsilon_p} \frac{\partial q_i}{\partial t} \quad 3.2$$

where c_i , $c_{p,i}$ and q_i correspond to the concentration of i^{th} component in the interstitial, stagnant and stationary phases. Four components are taken into consideration during the simulation, with i values ranging from 0 to 3, representing the salt ions, Rnase, cyt, and lyz; respectively. The variables t , z and r represent the time coordinate, axial position along the column direction, and the radial position along the particle radius; u denotes the superficial velocity, while ε_c and ε_p are the column porosity and particle porosity; D_{ax} , $k_{f,i}$, and $D_{p,i}$ represent axial dispersion, film mass transfer and pore diffusion coefficient of i_{th} component.

The Danckwerts' boundary conditions are implemented at the column inlet and outlet in Equations 3.3–3.4, along with the specified boundary conditions for the beads outlined in Equations 3.5–3.6 [206].

$$u c_{in,i}(t) = u c_i(t, 0) - D_{ax} \frac{\partial c_i}{\partial z}(t, 0) \quad 3.3$$

$$\frac{\partial c_i}{\partial z}(t, L) = 0 \quad 3.4$$

$$\varepsilon_p D_{p,i} \frac{\partial c_{p,i}}{\partial r^2}(\cdot, r_p) = k_{f,i} (c_i - c_{p,i}(\cdot, r_p)) \quad 3.5$$

$$\frac{\partial c_{p,i}}{\partial r^2}(\cdot, \cdot, 0) = 0 \quad 3.6$$

where $c_{in,i}(t)$ is the time-dependent inlet concentration of the i^{th} component; L and r_p represent the column length and the particle radius; respectively.

3.2.2.2 Steric Mass Action

Steric mass action was also implemented within CADET as shown in Equation 3.7 [206].

$$\frac{\partial q_i}{\partial t} = k_{a,i} c_{p,i} \left(\frac{\bar{q}_0}{q_{ref}} \right)^{\nu_i} - k_{d,i} q_i \left(\frac{c_{p,0}}{c_{ref}} \right)^{\nu_i} \quad 3.7$$

where c_{ref} and q_{ref} are the reference salt concentrations in the liquid and solid phases inside the beads, which are primarily utilized for the purpose of normalizing adsorption and desorption rates; for i^{th} component, $k_{a,i}$ and $k_{d,i}$ denote the adsorption and desorption coefficients, while ν_i is the characteristic charge; \bar{q}_0 is the number of available binding sites, which can be calculated from the number of bound counter ions q_0 by considering steric shielding, as depicted in Equation 3.8.

$$\bar{q}_0 = \Lambda - \sum_{j=1}^{N_{comp}-1} (\nu_j + \sigma_j) q_j = q_0 - \sum_{j=1}^{N_{comp}-1} \sigma_j q_j \quad 3.8$$

where q_0 could be derived from an electro-neutrality condition, as described in Equation 3.9; σ_j is the steric factor of i^{th} component, and Λ is the ionic capacity for the CEX resin. The total number of components, denoted as N_{comp} , should be 4 in this case study.

$$q_0 = \Lambda - \sum_{j=1}^{N_{comp}-1} \nu_j q_j \quad 3.9$$

For a more comprehensive understanding of the SMA model, publication [206] should be reviewed. The detailed description and corresponding values of all parameters related to column geometry, mass transport, and binding behavior of the proteins are summarized in Table 3.1.

3.2.3 Problem Statement

Productivity, yield, and purity are critical indicators for assessing the chromatographic column performance. The calculation of these metrics follows the methodology outlined in [90]. All three metrics are calculated within the pooling time interval t_p , starting at time τ . In this work, peak cutting for pooling is defined by setting the percentage of target protein concentration at the column outlet to a fixed threshold of 75%. While it is evident that the criteria for peak cutting can exert a substantial influence on the performance indicators of the column, it is noteworthy that this factor is not considered a decision variable in the optimization problem. Instead, various scenarios under different percentage thresholds are investigated to enhance our comprehension of the underlying process.

The productivity of cytochrome (Pr_{cyt}) is defined as the quantity of target protein produced per pooling time interval and per unit resin volume, as illustrated in Equation 3.10. The yield (Y_{cyt}) is determined by the withdrawn target product in relation to the corresponding feed mass, which can be calculated in Equation 3.11. The purity (Pu_{cyt}) is defined as the ratio of the collected target protein (cytochrome) to the cumulative sum of the three protein components during the pooling interval, as expressed in Equation 3.12.

$$Pr_{cyt} = \frac{Q \int_{\tau}^{\tau+t_p} c_{cyt}(t,L) dt}{t_p(1-\varepsilon_c)V_c} \quad 3.10$$

$$Y_{cyt} = \frac{\int_{\tau}^{\tau+t_p} c_{cyt}(t,L) dt}{t_{load} c_{in,cyt}} \quad 3.11$$

$$Pu_{cyt} = \frac{\int_{\tau}^{\tau+t_p} c_{cyt}(t,L) dt}{\sum_{i=1}^3 \int_{\tau}^{\tau+t_p} c_i(t,L) dt} \quad 3.12$$

where Q denotes the volumetric flowrate; V_c stands for the column volume; and t_{load} represents the loading time.

$$\min - Pr_{cyt}(x) \quad 3.13$$

$$s. t. Pu_{cyt}(x) \geq 90\% \quad 3.14$$

$$Y_{cyt}(x) \geq 80\% \quad 3.15$$

$$c_{salt}(x) \leq 1 M \quad 3.16$$

$$x^{lb} \leq x \leq x^{ub} \quad 3.17$$

In this study, productivity is selected as the performance variable for optimization, while the other two metrics (yield and purity) are required to meet certain predefined criteria. The goal of maximizing productivity is achieved by adjusting the multi-step elution strategies, as detailed in Section 3.2.1, where a total of 5 decision variables ($x = \{\Delta t_1, \Delta t_2, m_1, m_2, c_{init0}\}$) are presented. The overall optimization formulation problem with constraints can be found in Equations 3.13 – 3.17. The target values for purity and yield are considered to be 90% and 80%; respectively. It should be noted that these values are user-defined and can be adjusted according to specific requirements. Furthermore, it is essential to ensure that the salt concentration c_{salt} does not surpass 1 M during the elution process, aligning with practical applications. The lower and upper bounds of the decision variables in this optimization case study are summarized in Table 3.2, with reference to He et al. [90]. The upper bounds for certain variables (Δt_1 , Δt_2 , and m_2) are adjusted to smaller values. This adjustment is applied to m_2 to ensure the satisfaction of the salt constraint. It has been also observed that optimizing with large values for Δt_1 and Δt_2 does not yield any benefits, as indicated by the optimization results. A total process time of 15,000 s was adopted in the simulation to guarantee the comprehensive evaluation of different scenarios throughout the optimization process.

Table 3.2: Lower (lb) and upper bounds (ub) of decision variables in optimization case study.

Symbol	Description	lb	ub	Unit
Δt_1	Time interval for the 1 st linear elution	500	1000	<i>s</i>
Δt_2	Time interval for the 2 nd linear elution	1000	5000	<i>s</i>
m_1	Slope for the 1 st linear elution	1.0×10^{-3}	1.0×10^{-2}	<i>mol m⁻³ s⁻¹</i>
m_2	Slope for the 2 nd linear elution	1.0×10^{-3}	1.0	<i>mol m⁻³ s⁻¹</i>
c_{init0}	Initial salt concentration of elution	20	200	<i>mol m⁻³</i>

3.2.4 Optimization Framework

Model-based optimization facilitates the determination of optimal operating conditions, leading to a reduction in experimental efforts and advancements in process development [56]. Instead of employing mechanistic model, this study introduces a methodology that incorporates machine learning to enhance the optimization process, aiming to efficiently maximize the productivity of the IEX separation process. In this context, Gaussian Process Regression (GPR) model (also known as Kriging) is employed as a surrogate for the complicated mechanistic model [89].

The aforementioned optimization problem is characterized by multiple constraints, which would lead to multiple surrogates being required to represent each constraint individually. Consequently, a feasibility function in Equation 3.18, derived from feasibility analysis, is utilized to integrate all constraints, requiring only one GPR model to represent the entire feasible region [207]. Coupling with adaptive sampling

technique, it becomes feasible to reduce the sampling budget while maintaining a high level of accuracy in constructing the surrogate model [195].

$$\psi(x) = \max_{j \in J} \{f_j(x)\} \quad 3.18$$

where x , $f_j(x)$, $\psi(x)$ represent the operating variables, the j^{th} constraint function, and the feasibility function, respectively. Regions where $\psi(x)$ is less than or equal to 0 are regarded as the design space or feasible region of the defined process.

The initial GPR model is established using the initial sampling points generated from a space-filling design of experiment (DOE). Following this, an infill criterion guides the identification of the next sampling point with either high uncertainty or a low objective value, thereby saving the need for exhaustive sampling [208]. One of the frequently used infill criteria is the expected improvement (EI) function, as represented in Equation 3.19, which effectively balances the exploration and exploitation at the adaptive sampling phase. This iterative process continues until a predetermined stopping criterion is satisfied, enabling the utilization of the final accurate GPR model. This approach is also commonly recognized as Bayesian optimization [209].

$$EI(x) = (f_{min} - \hat{y})\Phi\left(\frac{f_{min} - \hat{y}}{\hat{s}}\right) + \hat{s}\phi\left(\frac{f_{min} - \hat{y}}{\hat{s}}\right) \quad 3.19$$

where \hat{y} and \hat{s} are the predicted GPR value and standard deviation at the point x , respectively; Φ and ϕ denote normal cumulative distribution function (CDF), and probability distribution function (PDF).

In addressing the constrained optimization case study, the Bayesian optimization framework can be directly applied to substitute both the objective function and feasibility function separately. Nevertheless, this method might inadvertently result in an increased sampling requirement. During the construction of the GPR model to characterize the feasible region formed by constraints, the search may be directed

toward the feasible region but with high objective values (in the context of a minimization problem). Conversely, when building the GPR for the objective, efforts may be focused on the region characterized by low objective values but within the infeasible region.

To mitigate this issue, the EI function for the constraints $EI_{feas}(x)$ in Equation 3.20 is penalized to account for the influence of the objective function value [210]. This is accomplished by multiplying $EI_{feas}(x)$ with the probability that the next sample point would yield an objective value $f(x)$ better than current optimum solution f_{min} , as illustrated in Equation 3.21. It is worth mentioning that the EI function for the feasibility analysis $EI_{feas}(x)$ is modified and derived based on Equation (3.19), with the objective of achieving a balance in sampling between regions that have not been adequately explored and regions where the probability of the predictor being equal to zero is maximized [211]. The EI for the objective $EI_{obj}(x)$ in Equation 3.22 is penalized to restrict the search within the feasible region, as expressed in Equation 3.23. This strategy is similarly applied by incorporating the probability of constraints being satisfied.

$$EI_{feas}(x) = \hat{s}_{feas} \phi\left(-\frac{\hat{y}_{feas}}{x}\right) = \hat{s}_{feas} \cdot \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{0.5\hat{y}_{feas}^2}{\hat{s}_{feas}^2}\right) \quad 3.20$$

$$EI_{feas}^p(x) = EI_{feas}(x)P[f(x) \leq f_{min}] = EI_{feas}(x)\Phi\left(\frac{f_{min}-\hat{y}_{obj}}{\hat{s}_{obj}}\right) \quad 3.21$$

where \hat{y}_{feas} and \hat{s}_{feas} represent the GPR predictor for feasibility function and the corresponding uncertainty at the point x , respectively.

$$EI_{obj}(x) = (f_{min} - \hat{y}_{obj}) \cdot \Phi\left(\frac{f_{min}-\hat{y}_{obj}}{\hat{s}_{obj}}\right) + \hat{s}_{obj} \cdot \phi\left(\frac{f_{min}-\hat{y}_{obj}}{\hat{s}_{obj}}\right) \quad 3.22$$

$$EI_{obj}^p(x) = EI_{obj}(x)P[\psi(x) \leq 0] = EI_{obj}(x)\Phi\left(\left(0 - \frac{\hat{y}_{feas}(x)}{\hat{s}_{feas}(x)}\right)\right) \quad 3.23$$

where \hat{y}_{obj} and \hat{s}_{obj} are the GPR predictor and uncertainty for the objective function at the point x ; $\psi(x) \leq 0$ defines the feasible region.

Moreover, the entire optimization process is divided into two stages: the feasibility stage and the optimization stage, as depicted in Figure 3.2. Generally speaking, the proposed framework starts with the feasibility stage, where the primary focus is on accurately identifying the feasible regions with small objective values. Upon meeting the stopping criteria for this stage, the algorithm transitions to the optimization stage to explore optimum points within the previously identified feasible region. Different stopping criteria are utilized and compared based on the computational time. The first criterion is determined solely by the maximum iteration number, while the second criterion is more advanced, taking into account not only the maximum iterations but also considering improvements in the El_{feas}^p and the objective function. Throughout these stages, the GPR models for the objective and feasibility function are continuously updated, and the optimal solution is returned from the final refined GPR models.

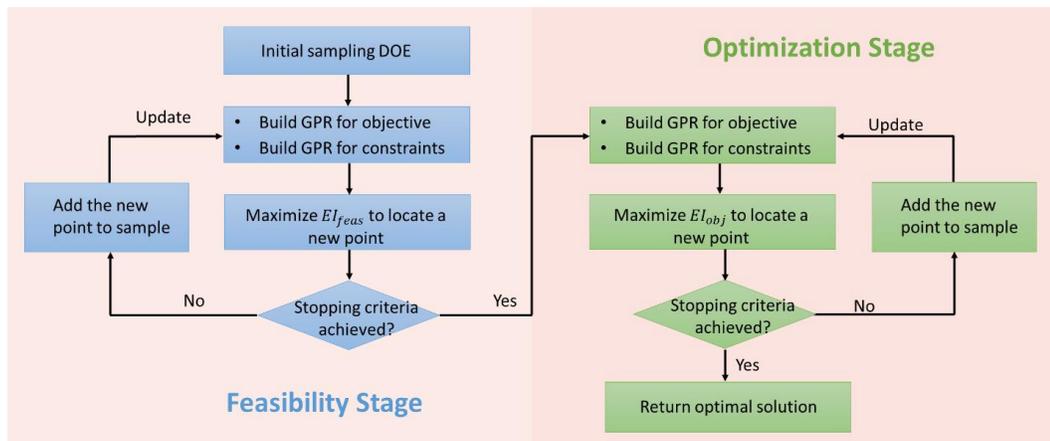


Figure 3.2: Proposed machine-learning (ML) based optimization framework.

In order to showcase the efficiency of the framework, we employed the commonly used global optimization solver, genetic algorithm (ga), to solve the case study for a comparative analysis. The function and constraint tolerances for ga were configured at 1×10^{-9} and 1×10^{-6} , respectively, to enhance the precision of the optimal results. All other options for ga were maintained at their default settings. The mechanistic model was solved using CADET, an open-source software available at <https://github.com/modsim/CADET>. To solve the coupled PDE equations, the axial column dimension was divided into 100 cells, while the radial bead was discretized into 20 cells. The Gaussian Process Regression model was developed using the DACE toolbox [194] built-in *MATLAB*. In the case of the feasibility function, an exponential correlation model was employed, with the hyperparameter being 0.1984. Regarding the objective function, a linear correlation model was selected with the hyperparameter being 0.3150. All the simulations were executed in *MATLAB R2019b* on Intel® Xeno® E-2274G CPU with 32.0 GB RAM.

3.3 Results and Discussion

3.3.1 Comparison of Optimization Results

3.3.1.1 Optimization Results using ga

To establish a baseline scenario for comparison, genetic algorithm (ga) was chosen to solve the constrained optimization problem, with the aim of identifying optimal elution strategies that maximize productivity while satisfying both the quality and operational constraints. Using ga, the optimized decision variables, objective value, column performance constraints, and simulation duration are summarized in Table 3.3. The optimization using pre-defined tolerances and computer settings as stated in the

previous section, required approximately 125.5 hours, yielding maximum productivity of $3.59 \times 10^{-3} \text{ mol m}^{-3} \text{ s}^{-1}$. The obtained yield and purity are 81.45% and 92.12%; respectively, meeting the specified constraints. The full and enlarged elution behavior of three proteins under this strategy is illustrated in Figures 3.3a and 3.3c, with the gray area representing the pooling interval. It could be clearly observed that the isocratic elution does not contribute to the separation and could be removed in practical experiments. It is worth noting that within the context of this case study, a higher number of elution steps does not necessarily result in a more favorable outcome. This observation suggests that the utilization of modeling and optimization can mitigate redundant experimental efforts and enhance our understanding of the underlying process.

Table 3.3: Optimized operating conditions and objectives using ga and ML-based framework.

Optimized operating conditions					
Approach	Δt_1 (s)	Δt_2 (s)	m_1	m_2	c_{init0} (mol m^{-3})
ga	4.83×10^3	2.01×10^3	2.2×10^{-3}	4.60×10^{-1}	51.5
ML-based Framework	2.90×10^3	1.00×10^3	5.3×10^{-3}	8.83×10^{-1}	65.0
Column performance metrics					
Approach	Yield (%)	Purity (%)	Productivity ($\text{mol m}^{-3} \text{ s}^{-1}$)	Simulation time (hrs)	
ga	81.45	92.12	3.59×10^{-3}	~125.5	
ML-based Framework	80.46	91.27	5.41×10^{-3}	~36.7	

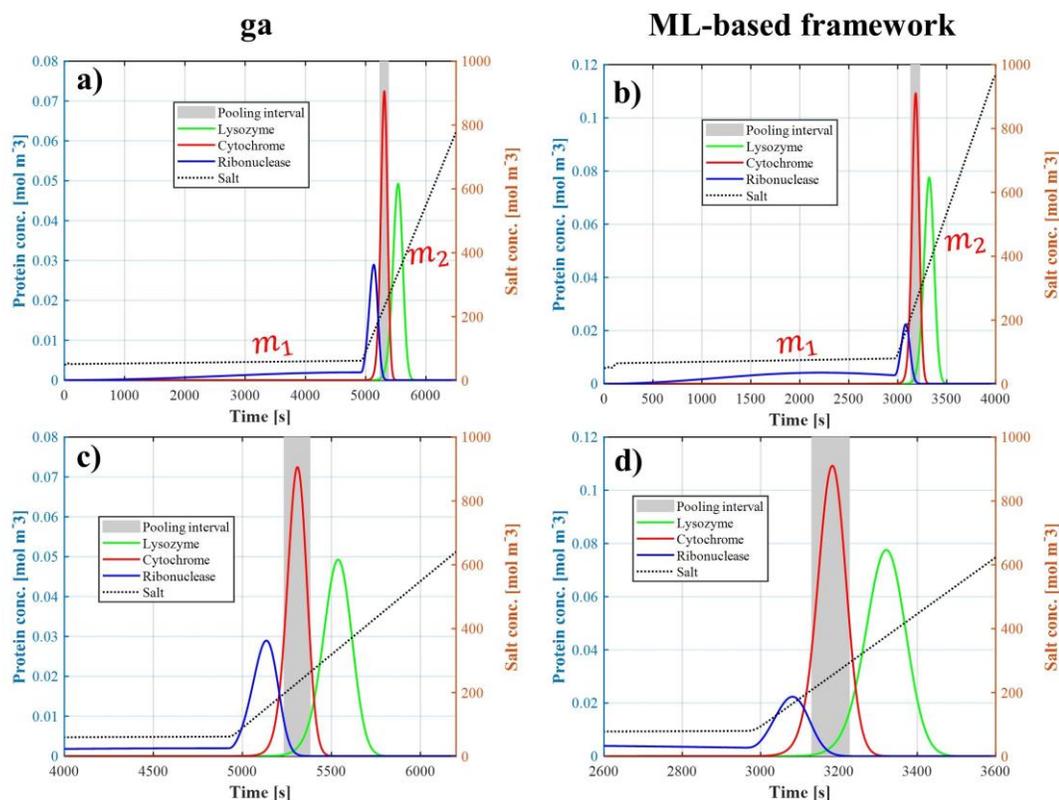


Figure 3.3: The full chromatograms at the optimized operating conditions obtained from *ga* (a) and the ML-based framework (b), with enlarged plots displayed in (c) and (d), respectively.

3.3.1.2 Optimization Results using ML-based Framework

To reduce the computational complexity and simplify the model, the ML-based framework was proposed and applied in the case study for the purification of ternary protein components. Specifically, 2000 points were sampled using Latin Hypercube Sampling (LHS) for the initial development of GPR models, followed by 500 iterations in the feasibility and optimization stages to improve the GPR models. In this work, a large number of sampling points were used to train the surrogate models to ensure the surrogate accuracy and optimization results. However, in real-case scenarios when the

process understanding is limited or unavailable, a promising approach is the use of hybrid modeling. In these cases, a neural network (NN) or different machine learning methods could be employed to represent the missing knowledge, followed by combining this NN with mechanistic model built on the available process understanding [199]. On the other hand, when no first principles models are available, physics-informed neural networks, by incorporating knowledge from physical laws or principles during the training process, could be utilized [139].

The detailed optimization results and corresponding chromatograms are illustrated in Table 3.3 and Figure 3.3b. It should be pointed out that the productivity achieved through the proposed framework is $5.41 \times 10^{-3} \text{ mol m}^{-3} \text{ s}^{-1}$, demonstrating a 50.1% improvement compared to that obtained using ga. Furthermore, the simulation time for the ML-based framework was only 36.7 hours, which is 29% of the time required by ga. Under the optimized conditions, the corresponding yield and purity are 80.46% and 91.27%; respectively. As the obtained yield approaches the boundary constraint (80%), it is important to note that the yield constraint is active under the current defined peak cutting criterion (75%). In contrast, the yield obtained from ga (81.45%) is not near the boundary, suggesting that the optimized results might be suboptimal. Although the simulation ends at 15,000 s, the total process time for the specified operating condition is defined as the duration until all the proteins are completely eluted from the column. It should be mentioned that although the multi-step (bilinear + isocratic) elution strategies are employed to maximize productivity, the final isocratic step does not yield any advantage in the separation process. This phenomenon could be attributed to the purity requirement, as will be discussed later. When analyzing the optimized elution behaviors using the two approaches, it is evident that the total

processing time to complete the separation process using the ML-enhanced framework (~3,600 s) is significantly shorter than that required by ga (~ 6,000 s).

From Figure 3.3 and Table 3.3, it is obvious that the first slope (m_1) is considerably smaller in comparison to the second slope (m_2) at the optimal solutions. In order to assess the potential impact of the initial linear elution step on the optimization results, the chromatograms after the removal of the first linear step are simulated based on the optimal operating conditions for the base case study, as illustrated in Figure 3.4. The comparison of the column performances obtained using the multi-step and single linear step strategies is provided in Table 3.4. It can be observed that the obtained purity and yield significantly decrease after removing the initial linear gradient elution step. This observation strongly indicates that the first elution step plays a crucial role in influencing the performance of the column.

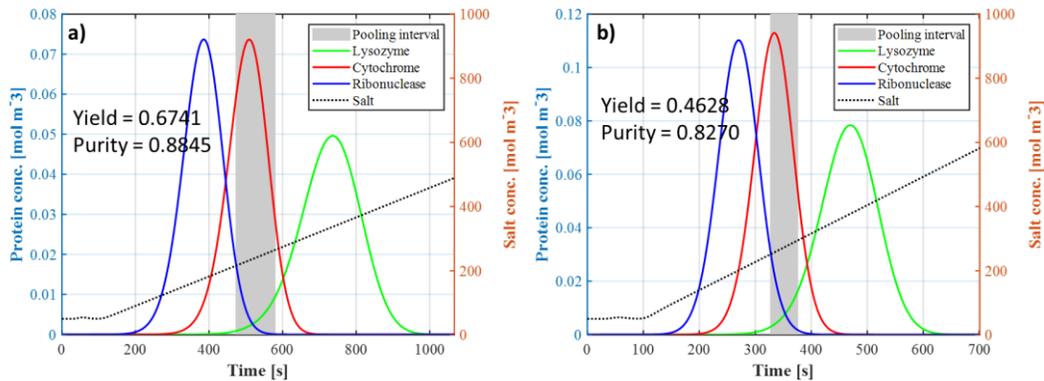


Figure 3.4: The new chromatograms generated after the removal of the first linear elution step at the optimal operating conditions obtained from ga (a) and the ML-based framework (b).

Table 3.4: Comparison of obtained yield and purity using multi-step and single linear elution strategies at the optimal operating conditions.

Optimization approach	Elution strategies	Yield (%)	Purity (%)
ga	Multi-step	81.45	92.12
	Single linear	67.41	88.45
ML-based Framework	Multi-step	80.46	91.27
	Single linear	46.28	82.70

To further investigate the impact of the first step on the elution profiles, various slope values were selected to simulate the chromatograms, as depicted in Figure 3.5. For additional quantitative analysis, the area under the curve for the impurity Ribonuclease at different elution steps is calculated and provided in Table 3.5. It is noticeable that with increasing values of the first slope, the eluted amount of Ribonuclease during this step increases, reflected by a rise in its area under the curve. As more impurity Ribonuclease is eluted out from the column during the initial step, an increase in the obtained purity and yield is observed.

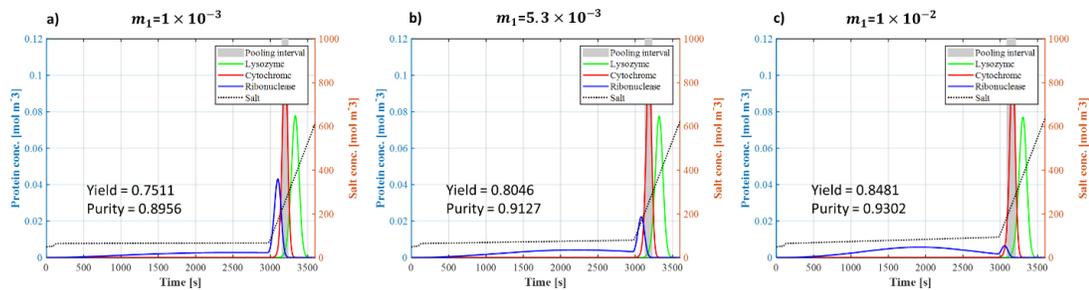


Figure 3.5: Different simulated chromatograms at different selected slope values for the first linear elution step.

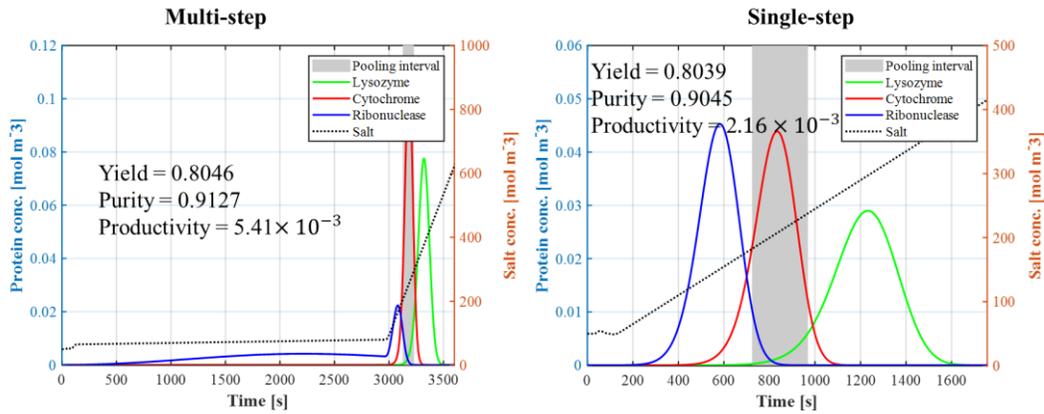


Figure 3.6: Comparison of the obtained chromatogram at the optimal solution for the multi-step and single step case study.

Table 3.5: Comparison of yield, purity, and area under the curve of Ribonuclease at various slope values.

m_1 ($\text{mol m}^{-3} \text{s}^{-1}$)	Yield (%)	Purity (%)	Area under the curve of Ribonuclease ($\text{mol m}^{-3} \text{s}$)	
			1 st linear gradient elution	2 nd linear gradient elution
1×10^{-3}	75.11	89.56	5.05	4.95
5.3×10^{-3}	80.46	91.27	7.45	2.55
1×10^{-2}	84.81	93.02	9.28	0.72

To provide further insights into the step elution strategies, an updated case study using only a single linear elution step was evaluated using the ML-based framework. This study also introduced the constraint for total processing time and incorporated peak cutting as a decision variable. It is found that the optimized productivity would decrease significantly from 5.41×10^{-3} to $2.16 \times 10^{-3} \text{ mol m}^{-3} \text{ s}^{-1}$ if multi-step elution strategies are simplified to a single linear gradient elution, as displayed in Figure 3.6. These results align well with our previous analyses, indicating that the first linear elution

step can have a significant impact on both purity and yield, consequently affecting the overall productivity.

In the optimization problem statement for the base case, where purity and yield requirements are set at 90% and 80%, respectively, the final isocratic elution stage does not contribute to the separation process, as previously mentioned. This may be attributed to the relatively low purity requirement. To explore the influence of purity constraints on the elution strategies, the purity requirement is raised to 95%, along with the decrease of yield requirement to 65% in order to ensure feasibility. All other process conditions remain unchanged, and the ML-based optimization framework is used. The obtained optimal chromatogram can be found in Figure 3.7. Given a higher purity specification, the slope value for the second gradient elution step (0.024) decreases remarkably, compared with the value (0.883) observed at a 90% purity requirement. In this scenario, the two linear gradient elution strategy targets the elution of the first impurity (Ribonuclease), and the final isocratic elution stage is crucial in separating the target protein (Cytochrome) and the additional impurity (Lysozyme). This result is consistent with the finding reported in [90]. From Figure 3.7b, it could be found that the separation is not completed when the total processing time is 15,000 s. However, it is important to note that this example serves primarily to investigate the impact of purity requirements on optimized elution strategies. If further exploration is desired with a higher purity criterion, additional adjustments can be made. For instance, increasing the peak cutting threshold and introducing constraints on eluted protein concentrations below a specified tolerance concentration, as detailed in the single elution case study previously.

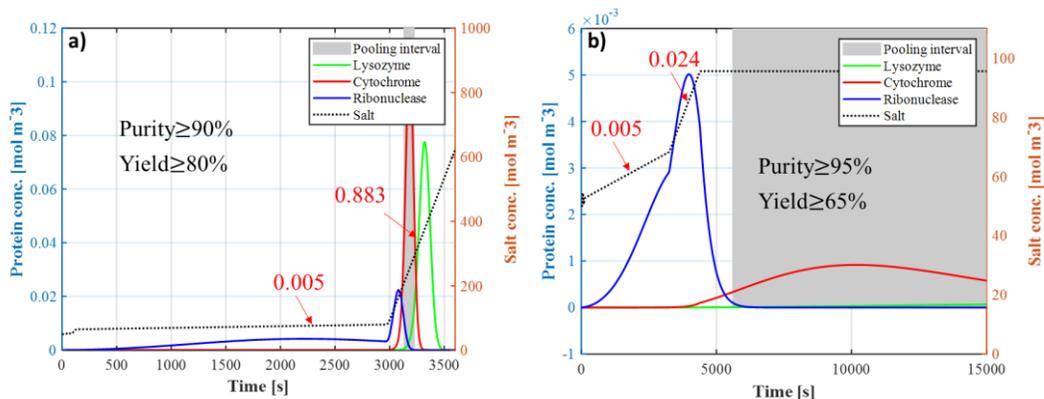


Figure 3.7: Comparison of the obtained optimal chromatograms under different purity and yield requirements: (a) Purity requirement of 90% and yield requirement of 80%; (b) Purity requirement of 95% and yield requirement of 65%.

To improve the efficiency of the ML-based framework, a more intelligent stopping criterion has been implemented and evaluated instead of solely relying on the maximal number of iterations. During the feasibility stage, the stopping criteria are defined by either reaching the maximum number of iterations (500) or achieving a low average of expected improvement for every 10 consecutive iterations (1×10^{-2}). Similarly, for the optimization stage, the stopping criteria are defined by either reaching the maximum number of iterations (500) or observing a low improvement in the objective function value every 10 iterations (1×10^{-6}). Under the updated stopping criterion, the optimized productivity reaches $5.21 \times 10^{-3} \text{ mol m}^{-3} \text{ s}^{-1}$, similar to the one ($5.41 \times 10^{-3} \text{ mol m}^{-3} \text{ s}^{-1}$) obtained using the maximum iterations as the stopping criteria. It is worth noting that the stochastic nature inherent in both the initial and adaptive sampling processes may lead to slightly different optimal solutions. The minor difference observed, with only a 3.7% deviation in the optimized productivity, indicates the robustness of the ML-based framework. Compared with the stopping criterion of

using maximum iterations, the updated stopping criterion results in a significant 48% reduction in the computational time required for the feasibility and optimization stages. This results in a total computational time reduction of 28.6% compared to exclusively depending on the maximum iterations.

3.3.2 Effect of the Peak Cutting Criteria

In the optimized case study, the yield constraint is active when the peak cutting criterion is set at a 75% target protein threshold (i.e., cytochrome). In an attempt to investigate the possibility of the active constraint shifting from yield to purity, the percentage threshold decreased from 75% to 65%. The updated optimization problem was solved using the same ML-based framework while maintaining the same operational and quality constraints. The optimized results at both percentage thresholds are summarized in Table 3.5.

As the percentage threshold for peak cutting decreases, the purity constraint becomes active, with the achieved purity (90.66%) reaching the boundary limit (90%). The relaxation of the percentage peak cutting criterion allows for the collection of more products, as evidenced by the higher yield (88.33%), but it is accompanied by a compromise in purity. In this particular case study, a reduction in the percentage threshold for the target protein results in a decreased productivity by 16.2%. The comparison of the chromatograms under different peak cutting criteria is illustrated in Figure 3.8. It could be clearly observed that in this specific scenario, a reduction in the percentage threshold leads to a delayed occurrence of the overall elution behavior at the optimal operating conditions.

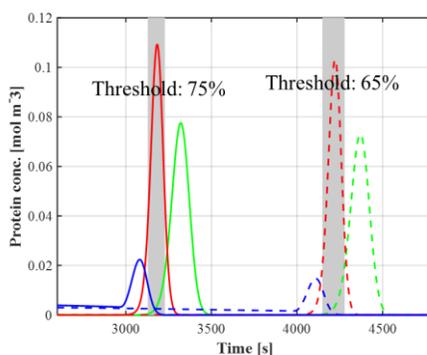


Figure 3.8: Comparison of the chromatograms under the optimized conditions at various percentage thresholds. The blue, red and green lines denote the column outlet concentration of lysozyme, cytochrome, and ribonuclease.

Table 3.5: Comparison of the optimized results at different percentage thresholds.

Optimized operating conditions					
Threshold	Δt_1 (s)	Δt_2 (s)	m_1	m_2	c_{init0} ($mol\ m^{-3}$)
75%	2.90×10^3	1.00×10^3	5.3×10^{-3}	8.83×10^{-1}	65.0
65%	3.88×10^3	1.07×10^3	1.0×10^{-3}	8.08×10^{-1}	67.4
Column performance metrics					
Threshold	Yield (%)	Purity (%)	Productivity ($mol\ m^{-3}\ s^{-1}$)		
75%	80.46	91.27	5.41×10^{-3}		
65%	88.33	90.66	4.53×10^{-3}		

3.3.3 Design Space Analysis

In previous sections, the optimization framework based on machine learning was employed to facilitate the identification of optimal operating conditions for the separation of a ternary mixture of proteins using CEX. The application of this framework could assist in finding an improved operating point characterized by higher

productivity using reduced computational resources. Moreover, this ML-based framework could contribute to a reduction in the sampling budget by guiding the algorithm to concentrate on feasible regions characterized by small objective values during the feasibility stage. This section aims to analyze the design space of the process derived using the proposed approach.

The design space, characterized by five decision variables, poses a challenge for direct visualization. Consequently, the GPR model for the feasibility function is utilized to generate 5D feasible regions, which are subsequently projected onto 2D subspaces at the optimal solutions. The resulting ten sets of 2D contour plots are illustrated in Figure 3.9. The boundary of the feasible region is represented by the line where the feasibility function value equals 0, visually represented by the red line, and the design space encompasses the area within this line. Within this context, the red square signifies the optimal point obtained using the framework and this point is located within the bounds of feasible regions. Under the optimal operating conditions, it is observed that m_2 exhibits a broader feasible region, suggesting the system's capacity to endure variations in the slope of the second linear elution step while maintaining robustness. The remaining variables have much smaller feasible ranges, mainly due to the constraint imposed by the salt limitation, especially at large m_2 value. This analysis contributes to an enhanced comprehension of the underlying processes and facilitates the identification of input variables with a higher likelihood of causing constraint violations.

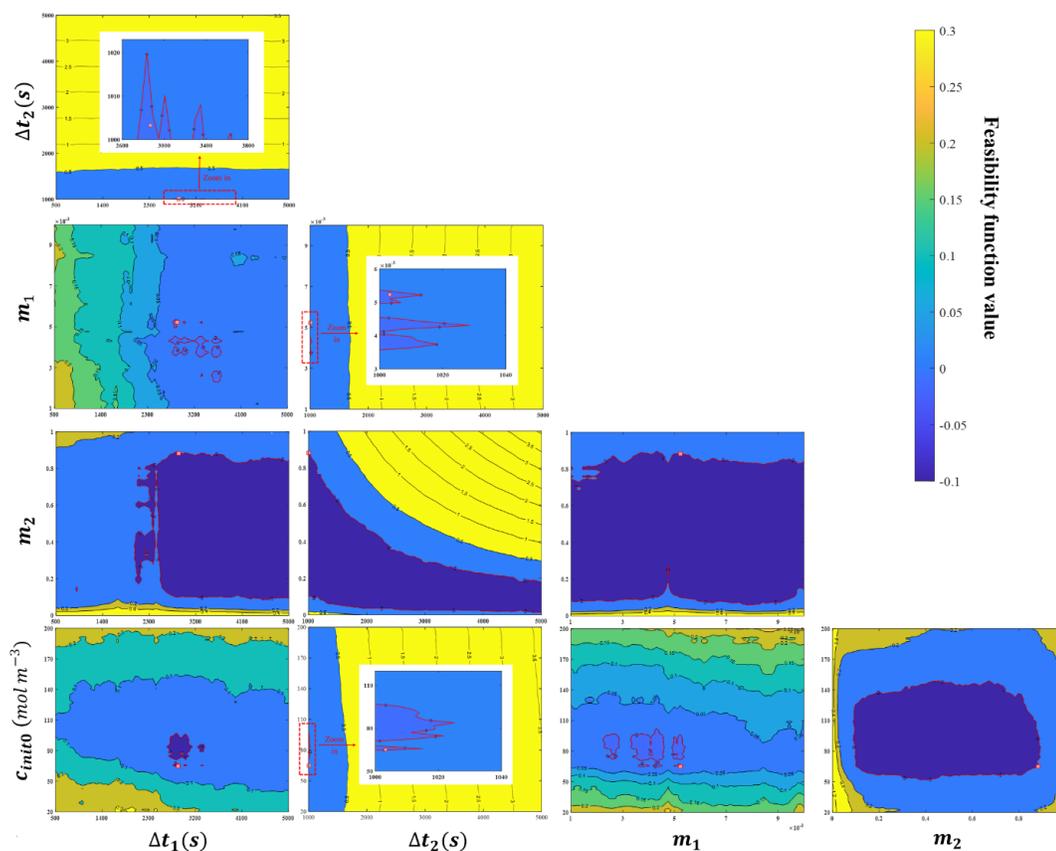


Figure 3.9: 2D contour plots of the design space projected from high dimensional space at the optimal solutions. The red line represents the feasible boundary identified using the GPR model, and the red square denotes the optimal solution obtained from the ML-based framework.

To provide a more comprehensive analysis, two specific contour plots are chosen from the set of ten for a detailed explanation, as shown in Figure 3.10. The red line indicates the salt constraint, defining the threshold where the salt concentration should be maintained below 1 M. According to our framework, the accuracy of feasible regions in proximity to optimal solutions is expected to be high, whereas regions far from the optimal solution may exhibit increased uncertainty and reduced accuracy.

Thus, points are sampled close to the feasible boundary, and the feasibility function values at these points are then calculated using both the trained GPR model and the original mechanistic model. Following this, the results are compared, and if both values exhibit the same sign, the point can be considered to be accurately identified. Given that a feasibility function value less than 0 signifies the infeasible regions and a value larger than 0 represents the feasible regions, the exact values are trivial and only the sign of the values holds significance.

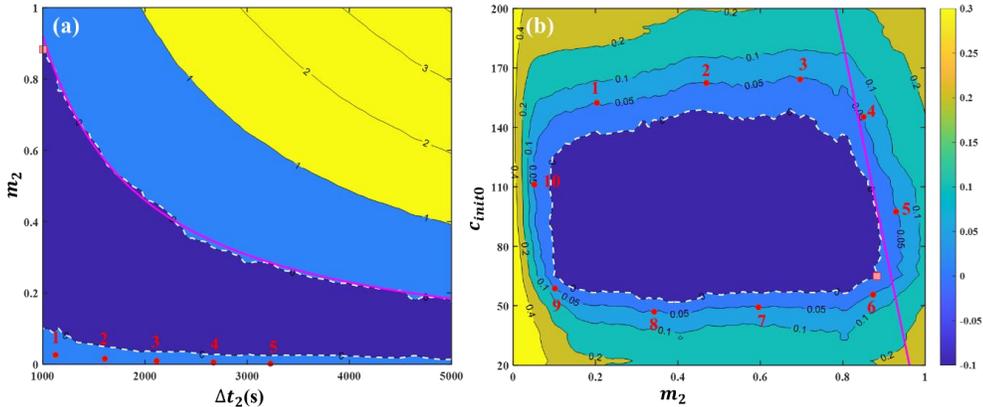


Figure 3.10: Selected 2D contour plots of the design space to illustrate the framework. The white dash line is the feasible boundary with the feasibility function value of 0, while the red solid line indicates the salt constraint. The red square and circles are optimal solution and sampled points, respectively.

Figure 3.10a shows the 2D contour plot between Δt_2 and m_2 under the optimized operating condition. In this scenario, salt constraint establishes one of the feasible boundaries, which is near the location of the optimal solution. Thus, all the points are sampled close to the other boundary, where the feasibility function value is approximately 0.3, and the comparison is summarized in Table 3.6. Out of the five

sampled points, all are identified as infeasible using the GPR model, while two of them are accurate according to the results from the mechanistic model. The same phenomena are observed for the sampled points in Figure 3.10b. From Table 3.7, it can be found that points 9 and 10, which are away from the optimal solution, exhibit low accuracy, whereas the remaining 8 points near the optimal solution demonstrate high accuracy. This observation aligns with the expectation of our framework, whereas the focus of sampling is centered towards the area close to the optimal solution. Given the primary objective of optimizing productivity under constraints, the analysis has demonstrated that the framework can yield an accurate design space around the optimum.

Table 3.6: Comparison of feasibility function values for the sampled points in Figure 3.10a.

	Feasibility function value		Matching?
	GPR model	Mechansitic	
Point 1		0.39 (+)	Yes
Point 2		0.13 (+)	Yes
Point 3	~0.3 (+)	-0.04 (-)	No
Point 4		-0.03 (-)	No
Point 5		-0.03 (-)	No

Table 3.7: Comparison of the sign of feasibility function values for the sampled points in Figure 3.10b.

	Sign of feasibility function value		Matching?
	GPR model	Mechansitic	
Points 1-8	+	+	Yes
Points 9-10	+	-	No

3.4 Conclusions

In this chapter, we introduced a ML-based optimization framework to address the nonconvex and nonlinear constrained optimization challenges encountered in biopharmaceutical separation. The framework utilized GPR models to replace the complex mechanistic model in handling constraints and defining objective function. To enhance efficiency and minimize sampling demands, the framework was strategically divided into feasibility and optimization stages. The feasibility stage concentrated on exploring the feasible region with smaller objective values, while the optimization stage focused on identifying optimal solutions within this defined feasible space.

To evaluate the efficacy of this framework, it was applied to a case study involving the separation of a ternary protein mixture comprising ribonuclease, cytochrome and lysozyme. The objective of this case study was to find the optimal elution strategies by maximizing productivity under the salt, yield and purity constraints. In comparison to the optimized productivity achieved through the genetic algorithm, this method not only demonstrated superior productivity but also accomplished the optimization process within a shorter simulation time. Specifically, the productivity optimized by the ML-based framework improved by 50.1%, accompanied by a 70.8% reduction in running time. The impacts of both the first slope values and purity requirements on the overall elution behaviors were comprehensively investigated, along with an examination of the stopping criteria.

The impact of varying the percentage peak cutting thresholds was examined to discern the critical constraint influencing the optimized conditions, which might help enhance our understanding of the process and facilitate potential improvements in the overall process. Through this analysis, it was observed that the active constraint transitions from yield to purity as the percentage threshold decreases from 75% to 65%.

Furthermore, a comprehensive analysis of the design space identification through our framework was carried out. The result demonstrated the framework's capability to accurately identify the design space near the optimal solution, which aligns with our expectations.

The case study examined in this work serves as an illustrative example to showcase the applicability of the proposed framework. The methodology presented in this study is generalizable and can be applied to other complicated case studies, providing valuable guidelines for addressing problems where identifying optimal solutions is challenging. However, it is important to note that the mechanistic model should be fully calibrated and validated before being utilized to generate the in-silico datasets as the accuracy of this framework highly depends on the accuracy of mechanistic model.

Chapter 4

HYBRID MODEL DEVELOPMENT FOR HYDROPHOBIC INTERACTION CHROMATOGRAPHY

4.1 Introduction

The biopharmaceutical industry has received considerable attention recently, as can be witnessed by the growing market demands and approvals of biological drugs by the EU and the US regulatory agencies [1-4, 155]. During biologics manufacturing, the formation of protein aggregates during production and separation processes remains a primary concern [212, 213]. Hydrophobic interaction chromatography (HIC) is one of the widely used techniques in downstream polishing steps for the separation of targeted monomeric forms of protein therapeutics from the dimeric and/or multimeric species [214]. HIC is an entropically-driven process that exploits the difference between the hydrophobicity and particle sizes of the monomer and aggregates to achieve separation [215]. Despite being commonly employed as an efficient purification strategy to remove aggregates, the mechanism for HIC adsorption is quite complex, relying on various process parameters, like pH, salt concentration, and adsorbent ligand hydrophobicity [216, 217]. Recently an automatic workflow for HIC method development is presented

Chapter 4 is adapted from the following journal article written by the author of this dissertation:

Ding, C., Gerberich, C., Ierapetritou, M., Hybrid model development for parameter estimation and process optimization of hydrophobic interaction chromatography. *Journal of Chromatography A*, 2023, 1703: 464113.

by combining screening with *in-silico* modeling by Barrientos et al. [218]. However, the current process development of HIC still largely depends on rules of thumb or tedious experimental screening.

Inspired by the quality by design (QbD) initiative and the need to reduce the effort of performing time- and resource-intensive experiments, mechanistic modeling has become an important tool for process characterization, development and optimization [13, 41, 189]. Mechanistic models are built based on the understanding of the chromatographic process physics, which mainly focuses on describing the solute transport in the mobile phase and adsorption in the stationary phase [74, 86, 89]. With different assumptions and simplifications, several variants of mechanistic models are used to describe the fluid flow inside the mobile phase, including equilibrium dispersive model, lumped kinetic model, and general rate model (GRM) [81-83]. In HIC modeling, the fluid dynamics are well-described by physics, but the underlying adsorption mechanism for the salt-dependent protein-ligand interaction is still unclear, which makes it challenging to postulate appropriate mathematical equations to describe the overall process [219]. To capture the role of salt, various adsorption laws like solvophobic and preferential interaction theories have been used [220-222]. Although the simple Langmuir isotherm is popular and frequently considered, this equation cannot account for the effect of salt concentration on adsorption behavior [74]. To overcome this limitation, the Langmuir isotherm is often modified to capture the dependence of adsorption on salt concentrations for the representation of the HIC adsorption process [223]. Moreover, Wang et al. [219] updated the isotherm by introducing the salt-dependent hydration number of ions and applied in modeling the system of glucose oxidase, bovine serum albumin, and lysozyme.

The challenge associated with the development of a precise mechanistic model is to identify the exact underlying physicochemical phenomena. Due to a limited understanding of the HIC adsorption mechanism, there is an absence of a reliable mechanistic model to describe the interaction between proteins and ligands under varying salt ions [79, 223]. Consequently, a hybrid modeling strategy is a promising alternative to accurately describe HIC process and reduce the model development effort as this approach can exploit the available information about the process and represent the missing knowledge by a data-driven component [126, 135]. By combining the mechanistic and data-driven models, the constructed hybrid model can retain the advantages of the mechanistic model (i.e., physical interpretability and generalizability) while being able to extract information from available data that cannot be captured by the first-principle knowledge [24, 127]. Neural network (NN) is a commonly used data-driven model, which is based on the neural structure of the human brain [147, 224]. NN is composed of many interconnected neurons in layers, with weights assigned to each interconnection. Narayanan et al. [136, 137] developed different hybrid models by combining varying degrees of process knowledge with NN and evaluated their performance in terms of the interpolation and extrapolation capabilities in Protein A chromatography process. It was found that the developed hybrid model outperformed the mechanistic model in terms of prediction accuracy and robustness. It should be noted though that the hybrid modeling strategy has not been applied to describe the complex salt-dependent adsorption mechanism for HIC process.

In this work, a mechanistic model is first developed to describe the HIC chromatographic process using an equilibrium dispersive model for hydrodynamics and modified isotherm derived by Wang et al. [219], followed by model validation with

experimental data. Due to the limited understanding of the underlying adsorption mechanism, a hybrid model is proposed by combining a simpler multi-component Langmuir isotherm (MCL) with a NN. Different methods to integrate the MCL with NN are investigated to find the appropriate hybrid model structure. During parameter estimation, a regularization strategy is incorporated to avoid overfitting and the effect of different NN structures and regularization rates is comprehensively investigated to acquire the hybrid model with the best performance. To ensure the generalizability of the developed hybrid model, an *in-silico* dataset is generated using the mechanistic model to test the extrapolation capability of the hybrid model. Finally, process optimization is conducted to find the optimal operating conditions under product quality constraints, and the optimal results obtained from the mechanistic and hybrid models are compared thoroughly.

4.2 Material and Methods

4.2.1 Experimental Setup

mAb A HIC load material was acquired, with a concentration of 11.5 mg/mL and an average dimer content of 1.2%. To enable binding to the HIC resin, 400 mM of salt was added to the load material. A HIC column with a height of 20.1 cm and an inner diameter of 0.486 cm was connected to an ÄKTA explorer (GE Healthcare) to purify the load material using a linear salt gradient, with the A buffer containing 425 mM salt and the B buffer containing 0 mM salt. The chromatographic sequence employed for this separation is shown in Table 4.1, and the resin loadings and gradient lengths were varied across four experiments as specified in Table 4.2. It should be noted that the 50 g/L loading runs exceeded the binding capacity of the HIC resin, so breakthrough was

observed. The linear salt gradient was fractionated, and then each fraction was analyzed using UV-vis spectroscopy to determine protein concentration and SEC-HPLC to determine monomer and dimer content. The flow rate for the four calibration runs was set at 300 cm/h for all steps in the sequence, and the validation run used a flow rate of 150 cm/h.

Table 4.1: HIC chromatographic sequence used in calibration and validation experiments.

Step	Volume (CV)	Buffer
Equilibration	3	425 mM modulating salt
Loading	Varies	mAb A material with 400 mM modulating salt added
Re-equilibration	3	425 mM modulating salt
Gradient	Varies	A Buffer: 425 mM modulating salt
Elution		B Buffer: 0 mM modulating salt
Strip	5	0 mM modulating salt
Cleaning	3	1.0 M NaOH
Storage	3	0.1 M NaOH

Table 4.2: Experimental operating conditions used for model calibration and validation.

	Experiment	Loading (g/L)	Gradient length (CV)	Flowrate (cm/h)
Calibration	Exp1	50	40	300
	Exp2	5	40	
	Exp3	50	10	
	Exp4	5	10	
Validation	Exp_Val	35	10	150

4.2.2 Mechanistic Model

Based on the hypothesis made and system characteristics, different models have been developed to describe the fluid dynamics and adsorption phenomena of the chromatographic process. Among those models, the equilibrium dispersive model (EDM) is the simplest one to reasonably simulate mass transfer [74]. It should be noted that such a model assumes the establishment of adsorption equilibrium, but a kinetic equation is chosen in this study to describe the complicated salt-dependent HIC isotherm. Thus, the kinetic-dispersive model (KDM), as a variant of EDM, is adopted as the mechanistic model. It is important to highlight that different papers may use different terms, such as transport dispersive model or lumped rate model, to describe the concepts that align with the assumptions of KDM [86, 188, 225]. Due to the complexity and limited understanding of the adsorption mechanism of HIC, finding an adequate isotherm to describe the system studied is a great challenge. In this work, a mechanistic HIC isotherm published by Wang et al. [219] was employed and further modified by adding a few parameters to better simulate the process.

4.2.2.1 Kinetic-Dispersive Model (KDM)

Similar to EDM, KDM is a lumped kinetic parameter model, with all mass transfer resistances and binding kinetics lumped into the isotherm parameters [196]. In this regard, the protein concentration in the bulk mobile phase is expected to be identical to the average concentration in the intra-particle mobile phase. The model also lumps axial dispersion and molecular diffusion into one single parameter (apparent axial dispersion coefficient) [226, 227].

Assuming a radially homogeneous column, the differential mass balance in the bulk moving phase can be represented by Equation 4.1.

$$\varepsilon_t \frac{\partial c_i}{\partial t} + v \frac{\partial c_i}{\partial z} + (1 - \varepsilon_t) \frac{\partial q_i}{\partial t} = D_{app} \varepsilon_e \frac{\partial^2 c_i}{\partial z^2} \quad 4.1$$

where c_i and q_i is the solute concentration of component i in the mobile and solid phases, v is the interstitial velocity, t and z denote the time and axial coordinate, ε_e and ε_t are the extra-particle and total porosity. Apparent axial dispersion coefficient D_{app} can be calculated by Equation 4.2, where N_t and H are the theoretical plate number and the height equivalent to a theoretical plate (HETP), respectively.

$$D_{app} = \frac{Lv}{2N_t} = \frac{Hv}{2} \quad 4.2$$

4.2.2.2 HIC Isotherm

To construct a complete mathematical formulation of the separation process, it is necessary to complement KDM with an adsorption isotherm. By considering the water structure, the salt-dependent HIC kinetic equation derived and developed by Wang et al. [219] was given in Equation 4.3, where $k_{a,i}$ and $k_{d,i}$ are the kinetic adsorption and desorption coefficients, $q_{max,i}$ is the maximum binding capacity, ν_i is the number of hydrophobic binding sites, $\beta_{0,i}$ is the ionic hydration number at infinite dilution, $\beta_{1,i}$ is the ionic hydration factor for the salt of component i , and c_s is the ionic salt concentration. It should be mentioned that the kinetic parameters $k_{a,i}$ and $k_{d,i}$ indirectly include mass transport resistances, but the ratio $k_{a,i}/k_{d,i}$ determines the equilibrium constant that is a thermodynamic parameter.

$$\frac{\partial q_i}{\partial t} = k_{a,i} c_i \left(1 - \sum_{j=1}^N \frac{q_j}{q_{max,j}} \right)^{\nu_i} - k_{d,i} q_i^{1+\nu_i \beta_{0,i} \exp(\beta_{1,i} c_s)} \quad 4.3$$

Inverse modeling was employed to fit this isotherm to the experimental data, but the fit was unacceptably poor. There were concerns that the model was not accurate enough to get usable process optimization results, so extensions of this isotherm were explored. To better capture the HIC system studied in this work, three additional

parameters (η_i , $q_{0,i}$, and $\beta_{2,i}$) are added to the above HIC isotherm, as shown in Equation 4.4.

$$\frac{\partial q_i}{\partial t} = k_{a,i} c_i^{\eta_i} \left(1 - \sum_{j=1}^N \frac{q_j}{q_{max,j}} \right)^{\nu_i} - k_{d,i} q_{0,i}^{1+\nu_i \beta_{0,i}} \left(\frac{q_i}{q_{0,i}} \right)^{1+\nu_i \beta_{0,i} \exp(\beta_{1,i} c_s + \beta_{2,i} c_m)} \quad 4.4$$

where η_i is the Freundlich index, $q_{0,i}$ is the reference concentration, $\beta_{2,i}$ is the ionic hydration factor for the monomer of i^{th} component, and c_m is the monomer concentration. The modified isotherm is an empirical model that can account for the dependence of the dimer adsorption on the monomer content, as can be seen from the last term c_m . By adding extra parameters, the adsorption model can help better represent the HIC separation process.

4.2.3 Hybrid Model

Since the fluid dynamics of the chromatographic process can be well described, KDM is used in its original formulation. The challenge in HIC mechanistic model development is to find an appropriate binding isotherm to describe the salt-dependent protein-resin interaction. Such development requires a high-level understanding of the process and significantly increases the effort required, as seen by the rather complex isotherm described in Section 4.2.2. Moreover, if the presumed mechanism differs from the true underlying mechanism, the formulated isotherm model may lose the capability to characterize the process and describe the experimental data. A potential approach to reduce such effort and increase the model flexibility is to employ a simpler isotherm, together with a universal approximator like a neural network (NN) to construct a hybrid model. The approximator utilizes the available data to infer the underlying mechanism of the process, which may reduce the effort in mechanistic model development phase [200, 228]. In such cases, even if the postulated simpler model differs significantly from

the experimental profiles, the mismatches may be reduced with the use of the NN. The gaps in process knowledge may therefore be abridged, saving time- and resource-intensive development efforts.

Based on this idea, multi-component Langmuir rate equation (MCL) [82] in Equation 4.5 is selected as the baseline adsorption mechanism as this equation is well-known and widely accepted. There are different strategies to include a NN in the MCL equation, leading to different hybrid model structures. In this work, two hybrid model structures are mainly considered and tested. The first hybrid model, referred to as the “Hybrid-MCL,” is established by adding a multiplier term (NN) into the whole Langmuir isotherm. As indicated in Equation 4.6, the NN is a function of mobile-phase concentration (c_i), solid-phase concentration (q_i), and salt concentration (c_s) at a given point in space and time. An alternative formulation of the hybrid model (referred to as the “Hybrid-Desorption”) is shown in Equation 4.7, in which only the desorption term is modified by multiplying a NN.

$$\frac{\partial q_i}{\partial t} = k_a c_i \left(1 - \sum_{j=1}^N \frac{q_j}{q_{max,j}} \right) - k_d q_i \quad 4.5$$

$$\frac{\partial q_i}{\partial t} = NN(c_i, q_i, c_s) \left[k_a c_i \left(1 - \sum_{j=1}^N \frac{q_j}{q_{max,j}} \right) - k_d q_i \right] \quad 4.6$$

$$\frac{\partial q_i}{\partial t} = k_a c_i \left(1 - \sum_{j=1}^N \frac{q_j}{q_{max,j}} \right) - k_d q_i NN(c_i, q_i, c_s) \quad 4.7$$

It is worth noting that since the incorporation of NN should not alter the physical meaning of the binding process, two different operators (e.g., exponential and absolute) are added to the NN. Unless otherwise stated, the absolute operator is enforced to the NN output.

4.2.4 Determination of Model Parameters

Based on the model mentioned in previous sections, the parameters could be classified into mass transfer and isotherm parameters. The first type of parameters, including ε_e , ε_t , and H , are used in the kinetic-dispersive model. The other type of model parameters is used to describe the adsorption behavior of the HIC process. For the mechanistic model, there are in total 9 adsorption parameters ($k_{a,i}$, $k_{d,i}$, $q_{max,i}$, η_i , ν_i , $q_{0,i}$, $\beta_{0,i}$, $\beta_{1,i}$, and $\beta_{2,i}$) for each component (monomer or dimer). In terms of the hybrid model, the number of isotherm parameters is related to the NN structures and will be discussed in Section 4.3.2.

4.2.4.1 Column Characteristics

To determine the column flow characteristics, two pulse injections were performed prior to running the purification experiments. A 100 μL pulse of 2.0 M NaCl (Sigma-Aldrich) was injected into a running solution of 0.4 M NaCl at a flow rate of 300 cm/h. The retention volume of this peak was used to determine ε_t , and the method of moments was also applied to this peak to determine H . To determine ε_e , a 100 μL pulse of 100 nm gold nanoparticle suspension (Sigma-Aldrich) was injected into a running solution of DI water at a flow rate of 300 cm/h, and the retention volume of the peak was calculated.

4.2.4.2 Parameter Estimation

The remaining adsorption parameters can be estimated by inverse fitting the experimental bind-and-elute curves. The flowchart of the procedure for parameter estimation using the hybrid model is illustrated in Figure 4.1. A similar scheme can also be used for the parameter estimation of mechanistic models by simply substituting the hybrid model block with the model of interest. Specifically, all parameters (adsorption

parameters like k_a , k_d , and q_{max} in MCL equation, NN parameters like weights and biases) are first randomly initialized, followed by solving the coupled PDE equations *via* built-in PDE solver `pdepe` in MATLAB 2019b [229]. The optimal isotherm parameters are obtained by iteratively solving the nonlinear optimization problem that minimizes the user-defined cost function. After parameter estimation, the developed hybrid model is validated to evaluate its predictive ability.

To find the optimal parameters and assess model performance, a normalized sum of squared error (NSSE) metric is used. Under this metric, the commonly used sum of squared error function is normalized across multiple experiments because the output concentrations are of different orders of magnitude. Since the constructed model is nonlinear and nonconvex, a global optimization strategy is used to avoid getting trapped in the local minimum. The differential evolution (DE) algorithm is frequently considered one of the most popular approaches to address such complicated optimization problems in many scientific and engineering problems [230]. DE is a stochastic population-based metaheuristic search algorithm that drives the population toward better solutions by applying the operators of mutation, crossover, and selection. As one of the most efficient DE variants, successful history-based adaptive differential evolution with linear population size reduction (L-SHADE) is adopted for parameter estimation in this work as this algorithm exhibits excellent performance [231]. Unless explicitly specified, the optimization results in Section 4.3 are obtained using L-SHADE. Other global optimization algorithms such as MultiStart `fmincon` [232], genetic algorithm (GA) [233] and particle-swarm optimization (PSO) [234] are also tested to estimate parameters of hybrid model. Due to the stochastic nature of those

global optimization algorithms, simulations for parameter estimation are run multiple times to ensure the robustness of the model.

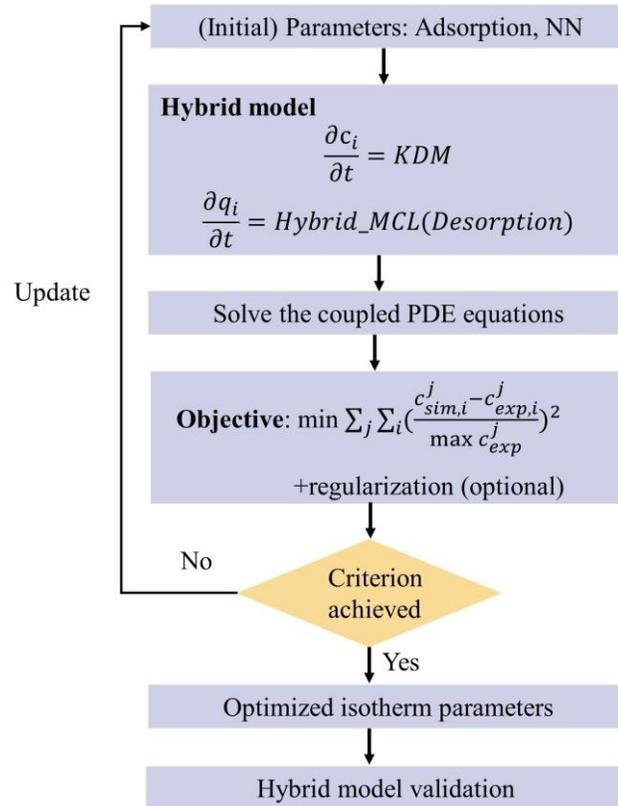


Figure 4.1: The overall schematic of parameter estimation using hybrid model.

Regularization strategy is employed to prevent overfitting and obtain a generalizable framework where the structural complexity of the NN model is also minimized. The complexity of NN could be quantified using L_2 regularization, which defines the regularization term as the sum of the squares of all weights. This term is added as a penalty term to the loss function to form a new cost function, as indicated in

Equation 4.8, where λ and β represent the regularization rate and feature weights of NN, respectively. It is worth noting that choosing an appropriate regularization rate is important to achieve the right balance between underfitting and overfitting. Accordingly, the effect of different regularization rates on the fitting results is evaluated.

$$\sum_j \sum_i \left(\frac{c_{sim,i}^j - c_{exp,i}^j}{\max c_{exp}^j} \right)^2 + \lambda \sum_{j=1}^p \beta_j^2 \quad 4.8$$

Critical components of NN model are the number of hidden layers, the number of nodes in each hidden layer and the form of the activation function that determines whether this node will be “activated” and how “active” it will be. The influence of NN structure is evaluated in order to obtain the optimal structure that minimizes the objective function value.

Table 4.3: The operating sequence used for the extrapolation test and optimization case study.

Step	Volume (CV)	Salt Conc. (mM)	Flowrate
Load	35 g/L	400	
Step elution	3	$c_{s,start}$	Q
Gradient elution	Gradient length	$c_{s,start}$ to $c_{s,end}$	

The mechanistic model is developed based on the underlined mechanism, so the model can be highly generalizable and physically interpretable. Although the data-driven component of the model based on NN can describe the process without prior knowledge, it mainly suffers from poor interpretability and generalizability. To test the extrapolation capability of the developed hybrid model, an *in-silico* dataset is generated using the mechanistic model outside the calibration datasets. The operating sequence

for the extrapolation test, as shown in Table 4.3, includes the step and gradient elution. It is worth noting that the wash step is not included in the test as it would only result in a time delay of the profile and will not impact the actual elution behavior from the modeling perspective.

4.2.5 Process Optimization

Model-based optimization is used to obtain the optimal operating conditions to reduce experimental efforts and improve process development [56, 235]. The sequence used for the optimization case study is chosen the same as the extrapolation test because the hybrid model has been validated to be suitable to describe this sequence. Specifically, the operating conditions, including the flow rate Q , gradient length, and starting and ending salt concentration for elution ($c_{s,start}$ and $c_{s,end}$), are optimized using the developed mechanistic and hybrid models. The primary objective considered in this work is to maximize the process yield under the purity requirement as it is an important metric to evaluate the column performance. Besides, the processing time is also an essential factor in measuring the separation efficiency in practical applications. Thus, the minimization of the processing time is considered as another objective. The two objectives can be combined into a single objective by adding weights (ω_1 and ω_2) as shown in Equation 4.9.

The overall optimization formulation is shown in Equations 4.9-4.15, where Q is flow rate, $t_{process}$ is the processing time, V_{Eluate} is the eluate volume, UV_1 and UV_2 are the optical density when the pooling starts and ends. In this study, 100 and 0.001 are selected for the weights of the objectives (ω_1 and ω_2), respectively. In biopharmaceutical manufacturing, at least 99.9% purity of monomer is required [90]. Due to the stringent product quality, a safety factor of 1.33 is added to the purity

requirement. The pooling time intervals are determined based on the peak cutting criteria, i.e., the optical density is at least 0.5 OD.

$$\min -\omega_1 \times yield + \omega_2 \times t_{process} \quad 4.9$$

$$Purity_{monomer} \geq 99.925\% \quad 4.10$$

$$c_s \leq 425 \text{ mM} \quad 4.11$$

$$100 \leq Q \leq 500 \text{ cm/h} \quad 4.12$$

$$t_{process} \leq 6 \text{ h} \quad 4.13$$

$$V_{Eluate} \leq 8 \text{ CV} \quad 4.14$$

$$UV_1, UV_2 \geq 0.5 \text{ OD} \quad 4.15$$

The model includes six decision variables to be optimized, i.e., Q , *gradient length*, $c_{s,start}$, $c_{s,end}$, UV_1 and UV_2 . Based on the mechanistic and hybrid model developed in Section 4.2 and 4.3, the optimization problem was solved using L-SHADE, and the obtained optimal parameters and objectives were compared.

4.3 Results and Discussion

4.3.1 Parameter Estimation using Mechanistic Model

In this work, a newly developed adsorption isotherm together with the KDM is used for the description of the HIC process. The resin and column packing parameters in KDM can be directly or indirectly measured based on the methodologies in Section 4.2.4.1. The total and extra-particle porosity, and the height equivalent to a theoretical plate (HETP) are given in Table 4.4.

Table 4.4: Resin and column packing parameters of the KDM in HIC process.

Symbol	Transport Parameter	Value	Units
ε_t	Total porosity	0.956	-
ε_e	Extra-particle porosity	0.415	-
H	HETP	4.5E-2	cm

The adsorption parameters in the salt-dependent HIC isotherm are obtained by inverse fitting the various experimental gradient elution curves via the L-SHADE optimization algorithm. The simulated and experimental chromatograms under varying operating conditions for model calibration are compared in Figure 4.2. It can be clearly observed that the proposed mechanistic model can capture the general trend of the elution behaviors accurately. Due to a limited process understanding of the HIC adsorption mechanism, there exist some deviations between the fitted and experimental data especially for Exp1 and Exp3.

To test the quality of the calibrated mechanistic model, an additional experiment was conducted under varying operating conditions as the validation dataset, as can be seen from Table 4.2. It is noteworthy that although the setpoints for loading and gradient length were within the range of the calibrated dataset, the flow rate used for validation (150 cm/h) is outside the calibrated condition (300 cm/h), which can help assess both the interpolation and extrapolation ability of the developed model. Figure 4.3a shows the predicted elution curve and experimental validation data. A good agreement between the simulated and measured data can be observed, especially for the monomer peak. The comparison of measured and predicted concentration data for monomer and dimer are shown in Figure 4.3b and 4.3c, respectively. For the monomer, the distribution is close to the reference line ($y = x$) with a high correlation coefficient of 0.9878 and R^2 value

of 0.9902, indicating the high accuracy of the model to characterize the monomer behavior. In terms of the correlation result for dimer validation data, the deviations are larger than that of monomer data but the error is acceptable ($R^2 = 0.9173$). It is noted that although significant biases can be found at the low concentration range, the absolute differences are still negligible as the concentration is close to 0.

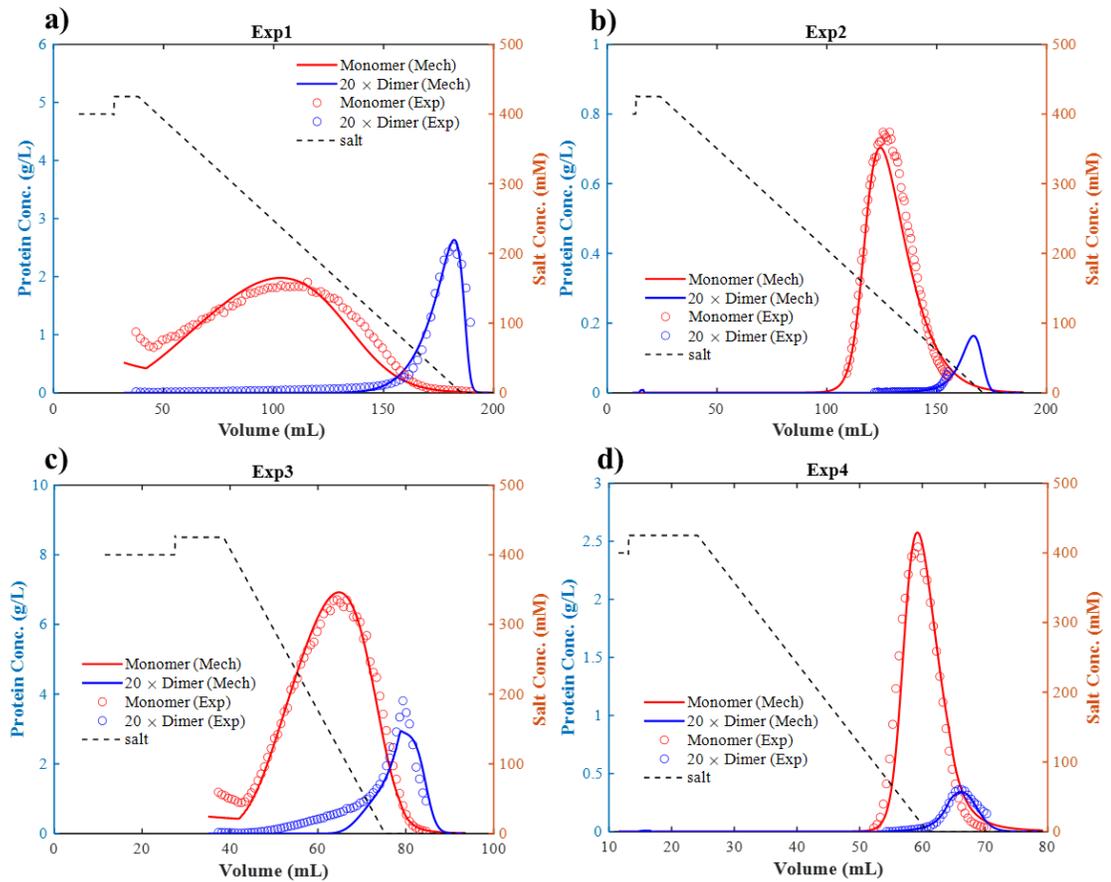


Figure 4.2: Comparison of the calculated bind-and-elute curves with the experimental dataset under different operating conditions during parameter estimation. All the experiments were conducted using a flow rate of 300 cm/h. Mech: mechanistic model; Exp: Experiment.

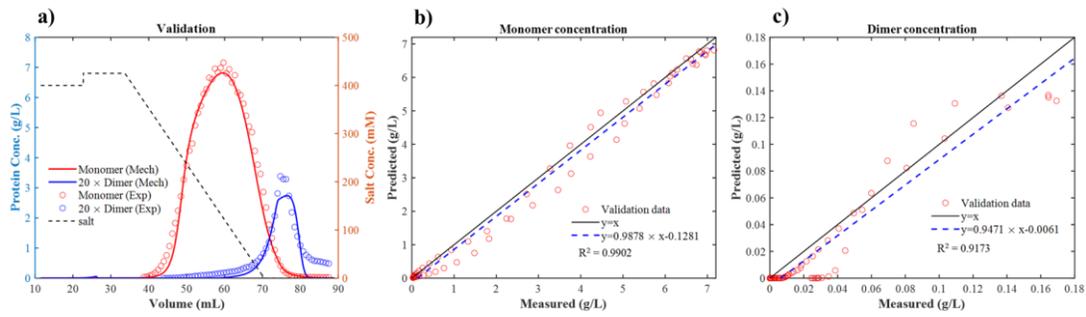


Figure 4.3: a) Comparison of the predicted bind-and-elute curves with the experimental dataset for model validation; Correlation between measured and predicted concentration for monomer (b) and dimer (c). Mech: mechanistic model; Exp: Experiment.

4.3.2 Hybrid Model Development and Validation

4.3.2.1 Parameter Estimation

Two different hybrid models (Hybrid_MCL and Hybrid_Desorption) are constructed to describe the HIC isotherm, followed by model calibration to test the applicability of the proposed hybrid model. Before parameter estimation is performed, the neural network structure should be determined in order to write the mathematical equations in closed form. The NN structure used in the hybrid model is illustrated in Figure 4.4, in which one hidden layer with 2 nodes is selected together with the hyperbolic tangent (tanh) as the activation function. For each component (monomer or dimer), a total of 14 parameters are required to be estimated, including the 3 parameters on the MCL equation and the 11 parameters at the NN model as illustrated in Figure 4.4.

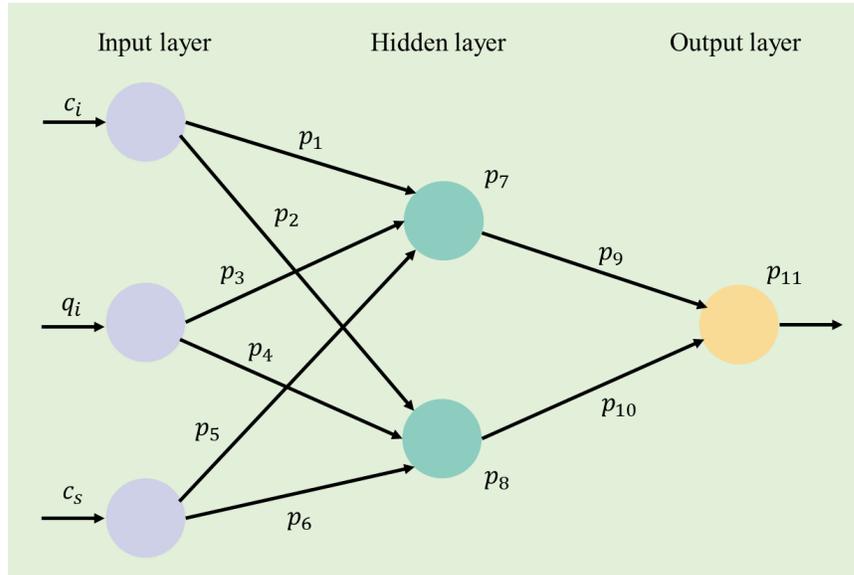


Figure 4.4: Illustration of neural network structure and parameters in hybrid model.

For the first hybrid model structure, Hybrid_MCL, the comparison of the simulated and experimental elution curves is displayed in Figure 4.5. It can be clearly discerned that this hybrid model lacks the capability to describe the HIC process. No benefit was observed by increasing the complexity of neural networks and running time, and testing different optimization algorithms, as can be seen in Table 4.5.

Table 4.5: Calibration NSSE using the Hybrid_MCL model for parameter estimation (1 hidden layer and tanh as the activation function).

# nodes	Operator	Optimization algorithms	Running time (days)			
			7	14	21	
2	absolute	L-SHADE	21.17	19.31	19.03	
3			29.60	27.42	27.38	
2	exponential		27.03	23.42	n/a	
3			37.45	20.84	n/a	
3	absolute		GA	29.49	n/a	n/a
	exponential			45.2	n/a	n/a

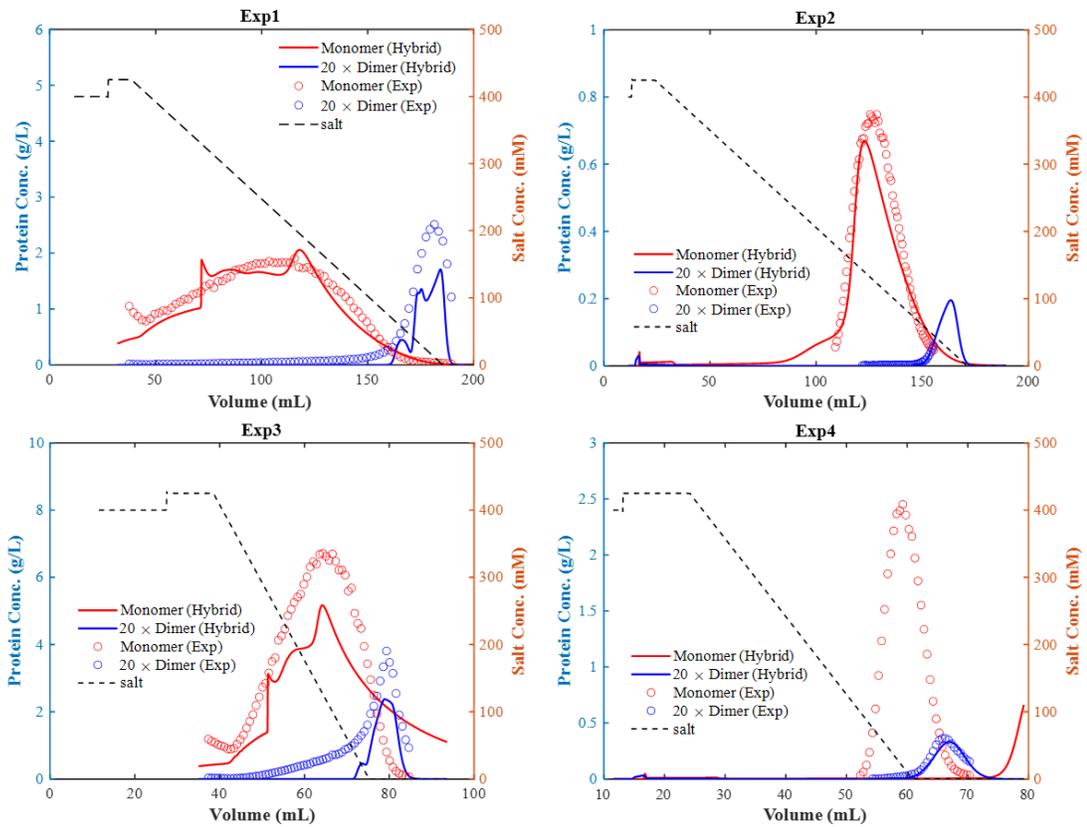


Figure 4.5: Comparison of simulated elution curves and experimental results using Hybrid_MCL model.

The results of the second hybrid model, Hybrid_Desorption, are shown in Figure 4.6. In general, this hybrid model can fit the experimental data quite well except for Exp1 and Exp3. Subsequently, the obtained model is used to predict the elution curves of the validation set, as can be seen in Figure 4.7. The results indicate that this hybrid model structure can reliably represent and simulate the HIC process studied in this work. As mentioned in Section 4.2.4.2, the global optimization algorithm L-SHADE used for parameter estimation is stochastic, which could result in different optimal values at each run. Multiple runs of estimating parameters were conducted to test the robustness of the

developed hybrid model. Table 4.6 lists the NSSE for calibration and validation data after 4 different runs. Although the obtained objective function values are slightly different after each run, the difference is in an acceptable range, indicating the robustness of our developed model. Based on Table 4.7, it can be observed that the average error for the calibration and validation data using the mechanistic model is pretty close. However, in terms of the fitting results based on the hybrid model, the average calibration NSSE for all runs is significantly lower than the validation error, suggesting the occurrence of overfitting.

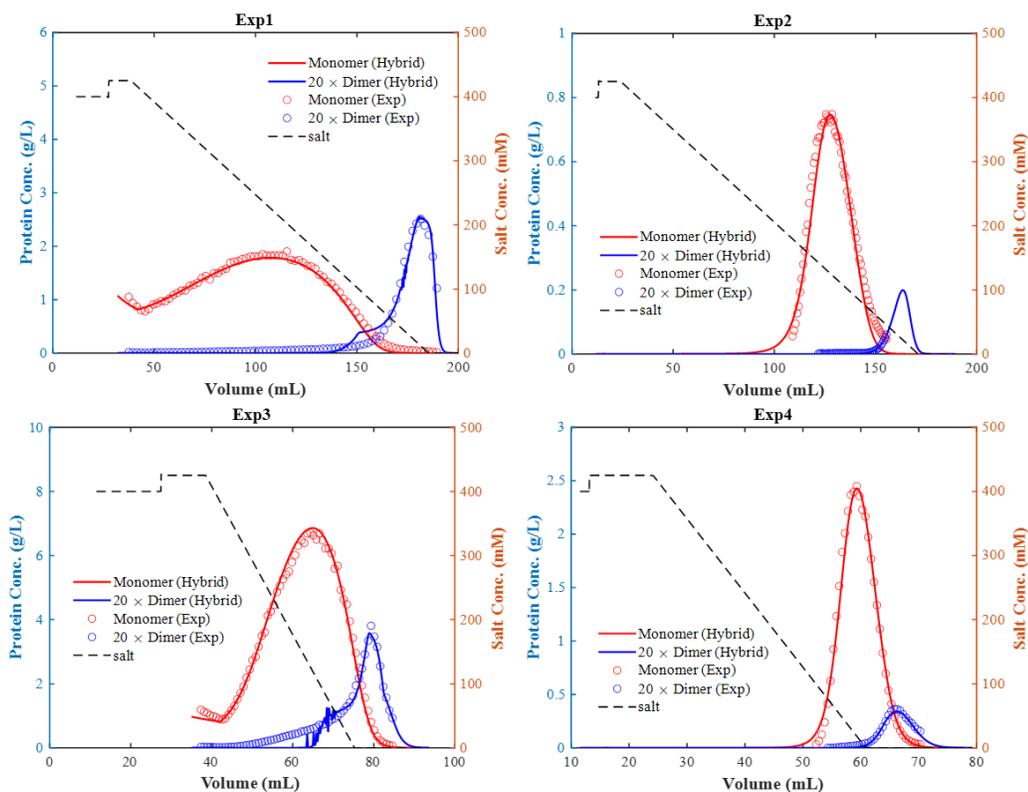


Figure 4.6: Comparison of simulated and experimental elution curves using Hybrid_Desorption model.

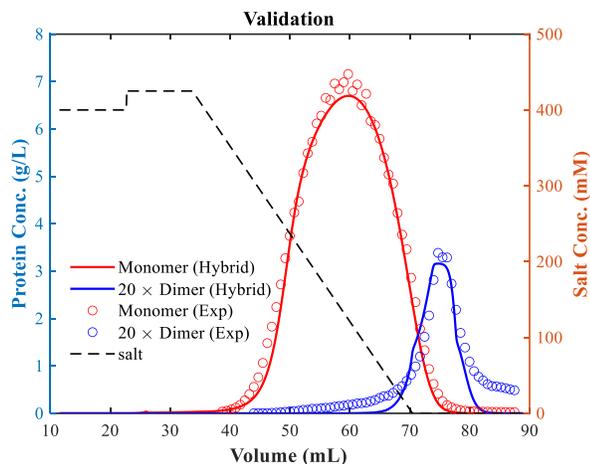


Figure 4.7: Comparison of predicted and experimental elution curves using Hybrid_Desorption model.

Table 4.6: Comparison of NSSE after each run.

Model type		NSSE_calibration	NSSE_validation
Hybrid model	1 st run	0.9344	0.6731
	2 nd run	1.6988	0.9981
	3 rd run	1.2864	0.6977
	4 th run	1.5821	0.6756
Mechanistic model	-	2.92	0.6252

Table 4.7: Comparison of average NSSE after each run.

Model type		Average NSSE	
		Calibration	Validation
Hybrid model	1 st run	0.2336	0.6731
	2 nd run	0.4247	0.9981
	3 rd run	0.3216	0.6977
	4 th run	0.3955	0.6756
Mechanistic model	-	0.73	0.6252

To avoid overfitting, L2 regularization term is added to the loss function with the regularization rate λ . The obtained error values under three different regularization rates ($1E - 4$, $2.3E - 4$, $5E - 4$) are compared in Figure 4.8. In general, the incorporation of regularization strategy should result in an increasing calibration error, but choosing the appropriate rate ($2.3E - 4$) could achieve a balance between the calibration and validation error to improve the performance of the hybrid model. It can be clearly noticed that a large regularization rate ($5E - 4$) results in the highest calibration error and lowest validation error probably because the NN structure is over-penalized whereas a small regularization rate ($1E - 4$) leads to a large validation error due to the small penalty on the NN weights. The simulated elution curves at a regularization rate of $2.3E - 4$ are shown in Figure 4.9, where no overfitting is observed.

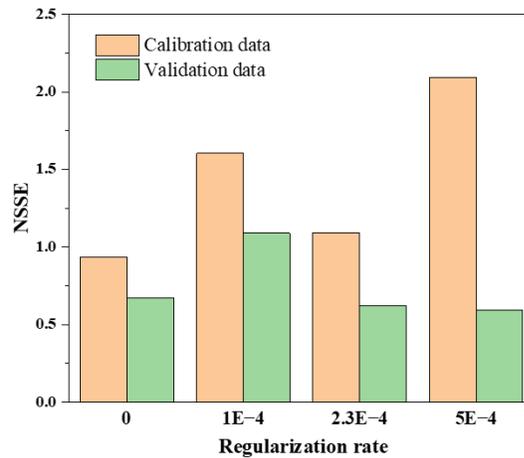


Figure 4.8: The NNSE for the calibration and validation dataset under different regularization rates.

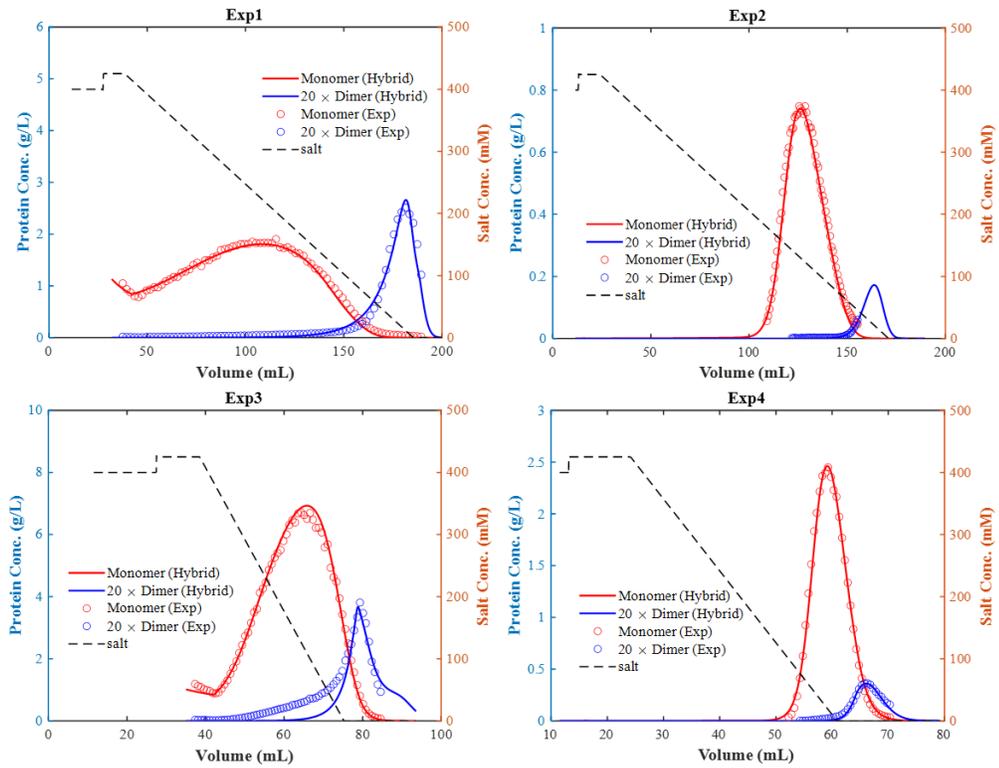


Figure 4.9: Comparison of simulated and experimental elution curves after the incorporation of regularization strategy ($\lambda = 2.3E - 4$).

Since the NN structure can affect the hybrid model expression, the effect of different activation functions and number of nodes on the model performance are investigated. According to previous analysis, since overfitting is found using a neural network with two nodes in one hidden layer to construct the hybrid model, no more than 2 nodes are adopted when examining the impact of NN structure. Additionally, as mentioned in Section 4.2.3, different operators can be added to NN output to maintain the physical meaning of the adsorption process. Table 4.8 summarizes the calibration and validation errors under different NN structures and operators. It should be noted that no regularization strategy is included in this step because the goal is to find the

optimal NN structure. When the simulations were carried out using different operators with tanh as the activation function, it was found that the exponential operator does not improve the fitting performance but increases the computational time. The NN with only 1 node is not sufficient to describe the behavior of the model as observed from the large calibration error (2.4086). Only the absolute operator and 2 nodes are thus considered when sigmoid is used as the activation function. The validation error obtained by using sigmoid as the activation function is the lowest among all the structures investigated.

Table 4.8: Comparison of calibration and validation error under different NN structures without the regularization strategy incorporated. All tested NN has one hidden layer.

Activation function	Operator	# nodes	NSSE_calibration	NSSE_validation
tanh	exponential	2	1.2856	0.6762
	absolute	1	2.4086	0.6849
	absolute	2	0.9344	0.6731
sigmoid	absolute	2	1.7462	0.4572

Based on these results, we hypothesized that the sigmoid activation function could contribute to better testing and validation performance after the application of the regularization strategy. The hybrid model performance under different regularization rates using sigmoid as an activation function and 1 hidden layer with 2 nodes is summarized in Table 4.9. It is observed that a regularization rate of $2.3E - 4$ yields a good performance on both the calibration and validation dataset. Under such circumstances, the calibrated and validated NSSE of the developed hybrid model are

1.1191 and 0.4286, while the errors of the mechanistic-based model are 2.92 and 0.6252, respectively. This hybrid model structure with $2.3E - 4$ as the regularization rate has the best overall performance in terms of the calibration and validation dataset, so the following analysis is obtained based on this model. The NSSE values for each experiment using the two models are provided in Table 4.10 and visually compared in Figure 4.10. It can be seen that the fitting performance for each experiment is enhanced with the use of hybrid model, ranging from 8.3% to 83.6%. Compared with the mechanistic model, the overall calibration performance is significantly improved by 62% and the validation accuracy is increased by 31.4%.

Table 4.9: Calibrated and validation error under different regularization rates using the hybrid model structure (sigmoid as activation function, 1 hidden layer with 2 nodes).

Model	λ	NSSE_calibration	NSSE_validation
Hybrid_Desorption	0	1.7462	0.4572
	1e-4	0.827	0.8088
	2.3e-4	1.1191	0.4286
	5e-4	1.1717	0.8666

Table 4.10: Comparison of SSE between mechanistic and best hybrid model.

Dataset	Experiment	Mechanistic model	Hybrid model	Improved accuracy
Calibration	Exp1	1.10	0.25	77.3%
	Exp2	0.73	0.12	83.6%
	Exp3	0.72	0.66	8.3%
	Exp4	0.37	0.08	78.4%
		2.92	1.11	62.0%
Validation	Exp_Val	0.63	0.43	31.7%

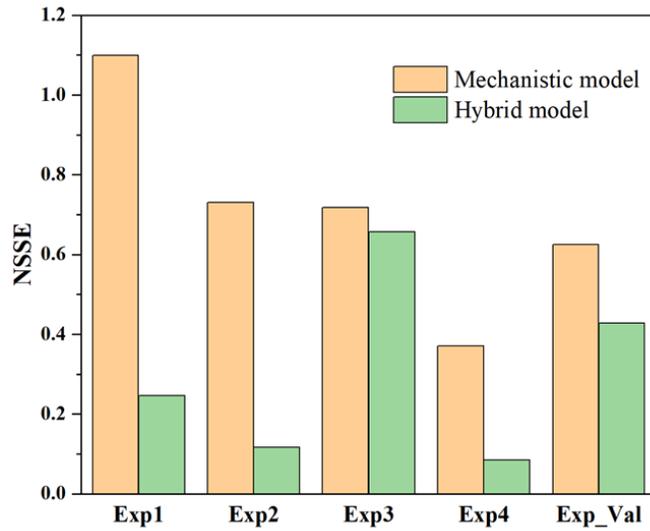


Figure 4.10: Comparison of NSSE for each experiment using mechanistic and hybrid models.

Table 4.11: Comparison of calibrated NSSE using different optimization algorithms.

Algorithm	NSSE_calibration
L-SHADE	1.119
GA	27.22
PSO + <i>fmincon</i>	2.237
MultiStart <i>fmincon</i>	75.04

Different optimization algorithms have been used for parameter estimation using this hybrid model. The calibrated error using various algorithms is given in Table 4.11 showing that L-SHADE performs best in this study. Based on the optimal identified parameters of the Hybrid_Desorption model, the comparison of simulated and experimental curves for model calibration is shown in Figure 4.11, and an excellent agreement is recognized between the model simulation and calibrated experimental data. Figure 4.12 illustrates the predicted concentration values for monomer and dimer

that lie close to the diagonal with a high R^2 value, an indication of the high quality and predictive accuracy of the developed hybrid model to describe the gradient elution in HIC process. It is worth noting that with such a small neural network size (1 hidden layer with 2 nodes), the hybrid model outperforms the mechanistic model, and can precisely describe the complicated salt-dependent interaction between the protein and ligand.

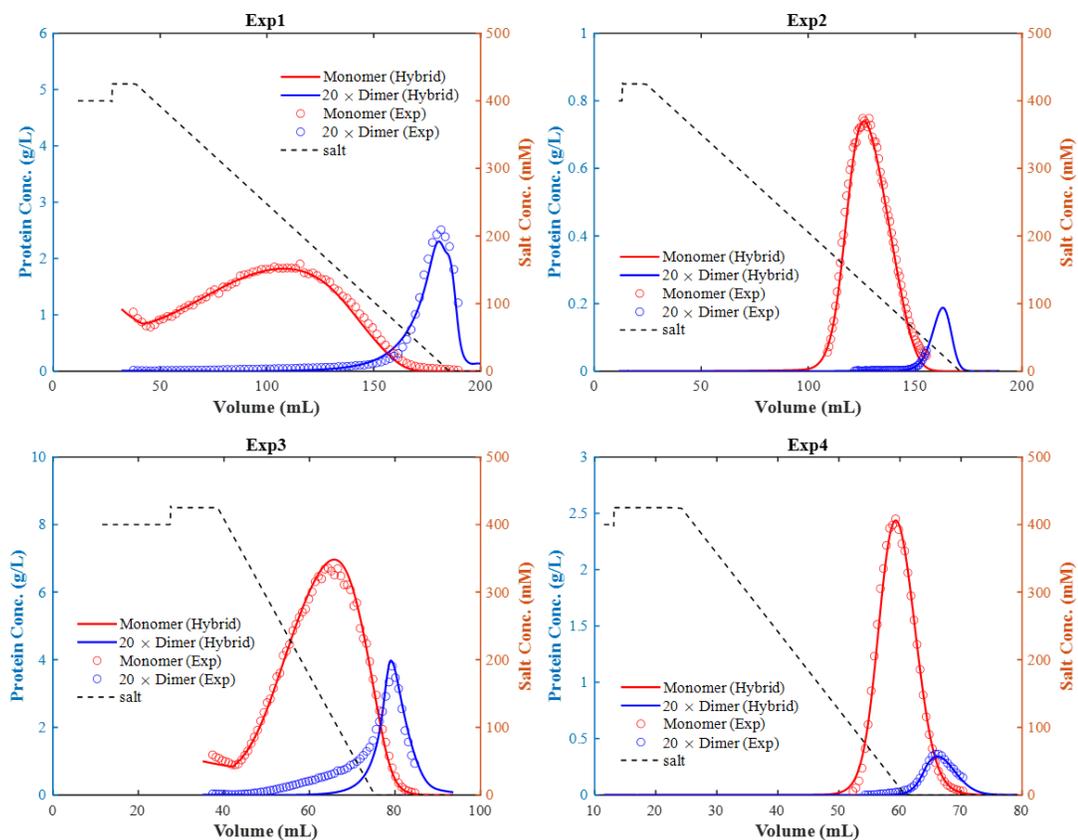


Figure 4.11: Comparison of the simulated bind-and-elute curves with the experimental dataset under different operating conditions during parameter estimation. All the experiments were conducted using a flow rate of 300 cm/h. Hybrid: Hybrid_Desorption model; Exp: Experiment.

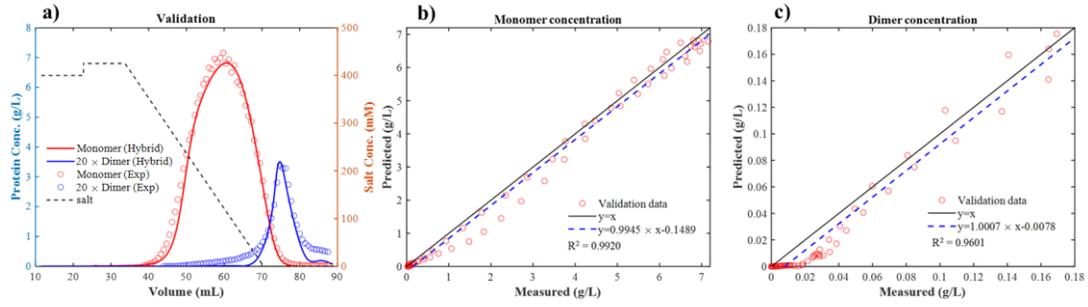


Figure 4.12: a) Comparison of the predicted bind-and-elute curves with the experimental dataset for model validation; Correlation between measured and predicted concentration for monomer (b) and dimer (c). Hybrid: Hybrid_Desorption model; Exp: Experiment.

4.3.2.2 Extrapolation Capability

Although the developed hybrid model has been validated using experimental datasets to test the extrapolation ability of the model, it is essential to see how the model performs in process conditions outside the range of conditions used for model building. Utilizing the mechanistic model, *in-silico* datasets can be generated to test the extrapolation capability of the hybrid model.

Table 4.12: Summarized operating conditions for extrapolation test under same loading conditions (35 g/L loading capacity and 0.2% aggregate content). Exp: experimental conditions for model calibration and validation.

	Flow rate (cm/h)	Gradient length (CV)	$c_{s,start}$ (mM)	$c_{s,end}$ (mM)
Exp	300/150	10/40	425	0
Test 1	100	5	300	30
Test 2	200	30	400	50
Test 3	250	15	200	10
Test 4	500	20	350	40

The extrapolation test is based on a new operation sequence in Table 4.12, along with four varying operating parameters (including flow rate, gradient length, and the starting and ending salt concentration). Four different test operating conditions for generating *in-silico* datasets are chosen and enumerated in Table 4.12. The predicted elution curves obtained from mechanistic and hybrid model are compared in Figure 4.13. Although the operating sequence (with step elution introduced) and conditions are different, no significant difference in the simulated results using the two different types of models is observed, indicating the good extrapolation capability and robustness of the established hybrid model. These simulations could help us investigate how the combination of step and gradient elution strategies and salt concentration would impact the overall shape of the elution curves, which indicates the importance of modeling approaches to help investigate such phenomena and increase process understanding with minimum experimental effort. In Figure 4.13a, the prediction for the dimer behavior under the first test condition seems greatly different, but it should be noted that this prediction is magnified by 20 times. The comparison of unmodified dimer concentration is displayed in Figure 4.14, and the difference is found to be negligible. Furthermore, monomers are the target and it is thus much more significant to achieve accurate predictions of the behavior of monomers than that of dimers.

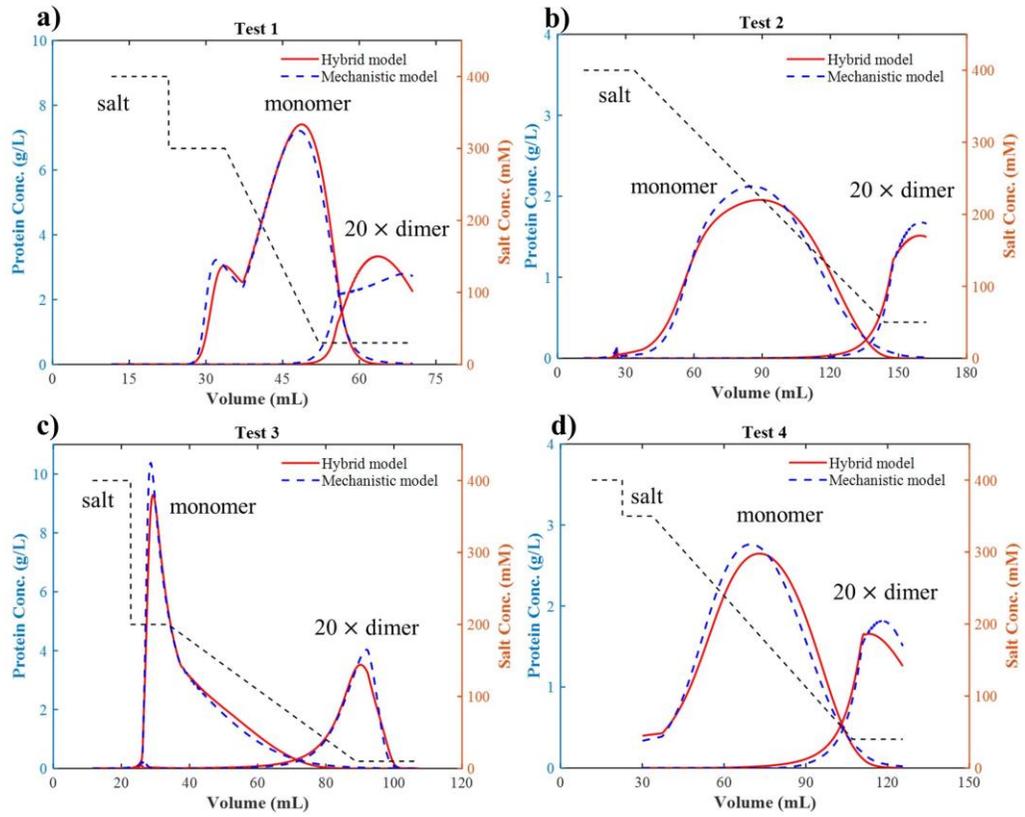


Figure 4.13: The comparison of predicted elution curves using hybrid and mechanistic models under four extrapolation tests.

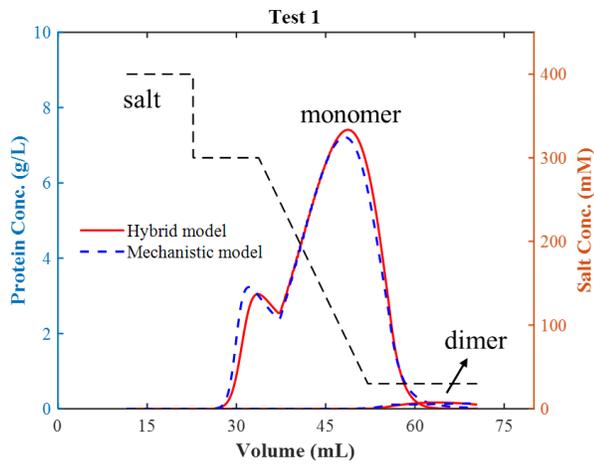


Figure 4.14: The comparison of extrapolation test 1 with unmodified dimer concentration.

4.3.3 Process Optimization

In this work, under product quality and operating constraints, yield is maximized while the processing time is minimized by varying the decision variables specified in Section 4.2.5. Since mechanistic and hybrid models have been developed to describe the HIC process, the two different models are applied in the optimization case study to identify the best operating conditions. The results after process optimization are listed in Table 4.13. The maximum yield acquired by the hybrid model is 99.45%, slightly higher than that (98.05%) from the mechanistic model, but the processing time of the hybrid model is a little bit longer. Because both models can reasonably describe the HIC process, there is no significant difference in the objective function values for the optimization case study. The optimal flow rates and peak cutting criteria are similar, but the obtained parameters during the elution steps are different. During the step elution phase with fixed elution length, with lower salt concentration, the shape of the band curve becomes sharper, resulting in more products eluted out, as shown in Figure 4.15. The following elution profile significantly changes when the gradient elution is introduced after the step elution. Moreover, it can be obviously found that the operating conditions for the gradient elution play an important role in affecting the elution behaviors. In this work, since the accuracy of the hybrid model is much higher than the mechanistic model, the optimization results from hybrid model are more reliable.

Table 4.13: Optimized operating conditions and objectives using mechanistic and hybrid model.

	Optimized operating conditions					
	Q (cm/hr)	Gradient length (CV)	$c_{s,start}$ (mM)	$c_{s,end}$ (mM)	UV_1 (OD)	UV_2 (OD)
Mechanistic	110.9	3.8	228.5	86.3	0.54	0.5
Hybrid	111.2	5.9	185.7	64.5	0.51	0.5

Table 4.13 continued.

Column performance metrics			
	Yield (%)	Purity (%)	$t_{process}$ (hr)
Mechanistic	98.05	99.925	3.41
Hybrid	99.45	99.925	3.76

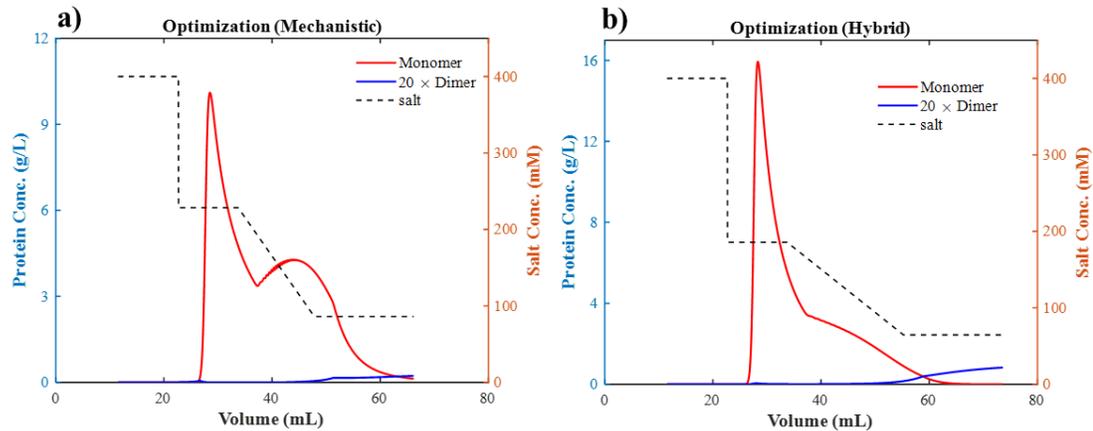


Figure 4.15: The predicted elution curves under optimized operating conditions obtained from mechanistic (a) and hybrid models (b).

4.4 Conclusions

This chapter focused on investigating the separation of monomers and dimers in hydrophobic interaction chromatography with an integrated experimental and modeling approach. At first, a modified isotherm derived from Wang et al. is built to reliably describe the HIC adsorption process, with an equilibrium dispersive model used to characterize the mass transport. However, since adsorption in HIC is highly related to the salt concentration and the mechanism is still unclear, the development of the isotherm requires a high-level process understanding and knowledge, which significantly increases the effort of model development.

Thus, a hybrid model approach is employed to reduce such effort and extract the missing relationships that cannot be captured by the mechanistic model. We proposed that the hybrid model can be constructed by combining a neural network with a simple but well-known isotherm (multi-component Langmuir). There are various ways to include a NN in the MCL equation, and it is found that how the two elements are integrated is an important component of the hybrid model performance. Once the appropriate hybrid model structure is identified (Hybrid_Desorption), the effects of different NN structures (number of hidden layers, number of nodes in each hidden layer, and activation function) and regularization rates are thoroughly examined to determine the hybrid model with the best performance. With a very small neural network structure (1 hidden layer with 2 nodes and sigmoid as the activation function), the accuracy of the obtained hybrid model is much higher than that of the mechanistic model, with an overall improvement of 62% and 31.4% in the calibration and validation datasets, respectively. Moreover, the developed hybrid model exhibits excellent extrapolation capability by comparing the *in-silico* simulations generated from the hybrid and mechanistic models. The methodology proposed in this work to construct hybrid model from a simplified known isotherm can help reduce the effort to identify the underlying mechanism and provide significant guidance to investigate the sophisticated process with limited insights.

Process optimization is performed based on the provided case study to maximize the yield and minimize the processing time, subject to quality and operating constraints. The optimal yield obtained by the hybrid model (99.45%) is higher than that (98.05%) from the mechanistic model, although the processing time of the hybrid model is longer. Although the developed models have been calibrated and validated by the experimental

data, a limitation of the optimization results is the lack of equipment validation. For future work, experiments could be performed to validate the optimization results. Additionally, the proposed methodology to develop hybrid models from a simplified isotherm can be applied to other types of chromatography.

Chapter 5

FLWSHEET MODELING FOR INTEGRATED CONTINUOUS BIOPHARMACEUTICAL PROCESS

5.1 Introduction

The commercial-scale production of biopharmaceuticals is currently performed in batches in which each unit operation is operated in sequence [164, 171]. However, due to the fast expansion of market demand, continuous production has become a promising alternative to producing mAbs, which has various advantages including higher and constant product quality, increased productivity and yield, smaller footprint, and rapid capacity adjustment [57]. Many other fields, such as food, petrochemical, chemical, mechanical, and small molecular pharmaceutical drug production, have achieved the transition from batch to continuous operation, but the fully integrated continuous biomanufacturing for the production of biotherapeutic products has not been implemented commercially [6, 13].

Nonetheless, some significant progress has been accomplished in the construction of continuous bioprocess in recent years [1, 9, 75]. Perfusion bioreactors with different cell retention devices (such as tangential flow filtration, TFF, and

Chapter 5 is adapted from the following journal article written by the author of this dissertation:

Ding, C., Ardeshta, H., Gillespie, C., Ierapetritou, M., Process design of a fully integrated continuous biopharmaceutical process for economic assessment and ecology impacts. *Biotechnology & Bioengineering*, 2022, 119(12):3567-3583.

alternating tangential filtration, ATF) are introduced as enabling technology for the continuous upstream process to maintain higher cell density and increased product titers [7, 236-238]. The economic comparison between fed-batch and perfusion bioreactors has also been comprehensively investigated based on dynamic simulation [8, 239] and mass balance [240]. A representative platform for the downstream purification process includes Protein A chromatography for primary capture, virus inactivation with a low-pH hold, one or more polishing chromatography steps, virus removal by nanofiltration, and final ultrafiltration/diafiltration [30]. Multi-column periodic counter-current chromatography (PCC) has been developed for the primary capture and polishing steps to effectively improve productivity, increase resin capacity utilization, and lower buffer consumption [82, 174, 177, 186]. Different reactors like coiled flow inversion reactor, tubular reactor, packed-bed reactor [241], and plug flow reactor [242] have been proposed to achieve continuous in-line virus inactivation. For the final formulation step for buffer exchange and/or product concentration, single-pass tangential flow filtration (SPTFF) has been designed to provide continuous ultrafiltration and diafiltration [243, 244]. There have been several successful examples of integrating perfusion cell culture with continuous downstream unit operations to achieve integrated continuous biopharmaceuticals reported from academia and industry [245-249]. In addition, single-use technology has also been developed, and it has been demonstrated that incorporating such technology into continuous biomanufacturing has significant economic benefits [250, 251].

The development of emerging technology such as the ones highlighted before is critical to building a truly continuous biomanufacturing line, but the demonstration of its economic viability is also significant to drive its commercialization [252]. Cost

evaluations of establishing continuous platforms to produce various monoclonal antibodies (mAbs) have been conducted by many research groups over the past decade [253-256]. Ou et al. [150] and Gupta et al. [156] compared the economic assessments between batch and continuous mAb production bioprocess, and cost savings of 35% to 68% have been determined to characterize the transition from batch to continuous operation, demonstrating the tremendous economic benefits and efficiencies of the continuous process. Different continuous platforms with varying unit procedures (such as perfusion technologies, capture scenarios, and polishing steps with membrane chromatography) have also been thoroughly evaluated [55, 59, 158, 257]. In addition to cost evaluations, Pollock et al. [258] compared the environmental impacts and operational feasibility of different manufacturing strategies, and it is found that continuous processing can offer ecological and operational robustness benefits. However, there are still many aspects that need to be considered. To our knowledge, few papers take the media and buffer preparation into consideration when building the continuous platforms although it has been shown that buffer preparation is significant in affecting the overall operational and scheduling activities [143, 259, 260]. The investigation of the shift of the bottlenecks is also limited when production increases. Moreover, most papers perform scenario analysis to examine the impact of process variables on production economics [150, 156, 261, 262], but few of them investigate the impacts on the environmental footprints [55, 263].

Thus, technical approaches are proposed in this work to solve the aforementioned challenges. Since our previous work has demonstrated the benefits of continuous bioprocess over batch operation [150], this work will focus on the implementation of the continuous process. For readers who are unfamiliar with the

background of the comparison between batch and continuous processes, we would like to recommend our previous work and those papers [55, 57, 59, 156, 158, 254, 258, 261, 262]. In this paper, mAb is chosen as representative molecules to build a detailed integrated continuous process with the addition of media and buffer preparation. Novel technologies including N-1 perfusion seed bioreactors, SPTFF, and single-use systems are incorporated into the process design. Scheduling is performed to achieve the implementation of real-time media and buffer addition, followed by performing a comprehensive economic analysis and evaluating ecology impacts by calculating environmental indicators (E-factors). Subsequently, scenario analysis is used to evaluate the influence of titer and bioreactor scale on process economics and environments. The shift of process bottlenecks when the upstream production capacity increases is also thoroughly investigated. Membrane chromatography is an emerging technology that can eliminate pore diffusion, improve bioseparation efficiency, and reduce buffer usage [264, 265], so the effect of membrane chromatography on the overall process performance is examined in terms of the economic and ecology impacts.

5.2 Methods

5.2.1 Process Description

The traditional batch platform to produce mAb therapeutics consists of cell culture, primary capture, virus inactivation, polishing, virus removal, and final formulation. The structural similarities between different mAbs make this platform widely accepted and regarded as an industry benchmark. Thus, this platform is adopted in this study to construct the continuous process, and the corresponding process flow diagram (PFD) is shown in Figure 5.1. It should be noted that there are multiple surge

tanks in the end-to-end integrated continuous process diagram, which are located before and after the multi-column continuous chromatography operations to eliminate the cyclic discontinuity. The introduction of surge vessels adds safety, flexibility, and robustness to a fully continuous biomanufacturing line by providing a temporal buffer [266, 267].

Instead of the conventional perfusion expansion stage, N-1 perfusion seed culture is considered in this work because this approach can provide higher target cell density (i.e., greater than 50×10^6 cells/mL) and shorten the culture process time (i.e., days not weeks) [268]. The cell density of N-1 seeding bioreactor is assumed to be 80×10^6 cells/mL, with two days for setting up and five days for perfusion. The volume of the seed bioreactor is estimated to be half that of the production bioreactor. Subsequently, the cells from the N-1 seed bioreactor are transferred to the production bioreactor, which has a setup duration of 1 day to ensure that the perfusion reaches a steady state. The culture is operated continuously for 28 days at a steady state with a titer of 1.6 g/L. Tangential flow filtration (TFF) is used as the cell retention device to separate cells from the liquid phase and recycle them for fermentation. The details of the upstream perfusion process are enumerated in Table 5.1.

In order to adjust the loading flowrates to the chromatography and distribute the product to different columns, a surge tank is placed to connect the upstream and downstream processes, as shown in the PFD. Three-column continuous Protein A affinity chromatography is chosen as the primary capture to remove culture media components, host cell proteins, and DNA. A surge vessel is installed after the Protein A chromatography to periodically collect the eluate and continuously transfer material out to the virus inactivation. A plug flow reactor (PFR) is chosen for continuous in-line low-

pH hold. The Protein A eluate pool is diluted in-line with hypochlorous acid to reach a low pH value before transferring into a plug flow reactor (PFR) with a residence time of 30 min. Following the reactor hold, the virally inactivated mixture will be neutralized to desired pH values by the continuous addition of Tris base before entering the polishing step. Two polishing chromatography steps are employed to further purify the protein by removing the product-related and residual process-related impurities in this work. The first polishing step is performed using twin-column flow-through anion-exchange chromatography (AEX), followed by the second step accomplished by three-column cation-exchange chromatography (CEX) operated in the bind and elute mode. There is a surge tank between AEX and CEX to adjust the flowrate and mix it with the titration buffer before the product goes into the CEX. The eluate from the polishing step is transported to the surge tanks and then to dead-end nanofiltration to remove the virus and further keep the material sterile before moving into the final formulation. Single-pass tangential flow filtration (SPTFF) module is utilized to offer continuous ultrafiltration and diafiltration with countercurrent flow considered. The basic idea of downstream process design is summarized above, and the corresponding detailed information is listed in Table 5.1.

In many papers, the annual throughput is initially defined with respect to the target product, followed by process design to satisfy these demands [150, 156]. However in this study, the process is designed based on the available commercial equipment and the yearly productivity is calculated from the designed process. This strategy is chosen for process design because there is currently no fully automated continuous biopharmaceutical plant, so the focus of this work is on the implementation

of continuous manufacturing process. Following this analysis, a scale-up/down approach can be used to change the annual production rate based on the market demand.

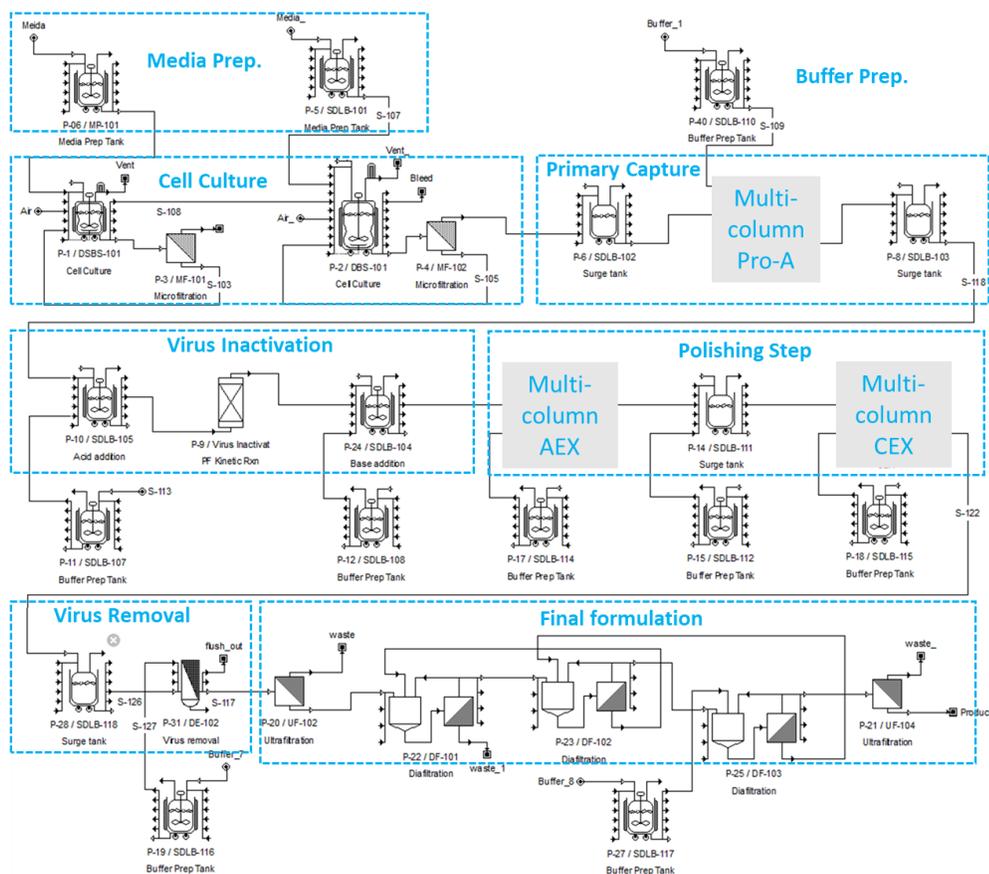


Figure 5.1: Process flow diagram (PFD) for continuous mAb production process.

Table 5.1: Details of the continuous platform for mAb production.

Unit operation	Parameter	Continuous
N-1 seed bioreactor	Total volume (L)	250
	Working volume (L)	200
	Perfusion rate (vvd)	1
	Cell density (Mcells/mL)	25
	Operating time (days)	5

Table 5.1 continued.

Perfusion bioreactor	Total volume (L)	500
	Working volume (L)	400
	Perfusion rate (vvd)	1.5
	Cell density (Mcells/mL)	120
	Operating time (days)	28
	mAb titer (g/L)	1.6
	Bleed flowrate (kg/h)	2.51
Protein A chromatography	No. of columns	3
	Yield (%)	90
	Resin binding capacity (g/L)	60
	Resin lifetime (cycles)	180
	Bed height (cm)	12
	Bed diameter (cm)	12.6
	Aspect ratio	0.952
	Residence time (min)	3.8
Virus inactivation	Residence time (min)	30
	Yield (%)	98
	PFR volume (L)	1.9
AEX	Operating mode	Flow through
	No. of columns	2
	Yield (%)	90
	Resin binding capacity (g/L)	150
	Resin lifetime (cycles)	175
	Bed height (cm)	20
	Bed diameter (cm)	8
	Residence time (min)	16.4
CEX	Operating mode	Bind and elute
	No. of columns	3
	Yield (%)	90
	Resin binding capacity (g/L)	40
	Resin lifetime (cycles)	300
	Bed height (cm)	12
	Bed diameter (cm)	10
	Residence time (min)	22.3

Table 5.1 continued

	No. of filters	4
Virus removal	Operating time of each filter (days)	7
	Yield (%)	95
	Filter size (m ²)	0.5
SPTFF	Flux (L/m ² /H)	54.7
	Yield (%)	95%
	Output flowrate (mL/min)	2.6
Buffer/media preparation	Surge tank volume (L)	50/100

5.2.2 Process Scheduling

Process scheduling is performed to determine the sequence and timing of each process in different units to ensure that the integrated production process runs properly. The main equipment occupancy chart for the designed process is shown in Figure 5.2, and it can be seen that this process is operated continuously for 28 days. The different color bars in the equipment occupancy chart represent different batches. The cycle time of the process can be determined by the length of one color bar. Only one column is displayed in each multi-column step because the occupancy chart of other columns is approximately the same on the one-year horizon. Scheduling for the media and buffer preparation steps is carried out to realize real-time addition. It is impractical to enumerate all the scheduling details in the implementation of media and buffer addition, so we chose the buffer addition in the primary capture step as an illustrating example to explain the idea behind it.

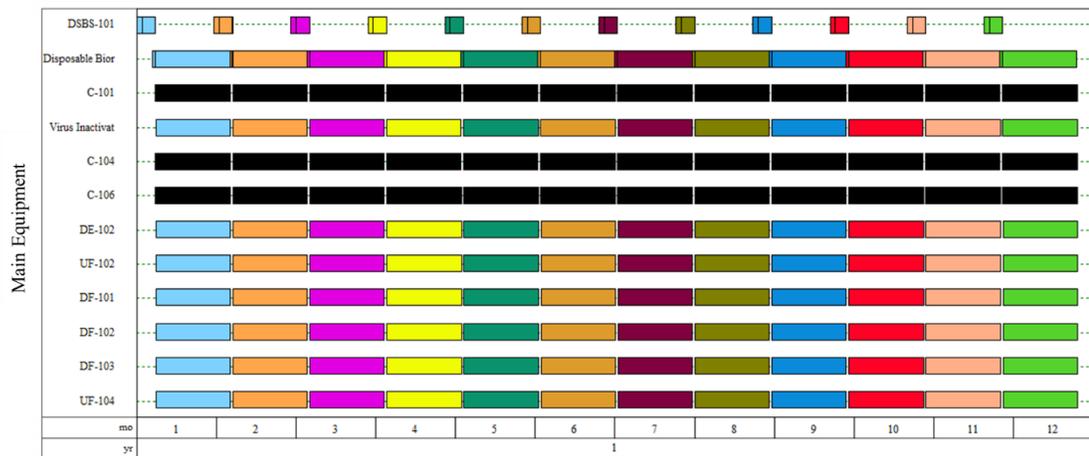


Figure 5.2: Main equipment occupancy chart for the continuous mAb production process. Legend: DSBS-101, seed bioreactor; Disposable Bior, production bioreactor; C-101, one protein A column in primary capture; Virus inactivation, plug flow reactor; C-104, one AEX column; C-106, one CEX column; DE-102–UF104: SPTFF.

The primary capture operation of Protein A chromatography includes the following procedures [150]: two pre-washing steps, loading, wash, elution, regeneration, and equilibrium. Each process buffer is filtered through a bioburden reduction filter to ensure sterility as shown in Figure 5.3. In primary capture, three buffer preparation lines are built for chromatographic operations. To simplify the occupancy chart, only three buffer hold tanks corresponding to three different lines, and three chromatography columns are shown in Figure 5.4. The three columns (C-102, C-101, and C-103) work sequentially according to the defined steps, with buffers drawn from the holding tanks. SDLB-114 provides buffer for two washing and equilibrate steps, SDLB-109 for elution and SDLB-111 for regeneration. It can be found that the addition of buffer follows the operation of the master equipment (i.e., chromatography) and the

tank is on hold if no operation is performed on the buffer. The cycle time of each holding tank is long enough to make sure each column finishes all the steps.

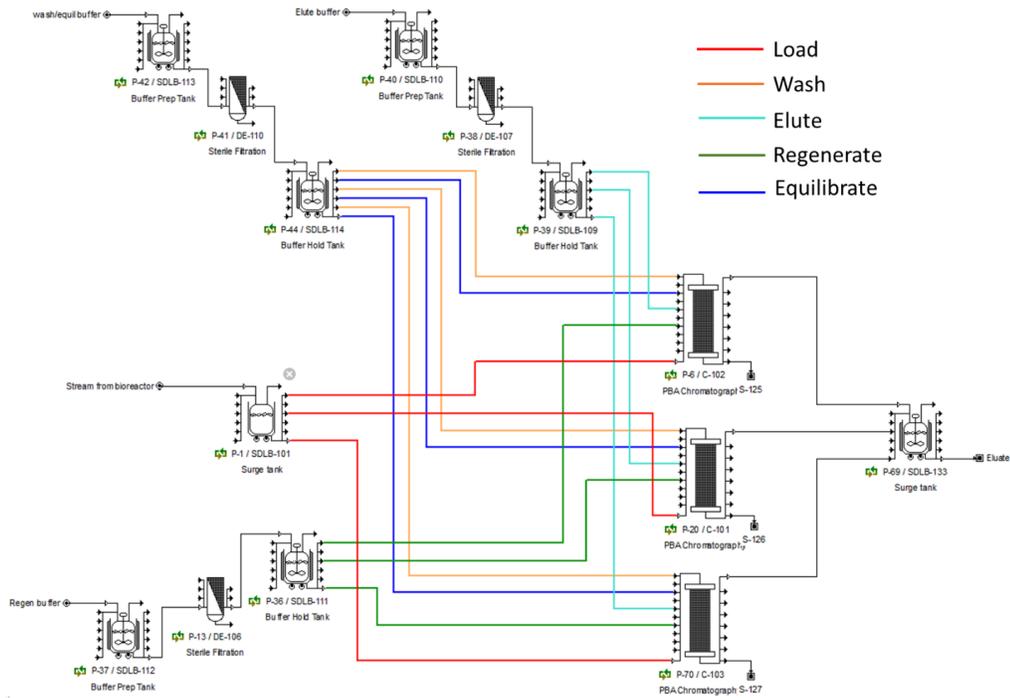


Figure 5.3: The schematic representation of buffer preparation step in primary capture. Red lines: loading step from the surge tank SDLB-101 between upstream and Protein A chromatography; orange lines: washing steps from washing buffer hold tank SDLB-114; turquoise lines: elution steps from elution buffer hold tank SDLB-109; green lines: regeneration steps from regeneration buffer hold tank SDLB-111; blue lines: equilibration steps from the washing buffer hold tank SDLB-114 (because the equilibration buffer and washing buffer are the same). DE-110, DE-107, DE-106: sterile filtration; SDLB-113: washing/equilibration buffer preparation tank; SDLB-110: elution buffer preparation tank; SDLB-112: regeneration buffer preparation tank; SDLB-133: surge tank after Protein A chromatography to cyclically collect the eluate.

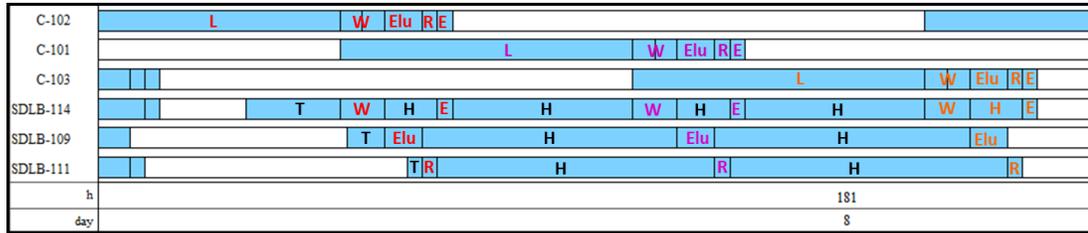


Figure 5.4: Scheduling of buffer preparation in primary capture. W: first and second washing steps; L: loading; Elu: elute; R: regeneration; E: equilibrate; H: hold; T: buffer transfer into the buffer hold tank. Red: operation in the first column (C-102); purple: operation in the second column (C-101); orange: operation in the third column (C-103). C: chromatography columns; SDLB-114: washing/equilibration buffer hold tanks; SDLB-109: elution buffer hold tanks; SDLB-111: regeneration buffer hold tanks.

5.2.3 Process Economics

In the cost evaluation, the production expenditure is comprised of capital expenditure (CapEX) and operating expenditure (OpEX) as can be seen in Figure 5.5. CapEX includes direct fixed capital (DFC), working capital, and start-up and validation cost. The DFC, referring to the fixed assets of an investment, can be calculated by summing up direct cost (DC), indirect cost (IC), and other cost (OC). The DC contains cost elements directly related to an investment, such as the equipment purchase cost, installation cost of equipment, process piping cost, instrumentation cost (like the transmitters, controllers, panels, control computers, etc.), insulation cost (insulation and painting), electrical cost, building cost (including process towers, control room, stairways, etc.), yard improvement cost (such as roads, fences, parking spaces, etc.), and auxiliary facilities cost (e.g., steam plant). Each cost component is proportional to the equipment purchase cost, and the proportional coefficients are listed in Table 5.2. Equipment costs are estimated using exponential method correlations as shown in Equation 5.1 [261]. The reference cost and size for all the equipment are obtained from

the literature or vendor resources and employed after consulting with our industrial partners in Table 5.3. It should be noted that the reference size of filters is based on the filter area instead of the membrane volumetric flux rate due to the limitation of SuperPro Designer.

$$cost = reference\ cost \times \left(\frac{size}{reference\ size} \right)^n \quad 5.1$$

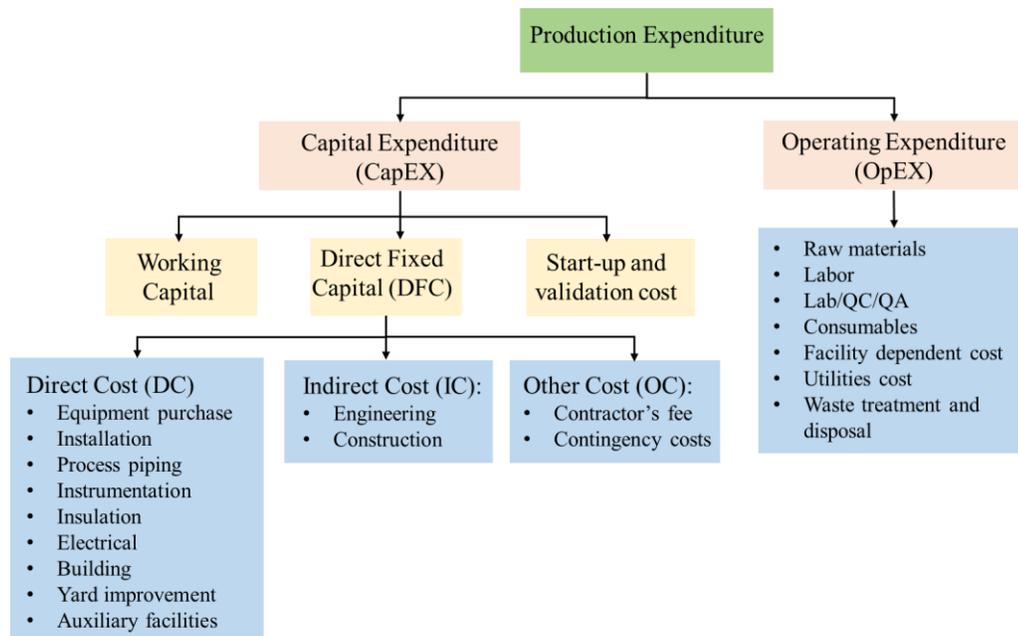


Figure 5.5: Production expenditure breakdown with capital expenditure (CapEX) and operating expenditure (OpEX).

The IC represents the costs indirectly associated with the investment, like the engineering and construction costs, which can be calculated by multiplying DC with specific factors. The last component of DFC is the miscellaneous or other costs, such as contractor's fees and contingencies, which are in proportion to the sum of DC and IC.

The second component of CapEX, working capital, which characterizes the assets required to operate the facility and is difficult to quantify directly, can be approximated by assuming one month of labor, raw material, utilities, and waste treatment cost. The last component includes the start-up and validation cost, which is estimated to be 5-10% of DFC.

The economic efficiency of different operation processes is compared by the cost of goods per gram (COG/g), which is defined in Equation 5.2 [150].

$$\text{COG/g} = \frac{\frac{\text{CapEX}}{\text{years of operation}} (\$/\text{yr}) + \text{annual OpEX} (\$/\text{yr})}{\text{annual production rate (g/yr)}} \quad 5.2$$

Table 5.2: Capital cost estimation in the economic analysis.

	Cost items	Reference Cost	Multiplier to Reference Cost
Direct Cost (DC)	Equipment Purchase Cost (PC)	PC	1
	Installation cost of equipment		Varies in different equipment
	Process piping cost		0.35
	Instrumentation cost		0.35
	Insulation cost		0.03
	Electrical cost		0.15
	Building cost		0.45
	Yard improvement cost		0.15
	Auxiliary facilities cost	0.50	
Indirect Cost (IC)	Engineering	DC	0.25
	Construction		0.35
Other Cost (OC)	Contractor's fee	DC+IC	0.05
	Contingency costs		0.10

Table 5.3: Equipment purchase cost estimation. The correlations were built with data from discussions with our industry partners, SuperPro Designer software, and various publications [150, 261, 269].

Equipment	Reference size	Reference Cost (\$)	Index (n)
Single-use bioreactor	2000 L	\$800,000	0.25
Filtration skid	1 m ²	\$257,420	0.22
PCC Chromatography skid	28.3 L	\$495,000	0
UFDF skid	1 m ²	\$103,250	0
Twin-column AEX skid	0.39 L	\$284,000	
	200 L	\$42,000	
Surge tank	100 L	\$40,000	
	50 L	\$32,000	-
PFR	1.9 L	\$71,000	
TFF	10.5 m ²	\$41,000	
	2.1 m ²	\$26,000	

The second part of production expenditure is OpEX, which covers the cost for raw materials, labor, Lab/QC/QA (laboratory/quality control/quality assurance), consumables, facility-dependent, utilities, and waste treatment and disposal. The costs for materials and consumables are referenced from the published data or the vendor sources, as shown in Table 5.4. Lab/QC/QA is assumed to account for 20-25% of labor cost. Facility-dependent cost refers to the cost associated with the use of facility, such as equipment maintenance, depreciation of the fixed capital cost and miscellaneous cost.

Table 5.4: Consumable cost estimation. The price is based on discussion with our industrial partners and SuperPro Designer software.

Consumables	Reference Cost (\$)
Protein A resin	\$18,600/L
CEX resin	\$2,400/L
AEX resin	\$1,750/L
Membrane absorber	\$15,000/m ²
TFF membrane	\$3,300/m ²
UF membrane	\$4,400/m ²
Dft Membrane	\$560 /m ²
Virus removal filter	\$10,920//m ²
SU bags (500 L)	\$8,000/item
SU bags (200 L)	\$6,000/item
SU bags for mixing (200 L)	\$820/item
SU bags for mixing (100 L)	\$690/item
SU bags for mixing (50 L)	\$600/item
SU bags for storage (100 L)	\$340/item
SU bags for storage (50 L)	\$310/item

5.2.4 Simulation Software

There are three commonly used simulation software packages for economic analysis and the detailed comparison of the three software can be found in our previous paper [150]. In this work, SuperPro Designer (Intelligen, Inc., Scotch Plains, NJ) is chosen for flowsheet modeling and process analysis of the continuous process. This software has many capabilities such as process design, process modeling and scheduling, economic analysis, throughput analysis, and process bottlenecking which we will utilize in the next sections.

5.2.5 Ecology Impacts

Besides cost analysis, the evaluation of ecological impact of a process is also critical to help raise awareness and promote ecological sustainability and facilitate the implementation of new technology using the stricter environmental regulations. The water and consumable consumption of different biopharmaceutical processes are compared by E-factors to assess their environmental burden because E-factor is the most commonly used metric in small-molecule drugs manufacturing [59]. E-factor is defined as the total amount of reagents, water, and consumables per unit product produced. Since single-use technology is incorporated into the process design, the cleaning-in-place (CIP) and steaming-in-place (SIP) requirements can be eliminated [250, 270]. Thus, E-factor values are only calculated for the consumption of process water (including cell culture media and process buffers) and consumables (such as single-use bags, resins, and membranes in Table 5.5) in this work. It is worth noting that process mass intensity (PMI) serves as a metric in other research papers [55, 271, 272] for assessing environmental impact. PMI is characterized as the total amount of water, raw materials, and consumables utilized per unit of active pharmaceutical ingredient produced in a process. Given these definitions, the calculation of E-factors and PMI remains consistent in this scenario.

Table 5.5: Consumable unit masses. The data is from Pollock's thesis [269] and adjusted based on our assumptions.

Consumables	Mass per Unit (kg)	Unit
Chromatographic resin	1.5	Liter of resin
Membrane	4	m ² of filter
SU bags for bioreactor	2.5–7 (varies at different scales)	item
SU bags for surge tanks	1.5–3.4 (varies at different scales)	item

5.2.6 Scenario Analysis

Scenario analysis is performed to evaluate the impacts of changing process parameters on the process economics and ecological impact. Parameters such as the varied bioreactor scales and upstream titer values are taken into consideration to assess the overall process impacts. The values of the different scenarios considered in the scenario analysis are provided in Table 5.6. Specifically, the bioreactor volumes range from 60 to 2000 L for throughput analysis. For upstream titer analysis, the facility size remains constant while the titer values are set to be $\pm 10\%$, $\pm 20\%$, and $\pm 30\%$ of the nominal value of 1.6 g/L.

Table 5.6: Summary of bioreactor scales and upstream titers considered for the base case and scenario analysis.

Varying Parameters	Base scenario	Scenario analysis
Bioreactor scale (L)	500	60, 1000, 2000
Upstream titer (g/L)	1.6	$\pm 10\%$, $\pm 20\%$, $\pm 30\%$

5.2.7 Process Debottlenecking

Process debottlenecking is performed by increasing the production capacity of existing facilities and identifying the bottleneck steps. Investigation of process bottlenecks can help improve the plant capacity with minimal effort and investment, which is of great significance. Annual throughput equals the number of batches multiplied by the batch size throughput. Based on this definition, there are two types of bottlenecks. One is the scheduling or time bottleneck, which is the equipment or resources that limit the number of batches per year. Cycle time reduction can help

address this bottleneck. Since this process is operated continuously, scheduling bottleneck cannot be studied as the cycle time cannot be reduced. Another type of bottleneck is defined by the capacity of the production steps and is referred to as the capacity bottleneck. In the designed continuous process, the upstream is the production stage while the downstream is only for purification, so the perfusion bioreactor in this case is the capacity bottleneck which limits the plant productivity.

5.2.8 Overall Methodology

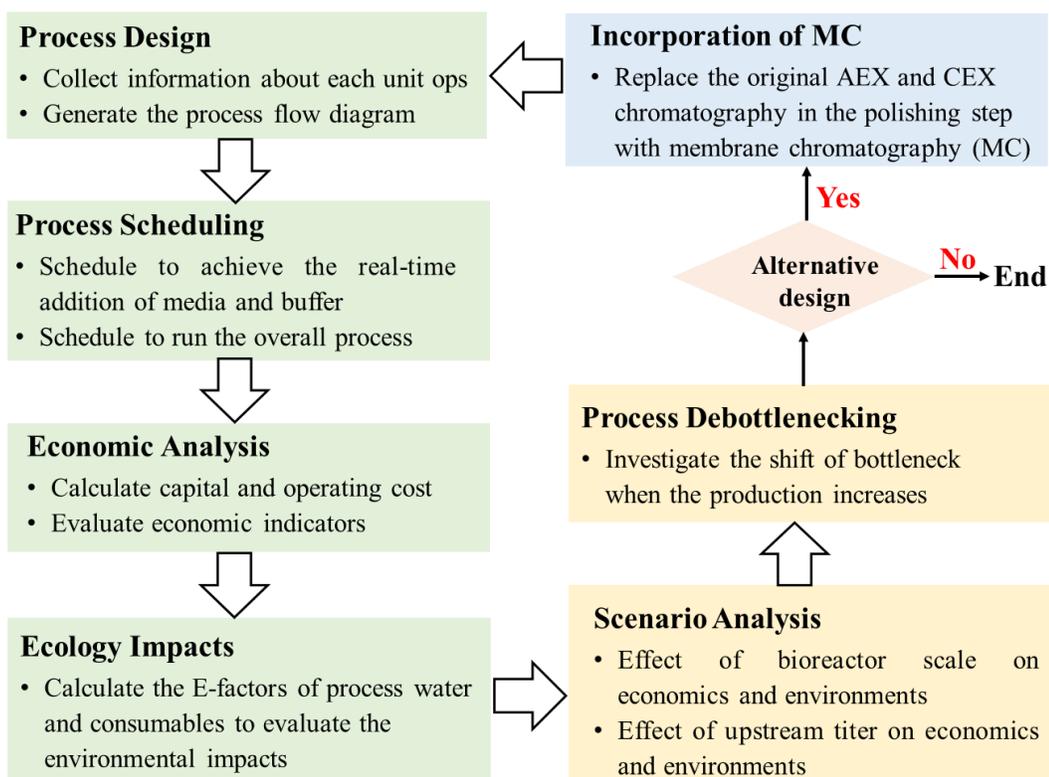


Figure 5.6: The overall framework for cost and ecology evaluation of continuous mAb production. The green box represents the repetitive procedures for an alternative design.

The overall methodology for economic and ecological evaluation of continuous bioprocesses is illustrated in Figure 5.6. Initially the process is designed based on commercial equipment sizes, followed by process scheduling to achieve real-time addition of media and buffer, and integration of the entire process operations. In the next step economic and ecology analyses are carried out to evaluate the cost and environmental impacts of the specific design. Scenario analysis is performed to assess the effects of the bioreactor scale and upstream titer on the economics and environmental impact indicators. Following that, process debottlenecking is conducted to investigate the shift of bottlenecks when the upstream production increases. Membrane chromatography is an emerging technology with significant economic benefits, so the incorporation of membrane chromatography in the process design is also considered, with the economic and ecology examination. Different design alternatives can be considered at this stage of the overall methodology.

5.3 Results and Discussion

5.3.1 Economic Analysis

This section provides the total cost evaluation results based on the base case scenario with the production bioreactor volume of 500 L and titer of 1.6 g/L. Under the current assumptions, the annual production rate is 143.71 kg/yr with the overall capital investment and the yearly operating cost being \$36.6 and \$12.2 million, respectively. The total cost of goods is calculated to be \$102.2/g, of which the operating COGs are \$85.2/g. To help visualize the contribution of each unit operation to the overall economics, the upstream process comprises media preparation and cell culture sections,

while the downstream is divided into primary capture, virus inactivation, polishing, virus removal, final formulation, and buffer preparation sections.

The cost breakdowns of capital investment and operating expenditure in each section are displayed in Figure 5.7. From Figure 5.7a, it can be observed that the cell culture section contributes most to the total capital expenditure, accounting for almost one-third (30%) of the overall investment. The second contributor is the buffer preparation section with a portion of 25%, as there are multiple tanks used to supply buffer, which also indicates the importance of investigating buffer preparation. The next contributor is the sections containing the continuous chromatography skids (i.e., 17% for polishing step and 10% for primary capture). In Figure 5.7b, the upstream process represents 40% of the entire operating cost, with 24% devoted to cell culture and 16% to media preparation. In other words, media preparation contributes as much as 40% of the upstream process, which also prompts us to look more closely at the media addition. The downstream process has the most significant contribution to the operating expenditure at 60%. Among them, buffer preparation (19%) contributes most to the downstream operating cost, followed by primary capture, polishing, virus removal, final formulation, and virus inactivation (14.5%, 12.8%, 7.7%, 3%, and 2.7%; respectively).

Figure 5.7b only offers us the economic contributions at a high level, so the detailed operating COG values at different unit operations are provided in Figure 5.8a, which is consistent with the analysis of Figure 5.7b. Figure 5.8b shows the operating COGs breakdown at different cost categories. It can be found that the unit operating cost of goods for facility-dependent cost (including equipment maintenance, depreciation of the fixed capital cost and miscellaneous cost) is \$45/g, which accounts for the largest portion (approximately 53%). Materials, labor-dependent, and

consumables costs are also significant to the operating expense, which has a percentage of 14.8%, 11.8% and 20.2%; respectively. Since facility-dependent, labor, consumables, and materials costs are significant to the operating cost, the breakdown of each expense for different sections is displayed in Figure 5.9.

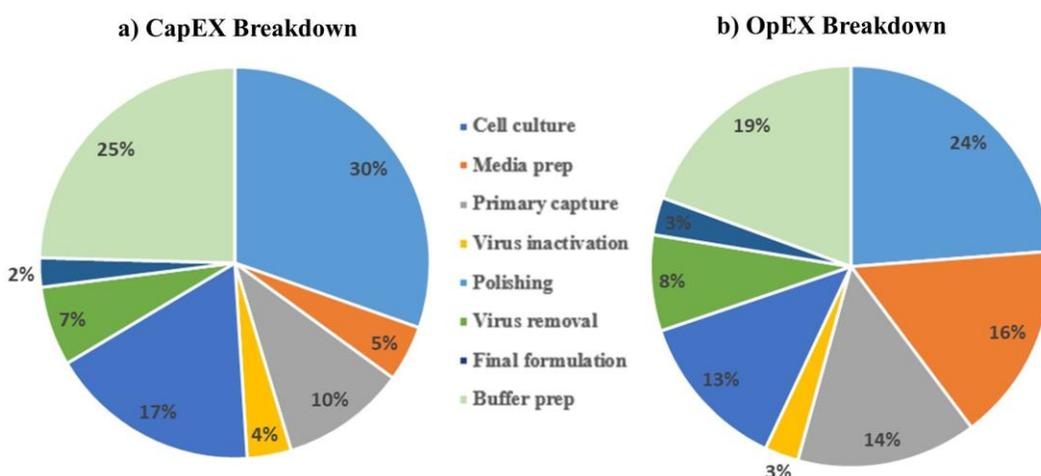


Figure 5.7: Capital cost (a) and operating expenditure (b) breakdown in different sections.

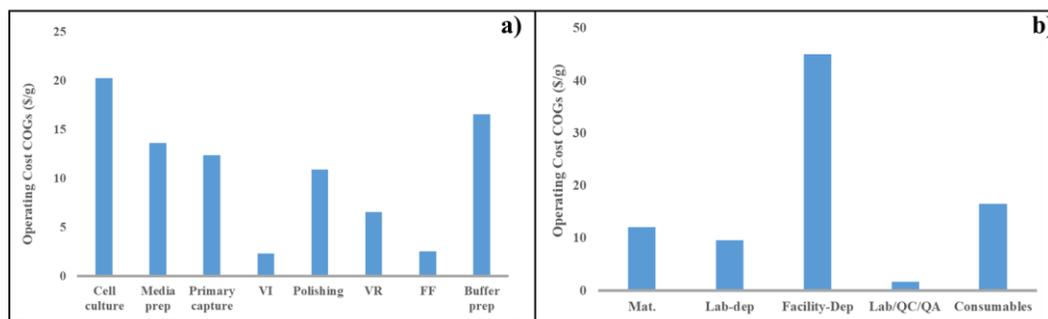


Figure 5.8: Operating cost of goods (COGs) breakdown in different sections (a) and cost categories (b). VI: virus inactivation; VR: virus removal; FF: final formulation.

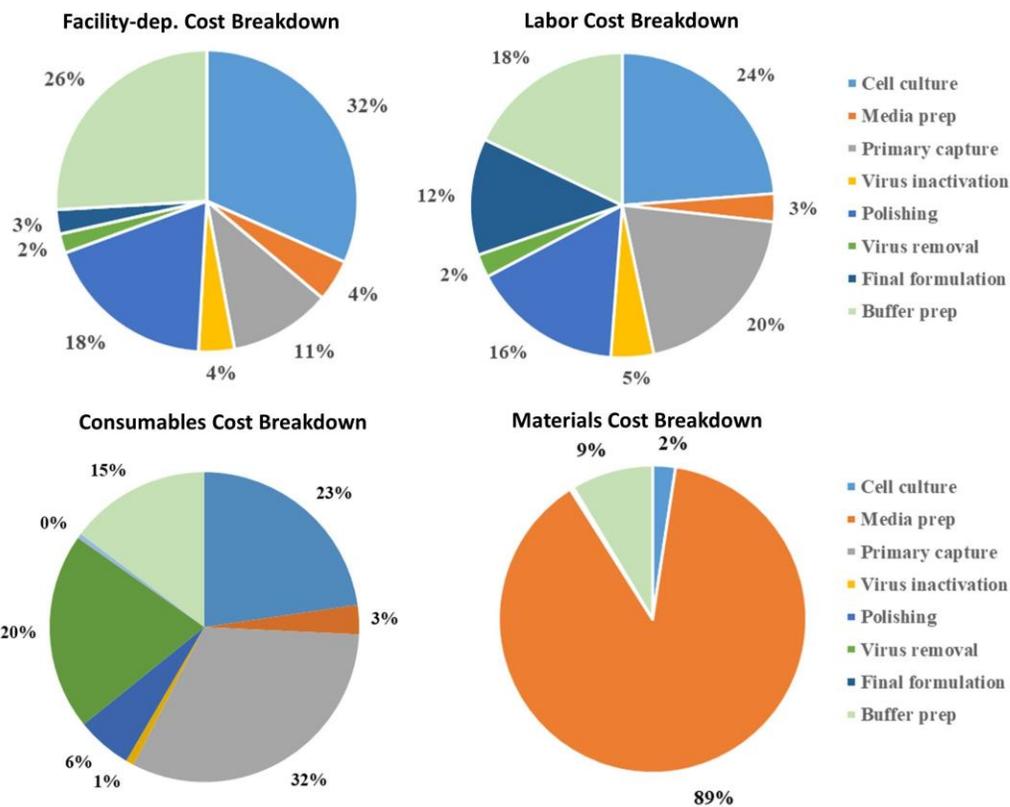


Figure 5.9: Cost breakdown with different sections.

The facility-dependent cost breakdown analysis result follows the same trend of capital investment breakdown because the facility-dependent cost is highly dependent on equipment cost. The total labor costs are estimated to be around \$1.4 million for the continuous process, similar to the value in Klutz's work [261]. The labor-cost breakdown analysis can give a general idea of the labor cost distribution in each section. It can be seen that cell culture, primary capture, buffer preparation, and polishing steps are significant contributors. From the consumable cost breakdown, primary capture is found to account for the highest portion because Protein A resin is very expensive, which is in line with our previous analysis [150]. Cell culture, virus removal, buffer

preparation, and polishing steps also play an essential role in the overall consumable cost as single-use technologies are incorporated into the process design. In terms of the materials cost breakdown, it can be seen that the media preparation contributes as much as 89% of the overall materials cost, conforming with what is observed in practice.

5.3.2 Ecological Analysis

Ecological analysis is performed to evaluate the environmental footprint of the designed continuous process on the basis of E-factor analysis, as shown in Table 5.7. E-factor values are expected to be negatively related to the environmental burden – a lower value represents fewer resources required to produce 1 kg of product. It can be found that water usage has the most significant contribution to the process E-factors, accounting for nearly 99.7%, which is in agreement with the findings that this is a water-intensive process operation [55, 263, 273]. In our simulation, 4865.6 kg water is required to make 1 kg mAbs, which is consistent with the findings in the literature that a standard mAb manufacturing process consumes 3,000 to over 7,000 kg water per kilogram product [258]. In addition, the upstream process consumes approximately 29.6% of the process water, similar to 20% to 25% consumption [258]. The E-factor for consumable consumption is estimated to be 11.1 kg/kg, which is in the range of 4.6 to 37.4 kg/kg reported in the literature [59]. Due to the dominance of water consumption by the media and buffer preparation sections, the breakdown of E-factors in each section is not provided. Instead, the comparison of process water usage between upstream and downstream processes is given.

Table 5.7: E-factors for process water and consumable usage in the upstream, downstream, and whole process.

	E-factors (kg/kg product)		
	Upstream	Downstream	Whole Process
Process water	1441.6	3424.0	4865.6
Consumables	4.2	6.9	11.1
Total E-factor	1445.8	3430.9	4876.7

5.3.3 Scenario Analysis

Parameter fluctuations and variations are inevitable in practice during pharmaceutical manufacturing. In order to evaluate the inherent risks and estimate the impacts of variable bias on the process, scenario analysis is performed in this chapter. Two variables – bioreactor scale and upstream protein titer – are assessed to investigate their effects on the economic and ecological results. The deviations of process variables in the scenario study are summarized in Table 5.6.

5.3.3.1 Throughput Analysis

Throughput analysis is conducted to evaluate the effect of bioreactor scales on production economics. Four different scales from 60 to 2000 L are chosen to perform the throughput study, as shown in Figure 5.10. It should be noted that the four volumes selected are commercialized equipment sizes. As can be seen from Figure 5.10a, the unit operating cost of goods (\$/g) decreases as the bioreactor scale increases, as would be expected from “economy of scale” [274]. Figure 5.10b provides the media consumption at different scales, which increases from 2,235 to 72,281 L/batch with the increase of the bioreactor volume from 60 to 2000 L. This analysis of media utilization would help decide the manufacturability of the process. Figure 5.10c shows the unit

operating cost breakdowns with different categories and sections. It can be found that facility-dependent cost significantly decreases as this cost is highly dependent on the capital cost. The total labor cost should remain unchanged in different cases, but there is a reduction in the unit labor cost as the production rate increases. The consumables and Lab/QC/QA costs per unit also slightly decrease, but there is negligible change in the unit material cost, probably because this cost is dominated by the media consumption independent of the bioreactor scale. In Figure 5.10d, it can be found that the unit operating cost of each section decreases, especially for the unit operations of buffer preparation, polishing, primary capture, and cell culture. Among them, buffer preparation has the largest cost reduction because the capital cost of buffer tanks per unit product remarkably decreases as the plant capacity increases, evidenced by the significant decrease in unit facility-dependent cost from \$82.1/g to \$3.2/g in buffer preparation.

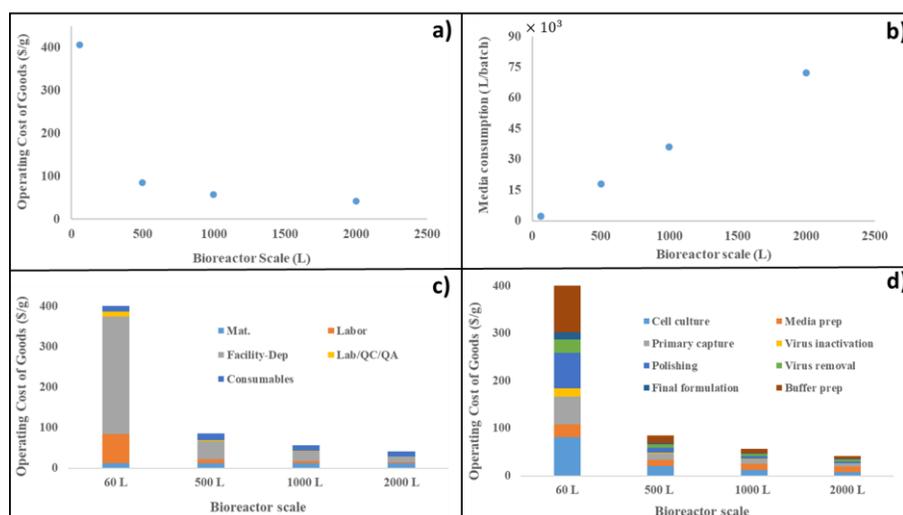


Figure 5.10: Change of total operating COGs (a), media consumption per batch (b), COGs breakdown at varying cost categories (c) and sections (d) with respect to different bioreactor scales.

In addition to the impacts of the bioreactor scale on the process economics, the effects on the ecological impacts are also investigated, as shown in Figure 5.11. Instead of comparing the absolute values of the E-factor, all the E-factors are normalized to the base case scenario, and the change of ratio to nominal value is compared. From Figure 5.11a, it can be seen that the consumable usage decreases when the bioreactor scale increases, especially at the small scale. The trend of process water consumption and total E-factor is the same, due to the significant contribution of water usage to the whole process. It is interesting to observe that the water utilization first increases and then slightly decreases with the increase of bioreactor scales. When the process is scaled up or down, each equipment size is adjusted to make sure that it is commercially available instead of just scaling based on the bioreactor volume. In the process design with a 60 L bioreactor, the working volume is set to 50 L because this volume is more likely to be used in practical applications. The ratio of working volume is 80% for the other bioreactor sizes. It is highly possible that the higher working volume ratio (83.3%) on a small scale results in lower water consumption as more products are produced. When the bioreactor scale increases from 500 to 2000 L, the water usage almost remains constant and only has a 2% reduction (decreasing from 4876.7 to 4775.6 kg/kg mAbs). This result makes sense as the materials used should be proportional to the plant capacity. Thus, there are few impacts of bioreactor scale on the process water consumption, but great effects on the consumable utilization on a small scale. Figure 5.11b shows the comparison of the percentage of water consumption between upstream and downstream operations, and it can be found that there are only minimal changes in the percentage when the bioreactor volume is changed.

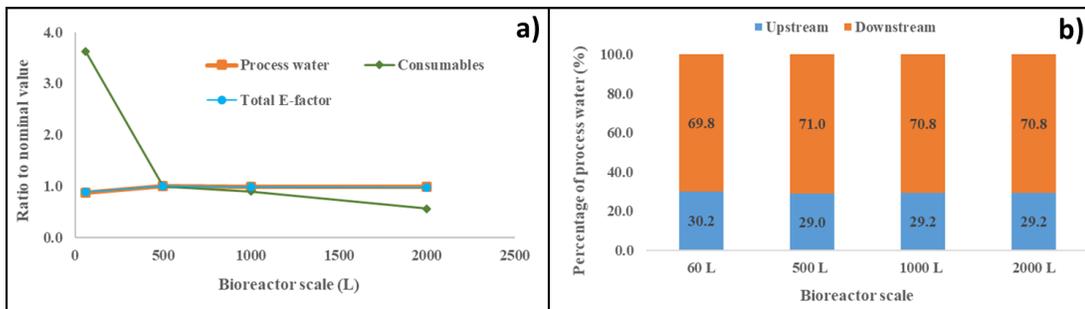


Figure 5.11: Change of the ratio to the normal value of different ecology indicators (a) and comparison of the percentage of process water between upstream and downstream (b) with respect to bioreactor scales.

5.3.3.2 Upstream Titer Analysis

Titer, representing the concentration of protein harvested from the perfusion bioreactor, has notable influences on productivity, further affecting the downstream operations like operating time and cycles of chromatography. Therefore, multiple scenarios with varying titers around the nominal value are designed to evaluate the effect of titer on process economics, as listed in Table 5.6. The facility and manufacturing line remain the same when performing upstream titer analysis.

First, it can be clearly seen that the overall unit operating cost (\$/g) decreases by 59% with the increase of titer from 1.12 to 2.08 g/L in Figure 5.12a, resulting from the increased annual production rate. The unit operating cost breakdown into different categories is shown in Figure 5.12b. The overall labor, facility-dependent, and Lab/QC/QA cost stay constant as the plant is fixed, but the corresponding unit operating cost reduction is observed as more product is produced. However, the total material and consumable costs increase, caused by the increased buffer consumption and usage of consumables like chromatography resins and filtration membranes. However, each corresponding unit cost decreases. The operating cost per unit in each section is also

inversely proportional to the titer due to the expansion of productivity to produce more mAbs.

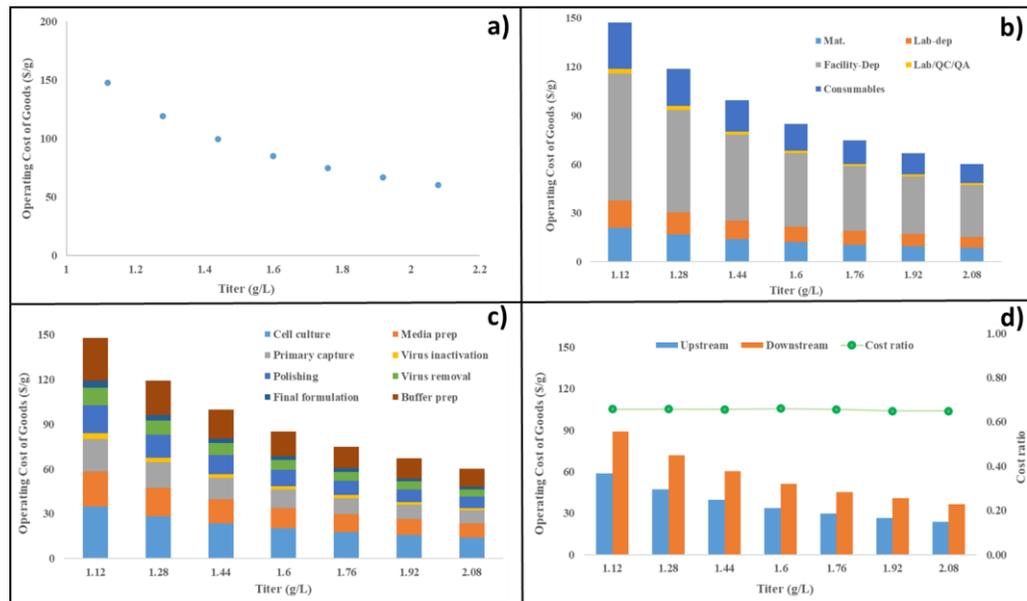


Figure 5.12: Change of total operating COGs (a), COGs breakdown at varying cost categories (b) and sections (c) versus the titer; Change of operating COGs in upstream, downstream, and the cost ratio (d).

Moreover, the full operating costs in primary capture, polishing, final filtration, and buffer preparation are observed to increase, consistent with the above analysis, as illustrated in Figure 5.12c. It is interesting to observe as shown in Figure 5.12d, that the operating cost ratio between upstream and downstream processes is almost constant with increasing titer values. In our previous work [150], the ratio decreases in the continuous operation because the process bottleneck emerges in the downstream operation when the titer increases from 1.5 to 5.5 g/L and the operating cycles and time of primary capture are adjusted to debottleneck, but in our case, there is no bottleneck

emerging with the increase of titer up to 2.08 g/L. Thus, the process scheduling does not change, resulting in a constant ratio.

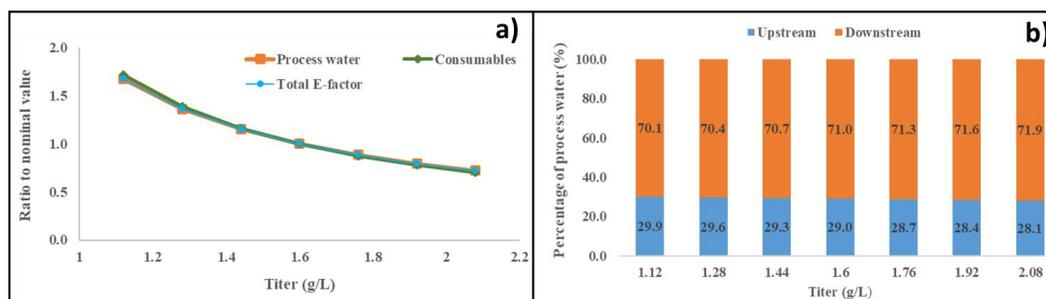


Figure 5.13: Change of the ratio to the normal value of different ecology indicators (a) and comparison of the percentage of process water between upstream and downstream (b) with respect to different titers.

The impact of upstream titers on the ecological footprint is also investigated. It can be seen from Figure 5.13a that the trend of different ecology indicators (water, consumables, and process consumption) with regard to the titer is almost the same. The ratio of the total E-factor to the base case nominal value decreases from 1.68 to 0.72 when the titer increases from 1.12 to 2.08 g/L. In other words, the total consumption of water and consumables was reduced by 57.1%, indicating the significant improvement that the upstream titer on the environmental burden. Under the same platform, more product is purified, leading to more efficient use of consumables. Moreover, it is interesting to find that the changing trend slows as the titer continues to increase, suggesting that the ecological impacts of the existing plant are at the minimum when the titer achieves the critical value (2.36 g/L, introduced in Section 5.3.4). Figure 5.13b shows the comparison of water consumption between upstream and downstream

operations when titer varies. It can be found that the percentage of upstream water usage slightly decreases from 29.9% to 28.1% when titer increases from 1.12 to 2.08 g/L. The total water consumption for cell culture should remain the same when the titer changes under the current platform, but the upstream E-factor decreases as the production rate increases. As more product is produced upstream and needed to purify in the downstream operations, the water consumption for downstream increases from 70.1% to 71.9%. It should be noted that the change in the percentage in the upstream and downstream is small, because there is no bottleneck emerging in the current titer range.

5.3.4 Process Debottlenecking

Under the current platform, the bottleneck is identified as the production bioreactor since this unit operation influences the production capacity, but the bottleneck will shift from the upstream to downstream unit operations as the upstream production increases. The annual throughput depends on the bioreactor scale, titer, and/or perfusion rate. Since the current platform (i.e., the equipment size and process flow) remains unchanged, only two variables (i.e., titer and perfusion rate) are considered to increase the production and evaluate the bottleneck shift.

The downstream operations including virus inactivation, virus removal, and final formulation are operated continuously for 28 days, so they will not become bottlenecks as the upstream production rate increases. The unit operations with cyclic behaviors like primary capture and polishing steps will potentially become bottlenecks. Our hypothesis is that by comparing the capacity utilization of the chromatography columns, we can determine which operation becomes bottleneck first. From Table 5.8, it can be found that protein A chromatography (ProA) has the highest capacity utilization (CU) while AEX has the lowest CU. Based on the previous assumption,

primary capture should become the bottleneck first when the production increases, followed by CEX and AEX. This hypothesis is tested by increasing the titer and perfusion rate to observe the occurrence and shifts of bottlenecks. The results are displayed in Table 5.9 and it is consistent with our presumption. Moreover, even if we push the titer or perfusion rate to the limit, no other unit operations will become bottlenecks, except those with cyclic processes.

Table 5.8: Summary of the capacity utilization in each chromatography.

Chromatography	Capacity utilization (CU)
Protein A (ProA) chromatography	66.4%
AEX chromatography	44.8%
CEX chromatography	52.7%

Table 5.9: Emerging of process bottlenecks with the increase of titer and perfusion rate.

Titer (g/L)	Process bottleneck	Perfusion rate (vvm)	Process bottleneck
1.6	--	1.5	--
2.4	ProA	2.5	ProA
3.0	ProA + CEX	3.0	ProA + CEX
3.6	ProA + CEX + AEX	3.5	ProA + CEX + AEX
4.0	ProA + CEX + AEX	4.0	ProA + CEX + AEX
5.0	ProA + CEX + AEX	5.0	ProA + CEX + AEX

The critical titer and perfusion rate are also investigated under the current platform, i.e., the primary capture would become a bottleneck if the titer or perfusion rate is further increased from the critical value. The operating cost per unit product is also compared in Table 5.10 and it can be concluded that the process is more economically efficient at critical titer compared with critical perfusion rate. The main reason is that increasing the perfusion rate will increase the media consumption per day, contributing to the increase of material and media preparation cost as shown in Figure 5.14.

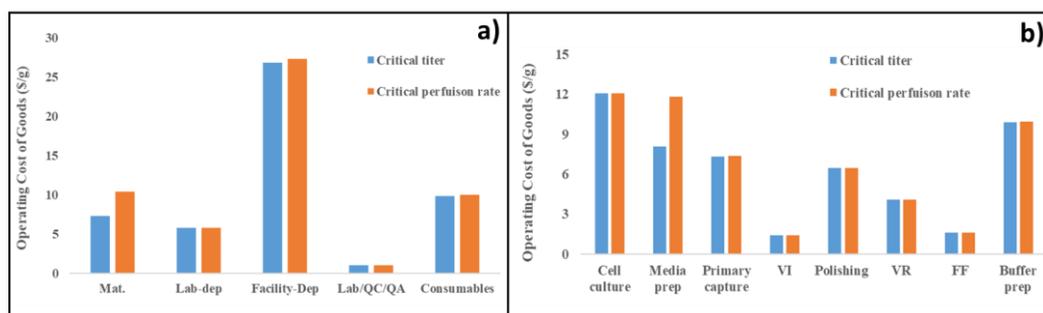


Figure 5.14: Comparison of unit operating COGs under different cost categories (a) and sections (b) at critical titer and perfusion rate. VI: virus inactivation; VR: virus removal; FF: final formulation.

Table 5.10: Unit operating COGs at critical titer and perfusion rate.

Variable	Critical value	Unit operating COGs (\$/g)
Titer	2.36 g/L	50.9
Perfusion rate	2.27 vvm	54.8

5.3.5 Alternative Design – Membrane Chromatography

Packed-bed column chromatography using resins has been widely applied in biopharmaceutical manufacturing, but it mainly suffers from the slow pore diffusion rate, severely limiting its separation efficiency [265]. An alternative technology is adsorptive membrane chromatography, which integrates membrane and liquid chromatography into a single-stage operation. The economic benefit of membrane chromatography has been demonstrated compared with resin based chromatography. Tressel et al. [275] and Muthukumar et al. [257] investigated the economic viability without incorporating the membrane chromatography into the whole process, and Varadaraju et al. [276] performed the economic evaluation based on a batch process by replacing the overall purification train with a membrane-only process. Nevertheless, its application into the continuous end-to-end bioprocess has not been examined yet.

Membrane chromatography, although not a new technology, offers many benefits, such as eliminating pore diffusion, lowering buffer consumption, reducing footprint, and saving the need of packing/unpacking the column [277, 278]. Moreover, the smaller bed volumes provided by membrane chromatography may result in reduced process time. All these benefits are of interest to the industry, so the incorporation of membranes into the continuous process is demonstrated. Moreover, membrane chromatography is still a relatively new and immature technology, restricting its commercialization due to the low binding capacity and uneven flow characteristics encountered during scale-up processes [264], so it is considered in the alternative design instead of the base case scenario. In this work, twin-column AEX and CEX membrane chromatography in flow-through mode are employed to replace the original polishing steps, referred to as “hybrid case.”

The economic and ecological comparisons of the base scenario and hybrid case are illustrated in Figure 5.15. The total operating COG/g decreases from \$85.2/g to \$75.4/g after the incorporation of the membrane chromatography in the continuous platform. As shown in Figure 5.15a, the cost reduction is mainly due to the polishing step and buffer preparation, demonstrating the economic benefits and low buffer usage of membrane chromatography, in line with the analysis provided in [257, 276]. Besides, the hybrid case shows the environmental benefits in comparison to the base case in Figure 5.15b. Although there is only a slight reduction from 11.1 to 10.5 in the consumable E-factor, the process water consumption significantly decreases from 4865.7 to 3508.7. Thus, investigation and incorporation of membrane chromatography into continuous bioprocess platforms can not only save operating costs but also alleviate the environmental burden by reducing water usage.

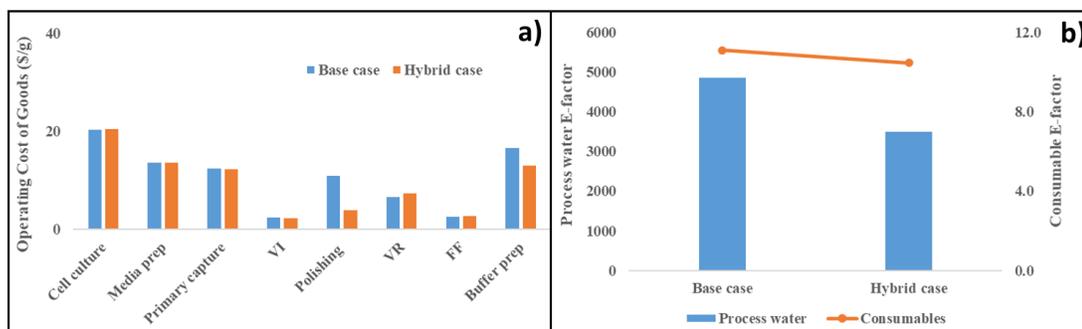


Figure 5.15: Economical and ecological comparisons of the base case and hybrid case scenarios. VI: virus inactivation; VR: virus removal; FF: final formulation.

5.4 Conclusions

There is a rising interest in the transition from batch to continuous process in biopharmaceutical manufacturing to increase production flexibility and reduce cost. To

investigate the feasibility of developing continuous bioprocesses, a fully integrated end-to-end continuous platform to produce mAbs is established in SuperPro Designer in this work. Some novel technologies are considered to be incorporated into process design, such as the N-1 seed bioreactor to obtain higher target cell density and reduce the expansion time and single-pass tangential flow filtration for the final formulation. Most importantly, media and buffer preparation steps are added, and the real-time supply of media and buffer to the master equipment is achieved within process scheduling. The process with a bioreactor volume of 500 L is regarded as the base-case scenario with an annual production rate of 143.71 kg/yr.

After building the benchmark platform, economic and environmental analyses are thoroughly performed. Economic assessment is on the basis of the cost of goods, and for the base case, the total COGs are calculated to be \$102.2/g, of which the operating COGs are \$85.2/g. A detailed analysis of the economics breakdown for different sections and cost categories is also provided. It is found that the most significant contributor to the capital expenditure is cell culture, followed by the buffer preparation accounting for 25%, suggesting that it is not reasonable to ignore the buffer preparation step for cost analysis. In addition, media preparation contributes as much as 40% to the upstream operating cost, and buffer preparation dominates the downstream operating expenditure. To evaluate the environmental impacts, E-factor is used as a quantitative metric. The simulation result shows that 4865.6 kg of process water and 11.1 kg of consumables are required to manufacture 1 kg of product, demonstrating the water-intensive nature of biopharmaceutical manufacturing.

Since process demand and productivity are variable, scenario analysis is conducted to examine the impacts of bioreactor scale and upstream titers on the process

economics and the environmental footprint. The unit operating COGs is found to decrease as either the bioreactor scale or upstream titer increases. The COGs breakdown at different cost categories and sections are also thoroughly analyzed. It is found that unit operating COGs of buffer preparation leads to significant cost reduction, especially when the bioreactor scale increases from 60 to 500 L. With the increase of titers from 1.12 to 2.08 g/L, the overall operating cost per gram decreases by 59%. In terms of the environmental impacts, the escalation of bioreactor volumes decreases the consumable usage, especially at the small scale, but has little effect on the process water consumption. However, it can be observed that the consumption of water and consumables reduces by 57.1% with the increase of upstream titer, suggesting that increasing titer will improve the environmental burden.

Moreover, the shift of bottlenecks is examined when the upstream production improves by increasing titer or perfusion rate. It was found that the bottleneck shifts from upstream to downstream cyclic chromatographic operations with the expansion of upstream production. Moreover, which operation becomes bottleneck first can be determined based on the capacity utilization of chromatography units. The critical titer and perfusion rate are also found, and the increase of titer is more economically efficient than that of perfusion rate.

Membrane chromatography has demonstrated many advantages, so it is incorporated into the continuous bioprocess platform to replace the original polishing steps. It is found that the integration of membrane chromatography is more economical and more environmentally friendly based on reduction of water consumption.

For future work, the developed model will be compared with the experimental results from the test bed team to further improve the prediction ability of the model. In

addition, the process failure rate can be considered in process design and economic evaluation as process failure occasionally happens due to filter clogging and fouling. To thoroughly investigate environmental sustainability, life cycle assessment (LCA) of the continuous bioprocess can be performed. With the incorporation of the above considerations, a more realistic and flexible continuous platform can be built.

Chapter 6

CONCLUSIONS AND PERSPECTIVES OF FUTURE WORK

6.1 Conclusions

The development of *in silico* strategies for development and implementation of new technologies in biopharmaceutical manufacturing processes can help reduce computational resources, save experimental efforts, and facilitate process development. In this dissertation, process systems engineering tools have been applied to downstream chromatography operations and the end-to-end integrated biopharmaceutical process.

The first part of the thesis focuses on individual unit operations, namely the different chromatography units, as illustrated in Chapters 2-4.

In Chapter 2, surrogate-based feasibility analysis is proposed to identify the design space of CaptureSMB in an effort to balance the computational complexity and model prediction accuracy. Based on the proposed framework, the computational time is reduced significantly by approximately 97%, compared to directly using mechanistic model. Through active set strategy, it is observed that productivity and yield constraints become active with the increase of process variables, while productivity and capacity utilization constraints are binding with the decrease of process variables. Additionally, the impacts of constraints, recovery-regeneration time, and column length are comprehensively investigated to acquire a profound understanding of the continuous capture process. The proposed framework is generalizable and can be adapted to any other specific processes for the identification of design space.

For Chapter 3, we introduced a ML-based optimization framework to address the nonconvex and nonlinear constrained optimization challenges encountered in biopharmaceutical separation. This framework is applied to a case study involving the separation of a ternary protein mixture. Compared to the results obtained through genetic algorithm, this approach enhanced productivity by 50.1% and reduced computation time by 70.8% simultaneously. The impacts of elution strategies and purity requirements on the overall chromatographic behaviors were comprehensively investigated. The impact of varying the peak cutting thresholds was also examined to discern the critical constraint influencing the optimized conditions, which might help enhance our process understanding and facilitate potential process improvements. The methodology presented in this study can provide valuable guidelines for addressing problems where identifying optimal solutions is challenging.

In Chapter 4, a hybrid model approach is employed for the HIC process, aiming to reduce investment effort required for the mechanistic model while extracting the missing relationships that cannot be captured by the mechanistic model. We proposed that the hybrid model can be constructed by combining a neural network with a simple but well-known isotherm (multi-component Langmuir). It is noted that the structure of the hybrid model is of critical importance to the accuracy of the developed model. Furthermore, a simple neural network with one hidden layer with two nodes and sigmoid as the activation function, significantly outperforms the mechanistic model, with a 62% improvement in accuracy in calibration and 31.4% in validation. The developed hybrid model demonstrates excellent extrapolation capability and identifies optimal operating conditions resulting in higher yield value compared to the mechanistic model.

The second part of the applications focuses on conducting POC study to evaluate the feasibility of transitioning from batch to continuous bioprocess. Therefore, in Chapter 5, we developed a fully integrated *in silico* end-to-end continuous platform to represent continuous mAb production in SuperPro Designer. The base-case scenario involves a bioreactor volume of 500 L, resulting in a total COGs at \$102.2/g. The environmental analysis reveals the water-intensive nature of biopharmaceutical manufacturing. Scenario analysis indicates that increasing bioreactor scale and mAb titer reduces operating COGs, with higher mAb titers showing greater ecological benefits. Debottlenecking study identifies multi-column chromatography as the process bottleneck, with the specific bottleneck dependent on capacity utilization for that step. The assessment of membrane chromatography demonstrates its cost-effectiveness and environmental friendliness, primarily attributed to its reduced water consumption.

6.2 Future Work

Opportunities for future research are outlined in this section.

Surrogate-based feasibility analysis is proposed in Chapter 2 to identify the design space with low computational demands. However, this framework does not account for uncertainties within the system, such as uncertainties from mechanistic model parameters or operating conditions [279-282]. Therefore, considering a probabilistic design space could be seen as a potential future direction.

The proposed two-stage optimization framework in Chapter 3 solely focuses on solving single-objective optimization problems, and the criteria to switch from feasibility to optimization stage could be further investigated. Consequently, there is an opportunity for extending this framework to address multi-objective challenges, such as the simultaneous minimization of cost and yield, along with a more intelligent algorithm

in stage-switching. This extension would better reflect the complexities inherent in the actual biopharmaceutical process.

The proposed methodology for constructing a hybrid model in Chapter 4 has been validated for the HIC process, with the challenge lying in finding an appropriate hybrid model structure. However, its suitability for other complex chromatographic operations, such as mixed-mode chromatography, remains uncertain. Therefore, future research could focus on implementing the hybrid model approach in other chromatographic processes to evaluate its effectiveness in real-world applications.

In terms of the integrated flowsheet model development in Chapter 5, the developed model is primarily based on mass balance, which is sufficient for the POC study during the early stage of process design. However, in order to gain a deeper understanding of the process, such as capturing critical process parameters and critical quality attributes, a more detailed flowsheet model that connects mechanistic models of various unit operations is desired. During the development of the detailed flowsheet model, machine learning algorithms and hybrid modeling strategies could be incorporated to reduce the computational complexity and improve the model accuracy.

REFERENCES

1. Narayanan, H., M. Sponchioni, and M. Morbidelli, *Integration and digitalization in the manufacturing of therapeutic proteins*. Chemical Engineering Science, 2022. **248**.
2. Walsh, G., *Biopharmaceutical benchmarks 2018*. Nature biotechnology, 2018. **36**(12): p. 1136-1145.
3. Gargalo, C., et al., *Towards the development of digital twins for the biomanufacturing industry*. Digital Twins, 2020: p. 1-34.
4. Kaplon, H., et al., *Antibodies to watch in 2020*. MAbs, 2020. **12**(1): p. 1703531.
5. Sokolov, M., *Decision making and risk management in biopharmaceutical engineering—opportunities in the age of covid-19 and digitalization*. Industrial & Engineering Chemistry Research, 2020. **59**(40): p. 17587-17592.
6. Khanal, O. and A.M. Lenhoff, *Developments and opportunities in continuous biopharmaceutical manufacturing*. MAbs, 2021. **13**(1): p. 1903664.
7. Bielser, J.M., et al., *Perfusion mammalian cell culture for recombinant protein manufacturing - A critical review*. Biotechnol. Adv., 2018. **36**(4): p. 1328-1340.
8. Jones, W. and D.I. Gerogiorgis, *Dynamic simulation, optimisation and economic analysis of fed-batch vs. perfusion bioreactors for advanced mAb manufacturing*. Computers & Chemical Engineering, 2022. **165**.
9. Lin, D.Q., Q.L. Zhang, and S.J. Yao, *Model-assisted approaches for continuous chromatography: Current situation and challenges*. J Chromatogr A, 2021. **1637**: p. 461855.
10. Rathore, A.S., G. Thakur, and N. Kateja, *Continuous integrated manufacturing for biopharmaceuticals: A new paradigm or an empty promise?* Biotechnol Bioeng, 2022.
11. 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*. Biotechnology progress, 2010. **26**(6): p. 1505-1527.
12. Ding, C., O. Yang, and M. Ierapetritou, *Towards Digital Twin for Biopharmaceutical Processes: Concept and Progress*, in *Biopharmaceutical Manufacturing: Progress, Trends and Challenges*. 2024, Springer. p. 179-211.
13. Chen, Y., et al., *Digital Twins in Pharmaceutical and Biopharmaceutical Manufacturing: A Literature Review*. Processes, 2020. **8**(9).

14. Grieves, M., *Intelligent digital twins and the development and management of complex systems*. Digital Twin, 2022. **2**(8): p. 8.
15. Grieves, M. and J. Vickers, *Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems*, in *Transdisciplinary perspectives on complex systems*. 2017, Springer. p. 85-113.
16. Grieves, M., *Digital twin: manufacturing excellence through virtual factory replication*. White paper, 2014. **1**(2014): p. 1-7.
17. Glaessgen, E. and D. Stargel. *The digital twin paradigm for future NASA and US Air Force vehicles*. in *53rd AIAA/ASME/ASCE/AHS/ASC structures, structural dynamics and materials conference 20th AIAA/ASME/AHS adaptive structures conference 14th AIAA*. 2012.
18. Kritzinger, W., et al., *Digital Twin in manufacturing: A categorical literature review and classification*. IFAC-PapersOnLine, 2018. **51**(11): p. 1016-1022.
19. Fuller, A., et al., *Digital twin: Enabling technologies, challenges and open research*. IEEE access, 2020. **8**: p. 108952-108971.
20. Cimino, C., E. Negri, and L. Fumagalli, *Review of digital twin applications in manufacturing*. Computers in Industry, 2019. **113**: p. 103130.
21. Bao, J., et al., *The modelling and operations for the digital twin in the context of manufacturing*. Enterprise Information Systems, 2019. **13**(4): p. 534-556.
22. Tao, F., et al., *Digital twins and cyber-physical systems toward smart manufacturing and industry 4.0: Correlation and comparison*. Engineering, 2019. **5**(4): p. 653-661.
23. Tao, F., et al., *Digital twin-driven product design, manufacturing and service with big data*. The International Journal of Advanced Manufacturing Technology, 2018. **94**(9): p. 3563-3576.
24. Sokolov, M., et al., *Hybrid modeling—a key enabler towards realizing digital twins in biopharma?* Current Opinion in Chemical Engineering, 2021. **34**: p. 100715.
25. Qi, Q., et al., *Digital twin service towards smart manufacturing*. Procedia Cirp, 2018. **72**: p. 237-242.
26. Li, X., et al., *A review of industrial wireless networks in the context of industry 4.0*. Wireless networks, 2017. **23**(1): p. 23-41.
27. Uhlemann, T.H.-J., et al., *The digital twin: demonstrating the potential of real time data acquisition in production systems*. Procedia Manufacturing, 2017. **9**: p. 113-120.
28. Damiani, L., et al., *Augmented and virtual reality applications in industrial systems: A qualitative review towards the industry 4.0 era*. IFAC-PapersOnLine, 2018. **51**(11): p. 624-630.
29. Zhuang, C., J. Liu, and H. Xiong, *Digital twin-based smart production management and control framework for the complex product assembly shop-floor*. The international journal of advanced manufacturing technology, 2018. **96**(1): p. 1149-1163.

30. Tripathi, N.K. and A. Shrivastava, *Recent Developments in Bioprocessing of Recombinant Proteins: Expression Hosts and Process Development*. Front. Bioeng. Biotechnol., 2019. **7**: p. 420.
31. Somasundaram, B., et al., *Progression of continuous downstream processing of monoclonal antibodies: Current trends and challenges*. Biotechnol. Bioeng., 2018. **115**(12): p. 2893-2907.
32. Rathore, A.S., et al., *Continuous processing for production of biopharmaceuticals*. Prep. Biochem. Biotechnol., 2015. **45**(8): p. 836-49.
33. Hogwood, C.E., et al., *The dynamics of the CHO host cell protein profile during clarification and protein A capture in a platform antibody purification process*. Biotechnol. Bioeng., 2013. **110**(1): p. 240-51.
34. Zydney, A.L., *Continuous downstream processing for high value biological products: A Review*. Biotechnol. Bioeng., 2016. **113**(3): p. 465-75.
35. Ramos - de - la - Peña, A.M., J. González - Valdez, and O. Aguilar, *Protein A chromatography: Challenges and progress in the purification of monoclonal antibodies*. J. Sep. Sci., 2019. **42**(9): p. 1816-1827.
36. Johnson, S.A., et al., *Virus filtration: A review of current and future practices in bioprocessing*. Biotechnology and Bioengineering, 2022. **119**(3): p. 743-761.
37. Shirataki, H., *Analysis of filtration with virus removal filters using the characteristic form of blocking model*. Biochemical Engineering Journal, 2022. **183**: p. 108460.
38. Liu, H.F., et al., *Recovery and purification process development for monoclonal antibody production*. MAbs, 2010. **2**(5): p. 480-99.
39. Briskot, T., et al., *Modeling the Gibbs–Donnan effect during ultrafiltration and diafiltration processes using the Poisson–Boltzmann theory in combination with a basic Stern model*. Journal of Membrane Science, 2022. **648**: p. 120333.
40. Steinebach, F., T. Muller-Spath, and M. Morbidelli, *Continuous counter-current chromatography for capture and polishing steps in biopharmaceutical production*. Biotechnol. J., 2016. **11**(9): p. 1126-41.
41. Smiatek, J., A. Jung, and E. Bluhmki, *Towards a Digital Bioprocess Replica: Computational Approaches in Biopharmaceutical Development and Manufacturing*. Trends Biotechnol., 2020. **38**(10): p. 1141-1153.
42. Pfizer, *M-Star - How Pfizer Leveraged Digital Twins to Create A Process Scale-Up Roadmap*. 2020: https://mstarafd.com/wp-content/uploads/2021/12/M-Star_Pfizer-CaseStudy.pdf.
43. SIEMENS, *Stepping up the pace in vaccine development and production*. 2021: <https://www.siemens.com/global/en/company/stories/industry/2021/pharma-vaccine-digitalization.html>.
44. GlaxoSmithKline, *Digital twin: using advanced technology to accelerate vaccine development*. 2022: <https://www.gsk.com/en-gb/behind-the-science->

- [magazine/digital-twin-using-advanced-technology-to-accelerate-vaccine-development/](#).
45. SIEMENS, *Siemens accelerated BioNTech Covid-19 vaccine production setup with automation and digitalization solutions*. 2021: <https://new.siemens.com/th/en/company/press-centre/siemens-accelerated-biontech-covid-19-vaccine-production-setup-.html>.
 46. DassaultSystèmes, *Dassault Systèmes Partners with Sanofi to Optimize Tech Transfer and Industrialization at Its Future “EVolutive Facilities”*. 2022: <https://investor.3ds.com/news-releases/news-release-details/dassault-systemes-partners-sanofi-optimize-tech-transfer-and>.
 47. Jones, D., et al., *Characterising the Digital Twin: A systematic literature review*. CIRP Journal of Manufacturing Science and Technology, 2020. **29**: p. 36-52.
 48. Chen, Y., M. Ierapetritou, and P. Bhalode, *Process Systems Engineering Tools toward Digital Twins of Pharmaceutical Continuous Manufacturing Processes*, in *Continuous Pharmaceutical Processing and Process Analytical Technology*. 2023, CRC Press. p. 449-475.
 49. Narayanan, H., et al., *Hybrid modeling for biopharmaceutical processes: advantages, opportunities, and implementation*. Frontiers in Chemical Engineering, 2023. **5**.
 50. Saltelli, A., et al., *Global sensitivity analysis: the primer*. 2008: John Wiley & Sons.
 51. Kotidis, P., et al., *Constrained global sensitivity analysis for bioprocess design space identification*. Computers & Chemical Engineering, 2019. **125**: p. 558-568.
 52. Rogers, A. and M. Ierapetritou, *Feasibility and flexibility analysis of black-box processes part 2: Surrogate-based flexibility analysis*. Chem. Eng. Sci., 2015. **137**: p. 1005-1013.
 53. Cooney, R.B., S.D. Jones, and L.L. Howard, *Quality By Design for Monoclonal Antibodies, Part 1*. Bioprocess Int., 2016. **14**(6): p. 28-35.
 54. Cooney, R.B., S.D. Jones, and L.L. Howard, *Quality By Design for Monoclonal Antibodies, Part 2*. Bioprocess Int., 2016. **14**(8): p. 24-33.
 55. Cataldo, A.L., et al., *Economics and ecology: Modelling of continuous primary recovery and capture scenarios for recombinant antibody production*. J Biotechnol, 2020. **308**: p. 87-95.
 56. Chen, Y., et al., *Optimization of key energy and performance metrics for drug product manufacturing*. Int. J. Pharm., 2023. **631**: p. 122487.
 57. Yang, O., M. Qadan, and M. Ierapetritou, *Economic Analysis of Batch and Continuous Biopharmaceutical Antibody Production: A Review*. J. Pharm. Innov., 2019. **14**: p. 1-19.
 58. Emara, Y., et al., *Life cycle management in the pharmaceutical industry using an applicable and robust LCA-based environmental sustainability assessment*

- approach. Designing Sustainable Technologies, Products and Policies: From Science to Innovation*, 2018: p. 79-88.
59. Pollock, J., S.V. Ho, and S.S. Farid, *Fed-batch and perfusion culture processes: economic, environmental, and operational feasibility under uncertainty*. *Biotechnol. Bioeng.*, 2013. **110**(1): p. 206-19.
 60. Wang, Z. and M. Ierapetritou, *Applications of optimization in the pharmaceutical process development*, in *How to Design and Implement Powder-To-Tablet Continuous Manufacturing Systems*. 2022. p. 271-299.
 61. Chopda, V., et al., *Recent advances in integrated process analytical techniques, modeling, and control strategies to enable continuous biomanufacturing of monoclonal antibodies*. *Journal of Chemical Technology & Biotechnology*, 2021. **97**(9): p. 2317-2335.
 62. Flevaris, K. and C. Chatzidoukas, *Facilitating the industrial transition to microbial and microalgal factories through mechanistic modelling within the Industry 4.0 paradigm*. *Current Opinion in Chemical Engineering*, 2021. **33**.
 63. Wang, Z., C. Wang, and G. Chen, *Kinetic modeling: A tool for temperature shift and feeding optimization in cell culture process development*. *Protein Expression and Purification*, 2022. **198**: p. 106130.
 64. 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.
 65. Kotidis, P., et al., *Model - based optimization of antibody galactosylation in CHO cell culture*. *Biotechnology and bioengineering*, 2019. **116**(7): p. 1612-1626.
 66. Alhuthali, S. and C. Kontoravdi, *Population balance modelling captures host cell protein dynamics in CHO cell cultures*. *PloS one*, 2022. **17**(3): p. e0265886.
 67. Llaneras, F. and J. Picó, *Stoichiometric modelling of cell metabolism*. *Journal of bioscience and bioengineering*, 2008. **105**(1): p. 1-11.
 68. 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**(Pt A): p. 9-20.
 69. Martínez, V.S., et al., *Dynamic metabolic flux analysis using B-splines to study the effects of temperature shift on CHO cell metabolism*. *Metabolic engineering communications*, 2015. **2**: p. 46-57.
 70. Brunner, M., et al., *Elevated pCO₂ affects the lactate metabolic shift in CHO cell culture processes*. *Engineering in life sciences*, 2018. **18**(3): p. 204-214.
 71. Leighty, R.W. and M.R. Antoniewicz, *Dynamic metabolic flux analysis (DMFA): a framework for determining fluxes at metabolic non-steady state*. *Metabolic engineering*, 2011. **13**(6): p. 745-755.
 72. Ahn, W.S. and M.R. Antoniewicz, *Towards dynamic metabolic flux analysis in CHO cell cultures*. *Biotechnology journal*, 2012. **7**(1): p. 61-74.

73. Nolan, R.P. and K. Lee, *Dynamic model of CHO cell metabolism*. Metabolic engineering, 2011. **13**(1): p. 108-124.
74. Kumar, V. and A.M. Lenhoff, *Mechanistic Modeling of Preparative Column Chromatography for Biotherapeutics*. Annu. Rev. Chem. Biomol. Eng., 2020. **11**: p. 235-255.
75. Gerstweiler, L., J. Bi, and A.P.J. Middelberg, *Continuous downstream bioprocessing for intensified manufacture of biopharmaceuticals and antibodies*. Chemical Engineering Science, 2021. **231**.
76. Gao, Z.-Y., et al., *Antibody capture with twin-column continuous chromatography: Effects of residence time, protein concentration and resin*. Separation and Purification Technology, 2020. **253**.
77. Shi, C., et al., *Model-based process development of continuous chromatography for antibody capture: A case study with twin-column system*. J. Chromatogr. A, 2020. **1619**: p. 460936.
78. Kumar, V., et al., *Robust mechanistic modeling of protein ion-exchange chromatography*. J Chromatogr A, 2021. **1660**: p. 462669.
79. Andris, S. and J. Hubbuch, *Modeling of hydrophobic interaction chromatography for the separation of antibody-drug conjugates and its application towards quality by design*. J Biotechnol, 2020. **317**: p. 48-58.
80. Halan, V., et al., *Multimodal chromatography for purification of biotherapeutics—a review*. Current Protein and Peptide Science, 2019. **20**(1): p. 4-13.
81. Baur, D., et al., *Optimal model-based design of the twin-column CaptureSMB process improves capacity utilization and productivity in protein A affinity capture*. Biotechnol. J., 2016. **11**(1): p. 135-45.
82. Khanal, O., et al., *Multi-column displacement chromatography for separation of charge variants of monoclonal antibodies*. J. Chromatogr. A, 2019. **1586**: p. 40-51.
83. Piątkowski, W., D. Antos, and K. Kaczmarski, *Modeling of preparative chromatography processes with slow intraparticle mass transport kinetics*. Journal of Chromatography A, 2003. **988**(2): p. 219-231.
84. Püttmann, A., et al., *Fast and accurate parameter sensitivities for the general rate model of column liquid chromatography*. Comput. Chem. Eng., 2013. **56**: p. 46-57.
85. Shekhawat, L.K. and A.S. Rathore, *An overview of mechanistic modeling of liquid chromatography*. Preparative biochemistry and biotechnology, 2019. **49**(6): p. 623-638.
86. Shekhawat, K.L. and S.A. Rathore, *An overview of mechanistic modeling of liquid chromatography*. Prep. Biochem. Biotechnol., 2019. **49**(6): p. 623-638.
87. Chen, Y.-C., S.-J. Yao, and D.-Q. Lin, *Parameter-by-parameter method for steric mass action model of ion exchange chromatography: Theoretical considerations and experimental verification*. Journal of Chromatography A, 2022. **1680**: p. 463418.

88. Dimartino, S., C. Boi, and G.C. Sarti, *Influence of protein adsorption kinetics on breakthrough broadening in membrane affinity chromatography*. J Chromatogr A, 2011. **1218**(26): p. 3966-72.
89. Ding, C. and M. Ierapetritou, *A novel framework of surrogate-based feasibility analysis for establishing design space of twin-column continuous chromatography*. Int. J. Pharm., 2021. **609**: p. 121161.
90. He, Q.-L., et al., *Model-based process design of a ternary protein separation using multi-step gradient ion-exchange SMB chromatography*. Comput. Chem. Eng., 2020. **138**.
91. Osberghaus, A., et al., *Optimizing a chromatographic three component separation: a comparison of mechanistic and empiric modeling approaches*. J Chromatogr A, 2012. **1237**: p. 86-95.
92. Bhojar, S., et al., *Predictive mechanistic modeling of loading and elution in protein A chromatography*. J Chromatogr A, 2024. **1713**: p. 464558.
93. Cebulla, D.H., et al., *Model - based optimization of an ion exchange chromatography process for the separation of von Willebrand factor fragments and human serum albumin*. Pamm, 2023: p. e202300027.
94. Umatheva, U., G. Chen, and R. Ghosh, *Computational fluid dynamic (CFD) simulation of a cuboid packed-bed chromatography device*. Chem. Eng. Res. Des., 2019. **152**: p. 393-401.
95. Schmidt, I., et al., *Estimation of Chromatographic Columns Performances using Computer Tomography and CFD Simulations*. Chemie Ingenieur Technik, 2011. **83**(1-2): p. 130-142.
96. Leweke, S. and E. von Lieres, *Chromatography Analysis and Design Toolkit (CADET)*. Computers & Chemical Engineering, 2018. **113**: p. 274-294.
97. CADET, *Two Dimensional General rate model (GRM2D)*. 2022: https://cadet.github.io/master/modelling/unit_operations/2d_general_rate_model.html#d-general-rate-model-model.
98. Erickson, J., et al., *End-to-end collaboration to transform biopharmaceutical development and manufacturing*. Biotechnol Bioeng, 2021. **118**(9): p. 3302-3312.
99. Coolbaugh, M.J., et al., *Pilot-scale demonstration of an end-to-end integrated and continuous biomanufacturing process*. Biotechnol Bioeng, 2021. **118**(9): p. 3287-3301.
100. Namila, N., *The Effects of Solution Condition on Virus Filtration Performance*. 2020: University of Arkansas.
101. Rathore, A.S., et al., *Mechanistic modeling of viral filtration*. Journal of Membrane Science, 2014. **458**: p. 96-103.
102. Suh, D., et al., *Virus filtration in biopharmaceutical downstream processes: key factors and current limitations*. Separation & Purification Reviews, 2022: p. 1-14.

103. Hadpe, S.R., et al., *ATF for cell culture harvest clarification: mechanistic modelling and comparison with TFF*. Journal of Chemical Technology & Biotechnology, 2017. **92**(4): p. 732-740.
104. Huter, M.J. and J. Strube, *Model-based design and process optimization of continuous single pass tangential flow filtration focusing on continuous bioprocessing*. Processes, 2019. **7**(6): p. 317.
105. Grote, F., H. Froehlich, and J. Strube, *Integration of ultrafiltration unit operations in biotechnology process design*. Chemical Engineering & Technology, 2011. **34**(5): p. 673-687.
106. Binabaji, E., et al., *Ultrafiltration of highly concentrated antibody solutions: Experiments and modeling for the effects of module and buffer conditions*. Biotechnology progress, 2016. **32**(3): p. 692-701.
107. Binabaji, E., et al., *Theoretical analysis of the ultrafiltration behavior of highly concentrated protein solutions*. Journal of membrane science, 2015. **494**: p. 216-223.
108. Bhosekar, A. and M. Ierapetritou, *Advances in surrogate based modeling, feasibility analysis, and optimization: A review*. Comput. Chem. Eng., 2018. **108**: p. 250-267.
109. Reis, M.S. and P.M. Saraiva, *Data-Driven Process System Engineering—Contributions to its consolidation following the path laid down by George Stephanopoulos*. Computers & Chemical Engineering, 2022. **159**: p. 107675.
110. Xie, Y., *Data-driven predictive modeling for cell line selection in biopharmaceutical production*. 2019, Massachusetts Institute of Technology.
111. Bashokouh, F., S. Abbasiliasi, and J.S. Tan, *Optimization of cultivation conditions for monoclonal IgM antibody production by M1A2 hybridoma using artificial neural network*. Cytotechnology, 2019. **71**(4): p. 849-860.
112. Alavijeh, M.K., et al., *Digitally enabled approaches for the scale up of mammalian cell bioreactors*. Digital Chemical Engineering, 2022: p. 100040.
113. Rathore, A.S., et al., *Chemometrics application in biotech processes: assessing comparability across processes and scales*. Journal of Chemical Technology & Biotechnology, 2014. **89**(9): p. 1311-1316.
114. Sokolov, M., et al., *Enhanced process understanding and multivariate prediction of the relationship between cell culture process and monoclonal antibody quality*. Biotechnology progress, 2017. **33**(5): p. 1368-1380.
115. Green, A. and J. Glassey, *Multivariate analysis of the effect of operating conditions on hybridoma cell metabolism and glycosylation of produced antibody*. Journal of Chemical Technology & Biotechnology, 2015. **90**(2): p. 303-313.
116. Tsang, V.L., et al., *Development of a scale down cell culture model using multivariate analysis as a qualification tool*. Biotechnology progress, 2014. **30**(1): p. 152-160.

117. Japel, R.C. and J.F. Buyel, *Bayesian optimization using multiple directional objective functions allows the rapid inverse fitting of parameters for chromatography simulations*. J Chromatogr A, 2022. **1679**: p. 463408.
118. Nikita, S., et al., *Reinforcement learning based optimization of process chromatography for continuous processing of biopharmaceuticals*. Chemical Engineering Science, 2021. **230**.
119. Wang, G., et al., *Estimation of adsorption isotherm and mass transfer parameters in protein chromatography using artificial neural networks*. J. Chromatogr. A, 2017. **1487**: p. 211-217.
120. Li, S., et al., *Using surrogate models for efficient optimization of simulated moving bed chromatography*. Computers & Chemical Engineering, 2014. **67**: p. 121-132.
121. Sachio, S., et al., *Computer-aided design space identification for screening of protein A affinity chromatography resins*. J Chromatogr A, 2024. **1722**: p. 464890.
122. Sachio, S., C. Kontoravdi, and M.M. Papathanasiou, *A model-based approach towards accelerated process development: A case study on chromatography*. Chemical Engineering Research and Design, 2023. **197**: p. 800-820.
123. Binger, Z.M. and A. Achilli, *Surrogate modeling of pressure loss & mass transfer in membrane channels via coupling of computational fluid dynamics and machine learning*. Desalination, 2023. **548**: p. 116241.
124. Zhang, B., et al., *Backwash sequence optimization of a pilot-scale ultrafiltration membrane system using data-driven modeling for parameter forecasting*. Journal of Membrane Science, 2020. **612**: p. 118464.
125. Von Stosch, M., et al., *Hybrid semi-parametric modeling in process systems engineering: Past, present and future*. Computers & Chemical Engineering, 2014. **60**: p. 86-101.
126. Chen, Y. and M. Ierapetritou, *A framework of hybrid model development with identification of plant - model mismatch*. AIChE J, 2020. **66**(10): p. e16996.
127. Tsopanoglou, A. and I.J. del Val, *Moving towards an era of hybrid modelling: advantages and challenges of coupling mechanistic and data-driven models for upstream pharmaceutical bioprocesses*. Current Opinion in Chemical Engineering, 2021. **32**: p. 100691.
128. Narayanan, H., et al., *A new generation of predictive models: the added value of hybrid models for manufacturing processes of therapeutic proteins*. Biotechnology and Bioengineering, 2019. **116**(10): p. 2540-2549.
129. 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.
130. Kotidis, P. and C. Kontoravdi, *Harnessing the potential of artificial neural networks for predicting protein glycosylation*. Metabolic engineering communications, 2020. **10**: p. e00131.

131. Möller, J., et al., *Model-assisted design of experiments as a concept for knowledge-based bioprocess development*. Bioprocess and biosystems engineering, 2019. **42**(5): p. 867-882.
132. 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.
133. Schinn, S.M., et al., *A genome - scale metabolic network model and machine learning predict amino acid concentrations in Chinese Hamster Ovary cell cultures*. Biotechnology and Bioengineering, 2021. **118**(5): p. 2118-2123.
134. Antonakoudis, A., et al., *Synergising stoichiometric modelling with artificial neural networks to predict antibody glycosylation patterns in Chinese hamster ovary cells*. Computers & Chemical Engineering, 2021. **154**: p. 107471.
135. Wang, G., et al., *Root cause investigation of deviations in protein chromatography based on mechanistic models and artificial neural networks*. J Chromatogr A, 2017. **1515**: p. 146-153.
136. Narayanan, H., et al., *Hybrid Models for the simulation and prediction of chromatographic processes for protein capture*. Journal of Chromatography A, 2021. **1650**: p. 462248.
137. Narayanan, H., et al., *Hybrid Models Based on Machine Learning and an Increasing Degree of Process Knowledge: Application to Capture Chromatographic Step*. Ind. Eng. Chem. Res., 2021. **60**(29): p. 10466-10478.
138. Feidl, F., et al., *Model based strategies towards protein A resin lifetime optimization and supervision*. J. Chromatogr. A, 2020. **1625**: p. 461261.
139. Tang, S.Y., et al., *Physics-informed neural networks to solve lumped kinetic model for chromatography process*. J Chromatogr A, 2023. **1708**: p. 464346.
140. Krippel, M., A. Dürauer, and M. Duerkop, *Hybrid modeling of cross-flow filtration: Predicting the flux evolution and duration of ultrafiltration processes*. Separation and Purification Technology, 2020. **248**: p. 117064.
141. Thiess, H., et al., *Module design for ultrafiltration in biotechnology: Hydraulic analysis and statistical modeling*. Journal of Membrane Science, 2017. **540**: p. 440-453.
142. Metta, N., et al., *Dynamic flowsheet model development and sensitivity analysis of a continuous pharmaceutical tablet manufacturing process using the wet granulation route*. Processes, 2019. **7**(4): p. 234.
143. Petrides, D., et al., *Biopharmaceutical Process Optimization with Simulation and Scheduling Tools*. Bioengineering (Basel), 2014. **1**(4): p. 154-187.
144. Gomis-Fons, J., et al., *Model-based design and control of a small-scale integrated continuous end-to-end mAb platform*. Biotechnol. Prog., 2020. **36**(4): p. e2995.
145. Liu, S. and L.G. Papageorgiou, *Optimal Antibody Purification Strategies Using Data-Driven Models*. Engineering, 2019. **5**(6): p. 1077-1092.

146. Zahel, T., et al., *Integrated process modeling—a process validation life cycle companion*. Bioengineering, 2017. **4**(4): p. 86.
147. Pirrung, S.M., et al., *Optimization of biopharmaceutical downstream processes supported by mechanistic models and artificial neural networks*. Biotechnology Progress, 2017. **33**(3): p. 696-707.
148. Sencar, J., N. Hammerschmidt, and A. Jungbauer, *Modeling the Residence Time Distribution of Integrated Continuous Bioprocesses*. Biotechnol. J., 2020. **15**(8): p. e2000008.
149. Rischawy, F., et al., *Connected mechanistic process modeling to predict a commercial biopharmaceutical downstream process*. Comput. Chem. Eng., 2023. **176**.
150. Yang, O., S. Prabhu, and M. Ierapetritou, *Comparison between Batch and Continuous Monoclonal Antibody Production and Economic Analysis*. Ind. Eng. Chem. Res., 2019. **58**(15): p. 5851-5863.
151. Taras, S. and A. Woinaroschy, *Simulation and multi-objective optimization of bioprocesses with Matlab and SuperPro Designer using a client–server interface*. Chem. Eng. Trans., 2011. **25**: p. 207-212.
152. Shi, C., et al., *Model-assisted process design for better evaluation and scaling up of continuous downstream bioprocessing*. Journal of Chromatography A, 2022. **1683**.
153. Fisher, A.C., et al., *The Current Scientific and Regulatory Landscape in Advancing Integrated Continuous Biopharmaceutical Manufacturing*. Trends Biotechnol, 2019. **37**(3): p. 253-267.
154. Godawat, R., et al., *End-to-end integrated fully continuous production of recombinant monoclonal antibodies*. J. Biotechnol., 2015. **213**: p. 13-19.
155. Ding, C., et al., *Process design of a fully integrated continuous biopharmaceutical process using economic and ecological impact assessment*. Biotechnol. Bioeng., 2022. **119**(12): p. 3567-3583.
156. Gupta, P., et al., *Economic assessment of continuous processing for manufacturing of biotherapeutics*. Biotechnol Prog, 2021. **37**(2): p. e3108.
157. Xenopoulos, A., *A new, integrated, continuous purification process template for monoclonal antibodies: process modeling and cost of goods studies*. Journal of biotechnology, 2015. **213**: p. 42-53.
158. Badr, S., et al., *Integrated design of biopharmaceutical manufacturing processes: Operation modes and process configurations for monoclonal antibody production*. Computers & Chemical Engineering, 2021. **153**.
159. FoodDrugAdministration, *Executive Order on Advancing Biotechnology and Biomanufacturing Innovation for a Sustainable, Safe, and Secure American Bioeconomy*. 2022: <https://www.whitehouse.gov/briefing-room/presidential-actions/2022/09/12/executive-order-on-advancing-biotechnology-and-biomanufacturing-innovation-for-a-sustainable-safe-and-secure-american-bioeconomy/>.

160. Huang, Q., *Intelligent manufacturing*, in *Understanding China's Manufacturing Industry*. 2022, Springer. p. 111-127.
161. Teixeira, J.E. and A.T.C. Tavares-Lehmann, *Industry 4.0 in the European union: Policies and national strategies*. Technological Forecasting and Social Change, 2022. **180**: p. 121664.
162. Vogg, S., T. Muller-Spath, and M. Morbidelli, *Design space and robustness analysis of batch and counter-current frontal chromatography processes for the removal of antibody aggregates*. J. Chromatogr. A, 2020: p. 460943.
163. Steinebach, F., et al., *Equilibrium Theory Based Design Space for the Multicolumn Countercurrent Solvent Gradient Purification Process*. Ind. Eng. Chem. Res., 2017. **56**(45): p. 13482-13489.
164. Zydney, A.L., *Perspectives on integrated continuous bioprocessing—opportunities and challenges*. Curr. Opin. Chem. Eng., 2015. **10**: p. 8-13.
165. Badman, C., et al., *Why We Need Continuous Pharmaceutical Manufacturing and How to Make It Happen*. J Pharm Sci, 2019. **108**(11): p. 3521-3523.
166. Rathore, A.S. and F. Shereef, *The influence of domestic manufacturing capabilities on biologic pricing in emerging economies*. Nat Biotechnol, 2019. **37**(5): p. 498-501.
167. Fletcher, N., *Turn batch to continuous processing*. Manuf Chem, 2010. **81**: p. 24-26.
168. Shukla, A.A., et al., *Evolving trends in mAb production processes*. Bioeng Transl Med, 2017. **2**(1): p. 58-69.
169. Amit K. Dutta, J.T., and Boris Napadensky, *Performance Optimization of Continuous Countercurrent Tangential Chromatography for Antibody Capture*. Biotechnol. Prog., 2016. **32**(2).
170. Gomis-Fons, J., et al., *Optimal loading flow rate trajectory in monoclonal antibody capture chromatography*. J Chromatogr A, 2021. **1635**: p. 461760.
171. Croughan, M.S., K.B. Konstantinov, and C. Cooney, *The future of industrial bioprocessing: batch or continuous?* Biotechnol. Bioeng., 2015. **112**(4): p. 648-51.
172. Angarita, M., et al., *Twin-column CaptureSMB: A novel cyclic process for protein A affinity chromatography*. J. Chromatogr. A, 2015. **1389**: p. 85-95.
173. Guo, J., M. Jin, and D. Kanani, *Optimization of Single-Column Batch and Multicolumn Continuous Protein A Chromatography and Performance Comparison Based on Mechanistic Model*. Biotechnol J, 2020: p. e2000192.
174. Sun, Y.N., et al., *Model-based process development and evaluation of twin-column continuous capture processes with Protein A affinity resin*. J. Chromatogr. A, 2020. **1625**: p. 461300.
175. Godawat, R., et al., *Periodic counter-current chromatography – design and operational considerations for integrated and continuous purification of proteins*. Biotechnol. J., 2012. **7**: p. 1496–1508.

176. Pollock, J., et al., *Optimising the design and operation of semi-continuous affinity chromatography for clinical and commercial manufacture*. J. Chromatogr. A, 2013. **1284**: p. 17-27.
177. Shi, C., et al., *Process development and optimization of continuous capture with three-column periodic counter-current chromatography*. Biotechnol Bioeng, 2021.
178. Gjoka, X., R. Gantier, and M. Schofield, *Transfer of a three step mAb chromatography process from batch to continuous: Optimizing productivity to minimize consumable requirements*. J. Biotechnol., 2017. **242**: p. 11-18.
179. Ng, C.K.S., et al., *Design of high productivity sequential multi-column chromatography for antibody capture*. Food Bioprod. Process, 2014. **92**(2): p. 233-241.
180. St Amand, M.M., et al., *Identifying a robust design space for glycosylation during monoclonal antibody production*. Biotechnol Prog, 2016. **32**(5): p. 1149-1162.
181. Winkle, A.S.R.H., *Quality by design for biopharmaceuticals*. Nature Biotechnology, 2009. **27**: p. 26.
182. Rathore, A.S., *QbD/PAT for bioprocessing: moving from theory to implementation*. Current Opinion in Chemical Engineering, 2014. **6**: p. 1-8.
183. Xiang, H., *Establishing Process Design Space for a Chromatography Purification Step: Application of Quality-by-Design Principles*. BioPharm. International, 2015. **28**(11).
184. Wang, Z., M.S. Escotet-Espinoza, and M. Ierapetritou, *Process analysis and optimization of continuous pharmaceutical manufacturing using flowsheet models*. Comput. Chem. Eng., 2017. **107**: p. 77-91.
185. Wang, Z. and M. Ierapetritou, *A novel feasibility analysis method for black-box processes using a radial basis function adaptive sampling approach*. AIChE Journal, 2017. **63**(2): p. 532-550.
186. Gomis-Fons, J., N. Andersson, and B. Nilsson, *Optimization study on periodic counter-current chromatography integrated in a monoclonal antibody downstream process*. J. Chromatogr. A, 2020. **1621**: p. 461055.
187. Zhu, M. and G. Carta, *Protein adsorption equilibrium and kinetics in multimodal cation exchange resins*. Adsorption, 2015. **22**(2): p. 165-179.
188. Ng, C.K., et al., *Design of high productivity antibody capture by protein A chromatography using an integrated experimental and modeling approach*. J. Chromatogr. B, 2012. **899**: p. 116-26.
189. Yang, O. and M. Ierapetritou, *mAb Production Modeling and Design Space Evaluation Including Glycosylation Process*. Processes, 2021. **9**(2): p. 324.
190. Hengl, T., G.B.M. Heuvelink, and D.G. Rossiter, *About regression-kriging: From equations to case studies*. Computers & Geosciences, 2007. **33**(10): p. 1301-1315.

191. Hengl, T., G.B.M. Heuvelink, and A. Stein, *A generic framework for spatial prediction of soil variables based on regression-kriging*. *Geoderma*, 2004. **120**(1-2): p. 75-93.
192. LIN, Q.Z.a.H.S., *Comparing Ordinary Kriging and Regression Kriging for Soil Properties in Contrasting Landscapes*. *Pedosphere*, 2010. **20**(5): p. 594-606.
193. Qian, J., et al., *A sequential constraints updating approach for Kriging surrogate model-assisted engineering optimization design problem*. *Engineering with Computers*, 2019. **36**(3): p. 993-1009.
194. Lophaven, S.N., H.B. Nielsen, and J. Søndergaard, *DACE—a Matlab Kriging toolbox; version 2; informatics and mathematical modelling*. Technical University of Denmark, Technical Report, 2002. **No. IMM-TR-2002-12**.
195. Wang, Z. and M. Ierapetritou, *Surrogate-based feasibility analysis for black-box stochastic simulations with heteroscedastic noise*. *J. Glob. Optim.*, 2018. **71**(4): p. 957-985.
196. Benedini, L.J., et al., *Modeling and simulation of anion exchange chromatography for purification of proteins in complex mixtures*. *J Chromatogr A*, 2020. **1613**: p. 460685.
197. Sun, Y.N., et al., *Model-assisted process development, characterization and design of continuous chromatography for antibody separation*. *J Chromatogr A*, 2023. **1707**: p. 464302.
198. Huuk, T.C., et al., *Model-based integrated optimization and evaluation of a multi-step ion exchange chromatography*. *Sep. Purif. Technol.*, 2014. **136**: p. 207-222.
199. Ding, C., C. Gerberich, and M. Ierapetritou, *Hybrid Model Development for Parameter Estimation and Process Optimization of Hydrophobic Interaction Chromatography*. *J Chromatogr A*, 2023. **1703**: p. 464113.
200. Rathore, A.S., et al., *Artificial intelligence and machine learning applications in biopharmaceutical manufacturing*. *Trends Biotechnol*, 2023. **41**(4): p. 497-510.
201. Daoutidis, P., et al., *Machine learning in process systems engineering: Challenges and opportunities*. *Comput. Chem. Eng.*, 2024. **181**.
202. Loonen, R.C., S. de Vries, and F. Goia, *Inverse design for advanced building envelope materials, systems and operation*, in *Rethinking Building Skins*. 2022, Elsevier. p. 377-402.
203. Osberghaus, A., et al., *Determination of parameters for the steric mass action model—a comparison between two approaches*. *J Chromatogr A*, 2012. **1233**: p. 54-65.
204. Osberghaus, A., et al., *Model-integrated process development demonstrated on the optimization of a robotic cation exchange step*. *Chem. Eng. Sci.*, 2012. **76**: p. 129-139.

205. Chen, G., A. Gerrior, and R. Ghosh, *Feasibility study for high-resolution multi-component separation of protein mixture using a cation-exchange cuboid packed-bed device*. J Chromatogr A, 2018. **1549**: p. 25-30.
206. von Lieres, E. and J. Andersson, *A fast and accurate solver for the general rate model of column liquid chromatography*. Comput. Chem. Eng., 2010. **34**(8): p. 1180-1191.
207. Boukouvala, F. and M.G. Ierapetritou, *Feasibility analysis of black-box processes using an adaptive sampling Kriging-based method*. Computers & Chemical Engineering, 2012. **36**: p. 358-368.
208. Jones, D.R., M. Schonlau, and W.J. Welch, *Efficient global optimization of expensive black-box functions*. Journal of Global optimization, 1998. **13**: p. 455-492.
209. Moćkus, J. *On Bayesian methods for seeking the extremum*. in *Optimization Techniques IFIP Technical Conference: Novosibirsk, July 1-7, 1974*. 1975. Springer.
210. Wang, Z., et al., *Surrogate-based Optimization for Pharmaceutical Manufacturing Processes*, in *27th European Symposium on Computer Aided Process Engineering*. 2017. p. 2797-2802.
211. Boukouvala, F. and M.G. Ierapetritou, *Derivative -free optimization for expensive constrained problems using a novel expected improvement objective function*. AIChE Journal, 2014. **60**(7): p. 2462-2474.
212. Shekhawat, L.K., M. Chandak, and A.S. Rathore, *Mechanistic modeling of hydrophobic interaction chromatography for monoclonal antibody purification: process optimization in the quality by design paradigm*. Journal of Chemical Technology & Biotechnology, 2017. **92**(10): p. 2527-2537.
213. Wasalathanthri, D.P., et al., *Technology outlook for real-time quality attribute and process parameter monitoring in biopharmaceutical development-A review*. Biotechnol. Bioeng., 2020. **117**(10): p. 3182-3198.
214. McCue, J.T., et al., *Modeling of protein monomer/aggregate purification and separation using hydrophobic interaction chromatography*. Bioprocess Biosyst Eng, 2008. **31**(3): p. 261-75.
215. Jandera, P. and P. Janás, *Recent advances in stationary phases and understanding of retention in hydrophilic interaction chromatography. A review*. Analytica Chimica Acta, 2017. **967**: p. 12-32.
216. McCalley, D.V., *Study of retention and peak shape in hydrophilic interaction chromatography over a wide pH range*. Journal of chromatography A, 2015. **1411**: p. 41-49.
217. Baumann, P., K. Baumgartner, and J. Hubbuch, *Influence of binding pH and protein solubility on the dynamic binding capacity in hydrophobic interaction chromatography*. J Chromatogr A, 2015. **1396**: p. 77-85.
218. Barrientos, R.C., et al., *Automated Hydrophobic Interaction Chromatography Screening Combined with In Silico Optimization as a Framework for*

- Nondenaturing Analysis and Purification of Biopharmaceuticals*. Analytical Chemistry, 2022. **94**(49): p. 17131-17141.
219. Wang, G., T. Hahn, and J. Hubbuch, *Water on hydrophobic surfaces: Mechanistic modeling of hydrophobic interaction chromatography*. J Chromatogr A, 2016. **1465**: p. 71-8.
 220. Xia, F., et al., *Evaluation of selectivity changes in HIC systems using a preferential interaction based analysis*. Biotechnol Bioeng, 2004. **87**(3): p. 354-63.
 221. Fekete, S., et al., *Hydrophobic interaction chromatography for the characterization of monoclonal antibodies and related products*. J Pharm Biomed Anal, 2016. **130**: p. 3-18.
 222. Creasy, A., et al., *Gradient elution behavior of proteins in hydrophobic interaction chromatography with U-shaped retention factor curves*. J Chromatogr A, 2018. **1547**: p. 53-61.
 223. Lietta, E., et al., *An Experimental and Modeling Combined Approach in Preparative Hydrophobic Interaction Chromatography*. Processes, 2022. **10**(5).
 224. Takahashi, M.B., et al., *Artificial neural network associated to UV/Vis spectroscopy for monitoring bioreactions in biopharmaceutical processes*. Bioprocess and biosystems engineering, 2015. **38**: p. 1045-1054.
 225. Kumar, V. and A. Rathore, *Mechanistic Modeling of Preparative Ion-Exchange Chromatography*. BioPharm International, 2017. **30**(4).
 226. Carta, G. and A. Jungbauer, *Protein chromatography: process development and scale-up*. 2020: John Wiley & Sons.
 227. Guiochon, G., A. Felinger, and D.G. Shirazi, *Fundamentals of preparative and nonlinear chromatography*. 2006: Elsevier.
 228. Luo, Y. and M. Ierapetritou, *Multifeedstock and Multiproduct Process Design Using Neural Network Surrogate Flexibility Constraints*. Industrial & Engineering Chemistry Research, 2023. **62**(5): p. 2067-2079.
 229. Skeel, R.D.a.M.B., *A Method for the Spatial Discretization of Parabolic Equations in One Space Variable*. SIAM Journal on Scientific and Statistical Computing, 1990. **11**: p. 1-32.
 230. Georgioudakis, M. and V. Plevris, *A Comparative Study of Differential Evolution Variants in Constrained Structural Optimization*. Frontiers in Built Environment, 2020. **6**.
 231. Piotrowski, A.P., *L-SHADE optimization algorithms with population-wide inertia*. Information Sciences, 2018. **468**: p. 117-141.
 232. Moles, C.G., P. Mendes, and J.R. Banga, *Parameter estimation in biochemical pathways: a comparison of global optimization methods*. Genome Res, 2003. **13**(11): p. 2467-74.
 233. Zakaria, A., et al., *Uncertainty models for stochastic optimization in renewable energy applications*. Renewable Energy, 2020. **145**: p. 1543-1571.

234. Zheng, Q.P., J. Wang, and A.L. Liu, *Stochastic Optimization for Unit Commitment—A Review*. IEEE Transactions on Power Systems, 2015. **30**(4): p. 1913-1924.
235. Kawajiri, Y., *Model-based optimization strategies for chromatographic processes: a review*. Adsorption, 2020.
236. Kamga, M.H., M. Cattaneo, and S. Yoon, *Integrated continuous biomanufacturing platform with ATF perfusion and one column chromatography operation for optimum resin utilization and productivity*. Prep. Biochem. Biotechnol., 2018. **48**(5): p. 383-390.
237. Wang, S., et al., *Shear contributions to cell culture performance and product recovery in ATF and TFF perfusion systems*. Journal of biotechnology, 2017. **246**: p. 52-60.
238. Walther, J., et al., *Perfusion cell culture decreases process and product heterogeneity in a head - to - head comparison with fed - batch*. Biotechnology journal, 2019. **14**(2): p. 1700733.
239. 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.
240. Xu, S., et al., *Bioreactor productivity and media cost comparison for different intensified cell culture processes*. Biotechnol Prog, 2017. **33**(4): p. 867-878.
241. Martins, D.L., et al., *Truly continuous low pH viral inactivation for biopharmaceutical process integration*. Biotechnol Bioeng, 2020. **117**(5): p. 1406-1417.
242. Gillespie, C., et al., *Continuous In-Line Virus Inactivation for Next Generation Bioprocessing*. Biotechnol J, 2019. **14**(2): p. e1700718.
243. Jabra, M.G., C.J. Yehl, and A.L. Zydney, *Multistage continuous countercurrent diafiltration for formulation of monoclonal antibodies*. Biotechnol Prog, 2019. **35**(4): p. e2810.
244. Brinkmann, A., et al., *Leveraging single - pass tangential flow filtration to enable decoupling of upstream and downstream monoclonal antibody processing*. Biotechnology progress, 2018. **34**(2): p. 405-411.
245. Konstantinov, K.B. and C.L. Cooney, *White Paper on Continuous Bioprocessing May 20–21 2014 Continuous Manufacturing Symposium*. J. Pharm. Sci., 2015. **104**(3): p. 813-820.
246. Kruse, T., et al., *An alternative downstream process based on aqueous two-phase extraction for the purification of monoclonal antibodies*. Biochem. Eng. J., 2020. **161**.
247. Karst, D.J., Steinebach, F., Soos, M., Morbidelli, M. , *Process performance and product quality in an integrated continuous antibody production process*. Biotechnol. Bioeng., 2017. **114**(2): p. 298-307.

248. Feidl, F., et al., *Process-wide control and automation of an integrated continuous manufacturing platform for antibodies*. *Biotechnol. Bioeng.*, 2020. **117**(5): p. 1367-1380.
249. Coffman, J., et al., *The Design Basis for the Integrated and Continuous Biomanufacturing Framework*. *Biotechnol Bioeng*, 2021.
250. Lalor, F., et al., *Sustainability in the biopharmaceutical industry: Seeking a holistic perspective*. *Biotechnol Adv*, 2019. **37**(5): p. 698-707.
251. Ötes, O., et al., *Moving to CoPACaPAnA: implementation of a continuous protein A capture process for antibody applications within an end-to-end single-use GMP manufacturing downstream process*. *Biotechnology Reports*, 2020. **26**: p. e00465.
252. Walther, J., et al., *The business impact of an integrated continuous biomanufacturing platform for recombinant protein production*. *J Biotechnol*, 2015. **213**: p. 3-12.
253. 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.
254. Arnold, L., et al., *Implementation of Fully Integrated Continuous Antibody Processing: Effects on Productivity and COGm*. *Biotechnol. J.*, 2019. **14**(2): p. e1800061.
255. Grilo, A.L., et al., *Monoclonal Antibodies Production Platforms: An Opportunity Study of a Non-Protein-A Chromatographic Platform Based on Process Economics*. *Biotechnol J*, 2017. **12**(12).
256. Nandi, S., et al., *Techno-economic analysis of a transient plant-based platform for monoclonal antibody production*. *MAbs*, 2016. **8**(8): p. 1456-1466.
257. Muthukumar, S., et al., *Economic benefits of membrane chromatography versus packed bed column purification of therapeutic proteins expressed in microbial and mammalian hosts*. *Journal of Chemical Technology & Biotechnology*, 2017. **92**(1): p. 59-68.
258. 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.
259. Toumi, A., et al., *Design and optimization of a large scale biopharmaceutical facility using process simulation and scheduling tools*. *Pharmaceutical Engineering*, 2010. **30**(2): p. 1-9.
260. 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.
261. Klutz, S., et al., *Cost evaluation of antibody production processes in different operation modes*. *Chem. Eng. Sci.*, 2016. **141**: p. 63-74.
262. 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.

263. Madabhushi, S.R., et al., *Quantitative assessment of environmental impact of biologics manufacturing using process mass intensity analysis*. Biotechnol Prog, 2018. **34**(6): p. 1566-1573.
264. Orr, V., et al., *Recent advances in bioprocessing application of membrane chromatography*. Biotechnol Adv, 2013. **31**(4): p. 450-65.
265. Liu, Z., S.R. Wickramasinghe, and X. Qian, *Membrane chromatography for protein purifications from ligand design to functionalization*. Separation Science and Technology, 2016. **52**(2): p. 299-319.
266. Thakur, G., et al., *Control of surge tanks for continuous manufacturing of monoclonal antibodies*. Biotechnology and Bioengineering, 2021. **118**(5): p. 1913-1931.
267. Rathore, A.S., et al., *Challenges in process control for continuous processing for production of monoclonal antibody products*. Current Opinion in Chemical Engineering, 2021. **31**: p. 100671.
268. Woodgate, J.M., *Perfusion N-1 Culture—Opportunities for Process Intensification*, in *Biopharmaceutical Processing*. 2018. p. 755-768.
269. Pollock, J.E., *Evaluating the Potential of Continuous Processes for Monoclonal Antibodies: Economic, Environmental and Operational Feasibility*. thesis, 2013.
270. Klutz, S., et al., *Developing the biofacility of the future based on continuous processing and single-use technology*. J Biotechnol, 2015. **213**: p. 120-30.
271. Budzinski, K., et al., *Introduction of a process mass intensity metric for biologics*. New biotechnology, 2019. **49**: p. 37-42.
272. Madabhushi, S.R., N.D. Pinto, and H. Lin, *Comparison of process mass intensity (PMI) of continuous and batch manufacturing processes for biologics*. New Biotechnology, 2022. **72**: p. 122-127.
273. Budzinski, K., et al., *Introduction of a process mass intensity metric for biologics*. N Biotechnol, 2019. **49**: p. 37-42.
274. Petrides, D., *Bioprocess Design and Economics*. 2015. **Chapter 11**.
275. Tressel, J.X.Z.a.T., *Basic Concepts in Q Membrane Chromatography for Large-Scale Antibody Production*. Biotechnol. Prog., 2006. **22**: p. 341–349.
276. Varadaraju, H., et al., *Process and economic evaluation for monoclonal antibody purification using a membrane-only process*. Biotechnol Prog, 2011. **27**(5): p. 1297-305.
277. Rathore, A.S., D. Kumar, and N. Kateja, *Recent developments in chromatographic purification of biopharmaceuticals*. Biotechnol Lett, 2018. **40**(6): p. 895-905.
278. Zobel-Roos, S., D. Stein, and J. Strube, *Evaluation of Continuous Membrane Chromatography Concepts with an Enhanced Process Simulation Approach*. Antibodies (Basel), 2018. **7**(1).
279. Heymann, W., et al., *Advanced error modeling and Bayesian uncertainty quantification in mechanistic liquid chromatography modeling*. J Chromatogr A, 2023. **1708**: p. 464329.

280. Briskot, T., et al., *Prediction uncertainty assessment of chromatography models using Bayesian inference*. J Chromatogr A, 2019. **1587**: p. 101-110.
281. Close, E.J., et al., *A model based approach for identifying robust operating conditions for industrial chromatography with process variability*. Chemical Engineering Science, 2014. **116**: p. 284-295.
282. Yamamoto, Y., T. Yajima, and Y. Kawajiri, *Uncertainty quantification for chromatography model parameters by Bayesian inference using sequential Monte Carlo method*. Chemical Engineering Research and Design, 2021. **175**: p. 223-237.

Appendix A

ACKNOWLEDGEMENT OF PUBLICATIONS AND COPYRIGHT PERMISSIONS

Sections of this dissertation have been published previously in scientific journals or book chapters. The following publications are acknowledged, followed by copyright permissions.

Chapter 1 includes extracts from the following published works of the author:

- Ding, C., Yang, O., Ierapetritou, M., Towards digital twin for biopharmaceutical processes: concept and progress. In: Pörtner, R. (eds) *Biopharmaceutical Manufacturing*. Cell Engineering, vol 11. Springer.
- Malinov, N., Raudenbush, K., Ding, C., Reddy, J., Ierapetritou, M., End-to-end process flowsheet modeling for biopharmaceutical production: current state and future potential. *Current Opinion in Chemical Engineering* [Under Review].

Chapter 2 includes extracts from the following published works of the author:

- Ding, C., Ierapetritou, M., A novel framework of surrogate-based feasibility analysis for establishing design space of twin-column continuous chromatography. *International Journal of Pharmaceutics*, 2021, 609: 121161.

Chapter 3 includes extracts from the following published works of the author:

- Ding, C., Ierapetritou, M., Machine learning-based optimization of a multi-step ion exchange chromatography for ternary protein separation. *Computers & Chemical Engineering*, 2024, 184:108642.

Chapter 4 includes extracts from the following published works of the author:

- Ding, C., Gerberich, C., Ierapetritou, M., Hybrid model development for parameter estimation and process optimization of hydrophobic interaction chromatography. *Journal of Chromatography A*, 2023, 1703: 464113.

Chapter 5 includes extracts from the following published works of the author:

- Ding, C., Ardeshta, H., Gillespie, C., Ierapetritou, M., Process design of a fully integrated continuous biopharmaceutical process for economic assessment and ecology impacts. *Biotechnology & Bioengineering*, 2022, 119(12):3567-3583.

Copyright permissions and license agreements for published materials are listed below.

1. Ding, C., Yang, O., Ierapetritou, M., Towards digital twin for biopharmaceutical processes: concept and progress. In: Pörtner, R. (eds) Biopharmaceutical Manufacturing. Cell Engineering, vol 11. Springer.

 **Marketplace**

This is a License Agreement between Chaoying Ding ("User") and Copyright Clearance Center, Inc. ("CCC") on behalf of the Rightsholder identified in the order details below. The license consists of the order details, the Marketplace Permissions General Terms and Conditions below, and any Rightsholder Terms and Conditions which are included below. All payments must be made in full to CCC in accordance with the Marketplace Permissions General Terms and Conditions below.

Order Date	25-Apr-2024	Type of Use	Republish in a thesis/dissertation
Order License ID	1477319-4	Publisher	Springer International Publishing
ISBN-13	9783031456688	Portion	Chapter/article

LICENSED CONTENT

Publication Title	Biopharmaceutical Manufacturing : Progress, Trends and Challenges	Rightsholder	Springer Nature BV
Article Title	Towards Digital Twin for Biopharmaceutical Processes: Concept and Progress. Chapter 6. Manufacturing of Recombinant Therapeutic Proteins	Publication Type	Book
Date	01/11/2024	Start Page	179
Language	English	End Page	211
Country	Switzerland	Volume	11

2. Ding, C., Ierapetritou, M., A novel framework of surrogate-based feasibility analysis for establishing design space of twin-column continuous chromatography. *International Journal of Pharmaceutics*, 2021, 609: 121161.

This is a License Agreement between Chaoying Ding ("User") and Copyright Clearance Center, Inc. ("CCC") on behalf of the Rightsholder identified in the order details below. The license consists of the order details, the Marketplace Permissions General Terms and Conditions below, and any Rightsholder Terms and Conditions which are included below. All payments must be made in full to CCC in accordance with the Marketplace Permissions General Terms and Conditions below.

Order Date	25-Apr-2024	Type of Use	Republish in a thesis/dissertation
Order License ID	1477319-1	Publisher	ELSEVIER BV
ISSN	0378-5173	Portion	Chapter/article

LICENSED CONTENT

Publication Title	International journal of pharmaceutics	Rightsholder	Elsevier Science & Technology Journals
Article Title	A novel framework of surrogate-based feasibility analysis for establishing design space of twin-column continuous chromatography	Publication Type	Journal
		Start Page	121161
		Volume	609
Date	01/01/1978		
Language	English		
Country	Netherlands		

3. Ding, C., Ierapetritou, M., Machine learning-based optimization of a multi-step ion exchange chromatography for ternary protein separation. *Computers & Chemical Engineering*, 2024, 184:108642.

This is a License Agreement between Chaoying Ding ("User") and Copyright Clearance Center, Inc. ("CCC") on behalf of the Rightsholder identified in the order details below. The license consists of the order details, the Marketplace Permissions General Terms and Conditions below, and any Rightsholder Terms and Conditions which are included below. All payments must be made in full to CCC in accordance with the Marketplace Permissions General Terms and Conditions below.

Order Date	26-Apr-2024	Type of Use	Republish in a thesis/dissertation
Order License ID	1477559-1	Publisher	JOHN WILEY & SONS, INC.
ISSN	1097-0290	Portion	Chapter/article

LICENSED CONTENT

Publication Title	Biotechnology and bioengineering	Publication Type	e-Journal
Article Title	Process design of a fully integrated continuous biopharmaceutical process using economic and ecological impact assessment	Start Page	3567
		End Page	3583
		Issue	12
Date	01/01/1962	Volume	119
Language	English	URL	http://onlinelibrary.wiley.com/journal/10...
Country	United States of America		
Rightsholder	John Wiley & Sons - Books		

4. Ding, C., Gerberich, C., Ierapetritou, M., Hybrid model development for parameter estimation and process optimization of hydrophobic interaction chromatography. *Journal of Chromatography A*, 2023, 1703: 464113.

This is a License Agreement between Chaoying Ding ("User") and Copyright Clearance Center, Inc. ("CCC") on behalf of the Rightsholder identified in the order details below. The license consists of the order details, the Marketplace Permissions General Terms and Conditions below, and any Rightsholder Terms and Conditions which are included below. All payments must be made in full to CCC in accordance with the Marketplace Permissions General Terms and Conditions below.

Order Date	25-Apr-2024	Type of Use	Republish in a thesis/dissertation
Order License ID	1477319-2	Publisher	ELSEVIER BV
ISSN	1873-3778	Portion	Chapter/article

LICENSED CONTENT

Publication Title	Journal of chromatography A	Rightsholder	Elsevier Science & Technology Journals
Article Title	Hybrid model development for parameter estimation and process optimization of hydrophobic interaction chromatography	Publication Type	e-Journal
Date	01/01/1993	Start Page	464113
Language	English	Volume	1703
Country	Netherlands		

5. Ding, C., Ardeshtna, H., Gillespie, C., Ierapetritou, M., Process design of a fully integrated continuous biopharmaceutical process for economic assessment and ecology impacts. *Biotechnology & Bioengineering*, 2022, 119(12):3567-3583.

This is a License Agreement between Chaoying Ding ("User") and Copyright Clearance Center, Inc. ("CCC") on behalf of the Rightsholder identified in the order details below. The license consists of the order details, the Marketplace Permissions General Terms and Conditions below, and any Rightsholder Terms and Conditions which are included below. All payments must be made in full to CCC in accordance with the Marketplace Permissions General Terms and Conditions below.

Order Date	25-Apr-2024	Type of Use	Republish in a thesis/dissertation
Order License ID	1477319-3	Publisher	PERGAMON
ISSN	0098-1354	Portion	Chapter/article

LICENSED CONTENT

Publication Title	Computers & chemical engineering	Country	United Kingdom of Great Britain and Northern Ireland
Article Title	Machine learning-based optimization of a multi-step ion exchange chromatography for ternary protein separation	Rightsholder	Elsevier Science & Technology Journals
Author/Editor	European Symposium on Computer Aided Process Engineering.	Publication Type	Journal
Date	01/01/1977	Start Page	108642
Language	English	Volume	184