Upstream process development for production of biotherapeutics using CHO cell culture

NEELESH GANGWAR

SCHOOL OF INTERDISCIPLINARY RESEARCH (SIRe)

INDIAN INSTITUTE OF TECHNOLOGY DELHI

FEBRUARY 2025

© **Indian Institute of Technology Delhi (IITD), New Delhi, 2025**

Upstream process development for production of biotherapeutics using CHO cell culture

by

Neelesh Gangwar

School of Interdisciplinary Research (SIRe)

Submitted

in fulfilment of the requirements of the degree of Doctor of Philosophy

to the

Indian Institute of Technology, Delhi

February 2025

Certificate

This is to certify that the thesis entitled "**Upstream process development for production of biotherapeutics using CHO cell culture**" being submitted by **Neelesh Gangwar** to the Indian Institute of Technology Delhi for the award of the degree of **Doctor of Philosophy** is a record of the original bonafide research work carried out by him under my guidance and supervision. The results contained in this thesis have not been submitted in part or in full to any other University or Institute for the award of any degree or diploma.

We certify that he has pursued the prescribed course of research.

Prof. Anurag S. Rathore
Department of Chemical Engineering
Indian Institute of Technology Delhi

Prof. James Gomes
Kusuma School of Biological Sciences
Indian Institute of Technology Delhi

Acknowledgements

I would like to express my deepest gratitude to the following individuals and organizations who have been instrumental in the completion of this Ph.D. thesis:

I am profoundly thankful to my supervisor, Prof. Anurag S. Rathore for his unwavering support and guidance. His expertise and insights have been invaluable throughout this research endeavour. I am grateful for the mentorship and encouragement he provided, which significantly contributed to the quality of this thesis. His valuable suggestions and thorough check of my work allowed me to complete this research on a high note. I am also grateful to my co-supervisor Prof. James Gomes, for the experience and guidance that he provided in compiling and analysing the data systematically.

I extend my appreciation to the members of my Ph.D. research committee, Prof. Sudip K. Pattanayek, Prof. Gaurav Goel, and Prof. Manidipa Banerjee for their valuable feedback and constructive criticism. Their periodic assessment and review of my work helped in going forward in the right direction, it greatly enriched my research and helped shape this thesis.

I am grateful to Council of Scientific & Industrial Research (CSIR), Department of Biotechnology (DBT), Industrial Research & Development Unit (IRD) and PG Section at IIT Delhi for providing the necessary financial support during my research tenure. Their funding enabled me to conduct experiments, attend conferences, and access vital resources, without which this thesis would not have been possible. I appreciate the support and resources provided by IIT Delhi, including access to libraries, laboratories, and research materials. These resources were indispensable to the successful completion of this research.

I would like to acknowledge my fellow researchers and colleagues at IIT Delhi for their stimulating discussions and collaborative spirit. To my family and friends, thank you for your unwavering encouragement, love, and understanding. Your support sustained me during the challenging times of this academic journey. In conclusion, I want to acknowledge everyone who has been a part of this Ph.D. journey. Your encouragement, wisdom, and support have been invaluable. This achievement is as much yours as it is mine.

NEELESH GANGWAR

Abstract

Throughout the past few decades, biotherapeutics have demonstrated significant clinical success and have been utilized successfully to treat a variety of serious illnesses. Monoclonal antibodies (mAbs) remain dominant in the category of biotherapeutics in terms of approvals of genuinely new biopharmaceuticals coming to market and market value contribution during 2018 – 2022. Chinese hamster ovary (CHO) cell is chosen to be default mammalian expression system with 89% of total products expressed in mammalian system. CHO cell is most favoured expression system because of its capabilities for post translational modifications (PTMs) of mAbs which closely resembles human mAbs. Despite being most efficient therapeutic molecules, production of mAbs in CHO cells imposes certain challenges like making them cost effective, maintaining batch-to-batch consistency, understanding process variability to make them suitable for a continuous production platform. Cell culture media forms the immediate environment of cells where they grow and propagate and eliminate their waste. Media is the water-based solution in which many components such as amino acids, minerals, vitamins, and salts are dissolved to make it suitable for cell growth. In this study we explored the role of the media in manipulating the critical quality attributes (CQAs) of mAb therapeutics.

In chapter 3 we investigate how vitamins and metal ions affect protein expression and CQA, specifically charge heterogeneity of mAb. The study examined the use of pyridoxine hydrochloride, biotin, folic acid, choline chloride, thiamine hydrochloride, and D-calcium pantothenate, as well as metal ions like Fe, Zn, Cu, Co, Mg, Mn, and Ni as addition to cell culture media. The findings suggest that pyridoxine enhances volumetric productivity, while Zn, Cu, Fe, Mn, and biotin influence charge heterogeneity. Fe, Mn, and Ni increase the production of acidic variants, whereas Cu and biotin inhibit it. Zn decreases the formation of basic variants, while biotin increases them. Further we demonstrate how we can achieve the charge variant profile of the reference molecule Herceptin® by utilizing metal ions as charge variant modulators. The in-house produced trastuzumab exhibited significantly lower acidic variants (17.64 ± 1.07% acidic and 12.86 ± 0.43% basic) compared to the reference product (24.97 ± 0.54% acidic and 11.41 ± 1.44% basic). After a thorough investigation, zinc was selected to regulate basic charge variations and iron was used to modify acidic charge variants. With optimal metal ion supplementation, the charge variant profile for in-house monoclonal antibody (24.7 ± 0.83% acidic and 14.4 ± 0.64% basic) closely matched that of the reference product. Although a minor reduction in viable cell density and product titer (~7%) was

observed, there was no significant impact on other quality attributes, including glycans and specific cell productivity.

In chapter 4, an extensive screening of various media additives (including metal ions, vitamins, sugars, and nucleosides) was conducted to evaluate their effects on glycosylation. This was followed by optimizing the concentration of the shortlisted components to control the galactosylation, afucosylation, mannosylation, and sialylation of monoclonal antibodies (mAbs). A one factor at a time (OFAT) approach was employed to screen the media additives for their effects. Six additives were shortlisted based on their impact and used to modulate the glycosylation profile of an in-house produced mAb (G0 2.38 ± 0.08%, G0F 75.58 ± 0.45%, G1F 10.07 ± 0.04%, G2F 0.54 ± 0.01%, G0F-N 5.84 ± 0.32%, sialylation 1.60 ± 0.33%, mannosylation 1.56 ± 0.39%) to match the glycan profile of a commercially available reference product (G0 2.49 ± 0.07%, G0F 37.83 ± 0.37%, G1F 34.77 ± 0.03%, G2F 4.87 ± 0.01%, G0F-N 2.34 ± 0.12%, sialylation 9.84 ± 0.30%, mannosylation 2.86 ± 0.29%). The proposed approach resulted in a glycan profile (G0 2.10 ± 0.07%, G0F 38.00 ± 0.49%, G1F 31.92 ± 0.09%, G2F 5.26 ± 0.54%, G0F-N 1.92 ± 0.02%, sialylation 10.28 ± 1.68%, mannosylation 3.12 ± 0.29%) that was nearly identical to the reference product. Importantly, other quality attributes, including charge variants, aggregates, titer, and cell viability, were not significantly impacted by the addition of these additives.

Due to the inherent complexity of cell culture media, understanding the impact of various media components on cell growth and CQAs is challenging. In chapter 5 an end-to-end machine learning framework for media component selection, CQA prediction, and optimization is presented. The preliminary dataset for feature selection was generated by conducting CHO-GS(-/-) cell culture in media formulations with varying metal ion concentrations. The acidic and basic charge variant composition of the reference product (24.97 ± 0.54% acidic and 11.41 ± 1.44% basic) was chosen as the target variable to evaluate the media formulations. Pearson's correlation coefficient and random forest-based techniques were used for feature ranking and selection to predict acidic and basic charge variants. Additionally, a global interpretation analysis using Shapley Additive exPlanations (SHAP) was utilized to select optimal features by evaluating the contributions of each feature in the extracted vectors. Finally, fifteen different regression models were employed to predict medium combinations, with grid search and random search cross-validation used for hyperparameter optimization. Experimental results demonstrate that Fe and Zn significantly impact the charge variant profile. This study aims to provide insights relevant to both innovators looking to establish a complete

pipeline for media development and optimization, and biosimilar manufacturers striving to demonstrate the analytical and functional biosimilarity of their products to the innovator.

In the Quality by Design (QbD) paradigm, controlling raw materials based on their impact on product qualities is a key aspect of developing a control strategy and maintain the consistency and quality of product. In chapter 6, a near-infrared (NIR) spectroscopy-based quantification method was developed for media additives known to be potential glycan modulators. A convolutional neural network (CNN)-based chemometric model was created to estimate galactose and uridine concentrations in various media formulations. By leveraging data augmentation, the CNN model demonstrated strong predictive performance (test $R^2 > 0.9$) for both analytes in real-time. This model was further combined with a Design of Experiments (DoE) based statistical model to predict glycosylation using the concentrations of media additives as inputs. The predicted glycosylation distributions aligned closely with the actual distributions, showing no significant differences in the investigated media formulations. The proposed approach can be utilized for media characterization and the implementation of process analytical technology (PAT) based control of mammalian cell culture raw materials.

Overall, this thesis aims to address the key challenges faced by the biopharmaceutical industry in upstream process development (USP) more specifically in media optimization. Results of these studies could be utilized for maintaining consistent CQAs (charge variant or glycosylation).

सार

पिछले कुछ दशकों में, बायोथेरेप्यूटिक्स ने महत्वपूर्ण नैदानिक सफलता प्रदर्शित की है और विभिन्न गंभीर बीमारियों के इलाज के लिए इसका सफलतापूर्वक उपयोग किया गया है। 2018-2022 के दौरान बाजार में आने वाले वास्तविक नए बायोफार्मास्यूटिकल्स की मंजूरी और बाजार मूल्य योगदान के मामले में मोनोक्लोनल एंटीबॉडी (एमएबीएस) बायोथेराप्यूटिक्स की श्रेणी में प्रमुख बने हुए हैं। चीनी हैम्स्टर अंडाशय (सीएचओ) सेल को स्तनधारी प्रणाली में व्यक्त कुल उत्पादों के 89% के साथ डिफ़ॉल्ट स्तनधारी अभिव्यक्ति प्रणाली के रूप में चुना गया है। CHO सेल mAbs के पोस्ट ट्रांसलेशनल संशोधनों (PTMs) के लिए अपनी क्षमताओं के कारण सबसे पसंदीदा अभिव्यक्ति प्रणाली है जो मानव mAbs से काफी मिलती जुलती है। सबसे कुशल चिकित्सीय अणु होने के बावजूद, सीएचओ कोशिकाओं में एमएबी का उत्पादन कुछ चुनौतियों का सामना करता है जैसे उन्हें लागत प्रभावी बनाना, बैच-टू-बैच स्थिरता बनाए रखना, निरंतर उत्पादन मंच के लिए उपयुक्त बनाने के लिए प्रक्रिया परिवर्तनशीलता को समझना। सेल कल्चर मीडिया कोशिकाओं का तात्कालिक वातावरण बनाता है जहां वे बढ़ते हैं, फैलते हैं और अपने अपशिष्ट को खत्म करते हैं। मीडिया जल-आधारित समाधान है जिसमें अमीनो एसिड, खनिज, विटामिन और नमक जैसे कई घटकों को कोशिका वृद्धि के लिए उपयुक्त बनाने के लिए घोल दिया जाता है। इस अध्ययन में हमने mAb चिकित्सीय की महत्वपूर्ण गुणवत्ता विशेषताओं (CQAs) में हेरफेर करने में मीडिया की भूमिका का पता लगाया।

अध्याय 3 में हम जांच करते हैं कि विटामिन और धातु आयन प्रोटीन अभिव्यक्ति और सीक्यूए को कैसे प्रभावित करते हैं, विशेष रूप से एमएबी की विषमता को चार्ज करते हैं। अध्ययन में सेल कल्चर मीडिया के अतिरिक्त पाइरिडोक्सिन हाइड्रोक्लोराइड, बायोटिन, फोलिक एसिड, कोलीन क्लोराइड, थायमिन हाइड्रोक्लोराइड और डी-कैल्शियम पैंटोथेनेट के साथ-साथ Fe, Zn, Cu, Co, Mg, Mn और Ni जैसे धातु आयनों के उपयोग की जांच की गई। निष्कर्षों से पता चलता है कि पाइरिडोक्सिन वॉल्यूमेट्रिक उत्पादकता को बढ़ाता है, जबकि Zn, Cu, Fe, Mn और बायोटिन चार्ज विविधता को प्रभावित करते हैं। Fe, Mn, और Ni अम्लीय वेरिएंट के उत्पादन को बढ़ाते हैं, जबकि Cu और बायोटिन इसे रोकते हैं। Zn बुनियादी वेरिएंट के गठन को कम करता है, जबकि बायोटिन उन्हें बढ़ाता है। इसके अलावा हम प्रदर्शित करते हैं कि हम चार्ज वेरिएंट मॉड्यूलेटर के रूप में धातु आयनों का उपयोग करके संदर्भ अणु हर्सेप्टिन® के चार्ज वेरिएंट प्रोफाइल को कैसे प्राप्त कर सकते हैं। घर में उत्पादित ट्रैस्टुज़ुमैब ने संदर्भ उत्पाद (24.97 ± 0.54% अम्लीय और 11.41 ± 1.44% बुनियादी) की तुलना में काफी कम अम्लीय वेरिएंट (17.64 ± 1.07% अम्लीय और 12.86 ± 0.43% बुनियादी) प्रदर्शित किया। गहन जांच के बाद, बुनियादी चार्ज भिन्नताओं

को विनियमित करने के लिए जस्ता का चयन किया गया और अम्लीय चार्ज वेरिएंट को संशोधित करने के लिए लोहे का उपयोग किया गया। इष्टतम धातु आयन अनुपूरण के साथ, इन-हाउस मोनोक्लोनल एंटीबॉडी (24.7 ± 0.83% अम्लीय और 14.4 ± 0.64% बुनियादी) के लिए चार्ज वैरिएंट प्रोफ़ाइल संदर्भ उत्पाद से निकटता से मेल खाती है। यद्यपि व्यवहार्य सेल घनत्व और उत्पाद अनुमापांक (~7%) में मामूली कमी देखी गई, लेकिन ग्लाइकेन और विशिष्ट सेल उत्पादकता सहित अन्य गुणवत्ता विशेषताओं पर कोई महत्वपूर्ण प्रभाव नहीं पड़ा।

अध्याय 4 में, ग्लाइकोसिलेशन पर उनके प्रभावों का मूल्यांकन करने के लिए विभिन्न मीडिया एडिटिव्स (धातु आयन, विटामिन, शर्करा और न्यूक्लियोसाइड सहित) की व्यापक जांच की गई थी। इसके बाद गैलेक्टोसाइलेशन, एफ़ुकोसिलेशन, मैन्नोसिलेशन और मोनोक्लोनल एंटीबॉडी (एमएबीएस) के सियालिलेशन को नियंत्रित करने के लिए शॉर्टलिस्ट किए गए घटकों की एकाग्रता को अनुकूलित किया गया। मीडिया एडिटिव्स के प्रभावों की स्क्रीनिंग के लिए एक समय में एक कारक (ओएफएटी) दृष्टिकोण को नियोजित किया गया था। छह एडिटिव्स को उनके प्रभाव के आधार पर शॉर्टलिस्ट किया गया था और इन-हाउस उत्पादित mAb (G0 2.38 ± 0.08%, G0F 75.58 ± 0.45%, G1F 10.07 ± 0.04%, G2F 0.54 ± 0.01%, G0F-N 5.84 ±) के ग्लाइकोसिलेशन प्रोफाइल को मॉड्यूलेट करने के लिए उपयोग किया गया था। व्यावसायिक रूप से उपलब्ध संदर्भ उत्पाद (जी0 2.49 ± 0.07%, जी0एफ 37.83 ± 0.37%, जी1एफ 34.77 ± 0.03%, जी2एफ) के ग्लाइकेन प्रोफाइल से मेल खाने के लिए 0.32%, सियालिलेशन 1.60 ± 0.33%, मैनोसाइलेशन 1.56 ± 0.39%) 4.87 ± 0.01%, जी0एफ-एन 2.34 ± 0.12%, सियालिलेशन 9.84 ± 0.30%, मैन्नोसिलेशन 2.86 ± 0.29%)। प्रस्तावित दृष्टिकोण के परिणामस्वरूप ग्लाइकेन प्रोफाइल (G0 2.10 ± 0.07%, G0F 38.00 ± 0.49%, G1F 31.92 ± 0.09%, G2F 5.26 ± 0.54%, G0F-N 1.92 ± 0.02%, सियालिलेशन 10.28 ± 1.68%, मैन्नोसिलेशन 3.12 ± 0.29%) जो लगभग संदर्भ उत्पाद के समान था। महत्वपूर्ण बात यह है कि चार्ज वेरिएंट, एग्रीगेट्स, टिटर और सेल व्यवहार्यता सहित अन्य गुणवत्ता विशेषताओं पर इन एडिटिव्स के शामिल होने से महत्वपूर्ण प्रभाव नहीं पड़ा।

सेल कल्चर मीडिया की अंतर्निहित जटिलता के कारण, सेल विकास और सीक्यूए पर विभिन्न मीडिया घटकों के प्रभाव को समझना चुनौतीपूर्ण है। अध्याय 5 में मीडिया घटक चयन, सीक्यूए भविष्यवाणी और अनुकूलन के लिए एक एंड-टू-एंड मशीन लर्निंग ढांचा प्रस्तुत किया गया है। फीचर चयन के लिए प्रारंभिक डेटासेट अलग-अलग धातु आयन सांद्रता के साथ मीडिया फॉर्मूलेशन में सीएचओ-जीएस (-/-) सेल संस्कृति का संचालन करके तैयार किया गया था। मीडिया फॉर्मूलेशन का मूल्यांकन करने के लिए संदर्भ उत्पाद की अम्लीय और बुनियादी चार्ज वैरिएंट संरचना (24.97 ± 0.54% अम्लीय और 11.41 ± 1.44%

बुनियादी) को लक्ष्य चर के रूप में चुना गया था। अम्लीय और बुनियादी चार्ज वेरिएंट की भविष्यवाणी करने के लिए फीचर रैंकिंग और चयन के लिए पियर्सन के सहसंबंध गुणांक और यादृच्छिक वन-आधारित तकनीकों का उपयोग किया गया था। इसके अतिरिक्त, निकाले गए वैक्टर में प्रत्येक सुविधा के योगदान का मूल्यांकन करके इष्टतम सुविधाओं का चयन करने के लिए शेपली एडिटिव एक्सप्लानेशंस (एसएचएपी) का उपयोग करके एक वैश्विक व्याख्या विश्लेषण का उपयोग किया गया था। अंत में, पंद्रह अलग-अलग प्रतिगमन मॉडल को मध्यम संयोजनों की भविष्यवाणी करने के लिए नियोजित किया गया था, जिसमें ग्रिड खोज और हाइपरपैरामीटर अनुकूलन के लिए यादृच्छिक खोज क्रॉस-सत्यापन का उपयोग किया गया था। प्रायोगिक परिणाम दर्शाते हैं कि Fe और Zn चार्ज वैरिएंट प्रोफ़ाइल को महत्वपूर्ण रूप से प्रभावित करते हैं। इस अध्ययन का उद्देश्य मीडिया विकास और अनुकूलन के लिए एक संपूर्ण पाइपलाइन स्थापित करने के इच्छुक इनोवेटर्स और इनोवेटर को अपने उत्पादों की विश्लेषणात्मक और कार्यात्मक बायोसिमिलरिटी प्रदर्शित करने का प्रयास करने वाले बायोसिमिलर निर्माताओं दोनों के लिए प्रासंगिक अंतर्दृष्टि प्रदान करना है।

डिजाइन द्वारा गुणवत्ता (क्यूबीडी) प्रतिमान में, उत्पाद की गुणवत्ता पर उनके प्रभाव के आधार पर कच्चे माल को नियंत्रित करना एक नियंत्रण रणनीति विकसित करने और उत्पाद की स्थिरता और गुणवत्ता बनाए रखने का एक महत्वपूर्ण पहलू है। अध्याय 6 में, संभावित ग्लाइकेन मॉड्यूलेटर के रूप में जाने जाने वाले मीडिया एडिटिव्स के लिए एक निकट-अवरक्त (एनआईआर) स्पेक्ट्रोस्कोपी-आधारित मात्रा निर्धारण विधि विकसित की गई थी। विभिन्न मीडिया फॉर्मूलेशन में गैलेक्टोज और यूरिडीन सांद्रता का अनुमान लगाने के लिए एक कन्वेन्शनल न्यूरल नेटवर्क (सीएनएन)-आधारित केमोमेट्रिक मॉडल बनाया गया था। डेटा संवर्द्धन का लाभ उठाकर, सीएनएन मॉडल ने वास्तविक समय में दोनों विश्लेषणों के लिए मजबूत पूर्वानुमानित प्रदर्शन (परीक्षण आर^2 > 0.9) प्रदर्शित किया। इनपुट के रूप में मीडिया एडिटिव्स की सांद्रता का उपयोग करके ग्लाइकोसिलेशन की भविष्यवाणी करने के लिए इस मॉडल को डिज़ाइन ऑफ एक्सपेरिमेंट्स (डीओई) आधारित सांख्यिकीय मॉडल के साथ जोड़ा गया था। अनुमानित ग्लाइकोसिलेशन वितरण वास्तविक वितरण के साथ निकटता से संरेखित हैं, जांच किए गए मीडिया फॉर्मूलेशन में कोई महत्वपूर्ण अंतर नहीं दिखा रहा है। प्रस्तावित दृष्टिकोण का उपयोग मीडिया लक्षण वर्णन और स्तनधारी कोशिका संस्कृति कच्चे माल के प्रक्रिया विश्लेषणात्मक प्रौद्योगिकी (पीएटी) आधारित नियंत्रण के कार्यान्वयन के लिए किया जा सकता है।

कुल मिलाकर, इस थीसिस का उद्देश्य विशेष रूप से मीडिया अनुकूलन में अपस्ट्रीम प्रक्रिया विकास (यूएसपी) में बायोफार्मास्युटिकल उद्योग के सामने आने वाली प्रमुख चुनौतियों का समाधान करना है।

इन अध्ययनों के परिणामों का उपयोग सुसंगत CQAs (चार्ज वेरिएंट या ग्लाइकोसिलेशन) को बनाए रखने के लिए किया जा सकता है।

Content

CERTIFICATE ..... **I**

ACKNOWLEDGEMENTS ..... **III**

ABSTRACT ..... **V**

CONTENT ..... **XIII**

LIST OF FIGURES ..... **XV**

LIST OF TABLES ..... **XVII**

LIST OF ABBREVIATIONS ..... **XIX**

SYMBOLS ..... **XXI**

CHAPTER 1 INTRODUCTION AND OBJECTIVES **1**

1.1 BACKGROUND 2

1.2 PROJECT OVERVIEW AND RESEARCH OBJECTIVES 5

CHAPTER 2 REVIEW OF LITERATURE **9**

2.1 RECOMBINANT PROTEIN PRODUCTION 10

2.2 RECOMBINANT PROTEIN PRODUCTION IN MAMMALIAN CELLS 12

2.3 MAMMALIAN CELL CULTURE 15

2.4 CELL CULTURE MEDIA AND CQAS 18

2.5 PROCESS ANALYTICAL TOOL (PAT) IN BIOPROCESS 27

CHAPTER 3 OPTIMIZATION OF MEDIA COMPONENTS TO TARGET CHARGE VARIANT PROFILE **29**

3.1 AIM 30

3.2 INTRODUCTION 30

3.3 MATERIALS AND METHODS 31

3.4 RESULTS AND DISCUSSION 37

3.5 CONCLUSION 50

CHAPTER 4 OPTIMIZATION OF MEDIA COMPONENTS TO TARGET GLYCOSYLATION PROFILE **51**

4.1 AIM 52

4.2 INTRODUCTION 52

4.3 MATERIALS AND METHODS 54

4.4 RESULTS AND DISCUSSION 61

4.5 CONCLUSION 67

CHAPTER 5 APPLICATION OF MACHINE LEARNING FOR CELL CULTURE MEDIA OPTIMIZATION **69**

5.1 AIM 70
5.2 INTRODUCTION 70
5.3 MATERIAL AND METHODS 71
5.4 RESULTS AND DISCUSSION 77
5.5 CONCLUSION 88

CHAPTER 6 NEAR-INFRARED SPECTROSCOPY (NIRS) AS PAT TOOL TO CHARACTERIZE MEDIA 89

6.1 AIM 90
6.2 INTRODUCTION 90
6.3 MATERIALS AND METHODS 92
6.4 RESULTS AND DISCUSSION 97
6.5 CONCLUSION 102

CHAPTER 7 CONCLUSION AND SCOPE OF FUTURE WORK 104

7.1 CONCLUSION 106
7.2 SCOPE OF FUTURE WORK 107

REFERENCES 109

APPENDIX 129

PUBLICATIONS 139

BIO DATA 143

List of Figures

Figure 1 Major steps involved in production of biotherapeutics 11
Figure 2 Steps involved in producing CHO cell lines producing protein of interest 14
Figure 3 Typical cell growth curve in batch mode. 17
Figure 4 Schematic representation of glycoforms. 23
Figure 5 Process monitoring probes/sensors in upstream 27
Figure 6 Definitive screening design analysis of vitamins and metal ions for their impact on titer in CHO cell culture 38
Figure 7 Definitive screening design analysis of effect of vitamins on basic charge variants in CHO cell culture 39
Figure 8 Definitive screening design analysis of effect of vitamins on acidic charge variants in CHO cell culture 40
Figure 9 Definitive screening design analysis of effect of metal ions on basic charge variants in CHO cell culture 41
Figure 10 Definitive screening design analysis of effect of metal ions on acidic charge variants in CHO cell culture 41
Figure 11 Charge variant profile distribution 42
Figure 12 Response surface methodology (CCD) analysis of supplement of Fe and Zn as metal ion supplementation in CHO cell culture 45
Figure 13 Charge variants distribution and culture profile of control and supplement of optimized metal ion concentration 46
Figure 14 SDS-PAGE profile of controlled culture and culture supplemented with optimized concentration of metal ions 47
Figure 15 Glycoforms distribution in controlled culture (Ctrl) and culture supplemented (Treated) with optimized concentration of metal ions (N = 2). 47
Figure 16 Glycosylation profile distribution represented using radar plot of control in Batch mode A) and Fed Batch mode B) was compared with Herclon™ reference product 61
Figure 17 Optimization for glycosylation distribution using definitive screen design (DSD) 63
Figure 18 Glycosylation distribution and culture profile of control culture and culture supplemented 64
Figure 19 Reactive oxygen species (ROS) assay of A) both control and supplemented culture in both batch and fed batch mode on day 6 B) both control and supplemented in fed batch on day 9 of culture. 66
Figure 20 Apoptosis assay of A1) negative control A2) positive control B) cells from day 6 of control culture C) cells from day 9 of culture D) cells from day 6 of supplemented culture and E) cells from day 9 of supplemented culture. 67
Figure 21 Proposed machine learning framework for prediction of Critical Quality Attributes 72
Figure 22 Comparison of charge variant profile of A) Acidic and B) Basic variants with respect to reference molecule 77
Figure 23 Feature ranking for acidic variants 78

Figure 24 Feature selection Basic variants 80
Figure 25 Spearman correlation scatter plots with linear regression 81
Figure 26 Box plots comparing the performance of different machine learning techniques in terms of mean absolute error 84
Figure 27 Prediction with extreme gradient boost regressor 84
Figure 28 Optimized medium cell culture and charge variant profile 86
Figure 29 Raw near-infrared spectra of ten random samples containing varying analyte concentration in CD CHO media 97
Figure 30 Training and Validation loss curve 98
Figure 31 CNN model predicted vs actual (HPLC) concentration value of (a) Uridine (b) Galactose 99
Figure 32 Schematic representation of integrated NIR based CNN model and DoE based statistical model used to predict glycosylation profile of mAbs in CHO cell culture 100
Figure 33 Actual vs predicted glycosylation profile of different media formulations Control B) Media formulation 1 (M1) C) Media formulation 2 (M2) B) Media formulation 3 (M3) 101

List of Tables

Table 1 Various host that are used to produce biologics 11
Table 2 Mammalian host cell lines used for production of biotherapeutic products 13
Table 3 Products produced in CHO cell lines 14
Table 4 CHO cell lines used in biotherapeutics production 15
Table 5 Cell growth stages and associated growth parameter 18
Table 6 Some classical media formulations 20
Table 7 Media components and other culture conditions modulating the quality attributes of mAbs 25
Table 8 Monitoring tools for available for upstream processing 28
Table 9 Distribution of vitamins in different commercially available media 32
Table 10 Distribution of metal ions in different commercially available media 32
Table 11 Analysis of variance (ANOVA) for model predicting acidic variant composition 44
Table 12 Analysis of variance (ANOVA) for model predicting basic variant composition 44
Table 13 Media components used in study for one factor at a time analysis (OFAT). Column LL shows the lower limit and UL shows the upper limit used for the study 56
Table 14 Analysis of variance (ANOVA) for model predicting different glycosylation population indexes (%) composition. 62
Table 15 Charge variants and aggregation profile of mAbs in control and supplemented culture 65
Table 16 Summary of the dataset used in ML study 74
Table 17 Summary of evaluation matrices of Regression model: Mean Absolute Error (MAE), Mean Squared error (MSE), Root mean squared error (RMSE), Coefficient of determination (R^2), Adj-R^2, Mean R^2 after cross-validation, and Standard deviation of R^2. 83
Table 18 Comparison of the proposed technique with state-of-the-art techniques for the detection of galactose and uridine 99
Table 19 Actual vs. predicted galactose and uridine concentration media formulations (M1, M2, M3) (N=2). 100
Table 20 Actual vs Predicted glycoforms in control (ctrl) media and in formulations with media additives (M1, M2, M3) (N=2) 101
Table 21 Charge variant profile and aggregates of control (ctrl) and other media formulations (M1, M2, M3) (N=2). 102

List of abbreviations

ANN	Artificial neural network
BHK	Baby hamster kidney
CD	Chemically defined
CHO	Chinese hamster ovary
CNN	Convolutional neural network
CQAs	Critical quality attributes
DoE	Design of experiment
E. coli	*Escherichia coli*
FcRn	Fc receptor
FDA	US food and drug administration
GBDT	Gradient boost decision tree
HAMAs	Human anti-mouse antibodies
HCP	Host cell protein
ICH	International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use
ICP-MS	Inductively coupled plasma mass spectrometry
IVCC	Integral of viable cell concentration
mAbs	Monoclonal antibodies
ML	Machine learning
MVDA	Multivariate data analysis
NIRS	Near-infrared spectroscopy
OFAT	One-factor-at-a-time
PAT	Process analytical technology
PoI	Protein of interest
PTMs	Post-translational modifications
QbD	Quality by design

Symbols

qp	Specific productivity of mAbs	[$gcell^{-1}day^{-1}$]
μ	Specific growth rate	[h^{-1}]