

Design of Online Estimator for Culture Monitoring and Media Development for *Bordetella pertussis*

by

Abhishek Mishra

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Chemical Engineering

Waterloo, Ontario, Canada, 2024

© Abhishek Mishra 2024

Examining Committee Members

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner: Dr. Vikramaditya Yadav
Associate Professor, Dept. of Chemical and Biological Engineering,
University of British Columbia, Canada

Supervisor: Dr. Hector Budman
Professor, Dept. of Chemical Engineering, University of Waterloo

Internal-External Member: Dr. Wayne Parker
Professor, Dept. of Civil Engineering, University of Waterloo

Internal Member: Dr. Eric Croiset
Professor, Dept. of Chemical Engineering, University of Waterloo

Internal Member:

Dr. Bill Anderson

Professor, Dept. of Chemical Engineering, University of Waterloo

Author's Declaration

“This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public”.

Statement of Contribution

Chapter 4 has been published as the paper below, author contributed as the first author in developing the idea of the paper, implementation of the method and writing the manuscript:

Abhishek Mishra, Melih Tamer, and Hector Budman. "In-line monitoring of Bordetella pertussis cultivation using fluorescence spectroscopy." *Bioprocess and Biosystems Engineering* 46.6 (2023): 789-802.

Chapter 5 has been published as the paper below, author contributed as the first author in developing the idea, implementation of the method and writing the manuscript:

Abhishek Mishra, Sarah Steinbach, Ibrahim M. Tamer, and Hector Budman. "Effect of oxidative stress on antigen productivity in B. Pertussis cultures." *Biochemical Engineering Journal* 210 (2024): 109421.

Results of chapter 6 has been presented in International Congress of Chemical and Process Engineering (CHISA), 2024.

Abstract

Whooping cough, or pertussis, is a contagious respiratory infection. Sanofi Pasteur's pertussis vaccine production involves fermentation using reactors of increasing sizes, where cells from each stage inoculate the next. The upstream product undergoes purification to extract the five antigens needed for the acellular vaccine. A major challenge is the variability in antigen yields, particularly for pertactin, which limits overall productivity. Despite controls for parameters like pH, temperature, agitation, aeration, and nutrient feed rates, there's no real-time monitoring to confirm growth and productivity during the batch run. This often leads to undetected deviations and potential yield losses, making fault detection methods essential for identifying variations and ensuring productivity.

The current work deals with the development of a method of bioprocess monitoring using soft sensors for *Bordetella pertussis* cultures and studying changes in media composition that impact the yield of antigens. Accordingly, the following objectives were pursued:

- Development of a real time fluorescence-based monitoring system.
- Investigation of possible sources of oxidative stress and its impact or correlation with antigen production.
- Development of a protocol for simultaneous monitoring of oxidative stress and antigen produced.
- Development of a mechanistic model for use in model-based filtering of an online fluorescence based sensor

This thesis presents the design and implementation of an in-line fluorescence spectroscopy system capable of real-time monitoring of critical bioprocess parameters. By utilizing dual excitation wavelength fluorometry, this method was able to track the dynamics of biomass, amino acids, and antigen production throughout the fermentation process. The fluorescence data, combined with statistical technique such as Partial Least Squares (PLS) regression was used for predicting key state variables. A significant enhancement in the predictive accuracy of the PLS models was observed when the models were calibrated for the bacterial strain and the media composition used in each cultivation process.

Previous research by our group ([Vitelli et al. \[2023c\]](#), [Zavatti et al. \[2016\]](#)) indicated a potential link between oxidative stress and antigen yield variability, particularly for pertactin. We hypothesized that glutamate, the most abundant carbon source, could be causing this oxidative stress. To test this, we developed a multi-parametric flow cytometry method to simultaneously monitor intracellular ROS and pertactin surface expression. Given that pertactin is an auto-transporter protein, understanding the relation between expression and secretion was crucial. Our findings showed a negative correlation between oxidative stress and pertactin surface expression. Using a tailored protein quantification method via affinity chromatography, combined with multi-color flow cytometry, we confirmed that higher glutamate concentrations induced higher oxidative stress and result in reduced pertactin secretion. The studies included both batch and fed-batch bioreactor experiments where the latter closely emulated the production environment at Sanofi. Hence, the findings demonstrated that variability in initial glutamate concentrations can have a major impact of productivity and may partially explain the variability observed in the manufacturing process.

The model from (Vitelli et al. [2023c]) was adapted to formulate a hybrid model that was used for model-based filtering of an online sensor. This mechanistic model incorporated the interactions between glutamate, ROS, and NADPH in neutralizing ROS. After calibration, it accurately predicted key variables under different oxidative stress conditions. In parallel, PLS regression models were developed using in-line fluorescence spectra which could predict OD, glutamate, and NADPH. By using a hybrid model that combines the mechanistic and PLS regression models via an Extended Kalman Filter, more accurate real-time estimates of key variables were obtained. The Akaike Information Criteria (AIC) confirmed that this hybrid model achieved a superior balance between complexity and accuracy as compared to purely mechanistic or PLS models.

Acknowledgements

First and foremost, I want to express my deepest gratitude to my advisor, Professor Hector Budman. His promptness and willingness to dedicate invaluable time to my progress have been instrumental throughout my research journey. His passion and unwavering commitment to research have been a constant source of inspiration, motivating me to persevere during challenging times. I am profoundly grateful for his continuous support and guidance, which has greatly contributed to my growth as a researcher. I would also like to express my gratitude for Dr. Boris Tartakovsky, who always guided me in my research and always helped me whenever I needed it. I also extend my heartfelt thanks to the Department of Chemical Engineering at the University of Waterloo for providing me with the opportunity to pursue my PhD in an environment rich with resources and a welcoming atmosphere. I am especially grateful to the technical staff, with particular thanks to Charles Dal Castel, Ravindra and Bert Habicher, whose patience and assistance have been invaluable throughout this journey.

I would also like to express my sincere appreciation to Dr. John Zhang and Dr. Yael Zilberman Simakov for never saying no to my TA applications. I will always treasure the enjoyable moments I shared while assisting both of them in handling undergrad labs. My gratitude also extends to Dr. Melih Tamer for his unwavering support throughout my PhD program, and for facilitating connections with domain experts at Sanofi Pasteur, which proved critical to my research. I am deeply thankful to my comprehensive examination committee members: Professor Wayne Parker, Professor Eric Croiset, and Professor Bill Anderson. Their insightful feedback provided direction to my research and enriched it with

new perspectives.

The warmth and camaraderie of my current and former colleagues and friends have made this journey memorable. My heartfelt thanks to Mousa, Subhranshu, Alok, Shubham Andkhau, Abhishek Panda, Saswat Nayak, Raghavan Gulti, Jainendra, Shubham, Birkishor, Vikas, Chandramukhi, Bhavana, Akanksha, Dr. Michael, Dr. Xin, Ana, Megala, Shilpa, Ipsita, Ahalya, Mohit, Yasmin, Pramoth, Ali, Alex, Zahra, Meghana, Arshiya, Adithya, Anshul, Rishabh Bajaj, Harish, Sudhan, Aaron, Madhuja, Kaycee, Haril, Avi, Mayur and Saurabh.

I also want to express my deepest gratitude to my girlfriend Priyanka, who was once my teacher during my undergraduate studies and is now my greatest support during my PhD. Her guidance then and her unwavering support now have been instrumental in my journey. I'm incredibly fortunate to have her by my side through it all. I owe a huge debt of gratitude to Mohammad Aghaei, Shuji Chang and Ittisak for being a steadfast support system, offering critical advice at pivotal moments, and always coming with me for coffee breaks. A special shoutout to Rishabh Tripathi for being a constant presence throughout these years, now a decade of friendship. My thanks also go to Ashirbad Pradhan and Dr. Piyush Agarwal for their unwavering support and care during my PhD.

Finally, this work would not have been possible without the constant support of my family. My endless gratitude goes to my parents for their unconditional love and care. I am also deeply indebted to my sister Nidhi for her affectionate love. Many others, knowingly or unknowingly, have contributed to the successful completion of this work, and I extend my heartfelt thanks to them all.

Dedication

*I dedicate this thesis to
my grandparents, parents and
my hometown Gorakhpur.*

Table of Contents

Examining Committee	ii
Author’s Declaration	iv
Statement of Contribution	v
Abstract	vi
Acknowledgements	ix
Dedication	xi
List of Figures	xix
List of Tables	xxvi
1 Introduction	1
1.1 Thesis Outline	7

2	Background and Literature Review	9
2.1	Whooping Cough and Vaccines	9
2.2	Virulence Factor Pertactin	12
2.2.1	Secretion Mechanism of Pertactin	12
2.3	Monitoring and Control of Bio-processes	13
2.4	Fluorescence	15
2.4.1	Theory of Fluorescence	15
2.4.2	Fluorophores	16
2.4.3	Fluorescence Spectroscopy	18
2.5	Fluorescence Spectroscopy for Bio-Process Monitoring	19
2.6	Oxidative Stress in Bacterial Cells	23
2.7	Principles of Flow Cytometry	26
2.7.1	Multi-colour Flow Cytometry	27
2.8	Column Chromatography	28
2.8.1	Affinity Chromatography	30
2.9	Multivariate Statistical Analysis	31
2.9.1	Data Pre-Processing	32
2.9.2	Principal Component Analysis	32
2.9.3	Partial Least Squares Regression	34

2.10	Mechanistic Models	35
2.11	Akaike Information Criteria	39
2.12	Extended Kalman Filter	39
3	Materials and Methods	41
3.1	<i>Bordetella pertussis</i> Flask Cultures	41
3.2	Batch Fermentation of <i>B. pertussis</i> in Bioreactors	42
3.2.1	Fed-Batch Runs	43
3.2.2	Preparation and formulation of media, growth factor and supplement	43
3.3	Selective Plating for <i>B. pertussis</i> Identification	45
3.4	Measurement Techniques	46
3.4.1	Optical density (OD)	46
3.4.2	Amino acid analysis using HPLC	46
3.4.3	Glutamate quantification using assay kit	47
3.5	Enzyme-Linked Immunosorbent Assay (ELISA)	48
3.6	Chromatographic Purification of Pertactin	49
3.6.1	Ligand Coupling and Column Preparation	49
3.6.2	Pertactin Purification	50
3.7	SDS PAGE Analysis	50

3.8	Fluorescence Spectroscopy	51
3.8.1	In-Line fluorometer	52
3.8.2	Off-Line fluorometry	53
3.9	Multi-Colour Flow Cytometry	54
3.9.1	Flow cytometry gating	54
3.9.2	Oxidative stress	55
3.9.3	Surface expression of pertactin	56
3.10	Fluorescence Microscopy	56
4	In-Line Monitoring of Bordetella Pertussis Cultivation Using Fluores-	
	cence Spectroscopy	58
4.1	Overview	58
4.2	Introduction	59
4.3	Materials and Methods	65
4.3.1	Culture Conditions	65
4.3.2	Offline measurements	67
4.3.3	Fluorescence Map Acquisition	68
4.3.4	Data preprocessing	69
4.3.5	Akaike Information Criteria	69
4.4	Results and Discussions	71

4.4.1	Qualitative description of the spectra	72
4.4.2	PLS model for biomass, amino acid and antigen prediction	74
4.4.3	Effect of cell line and media	79
4.4.4	Additional online measurements as model features in the soft sensor model	83
4.4.5	Akaike Information Criteria	86
4.5	Conclusions	87
5	Correlation Between Oxidative Stress and Growth & Antigen Produc- tivity in B. Pertussis cultures	90
5.1	Overview	90
5.2	Introduction	91
5.3	Materials and Methods	96
5.3.1	Flask Culture Conditions and Experimental Setup	96
5.3.2	Bioreactor Conditions and Experimental Setup	97
5.3.3	OD and Glutamate Quantification	97
5.3.4	Fluorescence Microscopy	98
5.3.5	Multi Colour Flow Cytometry for ROS and Surface Expression of PRN	99
5.3.6	Chromatographic Purification of Pertactin	101
5.3.7	SDS PAGE Analysis	103

5.3.8	Spectral Analysis of Culture Supernatant	103
5.4	Results	104
5.4.1	Fluorescence Microscopy	104
5.4.2	Cell growth, Oxidative Stress and Productivity	106
5.5	Discussion	113
5.6	Conclusions	121
6	A Mechanistic-Model Based Approach for Monitoring <i>B. pertussis</i> Fermentation	123
6.1	Introduction	123
6.2	Metabolic Model Development	127
6.3	Materials and Methods	133
6.3.1	<i>B. Pertussis</i> Fed-Batch Fermentations	133
6.3.2	Biomass and Glutamate Quantification	135
6.3.3	In-Line Fluorescence Acquisition	136
6.3.4	Fluorescence Spectroscopy	137
6.3.5	Flow Cytometry	137
6.3.6	Affinity Chromatography for Pertactin Quantification	137
6.3.7	Data Pre-processing and PLS Regression	138
6.3.8	Model Fitting	140

6.4	Results and Discussion	142
6.4.1	Cell growth, Oxidative Stress and Productivity	142
6.4.2	PLS Regression Model for State Estimation	144
6.4.3	Mechanistic Model Validation	150
6.4.4	Extended Kalman Filter	150
6.4.5	Akaike Information Criteria	157
6.5	Conclusions	158
7	Conclusions and Future Work	160
7.1	Conclusions	160
7.1.1	In-Line Monitoring of Bordetella Pertussis Cultivation Using Fluorescence Spectroscopy	161
7.1.2	Correlation between Oxidative Stress and Growth & Antigen Productivity in B. Pertussis cultures	162
7.1.3	Mechanistic model based monitoring of B. pertussis cultures	165
7.2	Future Work	166
	References	171

List of Figures

1.1	Reactor train setup for pertussis vaccine manufacturing at Sanofi. After reaching the desired biomass density, final batches are harvested. Downstream process includes multiple steps of protein purification to finally extract purified antigens needed for pertussis vaccine.	3
2.1	Jablonski diagram.	17
2.2	Schematic representation of a flow cytometer. BP: band pass, LP: long pass, SP: short pass. Half mirror (or beam splitter) passes 50 percent and reflects 50 percent of the light. 90/10 beam splitter reflects 90 percent and passes 10 percent of the light (taken with modifications from (da Silva et al. [2012])).	29
2.3	Principal components in PCA.	33
3.1	BG agar plate with <i>B. pertussis</i> growth.	46
3.2	Contaminated TSA plate.	46
3.3	TSA plate without contamination.	46

3.4	BG agar and TSA plates to identify and confirm contamination and pertussis growth.	46
3.5	Chromatography run on AKTA for purification of pertactin. Y-axis of the graph shows the parameters monitored during the chromatography run. . .	51
3.6	Optical probe and bioreactor setup with in-line probe for real-time fluorescence data acquisition.	53
3.7	Different flow cytometry gatings utilized to isolate <i>B. pertussis</i> population.	55
4.1	Emission spectra dynamics over the batch course for batch B.	74
4.2	Batch profile for agitation, dissolved oxygen and culture volume.	76
4.3	Performance of PLS regression model for experimental set 1 (Leave One Batch Out approach). Fluorescence area along with agitation, dissolved oxygen and culture volume at each time point were used as model predictor for the estimation of target variables such as OD, glutamate and proline consumption and pertactin antigen production over the batch course. Predicted and actual values (standardized) were plotted over the batch course for (a) OD (b) Glutamate (c) Proline (d) PRN. Figure (e), (f), (g), (h) show predicted vs measured values for corresponding variables.	79

4.4	Performance of PLS regression model for experimental set 2 (Leave One Batch Out approach). Predicted and actual values (standardized) were plotted over the course of the batch for (a) OD (b) Glutamate (c) Proline (d) PRN. Figure (e), (f), (g), (h) gives the plot of predicted vs measured values for corresponding variables.	80
4.5	Performance of PLS regression model for experimental set 3 (Leave One Batch Out . Predicted and actual values (standardized) were plotted over the course of the batch for (a) OD (b) Glutamate (c) Proline (d) PRN. Figure (e), (f), (g), (h) gives the plot of predicted vs measured values for corresponding variables.	81
4.6	PLS weight plots of first two principal components for model set 1, 2 and 3.	89
5.1	Fluorescence microscopy images illustrating the surface expression of PRN and the generation of Reactive Oxygen Species (ROS) by bacterial cells at different growth stages (100X zoom). During the initial lag phase, PRN expression is nearly absent (depicted by negligible red fluorescence) while ROS levels are high (depicted by green fluorescence). In the exponential phase, a substantial population of cells exhibits high PRN expression (depicted by red fluorescence), accompanied by a decrease in intracellular ROS levels (depicted by reduced green fluorescence). Subsequently, during the stationary phase, cells experiencing oxidative stress due to starvation display elevated ROS production (depicted by intensified green fluorescence) and reduced PRN expression (depicted by diminished red fluorescence). . .	105

5.2	Bacterial growth curve for the flask and bioreactor runs. Subfigure (5.2b) and (5.2d) shows the zoomed in growth curve to show the growth inhibition due to higher concentration of glutamate, which leads to less growth leading to lower antigen productivity.	107
5.3	Plot of weighted average of ROS and PRN vs time for the flask runs. . . .	108
5.4	Fluorescence Intensity Dynamics Across Emission Wavelengths in Three Flask Experiments. Sub-figures (5.4a), (5.4b), and (5.4c) represent the fluorescence intensity profiles at 18, 36, and 48 hours, respectively. Notably, the flask with elevated glutamate concentration consistently exhibits higher intensity levels across the entire emission spectrum.	110
5.5	Bacterial growth curve comparing area under the curve and pertactin yield for the bioreactor runs (shown in figure a) and agitation and dissolved oxygen profiles of the bioreactors (shown in figure b). Agitation rate varied from 200 to 700 RPM to maintain the dissolved oxygen concentration of 35 %. .	110
5.6	Plot of weighted average of ROS and PRN vs time for bioreactor runs. . .	113
5.7	Fluorescence Intensity Dynamics Across Emission Wavelengths in bioreactor Experiments. Sub-figures (5.7a), (5.7b), and (5.7c) represent the fluorescence intensity profiles at 6, 18, and 32 hours, respectively. Notably, the culture with elevated glutamate concentration consistently exhibits higher intensity levels across the entire emission spectrum.	114

5.8	Correlations between the rate of change of the weighted averages of surface pertactin expression and ROS (calculated as per equations 5.2 and 5.3) for flask and bioreactor experiments.	115
5.9	Correlation between secreted PRN and the area under the curve (AUC) for flask and bioreactor experiments.	116
5.10	Glutamate consumption rates for flask and bioreactor experiments.	119
5.11	SDS PAGE analysis of culture supernatant and fractions collected from affinity chromatography using 1 ml columns. Samples loaded in SDS gel includes supernatant material(SM), first flow through the column (load), wash fraction, eluate and strip fraction. Pure pertactin was also passed through the column and eluate fraction was collected to validate the working of the column (well 8). Well 9 contains pure pertactin (provided by Sanofi).	120
6.1	Schematic representation of the mechanism described in the model.	129
6.2	Model fitting results for (6.2a) optical density (OD), (6.2b) extracellular glutamate, (6.2c) extracellular NADPH, (6.2d) ROS and (6.2e) PRN, along with offline measurements.	141
6.3	Correlation between glutamate consumption rate to ROS ratio and pertactin secretion rate per log mean OD. The trend shows that secretion of pertactin is positively correlated to glutamate consumption rate and inversely correlated to ROS.	145

6.4	Experimental data for the state variables. Blue dot and line shows the control runs and orange line shows runs with 20 % higher glutamate. . . .	146
6.5	Performance of PLS regression model for different state variables. Figures (a), (b) and (c) shows the predicted vs time plots for OD, glutamate and extracellular NADPH respectively. Figures (d), (e) and (f) show the plots of true vs predicted values for the corresponding variables.	148
6.6	Performance of PLS regression model for different state variables. Figures (a) and (b) shows the predicted vs time plots for ROS and PRN and figure (c) and (d) shows predicted vs true values for the corresponding variables.	149
6.7	Mechanistic model validation for OD, glutamate, NADPH, ROS and PRN for the initial glutamate concentration of 10.3 g/L and initial OD of 0.3 (approx.). Experimental data from a batch and a fed-batch run which was not used in the calibration is shown by blue dots.	152
6.8	: Mechanistic model validation for OD, glutamate, NADPH, ROS and PRN for the initial glutamate concentration of 12.7 g/L and initial OD of 0.3 (approx.). Experimental data from a batch and a fed-batch run which was not used in the calibration is shown by blue dots.	153
6.9	Extended kalman filter validation model for 32 hour batch runs for the estimation of paramters such as OD, glutamate, NADPH, ROS and PRN for 10.3 g/L initial glutamate concentration and inoculation OD of of 0.3. PLS regression mode and mechanistic model are also plotted against validation batch for comparison.	155

6.10 Extended kalman filter validation model for the 48 hour fed-batch runs for the estimation of OD, glutamate, NADPH, ROS and PRN for the starting OD of 0.28 and 12.03 g/L of initial glutamate concentration. PLS regression estimates and mechanistic model estimates are also plotted against validation batches for comparison.	156
--	-----

List of Tables

2.1	Applications of fluorescence spectroscopy for in-line monitoring of bioprocesses.	24
3.1	Culture media components and their normalized composition for 2X media concentration.	44
3.2	Growth factor components (100X concentrated) and their normalized composition.	44
3.3	Supplement components and their normalized composition.	45
3.4	Gradient schedule for the HPLC runs using a ternary eluent system.	47
4.1	Culture Conditions.	66
4.2	Emission areas used for on-line monitoring	70
4.3	Prediction accuracy of 'Global' model. The model was calibrated by combining data from all the batch runs. The best result for LOBO approach is tabulated here. RMSE values along with coefficient of determination is calculated to evaluate the performance of regression models.	77

4.4	Modeling results of PLS regression with differing cell line and media used for training and prediction (using fluorescence, agitation rate, dissolved oxygen and volume data).	83
4.5	Prediction accuracy of PLS regression models for all sets of experiment with two different prediction strategies for set 1, 2 and 3. The regression model included fluorescence areas, agitation, dissolved oxygen and volume as model features.	84
4.6	Prediction efficiency of the regression model for all sets of experiment when the additional features (agitation, dissolved oxygen and volume changes) were not included with area.	85
4.7	Akaike Information Criteria (AIC) values for regression models.	87
5.1	Summary of batch runs.	97
6.1	Definition of variables used in the model.	133
6.2	Parameters defined in the model.	134
6.3	Summary of batch and fed-batch runs.	135
6.4	Excitation wavelength and emission ranges used for area calculation using fluorescence spectra.	139
6.5	Performance of the PLS regression model for different state variables. . . .	148
6.6	Calibrated parameters of the mechanistic model.	151

6.7	Performance of extended Kalman filter and mechanistic model for the estimation of different state variables for the validation of batch with 12.03 g/L of initial glutamate concentration.	155
6.8	AIC value comparison of mechanistic model, PLSR and EKF.	157

Chapter 1

Introduction

Whooping cough, also referred to as Pertussis, is a highly contagious respiratory tract disease. It is caused by the bacteria *Bordetella pertussis*, which is a gram-negative, aerobic, pathogenic, encapsulated coccobacillus of the genus *Bordetella* (Thalen [2008a]). The disease is very common and fatal in infants but can also be severe in children and adults. Even today, the disease continues to pose high threat to infants, with around 50 million cases and 300,000 deaths annually (Nieves and Heininger [2016b]), making it one of the world's leading cause for vaccine preventable deaths.

The vaccine available for pertussis may be cellular (containing whole cells) or acellular (containing fragments of cell or antigens). The vaccine produced by Sanofi Pasteur (industrial sponsor), referred to as 5-component pertussis, is an acellular vaccine which contains 4 different antigens and overall 5 components: *Fimbriae* (type 2 and 3), *Filamentous Hemagglutinin*, *Pertactin* and *Pertussis Toxin*. The upstream process at Sanofi involves a train

of reactors of increasing volume in which the final cell population from a reactor is used to inoculate the next reactor in the train. A suitable ratio between the inoculum size and the bioreactor volume ensures that cells will adapt efficiently to large-scale conditions, ensuring faster and more consistent growth. Thus, a train of reactors is needed to gradually amplify the cell mass while maintaining the ratio of inoculum size to volume. Further, the product from the upstream processes is then purified through several steps to obtain the five antigens that constitute the acellular vaccine. Figure (1.1) shows upstream setup for pertussis antigen production. A key challenge faced by the company in this process is the resulting variability in the yields of different antigens especially Pertactin (PRN) for different runs. Thus, the productivity of Pertactin is currently the bottleneck of the process. Also, due to lack of online informative sensors, if the fermentation process evolves abnormally, i.e. a fermentation that will result in low yield, and there is no way to know the outcome until the end of the purification when the antigen yield is finally quantified. So, developing online estimators can help in improving the bio-process control and can also aid in fault detection mid-run so that corrective measures can be taken during the remaining time or post run for the next batch.

Most bio-processes are operated in a sterile control volume containing a complex mixture of biochemicals and are controlled to maintain a desired level of the product. Following recent guidelines, the Food and Drug Administration (FDA), recommends that quality assurance and regulatory compliance will be accounted for during the process design phase. In the past, the manufacturing process was developed first and then quality control and assurance tests were then applied to ensure compliance. This new initiative by FDA aims to mitigate risk by reducing product and process variability and also providing for con-

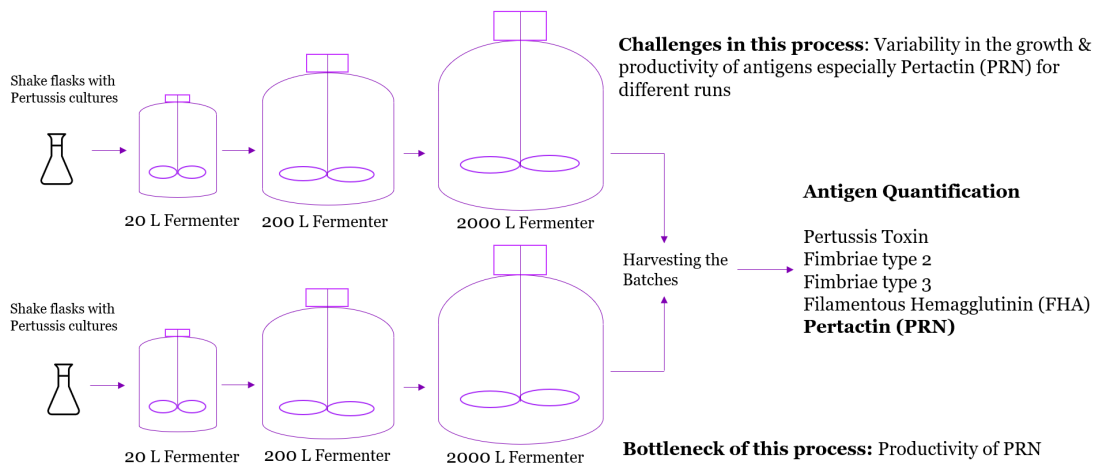


Figure 1.1: Reactor train setup for pertussis vaccine manufacturing at Sanofi. After reaching the desired biomass density, final batches are harvested. Downstream process includes multiple steps of protein purification to finally extract purified antigens needed for pertussis vaccine.

tinuous product improvement. Therefore, identifying the causes of variability, and taking measures to mitigate it, has become crucial for meeting regulatory requirements. In a bio-process, there can be multiple sources of variability. Some of the common causes of variability are changes in raw materials, faulty equipment, oxidative stress in cell cultures and faulty operation procedures. Therefore, the development of a robust instrumentation capable of monitoring the critical process parameters in real time is required for diagnosing sources of variability. This monitoring activity is expected to reveal the sources of abnormal operation based on the collected historical data.

Fluorescence spectroscopy is a well known technique that can be used to monitor fluorescent compounds in bioprocesses. Some cellular compounds such as proteins and metabolites have intrinsic fluorescence, and thus they can be monitored using a fluorometer. These

fluorescent compounds, called fluorophores, are part of the organism’s metabolism and can be correlated with growth and other bio-process variables. The development of online sensors based on fluorescence spectroscopy combined with chemometric methods can be used to control and optimize the process in real time. Since fluorescence spectroscopy is a non-invasive and highly sensitive method with detection ability of multiple components at low concentrations, it can be a promising method for bioprocess monitoring. The use of fluorescence spectroscopy for in-line monitoring is not very well established in the pharmaceutical industry. Hence, development of an in-line fluorometer for bioprocess monitoring and control is an attractive alternative to other complex and costly methods available in the market such as RAMAN and near infrared spectroscopy (NIR). Moreover, development of an in-line monitoring method also aligns with the objectives of PAT initiative of achieving quality by design. The development of an in-line fluorescence sensor is presented in Chapter (4) of this thesis.

With respect to the variability of the whooping cough vaccine process at Sanofi, previous studies by our group (Zavatti et al. [2019b, 2020b], Vitelli et al. [2023c]) found that oxidative stress can have a major impact on growth and productivity of batches. It was also found that profiles of pH, dissolved oxygen, temperature, agitation and aeration rates that are regularly monitored during fermentation are not informative about the presence of oxidative stress. Oxidative stress can be assessed by flow cytometry but it is not possible to do it on-line or in real time. (Zavatti et al. [2019b, 2020b]) found that the initial glutamate (main constituent of nutrient media and carbon source for bacteria) concentration plays a major role in inducing oxidative stress in bacterial cells. However, the studies (Zavatti et al. [2019b, 2020b]) were performed in flasks and thus they do not emulate the conditions in

the actual production scale where the bacteria is grown in bioreactors under tight oxygen, pH and temperature control. The study of oxidative stress in batch bioreactor operation is the subject of Chapter (5) of this thesis. Also, the flask experiments were conducted in batch mode while the bioreactors in the production environment are conducted in a mixed batch-fed-batch operation: batch for a period of approximately 24 hours followed by an additional 24 - 30 hours period of fed-batch operation. The batch-fed-batch operation of the bioreactor and the resulting growth and productivity are discussed in Chapter (6) of the thesis. One factor that can also lead to variability in measurements is the quantification method itself. It was found in our studies that enzyme-linked immunosorbent assay (ELISA) method used for quantification of pertactin is not very reliable and gives highly variable results. To make sure that the results are accurate and more consistent, a new method of protein quantification using affinity columns was developed by modifying a previously developed method of pertactin quantification by Sanofi. The quantification of pertactin by chromatography is presented in Chapter (3) that summarizes materials and methods in this thesis and it is further discussed in Chapter (5) and (6) where it was used to monitor the bioreactor operation in batch and fed-batch respectively.

One of the hypothesis of the current thesis is that reactive oxygen species (ROS) within a cell has direct impact on the cell's antigen productivity. Reactive oxygen species formed during respiration, that are highly reactive and cause damage to intracellular lipids and proteins. To demonstrate this hypothesis a 2D flow-cytometry method was developed to simultaneously measure the intracellular ROS and pertactin cell surface concentrations. This method is presented in Chapter (3) and the results of its application in bioreactor runs are further discussed in Chapters (5) and (6) of the thesis.

An additional limitation identified in previous studies by our group is that fluorescence sensors that are based on a combination of spectra and chemometric analysis are mainly accurate within the range of operating conditions used for chemometric model validation but they are less accurate outside the range of data used for calibration. Hence, mechanistic models based on first principles are increasingly applied to provide better predictions. These models can be used for both offline use, e.g. to determine optimal operating conditions, as well as online monitoring. For the latter application, the mechanistic models can be combined with data, e.g. fluorescence spectra, to develop Extended Kalman Filters (EKF) ([Ribeiro \[2004\]](#)) for filtering of measured states and estimation of unmeasured states. The use of mechanistic models to extend the range of validity of the fluorescence based on the in-line sensor is presented in Chapter ([6](#)).

In summary, this work addresses some of the limitations of the previous studies reviewed above with the goal of identifying the effect of oxidative stress in growth and productivity in bioreactor operation. It is worth emphasizing that all previous studies done by our group were done with either limited samples from the production environment or flask experiments. By performing experiments in 2-L bioreactors with a set of measurements that are currently not available in the production environment, the results in this thesis are expected to provide greater understanding of the phenomena as compared to the previous studies.

Following the above, the novel contributions of this work are enumerated as follows:

- Developed a method of in-line monitoring of *Bordetella pertussis* cultivation using dual excitation wavelength fluorometry.

- Investigated sources of reactive oxygen species (ROS) and its role as a causative agent for variability in batch productivity.
- Developed a 2-dimensional flow cytometry method to simultaneously monitor oxidative stress and productivity.
- Extended a chromatographic based separation method for Pertactin quantification in the bioreactor supernatant broth.
- Developed a mechanistic model to relate oxidative stress to cell growth and productivity.
- Developed a soft-sensor (EKF) based on a combination of online fluorescence and a mechanistic model.

1.1 Thesis Outline

The remaining of this thesis is organised as follows.

- Chapter (2) presents background and literature review on whooping cough, vaccines, bio-process monitoring, fluorescence spectroscopy, chemometric methods used in on-line monitoring system, the occurrence of oxidative stress in bacterial cells, and flow cytometry and its applications to the current study.
- Chapter (3) summarises the materials and methods used for analysis in experiments conducted.

- Chapter (4) summarises the use of fluorescence spectroscopy as a tool for on-line monitoring of *Bordetella Pertussis* cultivation in bioreactors.
- Chapter (5) outlines the experimental methods utilizing multi-parameter flow cytometry to investigate the role of media components (glutamate) in inducing high ROS, which is hypothesised as a causative agent for variability in antigen productivity.
- Chapter (6) presents the use of mechanistic model combined with an extended kalman filter to predict state variables using fluorescence measurements.
- Chapter (7) presents the summary and conclusions along with proposed future work.

Chapter 2

Background and Literature Review

This chapter provides an overview of the theories, current work and research gaps in the research on culture monitoring and performance for *Bordetella pertussis*.

2.1 Whooping Cough and Vaccines

Whooping cough, also known as pertussis, is a severe respiratory illness that predominantly affects infants and young children but can also occur in adults. This research aims to further understanding and propose improvements for the manufacturing process of the whooping cough vaccine. The work was conducted in partnership with the Manufacturing Technologies Laboratory at Sanofi Pasteur in Toronto. Whooping cough is caused by *Bordetella pertussis*, a gram-negative, rod-shaped, aerobic bacterium first isolated by Bordet and Gengou in 1906 (Kuchar et al. [2016]). The hallmark symptoms of pertussis include severe, spasmodic coughing accompanied by a distinctive whooping sound during

inhalation ([Nieves and Heininger \[2016a\]](#)). First documented in the 16th century, whooping cough has historically caused numerous infant deaths and was a leading cause of infant mortality until recently. Efforts to develop a vaccine began soon after the identification of the causative bacterium in 1906. Although a vaccine is now available, pertussis remains a significant health threat, particularly to infants, with approximately 50 million cases and 300,000 deaths annually, making it one of the world's leading causes of vaccine-preventable deaths ([Regan et al. \[2023\]](#)). Ongoing research aims to improve vaccine efficacy and understand the mechanisms of immunity and pathogen transmission to better combat this disease.

Vaccines are biological substances designed to prepare our body to combat diseases caused by pathogens. They achieve this by stimulating the immune system to produce a response, thereby creating a memory of the pathogen and enabling the body to develop immunity to the disease ([Szwejszer-Zawislak et al. \[2022\]](#)). The vaccines currently available on the market can be categorized into four types: inactivated (such as the polio vaccine), attenuated (such as the tuberculosis vaccine), acellular (such as the pertussis vaccine), and mRNA (such as the COVID-19 vaccine). Inactivated vaccines contain pathogen cells that have been killed, typically through chemical or radiation methods. Attenuated vaccines contain live pathogen cells that have been cultured in a way that reduces their virulence. Acellular vaccines consist of fragments of pathogen cells, which, when administered, elicit an immune response. mRNA vaccines, a recent development, work by instructing human cells to produce a protein that triggers immunity against the disease ([Barbier et al. \[2022\]](#)).

Sanofi's whooping cough vaccine is an acellular vaccine containing four virulent factors or antigens: Pertussis Toxin (PT), Fimbriae (type 2 and type 3), Filamentous Hemagglu-

tinin (FHA), and Pertactin (PRN) ([Smith et al. \[2001\]](#)). The acellular vaccine is preferred to the cellular one due to adverse side effects associated to the latter. The antigens composing the vaccine are known to elicit a strong immune response in the recipient. Pertussis toxin (PTX), one of the most complex bacterial protein toxins, is exclusively produced by *Bordetella pertussis*. Fimbriae, the plural form of fimbria, are proteinaceous, hair-like structures that extend from the cytoplasmic membrane of pertussis, aiding in bacterial adhesion to abiotic surfaces and biofilm formation. Filamentous Hemagglutinin (FHA) is a surface-associated adherence protein that serves as a primary appendage of pertussis, facilitating bacterial adherence to respiratory epithelial cells ([Belcher et al. \[2021\]](#)). Pertactin (PRN) is a surface-associated protein involved in adhesion to host cell surfaces; it is autotransported through the bacterial outer membrane and undergoes proteolytic cleavage ([Ma et al. \[2021\]](#)). These virulent factors, along with others, are incorporated into Sanofi's Adacel vaccine, which serves as a booster dose against diphtheria, pertussis, and tetanus.

This thesis deals with two key areas that can aid to the improvement of bio-process productivity:

- Real-time monitoring of the process.
- Elucidation of possible sources of process variability in the Sanofi whooping cough vaccine process.

In terms of potential sources of variability, the focus is on oxidative stress following observations made in previous studies by other students in our research group.

2.2 Virulence Factor Pertactin

Pertactin is a virulence factor produced by *Bordetella pertussis*, the bacterium responsible for whooping cough. Pertactin is an outer membrane protein and a key component of the acellular pertussis vaccines. It plays a crucial role in bacterial adhesion, helping *B. pertussis* bind to the ciliated epithelial cells in the human respiratory tract. By facilitating attachment, pertactin enhances the ability of *B. pertussis* to colonize the host, making it harder for the immune system to clear the infection (Kessie [2021]).

2.2.1 Secretion Mechanism of Pertactin

Pertactin is secreted by *Bordetella pertussis* through a type of autotransporter mechanism, a subtype of the type V secretion system (T5SS). This is a widely used secretion system by Gram-negative bacteria for the export of large proteins like pertactin (Miguelena Chamorro et al. [2023]). The process can be broken down into the following steps:

- Synthesis in the cytoplasm
- Translocation across the inner membrane
- Integration into the outer membrane
- Passenger domain cleavage and display

The autotransporter pathway, specifically the type V secretion system, is one of the most common mechanisms for secreting large proteins, including virulence factors, in Gram-negative bacteria like *Bordetella pertussis*. The efficiency and simplicity of this system are

notable because it requires minimal energy expenditure and components compared to other secretion systems.

2.3 Monitoring and Control of Bio-processes

The primary objectives of the vaccine manufacturing process are to achieve high productivity, reproducibility, and product quality. These parameters are significantly influenced by the design of the bioreactor, the operating conditions, and the composition of the chemical media. Process monitoring is a crucial aspect of the pharmaceutical industry as it is essential for understanding and optimizing bioprocesses. Certain process variables, such as dissolved oxygen, pH, and temperature, are routinely monitored in bioreactors because they are easy to measure online and provide feedback signals for process control. In contrast, measurements of cell growth and productivity that provide a direct quantification of process performance are more challenging to obtain in real-time and are often conducted off-line. Offline monitoring involves the periodic or batch-wise sampling of process samples for subsequent analysis in a laboratory or analytical facility. Key bioprocess parameters such as biomass, substrate concentration and product titer are mostly done off-line. Real-time monitoring can be further divided into three main categories:

- In-line monitoring
- On-line monitoring
- At-line monitoring

In-line monitoring involves real-time measurement and analysis of process parameters directly within the bioprocess stream without disrupting or removing samples from the process (Goker et al. [2020]). Typically, this method uses a non-invasive optical probe inserted into the bioreactor for data acquisition. In-line process analytics occur continuously, where as long as the production process is ongoing, automatic measurements are repeatedly conducted, enabling automated decisions and corrective actions based on the collected data.

On-line monitoring involves acquiring measurements through a built-in process branch or bypass, where further analysis takes place. After analysis, the sample is recirculated back into the process vessel, maintaining sterility and preventing contamination (Goker et al. [2020], Wasalathanthri et al. [2020]). The design of the built-in bypass or branch (flow-cell) must consider factors such as flow rate, path length, and cell volume. This method is used when the measuring device cannot withstand the harsh conditions within the actual process vessel, such as high pressures or temperatures. Although on-line measurement setups allow for continuous measurement, they have the limitation that the samples may not perfectly replicate the conditions of the main process. However, they permit the use of measurement techniques that are not feasible with in-line monitoring. For example, reagents can be automatically added to the sample for analysis, which would be unfeasible for in-line measurement as these reagents should not enter the final product.

At-line monitoring refers to the measurement and analysis of process parameters using samples temporarily withdrawn from the bioprocess stream and analyzed nearby or in close proximity to the process location (Wasalathanthri et al. [2020]).

2.4 Fluorescence

In this research, a fluorescence-based on-line probe is developed to monitor process variables in the bioreactor that cannot be measured with existing sensors.

2.4.1 Theory of Fluorescence

Fluorescence occurs when a molecule absorbs light at a specific wavelength and transitions from its ground electronic state to an excited electronic state. This process begins when a photon is absorbed, providing the energy necessary for the molecule to move to a higher energy level. Once in this excited state, the molecule remains there for a very short duration, typically on the order of nanoseconds, before it returns to its ground state (Wasalathanthri et al. [2020]). As the molecule relaxes back to its ground state, it emits a photon of light at a longer wavelength than the absorbed light, due to the loss of some energy as heat during the relaxation process. This emitted light is what we observe as fluorescence. Typically, excitation wavelengths range from 220 to 500 nm, and the emission spectra is recorded for wavelengths between 250 and 1000 nm (Nakazawa et al. [2022]). If the emission of photons occurs slowly, the phenomenon is called phosphorescence, which can also be observed with the naked eye.

Besides fluorescence, luminescence can also occur through other mechanisms as follows:

- **Photoluminescence** occurs when emission is stimulated by photons having wavelength in visible or ultraviolet spectrum.

- **Chemiluminescence** occurs when chemical energy stimulates the emission of a photon. This also includes bioluminescence which is chemiluminescence caused due to biological reactions.
- **Electroluminescence** occurs when electrical energy or a strong electric field stimulates the emission of a photon.

For fluorescence, the difference in energy between the absorbed and emitted photons is known as the Stokes shift. This configuration is unique for each fluorophore and thus the emission spectra is also unique to each of them. The Jablonski Diagram is used to represent the energy levels that can occur for a molecule. Figure (2.1) shows a simplified Jablonski diagram. S_0 is the ground electronic state, which typifies a molecule “at rest” and also higher energy electronic states (e.g. S_1 , S_2), which is achieved upon the absorption of light. Fluorescence intensity and wavelength are influenced by the molecular structure of the fluorophore, as well as the environment surrounding the fluorophore, such as solvent polarity and pH. Because of its sensitivity and specificity, fluorescence is widely used in various applications, including biological imaging, medical diagnostics, and chemical analysis.

2.4.2 Fluorophores

Fluorophores are molecules that absorb photons of specific wavelength, and emit photons of a longer wavelength. They can be termed as molecular transducers that can transfer energy from chemical, physical, or biological events into light that can be detected and analyzed to extract information that was inaccessible otherwise ([Lichtman and Conchello](#)

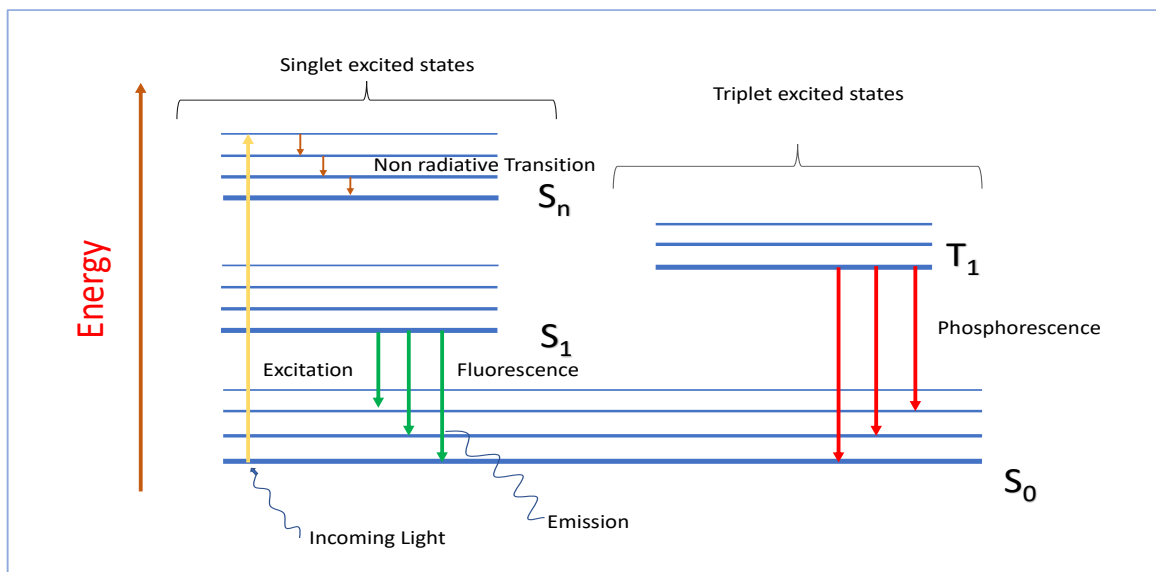


Figure 2.1: Jablonski diagram.

[2005]). Naturally occurring proteins, certain amino acids, small organic compounds or even a synthetic molecule can exhibit fluorescence phenomena. Based on their occurrence, they can be classified as intrinsic or extrinsic.

Intrinsic fluorophores are naturally present in the sample. In biological cells, naturally occurring fluorophores are amino acids such as phenylalanine, tyrosine and tryptophan, and co-factors such as reduced nicotinamide adenine dinucleotide phosphate (NADPH), reduced nicotinamide adenine dinucleotide (NADH), and the oxidized flavins: adenine dinucleotide (FAD) and the mononucleotide (FMN) (Lacowicz [2000]). Extrinsic fluorophores attach themselves to target samples and they produce a characteristic fluorescence emission based on how they react with their environment. Based on this interaction, they can be classified into three types:

- Fluorophores that show an increase in the fluorescence intensity after binding to a particular compound such as nucleic acids (e.g SYTOX or propidium iodide PI), proteins (some fluorescein derivatives such as FITC) and lipids (e.g Rhodamine 123, Nile Red, FSL-fluorescein).
- Fluorophores that fluoresce based on physiological state variables such as pH and membrane potential (example: SNARF (Seminaphtharhodafleur) dyes).
- Fluorophores whose fluorescence depends on enzymatic activity (example: Resorufin-based Substrates)

2.4.3 Fluorescence Spectroscopy

Fluorescence spectroscopy is a type of electromagnetic spectroscopic method which is used for analyzing the fluorescence from a sample (containing fluorophores) following excitation by a beam of light. All fluorescence instruments contain three basic items: a source of light to excite the molecule, a sample holder and a detector to read the emissions. The choice of the light source depends on the expected contents in the sample. The detector detects the emission from the sample for a range of wavelengths. In practice, to simplify the hardware while maintaining most information, only a small band of emitted wavelengths is examined and the incident light intensity is not kept very high to minimize the possible photo decomposition of the sample (Royer [1995b]). Most commonly used detectors in fluorescence spectroscopy are photo-multiplier tubes as they respond fast and are highly sensitive (Lakowicz and Lakowicz [1999]).

Some of the common instruments which utilize the principles of fluorescence spectroscopy are listed below.

- **Fluorescence microscopes** - They take advantage of fluorescence to generate a magnified image of the sample.
- **Spectrofluorometers and microplate readers** - that are used to estimate the average fluorescence signal of the sample.
- **Flow cytometers** - that can be used to measure the fluorescence of individual cells in a population of cells.

2.5 Fluorescence Spectroscopy for Bio-Process Monitoring

Real time monitoring of bio-processes has become increasingly important in recent years as a means of improving process control and product quality. With the establishment of the PAT initiative ([Hinz \[2006\]](#)) for innovative pharmaceutical development by the FDA, online monitoring of bio-processes has become a priority for improving process performance while meeting regulatory requirements ([Munson et al. \[2006\]](#)). From the analytical viewpoint, monitoring bioprocess variables other than pH, temperature and dissolved oxygen in real time is very challenging. Most bio-processes are operated in a sterile control volume containing a complex media involving a complex mixture of different biochemicals occurring

in different quantities. While some of these biochemicals may occur in very small quantities, they may have a large effect or they may be highly informative of the outcome of the process. Hence, small changes in the levels of these elements, that can typically occur in a production environment, may cause or reflect large variability among batches thus greatly motivating the requirement for online monitoring of culture parameters. In this regard, fluorescence spectroscopy is an attractive cost-effective option for in situ monitoring due to its inherent benefits. It is non-invasive, offering high sensitivity and it has the capability to detect multiple components even at low concentrations. Additionally, it does not consume analytes and/or additional reagents and does not require specialized sample preparation procedures (Faassen and Hitzmann [2015b]). Considering that some bioprocesses can span weeks, a fluorometer can perform measurements over an extended period without requiring recalibration.

Fluorescence spectroscopy has been applied for in-situ monitoring of bioprocesses in industrial settings, though such applications are not extensively reported. Fluorescent signals emitted by intracellular proteins (Horvath et al. [1993]) and vitamins (Hisiger and Jolicoeur [2005]) have shown promise in tracking microbial growth. Key fluorescent compounds involved in cellular growth and metabolic activity include cofactors like NADH and NADPH, fluorescent amino acids such as tryptophan, tyrosine, and phenylalanine, as well as various vitamins such as riboflavin and pyridoxine. These compounds exhibit maximum fluorescence signals at specific excitation/emission (ex/em) wavelength pairs. Consequently, using 2D fluorometry to scan the culture bulk during a bioreaction enables dynamic monitoring of all fluorophores, which directly or indirectly reflects changes in cel-

lular concentration, metabolome, and cellular environment ([Hisiger and Jolicoeur \[2005\]](#), [Surribas et al. \[2006\]](#)). Fluorescence spectroscopy is generally less expensive (almost half) than Raman spectroscopy due to simpler instrumentation and lower operational costs. It uses less costly light sources and detectors compared to those required for Raman system. NIR spectroscopy also tends to be more expensive than fluorescence spectroscopy due to the need for more complex optical components and data processing tools ([Rowland-Jones et al. \[2017\]](#)). In-situ use of fluorescence spectroscopy also offers high accuracy comparable to offline measurements, with the added benefits of increased data acquisition speed, reduced sample volume, and continuous monitoring capabilities. It's application in bioprocess monitoring provides a reliable and efficient alternative to traditional offline methods.

For example ([Teixeira et al. \[2009b\]](#)) utilized 2D fluorometry for in-line monitoring of mammalian cells in a fed-batch bioreactor system for IgG1 protein production. The monitoring was achieved by inserting a custom-developed probe into the bioreactor, which was connected to a spectrometer via optical fibers. Although the dynamics of individual fluorophores could not be directly correlated with target state variables, a chemometric approach allowed for accurate approximation of viable cell density and glycoprotein levels. Similarly, ([Graf et al. \[2019b\]](#)) implemented an LED-based 2D fluorescence spectroscopy system for inline monitoring of CHO cell cultures (total cell count, viable cell count, and cell viability) in a fed-batch setup. This new sensor reliably tracked the metabolic state of cells and quickly detected cellular distress as deviations from normal batch operation. Additionally, the spectral data offered insights into lactate metabolism shifts and cellular respiration during cultivation. The authors demonstrated that the 2D-fluorescence device is a highly sensitive tool for Process Analytical Technology applications in mammalian cell

cultures. In another instance, (Ödman et al. [2010b]) combined fluorescence spectroscopy with off-gas analysis for online monitoring of *Streptomyces coelicolor* fed-batch fermentations. Fluorescence spectra were collected using a spectrofluorometer that measured spectra through a borosilicate glass window in a recirculation loop connected to the bioreactor. A step-wise variable elimination method applied to the three-way fluorescence data resulted in simpler and more accurate N-PLS models, revealing four spectral profiles, three of which were identified as pyridoxine, NADPH, and flavin nucleotides. Similarly, (König et al. [2018]) used 2D fluorometry for inline monitoring of batch cultivations of wild-type *Escherichia coli* cultures in a 10 L bioreactor, using an immersion probe. The fluorometer measured the fluorescence of tryptophan, NADH, and FAD/FMN, fluorophores involved in the organism’s metabolism. Table (2.1) summarizes the fermentation systems, monitoring methods and applications of fluorescence spectroscopy for in-line monitoring of bioprocesses reported in literature.

While these applications highlight the potential of fluorescence spectroscopy for process monitoring, it is also important to consider the limitations that may impact its broader applicability and effectiveness. Despite many advantages, fluorescence spectroscopy is not well established in industry because some of the limitations discussed below. (Galinha et al. [2011]) observed that fluorescence signals can be highly sensitive to changes in bioreactor conditions such as pH, temperature, and ionic strength. These variations can affect the fluorescence intensity and spectra, complicating the interpretation of results and requiring stringent control of experimental conditions. Sometimes, the components of the bioreactor can exhibit autofluorescence which can interfere with the fluorescence of other

fluorescent components. Techniques like synchronous fluorescence spectroscopy and lock-in amplification have been developed to reduce background noise, but these add complexity and cost to the system (Teixeira et al. [2011b]). Sometimes, prolonged exposure to the excitation light can cause photo-bleaching, where the fluorescent molecules lose their ability to fluoresce. This results in a decrease in signal over time, which can compromise long-term monitoring and require frequent recalibration or replacement of fluorescent probes. Fluorescence spectroscopy uses intrinsic fluorescence of compounds that are present and are part of organisms' metabolic-pathways. Notably, most mammalian bioprocesses now rely on serum-free and protein-free media, although they do contain free amino acids and vitamin blends, which emit fluorescence. Unlike microbial fermentations, where cellular growth can be linked directly to the fluorescence of amino acids or vitamins, mammalian cells cannot produce these vitamins or fluorescent amino acids. Thus, analyzing the entire spectra using multivariate chemometric techniques becomes necessary in these scenarios (Teixeira et al. [2009b]).

2.6 Oxidative Stress in Bacterial Cells

Oxidative stress in bacteria arises from an imbalance between the production of reactive oxygen species (ROS) and the cells' ability to detoxify these harmful intermediates or repair the resultant damage. ROS, which include superoxide anions, hydrogen peroxide, and hydroxyl radicals, can originate from endogenous sources such as aerobic respiration and specific enzyme activities, or from exogenous sources like UV radiation, pollutants, and host immune responses (Bayr [2005]). These reactive molecules can inflict significant

Table 2.1: Applications of fluorescence spectroscopy for in-line monitoring of bioprocesses.

Fermentation system	Application	Monitoring method	Multivariate analysis method	References
Fermentation of <i>S. coelicolor</i> bacteria in 1.5 L fed-batch bioreactor system	Monitoring and prediction of cell dry weight and amino acids	In-line	PLS, locally weighted regression, and multilinear PLS (N-PLS)	Ödman et al.
Fed-batch fermentation of CHO cells in 2-L bioreactors	Monitoring of viable cell density and glycoprotein	In-line	PLS regression	Teixeira et al.
Batch cultures of CHO cell lines in 1-L bioreactor	Monitoring of cell viability and antibody concentration	In-line	PLS Regression	Teixeira et al.
Batch fermentations of <i>E. Coli</i> in 10 to 30 liter bioreactors	Monitoring of cell dry weight and biogenic fluorophores (tryptophan, NADH and riboflavin)	In-line	PCA, PLS regression	König et al.
Fed-batch cultivation of mammalian (CHO) cells in 10 Liter Bioreactors	Monitoring the concentration of BSA, FAD, NADH, NAD, pyroxyline and riboflavin	In-line	PCA, PLS regression	Graf et al.
Fed-batch cultivation of <i>Pichia Pastoris</i> cultures in a 14 liter fermentor	Monitoring of cell density, protein production, glycerol and methanol consumption	In-line	Multi-variate curve resolution method	Karakach et al.
Fed-batch cultivation of <i>B. Pertussis</i> cultivation in a 2 liter fermenter.	Monitoring and prediction of biomass, carbon source (glutamate) and antigen (pertactin) productivity	In-line	PLS regression	Mishra et al.
Cultivation of microalgae <i>Scenedesmus</i> AMDD in 300 liter continuous photobioreactor	Monitoring of cell growth and proteins	In-line	Multivariate curve resolution, PCA	Karakach et al.

damage on bacterial cells by modifying DNA, oxidizing proteins, and peroxidizing lipids, which compromises cellular integrity and function. The implications of oxidative stress are profound, influencing bacterial pathogenesis, antibiotic resistance, and biotechnological processes. While ROS production is a key host defense mechanism against pathogens, some bacteria have adapted to exploit oxidative environments to enhance their virulence ([Reniere \[2018\]](#)). Additionally, the ability to manage oxidative stress can affect the efficacy of certain antibiotics and the productivity of industrial fermentation processes. Understanding these dynamics offers valuable insights into bacterial physiology and can drive the development of new therapeutic strategies and biotechnological advancements.

To counteract the damage from ROS, bacteria have evolved various defense mechanisms involving enzymes like superoxide dismutase (SOD) and catalase, small molecule antioxidants such as glutathione, and robust DNA repair systems. Regulatory proteins, like OxyR and SoxR, play a critical role in sensing oxidative stress and up-regulating protective genes

(Cabisco et al. [2000b]). SODs convert the superoxides into hydrogen peroxide (H_2O_2), which is another major ROS species in the cell. The primary antioxidants which quench hydrogen peroxide are catalase and glutathione peroxidase (GPx) (Cabisco et al. [2000b]). Catalase is a very effective antioxidative enzyme and can not be saturated by (H_2O_2) at any concentration (Hong et al. [2019]). It converts (H_2O_2) into H_2O and molecular oxygen. GSH (Glutathione) system is another important anti-oxidative mechanism inside the cell which involves GSH, glutathione peroxidase and glutathione reductase, and it catalyzes the reduction of peroxides as per the reactions shown below. A critical component in these defense mechanisms is NADPH (Nicotinamide Adenine Dinucleotide Phosphate), which acts as a reducing agent. NADPH is essential for the regeneration of antioxidants like glutathione and thioredoxin, providing reducing power for detoxifying enzymes, and maintaining the redox balance of the cell (Nguyen et al. [2017]). It also supports biosynthetic reactions necessary for repairing and synthesizing cellular components during oxidative stress. Thus, NADPH plays a pivotal role in mitigating ROS damage while maintaining cellular integrity and growth, allowing bacteria to survive and adapt to oxidative environments. Oxidative stress has also been found to be correlated to NADPH production which in turn tends to diminish oxidative stress by facilitating reactions that promote a reducing environment (Lushchak [2001b]).



Glutathione reductase then reduces back oxidized glutathione (Glutathione Disulfide) or GSSG to reduced glutathione (GSH). Thioredoxin reductase is another mechanism for

countering H_2O_2 , which utilizes NADPH to catalyze the conversion of oxidized TRX into reduced TRX.

Among different sources of ROS, Excessive nutrients can induce oxidative stress in both mammalian cells and bacteria (Wellen and Thompson [2010c], Wang and Levin [2009b]). In a cell, the nutrients are oxidised in the tricarboxylic acid cycle (TCA) cycle to produce NADH and FADH_2 . These co-factors donate electrons to the electron transport chain (ETC) where they are used by molecular oxygen to form H_2O . Excess nutrient consumption in the TCA cycle can overwhelm the electron transfer chain (Wellen and Thompson [2010c]), causing excess ROS buildup inside the cell resulting in the generation of oxidative stress. This excess ROS is associated to incomplete reduction of oxygen in the ETC thus resulting in overflow of electrons which produce super-oxides in the presence of oxygen.

2.7 Principles of Flow Cytometry

Flow cytometry is a powerful analytical technique used to measure the physical and chemical characteristics of cells or particles as they flow in a fluid stream through a beam of light, typically a laser. The fundamental principles of flow cytometry involve the passage of individual cells or particles through a focused light source, where they scatter light and emit fluorescence that can be detected and quantified (Adan et al. [2017]). Forward scatter (FSC) and side scatter (SSC) are the primary parameters used to assess cell size and internal complexity, respectively. Fluorescent dyes and antibodies conjugated with fluorochromes can be used to label specific cellular components or markers, allowing for the simultaneous multiparametric analysis of various cellular properties (McKinnon [2018]).

The emitted fluorescence is collected by detectors, converted into electronic signals, and processed by a computer to generate data on cell populations, including size, granularity, and the presence of specific biomarkers. This enables detailed analysis of heterogeneous cell populations, making flow cytometry invaluable in research, clinical diagnostics, and various applications in biotechnology. The elements common in all flow cytometers are:

- A light source and a device to focus that light.
- A fluid line and controls to direct the liquid stream containing cells to pass through the light beam.
- A detector to detect the reflected or emitted light from the cells.
- A computer to process the signals coming from the electronic detector and to analyze it.

2.7.1 Multi-colour Flow Cytometry

Multi parameter flow cytometry is a technique used to identify defined cell types within a mixed cell population ([Adan et al. \[2017\]](#)). Using this method, it is possible to take simultaneous measurements of multiple parameters based on light scattering or fluorescence from molecules inside/on the surface of the cell. For example, intracellular ROS and NADPH within a cell population can be quantified simultaneously by multi-parameter flow cytometry using suitable dyes. Multi-colour combinations that can be monitored are determined by type and number of lasers and filters in a cytometer. These cytometers are

commonly available with multiple laser wavelengths such as red (633-nm), blue (488-nm), and violet (405-nm) laser. New machines also have lasers that can emit light at 561 nm, 532 nm, and 375 nm (ultraviolet). With more lasers available, it is possible to isolate many defined cell subsets for population analysis ([Adan et al. \[2017\]](#)). Figure (2.2) shows the schematic of a flow cytometer with multiple lasers and fluorescence channels which can detect the emissions from more than one fluorochrome. The parameters of interest can then be plotted as density plots in which each axis represents a particular marker that the sample had been stained for and can be used for further analysis.

2.8 Column Chromatography

Column chromatography is a fundamental and versatile technique used in the separation and purification of chemical as well as biological compounds. It is widely employed in chemical laboratories and industries for its ability in isolating individual components from complex mixtures. The technique operates on the principle of differential adsorption, where compounds in a mixture are separated based on their varying affinities for the stationary phase and the mobile phase ([Coskun \[2016\]](#)).

In column chromatography, a column is packed with a stationary phase, typically a solid adsorbent like silica gel or alumina. The mixture to be separated is dissolved in a suitable solvent and introduced into the column. As the mobile phase (the solvent or eluent) passes through the column, the different components of the mixture interact with the stationary phase to varying extents. Components with higher affinity for the stationary phase move more slowly, while those with lower affinity travel faster through the column. This results

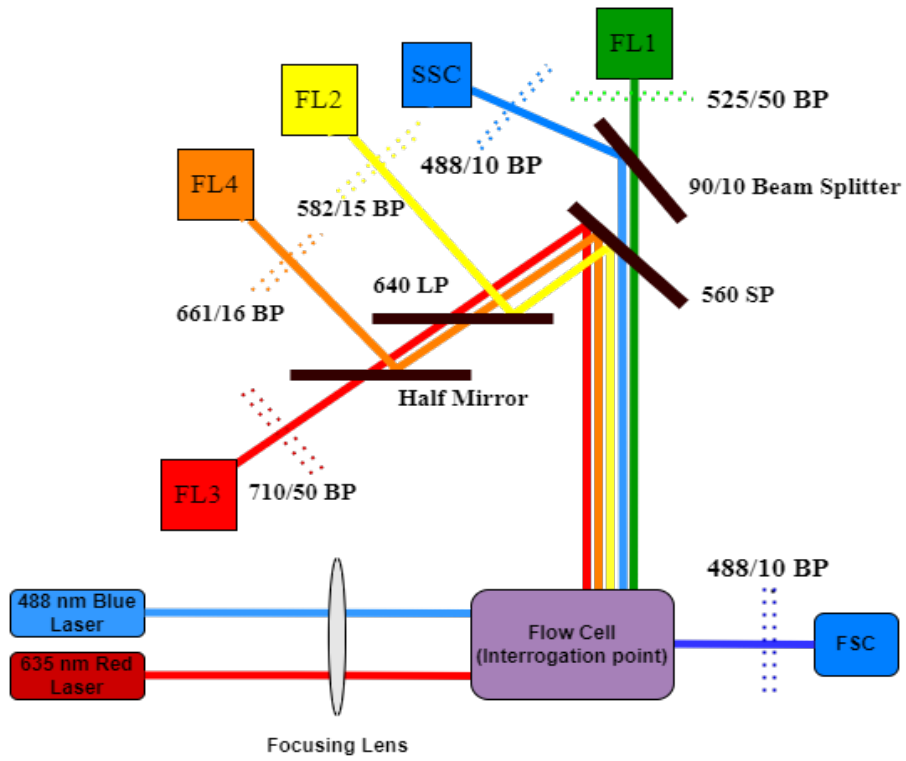


Figure 2.2: Schematic representation of a flow cytometer. BP: band pass, LP: long pass, SP: short pass. Half mirror (or beam splitter) passes 50 percent and reflects 50 percent of the light. 90/10 beam splitter reflects 90 percent and passes 10 percent of the light (taken with modifications from (da Silva et al. [2012])).

in the separation of the components as they elute at different times. The effectiveness of column chromatography is influenced by various factors including the choice of stationary phase, mobile phase, and the conditions under which the process is carried out. It is widely used in the purification of natural products, the isolation of proteins, and the analysis of complex mixtures in research and industrial settings.

2.8.1 Affinity Chromatography

Affinity chromatography is a specialized form of column chromatography that exploits the specific interactions between a target molecule and a ligand attached to the stationary phase. This technique is particularly powerful for the purification of biomolecules such as proteins, enzymes, antibodies, and nucleic acids due to its high specificity and efficiency. When a mixture containing the target molecule is introduced into the column, the target molecule binds selectively to the ligand due to these specific interactions, while non-target molecules pass through the column ([Hage \[1999\]](#)). In affinity chromatography, the stationary phase is typically a porous matrix (like agarose or sepharose) to which a specific ligand is covalently attached. The ligand is chosen based on its specific and reversible interaction with the target molecule. Common ligands include antibodies, antigens, enzymes, substrates, and cofactors. The column is then washed with a buffer to remove any non-specifically bound substances, ensuring that only the target molecule remains attached. To elute the bound target molecule, the conditions are altered, such as changing the pH, ionic strength, or introducing a competitive ligand, which disrupts the interaction between the target molecule and the ligand. The target molecule is then released from the stationary

phase and collected. This method's high specificity and efficiency allow for the purification of a single target molecule from complex mixtures, making affinity chromatography a powerful tool in biochemical research and pharmaceutical production.

In this work, initially, enzyme-linked immunosorbent assay (ELISA) was utilized for the purification of pertactin (PRN) antigen. The reason ELISA was replaced by affinity chromatography was because results from ELISA varied a lot as it is very sensitive and is susceptible to cross-reactivity. Cross-reactivity occurs when unintended molecules bind to the detection system, leading to false positives or skewed results. In ELISA, this can happen if antibodies used in the assay recognize not only the target protein but also other proteins with similar epitopes, especially in samples containing a mixture of proteins (Hosseini et al. [2018]). Affinity chromatography allows for highly selective binding and isolation of target proteins, resulting in a product with high purity.

2.9 Multivariate Statistical Analysis

Multivariate statistical analysis is a useful tool for understanding complex data sets which requires simultaneous examination of all variables. It provides the means for both describing and exploring data to extract the underlying patterns and structure.

2.9.1 Data Pre-Processing

Before using any data for multivariate statistical analysis, it is necessary to pre-process data for the purpose of regression analysis. Some of the commonly used methods for data pre-processing are:

- **Mean centering** the mean of each variable of dataset is calculated and subtracted from each datapoint of that variable.
- **Median Centering:** similar to mean centering but instead of the mean, the median of the data is used.
- **Scaling:** mean/median centering is conducted and followed by division of each datapoint in a column with the column-based standard deviation. After applying this approach, the mean and standard deviation of a pre-processed variable are zero and one respectively.

2.9.2 Principal Component Analysis

Principal component analysis, or PCA, is a statistical pre-treatment approach that is applied to analyze data that exhibits high correlation among the variables. It is one of the most popular multivariate statistical technique and it is used in almost all fields of science. PCA analyzes and compresses a data table with several correlated dependent variables by projecting the original data onto a set of new orthogonal variables, (non-correlated) referred to as latent variables. PCA can also be used to identify similarity between observations

by displaying them as points in 2D or 3D maps thus facilitating visualization (Abdi and Williams [2010]). Figure (2.3) shows the axes (principal components) along which the data points are projected. This axis is chosen by maximizing the distance among the projected points and at the same time, minimising the total distance from the original points to their corresponding projected points. Thus, PCA generates a representation that is as close to

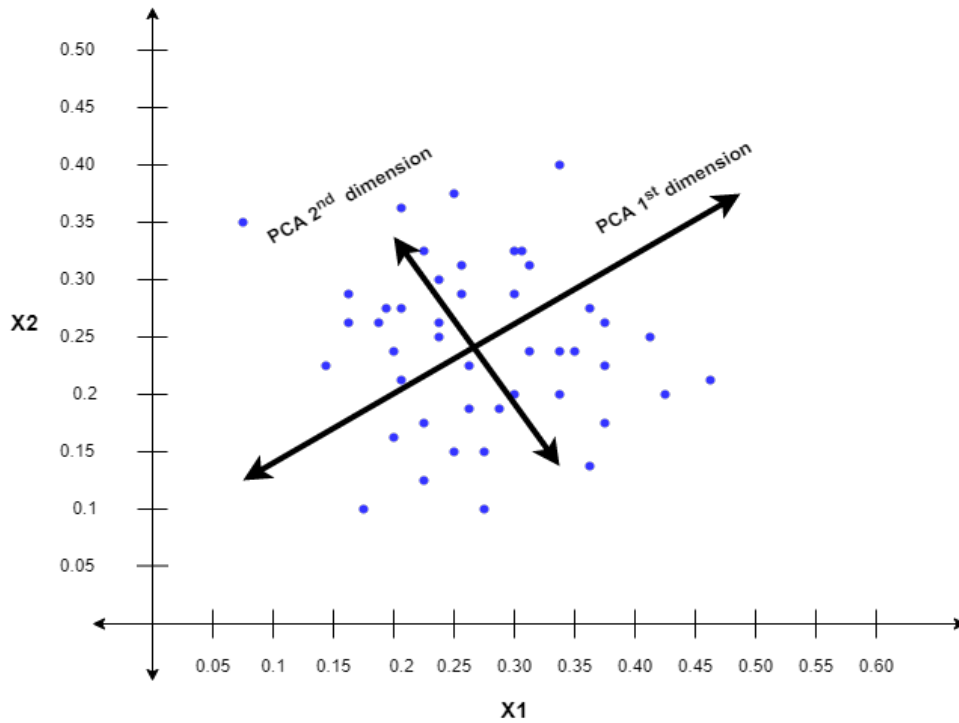


Figure 2.3: Principal components in PCA.

the original data as possible but the data is described along a reduced number of orthogonal directions to reduce sensitivity to noise. Equation (2.2) shows the generalised equation for PCA where the table of data can be represented as the summation of the products of S_i and P_i , and a residual matrix E that contains the unexplained variability in the data. This unexplained variability corresponds to the space spanned by the components that are

neglected in the PCA model.

$$X = \sum_{n=1}^n S_i * P_i + E \quad (2.2)$$

where n is the number of samples, S_i denotes the scores that contain information about the relation between the original samples and the principal components, and P_i represents the loadings that contain information about the contributions of the original variables to a principal component. In summary, PCA can extract strong patterns from large and highly correlated datasets by capturing the essence of the data in the space of a few principal components.

2.9.3 Partial Least Squares Regression

Partial least squares regression or PLSR is a regression method that is used to predict a set of dependent variables from a very large set of independent variables (predictors). Partial least squares projects the input data into latent variables (LVs) that have most effect on the outputs. The decomposition of matrix X (predictor) and matrix Y (response) can be described by the following equations.

$$X = TQ + E = \sum_{n=1}^m t_i * q_i + E \quad (2.3)$$

$$Y = UC + F = \sum_{n=1}^m u_i * c_i + F \quad (2.4)$$

Here, m represents the number of latent variables. Vectors t_i and u_i represent the scores. These scores contain information about the relation between the original samples (X and U respectively) and the latent variables. Vectors q_i and c_i are called loadings and they contain information on how T relates to X , and U relates to Y . E and F contains variables which go unexplained by the model. The algorithm maximizes the covariance between T and U thus ensuring that the projected input variables are most informative about the projected output variables.

2.10 Mechanistic Models

A mechanistic model is a type of mathematical model that aims to represent the underlying biological, chemical, or physical processes that drive a system. Unlike empirical models, which rely on fitting data to equations without necessarily understanding the underlying mechanisms, mechanistic models are built upon known principles and theories such as conservation of mass, energy, and momentum, as well as biochemical kinetics and thermodynamics. These models offer a detailed representation of systems, often at a molecular or cellular level, capturing dynamic interactions and feedback mechanisms. By incorporating these fundamental processes, mechanistic models have significant predictive power, allowing them to forecast system behavior under various conditions, including those that are not explicitly considered in the initially available data. The parameterization of these models involves parameters with physical or biological significance, which can often be measured experimentally or inferred from literature. Due to their comprehensive nature, mechanistic models can be quite complex, involving numerous equations and parameters ([Baker et al.](#)

[2018]).

Current literature shows that bioreactor monitoring have increasingly emphasized the integration of mechanistic models to enhance process understanding and control. These models, which are based on the fundamental principles of biological and chemical processes, offer a predictive framework that can simulate the complex dynamics within bioreactors. Recent advancements have focused on coupling mechanistic models with real-time data from advanced sensors and monitoring systems, enabling more accurate predictions of bioprocess behavior under varying conditions. This hybrid approach not only improves the precision of bioprocess control but also supports the development of digital twins i.e virtual replicas of bioprocesses that facilitate optimization and troubleshooting in-silico before modifications are made in the physical system. Additionally, there is a growing trend towards the use of mechanistic models in conjunction with machine learning techniques, allowing for enhanced model calibration and the handling of complex, nonlinear data. These innovations are driving more efficient, scalable, and robust bioprocesses, particularly in the fields of biotechnology and industrial microbiology. For example, (Golabgir and Herwig [2016]) combined a mechanistic model with a spectroscopic method (RAMAN) for the real time monitoring of fed-batch penicillin production. In that study, chemometric methods (PLS) were utilized for model building based on Raman measurements. then, a hybrid model based on the combination of the PLS model with a mechanistic model was used to predict biomass and penicillin productivity. The key advantage of using a hybrid model for process monitoring is that the resulting solution space is constrained by material balances, thus providing better extrapolation results for regions of the variable space

that were not used for model calibration. Thus, confining the solution space by satisfying a mechanistic model generally enhances the model’s predictive capabilities (Tsopanoglou and del Val [2021]). This framework can be used for spectroscopic methods (fluorescence, NIR, RAMAN) as well as statistical models (artificial neural networks, support vector machines etc.) as done in the current thesis.

In (Li et al. [2021]) a mechanistic model was developed to describe the dynamics of biomass, lactose, and lactate concentrations during the fermentation process. The model incorporated kinetic equations that account for the consumption of lactose by biomass, the production of lactic acid, and the subsequent impact on pH levels. The primary focus was on accurately predicting the pH as it directly influences the quality of the cream cheese. The mechanistic model served as the foundation for generating state variables (biomass, lactose, and lactic acid concentrations) over time. These state variables were then used as inputs to a Long Short-Term Memory (LSTM) neural network, a type of machine learning model, to predict the pH change during fermentation. This hybrid approach of combining a mechanistic model with a machine learning based model provided accurate predictions of fermentation end-time, making the process more predictable and efficient. In (González-Hernández et al. [2022]), authors used a mechanistic model to simulate the metabolic behavior of *Saccharomyces cerevisiae*, particularly its ability to switch between aerobic fermentation, anaerobic fermentation, and respiration based on glucose and oxygen availability. The model incorporates key metabolic pathways, calibrated and validated using experimental data, to predict metabolite concentrations over time. This model is then applied to optimize fermentation processes in industrial settings. In (Mears et al. [2016]), the authors used the mechanistic model to pilot scale filamentous fungal

fermentation systems for on-line application. A state estimator model was then developed based on stoichiometric balances in order to predict the biomass and product concentration.

Mechanistic models can be also used in combination with sensors within optimal on-line estimators such as Extended Kalman Filter (EKF) to enhance predictive accuracy. EKFs are particularly effective in estimating internal system states that are not directly measurable in real-time based on process data ([Dewasme et al. \[2009\]](#)). This combination is crucial for handling the inherent nonlinearity in bioprocesses, as EKF is based on a linearization of the nonlinear mechanistic models around current estimates. An EKF approach was used in the current thesis in combination with spectro-fluorescence measurements as discussed in Chapter (6) of this thesis.

Overall, mechanistic models are valuable in bioprocess monitoring for their detailed representation of biological, chemical, and physical processes, with few notable disadvantages. However, as these models are complex and computationally demanding, they often require a large dataset for calibration and validation. They are also sensitive to parameter values, meaning small errors can lead to significant deviations in predictions. The reliance on assumptions and simplifications can introduce inaccuracies, and once developed, these models are often rigid, making them less adaptable to changing process conditions. Additionally, the development of mechanistic models is time-consuming and requires specialized expertise, which can be a barrier for some organizations. Furthermore, these models may struggle to fully capture the complexity of biological interactions, limiting their effectiveness in dynamic environments. The EKF partially addresses these disadvantages since it updates estimates of process disturbances in real-time thus addressing potential model

inaccuracies.

2.11 Akaike Information Criteria

The Akaike Information Criteria (AIC), was used in chapters (4) and (6) for model comparison and selection. Introduced by Hirotugu Akaike in 1974, it is a widely used method for model selection in statistical and econometric applications. AIC helps identify the model that best balances goodness of fit with model complexity, thereby preventing overfitting. The AIC value for a given statistical model is calculated using the formula:

$$AIC = 2k - 2\ln(L) \quad (2.5)$$

where k is the number of parameters in the model and L is the likelihood of the model with respect to given the data. The formula comprises two components: the log-likelihood, which measures the fit of the model to the data, and the penalty term, which accounts for the number of parameters thus discouraging overly parametrized models (Vrieze [2012]). A lower AIC value indicates a better trade-off between fit and complexity, with the preferred model being the one with the lowest AIC among those compared.

2.12 Extended Kalman Filter

The extended kalman filter (EKF) was used in chapter (6) to develop a hybrid model using inputs from PLS regression model and mechanistic model, for state estimation. It is the

non-linear version of the Kalman filter which is a very popular estimator in discrete-time linear systems (Fujii [2013]). Using this algorithm, it is possible to estimate the dynamic evolution of the of a non linear system, given some initial information about noise and process disturbances. The main advantage of using EKF is that it only stores the states at previous time step and is very fast, which makes it well suited for real time state estimation. The formulation of EKF is described by the following equations:

$$x_k = f(x_{k-1}, u_k) + w_k \quad (2.6)$$

$$z_k = h(x_k) + v_k \quad (2.7)$$

where w_k and v_k represent the process and measurement noises which are assumed to have zero mean and they follow a multivariate Gaussian distribution. The function f predicts the states from the state estimated at previous time point and the function h is used to predict the measurement from the predicted state. While for linear systems, a Gaussian distribution at time k propagates onto another Gaussian distribution at time $k+1$, this is not true for non-linear functions. For simplicity, EKF makes a Gaussian approximation for joint distribution of state x and measurement z using a Taylor series transformation. Accordingly, at each time step, the non-linear system is linearized thus justifying the preservation of Gaussian distributions over time.

Chapter 3

Materials and Methods

This chapter details the experimental procedures and methodologies utilized to develop online estimators and enhance the productivity of *Bordetella pertussis* cultures through lab-scale experiments and Process Analytical Technology (PAT) techniques.

3.1 *Bordetella pertussis* Flask Cultures

Flask cell cultures were grown from frozen *Bordetella pertussis* seeds (B. pertussis wild type) in 500 ml polycarbonate, sterile flasks with a vented cap equipped with PTFE membrane having a 0.22 μm pore size from VWR, Canada, and a working volume of 120 ml. Due to the pathogenic nature of *Bordetella pertussis*, all experiments were conducted under bio-safety level 2 containment. The cultures were grown in modified Stainer-Scholte broth (cP media) supplemented with a mix of amino acids and salts as growth factors. Detailed media composition is not provided due to confidentiality. The incubation conditions were

set at 36°C with shaking at 200 RPM. The initial pH of the flask culture was adjusted to 7.2. Cultures were grown for a duration of 48 hours, with samples collected at 6, 12, 18, 24, 30, 36, 40 and 48 hours.

Flask experiments were conducted to study the impact of oxidative stress due to high initial glutamate concentration. They were also used to inoculate the *B. pertussis* bioreactor runs.

3.2 Batch Fermentation of *B. pertussis* in Bioreactors

Experiments were performed in 2-liter bioreactors (Biostat B plus, Sartorius) with a working volume of 1.6 liters. Bioreactor cultures were initiated, with cP media (control) and 20% higher glutamate concentration. The bioreactor runs lasted for 32 hours, with the temperature maintained at 36°C and the pH controlled at 7.2, similar to the initial conditions implemented in flask. Dissolved oxygen (DO) concentration was kept at 35% (of saturation value) by sequentially adjusting agitation and aeration. Initial agitation and aeration rates were set at 200 rpm and 2 liters per minute, respectively. Upper limit of agitation was set to 700 rpm. Batch fermentations were performed to study the impact of oxidative stress due to high initial glutamate concentration, on growth and antigen productivity of cultures.

3.2.1 Fed-Batch Runs

Fed-batch runs were performed to study the impact of oxidative stress on pertactin antigen productivity of *B. pertussis* cultures and to validate the mechanistic model based on the impact of oxidative stress on *B. pertussis* fermentations. Feed was started when the glutamate in the culture broth was almost consumed by the cells which was marked by an observed stall or slowing down of agitation due to lesser consumption of oxygen by the bacterial cells. Feed rate of the supplement media was around 18 ml/hr. Usually, all the glutamate was consumed after 28 hours of batch run. The agitation rate was kept between 200 to 600 rpm (Rushton impeller). A silicon-based antifoam was manually added to the bioreactors as per the batch requirement. Formulation of culture media, growth factors and supplement is discussed in the next section.

3.2.2 Preparation and formulation of media, growth factor and supplement

For starting a bacterial culture, 2X concentrated media and 100X concentrated growth factor was prepared and then diluted with sterilized water. The normalized composition of media, growth factor and supplement is tabulated in table 1, 2 and 3 respectively. Normalization is done for confidentiality.

Table 3.1: Culture media components and their normalized composition for 2X media concentration.

Sr. No.	Media Components	Concentration (Normalized)
1)	Mono Sodium Glutamate	1
2)	Sodium Chloride	0.25
3)	Potassium Phosphate Monobasic	0.05
4)	Magnesium Chloride Heptahydrate	0.01
5)	Potassium Chloride	0.02
6)	Tris Base	0.15
7)	Casamino Acid	1

Table 3.2: Growth factor components (100X concentrated) and their normalized composition.

Sr. No.	Growth Factor Components	Concentration (Normalized)
1	L- Cystine HCl	0.04
2	Nicotinic Acid	0.004
3	Ascorbic Acid	0.4
4	Reduced Glutathione	0.15
5	Heptakis Beta Cyclodextrin	1
6	FeSO ₄ .7H ₂ O	0.01
7	CaCl ₂ .2H ₂ O	0.02

Table 3.3: Supplement components and their normalized composition.

Sr. No.	Supplement Components	Concentration (Normalized)
1	Mono Sodium Glutamate	1
2	L-Cystine HCl	0.005
3	Reduced Glutathione	0.0187
4	FeSO ₄ .7H ₂ O	0.00125
5	Nicotinic Acid	0.0005
6	Ascorbic Acid	0.05

3.3 Selective Plating for *B. pertussis* Identification

Samples from bacterial cultures were taken and streaked on Bordet-Gengou (BG) agar and Tryptic (Trypticase) Soy Agar (TSA) plates under sterile conditions to check for contamination. Bordet-Gengou (BG) agar was utilized as it supports the growth of *B. pertussis*. In contrast, Tryptic (Trypticase) Soy Agar (TSA) plates were used to ensure that *B. pertussis* does not grow on this medium, thus helping to confirm the selectivity of the BG agar. Samples were plated on both BG agar and TSA plates. The growth observed on BG agar and the absence of growth on TSA plates were used as criteria to identify and confirm the presence of *B. pertussis* in the samples and lack of contamination. Figure (3.4) shows contaminated as well as non-contaminated TSA plates along with pertussis growth on BG agar plate.



Figure 3.1: BG agar plate with *B. pertussis* growth.



Figure 3.2: Contaminated TSA plate.

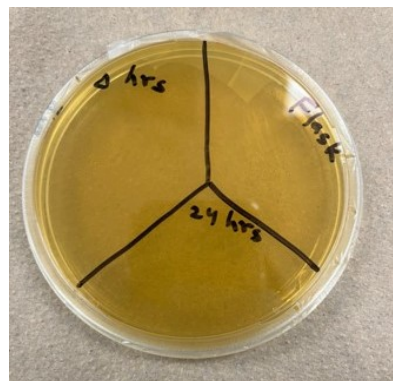


Figure 3.3: TSA plate without contamination.

Figure 3.4: BG agar and TSA plates to identify and confirm contamination and pertussis growth.

3.4 Measurement Techniques

3.4.1 Optical density (OD)

Samples were taken from the shake flask culture and bioreactor and diluted 20 times using a 0.9 saline solution. The optical density (OD) was then determined by measuring the absorbance at 600 nm with an optical spectrophotometer (Genesys, thermo-fisher).

3.4.2 Amino acid analysis using HPLC

The amount of glutamate in culture broth was quantified using a Waters 1525 chromatography system (Waters Corporation, MA, USA) equipped with a fluorescence detector (W2475; Waters Corporation, MA, USA). The method utilized a Waters AccQ.Tag™ Amino Acid Analysis Column C18 (4 μ m, 3.9 mm $\times$ 150 mm) set at 38°C. For each anal-

ysis, a sample volume of 5 μL was injected. The fluorescence detector was configured with an excitation wavelength of 248 nm and an emission wavelength of 395 nm. Eluent A consisted of an acetate-phosphate buffer solution prepared by mixing 100 mL AccQ.TagTM. Eluent A concentrate (Waters, MA, USA) with 1 liter of Millipore water. Eluent B was HPLC-grade acetonitrile, and eluent C was Millipore water. The flow rates of each eluent at various times during the HPLC run were adapted from Cohen’s work (Cohen [2000]) and are detailed in Table (3.4). Each HPLC analysis lasted for 40 minutes. Before each run and between sample runs, the column was equilibrated for 10 minutes using 99 % of eluent A and 1 % of eluent B at a flow rate of 1 mL/min.

Table 3.4: Gradient schedule for the HPLC runs using a ternary eluent system.

Time (Minutes)	Flow Rate (ml/min)	Eluent A (Buffer) %	Eluent B (Acetonitrile) %	Eluent C (Water) %	Gradient Curve
0	1	99	1	0	-
0.5	1	98	2	0	6
18	1	95	5	0	6
19	1	91	9	0	6
29.5	1	80	20	0	6
33	1	0	60	40	11
40	1	99	1	0	11

3.4.3 Glutamate quantification using assay kit

Glutamate concentration was also assessed at various time intervals using a fluorometric glutamate assay kit (Abcam, catalog number ab138883, Canada). A calibration curve for fluorometric quantification was established using glutamic acid stock solutions included

with the kit. Samples were filtered through 0.2 μm syringe filters, and the assay was performed according to the kit's protocol. A 25 μL aliquot of each sample was combined with 25 μL of the working reagent in a 96-well transparent microplate and incubated at room temperature for 60 minutes. Fluorescence (Ex/Em - 540/590 nm) was measured using a Biotek Gen5 microplate reader (Agilent Technologies, 2020, Canada). Measurements were repeated and the results were averaged.

3.5 Enzyme-Linked Immunosorbent Assay (ELISA)

Pertactin concentration was measured using ELISA with a Biotek Plate Reader Synergy H1 Mono RDR (2019). Microplates were coated with 50 μL /well of PRN polyclonal antibodies (1.5 $\mu\text{g}/\text{mL}$) purified from guinea pig blood and incubated for 20 hours. Following incubation, the plates were washed three times with 200 μL /well of a solution containing 10 mM PBS, 0.1 % Tween-20, and 0.1 % BSA. The plates were then blocked for 30 minutes at room temperature using 100 μL /well of 10 mM PBS and 0.1% BSA. After blocking, samples were centrifuged, filtered, diluted, and added to the wells for incubation. The plates were subsequently washed again with the PBS-Tween-BSA solution. Then, 100 μL of horseradish peroxidase-conjugated guinea pig PRN antibody was added to the plates, followed by a 1-hour incubation at room temperature. To stop the reaction, 50 μL of 2 N H_2SO_4 was added, and absorbance was measured at 450 nm. A standard curve was used to calculate PRN concentration, with absorbance being proportional to the PRN concentration.

3.6 Chromatographic Purification of Pertactin

Purification of pertactin was achieved using 1 ml affinity columns. The method is elaborated in the following subsections.

3.6.1 Ligand Coupling and Column Preparation

Affinity chromatography using a custom peptide (sourced from Bachem, Torrance, CA, USA) was employed to quantify pertactin in the supernatant. The protocol, originally developed by (Umatheva et al. [2020a]) for pertactin quantification, was modified for protein purification of samples from the fermentation broth. The amino acid sequence of the peptide is confidential. The peptide's primary amine ($-NH_2$) was coupled to N-hydroxysuccinimide (NHS) activated resin prepacked in a 1 mL column (Cytiva, catalogue no. - 17090601, MA, USA) following the manufacturer's instructions. The peptide was dissolved in a coupling buffer (200 mM $NaHCO_3$, 0.5 M NaCl, 0.1 % DMSO (v/v), pH 8.3). Prior to peptide loading, the column was washed with 6 column volumes (CVs) of 1 mM cold HCl at a flow rate of 0.5 drops/sec using a syringe. The coupling buffer containing the peptide was then passed through the column, allowing the peptide to react with the resin at room temperature for 4 hours. To remove and block any unreacted ligand, the column was washed with 6 CVs of 0.5 M ethanolamine (Thermo Fisher, Mississauga, Canada) in 0.5 M NaCl (pH 8.3) at a flow rate of 1 mL/min. Subsequently, the column was washed with 6 CVs of sodium acetate (Sigma Aldrich, Oakville, ON, Canada) in 0.5 M NaCl (pH 4.0) at the same flow rate. This blocking and washing sequence was repeated, and finally, the column was stored at 4 °C in 20 % ethanol.

3.6.2 Pertactin Purification

Chromatic purification of pertactin was performed using AKTA AVANT™ 25 (GE health-care, USA). Broth samples, filtered through a 0.2 μm syringe filter and collected at various time points during the culture, were used for protein purification. The column was initially washed with Milli-Q water and then equilibrated with 6 column volumes (CVs) of equilibration buffer A (50 mM Tris, pH 7.5). Next, the column was loaded with 10X concentrated feed material. After sample loading, buffer A was introduced into the column. The column was then washed with a wash buffer containing 50 mM Tris (pH 7.5), 0.1 % (v/v) Tween-80, and 1.0 M NaCl (Sigma Aldrich). Elution was performed using an eluting buffer (100 mM sodium citrate, pH 3.5). 1 ml sample of the eluate was collected and immediately neutralized with 0.125 mL of Tris-HCl (pH 8.0) to prevent pertactin denaturation. Figure (3.5) shows the elution profile and protein peak (during elution) for a sample run on AKTA.

3.7 SDS PAGE Analysis

Gel electrophoresis or SDS PAGE analysis was done to validate the pertactin binding by the column. The wash and eluate fractions collected from the column were analyzed using pre-cast 12% Bis-Tris SDS protein gels (Bio-Rad). Wash and eluate samples, along with pure pertactin (provided by Sanofi), were prepared according to the manufacturer's instructions under reducing conditions (with β -mercaptoethanol) and loaded into the gel wells. The gel was stained overnight with Coomassie blue stain (Bio-Rad) solution, which

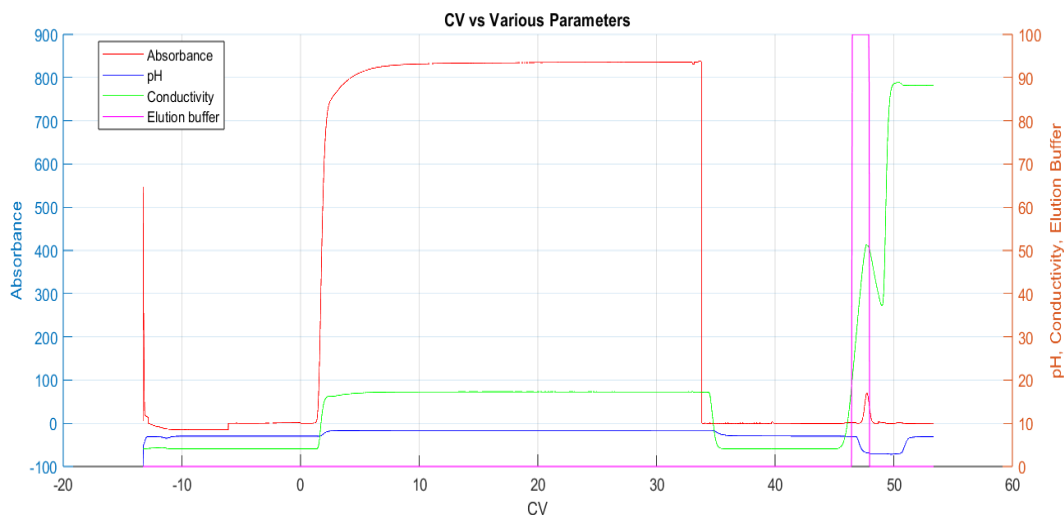


Figure 3.5: Chromatography run on AKTA for purification of pertactin. Y-axis of the graph shows the parameters monitored during the chromatography run.

consisted of 0.1 mg dye, 40% methanol, 40% 1 N acetic acid, and 20% Milli-Q water. After staining, the gels were destained using a solution containing 40% methanol, 40% 1 N acetic acid, and 20% Milli-Q water. The destained gels were then transferred to a densitometer (Bio-Rad) for image analysis.

3.8 Fluorescence Spectroscopy

Fluorescence spectroscopy was used for monitoring mainly NADPH and riboflavin secretion in the culture supernatant based on their unique fluorescence properties. Two different fluorescence spectroscopy equipments were used for monitoring NADPH levels in culture broth which are discussed in the following subsections.

3.8.1 In-Line fluorometer

An in-house developed fluorometer (by Dr. Boris Tartakovsky, National Research Council, Montreal) was utilized for data acquisition. This fluorometer was equipped with 365 nm and 405 nm LED lights, and an OceanOptics Inc. CCD array spectrometer capable of acquiring emission spectra within the 350-1100 nm range. It also featured a sterilizable fiber-optic reflection probe, comprising one excitation fiber and six read fibers. Importantly, the excitation lights were replaceable, allowing for optimal selection of excitation wavelengths. The fiber optic probe was inserted into the bioreactors through a standard top-mounted 9 mm bioreactor port. Real-time process monitoring was achieved by acquiring spectra in real time, with raw fluorescence data saved every minute. Emission data was collected in the 300-1200 nm range using the two excitation wavelengths (365 and 405 nm). The optimal instrument settings for the culture used in this study included a voltage of 5 V, an integration time of 100 ms, and a scanning rate of 1,000 nm/min. The fluorometer was controlled by a computer, which also recorded the acquired emission spectra. To minimize noise, three emission scans were averaged for each time point measurement. Each data acquisition was preceded by recording a background spectrum, which was subsequently subtracted from each emission scan to ensure accurate correction. Figure (3.6) shows the optical probe along with fluorometer setup used for in-line monitoring of *B. pertussis* cultures.



(a) Optical probe with fluorometer setup.



(b) Reactor setup with in-line probe.

Figure 3.6: Optical probe and bioreactor setup with in-line probe for real-time fluorescence data acquisition.

3.8.2 Off-Line fluorometry

The analysis of extracellular NADPH concentration in the supernatant was performed using fluorescence excitation-emission spectroscopy with a Cary Eclipse Fluorescence Spectrophotometer (Agilent Technologies). The supernatant samples were diluted 20-fold and analyzed in polymethylmethacrylate cuvettes with a slit width of 5 nm. Peak measurements were taken at photomultiplier tubes (PMT) voltage set at 800 V. To ensure that the NADPH concentration can be monitored, measurements were taken within the excitation range of 330-370 nm at 10 nm intervals and the emission range of 300-600 nm at 1 nm intervals. Phosphate buffered saline (PBS, 10 mM, pH 7.2) was used as a blank, and its spectra were subtracted from the sample spectra. The prominent intensity peaks observed in the emission spectra are attributed to first-order Rayleigh scattering, which occurs when the emission wavelength of oscillating molecules is close to the excitation wavelength of the incident light frequency.

3.9 Multi-Colour Flow Cytometry

Flow cytometry was used to monitor the levels of reactive oxygen species (ROS) and surface expression of pertactin (PRN) among the bacterial cell cultures. Experiments were conducted using an S3e Cell Sorter from Bio Rad. This instrument was equipped with two excitation lasers, a primary laser with an excitation wavelength of 488 nm and a secondary laser with excitation wavelength of 640 nm. It had emission filters to detect forward scattering (FSC) and side scattering (SSC), along with four channels for detecting fluorescence emission: 526/48 (FL1), 593/40 (FL2), 670/30 (FL3), and 700 LP (FL4). For each sample, a total of 20,000 events were measured at a rate of 1,000 events per second.

3.9.1 Flow cytometry gating

Before acquiring flow cytometry data, a gating strategy is necessary to eliminate noise, isolate *B. pertussis* cells, and exclude doublets. Figure (3.7) illustrates the flow cytometric gating protocol employed to measure single *B. pertussis* cells. In the first gate (3.7a), the side scattering area is plotted on the x-axis and the fluorescence area in the FL1 detector is plotted on the y-axis. This gating procedure aims to eliminate events that contribute to scattering intensity without corresponding fluorescence thus ensuring that the scattering arises from bacterial cells and is not due to electronic noise. The second gate (3.7b) plots the forward scattering area on the x-axis and the side scattering area on the y-axis. This gate identifies the region representing *B. pertussis* based on size and complexity and provides a rough check for contamination. The final gate (3.7c) excludes droplets containing more than one cell. The width of the side scattering is plotted on the x-axis and the height of

the side scattering is plotted on the y-axis. Due to the dilute sample solution ($OD_{600} = 0.1$), a narrow gate can be used at this location.

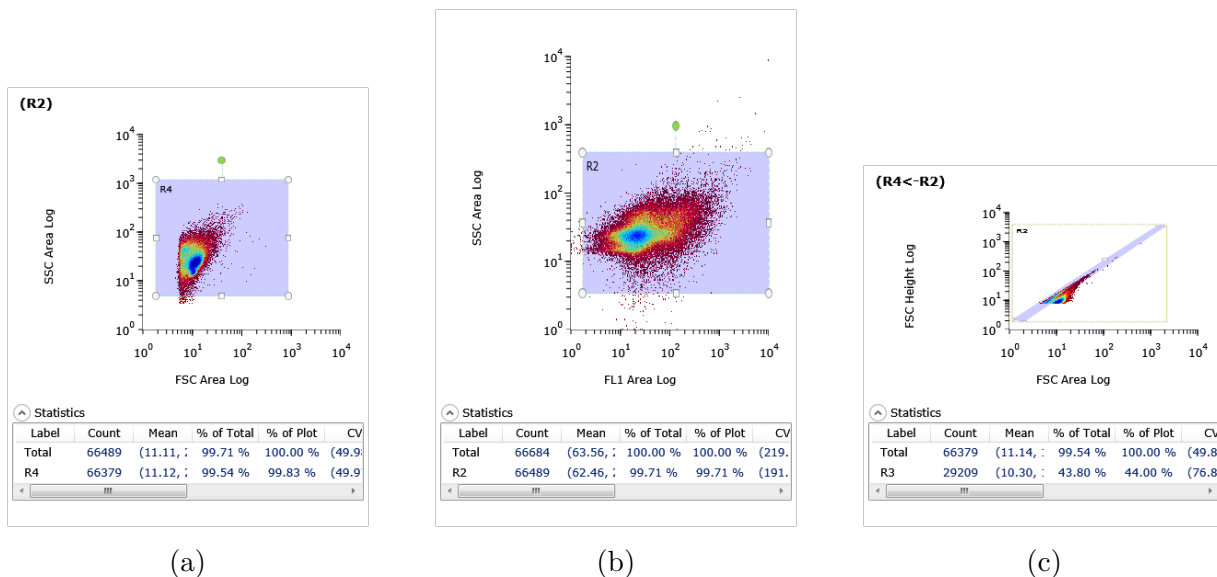


Figure 3.7: Different flow cytometry gatings utilized to isolate *B. pertussis* population.

3.9.2 Oxidative stress

Detection and analysis of reactive oxygen species (ROS) was done to study the correlation between ROS and initial glutamate concentration. ROS levels in cell culture were assessed using the fluorescein derivative carboxy-2',7'-dichlorodihydrofluorescein diacetate (carboxy-H2DCFDA) obtained from Sigma-Aldrich (Canada). A 1 mM solution of this reagent was prepared by dissolving it in DMSO and then adding it to the samples to achieve a final concentration of 50 μ M. This reagent is non-fluorescent and cell-permeable but becomes fluorescent DCF after being deacetylated by intracellular esterases and react-

ing with hydrogen peroxide and other reactive oxygen species (Halliwell and Whiteman, 2004). The FL1 channel was utilized to measure ROS distribution through the emission of DCF, which exhibits green fluorescence at an excitation/emission pair of 503/523 nm.

3.9.3 Surface expression of pertactin

To study the correlation between cell surface expression of pertactin, and oxidative stress, surface pertactin was assessed using Cy5 dye (Ex/Em - 640/700 nm) (abcam, catalogue no. - 188288) which was conjugated to a pertactin antibody (provided by Sanofi). To monitor the cell surface expression of pertactin, 1 ml of cell culture was washed, diluted to an OD of 0.1 and re-suspended in PBS. 2 μ L conjugated Cy5 dye was then added to the washed sample and then incubated for 45 minutes. Cells were then washed and re-suspended in 1 mL PBS. FL3 channel of the cytometer was then utilized to measure the fluorescence intensity of 20,000 cells stained with Cy5 dye, which exhibits red fluorescence at an excitation/emission pair of 638/693 nm.

3.10 Fluorescence Microscopy

To visualize the effect of ROS on cell surface pertactin expression, fluorescence microscopy was employed using a Zeiss LSM 700 confocal microscope (Carl Zeiss, 2019, Germany) equipped with four excitation lasers (405 nm, 488 nm, 555 nm, and 633 nm). Cell samples were taken from the flasks, centrifuged at 10,000 RPM for 3 minutes, washed, and re-suspended in PBS. The washed cell cultures were then diluted to an OD of 0.01. For

detecting ROS in the cell cultures, a fluorescein derivative, 2',7'-dichlorodihydrofluorescein diacetate (H₂DCFDA) (Thermo Fisher, catalogue no. D399, ex/em - 503/523 nm), was used. A 1 mM stock solution of H₂DCFDA was prepared in DMSO, and 5 μ L of this stock solution was added to the washed sample for staining. To visualize the surface expression of pertactin, a Cy5-conjugated pertactin antibody (provided by Sanofi, abcam catalogue no. 188288) was used. For this, 2 μ L of Cy5-conjugated pertactin antibody was added to 1 mL of the washed and diluted cell culture. The sample was incubated for 45 minutes, then washed and re-suspended in PBS. Stained cells were visualized under the microscope using glass slides (VWR, catalogue no. 16004422) coated with 0.1 % (v/v) poly-L-lysine solution.

Chapter 4

In-Line Monitoring of Bordetella Pertussis Cultivation Using Fluorescence Spectroscopy

4.1 Overview

¹ In this work, the use of advanced fluorescence spectroscopy as a non-invasive process analytical technology (PAT) tool for real-time monitoring of Bordetella pertussis cultures is evaluated. PAT initiatives in the pharmaceutical industry are integrated to gather detailed information about the chemical composition, physical properties, and dynamics of processes without disrupting the production line. We employed a 2-D fluorometer equipped with two

¹Adapted from (Mishra et al.). “In-line monitoring of Bordetella pertussis cultivation using fluorescence spectroscopy. Bioprocess and Biosystems Engineering, 46(6), pp.789-802, 2023”

specific excitation wavelengths (365 nm and 405 nm) to capture fluorescence emission spectra across a range of 350–850 nm. This setup was integrated directly into the bioreactor, allowing for continuous in-line monitoring during both batch and fed-batch fermentation processes. The study focused on monitoring several critical parameters, including cell biomass, concentrations of key amino acids (glutamate and proline), and the production of Pertactin, an important antigen in the *Bordetella pertussis* vaccine. To interpret the complex fluorescence data, the authors developed Partial Least Squares (PLS) regression models, which were used to correlate the fluorescence signals with the bioprocess variables. The results showed that accurate predictions were possible, particularly when the models were specifically calibrated for each strain of *Bordetella pertussis* and each type of nutrient media used in the experiments.

The study concluded that separate models for different strains and media significantly improved prediction accuracy, particularly for biomass, glutamate, and proline concentrations. However, predicting Pertactin was more challenging, likely due to noise in the offline measurements used for validation.

4.2 Introduction

Bioprocess monitoring is crucial in modern biochemical industries, as productivity can be significantly improved by applying advanced monitoring and control techniques. Most bioprocesses are operated aseptically in a sterile control volume containing a complex mixture of biochemicals and they are conducted to obtain a desired level of productivity. Offline analytical monitoring methods generally introduce significant time delays between sam-

ple collection to sample analysis thus reducing their efficacy for process monitoring and control. Online sensors on the other hand, can be advantageous as they can be used for real time monitoring as well as for implementing novel machine learning algorithms for fault detection and diagnostics (Agarwal et al. [2021, 2019]). Hence, the use of online monitoring for microbial cultivation is of particular interest for controlling and increasing batch-to-batch repeatability. Currently, there are several techniques available for online monitoring of bio-processes (Mandenijs et al. [2013]). For effective online monitoring of a bio-process, the method has to be non-invasive and specific enough to measure multiple target variables in a single run. Spectroscopic sensors, including UV (Korshin et al. [1997]), mid and near infrared (Kara et al. [2011]), Raman (Long [1977]) and fluorescence spectroscopy (Royer [1995a]), meet these criteria. Since spectra obtained from these methods are corrupted by noise and are not trivially correlated to the process variables of interest, different mathematical techniques are used to mine meaningful information from the spectra. For example, chemometric methods like principal component analysis (PCA) (Wold et al. [1987a]), partial least squares regression (PLSR) (Höskuldsson [1988]) and artificial neural networks (ANN) (Braspenning et al. [1995]) are used along with data pre-processing methods for data mining (Yi et al. [2016]).

Near Infrared (NIR) Spectroscopy is a non-invasive method that can take rapid spectral measurements but has a low sensitivity for compounds with concentration below 0.1 percent (w/w) (Abbas et al. [2020]). It also has a higher absorbance in aqueous solutions, which creates broader peaks in NIR, leading to low detection accuracy of compounds that are present in low concentrations (Kara et al. [2011], Abbas et al. [2020]). While NIR exhibits

broad overlapping peaks, mid infrared spectroscopy (MIR) produces distinct peaks due to absorbance from characteristic bonds which permits quantification of bio-process variables. The major limitation of MIR is that the background solvent has to be transparent within the spectral region of interest (Haas and Mizaikoff [2016]). Thus, similar to NIR, MIR has a higher absorbance in aqueous solution leading to low detection of trace compounds. The other limitation of MIR is that the required equipment is expensive and needs a lot of data for model calibration (Abbas et al. [2020]). Several examples of NIR and MIR spectroscopy applications for online monitoring have been reported. For example, (Fazenda et al. [2013]) used NIR in situ combined with metabolic flux analysis model with PLS as a multivariate analysis method for real time monitoring of biomass and glycerol in *Pichia pastoris* culture.

Raman spectroscopy is another powerful spectroscopy method, which can be used for online monitoring. Unlike NIR and MIR, Raman spectroscopy has a weaker absorbance for polar molecules such as water and can be easily applied to aqueous batches. However, the use of Raman spectroscopy as a real time bio-process monitoring tool is limited because it requires low frequency lasers to mitigate fluorescence related effects. Also, the use of low frequency lasers in Raman spectroscopy can cause heating effect due to long exposure time needed for such lasers, which may damage or alter the sample (Graves and Gardiner [1989]). For example, (Shaw et al. [1999]) used Raman spectroscopy to monitor the changes in substrate and metabolite concentrations as well as product formation in yeast fermentation. The authors also suggested the use of a by-pass filter to remove cells as they were causing interference to the photon scattering process.

Fluorescence spectroscopy is gaining popularity as a method for in situ monitoring because of its inherent advantages. It is non-invasive, highly sensitive, and can also detect multiple components at low concentrations (Royer [1995a], Mandenius et al. [2013]). Also, the equipment needed for fluorescence spectroscopy is less costly as compared to Raman or NIR (Faassen and Hitzmann [2015a]). Furthermore, there is no analyte consumption, no additional reagent requirement and no special procedure for sample preparation. Some bio-processes can run for weeks. Once set, a fluorometer can take measurements for a long time without the need of recalibration.

(Ödman et al. [2010a]) demonstrated the use of fluorescence spectroscopy combined with off-gas analysis for at-line monitoring of fed-batch cultivations of *Streptomyces coelicolor* for the production of antibiotic actinorhodin. Fluorescence spectra were collected through a borosilicate glass window located in a recirculation loop connected to the bioreactor. Partial least squares (PLS), locally weighted regression, and multilinear PLS (N-PLS) models were built for prediction of biomass and substrate (casamino acids) concentrations but they were not able to achieve continuous prediction of variables for new test batches. (Teixeira et al. [2009a]) used in-situ 2D fluorometry with a range of excitation/emission wavelengths for monitoring key bioprocess variables in mammalian cell cultures, namely the concentration of viable cells and recombinant proteins. It was observed that the intensity of fluorescence signals in the excitation/emission wavelength range of fluorescing amino acids (tyrosine, tryptophan and phenylalanine), certain vitamins and Nicotinamide Adenine Dinucleotide Phosphate (NADPH) changed along cultivation time. Raw spectrum was treated with PCA to reduce noise and for dimension reduction of data. Then, a partial

least squares (PLS) regression was applied to correlate the reduced fluorescence maps with target state variables. However, 30 minutes were required to acquire a single spectra, which can be too long in online monitoring as some dynamic effects such as changes in culture's redox state are fast and thus important information can go undetected. In another work, (Teixeira et al. [2011a]) used synchronous fluorescence spectroscopy (SFS) for real-time bio-process monitoring of animal cell cultures. SFS used a range of both excitation and emission wavelengths. Bioreactor cultures of three monoclonal antibody-producing CHO cell lines were followed in situ by both 2D and synchronous fluorometry techniques. Partial least squares regression models were then independently developed for viable cell density and antibody levels on the basis of the different fluorescence signals.

(Graf et al. [2019a]) developed a 2D fluorescence sensor system for in-line monitoring of mammalian cell cultures. The spectroscopy device utilized three high-power LEDs as excitation sources in combination with a back-thinned CCD-spectrometer for fluorescence detection. Different excitation wavelengths were used to acquire emission spectra. The acquired fluorescence spectra of several batches were evaluated using multivariate methods. The resulting batch evolution models were specifically used to detect deviations from "golden batch" runs but they were not used to predict process variables of interest. (Karakach et al. [2019a]) utilised multi-variate curve resolution (MCR) analysis for real-time study of fluorescent compounds present in *Pichia Pastoris* cultivation in a bioreactor. Emission spectra was acquired using a dual excitation fluorometer. MCR derived 'pure concentration' profiles were used for real-time process diagnostics and fault detection but the authors could not demonstrate the online prediction of target variables.

In summary, previous work on online fluorescence-based monitoring has mostly focused on mammalian cultures while the prediction ability of microbial cultures, such as *B. pertussis* studied in the current work, have not been thoroughly addressed. There are significant differences between mammalian and microbial cultures in terms of cell size, media compositions and thus the intrinsic fluorescence measured in these cultures are expected to be different. Also, previous studies have not addressed differences in media and cell strains. It is important to underline that only particular molecules in the cell culture exhibit intrinsic fluorescence, e.g. NADPH, some amino acids, and thus many of the species that are quantified in reported studies were in fact not directly measured from fluorescence but are rather inferred from their correlations with fluorescing molecules. However, such correlation will depend on the cell metabolism and the growth medium composition. Accordingly, it is important to assess the effect of changes in media and cell strains on these inferences as done in the current study.

In this work, a multi-wavelength fluorometer, which was previously used by (Karakach et al. [2019a]) for culture monitoring and fault detection of *Pichia pastoris* cultures, was used for in-line monitoring of bacterial (*Bordetella pertussis*) fermentation for antigen (pertactin) production. The fluorometer developed is compact and robust and can be easily installed in an industrial environment. Instead of using a range of excitation wavelengths, the fluorometer combined only two LED excitation lights (365 and 405 nm) with a CCD array spectrometer for fast and accurate acquisition of fluorescence spectra. The acquired fluorescence data was used to develop regression models to predict biomass growth, glu-

tamate consumption and production of pertactin over the course of the batch. Also, it is shown that combining the fluorescence spectra with real time agitation, dissolved oxygen and culture volume data resulted in improved prediction of target variables. A comparison between different regression models used for prediction and also with models trained separately for different cell lines and media is presented. The comparison is based on the Akaike Information Criterion (AIC) to account for the different number of parameters of the different models.

4.3 Materials and Methods

4.3.1 Culture Conditions

Fermentations were performed in 2 liters bioreactors with a working volume of 1.5 to 1.6 liter. Two different bacterial strains were used: a strain currently used in vaccine manufacturing and a hybrid strain, hereafter referred to as “Wild” and “Recombinant”, respectively. Two different culture media were used and are hereafter referred to as “cP” (modified Stainer Scholte broth) and “F11” (synthetic medium containing Amino acids + salts) media. A total of 11 fed-batch fermentations were performed. Cultures were maintained at 36 °C, 7.2 pH, with a controllable Dissolved Oxygen at 35 % (of saturation value) by maintaining constant air-flow (at 2 lpm). The agitation rate was kept between 200 to 600 rpm (Rushton impeller). Antifoam was manually added to the bioreactors as per the batch requirement. Table (4.1) gives a summary of each run as well as batch to

Table 4.1: Culture Conditions.

Set	Batch ID	Batch Run Time (hr)	Initial Glutamate Concentration (mmol/L)	Initial Proline Concentration (mmol/L)	Feed start Time (hr)	Bacterial Strain	Media Type	Maximum OD (600 nm) (standardized)	Productivity (PRN) (standardized)
1	A	71.2	62.47	50.1	22.3	Wild	F11	1.276	1.58
1	B	72.08	69.06	51.12	21.5	Wild	F11	1.286	1.31
1	C	72	63.37	53.52	24	Wild	F11	1.47	1.44
1	D	70.1	74.75	54.11	25.6	Wild	F11	1.75	1.42
2	E	70.2	66.50	50.86	24.2	Hybrid	F11	1.46	2.59
2	F	72.6	69.80	50.88	21.5	Hybrid	F11	1.08	2.52
2	G	71.6	67.28	51.63	24.1	Hybrid	F11	0.95	3.11
3	H	72	64.17	36.8	25	Hybrid	cP	0.98	2.64
3	I	73.65	58.14	41.52	26.9	Hybrid	cP	2.08	2.99
3	J	74	66.91	38.91	24.1	Hybrid	cP	1.60	3.02
3	K	71	67.47	40.97	26.2	Hybrid	cP	1.95	3.13

batch differences in terms of initial glutamate feed, maximum optical density (OD) and productivity. The values of OD and antibody concentration are standardized with mean and standard deviation values for intellectual property reasons. For set 1 that used the wild strain and F11 media, a total of 52 data points for OD, 40 for amino acids concentration (glutamate and proline) and 20 for pertactin concentration were collected. For set 2, using the recombinant strain and F11 media, 40 data points for OD, 30 for amino acids and 14 for pertactin (PRN) antigen were available. Lastly, for set 3 that used hybrid strain and cP media formulation, 85 data points for OD, 45 for amino acids concentration and 22 for PRN concentration were measured. The amount of data points that could be used for PRN was lower than for the other variables because this antigen is produced in measurable amounts only in later stages of bacterial fermentation (usually after the fed batch process starts) and also because of the complexity of the ELISA analysis.

4.3.2 Offline measurements

Biomass samples were collected from the bioreactor at time intervals between 2 to 12 hours for the determination of OD, glutamate consumption and amount of PRN antigen produced for each batch. Optical density was determined using a spectrophotometer (Thermo Scientific, GENESYS 40/50 Vis/UV-Vis (2020)). Glutamate and proline concentration were measured using HPLC (Thermo Fisher UltiMate (2019)). A high efficiency AEX column (ProPac Bio LC SAX-10, 4x250 mm, Thermo Fisher, catalogue no. 054997) was utilised. Precolumn derivatization of the amino acids was done using AccQ Fluor (6-aminoquinolyl-N-hydroxysuccinimidyl carbamate (AQC)). The derivatives were then monitored using fluorescence detection method. The standard error of measurement was below 5 percent.

PRN was quantified by two different methods. For set 1 and set 2, antigen concentration was measured using ELISA (Biotek Plate Reader Synergy H1 Mono RDR (2019)) by coating the micro plates with 50 μL /well of PRN polyclonal antibodies (1.5 $\mu\text{g}/\text{mL}$) purified from guinea pig blood. After that, incubation of plates was done for 20 hours followed by washing it 3 times with 200 μL /well of a solution containing 10 mM PBS, 0.1 % tween-20 and 0.1 % BSA. Blocking was done for 30 minutes at room temperature with 100 μL /well of 10 mM PBS and 0.1 % BSA. After this, the samples were centrifuged, filtered, diluted and added to wells for incubation. Then, plates were washed again with PBS-Tween-BSA solution. 100 μL of horseradish per-oxidase conjugated guinea pig PRN antibody was added to the plates followed by 1 hour of incubation at room temperature. 50 μL of 2 N H_2SO_4 was added to stop the reaction and absorbance was measured at 450 nm. Standard

curve was used to calculate PRN concentration. Absorbance was proportional to PRN concentration. All the chemicals and solutions used for HPLC and ELISA were obtained from Thermo Fisher Scientific. For set 3, HPLC was used instead of ELISA since the latter was found to have high variability for supernatant samples.

For set 3, reverse phase HPLC was utilised for pertactin quantification. A 2.1 x 100 mm Thermo PabPac RP column was used. The volume of the column was around 0.35 ml. Solvent A was water + 0.1 % TFA and solvent B was 90 % acetonitrile + 10 % water + 0.1 % TFA. The flow rate was 0.5 ml/min. Detection was done using intrinsic fluorescence with excitation at 280 nm and emission at 335 nm.

4.3.3 Fluorescence Map Acquisition

A multi-wavelength (dual excitation-full emission) fluorometer was used for data acquisition. The fluorometer was equipped with 365 nm and 405 nm LED lights, a CCD array spectrometer (OceanOptics Inc.), which can acquire emission spectra in a range of 350-1100 nm and a sterilizable fiber-optic reflection probe which had one excitation fiber and 6 read fibers. Notably, the excitation lights are replaceable thus enabling optimal selection of excitation wavelengths. The fiber optic probe was inserted into the bioreactors via a standard top-mounted 9 mm bioreactor port. Real-time process monitoring was accomplished by spectra acquisition at an interval of 1 min, while raw fluorescence data was saved every 6 minutes for set 1 and set 2 and every 10 minute for set 3. Two excitation wavelengths (365 and 405 nm) were used and emission was acquired in the range of 400-850 nm. A voltage of 5 V, integration time 100 ms, and a scanning rate of 1,000 nm/min were identified as

optimum instrument settings for the culture employed in this study. The fluorometer was controlled by a computer, which also recorded the acquired emission spectra. For measurement at each time point, an average of three emission scans were taken to reduce noise. The acquisition of data at each time point was preceded by the recording of a background spectrum that was later subtracted from each emission scan acquired for correction.

4.3.4 Data preprocessing

Spectral areas were calculated using the acquired fluorescence spectra. The emission spectra were divided into 9 sets corresponding to fluorescence areas. Table (4.2) describes the excitation/emission wavelengths associated with each area and the fluorophores corresponding to the emission peaks. The fluorescence areas of each batch were normalized by the areas corresponding to time $t = 0$ hour which is collected immediately after the inoculation to eliminate the batch to batch variability which may arise from different geometrical position of system components (like position of fluorometer in bioreactor) in each batch. MATLAB (release R2020b) was used for all chemometric calculations.

4.3.5 Akaike Information Criteria

The Akaike Information Criteria or AIC can be used to assess the relative accuracy of statistical models by accounting for the trade-off between accuracy and number of model parameters (Hu [2007]). Given multiple models for the data, AIC can be used to evaluate the quality of each model when compared to each other, thus providing a means for model selection. The model performance is measured by checking the model fit and penalizing

Table 4.2: Emission areas used for on-line monitoring

Notation	Excitation nm	Emission nm	Comments
A1	365	450-475	Presumed NADH/NADPH peak
A2	365	475-550	Presumed Riboflavin peak
A3	365	600-640	Porphyrins
A4	365	660-695	
A5	365	700-800	Light scattering (secondary)
A6	405	500-580	Presumed Riboflavin peak
A7	405	590-640	Porphyrins
A8	405	660-695	
A9	405	750-850	Light scattering (secondary)

for complexity of the model where the latter is related to the number of model parameters. The model with a lower AIC value is regarded as the one that provides a better balance of model fit with model complexity. The calculation of the AIC test criterion involves the following assumptions:

1. Same data is used among multiple models.
2. Same response variables are measured among multiple models.
3. The data sample size is infinite

Since the data is never infinite, a correction term is added. The AIC for an infinite data set is given as follows:

$$AIC = -2\ln(L) + 2K \quad (4.1)$$

where L is the likelihood and K is the number of parameters in the model. The log likelihood is used to describe how likely one can see their observed data. Models with high log likelihoods have low Akaike value. For Least squares model fitting and lower number of datapoints, the AIC calculation is modified as follows:

$$AIC = n * \ln(\sigma^2) + 2K + \frac{2K(K+1)}{(n-K-1)} \quad (4.2)$$

where n and K are number of data points and parameters in the model respectively and σ equals the Residual Sum of Squares/ n . Once the AIC is calculated for each model, the model with the lowest AIC is viewed as the best one among the compared models. However, it should be noted that the AIC only measures the relative quality of models, which means that all the models that are being compared can exhibit poor fit (Cavanaugh and Neath [2019]). Therefore, other measurements such as root mean squared error (RMSE) or mean absolute percentage error (MAPE) should also be calculated to assess accuracy since these values can be a better indicator of individual model performance.

4.4 Results and Discussions

In this study, fluorescence-based monitoring of *Bordetella pertussis* was carried out to achieve real-time estimation of key cultivation parameters: optical density (OD), main carbon sources (glutamate and proline) concentration and productivity of pertactin that is one of the 4 antigens that make up the vaccine (pertactin concentration). Optical density provided information about growth and growth rates. Glutamate and proline are

the main nutrients in *B. pertussis* fermentation. It has been reported that in terms of amino acids content, glutamate and proline are sufficient for inducing pertussis growth while all other needed amino acids can be bio-synthesized (Proom and Knight [1955]). Pertactin is one of 4 antigens contained in the acellular vaccine but, since it is produced in very small amounts, it is the limiting compound for overall vaccine productivity. Partial Least Squares regression models were developed for the prediction of target variables. The prediction accuracy of the model was determined by testing the model on a completely new data set to predict biomass growth, consumption of glutamate and pertactin productivity. Finally, the regression models were tested for over fitting using the Akaike Information Criteria (AIC).

4.4.1 Qualitative description of the spectra

Table (4.2) shows the range of excitation and emission wavelengths defining each of the areas under the spectra that were used as inputs to the regression model and the corresponding species that exhibit intrinsic fluorescence in each of these excitation/emission wavelength ranges. Although the experimental runs were conducted with different media and cell lines, the fluorescence spectra obtained over time exhibited common patterns for the emission signals of all batches. Figure (4.1) shows the dynamics of emission spectra observed for two excitation wavelengths, at different times during the course of fermentation. The spectra were saturated in the range of emission wavelengths of 350 - 440 nm because of the excitation light reflection by cells and reactor walls. The first peak for 365 nm excitation wavelength and 450 nm emission wavelength corresponds to NADPH

([Blacker and Duchen \[2016\]](#)). The peak rises as the fermentation proceeds. The second and main peak in the excitation 405 nm and emission range of 530-540 nm is associated with riboflavin which also shows a gradual increase over batch course. The increase of NADPH observed at the start of the fermentation is believed to be caused by the secretion of NADPH by the cells to counter the oxidative stress which arises due to excess nutrients present in the culture. It was noticed that the peak of NADPH increased monotonically during the fermentation since it is secreted by the cells thus accumulating into the extracellular broth. Similarly, Riboflavin also has an important antioxidant role since the glutathione redox cycle requires riboflavin in the FAD coenzyme form for its activity ([Ashoori and Saedisomeolia \[2014\]](#)). The nutrients in the cell are oxidized in tricarboxylic acid cycle (TCA) cycle to produce NADH and FADH. These co-factors donate electrons to the electron transport chain (ETC) where they are used by molecular oxygen to form H_2O . Incomplete reduction of oxygen in the ETC also occurs due to excess of electrons because when oxygen receives more electrons than it can efficiently handle in a controlled manner, it tends to form partially reduced oxygen species, which leads to formation of superoxides. Thus, excess nutrient consumption in the TCA cycle can overwhelm the electron transfer chain ([Wellen and Thompson \[2010a\]](#)). Then, excess ROS buildup inside the cell occurs resulting in oxidative stress. While the use of fluorescence to track oxidative stress is beyond the scope of the current study, we are currently investigating the ability to infer such stress from NADPH fluorescence based measurements. The peaks around 740 nm emission corresponds to secondary reflections which gradually rise as the fermentation proceeds and then slightly decrease around 48 hrs. The gradual increase in peak height are strongly correlated to the biomass growth trend.

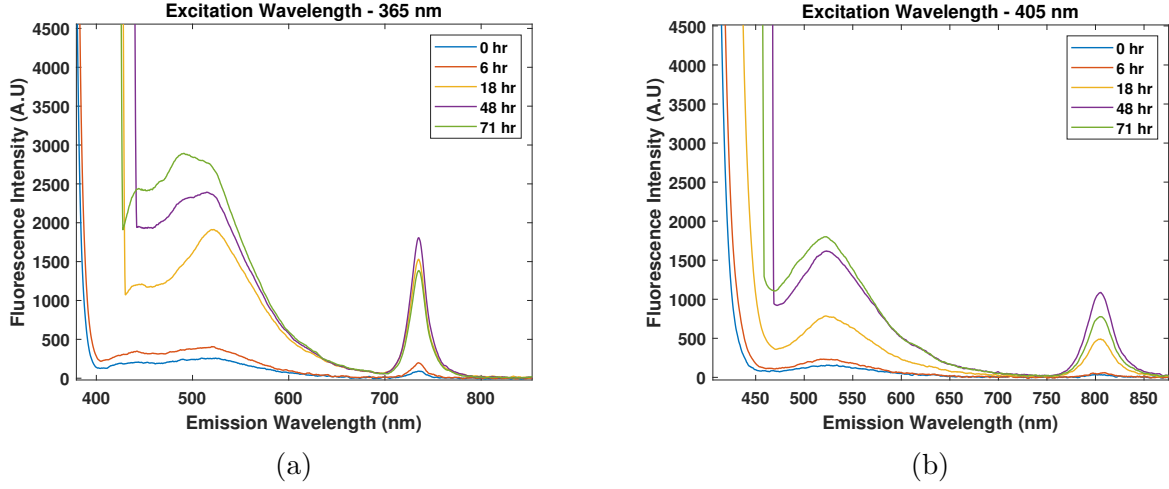


Figure 4.1: Emission spectra dynamics over the batch course for batch B.

4.4.2 PLS model for biomass, amino acid and antigen prediction

In this experiment, along with fluorescence, many other batch parameters such as temperature, dissolved oxygen, agitation, pH and culture volume changes were also monitored continuously. Figure (4.2) shows the variation of agitation, dissolved oxygen and culture volume change over the course of batch A. Most of these measured variables are correlated to each other since they are either manipulated or controlled variables in closed loop controllers. pH is controlled by manipulation of acid feed rate, temperature by manipulation of jacket temperature and dissolved oxygen by manipulation of agitation rate. It was hypothesized that some of these measurements contain information about the four estimated variables (OD, glutamate, proline and pertactin) and thus they could be combined together with fluorescence to improve the accuracy of the soft sensor. To avoid increasing the number of tuning parameters in the soft sensor model, a preliminary investigation was conducted using step-wise regression to determine which measurements contribute the most

to the soft sensor accuracy. Following this analysis it was determined that only agitation, dissolved oxygen and culture volume contribute significantly to the estimation problem. This result can be explained by the performance of the closed loop controllers deployed in the bioreactor. On the one hand, pH and temperature were tightly controlled and they did not exhibit significant variations during the culture thus explaining why they were not informative. On the other hand, the control of dissolved oxygen was less strict as evident by variation of dissolved oxygen during the first 10 hours of the experiments shown in Figure (4.2) until the set point of 35 percent was reached. The aeration rate was constant and the agitation rate could not be changed during this period to avoid high stress on the cells. Accordingly, changes in rate of consumption of oxygen at a constant aeration and agitation rates were expected to be correlated to the initial growth rates. After the set point was achieved, the agitation rate changed to maintain the dissolved oxygen at its set-point. Thus, the changes in agitation rate become informative about the oxygen requirements of the culture and thus can provide indirect information about growth and nutrient depletion. The volume changes reflected the transition between batch and fed-batch operation and thus, the inclusion of volume as additional input provided information about the changes in the culture operation. Following the above analysis, dissolved oxygen, agitation rate and culture volume were combined with the fluorescence areas as additional features in a new Partial Least Squares regression model.

A Partial Least squares (PLS) based regression model was employed to correlate the fluorescence spectra and areas (spectra integrals) along with fermentation data such as agitation, dissolved oxygen and volume with the target variables i.e. biomass, glutamate,

proline and pertactin (antigen). Initially, one single model was developed by combining data from all the batches. This single model will be referred to as the 'global model' to differentiate it from models developed for each media or each cell line presented later.

Two different approaches were used for selecting the data for training and prediction as

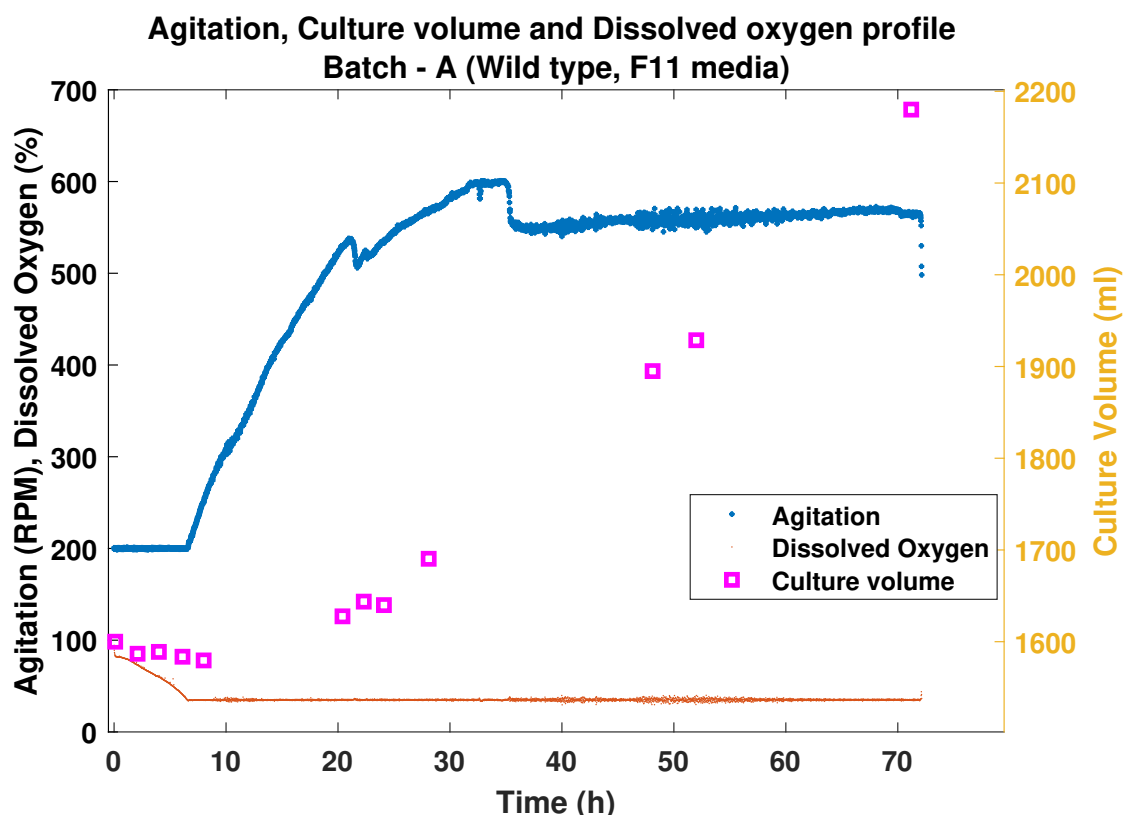


Figure 4.2: Batch profile for agitation, dissolved oxygen and culture volume.

follows:

1. The Leave One Batch Out (LOBO) approach was used to select all except one fermentation batch of a particular experimental set for calibrating the model. All batches except one (in each set) were used for calibration of the model, and the remaining

Table 4.3: Prediction accuracy of 'Global' model. The model was calibrated by combining data from all the batch runs. The best result for LOBO approach is tabulated here. RMSE values along with coefficient of determination is calculated to evaluate the performance of regression models.

Prediction Strategy	Variable	Global Model			
		R ² -c	R ² -p	RMSE-c	RMSE-p
Leave One Batch Out (LOBO)	OD	0.90	0.81	0.33	0.52
	Glutamate	0.88	0.56	0.42	0.96
	Proline	0.83	0.55	0.39	0.88
	Pertactin	0.79	0.43	0.48	1.44
Random Partition of Data Set (50-50 %)	OD	0.91	0.78	0.43	1.05
	Glutamate	0.75	0.43	0.76	1.12
	Proline	0.73	0.44	0.70	1.56
	Pertactin	0.82	0.33	0.54	1.98

batch was used to test the model. The calibration/test ratio in terms of data was approximately 70/30 in the leave one batch out approach.

2. Random partition of the data into calibration and testing sets was conducted for each experimental set to test the model. In this approach half of the data set was used for model calibration and the other half was used to test the model.

Two sets of inputs were considered. Firstly, raw spectra combined with agitation, dissolved oxygen and volume was used for PLS model calibration. Secondly, areas of fluorescence spectra, i.e. areas under the curve of the fluorescence spectra in different ranges of emission wavelengths, combined with agitation, dissolved oxygen and volume was used to calibrate PLS models for each of these sets. The resulting models were used to predict target variables using a leave one batch out (LOBO) and random partition strategy. The LOBO approach consists in leaving one batch out, then calibrating the model with the

remaining batches and predicting the target variables for the batch that was left out. Following the application of this procedure, it was found that the prediction accuracy slightly improve with the model that uses fluorescence areas as inputs as compared to the model that uses raw spectra. So, only fluorescence areas combined with agitation, dissolved oxygen and volume were used for model calibration and prediction. This slight improvement of accuracy when using fluorescence areas versus raw spectra was expected due to the filtering of noise present in the spectra by using areas instead of raw data.

For the global model, data from both the cell line and media was combined and used for model calibration for both the strategies. The number of principal components were chosen based on the minimum RMSE values for prediction. Fluorescence areas along with agitation, dissolved oxygen and culture volume were used for calibrating the model. Also, the model was trained with different batches having high variability to ensure that the range of variability of the data was captured. The same approach was used for models developed for each media or each cell line, the results of which will be discussed in a later section. Table (4.3) presents the coefficient of determination (R^2) and Root Mean-Squared Error (RMSE) values for the global model. The R-squared values give a measure of goodness-of-fit and combined with RMSE value, they can be used to evaluate the performance of regression models. R-squared values for the LOBO approach were slightly better compared to the random partition of data approach. Table (4.3) shows that biomass i.e OD obtained the highest R-squared values for both calibration and prediction, i.e. 0.9 for calibration and 0.81 for prediction, while the other variables resulted in lower R-squared values for prediction. These results were somewhat expected and they corroborated the hypothesis

that data collected for different cell lines and media cannot be properly modelled with one single (global) model. Thus, the use of separate models was investigated as discussed below.

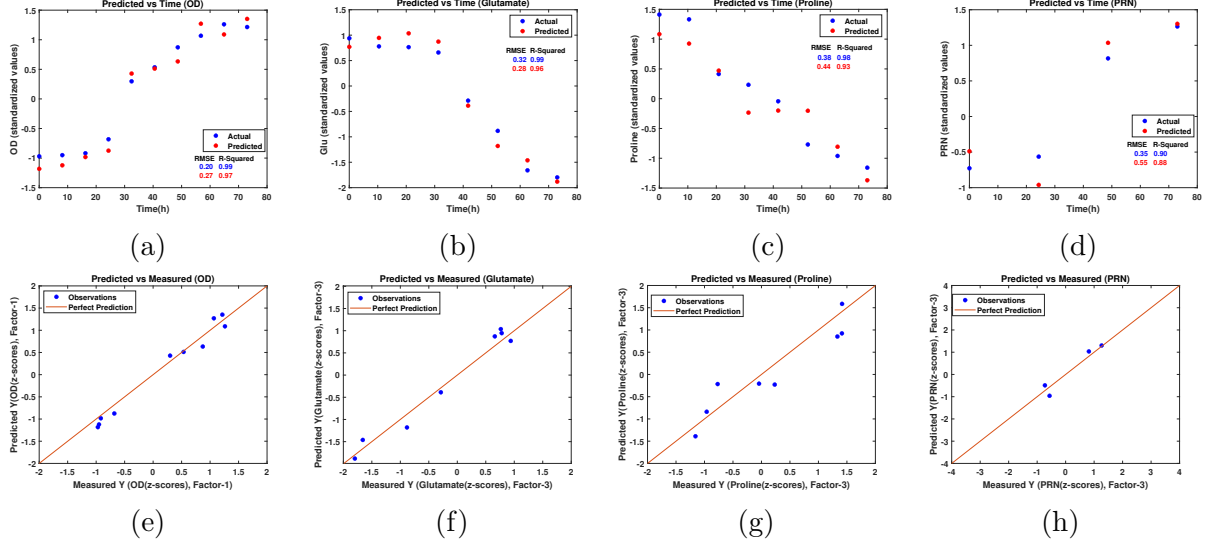


Figure 4.3: Performance of PLS regression model for experimental set 1 (Leave One Batch Out approach). Fluorescence area along with agitation, dissolved oxygen and culture volume at each time point were used as model predictor for the estimation of target variables such as OD, glutamate and proline consumption and pertactin antigen production over the batch course. Predicted and actual values (standardized) were plotted over the batch course for (a) OD (b) Glutamate (c) Proline (d) PRN. Figure (e), (f), (g), (h) show predicted vs measured values for corresponding variables.

4.4.3 Effect of cell line and media

A main point investigated in the current research was the effect of changing either the cell line or the media composition on the predictive ability of the fluorometry-based sensor.

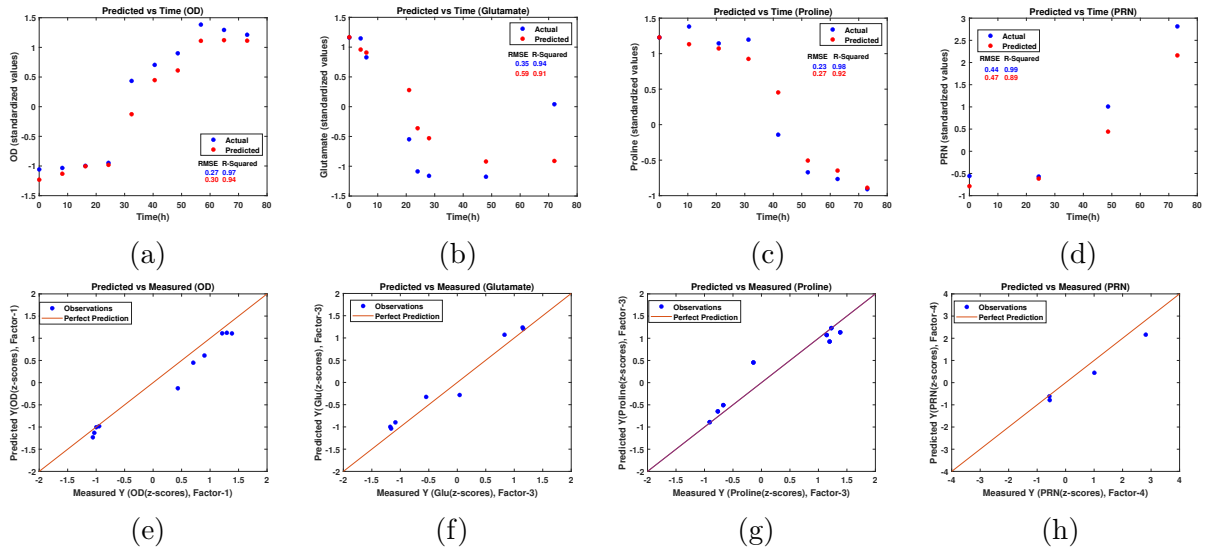


Figure 4.4: Performance of PLS regression model for experimental set 2 (Leave One Batch Out approach). Predicted and actual values (standardized) were plotted over the course of the batch for (a) OD (b) Glutamate (c) Proline (d) PRN. Figure (e), (f), (g), (h) gives the plot of predicted vs measured values for corresponding variables.

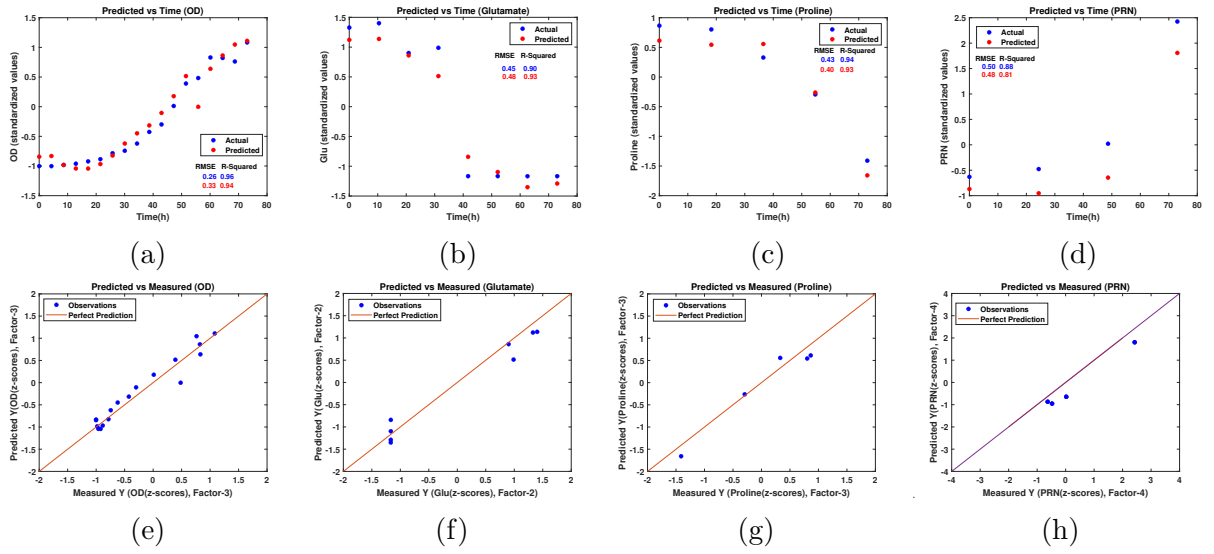


Figure 4.5: Performance of PLS regression model for experimental set 3 (Leave One Batch Out). Predicted and actual values (standardized) were plotted over the course of the batch for (a) OD (b) Glutamate (c) Proline (d) PRN. Figure (e), (f), (g), (h) gives the plot of predicted vs measured values for corresponding variables.

As mentioned above, none of the process variables evaluated in this work , i.e. OD, glutamate, proline and pertactin concentration, have intrinsic fluorescence. Thus, the values predicted from the fluorescence spectra are only an inference based on the correlation of the four estimated process variables to molecules in the cell culture that have intrinsic fluorescence, e.g. NADH/NADPH and riboflavin, or measurements of scattering that are also part of the measured spectra. These correlations are expected to change if the cell metabolism changes. Since changes in cell line (cell strain) and media composition are expected to impact the cell metabolism, the correlation between the estimated variables and the fluorescing compounds are expected to change accordingly. Based on this rationale, it was hypothesized that by developing separate models for different combinations of media and cell lines as described in Table (4.1), the predictive accuracy could be improved.

Then, to confirm that batches with different cell strains cannot be described by the same model, a model was calibrated for data set 1 which uses the wild strain and F11 media and we tested the model with data set 2 that uses the hybrid strain with F11 media. Hence, the two data sets used for calibration and testing differed in terms of cell strain but used the same media. Table (4.4) presents the results of the PLS regression for this case. The fact that the predicted R-squared values were very low for all four variables confirmed that different strains could not be monitored using a single model calibrated with one of the strains. Hence, separate models should be designed for different strains.

Since each data set were obtained from experiments using the same combination of media and cell line, one model was developed for each set that will be referred to as models

Table 4.4: Modeling results of PLS regression with differing cell line and media used for training and prediction (using fluorescence, agitation rate, dissolved oxygen and volume data).

Variable	Training - batch A,B,C,D (wild strain), F11 media Prediction - batch E, F (hybrid strain), F11 media				Training - batch E,F,G (F11 media), hybrid strain Prediction - batch H (cP media), hybrid strain			
	R ² -c	R ² -p	RMSEc	RMSEp	R ² -c	R ² -p	RMSEc	RMSEp
OD	0.92	0.61	0.33	1.11	0.95	0.71	0.23	1.21
Glu	0.90	0.66	0.52	1.94	0.88	0.62	0.29	1.25
Proline	0.94	0.69	0.28	1.98	0.85	0.60	0.37	1.06
PRN	0.82	0.44	0.63	2.23	0.81	0.56	0.55	1.31

'data set 1 model', 'data set 2 model' and 'data set 3' model hereafter. The results for these 3 models are summarized in Table (4.5). Using these separate models, very good predictions were obtained with both the LOBO and random data partition approaches. The LOBO approach gave slightly better results because the calibration model was trained with data which accounted for batch to batch variability. For the 3 data sets, the R²-p with the LOBO approach for OD were between 0.94-0.97, between 0.94-0.96 for glutamate, between 0.93-0.92 for proline and between 0.88-0.81 for PRN. It is not surprising that the prediction accuracy is lower for PRN as compared to the other variable as measurements of the antigen by either ELISA or HPLC introduced more noise as compared to the other measurements. Figures (4.3), (4.4) and (4.5) summarises the results of PLS regression for set 1, 2 and 3.

4.4.4 Additional online measurements as model features in the soft sensor model

To test whether combining additional variables with fluorescence areas, i.e. dissolved oxygen, agitation rate and cell culture volume could contribute significantly to the accuracy

Table 4.5: Prediction accuracy of PLS regression models for all sets of experiment with two different prediction strategies for set 1, 2 and 3. The regression model included fluorescence areas, agitation, dissolved oxygen and volume as model features.

Prediction strategy	Variable	Set 1				Set 2				Set 3			
		R ² -c	R ² -p	RMSEc	RMSEp	R ² -c	R ² -p	RMSEc	RMSEp	R ² -c	R ² -p	RMSEc	RMSEp
Leave One Batch Out (LOBO)	OD	0.99	0.97	0.20	0.27	0.97	0.94	0.27	0.30	0.96	0.94	0.26	0.33
	Glu	0.99	0.96	0.32	0.28	0.94	0.91	0.35	0.59	0.90	0.93	0.45	0.48
	Pro	0.98	0.93	0.38	0.44	0.98	0.92	0.23	0.27	0.94	0.93	0.43	0.40
	PRN	0.88	0.81	0.50	0.48	0.99	0.89	0.44	0.47	0.90	0.88	0.35	0.55
Random Partition of Dataset (50-50%)	OD	0.95	0.92	0.24	0.51	0.90	0.92	0.29	0.48	0.96	0.95	0.19	0.39
	Glu	0.97	0.88	0.24	0.53	0.88	0.83	0.31	0.56	0.98	0.88	0.22	0.38
	Pro	0.92	0.87	0.31	0.49	0.93	0.88	0.29	0.55	0.86	0.71	0.33	0.61
	PRN	0.85	0.67	0.27	0.64	0.86	0.71	0.33	0.61	0.92	0.84	0.30	0.52

of the soft sensor, separate models were designed for each data set as done in the previous section without including the additional input variables in the PLS regression models. Table (4.6) summarises the prediction accuracy without these additional features for all sets of experiments. It was found that combining agitation, dissolved oxygen and culture volume data improved the prediction accuracy of PLS regression. The R-squared values for prediction were consistently higher by almost 10 percent when the additional variables were introduced (see Table (4.4)) as compared to the R-squared values for prediction shown in Table (4.6). The RMSE values are lower for the models with additional inputs (see Table (4.4) as compared to the models that are based only on fluorescence (see Table (4.6)). As stated above, fluorescence is only indirectly correlated to the species that are being estimated and the addition of fermentation profiles, such as volume, dissolved oxygen and agitation rate, improve the prediction because they contribute further information regarding the cell culture.

To determine the magnitude and direction of the contributions of the input variables towards each principal component of the PLSR model depends, the PLSR weights were

Table 4.6: Prediction efficiency of the regression model for all sets of experiment when the additional features (agitation, dissolved oxygen and volume changes) were not included with area.

Variable	Set 1				Set 2				Set 3			
	R ² -v	R ² -p	RMSEv	RMSEp	R ² -v	R ² -p	RMSEv	RMSEp	R ² -v	R ² -p	RMSEv	RMSEp
OD	0.92	0.84	0.28	0.31	0.98	0.78	0.15	0.46	0.85	0.72	0.41	0.69
Glu	0.93	0.81	0.41	0.78	0.92	0.75	0.37	0.58	0.82	0.68	0.50	0.64
Proline	0.98	0.76	0.31	0.45	0.97	0.70	0.28	0.43	0.92	0.75	0.34	0.47
PRN	0.81	0.31	0.87	1.08	0.99	0.79	0.31	0.63	0.84	0.45	0.60	0.63

plotted for models set 1, 2 and 3 for PC1 and PC2 respectively. Figure (4.6) shows the PLS weights as percentages of the total for all four output variables, i.e. OD, glutamate, proline and antigen (PRN). Since most of the variability in the data is explained by the first principal component, the interpretation of the contribution plots is referred to this component. For PC1, the optical density (biomass) is positively correlated with all fluorescence areas in similar degree which can be explained by the monotonic increase of the intensities of the spectra as the fermentations progress and the biomass levels increase. From the contribution plots, it can be seen that OD is negatively correlated to DO indicating that a slower growth results in a lower oxygen consumption. The concentrations of glutamate and proline are negatively correlated to the fluorescence areas indicating that as biomass grows during the batch the nutrients are consumed. Both glutamate and proline concentrations are positively correlated with dissolved oxygen indicating less oxygen consumption corresponding to a lower consumption of nutrients. The negative correlation of agitation with nutrients' concentration indicates that a larger agitation rate is needed to maintain the DO target when oxygen consumption is higher. PRN weights related to oxygen concentration follow similar trend as biomass, i.e. increase in fluorescence correlates with an increase in antigen, thus indicating strong correlation between biomass and antigen increases.

4.4.5 Akaike Information Criteria

Since our study involves models that apply to different number of samples, e.g. the global model and the separate models developed for each data set, or models that have different number of parameters resulting from inclusion or exclusion of additional inputs, a formal comparison of PLS regression models was conducted based on Akaike Information Criteria. Regression models were evaluated on the basis of AIC values calculated from equation (4.2) that accounts for differences in number of model parameters or data samples. The model with lower Akaike value is considered better than other models in terms of balance between model-fitting and complexity. Akaike values were calculated for all sets of data and for all target variables. Table (4.7) provides a summary of AIC values calculated for the regression models developed for all four variables, including regression models that integrated the additional parameters. The Table compares the global model that was a single model developed for all the data sets together with the models that were designed to account for the effect of different media and cell line used. For OD, the AIC values for the separate models were lower than the AIC value of the global model. Also, the AIC value of OD with fluorescence and additional inputs (dissolved oxygen, agitation rate and culture volume) were lower than values of corresponding models with only fluorescence areas suggesting that including additional parameters improves the performance of the model. Similarly, for amino acids (glutamate and proline) the Akaike values for all the separate models were lower than the AIC for the global model. Models with additional inputs had lower Akaike value than models with just fluorescence areas except for set 3 of proline where Akaike value for model with additional inputs is slightly higher than the model which uses just

fluorescence suggesting that the additional inputs in set 3 slightly over-fit the data because of the additional parameters required when using additional inputs in the PLS regression. As for the prediction of the antigen (pertactin), the results are less consistent. The Akaike value for the 'Global' model with only fluorescence is larger compared to regression model for sets 2 and 3 but smaller than the AIC for step 1. The AIC is lower or similar for the separate models versus the global model when additional inputs were used for set 1 and 2 but not for set 3. The less consistent results for pertactin may be explained by the fact that the measurements of the antigen by either ELISA or HPLC are expected to be highly noisy due to the presence of many proteins in the extracellular environment of the cell culture.

Table 4.7: Akaike Information Criteria (AIC) values for regression models.

Variable	Global		Set 1		Set 2		Set 3	
	Areas	Areas + additional parameters	Areas	Areas + additional parameters	Areas	Areas + additional parameters	Areas	Areas + additional parameters
OD	36.6	37.8	31.3	31.8	36.4	35.0	33.3	35.6
Glu	38.8	29.76	31.7	31.5	35.4	28.9	32.7	28.2
Pro	40.8	30.88	34.4	32.3	36.6	35.1	33.7	34.9
PRN	50.1	48.2	51.1	46.8	12.5	24.1	48.2	53.2

4.5 Conclusions

In this work, the capability of fluorescence spectroscopy as an in-line monitoring tool was demonstrated using a 2-D fluorometer. The robust and non-invasive design of the fluorometer with just two light sources makes it a practical solution in an industrial setup.

The technique was used to monitor 4 target variables in the bacterial cell cultures. The fluorophores which were part of cell's redox state were monitored using the online fluorometer. Fluorescence maps were acquired in-line over the batch course. Fluorescence areas obtained from these maps along with commonly measured online parameters such as agitation, dissolved oxygen and culture volume were used in regression models. The performance of the regression models was found to be dependent on the cell line and media used for the model calibration. Using separate models for different media and different cell lines was crucial for improving the predictions. Also, the inclusion of additional measured fermentation variables such as agitation, culture volume change over time and dissolved oxygen in the model improved the prediction accuracy. The models were formally compared by the Akaike Information Criteria to take into account the differences in number of samples and number of model parameters considered for the development of each model. While the Akaike criterion corroborated the superiority of the separate models with additional inputs to predict OD, glutamate and proline, the results were less conclusive for prediction of the antigen (pertactin). The lower accuracy obtained for pertactin may be due to the noise in the off-line measurements of this variable by either ELISA or HPLC. Overall, these results suggest that 2-dimensional fluorometry combined with multivariate models is a promising Process Analytical Technology for monitoring microbial cultures in the pharmaceutical industry.

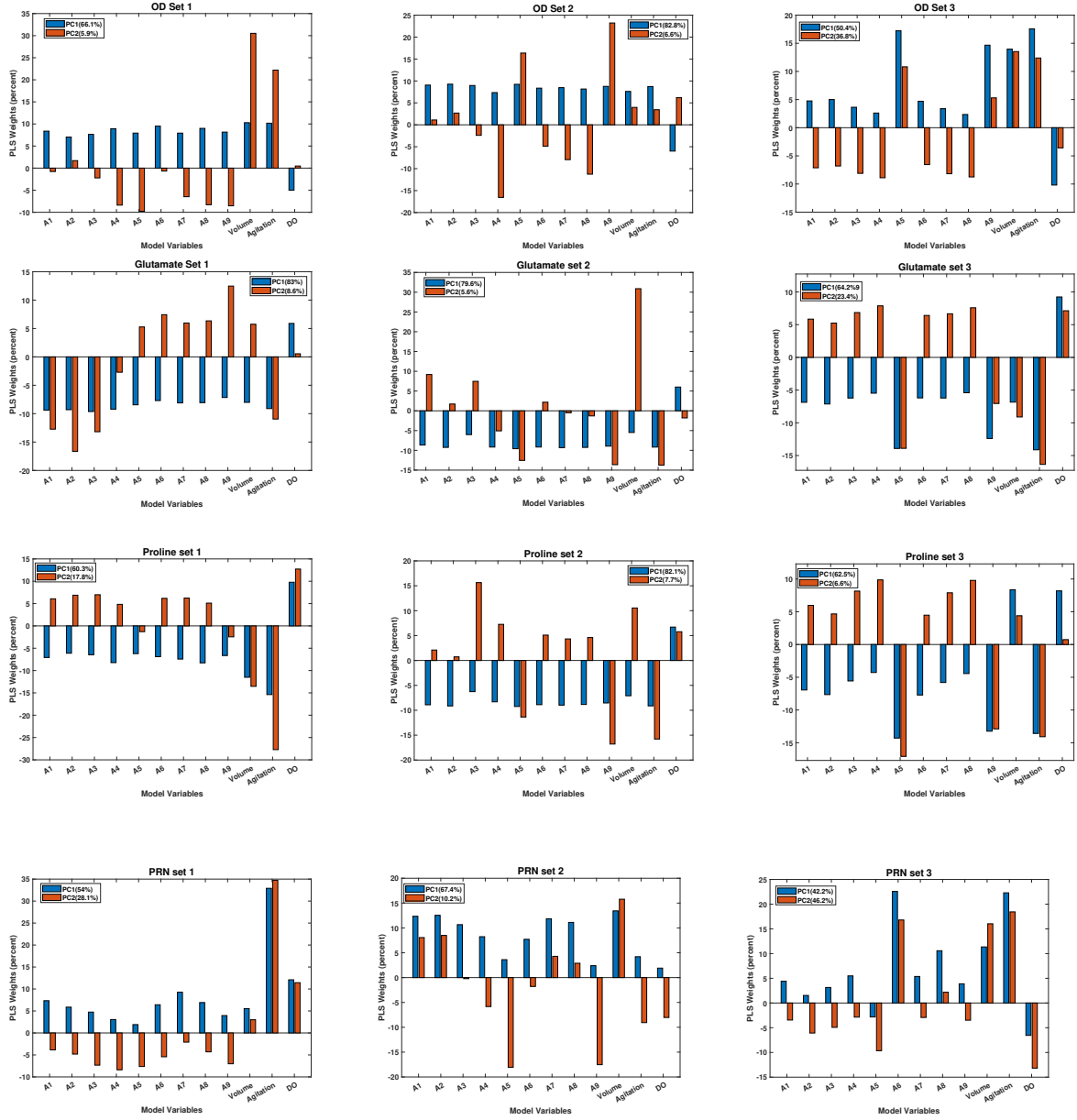


Figure 4.6: PLS weight plots of first two principal components for model set 1, 2 and 3.

Chapter 5

Correlation Between Oxidative Stress and Growth & Antigen Productivity in *B. Pertussis* cultures

5.1 Overview

¹ In this work, we explore correlation between initial glutamate concentration, growth and pertactin productivity, a key antigen in the acellular pertussis vaccine for *B. pertussis* . The study employs a combination of cytometry, chromatography, and fluorescence spectroscopy to investigate the relationship between oxidative stress and antigen productivity. We found that higher initial glutamate levels lead to increased oxidative stress, resulting

¹Adapted from (Mishra et al. [2024]). “Effect of oxidative stress on antigen productivity in *B. Pertussis* cultures. Biochemical Engineering Journal 210 (2024): 109421.”

in prolonged lag phases, slower bacterial growth, and reduced pertactin yield. A strong inverse correlation between oxidative stress and pertactin synthesis is observed, indicating that elevated stress conditions significantly impair antigen production. The effects are particularly pronounced in bioreactor experiments, where closed-loop oxygen control exacerbates the reduction in productivity compared to flask cultures. This research underscores the importance of controlling glutamate concentrations to minimize oxidative stress and ensure consistent, high-yield production of pertactin for vaccine manufacturing.

5.2 Introduction

Bordetella pertussis, commonly known as *pertussis*, is a gram-negative, aerobic, pathogenic, encapsulated coccobacillus bacterium that causes whooping cough ([Organization et al. \[2014\]](#)). This highly contagious disease causes infection in the respiratory tract and it is manifested by severe coughing fits. It can be especially dangerous for infants and young children, often leading to severe complications and even death. Whooping cough remains a significant danger to infants, causing approximately 50 million cases and more than 100,000 deaths every year ([Thalen \[2008b\]](#)) thus making it one of the primary causes of vaccine preventable deaths ([Mooi et al. \[2009\]](#)).

The acellular pertussis vaccine produced by Sanofi contains a total of five antigens: *Fimbriae* (type 2 and 3), *Filamentous Hemagglutinin*, *Pertactin*, and *Pertussis toxin* ([Borisova et al. \[2007\]](#)). These antigens are isolated and purified after harvesting the bacteria from an upstream fermentation process. One of the major process challenges is the low and variable

yield of pertactin, making it the bottleneck of the pertussis vaccine manufacturing process. Pertactin is an auto-transporter protein found on the surface of *Bordetella pertussis* and it is one of the key component of DTaP (diphtheria, tetanus, and acellular pertussis) vaccine of Sanofi (Zlamy [2016]). Auto-transporter proteins, including pertactin (PRN) found in Gram-negative pathogens, consist of three functional parts: the N-terminal signal sequence, the passenger domain, and the C-terminal auto-transporter domain. The signal sequence directs the passenger and transporter domains to the periplasm, where the transporter domain creates a pore in the outer membrane which allows the passenger domain to move to the cell surface. PRN passenger domain is then cleaved in a unique intra-barrel auto-catalytic reaction. Following cleavage and release of the passenger domain, only a small fragment of the original alpha helix remains inside the barrel. This cleavage happens either by an uncharacterized periplasmic/outer membrane protease or by an auto-catalytic mechanism that is not easily predicted from the protein sequence. (Girard and Mourez [2006], Noofeli [2008]).

Based on previous studies (Zavatti et al. [2019a, 2020a], Vitelli et al. [2023a], Agarwal [2022]), we hypothesize that oxidative stress can be one of the causes of variability in productivity of PRN. Oxidative stress arises from an imbalance between production and neutralization of reactive oxygen species (ROS) (Cabiscol Català et al. [2000]). Reactive oxygen species are produced during respiration within the cells. ROS, including superoxide radicals (O_2^-), hydrogen peroxide (H_2O_2), and hydroxyl radicals ($\cdot OH$), can be harmful to cells as they can cause oxidative damage to proteins, DNA, and lipids (Lushchak [2001a], Varatnitskaya et al. [2021]). Bacterial cells have evolved mechanisms for quenching ROS

by reactions mediated by enzymes such as Superoxide Dismutase (SOD), Catalase and Glutathione peroxidase (Lushchak [2001a]). These defense mechanisms against ROS are crucial for bacterial survival, especially in environments where oxidative stress is prevalent, such as during host-pathogen interactions (Varatnitskaya et al. [2021]). The state of imbalance between the high ROS production during respiration and antioxidant production to quench ROS can lead to intracellular accumulation of reactive oxygen species thus causing oxidative stress (Fasnacht and Polacek [2021], Storz and Imlay [1999]). Previous studies by our group have shown a correlation between ROS levels and final productivity measured after purification. Also, high levels of NADPH that plays a key role in antioxidant reactions were found in samples collected for batches of reduced productivity which further hinted at high ROS levels. However, since these studies were solely based on manufacturing samples from the end point of the process after purification, we could not clearly identify the impact of each stage of the manufacturing process on the variability and reduction in productivity. For instance, while fermentation was suspected to have severe impact on productivity, this effect could not be systematically studied. The study of productivity during fermentation was further challenged by the high variability of the ELISA measurements of antigen performed at that stage.

Excess nutrient media in the culture can be one of the causes of elevated oxidative stress conditions during the initial phases of the batch cultures. The most abundant amino acid utilized by *B. pertussis* is glutamate, which serves as a major carbon source for its metabolism. While *B. pertussis* can also utilize other amino acids like proline and arginine for growth (Thalen et al. [1999]), it has been shown that it can grow solely with glutamate.

Researchers in (Wellen and Thompson [2010b]) showed that for a micro-algal culture, nutrient concentration beyond the normal levels required for physiological functions leads to overproduction of ROS due to factors such as oncogene activation or prolonged exposure to growth factors in the presence of abundant nutrients. To cope with elevated levels of intracellular metabolites, cells have developed various defense mechanisms such as utilizing ROS to promote signaling and proliferation, activating the mTOR pathway, and regulating signaling and transcription through modifications that are sensitive to metabolites. In (Wang and Levin [2009a]), the authors described how bacteria cope with nutrient stress by adjusting their cell cycle parameters, such as delayed initiation of chromosome replication and the elongation and segregation of replicated chromosomes. The correlation between oxidative stress and productivity has been reported for some bacteria but the relative effect of growth versus protein synthesis rate on antigen productivity has not been clearly elucidated.

The current work focuses on the role of high glutamate concentration in inducing oxidative stress among the *B. pertussis* fermentation and on how this stress correlates with productivity of pertactin (antigen). While in earlier studies (Zavatti et al. [2019a]) hydrogen peroxide (H_2O_2) was used to artificially induce oxidative stress, in the current study the stress was induced by varying concentration of glutamate. This was done to emulate reported manufacturing conditions involving variability in the initial glutamate concentration. While in our earlier studies only the effect of stress on growth was assessed, the current study focuses on the effect of stress on both growth and productivity. To this purpose, three different methodologies were used:

- A multi-parametric flow cytometry method to simultaneously analyse internal ROS and membrane bound pertactin.
- A chromatography based method to quantify secreted pertactin in the broth based on a peptide specific binding to pertactin.
- A spectro-fluorometric method was used for qualitative analysis of NADPH and riboflavin in the supernatant.

While the first method provided an instantaneous status of cell surface pertactin and intracellular ROS levels, the second method permits quantification of overall antigen secretion into the fermentation supernatant over time. The spectro-fluorometric measurement serves to corroborate the occurrence of high levels of NADPH for high ROS conditions found in our earlier studies. The cytometric measurements also permits the investigation of the heterogeneity of the cell population with respect to membrane bound pertactin and ROS levels at different stages of the culture.

The experimental results indicate a strong correlation between glutamate-induced oxidative stress and low productivity of *B. Pertussis* batches. It is shown that a higher oxidative stress is associated to a longer lag phase which eventually leads to slower growth and lower pertactin productivity. Using the particular analytical techniques mentioned above, we are able to differentiate between the respective correlations of growth versus antigen synthesis rate to the overall reduction in secreted antigen in the presence of high ROS. It will be shown that productivity is mainly related to lower growth where the latter is correlated to high ROS. Experiments were performed in both flasks and bioreactors in

order to study the effect of the dissolved oxygen and pH closed loop controllers that are available for the bioreactor but are not available in the flasks. The comparison between flask and bioreactor operation shows that under closed loop oxygen control, the effects of increased glutamate and resulting high stress on productivity is further exacerbated due to reduced agitation rate.

5.3 Materials and Methods

5.3.1 Flask Culture Conditions and Experimental Setup

Flask cell cultures were grown from frozen *Bordetella pertussis* seeds (*B. pertussis wild type*) in 500 ml flasks (120 ml working volume). Since *Bordetella pertussis* is pathogenic, all the experiments were performed in a biosafety level 2 containment level. Modified Stainer-Scholte broth (cP media) along with growth factor (a mix of amino acids and salts) was used for culturing pertussis. Cultures were maintained in an incubator at a temperature of 36 °C and shaken at 200 RPM. The initial pH of the flask culture was 7.2. The cultures were grown for 48 hrs and sampling was done at the end of 6, 18, 36 and 48 hrs. Sampling times were chosen to monitor the cultures during their lag phase, onset of exponential phase, exponential phase and stationary phase where all the carbon source is depleted. Three cultures with one repeat each were conducted: one with normal cP media, one with 20 % higher glutamate and one with 40 % higher glutamate. Table (5.1) summarizes each culture along with differences in the culture in terms of maximum optical density, initial glutamate concentration and pertactin productivity. The values of

OD, initial glutamate and pertactin yield are standardized due to confidentiality.

Table 5.1: Summary of batch runs.

<u>Batch ID</u>	<u>Initial glutamate concentration (Normalized)</u>	<u>Batch age (Hrs)</u>	<u>OD at 32 hrs (Normalized)</u>	<u>Cumulative PRN yield (Normalized)</u>
A (flask)	0.72	48	0.94	1
B (flask)	0.86	48	0.83	0.74
C (flask)	1	48	0.80	0.67
D (bioreactor)	0.71	32	1	0.89
E (bioreactor)	0.86	32	0.73	0.61

5.3.2 Bioreactor Conditions and Experimental Setup

Experiments were also conducted in 2 liter batch reactors (Biostat B plus, Sartorius) with 1.6 liter working volume. Four bioreactor cultures were started, two repeats with cP media (control) and two repeats with cP media with 20 % higher glutamate. The bioreactor runs were operated for 32 hours, and the temperature and pH were controlled at 36 °C and a pH of 7.2 respectively, similar to the conditions used in flasks. Dissolved oxygen (DO) concentration was maintained at 35 % by sequential manipulation of agitation and aeration. Initial agitation and aeration rates were set to 200 rpm and 2 liters per minute respectively.

5.3.3 OD and Glutamate Quantification

Samples were collected at four different time points. Optical density was measured using a spectrophotometer (Thermo Scientific, GENESYS 40/50 Vis/UV–Vis (2020), Canada).

Glutamate concentration was measured at different time intervals using a fluorometric glutamate assay kit (Abcam, catalogue - ab138883, Canada). A calibration curve for fluorometric quantification was obtained using glutamic acid stock solutions provided with the kit. Samples were filtered using a 0.2 μm syringe filters and the assay was performed as per the protocol provided with the kit. 25 μL aliquot of each sample was mixed with 25 μL of working reagent in a 96-well transparent microplate and incubated at room temperature for 60 min. The fluorescence (Ex/Em - 540/590 nm) was measured using a Biotek Gen5 microplate reader (Agilent technologies, (2020), Canada). Repeats of the readings were taken and averaged.

5.3.4 Fluorescence Microscopy

To visualise the effect of ROS on the expression of cell surface pertactin, fluorescence microscopy was utilized. A Zeiss LSM 700 confocal microscope (Carl zeiss, (2019), Germany) with four excitation lasers (405 nm, 488 nm, 555 nm and 633 nm) was used to obtain the images. Cell samples were taken from the flasks and centrifuged at 10000 RPM for 3 minutes, washed and re-suspended in PBS. Washed cell cultures were diluted to an OD of 0.01. For detection of ROS among cell cultures using fluorescence microscopy, a fluorescein derivative (2',7'-dichloro dihydrofluorescein diacetate (H_2DCFDA)) (Thermo Fisher, catalogue no. D399, ex/em - 503/523 nm) was used. 1 mM stock solution of (H_2DCFDA) was prepared using DMSO as a solvent. 5 μL of the stock solution was then used to dye the washed sample. For visualising the surface expression of pertactin, Cy5 (abcam, catalogue - 188288) conjugated with pertactin antibody (provided by Sanofi) was used. 2 μL

of Cy5 conjugated with pertactin antibody was used to dye 1 ml of washed and diluted cell culture. The sample was then incubated for 45 minutes. After incubation, the sample was washed and re-suspended in PBS. Glass slides (VWR, catalogue no. 16004422) coated with 0.1 percent (v/v) poly- L- lysine solution were used to visualize stained cells under the microscope for a qualitative analysis.

5.3.5 Multi Colour Flow Cytometry for ROS and Surface Expression of PRN

Detection of reactive oxygen species (ROS) and analysis of surface expression of PRN of the bacterial cell population was done using a cell sorter (S₃E cell sorter, catalogue no. 12007059, Bio-Rad Laboratories, CA, USA). The instrument was equipped with 2 excitation lasers at 488 nm (primary) and 640 nm (secondary), along with emission filters for forward scattering (FSC) and side scattering (SSC) detection (488/10 band width filters), and fluorescence emission detection: 526/48 (FL1), 593/40 (FL2), 670/30 (FL3) and 700 LP (FL4). In this analysis, the FL1 detector was used for green fluorescence (H₂DCFDA), the FL3 detector for red fluorescence (Cy5 conjugated with pertactin antibody, ex/em - 640/700 nm). While (H₂DCFDA) is widely used for measuring ROS, it should be used cautiously due to its susceptibility to auto-oxidation and potential interference from other cellular components ([Halliwell and Gutteridge \[2015\]](#)). The use of various ROS measuring methods were compared in (Zavatti et al, 2019) and the carboxy- H₂DCFDA based method was found to work best for B. pertussis. The total number of events per sample was 20,000 acquired at a rate of 500-700 events/sec. (H₂DCFDA)) (Thermo Fisher, catalogue no.

D399, ex/em - 503/523 nm) was used for the detection of ROS. 1 mM of the reagent was prepared by dissolving it in DMSO and the final concentration of reagent per sample (1 ml) was 50 μ M. Cell samples were washed, re-suspended in PBS, stained with both the reagents (H₂DCFDA and Cy5) and then incubated for 1 h at room temperature. For the analysis of surface expression of PRN, Cy5 dye (Ex/Em - 640/700 nm) (abcam, catalogue no. - 188288) was conjugated with pertactin antibody (provided by Sanofi). 1 ml of cell culture was washed, diluted to an OD of 0.1 and re-suspended in PBS.

Flow cytometry was used to simultaneously monitor the general trend of ROS and surface pertactin expression among the cell populations in different growth phases. The average level of ROS and surface PRN of the cell population was determined using a weighted average (WA) calculation. The weighted averages for ROS and PRN were calculated as per the following equation:

$$WA = \sum_{i=1}^n \frac{C_i}{C_t} * I_i \quad (5.1)$$

Where C_i is the number of cells that have fluorescence intensity I_i , and C_t is the total number of cells analyzed.

The rates of change of the weighted averages of ROS and PRN were calculated according to the following formulae:

$$\text{WA rate (ROS)} = \frac{WA_{\text{ROS},t} - WA_{\text{ROS},t-1}}{T_t - T_{t-1}} \quad (5.2)$$

where $WA_{\text{ROS},t}$ is the weighted average of ROS at time point t , $WA_{\text{ROS},t-1}$ is the weighted

average at time point $t-1$, and T_t & T_{t-1} are time (in hours) at time points t and $t-1$.

And,

$$\text{WA rate (PRN)} = \frac{\text{WA}_{\text{PRN},t} - \text{WA}_{\text{PRN},t-1}}{T_t - T_{t-1}} \quad (5.3)$$

where $\text{WA}_{\text{PRN},t}$ is the weighted average of PRN at time point t , $\text{WA}_{\text{PRN},t-1}$ is the weighted average at time point $t-1$, and T_t & T_{t-1} are time (in hours) at time points t and $t-1$.

5.3.6 Chromatographic Purification of Pertactin

Purification of pertactin was achieved using 1 ml affinity columns. The method is elaborated in the following sections.

Ligand Coupling and Column Preparation

Affinity chromatography using a custom peptide (obtained from Bachem, Torrance, CA, USA) was used for quantification of pertactin in supernatant. The protocol was developed by (Umatheva et al. [2020b]) for quantifying pertactin and was modified and used to achieve protein purification of samples collected from the fermentation broth. The amino acid sequence of the affinity peptide is not disclosed here due to confidentiality. The primary amine ($-\text{NH}_2$) of the peptide molecule was coupled to N-hydroxysuccinimide (NHS) activated resin prepacked in a 1 mL column (Cytiva, catalogue no. - 17090601, MA, USA) according to the manufacturer's protocol. The peptide was dissolved in coupling buffer (200mM NaHCO_3 , 0.5M NaCl , 0.1 percent DMSO (v/v), pH 8.3). Before passing the peptide through the column, it was washed with 6 column volumes (CVs) of 1 mM cold HCl ,

using a syringe at a flowrate of 0.5 drops/sec. Coupling buffer containing peptide was then passed through the column and was allowed to react with the resin at room temperature for 4 hours. To remove and block any unreacted ligand, the column was washed with 6 CVs of 0.5M ethanolamine (Thermo Fisher, Mississauga, Canada) prepared in 0.5M NaCl (pH 8.3) at 1 mL/min flow rate. The column was then washed at 1 mL/min flow rate using 6 CVs of sodium acetate (Sigma Aldrich, Oakville, ON, Canada) in 0.5M NaCl (pH 4.0). This blocking and washing sequence was repeated and finally the column was stored at 4 °C in 20 % ethanol.

Antigen Purification

Chromatic purification of pertactin was performed using the affinity column coupled with the ligand. The samples fed for protein purification were broth samples filtered with 0.2 um syringe filter collected at different times during the culture. The column was washed with milli-Q water and was equilibrated with 6 CV of equilibration buffer A: 50 mM Tris (pH 7.5). Then, the column was loaded with 10X concentrated feed material. After sample loading, buffer A was fed into the column. Subsequently, the column was washed with wash buffer consisting of 50 mM Tris pH 7.5, 0.1 percent (v/v) tween-80, 1.0 M NaCl (sigma aldrich). Elution was achieved using the eluting buffer (100 mM sodium citrate, pH 3.5). A 1 ml sample of eluate was collected and was immediately neutralised using 0.125 ml of Tris-HCl (pH 8.0) to prevent pertactin denaturation.

5.3.7 SDS PAGE Analysis

The wash and eluate fractions collected from the column were analyzed using pre-cast 12 % Bis-Tris SDS protein gels (Bio-Rad). Wash and eluate samples along with pure pertactin (from Sanofi) were prepared according to the manufacturer's instructions under the reducing condition (in presence of β -mercaptoethanol) and loaded into the gel wells. The gel was then stained and kept overnight with coomassie blue stain (BioRad) solution (0.1 mg dye, 40% methanol, 40% 1 N acetic acid and 20% milliQ water). The gels were then destained using the destaining solution (40% methanol, 40% 1 N acetic acid and 20% milliQ water). Destained gels were then transferred to a densitometer (BioRad) for image analysis. SDS PAGE result is presented in figure (5.11) showing the pertactin band obtained after affinity chromatography.

5.3.8 Spectral Analysis of Culture Supernatant

Spectral analysis of culture supernatant was conducted using a fluorescence spectrophotometer (Agilent technologies, CA, USA). Cell cultures were centrifuged and then filtered using 0.2 μ m syringe filter. Filtered supernatants were diluted and loaded into fluorometric quartz cuvettes and their fluorescence intensity was analyzed at different sampling points. Excitation range of 350 - 370 nm was used with the emission range of 200 - 700 nm.

5.4 Results

5.4.1 Fluorescence Microscopy

Fluorescence microscopy was used as a preliminary step to visualize the surface expression of PRN and reactive oxygen species generated by the cells. The observations are qualitative since they are based on cell counting for individual images. Figure (5.1) shows the surface expression of pertactin (red) and reactive oxygen species (green) during lag, exponential and death phase of bacterial cells. It was observed that surface pertactin expression is negligible in the early growth phases of *B. pertussis*. In the lag phase, the bacteria are adapting to the new environment and undergoing rapid replication. During this lag phase and early exponential phase, cells focus on replicating their genetic material and increasing their population. Hence, they may not yet be exposed to the specific environmental conditions that trigger the production of virulence factors like pertactin (Madar et al. [2013]). We hypothesized that a longer lag phase and slower growth due to initial high concentration of glutamate causes a delayed expression of pertactin which may eventually lead to lower productivity of batches. In the exponential phase, the ROS generated by the cells is lower when compared to the initial growth phase and also, a larger fraction of cell population has high surface expression of pertactin. Again, in the stationary phase, when the cells are running out of nutrients, an increase in the ROS production by cells and decrease in the PRN expression are observed. Since the observations based on fluorescence microscopy are qualitative, more rigorous quantification of surface PRN and intracellular ROS was achieved via flow cytometry as explained below.

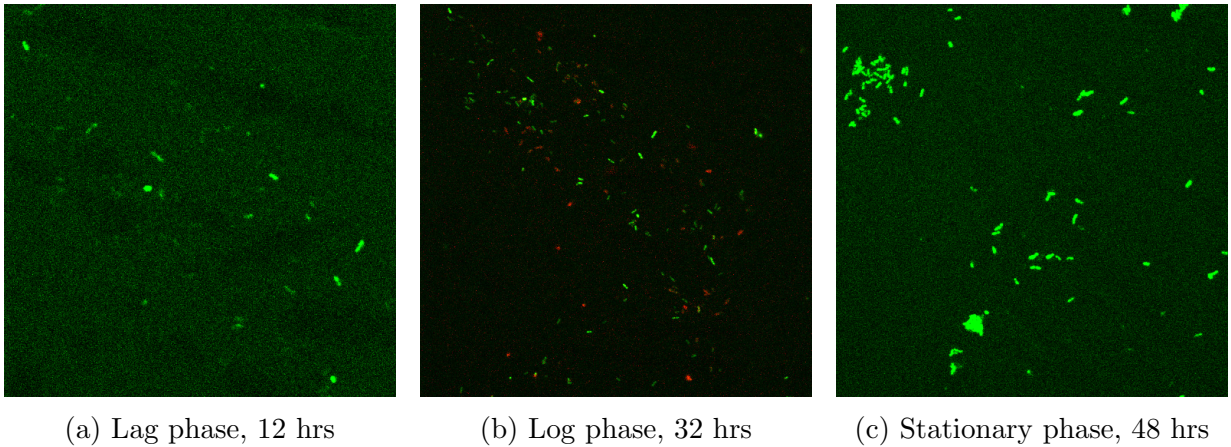


Figure 5.1: Fluorescence microscopy images illustrating the surface expression of PRN and the generation of Reactive Oxygen Species (ROS) by bacterial cells at different growth stages (100X zoom). During the initial lag phase, PRN expression is nearly absent (depicted by negligible red fluorescence) while ROS levels are high (depicted by green fluorescence). In the exponential phase, a substantial population of cells exhibits high PRN expression (depicted by red fluorescence), accompanied by a decrease in intracellular ROS levels (depicted by reduced green fluorescence). Subsequently, during the stationary phase, cells experiencing oxidative stress due to starvation display elevated ROS production (depicted by intensified green fluorescence) and reduced PRN expression (depicted by diminished red fluorescence).

5.4.2 Cell growth, Oxidative Stress and Productivity

Flask Cultures

Growth of the bacteria was monitored in three flask experiments with different initial glutamate concentration. Three repeats were done for each flask. Error bars for reported results are based on these repeats. Figure (5.2a) and (5.2b) shows the growth curve for the three cultures conducted with normal cP media (shown in blue), with 20 % higher glutamate (orange) and culture with 40 % higher initial glutamate (yellow) respectively. A slower growth in cultures with higher initial glutamate levels was observed, in comparison to the flask with defined cP media (control). The slower growth may also result from a possibly longer lag phase since the latter cannot be measured accurately. Higher oxidative stress levels were observed in the cultures with higher initial glutamate concentration and possibly longer lag phase duration. Also, the pertactin productivity for the cultures with high initial glutamate concentration was lower than control by the end of the initial lag phase. It has been established in literature that excessive nutrients can cause oxidative stress in both mammalian and bacterial cells ([Wellen and Thompson \[2010b\]](#), [Wang and Levin \[2009a\]](#)). This happens when the tricarboxylic acid (TCA) cycle oxidizes nutrients, generating electrons that form NADH and FADH₂. These electrons are usually transferred to molecular oxygen through the electron transport chain (ETC). However, if this process is too fast due to nutrient excess, incomplete reduction can result in the production of super-oxides. Consequently, when TCA cycle components break down faster than the electron chain cycle can handle, such as during nutrient excess, high levels of ATP and other energy-rich molecules can be generated. This may lead to feedback inhibition of enzymes in the

ETC, causing an increase of reactive oxygen species (ROS), which leads to oxidative stress (Imlay [2015]). A higher oxidative stress can affect the cellular redox balance contributing to metabolic disruption and enzyme inactivation (Imlay [2013a]). Thus, we argue that there is a correlation between higher levels of stress and combined slower growth and possibly longer lag phase occurring for the cultures with higher initial glutamate concentrations.

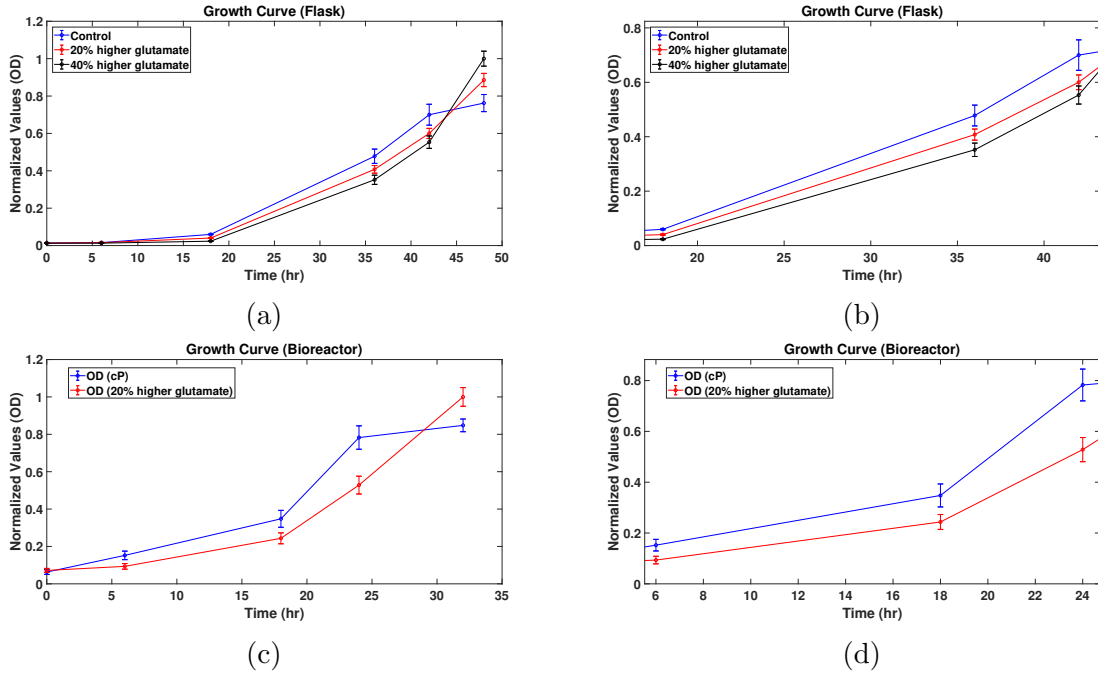


Figure 5.2: Bacterial growth curve for the flask and bioreactor runs. Subfigure (5.2b) and (5.2d) shows the zoomed in growth curve to show the growth inhibition due to higher concentration of glutamate, which leads to less growth leading to lower antigen productivity.

Figure (5.3) shows the weighted average distribution of fluorescence intensities of ROS and PRN over time, for the flask runs. Weighted average (WA) is the average of fluorescence intensities acquired by flow cytometry for 20,000 cells with ROS and PRN fluorophores. It

was observed that initially (during the lag phase), the weighted average of ROS is high and as the cells go into the exponential phase, ROS decreases and WA of PRN goes up. This inverse correlation between WA of ROS and PRN was also observed for cultures with 20% and 40% higher glutamate cultures. It should be noted that for the cultures with higher glutamate, the initial WA for ROS was higher than the WA for the culture with cP media indicating that the oxidative stress in the lag phase is higher in the cultures with higher glutamate concentration.

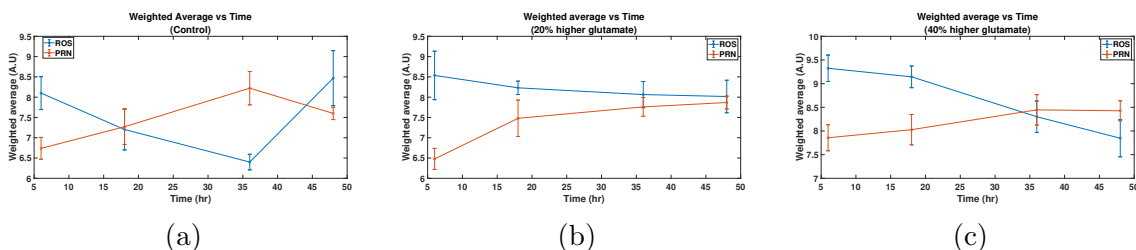


Figure 5.3: Plot of weighted average of ROS and PRN vs time for the flask runs.

The secretion of pertactin into the supernatant was quantified using affinity chromatography. The amount of pertactin in the supernatant was analyzed at each time point and it was observed that cultures with higher growth has higher quantity of secreted pertactin. In other words, flasks with lower oxidative stress have higher secreted pertactin productivity. Normalized pertactin yield for each flask run is summarised in table (5.1). The flask run with 40% higher glutamate had the lowest yield while the flask with cP media (control) had the highest pertactin yield.

Fluorescence spectra of the culture supernatant was analysed to detect the fluorescent compounds secreted by the bacteria. The nutrient media for pertussis does not contain

any fluorescent molecules but molecules such as NADPH fluoresce. Figure (5.4) shows the fluorescence emission spectra for the flask runs at the excitation wavelength of 340 nm. Emission spectra was acquired for 18, 36 and 48 hrs for the flask runs. Two distinct peaks were observed in the spectra which can be related to NADPH (emission peak around 440-450 nm) and riboflavin (emission peak around 530-550 nm) secretions (Mishra et al. [2023a], Odanaka et al. [2017]). These two compounds are essential co-factors involved in redox reactions. Riboflavin is a precursor for the synthesis of flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD). These flavins play a role in various enzymatic processes, including those related to the electron transport chain and redox homeostasis (Green and Paget [2004]). A higher concentration of NADPH and riboflavin in the culture supernatant have been found correlated with higher ROS among cell cultures. Higher concentrations of glutamate can influence the balance of NADPH and NADP^+ (oxidized form of NADPH) in the cell, potentially leading to a more reduced intracellular environment. This reduction could contribute to increased oxidative stress, as certain metabolic pathways may generate ROS under such conditions (Lemire et al. [2017]).

Bioreactor runs

Experiments under similar conditions to the ones used in flasks were also conducted in 2-liter bioreactors to assess the impact of varied glutamate concentrations on both growth and pertactin yield. A key difference between the flask and bioreactor conditions, is that in the former the DO and pH are not controlled whereas in the latter they are kept at fixed set points with closed loop controllers. Control of pH is achieved by acid addition while control of DO is achieved by sequential manipulation of agitation and aeration rates. The growth

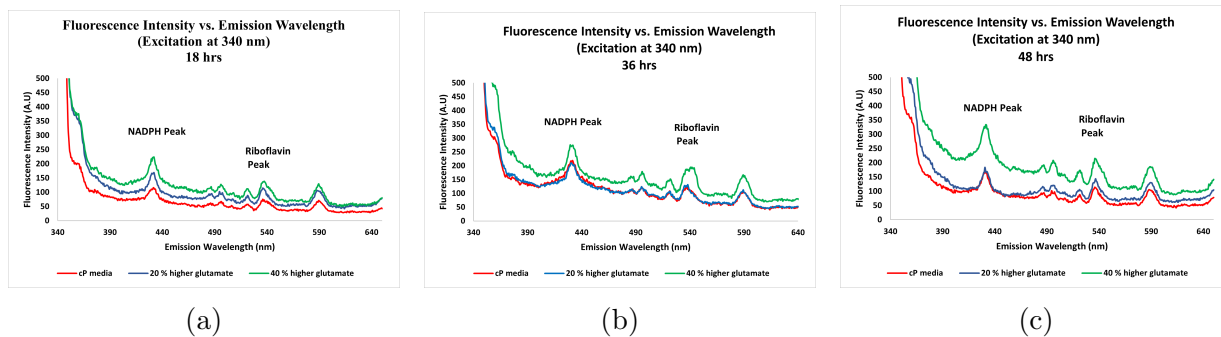


Figure 5.4: Fluorescence Intensity Dynamics Across Emission Wavelengths in Three Flask Experiments. Sub-figures (5.4a), (5.4b), and (5.4c) represent the fluorescence intensity profiles at 18, 36, and 48 hours, respectively. Notably, the flask with elevated glutamate concentration consistently exhibits higher intensity levels across the entire emission spectrum.

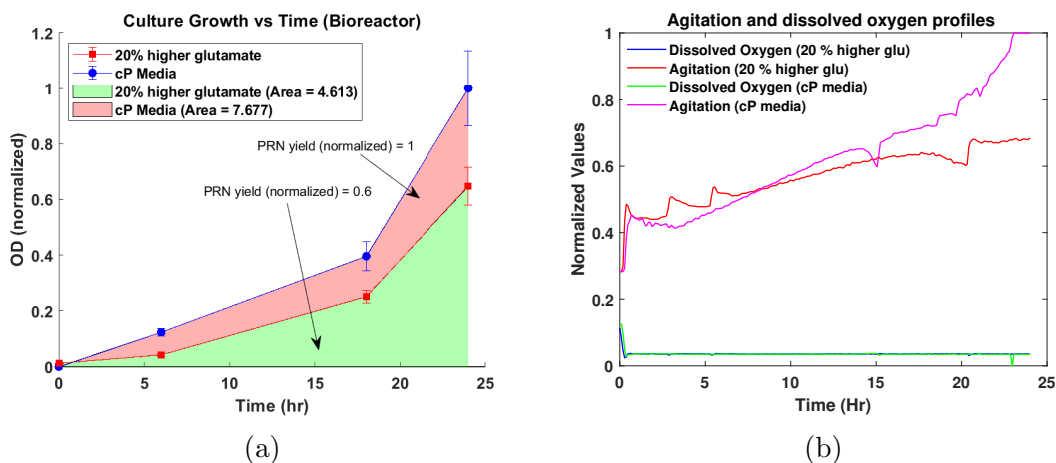


Figure 5.5: Bacterial growth curve comparing area under the curve and pertactin yield for the bioreactor runs (shown in figure a) and agitation and dissolved oxygen profiles of the bioreactors (shown in figure b). Agitation rate varied from 200 to 700 RPM to maintain the dissolved oxygen concentration of 35 %.

curves of two runs are depicted in the figures (5.2c) and (5.2d), for both conventional cP media (control) and with a 20 % higher glutamate concentration. Two repeats are conducted for each of these two experimental conditions and error bars are obtained from these repeats. Consistent with the trends observed in flask experiments, the bioreactor results demonstrated a slower growth for the culture subjected to elevated initial glutamate levels.

The values for weighted average of ROS and PRN were calculated and plotted over time. The graph depicts the evolution of the weighted average of reactive oxygen species (ROS) and surface pertactin (PRN) fluorescence intensities over the course of the experiment. Figure (5.6) shows the plot of weighted average vs time. The blue line on the graph illustrates the dynamic changes in the weighted average value of ROS, while the orange line represents the corresponding weighted average value of surface PRN. During the initial batch stage, a notable increase in the weighted average of ROS is observed, gradually declining over time. Simultaneously, the weighted average of surface PRN shows an inverse trend to ROS, i.e. increasing as ROS levels decrease. In the subsequent starvation phase, an increase in the weighted average of ROS corresponds to a decline in the weighted average value of PRN. Particularly noteworthy is the observation that cultures with 20 % higher glutamate concentration exhibit, at the 6-hour mark, a higher weighted average value of ROS compared to the control, similar to flask experiments. This suggests a noticeable impact of elevated glutamate levels on the oxidative stress response within the cell population. Weighted average of ROS for culture with control media increases again at the end of the batch as the carbon source is depleted completely inducing stress due to less nutrients (Niki et al. [2005]). During starvation, bacteria experience nutrient limitations.

This leads to an imbalance in cellular redox reactions, resulting in the overproduction of reactive oxygen species (ROS) by lipid peroxidation, a process where ROS attack unsaturated fatty acids in bacterial cell membranes. As the cell's defense mechanisms weaken due to nutrient scarcity, ROS-induced lipid-peroxidation cannot be properly balanced by ant-oxidant (Imlay [2013b]). In the cultures with 20 % higher glutamate, the weighted average of ROS is still lower and decreasing when compared to control as they still have glutamate left at the end of 32 hrs.

Moreover, the antigen yield from this culture was notably lower as compared to the culture maintained in normal media (normalized pertactin yield shown in table (5.1)). Figure (5.5a) shows the growth curve and pertactin yield in bioreactors for 24 hrs. As further discussed in the following section, the differences between the bioreactor runs may be attributed to the differences in operating conditions due to closed loop control. For example, Figure (5.5b) illustrates a comparison of agitation profiles between the two culture runs. Notably, the culture with higher initial glutamate exhibited significant reduction in agitation rate needed to maintain the same dissolved oxygen set point, a consequence of slower cell growth and lower oxygen consumption.

Samples from the bioreactor culture supernatant at 6, 18 and 32 hrs were also analyzed with the spectrofluorometer. The observed fluorescence spectra trends were similar to flask cultures as shown in figure (5.7). Accordingly, the cultures with higher glutamate concentration exhibited higher secretion of NADPH in cultures with higher glutamate concentration. This is part of the cellular defense mechanism to counteract the damaging effects of ROS which is induced due to initial high nutrient concentrations (Wellen and

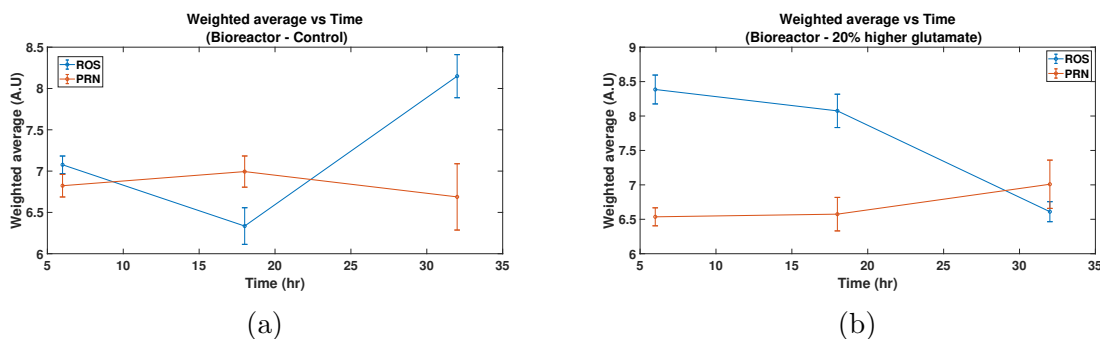


Figure 5.6: Plot of weighted average of ROS and PRN vs time for bioreactor runs.

Thompson [2010b]).

When comparing fermentation results of flask and reactor setup with high initial glutamate concentrations, we observed lower productivity in both, which is attributable to oxidative stress. However, the impact of oxidative stress appears more pronounced in the bioreactor. Specifically, the reduction in pertactin levels in the bioreactor is notably more drastic when compared to flask values.

5.5 Discussion

The experimental results from flask and bioreactor showed clear correlations between high initial glutamate concentration and ROS with growth and productivity. We observed that a high initial glutamate concentration in the bacterial culture is correlated to high oxidative stress. A higher oxidative stress in the culture was found inversely correlated with growth and productivity. Also, we observed that the use of dissolved oxygen control in the bioreactors can also impact the productivity as compared to flask operation where

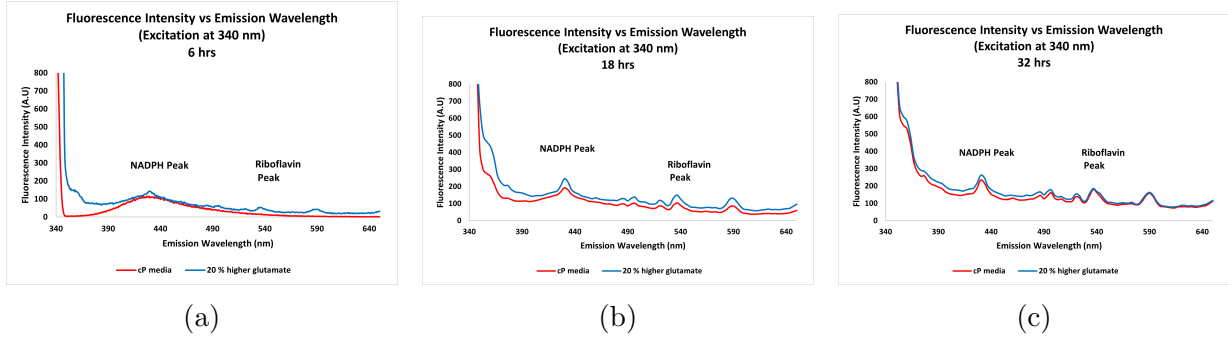


Figure 5.7: Fluorescence Intensity Dynamics Across Emission Wavelengths in bioreactor Experiments. Sub-figures (5.7a), (5.7b), and (5.7c) represent the fluorescence intensity profiles at 6, 18, and 32 hours, respectively. Notably, the culture with elevated glutamate concentration consistently exhibits higher intensity levels across the entire emission spectrum.

dissolved oxygen is not controlled. Furthermore, the application of two different analytical techniques, i.e. flow cytometry and chromatography, to quantify productivity can provide additional information on the relative contributions of growth rate and antigen synthesis rate to the total secreted antigen in the presence of high ROS. These contributions can be inferred from the following two correlations:

- In the current study for measuring surface pertactin, a direct high positive correlation was identified between the area under the biomass growth curve and secreted pertactin in the supernatant for both flask and bioreactor experiments. The significance of this correlation ($R^2 = 0.909$ for flask and $R^2 = 0.856$ for bioreactor) reflects the importance of biomass growth in determining the yield of secreted pertactin. Furthermore, the relation of secreted pertactin with respect to the integration of the biomass curve (cumulative cell hours) implies that the specific pertactin secretion rate is not highly variable along the culture. Figure (5.8b) shows the correlation

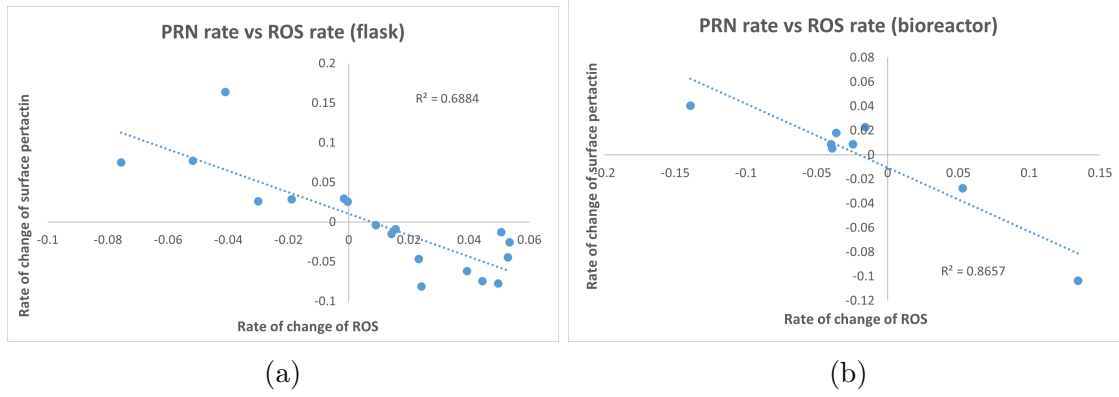


Figure 5.8: Correlations between the rate of change of the weighted averages of surface pertactin expression and ROS (calculated as per equations 5.2 and 5.3) for flask and bioreactor experiments.

between secreted PRN and AUC. Specifically, flasks in which growth was impeded by elevated concentrations of glutamate resulted in slower growth and subsequently lower areas under the growth curve compared to control conditions. This observed reduction in the growth of cultures corresponded with a decrease in pertactin productivity, suggesting that oxidative stress induced as a result of higher initial glutamate concentrations is highly correlated with reduced growth and overall yield of pertactin.

- A strong inverse correlation was found between the rate of change of weighted average of surface pertactin and rate of change of weighted average of ROS as shown in figure (5.8a) for flask runs and in figure (5.9a) for bioreactor. It was found that as the rate of change of ROS among the cell population increased, there was a corresponding decrease in the rate of change of surface pertactin expression. This negative correlation ($R^2 = 0.688$ for flask and $R^2 = 0.865$ for bioreactor) suggests a regulatory relationship

whereby high oxidative stress negatively impacts the production of surface pertactin. While such inverse correlation was hypothesized in our earlier studies, it was not proven before due to unavailability of cytometry measurements.

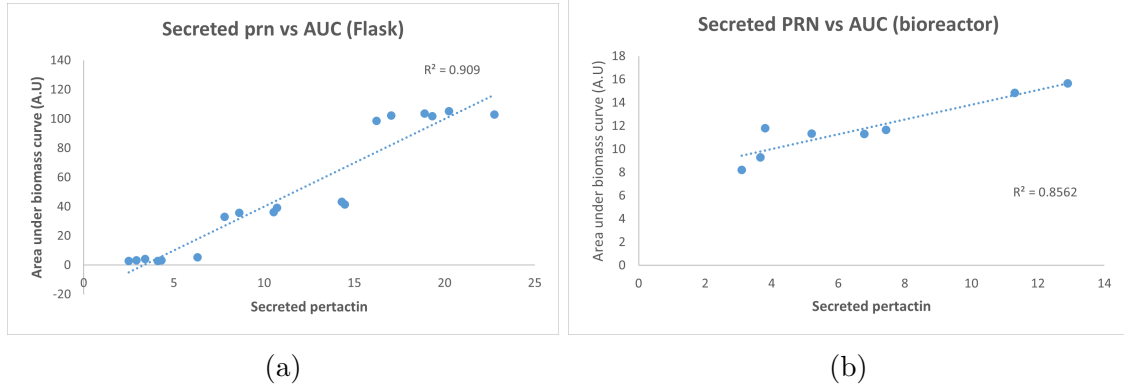


Figure 5.9: Correlation between secreted PRN and the area under the curve (AUC) for flask and bioreactor experiments.

The relationships between rate of change of ROS & surface pertactin and secreted pertactin & AUC were similar in both flask and bioreactor experiments. The replication of these correlations in the bioreactor underlines the robustness of the observed regulatory relationship between oxidative stress and surface pertactin accumulation.

The correlation of oxidative stress with growth and pertactin synthesis rate can be further elucidated by using balance equations as follows. Regarding the correlation between secreted PRN and area under the growth curve (AUC) in figure (5.8b), the secreted pertactin can be described by the following equations:

$$\frac{d(PRN_{\text{secreted}})}{dt} = \alpha * X_V \quad (5.4)$$

Where $d(PRN_{secreted})/dt$ is the secreted PRN rate, α is the secretion rate and X_V is the biomass. If α is assumed constant, Equation (5.5) can be modified and written as follows:

$$PRN_{secreted} = \alpha \int X_V * dt \quad (5.5)$$

Equation (5.5) describes a linear mathematical relation between secreted PRN and AUC that results from the assumption of constant specific secretion rate. Since figure (5.8b) shows an almost linear relation between pertactin and AUC, it can be concluded that the secretion rate α is approximately constant. No statistically significant correlations were found between α and either ROS or surface PRN concentrations.

On the other hand, a balance Equation (5.6) can be used to describe the relationship between the oxidative stress and surface PRN as follows:

$$\frac{d(PRN_{surface})}{dt} = \frac{d(PRN_{cell})}{dt} - \frac{d(PRN_{secreted})}{dt} \quad (5.6)$$

where $d(PRN_{surface})/dt$ is accumulation rate of pertactin, $d(PRN_{cell})/dt$ is the transport rate from cell to surface and $d(PRN_{secreted})/dt$ is the secretion rate. If accumulation is assumed only at the surface, the transport rate to the surface would be equal to the synthesis rate of PRN within the cell. The accumulation rate is equivalent to the time derivative of the surface pertactin concentration. Then, based on (5.8) and since the secretion rate α is approximately constant, equation (5.6) implies that the transport rate is directly correlated to the accumulation rate. Consequently the transport rate of pertactin to the surface is inversely correlated with ROS as shown in figure (5.8a) and figure (5.8b) for flask and

bioreactor experiments respectively.

Furthermore, following the lack of correlation between the secretion rate α with ROS and/or surface pertactin, it can be concluded that the inverse correlation between growth and ROS is the dominant factor affecting the total secreted pertactin. Thus, while the surface pertactin concentration is found to be inversely correlated with ROS, this does not affect significantly the secreted amount of antigen in the range of experimental conditions that was considered.

In the context of pertactin production, these findings emphasizes the need for optimizing glutamate concentrations for achieving high growth with resulting high productivity.

A comparison of the flask and bioreactor experiments show a higher impact of ROS on growth and productivity as shown in Table (5.1). These differences can be explained by observed differences in glutamate consumption rates in the experiments. In the flask experiment, for the cultures with higher initial glutamate, we observed a higher rate of glutamate consumption compared to the flask with cP media as shown in figure . However, in the bioreactor, the trend was reversed since the glutamate consumption rate for the cP media (control) was lower than that for the culture with higher glutamate as shown in figure (5.10).

This heightened effects of ROS in the bioreactor can be attributed to closed-loop control. With higher initial glutamate concentration there is higher ROS and slower growth. When

oxygen is maintained constant at set point, the balance of dissolved oxygen is given as follows:

$$q_O = k_L \cdot (C_{\text{setpoint}} - C_{\text{saturation}}) \quad (5.7)$$

where q_O is metabolic consumption rate of oxygen, k_L is the mass transfer coefficient for oxygen in the bioreactor, C_{setpoint} is the desired dissolved oxygen concentration (setpoint) and $C_{\text{saturation}}$ is the current dissolved oxygen concentration in the bioreactor.

To maintain this balance, a significant decrease in agitation rate as shown in figure (5.5) is imposed by the controller thus reducing the mass transfer coefficient (kLa). This, in turn, results in a smaller corresponding glutamate metabolic consumption rate that contributes to significantly lower growth and antigen productivity in the reactor as compared to the flask where oxygen is not controlled.

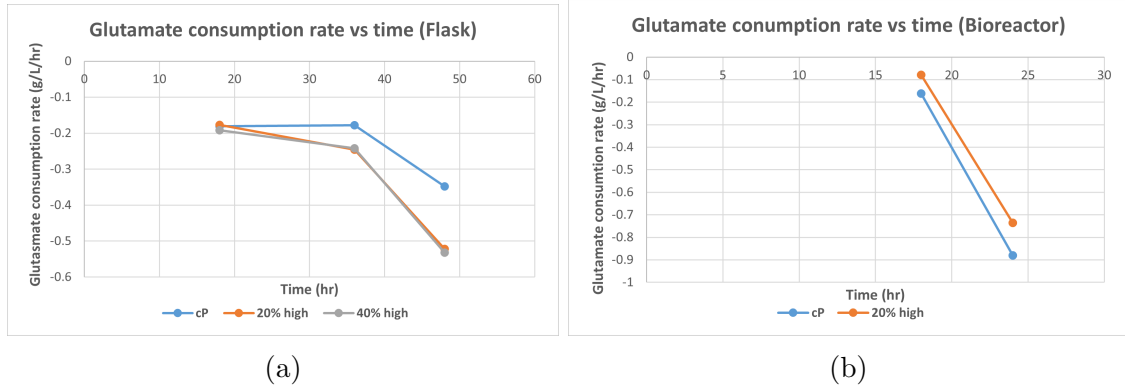


Figure 5.10: Glutamate consumption rates for flask and bioreactor experiments.

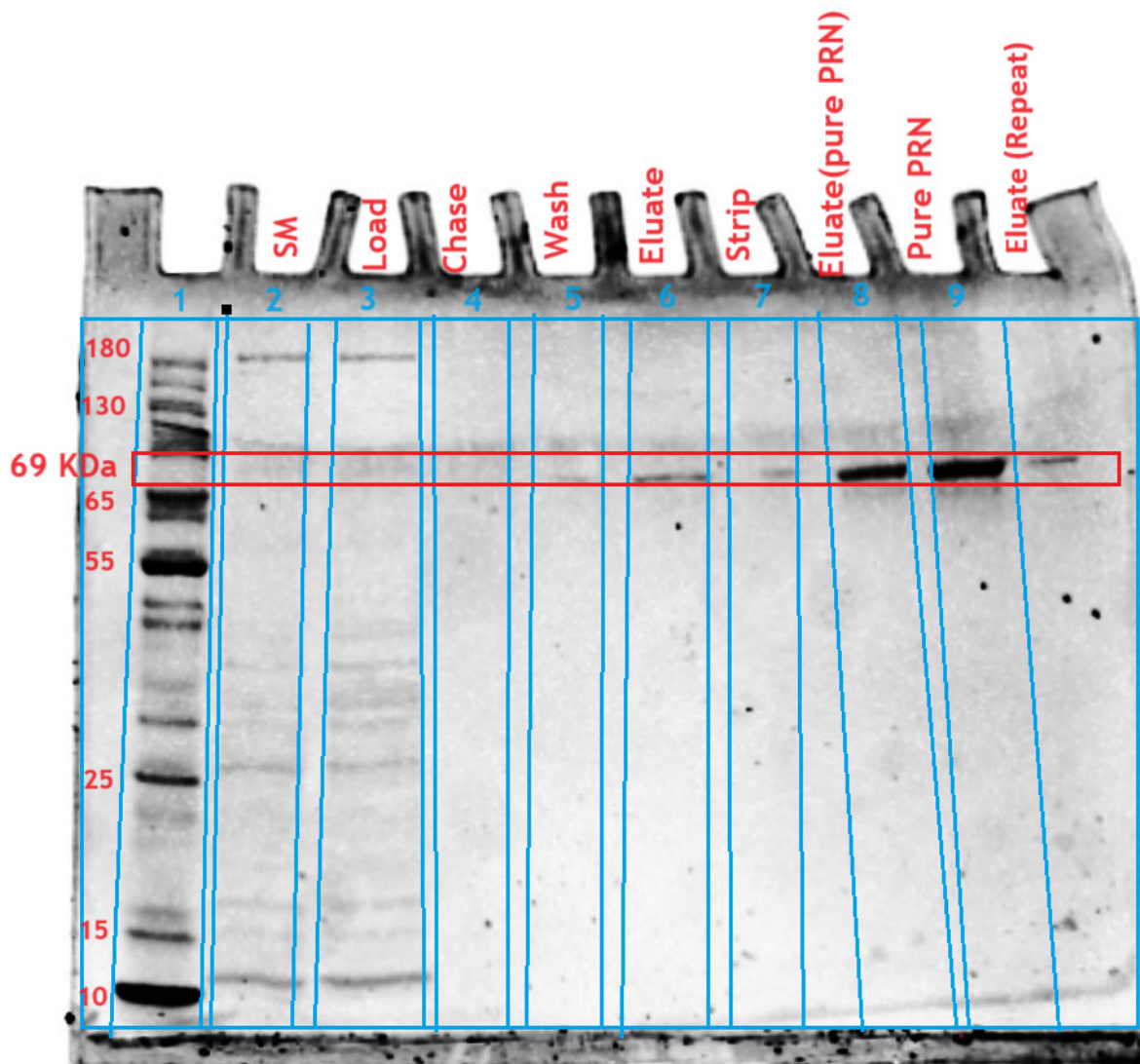


Figure 5.11: SDS PAGE analysis of culture supernatant and fractions collected from affinity chromatography using 1 ml columns. Samples loaded in SDS gel includes supernatant material(SM), first flow through the column (load), wash fraction, eluate and strip fraction. Pure pertactin was also passed through the column and eluate fraction was collected to validate the working of the column (well 8). Well 9 contains pure pertactin (provided by Sanofi).

5.6 Conclusions

The effect of oxidative stress induced by higher concentration of glutamate was investigated. In particular, the impact of oxidative stress on growth and pertactin productivity was studied. Surface expression of pertactin and the intracellular expression of reactive oxygen species (ROS) were monitored simultaneously by cytometry revealing an inverse correlation between oxidative stress and surface expression of pertactin. The weighted average of ROS exhibited a distinct pattern, initially peaking during the lag phase, decreasing during the exponential phase, and then rising again during the starvation phase. These trends were also corroborated by fluorescence microscopy. Cultures with elevated glutamate concentrations exhibited higher weighted average for ROS, emphasizing the direct link between nutrient composition, oxidative stress, and cellular responses.

The analysis also showed an inverse correlation between the rates of change of surface pertactin and ROS. This suggests a synchronized response, where alterations in surface pertactin expression align closely with changes in ROS levels. The quantification of secreted pertactin in culture supernatant was done using affinity chromatography. A robust correlation was established between the secreted pertactin and the area under the biomass curve, indicating a direct relationship between cellular growth and pertactin productivity. This correlation implies that increased biomass is associated with higher pertactin secretion, reinforcing the importance of optimizing growth conditions for enhanced bio-production. Since no significant correlation were found between the secretion rate and ROS or surface pertactin, it was concluded that growth and corresponding cumulative cell hours are

the determinant factor for the total antigen secretion. Higher secretion of antioxidants (NADPH) were observed as a cellular response to oxidative stress. The higher secretion of NADPH observed in cultures under increased oxidative stress, compared to control runs, underscores the cell's attempt to counterbalance the deleterious effects of oxidative stress through enhanced antioxidant generation and release. Finally, the study illustrated the effect of dissolved oxygen control in the presence of higher stress and corresponding reduced growth. The controller significantly reduced the agitation rate to maintain the oxygen set point in the presence of reduced growth. This resulted in significant reduction in growth and productivity as compared to the flask where control is not used.

In summary, excess initial glutamate, the key nutrient in pertussis cultures, was found to be a main cause of increased ROS and reduced productivity. This study corroborates the empirical observations made in our earlier studies with manufacturing samples where high ROS in fermentation resulted in lower productivity at the end of the downstream process. Hence, avoiding excess variability in initial glutamate is crucial to ensure satisfactory productivity. While factors other than glutamate may also result in high ROS and reduced productivity, the current work provides a systematic set of measurements and interpretations to assess these effects.

Chapter 6

A Mechanistic-Model Based Approach for Monitoring *B. pertussis* Fermentation

6.1 Introduction

Whooping cough, or pertussis, is a highly contagious respiratory illness caused by the gram-negative, aerobic bacterium *Bordetella pertussis*. Pertactin, a key virulence factor of *B. pertussis*, is a significant component of the pertussis vaccine. However, harvesting pertactin presents challenges in vaccine production due to its low concentration in fermentation broth and its variability in the final yield. Previous studies ([Zavatti et al. \[2019b\]](#), [Mishra et al. \[2024\]](#)) have shown a strong correlation between oxidative stress and the

growth and productivity of *B. pertussis* fermentations. Also, initially elevated concentration of the carbon source in the media was found to be correlated to higher oxidative stress levels. Hence, variability in the level of nutrients in media correlates with significant variability in growth and antigen productivity.

Oxidative stress is caused by an imbalance between the production of reactive oxygen species and a bacterial cell's ability to detoxify these reactive intermediates or repair the resulting damage. ROS may be described as extremely reactive oxygen-containing molecules: super-oxide anion ($O_2^{\bullet-}$), hydrogen peroxide (H_2O_2), and hydroxyl radical ($OH^{\bullet}$). These reactive oxygen species can damage and affect different cellular components, such as proteins, lipids, and nucleic acids. ROS exerts severe impacts on bacteria by inducing protein damage through the oxidation of amino acids, lipid peroxidation that damages cell membranes, and DNA damage from mutations, strand breaks, and base modifications, leading to genomic instability. In order to cope with oxidative stress, bacteria have developed several mechanisms of defense. Enzymatic antioxidants include superoxide dismutase (SOD), converting superoxide anions into hydrogen peroxide and oxygen, catalase decomposes hydrogen peroxide into water and oxygen, and peroxidases reduce hydrogen peroxide and organic peroxides to water and alcohols (Cabiscol Català et al. [2000]). NADPH, or nicotinamide adenine dinucleotide phosphate, is a coenzyme involved in a wide range of metabolic reactions including those that regulate oxidative stress. In *E. coli*, it is primarily produced by the pentose phosphate pathway and acts as a reducing agent for several biochemical processes. It plays a vital role in antioxidant defense by providing the reducing power needed to regenerate antioxidants such as glutathione and thioredoxin, which neutralizes

reactive oxygen species (ROS). Several studies on bacteria have found that inhibiting glycolysis pathway enables cells to redirect metabolic flux into the pentose phosphate pathway. This shift promotes NADPH synthesis, which boosts the cells' protection against oxidative stress (Singh et al. [2007]). Conversely, *B. pertussis* lacks an active glycolysis pathway and relies on gluconeogenesis to utilize the products of the TCA cycle, with glutamate serving as the primary nutrient source. In fact, *B. pertussis* can synthesize most amino acids and grow in media primarily containing glutamate and proline (which converts to glutamate). NADPH is produced from the conversion of isocitrate and glutamate to alpha-ketoglutarate and malate to pyruvate, rather than through the pentose phosphate pathway as observed in *E. coli* (Kanehisa and Goto [2000]).

In chapter (5), it was shown that glutamate concentration is highly correlated to growth and productivity of *B. pertussis* cultures. Previous work by (Wellen and Thompson [2010c], Wang and Levin [2009b]), and (Mishra et al. [2024]) showed that for batch cultures, higher initial concentration of glutamate (carbon source) can induce higher oxidative stress among *B. pertussis* cultures, leading to slower growth and lower antigen productivity. In this chapter, a mechanistic model is used to describe the effect of ROS on growth and antigen productivity in *B. pertussis* batch followed by fed-batch operation. The final goal of the current study is developing a model based sensor in combination with fluorescence spectra acquired on-line. The mechanistic model adapts Himeoka and Kaneko's theoretical framework (Himeoka and Kaneko [2017]), which originally described growth inhibition from the interaction between growth promoters (ribosomes) and waste products (misfolded proteins). Our model specifically addresses *B. pertussis* growth inhibition and lower pertactin

(antigen) productivity caused by ROS (inhibitor) and the role of NADPH (byproduct) in quenching ROS (Singh et al. [2007]). The model is adapted from an earlier work in our group by (Vitelli et al. [2023c]) that was applied to batch operation in flasks but it only predicted biomass growth and ROS. The current proposed model considers fed-batch operation in bioreactors and it also predicts pertactin productivity, along with ROS, NADPH and glutamate concentration. Hence, in contrast with the earlier study, the model proposed here correctly mimics the production environment in terms of sequential batch and fed-batch operations and it also addresses pertactin (antigen) production which is the key target of this manufacturing process. Experiments were conducted in 2-liter fed-batch bioreactors. NADPH and riboflavin (antioxidant) levels in culture supernatant were monitored using an in-house developed fluorometer (in-line, dual excitation) which was also used for real time monitoring of biomass, glutamate, ROS and pertactin concentrations by combining it with metabolic model. Offline measurements of biomass and NADPH were done using a spectro-photometer and fluorescence spectroscopy respectively. ROS and surface expression of pertactin were assessed using flow cytometry and secreted pertactin was quantified using affinity chromatography. Fed-batch fermentations of *B. pertussis* were started with varying concentrations of initial glutamate to observe its impact on ROS levels. The culture behaviour for varying oxidative stress conditions caused by different initial nutrient levels is of particular industrial interest since variability of nutrients in media has been reported for the production scale operation.

In summary, two different applications of the proposed mechanistic model are presented:

- The model is calibrated to off-line measurements with a subset of the available data and it is validated with the remaining data that was not used for validation.
- Spectra acquired with in-line fluorometer combined with the mechanistic model is used for developing an real-time estimator (soft sensor) of growth, oxidative stress level, glutamate, and antigen productivity.

The availability of an on-line sensor is expected to lead to other future applications as follows:

- Real-time diagnosis of bad batches in the industrial process,
- Potential implementation of mid-points corrective actions during the fermentation.
- Optimizing cell growth through model-based strategies
- Identifying potential sources of ROS by integrating additional metabolic pathways known to cause oxidative stress. On-line estimation of ROS levels via the model to predict low/high productivity earlier on during the batch.

6.2 Metabolic Model Development

The original model of Himeoka and Kaneko's was used to explain various phases of a cell population - lag, exponential growth, stationary, and death - without delving into detailed metabolic mechanisms that correlate with these events. Hence, the original model was not used to explain oxidative stress mechanisms. The model was applied by (Viteli et al.

[2023c]) to explain oxidative stress events based on observed experimental correlations between oxidative stress with the durations of different phases of growth, i.e. lag, death etc. Vitelli's model involves the interaction of two key components which are formed when the cell consumes glutamate: component A, which promotes cell growth auto-catalytically, and component B, which inhibits this process by forming a complex (component C) with component A. This interaction is illustrated schematically in Figure (6.1). Here, component A is NADPH and component B is ROS. Component A enhances cell growth through an auto-catalytic mechanism, and component B hinders this auto-catalytic process by interacting with component A to form a complex C (NADP^+). This process is illustrated schematically in Figure (6.1). The assignment of species in the model to particular compounds such as glutamate, ROS and NADPH is based on prior knowledge about the cell metabolism as reviewed in the Introduction. For example, glutamate is metabolized in the TCA cycle to produce ROS through generation of NADH and FADH and NADPH through the isocitrate to oxoglutarate reaction. The contributions of NADPH to growth are related to synthesis of fatty acids and nucleotides (Chen et al. [2019]) that are both contributors to cell growth.

NADPH, or Nicotinamide Adenine Dinucleotide Phosphate Hydrogen, is an important cofactor in various cellular processes, particularly in anabolic reactions like lipid and nucleic acid synthesis, where it functions as a reducing agent by donating electrons. This electron-donating ability is also key to its role in protecting cells from oxidative stress. One of the main mechanisms by which NADPH quenches oxidative stress is by maintaining levels of reduced glutathione (GSH), a major cellular antioxidant. The enzyme glutathione reductase depends on NADPH to convert oxidized glutathione (GSSG) back to its active

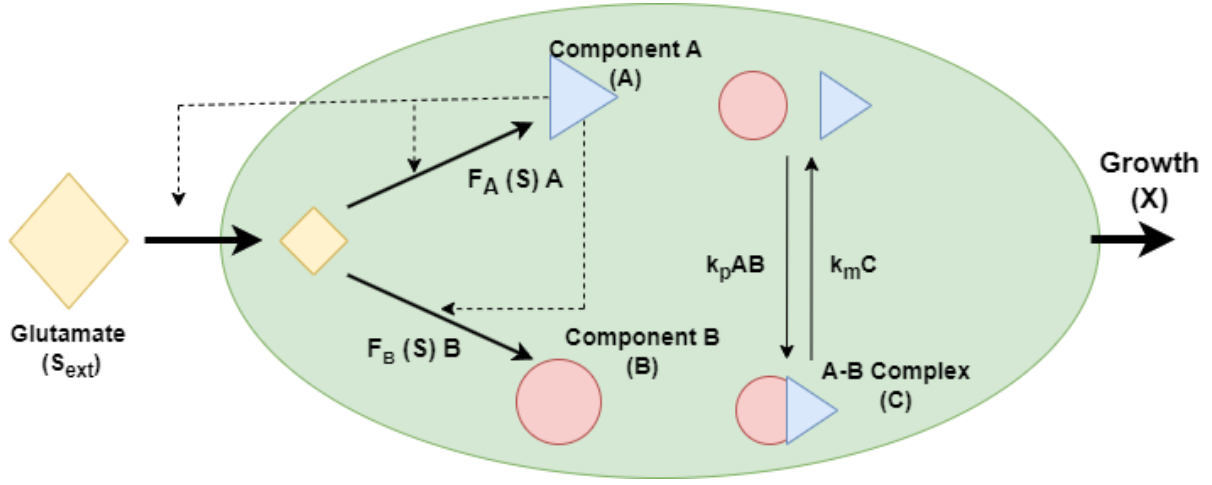


Figure 6.1: Schematic representation of the mechanism described in the model.

form, GSH, which then neutralizes harmful reactive oxygen species (ROS) that can cause cellular damage. Beyond its role in glutathione maintenance, NADPH directly reduces ROS through various enzymatic reactions. Additionally, NADPH is crucial for the thioredoxin system, another antioxidant pathway that helps maintain a reduced environment in cells thus ensuring that proteins function properly under oxidative stress (Cabisco et al. [2000a]).

$$\frac{dS_{\text{ext}}}{dt} = -\frac{\nu \cdot X}{Y_{xs}} + \frac{F}{V} \cdot (S_{\text{ext, inlet}} - S_{\text{ext}}) \quad (6.1)$$

$$\frac{dS}{dt} = -F_A \cdot A - F_B \cdot A + A \cdot (S_{\text{ext}} - S) - \nu \cdot S \quad (6.2)$$

$$\frac{dA}{dt} = F_A \cdot A - k_p \cdot A \cdot B + k_m \cdot C - \nu \cdot A - K_m \cdot A \quad (6.3)$$

$$\frac{dB}{dt} = F_B \cdot A - k_p \cdot A \cdot B + k_m \cdot C - \nu \cdot B - d_b \cdot B \quad (6.4)$$

$$\frac{dC}{dt} = k_p \cdot A \cdot B - k_m \cdot C - \nu \cdot C \quad (6.5)$$

$$\frac{dX}{dt} = \nu \cdot X - D \cdot X - \frac{F}{V} \cdot X \quad (6.6)$$

$$\frac{dA_{\text{ext}}}{dt} = (K_m + D) \cdot A \cdot X \quad (6.7)$$

$$\frac{dPRN}{dt} = \frac{p \cdot \left(\frac{dS_{\text{ext}}}{dt} - \frac{F}{V} \cdot (S_{\text{ext, inlet}} - S_{\text{ext}}) \right)}{\overline{X} \cdot (Y_p + B)^2} \quad (6.8)$$

$$F_A = \left(\frac{\mu}{1 + \frac{S}{K_{si}}} \right) \cdot \left(\frac{S}{K \cdot X + S} \right) \cdot \left(\frac{S}{K_t + S} \right)^{1.5} \quad (6.9)$$

$$F_B = \left(\frac{\mu}{1 + \frac{S}{K_{si}}} \right) \cdot \left(\frac{S}{K \cdot X + S} \right) \cdot \left(\frac{K_t}{K_t + S} \right)^{1.5} \quad (6.10)$$

$$\nu = F_A \cdot A \quad (6.11)$$

$$V = V_0 + F \cdot (t - 28) \quad (6.12)$$

The model described by Equations (6.1 - 6.8) is derived from transient mass balances for various species and the rate expressions for the reactions depicted schematically in Figure (6.1). The symbols used to represent the concentrations of different species are explained in Table (6.1), while Table (6.2) defines the model parameters. Following Figure (6.1), glutamate (substrate S) is involved in the production of both NADPH (denoted as A) and ROS (denoted as B) through the terms $F_A A$ (which equals the growth rate μ) and $F_B A$, respectively. In this model, these terms introduce auto-catalytic effects on A (NADPH), which in turn promote growth and increase substrate consumption, thereby enhancing the growth rate μ . The terms F_A and F_B are both dependent on the intracellular concentration of the substrate S . Additionally, the term $k_p AB$ represents the co-metabolism of NADPH and ROS which is related to the antioxidant reactions' mechanisms described above. Based on experimental results, the following modifications were introduced by Vitelli to the original model proposed by Himeoka and Kaneko: a substrate inhibition term and exponents in the expressions for F_A and F_B (as seen in Equations (6.9) and (6.10)), a degradation term for ROS (equation (6.4)), and a term accounting for the mass transport of NADPH into the supernatant (Equations (6.3) and (6.7)). The ROS degradation term is introduced because ROS are highly reactive and will interact with other compounds if not neutralized by an antioxidant. The mass transfer step has been added to represent the secretion of NADPH through the cell membrane into the supernatant.

The model of Vitelli was developed for flask cultures of pertussis without any feed or closed loop control. In this work, the model has been further modified for its application to a fed-batch bioreactor operation as follows:

- Volume changes due to feeding and dilution factors are introduced in equation (6.1) and (6.6).
- Substrate feeding to the bioreactor was started at 28 hours and was maintained until the end of the run (6.12).
- Product concentration (pertactin) is predicted by including a differential equation to predict the concentration of secreted pertactin (6.8). This function of the form $f(ROS, glutamate, Biomass)$ in equation (6.8) is the specific secretion rate of pertactin into the supernatant.

The dependency of (f) with respect to ROS, Biomass and glutamate is further discussed and justified in the results section based on experimental observations.

Furthermore, the model of Vitelli had to be re-calibrated since the experiments in the current study were performed in bioreactors with closed loop control of dissolved oxygen, pH and temperature while Vitelli performed all his experiments in flasks. In closed-loop control, parameters like dissolved oxygen (DO) and pH are tightly regulated, which differs significantly from flask runs where these parameters are not controlled. DO levels are consistently monitored and adjusted through controlled aeration and agitation, maintaining a stable oxygen supply. Similarly, pH in bioreactors is controlled through the automatic addition of acid or base, ensuring an optimal and stable environment for cell growth and

metabolism. lack of closed loop control in flasks can lead to variable DO and pH levels. Fluctuating pH in flasks can disrupt cellular metabolism and impair antioxidant defenses, resulting in reactive oxygen species (ROS) production (Maurer et al. [2005]). On the other hand, closed loop DO control is not always beneficial and can exacerbate stress as explained in the previous chapter.

Table 6.1: Definition of variables used in the model.

Model Variable	Definition
Sext	Extracellular glutamate concentration (mmol/L)
S	Intracellular glutamate concentration (mmol/L)
A	Intracellular NADPH concentration (mmol/L)
B	ROS concentration (mmol/L)
C	NADP+ concentration (mmol/L)
X	Biomass intensity (OD)
Aext	Extracellular NADPH concentration (mmol/L)
PRN	Pertactin concentration (mmol/L)

6.3 Materials and Methods

6.3.1 *B. Pertussis* Fed-Batch Fermentations

Fed-batch fermentation of *B. pertussis* culture were performed in 3-Liter bioreactors (Biostat B plus, Sartorius) with a working volume of 1.6 - 2 Liters. Inoculum was prepared in 500 ml shake flask with a working volume of 120 ml and inoculation was done after the flask age of 44 hours. A total of 5 fed-batch fermentations, and 4 batch fermentations were

Table 6.2: Parameters defined in the model.

Symbol	Meaning	Units
μ	Rate constant for biomass growth	h^{-1}
$Y_{x/s}$	Yield coefficient	intensity biomass / g glutamate
D	Rate constant for biomass death	intensity biomass / h^{-1}
F_A	Flux of substrate conversion into A	h^{-1}
F_B	Flux of substrate conversion into B	h^{-1}
k_p	Rate constant for $A + B \rightarrow C$	$mM^{-1}h^{-1}$
k_m	Rate constant for $C \rightarrow A + B$	h^{-1}
d	Rate of degradation of B (ROS)	h^{-1}
v	Rate constant for the generation of NADPH	$mM \cdot h^{-1}$
K_{si}	Rate constant for inhibition	h^{-1}
K	Saturation constant for F_A and F_B	mM
K_t	Phenomenological constant changing A/B balance	mM
K_m	Transport function for A across the cell membrane into the supernatant	intensity biomass / h^{-1}
α	Unit conversion constant (=1)	mM^{-1}
p	Yield coefficient for PRN	intensity biomass / mmol PRN
Y_p	Inhibition constant	mmol/L

conducted which includes batches with cP media, batches with 20 % higher initial glutamate concentration and their repeats. The cultures were maintained at 36°C and a pH of 7.2, with the dissolved oxygen level controlled at 35 % by maintaining a constant air flow of 2 lpm. Agitation rates were varied between 200 and 700 rpm using a Rushton impeller. Antifoam was manually added to the bioreactors as needed for each batch. Supplement feed with a rate of 20 ml/hours was started at the end of 28 hours when all the nutrients in the batch run was consumed. Table (6.3) provides a summary of each run, highlighting the batch-to-batch variations in terms of initial glutamate feed, maximum optical density (OD), and productivity. Total time for batch fermentation was 32 hours, for fed-batch runs it was 48 hours and samples were taken every 6 hour for analysis. For intellectual property reasons, the values of OD, glutamate and antibody concentration have been normalized between 1 and 2. Another reason for normalizing data between 1 to 2 was avoid values close to zero as encountering a value of 0 can lead to significant complications. Specifically,

when a variable normalized to zero is used in expressions involving division, logarithmic transformations, or root extractions, it can generate undefined or complex values.

Table 6.3: Summary of batch and fed-batch runs.

Batch ID	Initial glutamate concentration (Normalized)	Batch age (Hrs)	OD at 32 hrs (Normalized)	Cumulative PRN yield (Normalized)
A	1.064	48	1.36	2
B	2.0	48	1.08	1.28
C	1.02	48	1.48	1.78
D	1.97	48	1.0	1.0
E	1.08	48	1.76	1.67
F	1.09	32	1.88	1.88
G	1.95	32	1.70	1.39
H	1.0	32	2.0	1.92
I	1.965	32	1.77	1.47

6.3.2 Biomass and Glutamate Quantification

Biomass samples were taken from the bioreactor at intervals ranging from 4 to 8 hours to measure optical density (OD), and glutamate concentration. Optical density was measured using a Thermo Scientific GENESYS 40/50 Vis/UV-Vis spectrophotometer (2020). Concentration of extracellular glutamate was measured using reverse phase HPLC method. HPLC was performed using a Waters 1525 chromatography system (Waters Corporation, MA, USA) equipped with a fluorescence detector (W2475; Waters Corporation, MA, USA). The method employed a Waters AccQ.Tag™ Amino Acid Analysis Column C18 (4 μ m, 3.9 mm $\times$ 150 mm), maintained at 38°C. For each analysis, a 5 μ L sample was injected. The fluorescence detector was set to an excitation wavelength of 248 nm and an emission wavelength of 395 nm. The mobile phase consisted of Eluent A, an acetate-phosphate buffer

solution prepared by mixing 100 mL AccQ.Tag[™] concentrate (Waters, MA, USA) with 1 liter of Millipore water. Eluent B was HPLC-grade acetonitrile, and Eluent C was Millipore water. Each HPLC run lasted for 40 minutes. The column was equilibrated for 10 minutes before each run and between sample runs, using a mixture of 99 % Eluent A and 1 % Eluent B at a flow rate of 1 mL/min.

6.3.3 In-Line Fluorescence Acquisition

A multi-wavelength (dual excitation, full emission) fluorometer was used for data acquisition. The fluorometer was equipped with 365 nm and 405 nm LED lights, a CCD array spectrometer from OceanOptics Inc. capable of capturing emission spectra within the 350-1100 nm range, and a sterilizable fiber-optic reflection probe with one excitation fiber and six read fibers. The excitation lights are replaceable, allowing for the optimal selection of excitation wavelengths. The fiber-optic probe was inserted into the bioreactors through a standard 9 mm top-mounted bioreactor port. Real-time process monitoring was achieved by acquiring spectra at 1-minute intervals. Emission spectra were captured within the 300-850 nm range. The optimal instrument settings identified for this study were a voltage of 4 V, an integration time of 100 ms, and a scanning rate of 1,000 nm/min. The fluorometer was controlled by a computer, which also logged the acquired emission spectra. To minimize noise, an average of three emission scans was taken at each time point. Each data acquisition was preceded by the recording of a background spectrum, which was later subtracted from the emission scans for correction.

6.3.4 Fluorescence Spectroscopy

Offline fluorescence spectroscopy was performed at an excitation of 340-360 nm for the monitoring of NADPH levels in culture supernatant. The detailed procedure is described in section (3.8.2) of chapter (3).

6.3.5 Flow Cytometry

Flow cytometry was used for assessing the ROS and levels of surface expression of pertactin among bacterial cell populations. A cell sorter (S3e cell sorter, Bio-Rad laboratories, CA) equipped with two lasers (488 nm, 640 nm) was used. A detailed description of the methods used for flow cytometry measurements is provided in Chapter 3 section (3.9) of this thesis. Data collected using the flow cytometer is then used to calculate the weighted average (WA) of ROS and surface pertactin fluorescence intensity, using the following formula:

$$WA = \sum_{i=1}^n \frac{C_i}{C_t} * I_i \quad (6.13)$$

Where C_i is the number of cells that have fluorescence intensity I_i , and C_t is the total number of cells analyzed.

6.3.6 Affinity Chromatography for Pertactin Quantification

Affinity chromatography was used for secreted pertactin (antigen) quantification. A detailed method of column preparation and procedure for protein purification is given in

chapter (3) section (3.6.2).

6.3.7 Data Pre-processing and PLS Regression

Partial least square regression method was used as a chemometric tool for estimating the state variables. Fluorescence maps collected using the in-line fluorometer were used to calculate the area under the peak before their use for regression analysis. Table (6.4) shows the emission ranges used for calculating the areas and the fluorophores which produce fluorescence peaks in that range. Further, the fluorescence areas for each batch were normalized to the areas at time $t = 0$ hours, which were collected immediately after inoculation. This normalization was done for each batch to eliminate batch-to-batch variability that could arise from different geometric positions of system components, such as the placement of the fluorometer probe in the bioreactor. MATLAB (release R2023b) was used for all chemometric calculations. A Partial Least Squares (PLS) regression model was utilized to establish correlations between the fluorescence areas (spectral integrals) along with fermentation parameters such as agitation, dissolved oxygen, and volume, with the target variables—namely, biomass intensity (OD), glutamate, Extracellular NADPH, weighted average of ROS and pertactin (antigen) (see chemometric analysis presented in Chapter 4 of this thesis).

Fermentation variables can be directly predicted with a Partial Least Squares (PLS) regression model based on fluorescence spectra. However, the high dimensionality and inherent collinearity of fluorescence spectra data pose numerical challenges in model de-

velopment. To address these challenges, proper data pre-treatment is essential. Due to colinearity the regressor matrix is almost singular and thus, when performing regression between the target process variables and the spectra, the resulting regression parameters are highly inaccurate. Hence, after obtaining the fluorescence maps, compression of the multivariate data into a lower-dimensional space is required to mitigate the effect of high correlation in the data. In such cases, multivariate chemometric techniques can be used to correlate the spectra data with target state variables since they reduce the impact of the colinearity of the regression matrix as compared to conventional multivariate least squares regression.

Table 6.4: Excitation wavelength and emission ranges used for area calculation using fluorescence spectra.

Notation	Excitation (nm)	Emission (nm)	Comments
A1	365	410-425	-
A2	365	430-470	Presumed NADPH peak
A3	365	475-560	Presumed Riboflavin peak
A4	365	590-650	Porphyrins
A5	365	700-800	Light scattering (secondary)
A6	405	450-480	Presumed Riboflavin peak
A7	405	490-560	Riboflavin peak
A8	405	565-620	Porphyrins
A9	405	635-700	Light scattering (secondary)
A10	405	790-850	Light scattering (secondary)

6.3.8 Model Fitting

The mechanistic model described in section (6.2) was calibrated using genetic algorithm (ga) function in MATLAB. This function is a part of MATLAB's Global Optimization Toolbox, and it provides a flexible and customizable environment for solving a wide range of optimization problems. The function works by evolving a population of potential solutions through a series of generations. At each generation, the algorithm evaluates the fitness of each individual, selects the best-performing individuals, and applies genetic operators such as crossover, mutation, and selection to produce a new population. The advantage of using genetic algorithm lies in its ability to search large, complex spaces and find near-optimal solutions in cases where conventional optimization techniques might fail since they lead to local optima. The function also allows tailoring the algorithm to a specific problem by adjusting parameters such as population size, selection criteria, crossover and mutation rates, and stopping conditions. To fit the model, the measured variables (Biomass (X), extracellular glutamate (Sext), extracellular NADPH (Aext), ROS (B), and pertactin (PRN)) were first normalized to avoid biasing the model with respect to any particular variable. The objective function to be minimized for model fitting was the sum of squared errors of the normalized experimental values. For this purpose, offline data collected for the 7 fermentations (which includes batch and fed-batch runs for both cP media and 20 % higher glutamate) with 58 data-points were used for calibration. Table (6.2) defines the parameters used in model fitting. Figure (6.2) shows the calibrated model predictions for different state variables along with offline measurements for normal runs (with cP media). After calibration, the model was validated with batches that were not used for calibration.

After calibration of the mechanistic model using the genetic algorithm, the prediction

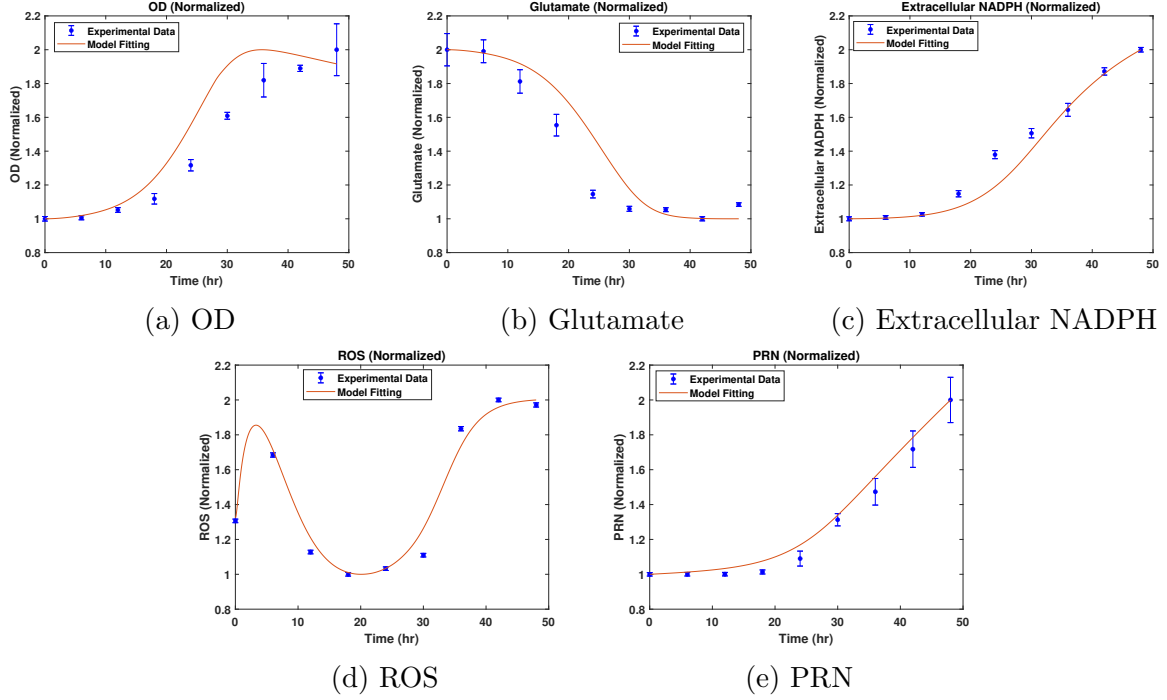


Figure 6.2: Model fitting results for (6.2a) optical density (OD), (6.2b) extracellular glutamate, (6.2c) extracellular NADPH, (6.2d) ROS and (6.2e) PRN, along with offline measurements.

accuracy was compared with a purely empirical PLS model. One of the standard methods for model comparison in statistical modeling is the Akaike Information Criterion (AIC). The AIC provides a measure of the relative quality of the model by balancing the goodness of fit with model complexity. These criteria assess whether a fitted model maintains an optimal balance between its predictive accuracy and the number of fitting parameters. The

formula for AIC is given by:

$$\text{AIC} = -n \log \left(\frac{\text{RSS}}{n} \right) + 2k + \frac{2k(k+1)}{n-k-1} \quad (6.14)$$

where:

- n is the number of data points,
- RSS is the residual sum of squares,
- k is the number of parameters in the model.

6.4 Results and Discussion

6.4.1 Cell growth, Oxidative Stress and Productivity

Fed-batch fermentations with two different concentrations of glutamate were performed to study their effect on growth and pertactin productivity. Figure (6.4) shows the growth profiles, glutamate consumption, extracellular NADPH concentration, weighted average ROS levels and pertactin productivity for the corresponding batches. Cultures with 20 % higher initial glutamate levels exhibited slower growth compared to those in the control run with defined cP media. As discussed in chapter (5), this reduced growth rate in the lag phase could be attributed to a combination of an inhibitory mechanism of glutamate consumption and slower growth due to higher oxidative stress as shown in figure (6.4e). A longer lag phase could be an additional explanation for the observations at the

beginning of the batch. ROS levels among cell cultures decrease as the bacterial growth reaches exponential phase and increases again towards the end of fermentation when the cell growth begins to stall, possibly due to slow consumption rates of glutamate when the latter reaches low concentrations. It should also be noted that in cultures with higher glutamate levels, the initial WA for ROS was greater than that in the culture with cP media. This suggests that oxidative stress during the lag phase is more pronounced in cultures with elevated glutamate concentrations. Additionally, NADPH level in the fermentation broth (shown in figure (6.4c)) was higher in cultures with high initial glutamate concentration, when compared to the control batch. It should be noticed that NADPH is measured in the supernatant (Zavatti et al. [2016]). This accumulation of NADPH in the supernatant is related to growth inhibition since biomass consumes NADPH. When growth is inhibited, NADPH is not consumed thus reaching higher intracellular levels and promoting higher secretion into the supernatant. Pertactin productivity in cultures with high initial glutamate concentrations was also lower than that in the control by the end of the initial lag phase. This is primarily because of the lower growth as growth is highly correlated to pertactin secretion which we demonstrated in the previous chapter. Also, it has been previously found that higher levels of NADPH were accompanied by lower pertactin productivity (Lushchak [2001b]).

In the previous chapter, we demonstrated that high initial concentrations of glutamate can induce oxidative stress, leading to slower growth in bacterial cultures. It was also observed that cultures with slower growth rates consumed less glutamate, resulting in lower oxygen uptake. Additionally, we found that oxidative stress, correlated with elevated ini-

tial glutamate levels, was inversely related to both growth and productivity. Furthermore, controlling dissolved oxygen levels in bioreactors was shown to impact productivity compared to flask operations, where dissolved oxygen remains uncontrolled. The results of the fed-batch runs follow the previous trend and reinforced our earlier observations, showing a strong correlation between growth, productivity and oxidative stress. Based on these experimental observations, the secreted pertactin concentration was hypothesized to be a function of biomass X , ROS B and glutamate concentration S_{ext} as follows:

$$f(X, B, \text{Glu}) = \frac{k_1 \frac{d\text{Glu}}{dt} \cdot X}{k_2 + \text{ROS}} \quad (6.15)$$

To validate this relation, the ratio of glutamate consumption rate and ROS per log mean biomass was plotted against pertactin secretion rate per log mean OD as shown in figure (6.3). A correlation was found ($R^2 = 0.6001$). It was observed that for a higher ratio of glutamate consumption rate to ROS per log mean OD, the pertactin secretion rate is also higher. The dependency of the secretion rate with glutamate consumption rate was hypothesized based on the fact that glutamate is the main nutrient and its faster consumption rate will be correlated to higher synthesis of protein. Then, faster protein synthesis will lead to higher secretion rate. Based on these observations, equation (6.8) was used in the mechanistic model for describing pertactin secretion.

6.4.2 PLS Regression Model for State Estimation

A Partial Least Squares (PLS) regression model was utilized to establish a relationship between the fluorescence spectra integrals (spectra areas) collected in-line, along with fer-

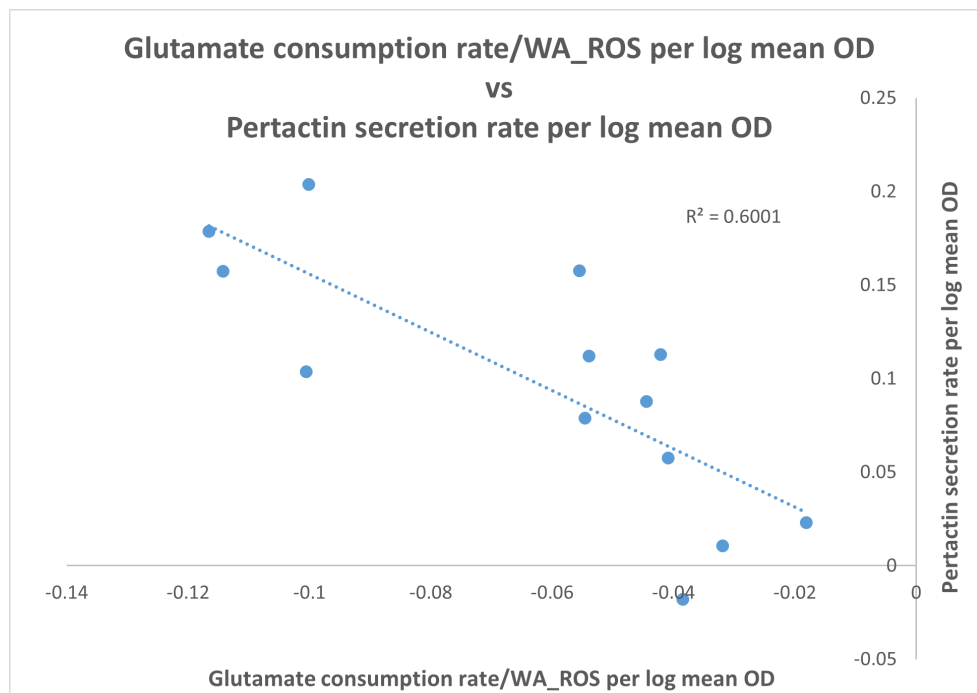


Figure 6.3: Correlation between glutamate consumption rate to ROS ratio and pertactin secretion rate per log mean OD. The trend shows that secretion of pertactin is positively correlated to glutamate consumption rate and inversely correlated to ROS.

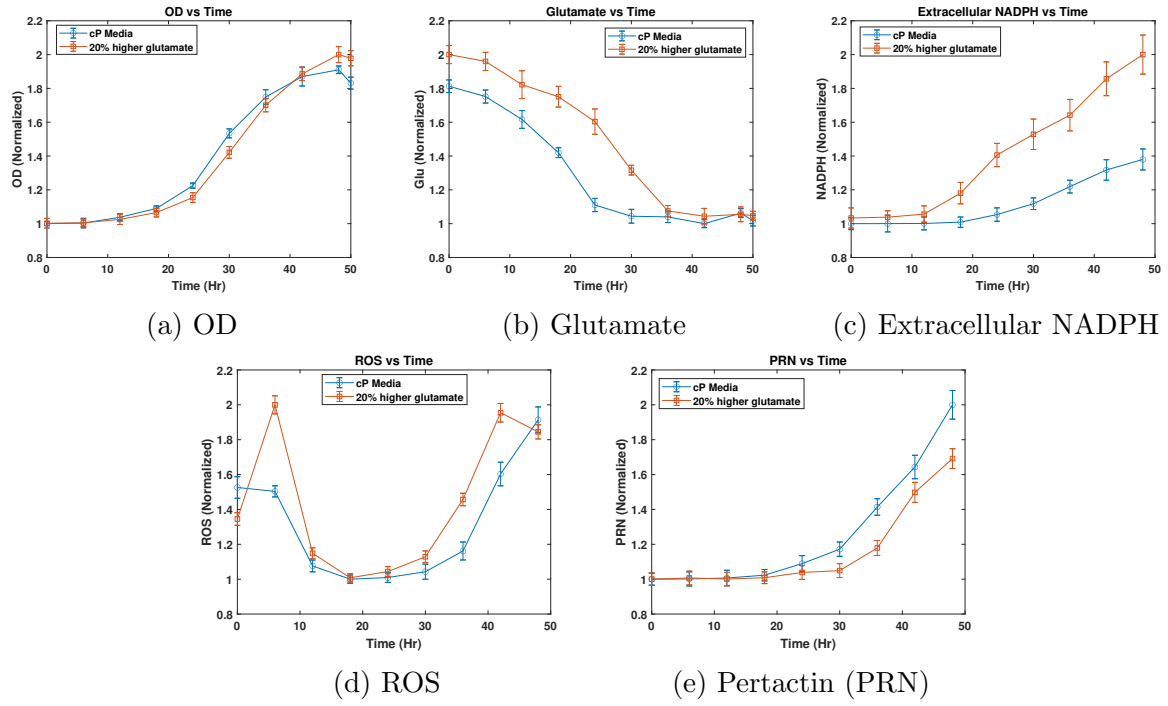


Figure 6.4: Experimental data for the state variables. Blue dot and line shows the control runs and orange line shows runs with 20 % higher glutamate.

mentation parameters such as agitation, dissolved oxygen, and volume, with the target variables, which include biomass, extracellular glutamate, weighted average of ROS, and cumulative pertactin (antigen) concentration. Table (6.5) summarizes the performance of the PLS regression model. The model demonstrates strong performance across most of the target variables, with particularly high accuracy in predicting optical density (OD) and extracellular glutamate. The R^2 values for calibration and validation for OD (0.9953 and 0.9652) and for glutamate (0.9814 and 0.9014) indicate that the PLS model explains a substantial portion of the variance, while the low RMSE values confirm its precision. The model also performs well for extracellular NADPH and ROS, with R^2 values over 0.80 for both calibration and prediction, though the predictive power slightly decreases as indicated by a higher RMSE-p values. In contrast, the model's performance for extracellular pertactin (PRN) and ROS is lower, with an R^2 -p of 0.8187 and a higher RMSE-p of 0.1726 for PRN and 0.8474 and 0.1652 for ROS respectively. One of the reasons of highly accurate biomass and glutamate predictions is due to the secondary scattering of light in the emission range of 600 - 800 nm as presented in table (6.4). Scattering of light is directly related to the biomass concentration and as a result, higher the biomass concentration, higher is the peak area for secondary scattering. Also, biomass growth is directly correlated to glutamate consumption as higher the biomass growth, more glutamate is consumed by the bacterial cells leading to a high correlation between fluorescence areas and biomass and glutamate concentrations. Fluorescence areas in the range of 425 - 450 nm corresponds to NADPH and area in the range of 475 - 550 nm emission range corresponds to riboflavin. A significant correlation is therefore observed for NADPH and ROS.

Table 6.5: Performance of the PLS regression model for different state variables.

Variable	R^2 -c	R^2 -p	RMSE-c	RMSE-p
Glutamate (Extracellular)	0.9814	0.9014	0.0863	0.1617
OD	0.9953	0.9652	0.0356	0.0842
ROS	0.9537	0.8474	0.0984	0.1652
NADPH (Extracellular)	0.9646	0.8165	0.0512	0.1901
PRN (Extracellular)	0.9354	0.8187	0.0965	0.1726

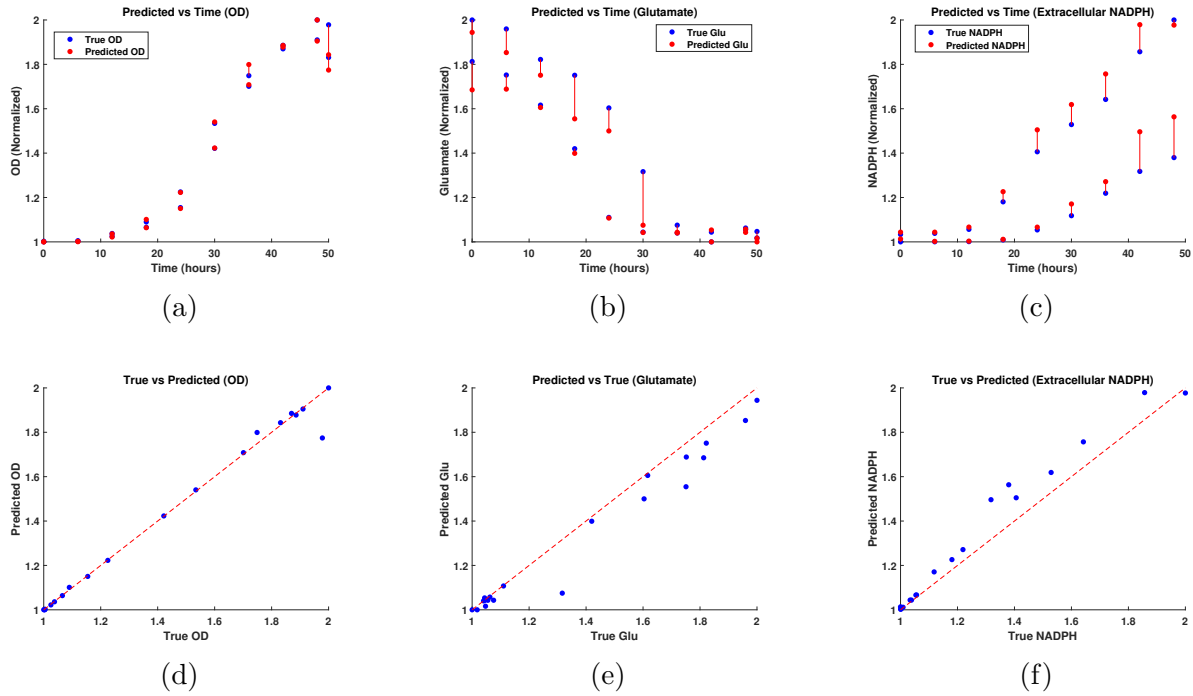
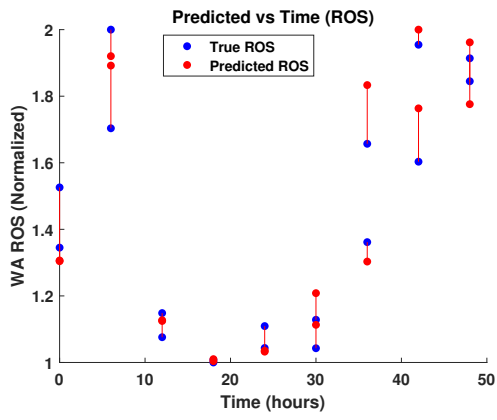
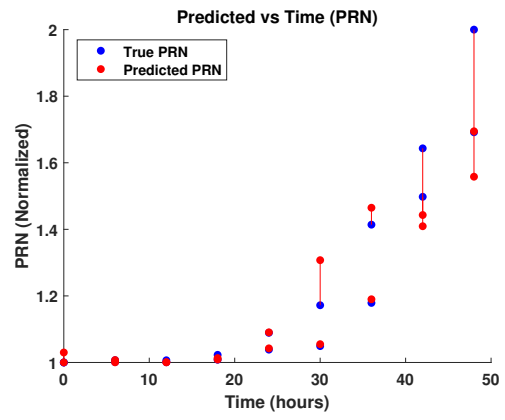


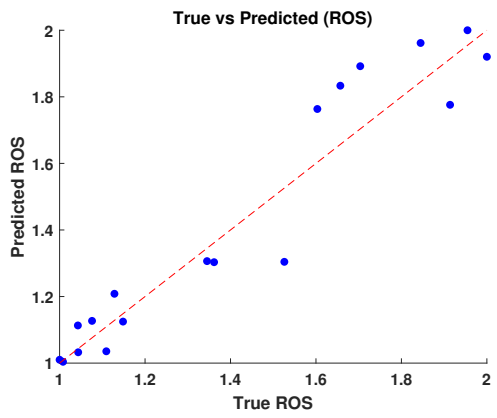
Figure 6.5: Performance of PLS regression model for different state variables. Figures (a), (b) and (c) shows the predicted vs time plots for OD, glutamate and extracellular NADPH respectively. Figures (d), (e) and (f) show the plots of true vs predicted values for the corresponding variables.



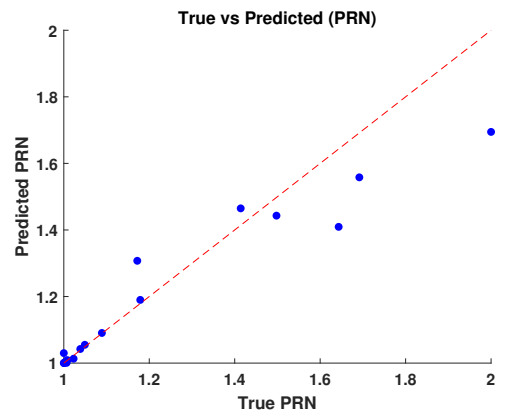
(a)



(b)



(c)



(d)

Figure 6.6: Performance of PLS regression model for different state variables. Figures (a) and (b) shows the predicted vs time plots for ROS and PRN and figure (c) and (d) shows predicted vs true values for the corresponding variables.

6.4.3 Mechanistic Model Validation

The oxidative stress model was calibrated using a genetic algorithm on experimental data by fitting it to experimental data collected from batch and fed batch runs with control and 20 % higher glutamate runs and was validated using data from one batch and one fed-batch run for both the conditions. The system of equations (Equations (6.1)–(6.12)) that make up the model is detailed in section (6.2), and the definition of each parameter determined by fitting the model is provided in Table (6.6). A comparison of the experimental data and the model calibration curves for optical density (OD), extracellular glutamate concentration, extracellular NADPH levels, pertactin concentration and intracellular ROS concentrations is presented in Figures (6.7) and (6.8). These figures illustrate the results for both the control conditions and 20 % higher glutamate condition. Both the conditions (of differing glutamate conditions) were accurately described by the calibrated model for OD, glutamate, NADPH, ROS and PRN. Table (6.7) shows the RMSE values for calibration and validation of mechanistic model. For OD, glutamate and PRN, the RMSE-validation values are less than 10 % of the total range of these variables while for ROS and NADPH, they are 18.8 % and 14.5 % of their total range of values.

6.4.4 Extended Kalman Filter

Extended kalman filtering (EKF) was utilized to improve the prediction accuracy as compared to either the pure mechanistic model or the purely empirical PLS model. Figure (6.10) demonstrate the performance of extended kalman filter (EKF) compared to the mechanistic model predictions and PLS models as compared to offline measurements. For

Table 6.6: Calibrated parameters of the mechanistic model.

Parameter	Calibrated Value
μ	2.68 mM/h
K	2.781 mM
K_t	1.354 mM
k_p	0.2137 1/mM.h
k_m	0.0001 1/h
Y_{xs}	0.657 g biomass/g glutamate
D	0.0010 1/g biomass.h
K_{si}	1.9528 mM
dB	0.0449 1/h
K_m	0.00001251 1/g biomass.h
p	3.9901 L/mmol ²
Y_p	380.6309 mmol/L

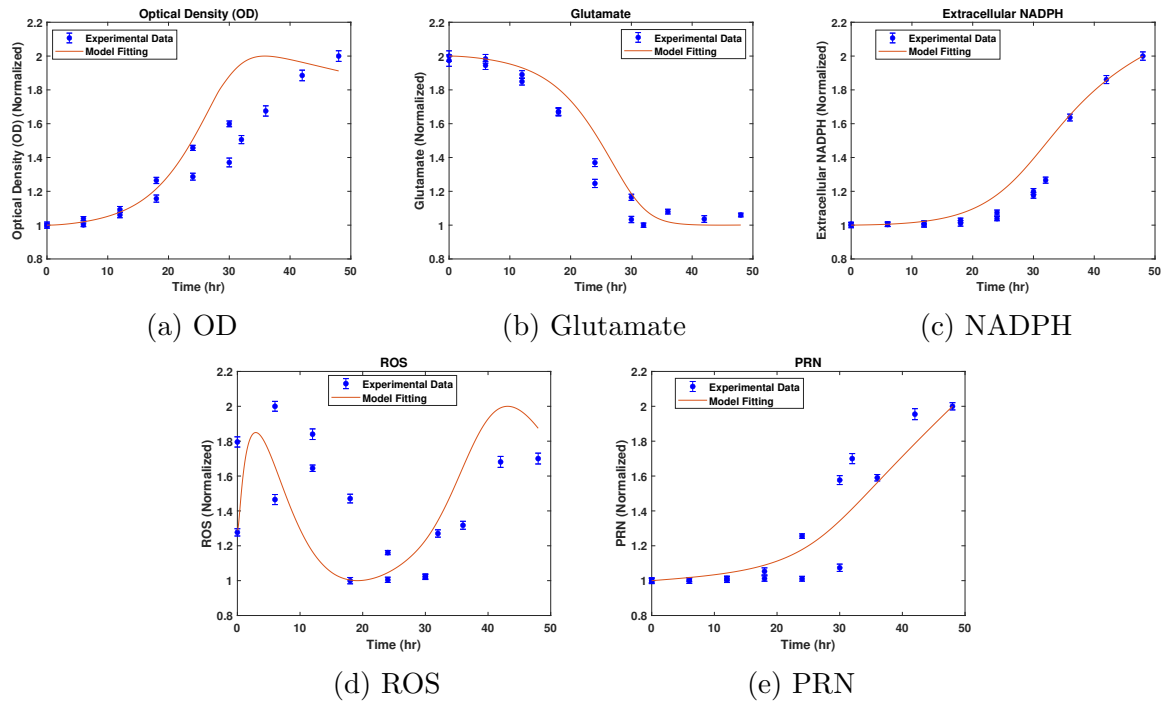


Figure 6.7: Mechanistic model validation for OD, glutamate, NADPH, ROS and PRN for the initial glutamate concentration of 10.3 g/L and initial OD of 0.3 (approx.). Experimental data from a batch and a fed-batch run which was not used in the calibration is shown by blue dots.

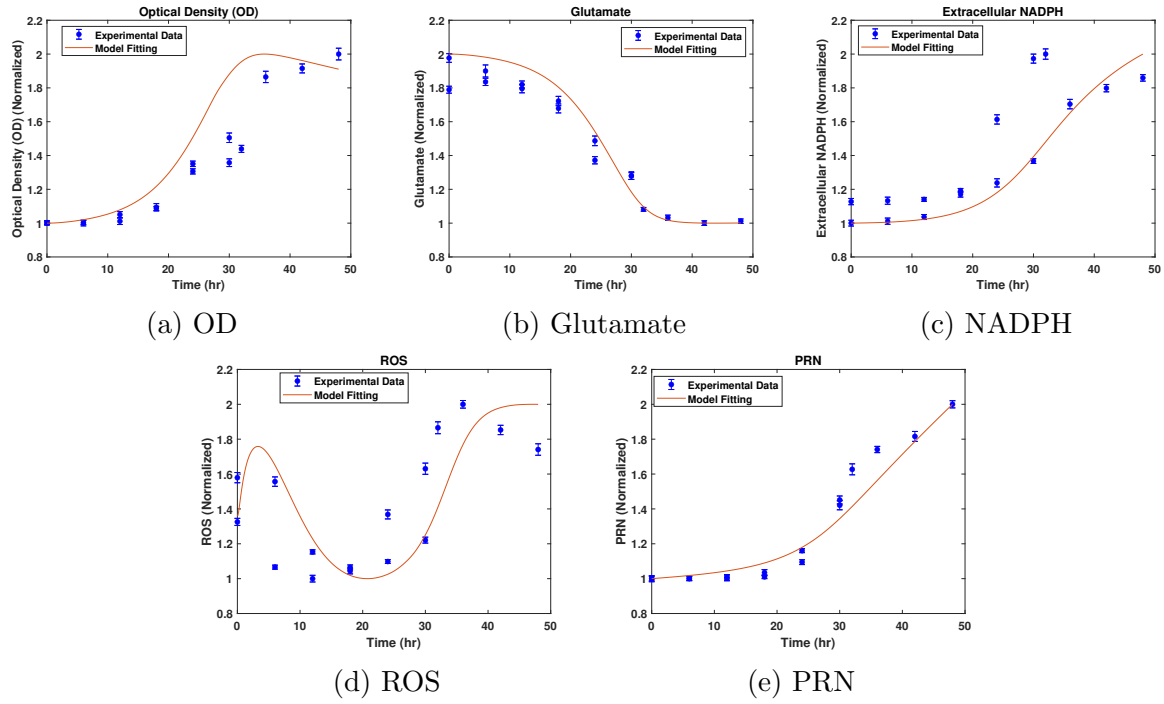


Figure 6.8: : Mechanistic model validation for OD, glutamate, NADPH, ROS and PRN for the initial glutamate concentration of 12.7 g/L and initial OD of 0.3 (approx.). Experimental data from a batch and a fed-batch run which was not used in the calibration is shown by blue dots.

biomass growth, extracellular glutamate, and ROS, the EKF estimates closely agree with experimental data, with significant improvements as compared to the mechanistic model and the PLS model predictions. In the case of ROS, the EKF effectively captures the dynamic fluctuations observed experimentally, offering a more accurate prediction than the other methods. Similarly, for glutamate prediction, the EKF estimate shows a marked improvement, closely tracking the experimental results and further validating the EKF's ability to improve model predictions across all examined states' variables. It can be noticed that NADPH secretion is higher towards the end of the run for all models. A possible explanation can be that this is due to NADPH contained in the cells which is released into the supernatant due to cell death. This effect was not included into the model at the current stage to avoid over-parameterization. For pertactin, the EKF prediction is similar to PLS regression model prediction. Figures (6.9 and 6.10) show the validation results of EKF model for two different initial glutamate concentration. A summary of model performance in terms of prediction accuracy is presented in table (6.7). The EKF achieved an RMSE-v of 0.0667 for OD compared to 0.1025 for the Mechanistic Model, and 0.0744 for Glutamate versus 0.0821. For Pertactin (PRN), hybrid (EKF) model demonstrated better performance with an RMSE of 0.0620 as compared to the mechanistic model's 0.0831. It can be noticed that by manipulating the process noise initial values it is possible to improve accuracy in some variable at the cost of deterioration of accuracy in other variables. Following this rationale, we attempted to tune the process noise to improve the accuracy of the pertactin predictions but this resulted in significant deterioration in prediction of the other variables and thus, this approach was not pursued further.

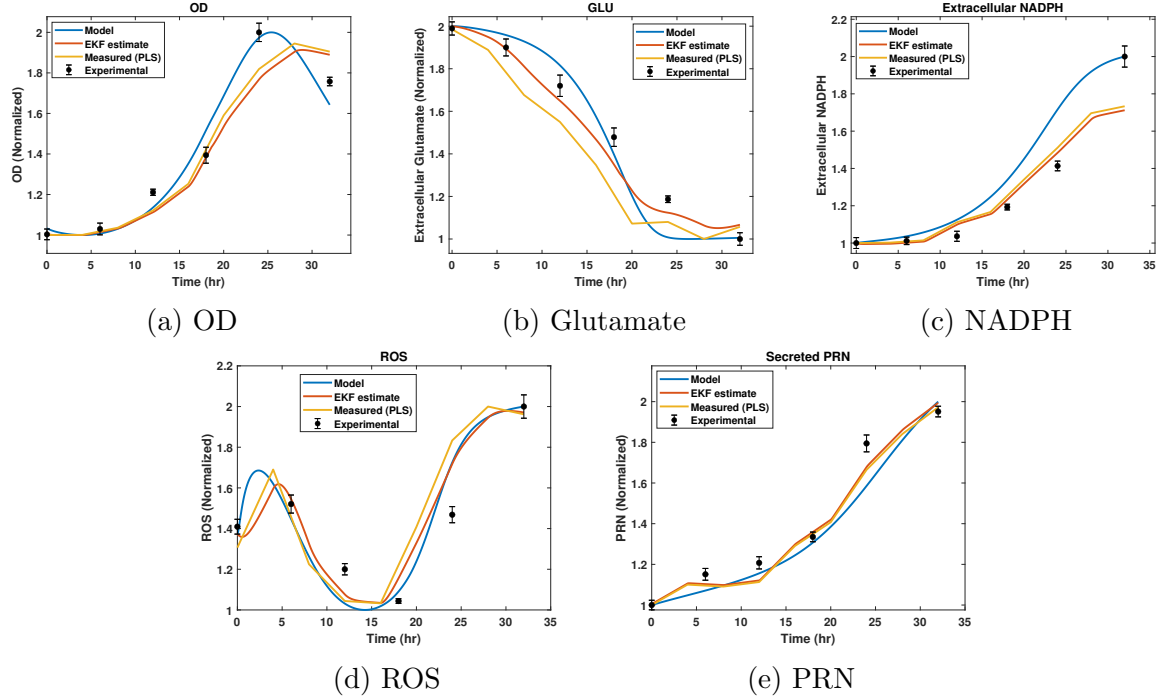


Figure 6.9: Extended kalman filter validation model for 32 hour batch runs for the estimation of paramters such as OD, glutamate, NADPH, ROS and PRN for 10.3 g/L initial glutamate concentration and inoculation OD of of 0.3. PLS regression mode and mechanistic model are also plotted against validation batch for comparison.

Table 6.7: Performance of extended Kalman filter and mechanistic model for the estimation of different state variables for the validation of batch with 12.03 g/L of initial glutamate concentration.

Parameter	OD	Glutamate	ROS	Extracellular NADPH	PRN
RMSE-c - EKF	0.0380	0.0433	0.1055	0.0757	0.0708
RMSE-v - EKF	0.0667	0.0844	0.1006	0.1393	0.0620
RMSE-c - Mechanistic Model	0.0412	0.0658	0.1203	0.0902	0.0804
RMSE-v - Mechanistic Model	0.1025	0.1821	0.1293	0.1254	0.0831

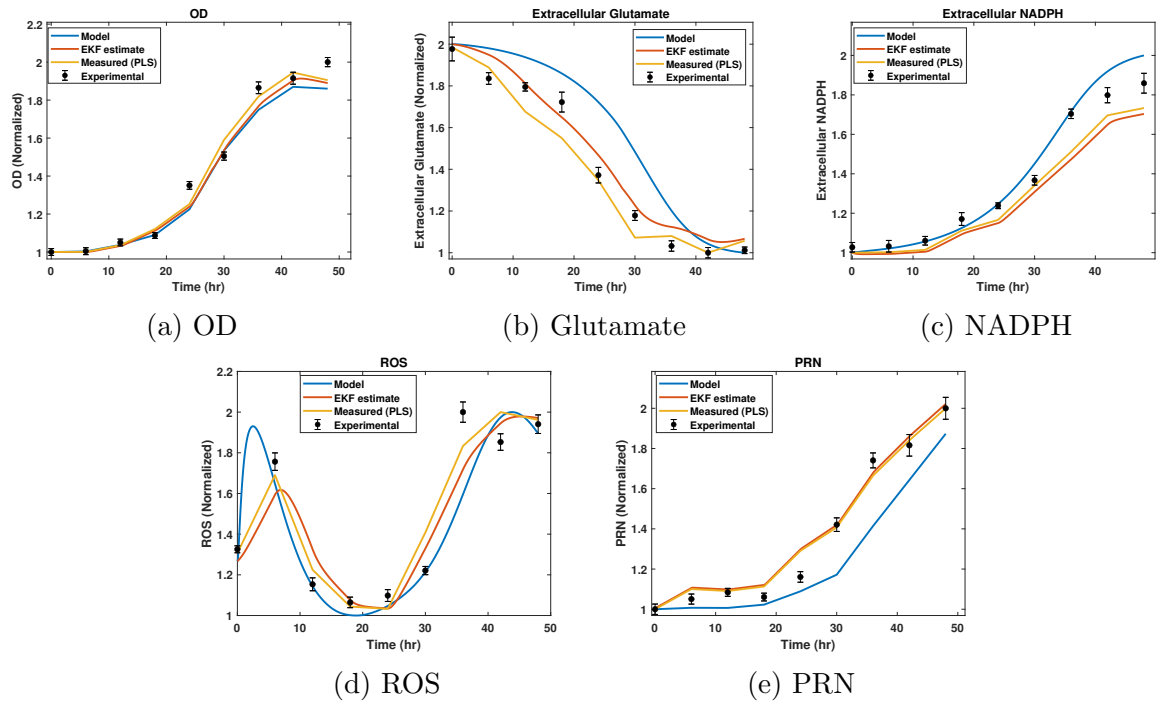


Figure 6.10: Extended kalman filter validation model for the 48 hour fed-batch runs for the estimation of OD, glutamate, NADPH, ROS and PRN for the starting OD of 0.28 and 12.03 g/L of initial glutamate concentration. PLS regression estimates and mechanistic model estimates are also plotted against validation batches for comparison.

6.4.5 Akaike Information Criteria

The Akaike Information Criterion (AIC) values for the Mechanistic Model, PLS Regression Model, and Extended Kalman Filter were compared to evaluate the performance of these models in the context of predicting different state variables in the Bordetella pertussis cultivation process. The AIC is a measure used to assess the quality of statistical models, with lower values indicating a better fit to the data while penalizing for model complexity. A summary of the model performance in terms of AIC values is given in table (6.8). Following this comparison, the Extended Kalman Filter (EKF) shows the lowest AIC value of 201.17, suggesting that it provided the best balance between model accuracy and complexity. The Mechanistic Model followed with an AIC value of 231.34, indicating a slightly less optimal fit but still better than the PLS Regression Model, which had the highest AIC value of 325.67.

These results clearly indicate that the EKF was the most effective model for accurately capturing the dynamics of the system with minimal over-fitting, making it a more reliable choice for real-time monitoring and control of the bioprocess.

Table 6.8: AIC value comparison of mechanistic model, PLSR and EKF.

Parameter	Mechanistic Model	PLS Regression Model	Extended Kalman Filter
AIC Value	231.34	325.67	201.17

6.5 Conclusions

A metabolic model was designed to predict the growth and pertactin (antigen) productivity of *B. pertussis* under different oxidative stress levels was calibrated and validated using data from batch and fed-batch fermentation experiments. The focus was on integrating in-line fluorescence spectroscopy with the mechanistic model, thus incorporating both mechanistic and data-driven elements to improve the accuracy of state estimation in *B. pertussis* fermentation. Experiments included two different initial glutamate concentrations as it was found that a higher glutamate concentration induces higher ROS in the fermentation batches that leads to lower growth and productivity. The experimental data provided clear evidence of substrate inhibition due to glutamate, which increased oxidative stress, and the mechanistic model accurately depicted these effects. Most importantly, the high levels of oxidative stress persisted during the fed-batch period almost until the end of the run. This caused a consistent and significant reduction in pertactin productivity of similar magnitude as observed in the manufacturing scale process. This is a first lab scale demonstration of the variability observed in production under similar conditions to the ones imposed at the large scale.

Fermentation batches were monitored using an in-line fluorometer which was previously used for in-line monitoring of *B. pertussis* fermentation in chapter (4). PLS regression model was developed using fluorescence spectra integrals which accurately predicted key parameters such as biomass, glutamate and NADPH. To further improve the accuracy of other state variables, mechanistic model was coupled to PLS regression model using

extended kalman filter (EKF) for bioreactor monitoring. Comparisons of model predictions to data with different model structures concluded that the hybrid model, combining fluorescence spectroscopy data with the mechanistic model, can significantly improve the monitoring and control of *B. pertussis* fermentation. The EKF model performed better than both mechanistic and regression model for OD, glutamate, NADPH, ROS and PRN. The predictive accuracy of the mechanistic model for ROS and extracellular NADPH may be improved with more precise modeling of ROS levels during lag, log and death phase.

To compare the model in terms of model accuracy and over-fitting, AIC values for PLS regression model, mechanistic model and EKF models were calculated and compared. The comparison showed that the hybrid model (EKF) accurately captured system dynamics with lowest over-fitting. In conclusion, this work demonstrates the successful integration of mechanistic modeling with a data-driven approach to enhance the predictive accuracy and robustness of *Bordetella pertussis* fermentation monitoring.

Chapter 7

Conclusions and Future Work

7.1 Conclusions

This thesis addressed different aspects of the closed loop bioreactor operation of the whooping cough vaccine manufacturing process as follows:

- Investigation of the correlation between productivity and culture conditions, in particular the interplay between high initial nutrient concentration, oxidative stress, growth and antigen productivity,
- Mechanistic modelling of these phenomena in closed loop bioreactor operation.
- Development of online monitoring method for *B. pertussis* that will serve in the future for continuous culture monitoring and for informing future process changes and possible corrective actions.

7.1.1 In-Line Monitoring of Bordetella Pertussis Cultivation Using Fluorescence Spectroscopy

The findings from this part of the research are as follows:

1. **High accuracy of biomass and amino acid prediction:** The Partial Least Squares (PLS) regression models developed using the fluorescence spectra acquired using a 2D in-line fluorometer demonstrated highly accurate predictions of biomass and amino acid concentrations, specifically glutamate and proline. This result was significant because it validated the potential of fluorescence spectroscopy as a reliable tool for real-time monitoring of these critical parameters for *B. pertussis* cultivation. This method can also be applied to other bacterial cultures. The ability to accurately predict biomass and main nutrient (glutamate) in real-time is expected to allow for more precise control of the bioprocess, leading to improved consistency in production outcomes.
2. **Strain-specific calibration enhanced predictive Performance:** The study found that the predictive performance of the PLS models improved significantly when they were calibrated specifically for the bacterial strain and media composition used in each cultivation process. This result highlighted the importance of customizing monitoring tools to the specific conditions of each bioprocess. The strain-specific models were needed to account for the distinct behavior of different bacterial strains, leading to more accurate predictions of target variables. This finding suggested that bioprocess monitoring systems should be tailored to the specific strains and condi-

tions of each production run to optimize performance.

3. **Challenges in pertactin prediction due to measurement noise:** One of the key challenges identified in the study was the difficulty in accurately predicting pertactin levels using the fluorescence-based PLS models. Noise in the offline data acquired with either ELISA and HPLC motivated the need for a more reliable quantification of the antigen. At that stage of the study, lack of a better quantification method of pertactin was found as the key obstacle for both gaining further understanding about the process and developing reliable online estimators of pertactin. The chromatographic technique developed in Chapter 4 partially addressed this challenge as explained below.

7.1.2 Correlation between Oxidative Stress and Growth & Antigen Productivity in *B. Pertussis* cultures

The key findings from this section of the research are as follows:

1. **Chromatography based quantification of antigen** A more precise quantification of pertactin in the fermentation broth was instrumental for further elucidation of the effect of culture conditions on antigen productivity. The chromatography method was based on a method previously developed for downstream purified samples.
2. **Influence of glutamate composition on oxidative stress:** The study found that the composition of the growth media, particularly the initial concentration of glutamate, is accompanied by high oxidative stress in *Bordetella pertussis* cultures.

Higher glutamate levels were directly associated with increased ROS production. At this stage it is not possible to assess whether oxidative stress causes growth inhibition or whether growth inhibition results in higher stress but both phenomena, i.e. growth inhibition and high stress appear to be highly correlated. This result underscores the importance of optimizing media composition to enhance the efficiency of the bioprocess. By carefully adjusting nutrient levels, it is possible to reduce one of the key sources of variability in process performance. While this research only investigated the effect of high glutamate since it is the main nutrient in the media, variability in other amino-acids may have similar effect.

3. Impact of oxidative stress on antigen productivity: The study found a significant correlation between oxidative stress and the reduction in antigen productivity, specifically Pertactin (PRN), in *Bordetella pertussis* cultures. Elevated levels of Reactive Oxygen Species (ROS) were shown to inversely correlate with the productivity of the bacteria. This was particularly evident in cultures with higher initial concentrations of glutamate, which led to increased oxidative stress. This result emphasizes the importance of monitoring and controlling oxidative stress in bioreactor operations to improve overall process outcome. For example, autoclaving of media before fermentation may result in differing amounts of water evaporation thus resulting in variability in initial glutamate levels. Such events should be monitored and mitigated to avoid negative impacts on productivity.

4. Differential effects of oxidative stress in flask vs. bioreactor conditions: Another key finding of the study was the differential impact of oxidative stress ob-

served between flask and bioreactor conditions. While oxidative stress was present in both setups, its effects on antigen productivity and bacterial growth were more pronounced in the bioreactor environment. This difference was found to be related to closed loop manipulation in agitation and aeration rates for oxygen control in response to variations in growth and glutamate consumption rates among different runs.

5. **Corroboration of correlation between high levels of NADPH to high ROS and low productivity:** This correlation was first observed by a previous PhD from a large set of manufacturing samples. Thus, it was very important to reproduce this observation in laboratory bioreactor experiments under control conditions. It should be noticed that the observed NADPH concentration is from the fermentation broth. The explanation for such higher levels is that under stress conditions, growth is inhibited and thus NADPH it is not consumed towards biomass. As per the analysis in Chapter 6, NADPH is used towards growth and thus in the absence of growth it tends to reach higher intracellular levels thus resulting in higher secretion. Accumulation of NADPH in the supernatant was also observed under conditions of glutamate depletion. Again, in the absence of glutamate, there is no growth and thus NADPH cannot be consumed towards biomass. Then, NADPH reaches higher intracellular levels leading to added secretion.

7.1.3 Mechanistic model based monitoring of *B. pertussis* cultures

The findings from this part of the research are as follows:

1. **Development and validation of a mechanistic model:** A mechanistic model based on the relationship between ROS-biomass growth relation and role of NADPH in quenching ROS among bacterial cultures was successfully developed to simulate the *Bordetella pertussis* fermentation process, incorporating key state variables such as optical density (OD), glutamate concentration, reactive oxygen species (ROS), extracellular NADPH, and Pertactin (PRN) production. The model was validated against experimental data, showing a high degree of accuracy in predicting these variables. This result is significant as it demonstrates the model's capability to replicate the complex dynamics of the fermentation process, providing a robust tool for understanding and optimizing bioprocesses in real-time.
- 2.
3. **Improved predictive performance with PLS regression models:** Partial Least Squares (PLS) regression models were employed to predict key state variables based on fluorescence spectroscopy data. The study revealed that PLS regression, when combined with appropriate data preprocessing and feature selection, could effectively predict variables such as OD, glutamate concentration, and ROS levels. However, the PLS models were found to be less robust compared to the mechanistic models, particularly under conditions outside the calibration data range. This finding em-

phasizes the usefulness of PLS regression for real-time monitoring but also highlights its limitations when dealing with complex, nonlinear bioprocess dynamics.

4. Integration with extended kalman filter (EKF) for real-time Monitoring:

Mechanistic model was integrated with an Extended Kalman Filter (EKF) to enable real-time estimation of key state variables during fermentation. This integration allowed for continuous monitoring and adjustment of the process, significantly improving the ability to detect and respond to deviations from optimal conditions. The EKF-enhanced monitoring system showed better performance in predicting the states especially the product PRN, which was not very well estimated by either of the individual models. The hybrid model exhibited the lowest AIC values, indicating the best balance between model accuracy and complexity. This result reinforces the superiority of the mechanistic model in capturing the nuances of the fermentation process and its potential as a preferred method for bioprocess monitoring and control.

7.2 Future Work

Building on the conclusions of this research, this section presents potential directions for future studies both in terms of methodologies to isolate the causes of process variability as well as improving statistical and mechanistic models for better monitoring and process control. The following research avenues can be explored to improve on the results of this research work:

1. The current fluorescence-based in-line monitoring system could be further refined to

enhance its sensitivity and specificity. Fluorescence data acquired from a bioreactor broth which is continuously agitated are corrupted with higher noise as compared with fluorescence data acquired off-line. The location of the fluorometer can be optimized so as to reduce the level of noise while maintaining high sensitivity.

2. In chapter (4) it was shown that a single PLS model was not able to explain the variability among different strains and media. To deal with this, non-linear models such as artificial neural networks (ANN) and support vector machines (SVM) could be used to develop a universal model that applies to different media and strains. Artificial Neural Networks (ANNs) and Partial Least Squares (PLS) models are powerful tools for modeling and predicting outcomes in processes exhibiting nonlinear responses.
3. The current work focused on one particular source of variability in productivity, i.e. the impact of glutamate concentration in inducing oxidative stress. However, media for pertussis fermentation consists of more than 20 components and the vaccine contains a total of 5 antigens. Future work should explore the effect of variations of other media components and investigate their role in inducing oxidative stress, growth inhibition and their impact on productivity. For example, it is suspected that sodium concentration may also have an effect on growth and productivity. This is particularly important in view that glutamate is fed as monosodium glutamate and thus variations in msg levels will be accompanied by changes in the levels of sodium. Variations in salt concentrations will alter the osmotic balance across the cell membrane thus resulting in gradients between the extracellular and intracellular

environments.

4. Multi parameter flow cytometry was utilized to monitor ROS and surface expression of PRN. An effort to monitor intra-cellular NADPH along with ROS and surface PRN was also made using cytometry but due to lack of proper equipment, i.e. availability of proper laser, this could not be achieved. Intra-cellular NADPH concentrations can be measured using flow cytometry with commercially available dyes or by taking advantage of auto-fluorescence when excited with the appropriate laser wavelength. Utilizing such dyes may allow for the simultaneous monitoring of intracellular ROS and NADPH levels, which could help to confirm the relationship between these two species.
5. Bioreactor fermentations for *B. pertussis* were performed in closed loop where pH, temperature and dissolved oxygen were tightly controlled. Closed loop control of dissolved oxygen was found to exacerbate the effects of oxidative stress effect among the cultures by forcing a decrease in oxygen mass transfer and oxygen uptake rate. It is of interest to investigate the results of growth and productivity when the fermentations are performed under open-loop control of dissolved oxygen. While open loop operation is currently not allowed within the certified process specifications, it could be possible to strongly detune the controller. Such detuning, i.e. sluggish tuning, could permit to maintain slow closed loop operation while avoiding drastic reduction in agitation and aeration rates during periods of growth inhibition.
6. The mechanistic model developed for the monitoring of *B. pertussis* fermentation is based on the relationship between substrate consumption, ROS production and

quenching by NADPH. The model can be made more comprehensive by including reactions for other anti-oxidative reactions, adding equations to capture the bioprocess dynamics of all aminoacids, antigen, NADPH and ROS. Such comprehensive model could then be used for media optimization or for model based fault detection.

7. Polymerase Chain Reaction (PCR) can be used to determine which genes are active in *Bordetella pertussis* under significant oxidative stress. The expression levels of the genes of interest (genes related to oxidative stress) can be studied under different stress inducing conditions to see which factors cause oxidative stress among the bacterial cultures.
8. LC-MS (Liquid Chromatography-Mass Spectroscopy) currently available in our department can be used with isotopes to map the fluxes under conditions of oxidative stress. This could help elucidate which reactions are responsible for growth and glutamate consumption inhibition during high ROS conditions.
9. it was hypothesized in this thesis that variability in media autoclaving process in the production scale may cause variability in the concentrations of main nutrients such as glutamate. Since variability in initial glutamate concentration was identified as a possible source of variability in antigen productivity, it could be possible to emulate at the lab scale the variable autoclaving conditions and test the effect of media after auto-claving on the process performance. It is possible that the auto-claving of media causes degradation in some components which also may result in additional process variability.
10. The switch between batch to fed-batch operation is believed to have great effect on the

final productivity, The instant of the switch is identified at Sanofi by a corresponding spike in oxygen concentration that is caused when glutamate in the batch is almost depleted. In the fed-batch lab experiments we were not be able to identify that oxygen spike, possibly because of differences in mass transfer conditions and in controller tuning parameters' values between the small and large scale bioreactors. This should be studied further. In particular it could be interesting to perform experiments where the fed-batch operation is started sometime before the depletion of glutamate. The complete consumption of glutamate was observed to result in large ROS production rates.

References

- Ouissam Abbas, Audrey Pissard, and Vincent Baeten. Near-infrared, mid-infrared, and raman spectroscopy. In *Chemical Analysis of Food*, pages 77–134. Elsevier, 2020.
- Hervé Abdi and Lynne J Williams. Principal component analysis. *Wiley interdisciplinary reviews: computational statistics*, 2(4):433–459, 2010.
- Aysun Adan, Günel Alizada, Yağmur Kiraz, Yusuf Baran, and Ayten Nalbant. Flow cytometry: basic principles and applications. *Critical reviews in biotechnology*, 37(2): 163–176, 2017.
- Piyush Agarwal. *Application of Deep Learning in Chemical Processes: Explainability, Monitoring and Observability*. PhD thesis, University of Waterloo, 2022.
- Piyush Agarwal, Melih Tamer, M Hossein Sahraei, and Hector Budman. Deep learning for classification of profit-based operating regions in industrial processes. *Industrial & Engineering Chemistry Research*, 59(6):2378–2395, 2019.
- Piyush Agarwal, Melih Tamer, and Hector Budman. Explainability: Relevance based

- dynamic deep learning algorithm for fault detection and diagnosis in chemical processes. *Computers & Chemical Engineering*, 154:107467, 2021.
- Marziyeh Ashoori and Ahmad Saedisomeolia. Riboflavin (vitamin b2) and oxidative stress: a review. *British journal of nutrition*, 111(11):1985–1991, 2014.
- Ruth E Baker, Jose-Maria Pena, Jayaratnam Jayamohan, and Antoine Jérusalem. Mechanistic models versus machine learning, a fight worth fighting for the biological community? *Biology letters*, 14(5):20170660, 2018.
- Ann J Barbier, Allen Yujie Jiang, Peng Zhang, Richard Wooster, and Daniel G Anderson. The clinical progress of mrna vaccines and immunotherapies. *Nature biotechnology*, 40(6):840–854, 2022.
- Hülya Bayr. Reactive oxygen species. *Critical care medicine*, 33(12):S498–S501, 2005.
- Thomas Belcher, Violaine Dubois, Alex Rivera-Millot, Camille Locht, and Françoise Jacob-Dubuisson. Pathogenicity and virulence of bordetella pertussis and its adaptation to its strictly human host. *Virulence*, 12(1):2608–2632, 2021.
- Thomas S Blacker and Michael R Duchon. Investigating mitochondrial redox state using nadh and nadph autofluorescence. *Free Radical Biology and Medicine*, 100:53–65, 2016.
- Olga Borisova, Svetlana Yu Kombarova, Nelli S Zakharova, Marjolein van Gent, Vladimir A Aleshkin, Isabella Mazurova, and Frits R Mooi. Antigenic divergence between bordetella pertussis clinical isolates from moscow, russia, and vaccine strains. *Clinical and Vaccine Immunology*, 14(3):234–238, 2007.

- Petrus J Braspenning, Frank Thuijsman, and Antonius Jozef Martha Maria Weijters. *Artificial neural networks: an introduction to ANN theory and practice*, volume 931. Springer Science & Business Media, 1995.
- E Cabiscol, J Tamarit, and J Ros. Oxidative stress in bacteria and protein damage by reactive oxygen species. *International microbiology: the official journal of the Spanish Society for Microbiology*, 3(1):3–8, 2000a.
- E Cabiscol, J Tamarit, and J Ros. Oxidative stress in bacteria and protein damage by reactive oxygen species. *International microbiology: the official journal of the Spanish Society for Microbiology*, 3(1):3–8, 2000b.
- Elisa Cabiscol Català, Jordi Tamarit Sumalla, and Joaquim Ros Salvador. Oxidative stress in bacteria and protein damage by reactive oxygen species. *International Microbiology*, 2000, vol. 3, núm. 1, p. 3-8, 2000.
- Joseph E Cavanaugh and Andrew A Neath. The akaike information criterion: Background, derivation, properties, application, interpretation, and refinements. *Wiley Interdisciplinary Reviews: Computational Statistics*, 11(3):e1460, 2019.
- Li Chen, Zhaoyue Zhang, Atsushi Hoshino, Henry D Zheng, Michael Morley, Zoltan Arany, and Joshua D Rabinowitz. Nadph production by the oxidative pentose-phosphate pathway supports folate metabolism. *Nature metabolism*, 1(3):404–415, 2019.
- Steven A Cohen. Amino acid analysis using precolumn derivatization with 6-aminoquinolyl-n-hydroxysuccinimidyl carbamate. *Amino acid analysis protocols*, pages 39–47, 2000.

- Ozlem Coskun. Separation techniques: chromatography. *Northern clinics of Istanbul*, 3(2):156, 2016.
- Teresa Lopes da Silva, Jose Carlos Roseiro, and Alberto Reis. Applications and perspectives of multi-parameter flow cytometry to microbial biofuels production processes. *Trends in biotechnology*, 30(4):225–232, 2012.
- Laurent Dewasme, Philippe Bogaerts, and Alain Vande Wouwer. Monitoring of bioprocesses: mechanistic and data-driven approaches. In *Computational Intelligence Techniques for Bioprocess Modelling, Supervision and Control*, pages 57–97. Springer, 2009.
- Evgeniy Eruslanov and Sergei Kusmartsev. Identification of ros using oxidized dcfda and flow-cytometry. In *Advanced protocols in oxidative stress II*, pages 57–72. Springer, 2010.
- Saskia M Faassen and Bernd Hitzmann. Fluorescence spectroscopy and chemometric modeling for bioprocess monitoring. *Sensors*, 15(5):10271–10291, 2015a.
- Saskia M Faassen and Bernd Hitzmann. Fluorescence spectroscopy and chemometric modeling for bioprocess monitoring. *Sensors*, 15(5):10271–10291, 2015b.
- Michel Fasnacht and Norbert Polacek. Oxidative stress in bacteria and the central dogma of molecular biology. *Frontiers in Molecular Biosciences*, 8:671037, 2021.
- Mariana L Fazenda, Joao ML Dias, Linda M Harvey, Alison Nordon, Ruan Edrada-Ebel, David LittleJohn, and Brian McNeil. Towards better understanding of an industrial cell factory: investigating the feasibility of real-time metabolic flux analysis in *pichia pastoris*. *Microbial cell factories*, 12(1):1–14, 2013.

Keisuke Fujii. Extended kalman filter. *Refernce Manual*, 14:41, 2013.

Cláudia F. Galinha, Gilda Carvalho, C. Portugal, G. Guglielmi, G. Guglielmi, Rui Oliveira, J. Crespo, and M. Reis. Real-time monitoring of membrane bioreactors with 2d-fluorescence data and statistically based models. *Water science and technology : a journal of the International Association on Water Pollution Research*, 63 7:1381–8, 2011. doi: 10.2166/wst.2011.195.

Victoria Girard and Michael Mourez. Adhesion mediated by autotransporters of gram-negative bacteria: structural and functional features. *Research in microbiology*, 157(5): 407–416, 2006.

Meric Goker, Sevinc Kurbanoglu, Pinar Yilgor Huri, Mustafa Kemal Sezginürk, and Sibel A Ozkan. Bioprocess monitoring by biosensor-based technologies. In *Commercial Biosensors and Their Applications*, pages 259–285. Elsevier, 2020.

Aydin Golabgir and Christoph Herwig. Combining mechanistic modeling and raman spectroscopy for real-time monitoring of fed-batch penicillin production. *Chemie Ingenieur Technik*, 88(6):764–776, 2016.

Yusmel González-Hernández, Emilie Michiels, and Patrick Perré. A comprehensive mechanistic yeast model able to switch metabolism according to growth conditions. *Fermentation*, 8(12):710, 2022.

Alexander Graf, Jens Claßen, Dörte Solle, Bernd Hitzmann, Karsten Rebner, and Marek Hoehse. A novel led-based 2d-fluorescence spectroscopy system for in-line monitoring

- of chinese hamster ovary cell cultivations–part i. *Engineering in life sciences*, 19(5): 352–362, 2019a.
- Alexander Graf, Jens Claßen, Dörte Solle, Bernd Hitzmann, Karsten Rebner, and Marek Hoehse. A novel led-based 2d-fluorescence spectroscopy system for in-line monitoring of chinese hamster ovary cell cultivations–part i. *Engineering in life sciences*, 19(5): 352–362, 2019b.
- PRGDJ Graves and D Gardiner. Practical raman spectroscopy. *Springer*, 1989.
- Jeffrey Green and Mark S Paget. Bacterial redox sensors. *Nature Reviews Microbiology*, 2(12):954–966, 2004.
- Julian Haas and Boris Mizaikoff. Advances in mid-infrared spectroscopy for chemical analysis. *Annual Review of Analytical Chemistry*, 9:45–68, 2016.
- David S Hage. Affinity chromatography: a review of clinical applications. *Clinical chemistry*, 45(5):593–615, 1999.
- Barry Halliwell and John MC Gutteridge. *Free radicals in biology and medicine*. Oxford university press, USA, 2015.
- Yusuke Himeoka and Kunihiro Kaneko. Theory for transitions between exponential and stationary phases: universal laws for lag time. *Physical Review X*, 7(2):021049, 2017.
- Dirk C Hinz. Process analytical technologies in the pharmaceutical industry: the fda’s pat initiative. *Analytical and bioanalytical chemistry*, 384(5):1036–1042, 2006.

- Steve Hisiger and Mario Jolicoeur. A multiwavelength fluorescence probe: Is one probe capable for on-line monitoring of recombinant protein production and biomass activity? *Journal of biotechnology*, 117(4):325–336, 2005.
- Yuzhi Hong, Jie Zeng, Xiuhong Wang, Karl Drlica, and Xilin Zhao. Post-stress bacterial cell death mediated by reactive oxygen species. *Proceedings of the National Academy of Sciences*, 116(20):10064–10071, 2019.
- John J Horvath, Scott A Glazier, and Christopher J Spangler. In situ fluorescence cell mass measurements of *saccharomyces cerevisiae* using cellular tryptophan. *Biotechnology progress*, 9(6):666–670, 1993.
- Agnar Höskuldsson. Pls regression methods. *Journal of chemometrics*, 2(3):211–228, 1988.
- Samira Hosseini, Patricia Vázquez-Villegas, Marco Rito-Palomares, Sergio O Martinez-Chapa, Samira Hosseini, Patricia Vázquez-Villegas, Marco Rito-Palomares, and Sergio O Martinez-Chapa. Advantages, disadvantages and modifications of conventional elisa. *Enzyme-Linked Immunosorbent Assay (ELISA) from A to Z*, pages 67–115, 2018.
- Shuhua Hu. Akaike information criterion. *Center for Research in Scientific Computation*, 93:42, 2007.
- James A Imlay. The molecular mechanisms and physiological consequences of oxidative stress: lessons from a model bacterium. *Nature Reviews Microbiology*, 11(7):443–454, 2013a.
- James A Imlay. The molecular mechanisms and physiological consequences of oxidative

- stress: lessons from a model bacterium. *Nature Reviews Microbiology*, 11(7):443–454, 2013b.
- James A Imlay. Diagnosing oxidative stress in bacteria: not as easy as you might think. *Current opinion in microbiology*, 24:124–131, 2015.
- Minoru Kanehisa and Susumu Goto. Kegg: kyoto encyclopedia of genes and genomes. *Nucleic acids research*, 28(1):27–30, 2000.
- Selin Kara, Jakob Mueller, and Andreas Liese. Online analysis methods for monitoring of bioprocesses. *Chimica oggi*, 29:38–41, 04 2011.
- Tobias K Karakach, Patrick J McGinn, Jerome Choi, Scott P MacQuarrie, and Boris Tartakovsky. Real-time monitoring, diagnosis, and time-course analysis of microalgae *scenedesmus amdd* cultivation using dual excitation wavelength fluorometry. *Journal of Applied Phycology*, 27:1823–1832, 2015.
- Tobias K Karakach, Aurore Dachon, Jerome Choi, Carlos Miguez, Luke Masson, and Boris Tartakovsky. Fluorescence-based real time monitoring and diagnostics of recombinant *pichia pastoris* cultivations in a bioreactor. *Biotechnology Progress*, 35(2):e2761, 2019a.
- Tobias K Karakach, Aurore Dachon, Jerome Choi, Carlos Miguez, Luke Masson, and Boris Tartakovsky. Fluorescence-based real time monitoring and diagnostics of recombinant *pichia pastoris* cultivations in a bioreactor. *Biotechnology Progress*, 35(2):e2761, 2019b.
- David Komla Kessie. *Characterisation of Bordetella pertussis virulence mechanisms using engineered human airway tissue models*. Bayerische Julius-Maximilians-Universitaet Wuerzburg (Germany), 2021.

- Jan C König, Tobias Steinwedel, Dörte Solle, Patrick Lindner, Ingo de Vries, Thorleif Hentrop, Michael Findeis, Gernot T John, Thomas Scheper, and Sascha Beutel. Development and characterisation of a new fluorescence sensor for online monitoring of bioprocesses. *Journal of Sensors and Sensor Systems*, 7(2):461–467, 2018.
- Gregory V. Korshin, Chi-Wang Li, and Mark M. Benjamin. Monitoring the properties of natural organic matter through uv spectroscopy: A consistent theory. *Water Research*, 31(7):1787–1795, 1997. ISSN 0043-1354.
- Ernest Kuchar, M Karlikowska-Skwarnik, S Han, and A Nitsch-Osuch. Pertussis: history of the disease and current prevention failure. *Pulmonary dysfunction and disease*, pages 77–82, 2016.
- Joseph R Lacowicz. *Protein fluorescence*, volume 6. Springer Science & Business Media, 2000.
- Joseph R Lakowicz and Joseph R Lakowicz. Instrumentation for fluorescence spectroscopy. *Principles of fluorescence spectroscopy*, pages 25–61, 1999.
- J Lemire, A Alhasawi, VP Appanna, S Tharmalingam, and VD Appanna. Metabolic defence against oxidative stress: the road less travelled so far. *Journal of applied microbiology*, 123(4):798–809, 2017.
- Bing Li, Yinzi Lin, Wei Yu, David I Wilson, and Brent R Young. Application of mechanistic modelling and machine learning for cream cheese fermentation ph prediction. *Journal of Chemical Technology & Biotechnology*, 96(1):125–133, 2021.

- Jeff W Lichtman and José-Angel Conchello. Fluorescence microscopy. *Nature methods*, 2 (12):910–919, 2005.
- Derek Albert Long. Raman spectroscopy. *New York*, 1, 1977.
- VI Lushchak. Oxidative stress and mechanisms of protection against it in bacteria. *Biochemistry (Moscow)*, 66:476–489, 2001a.
- VI Lushchak. Oxidative stress and mechanisms of protection against it in bacteria. *Biochemistry (Moscow)*, 66:476–489, 2001b.
- Longhuan Ma, Amanda Caulfield, Kalyan K Dewan, and Eric T Harvill. Pertactin-deficient bordetella pertussis, vaccine-driven evolution, and reemergence of pertussis. *Emerging infectious diseases*, 27(6):1561, 2021.
- Daniel Madar, Erez Dekel, Anat Bren, Anat Zimmer, Ziv Porat, and Uri Alon. Promoter activity dynamics in the lag phase of escherichia coli. *BMC systems biology*, 7:1–13, 2013.
- Carl-Fredrik Mandenius, Nigel J Titchener-Hooker, et al. *Measurement, monitoring, modelling and control of bioprocesses*, volume 132. Springer, 2013.
- Lisa M Maurer, Elizabeth Yohannes, Sandra S Bondurant, Michael Radmacher, and Joan L Slonczewski. ph regulates genes for flagellar motility, catabolism, and oxidative stress in escherichia coli k-12. *Journal of bacteriology*, 187(1):304–319, 2005.
- Katherine M McKinnon. Flow cytometry: an overview. *Current protocols in immunology*, 120(1):5–1, 2018.

- Lisa Mears, Stuart M Stocks, Mads O Albaek, Gürkan Sin, and Krist V Gernaey. Mechanistic models for process development and optimization of fed-batch fermentation systems. In *Computer Aided Chemical Engineering*, volume 38, pages 1311–1316. Elsevier, 2016.
- Beatriz Miguelena Chamorro, Karelle De Luca, Gokul Swaminathan, Stéphanie Longet, Egbert Mundt, and Stéphane Paul. Bordetella bronchiseptica and bordetella pertussis: similarities and differences in infection, immuno-modulation, and vaccine considerations. *Clinical Microbiology Reviews*, 36(3):e00164–22, 2023.
- Abhishek Mishra, Melih Tamer, and Hector Budman. In-line monitoring of bordetella pertussis cultivation using fluorescence spectroscopy. *Bioprocess and Biosystems Engineering*, 46(6):789–802, 2023a.
- Abhishek Mishra, Melih Tamer, and Hector Budman. In-line monitoring of bordetella pertussis cultivation using fluorescence spectroscopy. *Bioprocess and Biosystems Engineering*, 46(6):789–802, 2023b.
- Abhishek Mishra, Sarah Steinbach, Ibrahim M Tamer, and Hector Budman. Effect of oxidative stress on antigen productivity in b. pertussis cultures. *Biochemical Engineering Journal*, 210:109421, 2024.
- Frits R Mooi, Inge HM Van Loo, Marjolein Van Gent, Qiushui He, Marieke J Bart, Kees J Heuvelman, Sabine C De Greeff, Dimitri Diavatopoulos, Peter Teunis, Nico Nagelkerke, et al. Bordetella pertussis strains with increased toxin production associated with pertussis resurgence. *Emerging infectious diseases*, 15(8):1206, 2009.

- James Munson, C Freeman Stanfield, and Bir Gujral. A review of process analytical technology (pat) in the us pharmaceutical industry. *Current Pharmaceutical Analysis*, 2(4):405–414, 2006.
- Tormod Næs and Harald Martens. Principal component regression in nir analysis: view-points, background details and selection of components. *Journal of chemometrics*, 2(2): 155–167, 1988.
- Eiichiro Nakazawa, Shigetoshi Nara, Sumiaki Ibuki, Hajime Yamamoto, and Pieter Dorenbos. Fundamentals of luminescence. In *Phosphor Handbook*, pages 17–88. CRC Press, 2022.
- Giang T Nguyen, Erin R Green, and Joan Mecsas. Neutrophils to the roscue: mechanisms of nadph oxidase activation and bacterial resistance. *Frontiers in cellular and infection microbiology*, 7:373, 2017.
- Delma J Nieves and Ulrich Heininger. Bordetella pertussis. *Emerging Infections* 10, pages 311–339, 2016a.
- Delma J Nieves and Ulrich Heininger. Bordetella pertussis. *Emerging Infections* 10, pages 311–339, 2016b.
- Etsuo Niki, Yasukazu Yoshida, Yoshiro Saito, and Noriko Noguchi. Lipid peroxidation: mechanisms, inhibition, and biological effects. *Biochemical and biophysical research communications*, 338(1):668–676, 2005.
- Mojtaba Noofeli. *Genetic analysis and characterisation of the BapC autotransporter of bordetella pertussis*. PhD thesis, University of Glasgow, 2008.

- Keita Odanaka, Masato Iwatsuki, Tomomitsu Satho, and Mineo Watanabe. Identification and characterization of a brilliant yellow pigment produced by *bordetella pertussis*. *Microbiology and immunology*, 61(11):490–496, 2017.
- Peter Ödman, Claus Lindvald Johansen, Lisbeth Olsson, Krist V Gernaey, and Anna Eliasson Lantz. Sensor combination and chemometric variable selection for online monitoring of *streptomyces coelicolor* fed-batch cultivations. *Applied microbiology and biotechnology*, 86(6):1745–1759, 2010a.
- Peter Ödman, Claus Lindvald Johansen, Lisbeth Olsson, Krist V Gernaey, and Anna Eliasson Lantz. Sensor combination and chemometric variable selection for online monitoring of *streptomyces coelicolor* fed-batch cultivations. *Applied microbiology and biotechnology*, 86:1745–1759, 2010b.
- World Health Organization et al. Laboratory manual for the diagnosis of whooping cough caused by *bordetella pertussis*/*bordetella parapertussis*: update 2014. Technical report, World Health Organization, 2014.
- H Proom and BCJG Knight. The minimal nutritional requirements of some species in the genus *bacillus*. *Microbiology*, 13(3):474–480, 1955.
- Annette K Regan, Hannah C Moore, Michael J Binks, Lisa McHugh, Christopher C Blyth, Gavin Pereira, Karin Lust, Mohinder Sarna, Ross Andrews, Damien Foo, et al. Maternal pertussis vaccination, infant immunization, and risk of pertussis. *Pediatrics*, 152(5), 2023.

- Michelle L Reniere. Reduce, induce, thrive: bacterial redox sensing during pathogenesis. *Journal of bacteriology*, 200(17):10–1128, 2018.
- Maria Isabel Ribeiro. Kalman and extended kalman filters: Concept, derivation and properties. *Institute for Systems and Robotics*, 43(46):3736–3741, 2004.
- Ruth C Rowland-Jones, F. V. D. van den Berg, A. Racher, Elaine B. Martin, and C. Jaques. Comparison of spectroscopy technologies for improved monitoring of cell culture processes in miniature bioreactors. *Biotechnology Progress*, 33:337 – 346, 2017. doi: 10.1002/btpr.2459.
- Catherine A Royer. Fluorescence spectroscopy. *Protein stability and folding*, pages 65–89, 1995a.
- Catherine A Royer. Fluorescence spectroscopy. *Protein stability and folding: Theory and practice*, pages 65–89, 1995b.
- Howard M Shapiro. *Practical flow cytometry*. John Wiley & Sons, 2005.
- Adrian D Shaw, Naheed Kaderbhai, Alun Jones, Andrew M Woodward, Royston Goodacre, Jem J Rowland, and Douglas B Kell. Noninvasive, on-line monitoring of the biotransformation by yeast of glucose to ethanol using dispersive raman spectroscopy and chemometrics. *Applied spectroscopy*, 53(11):1419–1428, 1999.
- Ranji Singh, Ryan J Mailloux, Simone Puiseux-Dao, and Vasu D Appanna. Oxidative stress evokes a metabolic adaptation that favors increased nadph synthesis and decreased nadh production in pseudomonas fluorescens. *Journal of bacteriology*, 189(18):6665–6675, 2007.

- Adam M Smith, Carlos A Guzmán, and Mark J Walker. The virulence factors of bordetella pertussis: a matter of control. *FEMS microbiology reviews*, 25(3):309–333, 2001.
- Gisela Storz and James A Imlay. Oxidative stress. *Current opinion in microbiology*, 2(2):188–194, 1999.
- Gisela Storz, Louis A Tartaglia, Spencer B Farr, and Bruce N Ames. Bacterial defenses against oxidative stress. *Trends in Genetics*, 6:363–368, 1990.
- Anna Surribas, David Geissler, Alexander Gierse, Thomas Scheper, Bernd Hitzmann, José Luis Montesinos, and Francisco Valero. State variables monitoring by in situ multi-wavelength fluorescence spectroscopy in heterologous protein production by pichia pastoris. *Journal of Biotechnology*, 124(2):412–419, 2006.
- Ewa Szwejszer-Zawislak, Mieszko M Wilk, Piotr Piszczek, Justyna Krawczyk, Daria Wilczyńska, and Daniela Hozbor. Evaluation of whole-cell and acellular pertussis vaccines in the context of long-term herd immunity. *Vaccines*, 11(1):1, 2022.
- Ana P Teixeira, Carla AM Portugal, Nuno Carinhas, João ML Dias, João P Crespo, Paula M Alves, MJT Carrondo, and Rui Oliveira. In situ 2d fluorometry and chemometric monitoring of mammalian cell cultures. *Biotechnology and bioengineering*, 102(4):1098–1106, 2009a.
- Ana P Teixeira, Carla AM Portugal, Nuno Carinhas, João ML Dias, João P Crespo, Paula M Alves, MJT Carrondo, and Rui Oliveira. In situ 2d fluorometry and chemometric monitoring of mammalian cell cultures. *Biotechnology and bioengineering*, 102(4):1098–1106, 2009b.

- Ana P Teixeira, Tiago M Duarte, MJT Carrondo, and Paula M Alves. Synchronous fluorescence spectroscopy as a novel tool to enable pat applications in bioprocesses. *Biotechnology and bioengineering*, 108(8):1852–1861, 2011a.
- Ana P Teixeira, Tiago M Duarte, MJT Carrondo, and Paula M Alves. Synchronous fluorescence spectroscopy as a novel tool to enable pat applications in bioprocesses. *Biotechnology and bioengineering*, 108(8):1852–1861, 2011b.
- M. Thalen. *Whooping Cough Vaccines: Production of Virulent B. Pertussis*. PhD thesis, Wageningen University, 2008a.
- Marcel Thalen. *Whooping cough vaccines: production of virulent B. pertussis*. Wageningen University and Research, 2008b.
- Marcel Thalen, Jan van den IJssel, Wim Jiskoot, Bert Zomer, Paul Roholl, Cornelis de Gooijer, Coen Beuvery, and Johannes Tramper. Rational medium design for bordetella pertussis: basic metabolism. *Journal of biotechnology*, 75(2-3):147–159, 1999.
- Randall D Tobias et al. An introduction to partial least squares regression. In *Proceedings of the twentieth annual SAS users group international conference*, volume 20. Citeseer, 1995.
- Apostolos Tsopanoglou and Ioscani Jiménez 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*, 32:100691, 2021.

- Umatheny Umatheva, Braden Sweeting, Léo Sauvaget, Nerissa Dela Rosa, John Riley, Melih Tamer, and Raja Ghosh. Purification of bacterial virulence factor pertactin using high affinity ligands. *Biochemical Engineering Journal*, 164:107760, 2020a.
- Umatheny Umatheva, Braden Sweeting, Léo Sauvaget, Nerissa Dela Rosa, John Riley, Melih Tamer, and Raja Ghosh. Purification of bacterial virulence factor pertactin using high affinity ligands. *Biochemical Engineering Journal*, 164:107760, 2020b.
- Marharyta Varatnitskaya, Adriana Degrossoli, and Lars I Leichert. Redox regulation in host-pathogen interactions: thiol switches and beyond. *Biological Chemistry*, 402(3): 299–316, 2021.
- Michael Vitelli, Ibrahim Melih Tamer, Mark Pritzker, and Hector Budman. Modeling the effect of oxidative stress on bordetella pertussis fermentations. *Biotechnology Progress*, page e3335, 2023a.
- Michael Vitelli, Ibrahim Melih Tamer, Mark Pritzker, and Hector Budman. Modeling the effect of oxidative stress on bordetella pertussis fermentations. *Biotechnology Progress*, 39(3):e3335, 2023b.
- Michael Vitelli, Ibrahim Melih Tamer, Mark Pritzker, and Hector Budman. Modeling the effect of oxidative stress on bordetella pertussis fermentations. *Biotechnology Progress*, 39(3):e3335, 2023c.
- Scott I Vrieze. Model selection and psychological theory: a discussion of the differences between the akaike information criterion (aic) and the bayesian information criterion (bic). *Psychological methods*, 17(2):228, 2012.

- Jue D Wang and Petra A Levin. Metabolism, cell growth and the bacterial cell cycle. *Nature Reviews Microbiology*, 7(11):822–827, 2009a.
- Jue D Wang and Petra A Levin. Metabolism, cell growth and the bacterial cell cycle. *Nature Reviews Microbiology*, 7(11):822–827, 2009b.
- Dhanuka P Wasalathanthri, Matthew S Rehmann, Jay M West, Michael C Borys, Julia Ding, and Zhen Jian Li. Paving the way for real time process monitoring in biomanufacturing. *Am Pharm Rev*, 23(5):54–58, 2020.
- Kathryn E Wellen and Craig B Thompson. Cellular metabolic stress: considering how cells respond to nutrient excess. *Molecular cell*, 40(2):323–332, 2010a.
- Kathryn E Wellen and Craig B Thompson. Cellular metabolic stress: considering how cells respond to nutrient excess. *Molecular cell*, 40(2):323–332, 2010b.
- Kathryn E Wellen and Craig B Thompson. Cellular metabolic stress: considering how cells respond to nutrient excess. *Molecular cell*, 40(2):323–332, 2010c.
- Svante Wold, Kim Esbensen, and Paul Geladi. Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(1-3):37–52, 1987a.
- Svante Wold, Kim Esbensen, and Paul Geladi. Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(1-3):37–52, 1987b.
- Lunzhao Yi, Naiping Dong, Yonghuan Yun, Baichuan Deng, Dabing Ren, Shao Liu, and Yizeng Liang. Chemometric methods in data processing of mass spectrometry-based metabolomics: A review. *Analytica chimica acta*, 914:17–34, 2016.

- Vanessa Zavatti and Hector Budman. Monitoring of an antigen manufacturing process. *Bioprocess and biosystems engineering*, 39(6):855–869, 2016.
- Vanessa Zavatti, Hector Budman, Raymond Legge, and Melih Tamer. Monitoring of an antigen manufacturing process. *Bioprocess and biosystems engineering*, 39:855–869, 2016.
- Vanessa Zavatti, Hector Budman, Raymond L Legge, and Melih Tamer. Impact of oxidative stress on protein production by bordetella pertussis for vaccine production. *Biochemical Engineering Journal*, 151:107359, 2019a.
- Vanessa Zavatti, Hector Budman, Raymond L Legge, and Melih Tamer. Impact of oxidative stress on protein production by bordetella pertussis for vaccine production. *Biochemical Engineering Journal*, 151:107359, 2019b.
- Vanessa Zavatti, Hector Budman, Raymond L Legge, and Melih Tamer. Investigation of the effects of oxidative stress-inducing factors on culturing and productivity of bordetella pertussis. *Biotechnology Progress*, 36(1):e2899, 2020a.
- Vanessa Zavatti, Hector Budman, Raymond L Legge, and Melih Tamer. Investigation of the effects of oxidative stress-inducing factors on culturing and productivity of bordetella pertussis. *Biotechnology progress*, 36(1):e2899, 2020b.
- Manuela Zlamy. Rediscovering pertussis. *Frontiers in pediatrics*, 4:52, 2016.