

**NEXT GENERATION PLATFORMS FOR  
ANALYTICAL CHARACTERIZATION OF  
BIOSIMILARS**

**ANUJ SHRIVASTAVA**



**DEPARTMENT OF CHEMICAL ENGINEERING  
INDIAN INSTITUTE OF TECHNOLOGY DELHI**

**MAY 2025**

© Indian Institute of Technology New Delhi 2025

**NEXT GENERATION PLATFORMS FOR  
ANALYTICAL CHARACTERIZATION OF  
BIOSIMILARS**

by

**ANUJ SHRIVASTAVA**

Submitted

In fulfilment of the requirements of the degree of Doctor of  
Philosophy

To



**Department Of Chemical Engineering**

**Indian Institute of Technology Delhi**

**May 2025**

*Dedicated to my family...*

## CERTIFICATE

This is to certify that the thesis entitled “**NEXT GENERATION PLATFORMS FOR ANALYTICAL CHARACTERIZATION OF BIOSIMILARS**”, being submitted by **ANUJ SHRIVASTAVA** 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 our 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.

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

**Prof. Anurag S. Rathore**

Department of Chemical Engineering

Indian Institute of Technology Delhi

## Acknowledgements

First and foremost, I would like to express my deepest gratitude and reverence to my PhD supervisor, Prof. Anurag S. Rathore, for his constant support and guidance throughout the course of this work and for providing me with intellectual freedom and independence during my research work. I feel indebted to him for his constructive feedback, criticism, and moral support, which helped me carry my work in the right direction. In the realm of scientific collaboration, I would also like to express my gratitude to Prof. Andras Guttman, University of Debrecen, Hungary, for providing me with insightful feedback, especially on manuscript drafting.

I would like to thank my PhD research committee members, Prof. Sudip K. Pattanayek, Prof. James Gomes, and Prof. Manoj Ramteke, whose periodic assessment and review of my work helped me to go forward in the right direction. I am also grateful to Department of Chemical Engineering, IIT Delhi for providing me a suitable environment for conducting the research and knowledge on various chemical engineering subjects, which were new to me during my initial stages of PhD.

Words are insufficient to thank my friends, Dr. Akshit, Dr. Shantanu, Dr. Rajat, and Dr. Himanshu, for their fruitful and insightful discussions in both professional and personal matters. My special thanks to my seniors, Dr. Shyamapada, Dr. Nikita, Dr. Jaya, Dr. Sunil, Dr. Sharvan, Dr. Sanghati, Dr. Avinash, Dr. Babji, Dr. Ramesh, Dr. Shristi, Dr. Keerthiveena and Dr. Nupur, and Dr. Rozaleen, who provided me crucial guidance and advice during chemometric method development and troubleshooting during experimental runs.

I thank my colleagues and juniors Dr. Neelesh, Dr. Rucha, Naveen, Deepika, Kunal, Sanjeet, and Dr. Nitika, and the whole Centre of Excellence – Centre of Biotechnology (COE-CBT) team members with whom I worked on several projects and had invaluable discussions. I have no words to express my thanks to my lab staff members Mohit ji, Nitin ji, Naresh ji, Ranjeet ji and Surender ji for their immense co-operation in the administrative tasks.

I would also like to express my heartfelt gratitude to my friends Dr. Anand, Dr. Vivek, Akhilesh, Dr. Prateek, Dr. Ketan, Dr. Faisal, Dr. Priyank, and Dr. Ankit for their unwavering support and for providing me with wonderful companionship throughout my PhD journey. Their constant encouragement and shared moments of joy and learning have made this journey truly memorable.

Finally, I am deeply indebted to my parents, Mr. Brahma Swaroop Shrivastava, and Mrs. Manorama Shrivastava, my elder brother Ankit Shrivastava and my younger brother Aman Shrivastava for giving me unconditional love, support and faith to my ambitions and their constant encouragement to pursue higher studies. I would also like to extend my heartfelt thanks to my sisters-in-law, Shalu Shrivastava and Deepika Shrivastava, for their kindness and support, which have been a great source of strength during this time.

I would like to express my heartfelt gratitude to my wife, Abhilasha, whose unwavering love, support, and understanding have been a constant source of strength throughout this journey. Her encouragement has helped me navigate the highs and lows with resilience.

I am equally grateful to my in-laws for their steadfast support and encouragement, which has been invaluable during this time. Their belief in me has made this journey all the more meaningful.

Finally, I dedicate this work to the Almighty, whose blessings, grace, and guidance have been my eternal source of strength and inspiration throughout this scientific adventure. I would also like to end this section of thesis by acknowledging again the privilege that I had at IIT Delhi for understanding this research. I hope my work contributes meaningfully to the field of therapeutic proteins.

ANUJ SHRIVASTAVA

## Abstract

Monoclonal antibodies (mAbs) are vital biotherapeutics widely used in treating various diseases. Their therapeutic efficacy and stability are significantly influenced by glycosylation patterns and aggregation behaviour. The expiration of patents for innovator mAbs has accelerated the development of biosimilars, driving the need for rapid, cost-effective analytical techniques. Traditional methods for glycosylation and aggregation analysis are resource-intensive and time-consuming, prompting the demand for development of next generation platforms. This thesis work involved rapid quantification and characterization of N-glycans and aggregates by integrating machine learning (ML) algorithms and automated tools with advanced analytical methods. Additionally, it involves the investigation of mAb stability under various buffer systems and storage conditions, providing insights to optimize manufacturing workflows and ensure therapeutic reliability.

The resource intensive nature of Liquid chromatography-mass spectrometry (LC-MS) for identification and quantification of N-glycans, necessitates the development of rapid and cost-effective orthogonal analysis approaches. First objective (Chapter 3) was to develop an online method utilizing the Extreme Gradient Boosting (XGBoost) - ML algorithm for real time quantification of InstantPC labelled N-glycans by LC – fluorescence detector (FLD). The LC-FLD profile was pre-processed for baseline correction and noise reduction prior to being fed to the ML algorithm. The algorithm had been successfully tested for commercial and inhouse developed mAbs and validated using LC-MS quantification as reference. The LC-FLD-ML model predicted values were at par with the LC-MS values with root mean square error of <0.5 and R2 of >0.95. The average error using ML model (1.80 %) were reduced by a minimum of 28 % and 40 % for origin (1.5 %) and manual (1.07 %) based integration, respectively. The approach reduces the data analysis time per sample by ~70 % (from ~5 min to ~1.5 min), thereby offering a time and resource efficient orthogonality with LC-MS for quantification of N-glycans in mAbs.

Glycosylation, a critical quality attribute, makes glycosimilarity assessment pivotal for biosimilar development. The next study (Chapter 4) proposed a Python-based automated tool for rapid estimation of the Glycosimilarity Index (GI). A comprehensive analytical glycosimilarity comparison of the trastuzumab originator product, Herclon (Roche), with five marketed biosimilars: Trasturel (Reliance Life Sciences), Canmab (Biocon), Vivitra (Zydus Ingenia), Hertraz (Mylan), and Biceltis (Cipla) has been performed. Similarly, a comparison

between the bevacizumab originator product, Avastin (Roche), and its five biosimilars: Abevmy (Mylan), Krabeva (Biocon), Ivzumab (RPG LifeSciences), Bryxta (Zydus), and Advamab (Alkem Labs) was also presented. Glycosimilarity had been assessed using liquid chromatography-fluorescence detection with data integrated using the XGBoost-ML algorithm. The GI had been calculated by combining profile similarity and compositional similarity, estimated based on the criticality and tolerance of each glycan. The tool enabled rapid GI estimation (<1 min/sample) with reduced errors compared to Excel (> 10 min/sample). Biosimilars exhibited high GI (> 95%), with the lowest GI observed at 87.80% for trastuzumab and 92.39% for bevacizumab. The Python-based tool offers a high throughput and a reliable platform for glycosimilarity assessment, outperforming Excel methods. Minor variations in glycosylation patterns were observed among the biosimilars, suggesting a modest glycosimilarity variation (GI range between 80-100%).

In this next study (Chapter 5), presented a novel dynamic light scattering (DLS)-based approach that allowed to quantify the relative percentage of multimers (monomer, dimer, trimer, and tetramer) in a mAb therapeutic product. The proposed approach used a ML algorithm and regression to model the system and predicted the amount of relevant species such as monomer, dimer, trimer, and tetramer of a mAb. The proposed DLS-ML technique compared favourably to all potential alternatives with respect to the key method attributes, including per sample cost of analysis, per sample time of data acquisition along with ML-based aggregate prediction (<2 min), sample requirements (<3 µg), and user-friendliness of analysis. The proposed rapid method can serve as an orthogonal tool to size exclusion chromatography, which is the current industry workhorse for aggregate assessment.

In the fourth study (Chapter 6), the stability and kinetics of degradation when mAb samples are stored in commonly used Protein A elution buffers, including citrate, acetate, and glycine, at varying pre-existing aggregates levels (low: 1-5%, moderate: 5-15% and high: 15-25%) at 4°C and 30°C to simulate standard and worst-case conditions. mAb samples were subjected to thermal stress to achieve different levels of initial aggregates. These pre-aggregated samples were then incubated in different buffers at 4°C and 30°C to assess aggregation rates and stability. Aggregates were quantified using DLS. At 30°C, half-life reductions for citrate, acetate, and glycine buffers were 6.30-fold, 6.48-fold, and 9.64-fold, respectively, compared to 4°C, with glycine buffer offering the most protection against aggregation and citrate buffer the least. At higher initial aggregate levels, half-lives decreased by 2.15-, 1.95-, and 1.73-fold for citrate, acetate, and glycine buffers, respectively, compared to lower initial aggregates. This

shows that with the increase in initial aggregation levels, the aggregation rate increases. Further, while second-order kinetics was primarily observed with samples having lower initial aggregate levels, first-order kinetics was more common in samples with medium and high initial aggregate levels. Amongst all the conditions that were explored in this study, glycine buffer at 4°C with 1–5% initial aggregates achieved the highest half-life of 129 days. In contrast, at 30°C, citrate buffer with high initial aggregate levels exhibited the lowest stability, with a half-life of just 3.5 days. The findings highlight the significance of using optimal buffer systems and storage conditions for in-process intermediates during mAb manufacturing so as to have a robust process and safe and efficacious biotherapeutic product.

## सारांश

मोनोक्लोनल एंटीबॉडी (mAbs) महत्वपूर्ण बायोथेरेप्यूटिक्स हैं जिनका व्यापक रूप से विभिन्न रोगों के उपचार में उपयोग किया जाता है। उनकी चिकित्सीय प्रभावकारिता और स्थिरता ग्लाइकोसिलेशन पैटर्न और एकत्रीकरण व्यवहार से काफी प्रभावित होती है। इनोवेटर mAbs के लिए पेटेंट की समाप्ति ने बायोसिमिलर के विकास को गति दी है, जिससे तेज़, लागत प्रभावी विश्लेषणात्मक तकनीकों की आवश्यकता बढ़ गई है। ग्लाइकोसिलेशन और एकत्रीकरण विश्लेषण के लिए पारंपरिक तरीके संसाधन-गहन और समय लेने वाले हैं, जिससे अगली पीढ़ी के प्लेटफ़ॉर्म के विकास की मांग बढ़ रही है। इस थीसिस कार्य में मशीन लर्निंग (ML) एल्गोरिदम और स्वचालित उपकरणों को उन्नत विश्लेषणात्मक विधियों के साथ एकीकृत करके N-ग्लाइकन और समुच्चय का तेजी से परिमाणीकरण और लक्षण वर्णन शामिल था। इसके अतिरिक्त, इसमें विभिन्न बफर सिस्टम और भंडारण स्थितियों के तहत mAb स्थिरता की जांच शामिल है, जो विनिर्माण वर्कफ़्लो को अनुकूलित करने और चिकित्सीय विश्वसनीयता सुनिश्चित करने के लिए अंतर्दृष्टि प्रदान करता है। एन-ग्लाइकन की पहचान और मात्रा का पता लगाने के लिए लिक्विड क्रोमैटोग्राफी-मास स्पेक्ट्रोमेट्री (एलसी-एमएस) की संसाधन गहन प्रकृति, त्वरित और लागत प्रभावी ऑर्थोगोनल विश्लेषण दृष्टिकोणों के विकास को आवश्यक बनाती है। पहला उद्देश्य (अध्याय 3) एलसी-फ्लोरोसेंस डिटेक्टर (एफएलडी) द्वारा इंस्टेंटपीसी लेबल वाले एन-ग्लाइकन की वास्तविक समय मात्रा का पता लगाने के लिए एक्सट्रीम ग्रेडिंट बूस्टिंग (एक्सजीबूस्ट) - एमएल एल्गोरिदम का उपयोग करके एक ऑनलाइन विधि विकसित करना था। एलसी-एफएलडी प्रोफाइल को एमएल एल्गोरिदम में फीड करने से पहले बेसलाइन सुधार और शोर में कमी के लिए प्री-प्रोसेस किया गया था। एल्गोरिदम का व्यावसायिक और इनहाउस विकसित mAbs के लिए सफलतापूर्वक परीक्षण किया गया था और संदर्भ के रूप में एलसी-एमएस मात्रा का उपयोग करके मान्य किया गया था। एलसी-एफएलडी-एमएल मॉडल की भविष्यवाणी के मान एलसी-एमएस मानों के बराबर थे, जिसमें रूट माध्य वर्ग त्रुटि  $<0.5$  और  $R^2 >0.95$  थी। एमएल मॉडल (1.80%) का उपयोग करने वाली औसत त्रुटि क्रमशः मूल (1.5%) और मैनुअल (1.07%) आधारित एकीकरण के लिए न्यूनतम 28% और 40% तक कम हो गई। यह दृष्टिकोण प्रति नमूने डेटा विश्लेषण समय को ~70% (~5 मिनट से ~1.5 मिनट तक) कम करता है, जिससे mAbs में N-ग्लाइकन की मात्रा का पता लगाने के लिए LC-MS के साथ समय और संसाधन कुशल ऑर्थोगोनैलिटी की पेशकश की जाती है। ग्लाइकोसिलेशन, एक महत्वपूर्ण गुणवत्ता विशेषता, ग्लाइकोसिमिलैरिटी मूल्यांकन को बायोसिमिलर विकास के लिए महत्वपूर्ण बनाती है। अगले अध्ययन (अध्याय 4) ने ग्लाइकोसिमिलैरिटी इंडेक्स (जीआई) के तेजी से अनुमान के लिए पायथन-आधारित

स्वचालित उपकरण का प्रस्ताव दिया। ट्रेस्टुजुमाब मूल उत्पाद, हर्कलॉन (रोश) की पांच बाजार में बिकने वाली बायोसिमिलर: ट्रेस्टुरेल (रिलायंस लाइफ साइंसेज), कैनमैब (बायोकॉन), विविट्रा (ज़ाइडस इंजेनिया), हर्ट्राज (माइलन) और बिसेल्टिस (सिप्ला) के साथ एक व्यापक विश्लेषणात्मक ग्लाइकोसिमिलरिटी तुलना की गई है। इसी तरह, बेवाकिजुमैब मूल उत्पाद, एवास्टिन (रोश) और इसके पांच बायोसिमिलर: एबेवीमी (माइलन), क्रैबेवा (बायोकॉन), इवजुमाब (आरपीजी लाइफसाइंसेज), ब्रायक्स्टा (ज़ाइडस) और एडवामैब (एल्केम लैब्स) के बीच तुलना भी प्रस्तुत की गई। ग्लाइकोसिमिलरिटी का मूल्यांकन लिक्विड क्रोमैटोग्राफी-फ्लोरोसेंस डिटेक्शन का उपयोग करके किया गया था, जिसमें एक्सजीबूस्ट-एमएल एल्गोरिदम का उपयोग करके डेटा एकीकृत किया गया था। जीआई की गणना प्रोफ़ाइल समानता और संरचनागत समानता को मिलाकर की गई थी, जिसका अनुमान प्रत्येक ग्लाइकेन की गंभीरता और सहनशीलता के आधार पर लगाया गया था। इस उपकरण ने एक्सेल (> 10 मिनट/नमूना) की तुलना में कम त्रुटियों के साथ तेजी से जीआई आकलन (<1 मिनट/नमूना) सक्षम किया। बायोसिमिलर ने उच्च जीआई (> 95%) प्रदर्शित किया, जिसमें सबसे कम जीआई ट्रेस्टुजुमाब के लिए 87.80% और बेवाकिजुमैब के लिए 92.39% देखा गया। पायथन-आधारित उपकरण एक्सेल विधियों से बेहतर प्रदर्शन करते हुए ग्लाइकोसिमिलरिटी आकलन के लिए एक उच्च थ्रूपुट और एक विश्वसनीय प्लेटफ़ॉर्म प्रदान करता है। बायोसिमिलर के बीच ग्लाइकोसिलेशन पैटर्न में मामूली बदलाव देखे गए, जो एक मामूली ग्लाइकोसिमिलरिटी भिन्नता (80-100% के बीच जीआई रेंज) का सुझाव देते हैं।

इस अगले अध्ययन (अध्याय 5) में, एक नया गतिशील प्रकाश बिखराव (डीएलएस)-आधारित दृष्टिकोण प्रस्तुत किया गया, जिसने एमएबी चिकित्सीय उत्पाद में मल्टीमर्स (मोनोमर, डिमर, ट्रिमर और टेट्रामर) के सापेक्ष प्रतिशत को निर्धारित करने की अनुमति दी। प्रस्तावित दृष्टिकोण ने सिस्टम को मॉडल करने के लिए एक एमएल एल्गोरिदम और प्रतिगमन का उपयोग किया और एक mAb के मोनोमर, डिमर, ट्रिमर और टेट्रामर जैसी प्रासंगिक प्रजातियों की मात्रा का अनुमान लगाया। प्रस्तावित डीएलएस-एमएल तकनीक ने मुख्य विधि विशेषताओं के संबंध में सभी संभावित विकल्पों की तुलना में अनुकूल प्रदर्शन किया, जिसमें प्रति नमूना विश्लेषण की लागत, प्रति नमूना डेटा अधिग्रहण का समय और साथ ही एमएल-आधारित समग्र भविष्यवाणी (<2 मिनट), नमूना आवश्यकताएं (<3 µg), और विश्लेषण की उपयोगकर्ता-मित्रता शामिल है। प्रस्तावित तीव्र विधि आकार बहिष्करण क्रोमैटोग्राफी के लिए एक ऑर्थोगोनल उपकरण के रूप में काम कर सकती है, जो समग्र मूल्यांकन के लिए वर्तमान उद्योग का काम है।

चौथे अध्ययन (अध्याय 6) में, जब mAb नमूनों को सामान्य रूप से उपयोग किए जाने वाले प्रोटीन ए इल्यूशन बफर्स में संग्रहीत किया जाता है, जिसमें साइट्रेट, एसीटेट और ग्लाइसिन शामिल हैं, तो गिरावट की स्थिरता और गतिकी, अलग-अलग पूर्व-मौजूदा समुच्चय स्तरों (निम्न: 1-5%, मध्यम: 5-15% और उच्च: 15-25%) पर 4°C और 30°C पर मानक और सबसे खराब स्थिति की स्थिति का अनुकरण करने के लिए। mAb नमूनों को प्रारंभिक समुच्चय के विभिन्न स्तरों को प्राप्त करने के लिए थर्मल तनाव के अधीन किया गया था। इन पूर्व-एकत्रित नमूनों को एकत्रीकरण दरों और स्थिरता का आकलन करने के लिए 4°C और 30°C पर विभिन्न बफर्स में इनक्यूबेट किया गया था। समुच्चयों को गतिशील प्रकाश बिखराव (DLS) का उपयोग करके मात्राबद्ध किया गया था। 30 डिग्री सेल्सियस पर, साइट्रेट, एसीटेट और ग्लाइसिन बफर्स के लिए अर्ध-आयु में कमी क्रमशः 6.30 गुना, 6.48 गुना और 9.64 गुना थी, जबकि 4 डिग्री सेल्सियस पर यह कमी थी, जिसमें ग्लाइसिन बफर एकत्रीकरण के खिलाफ सबसे अधिक सुरक्षा प्रदान करता है और साइट्रेट बफर सबसे कम सुरक्षा प्रदान करता है। उच्च प्रारंभिक समुच्चय स्तरों पर, कम प्रारंभिक समुच्चय की तुलना में साइट्रेट, एसीटेट और ग्लाइसिन बफर्स के लिए अर्ध-आयु क्रमशः 2.15-, 1.95- और 1.73 गुना कम हो गई। यह दर्शाता है कि प्रारंभिक एकत्रीकरण स्तरों में वृद्धि के साथ, एकत्रीकरण दर बढ़ जाती है। इसके अलावा, जबकि द्वितीय-क्रम गतिकी मुख्य रूप से कम प्रारंभिक समुच्चय स्तर वाले नमूनों के साथ देखी गई थी, मध्यम और उच्च प्रारंभिक समुच्चय स्तर वाले नमूनों में प्रथम-क्रम गतिकी अधिक आम थी। इस अध्ययन में जिन सभी स्थितियों का पता लगाया गया, उनमें से 1-5% प्रारंभिक समुच्चय के साथ 4 डिग्री सेल्सियस पर ग्लाइसिन बफर ने 129 दिनों का उच्चतम अर्ध-जीवन प्राप्त किया। इसके विपरीत, 30 डिग्री सेल्सियस पर, उच्च प्रारंभिक समुच्चय स्तरों वाले साइट्रेट बफर ने सबसे कम स्थिरता प्रदर्शित की, जिसका अर्ध-जीवन केवल 3.5 दिन था। निष्कर्ष mAb निर्माण के दौरान इन-प्रोसेस इंटरमीडिएट के लिए इष्टतम बफर सिस्टम और भंडारण स्थिति का उपयोग करने के महत्व को उजागर करते हैं ताकि एक मजबूत प्रक्रिया और सुरक्षित और प्रभावोत्पादक बायोथेरेप्यूटिक उत्पाद प्राप्त किया जा सके।

## Table of Contents

|  |              |
|--|--------------|
| <b>CERTIFICATE</b> .....   | <b>i</b>     |
| <b>Acknowledgements</b> .....  | <b>ii</b>    |
| <b>Abstract</b> .....  | <b>iv</b>    |
| <b>सारांश</b> .....  | <b>vii</b>   |
| <b>List of Figures</b> .....   | <b>xiv</b>   |
| <b>List of Tables</b> .....  | <b>xviii</b> |
| <b>List of Abbreviations</b> .....   | <b>xix</b>   |
| <b>CHAPTER 1 INTRODUCTION</b> .....  | <b>1</b>     |
| 1.1 Background .....   | 3            |
| 1.2 N-glycosylation.....   | 5            |
| 1.2.1 Sample preparation .....   | 7            |
| 1.2.2 Data acquisition .....   | 8            |
| 1.2.3 Data analysis .....  | 8            |
| 1.3 Aggregation of mAb .....   | 9            |
| 1.3.1 Different size range of mAb and their characterization techniques.....       | 9            |
| 1.3.2 Factors and mechanism behind aggregates formation .....                      | 10           |
| 1.3.3 Stability assessment approaches for mAbs .....                               | 11           |
| 1.4 Purpose of the current study .....   | 12           |
| <b>CHAPTER 2 LITERATURE REVIEW</b> .....   | <b>13</b>    |
| 2.1 Glycosylation .....  | 15           |
| 2.1.1 High throughput approaches in sample preparation and data acquisition: ..... | 15           |
| 2.1.2 Progress in data analysis and interpretation software .....                  | 17           |
| 2.1.3 Advancements in chemometric approaches for assessing biosimilarity: .....    | 18           |
| 2.2 Aggregation.....   | 20           |
| 2.2.1 Techniques for size and aggregation analysis .....                           | 21           |
| 2.2.2 Spectroscopic methods for structural and stability assessment .....          | 23           |

|  |           |
|--|-----------|
| 2.2.3 Two-dimensional (2D) settings .....  | 24        |
| 2.2.4 Insights into mAb aggregation.....   | 26        |
| 2.3 Summary of the literature review .....   | 29        |
| <b>CHAPTER 3 MACHINE LEARNING TOOL FOR RAPID QUANTIFICATION OF<br/>N-GLYCANS USING FLUORESCENCE DETECTION.....</b>       | <b>31</b> |
| 3.1 Introduction.....  | 33        |
| 3.2 Material and Methods .....   | 36        |
| 3.2.1 Materials .....  | 36        |
| 3.2.2 Sample preparation .....   | 36        |
| 3.2.3 Data acquisition .....   | 37        |
| 3.2.4 Data analysis .....  | 38        |
| 3.3 Results.....   | 41        |
| 3.3.1 Observation using InstantPC kit-based sample preparation .....   | 42        |
| 3.3.2 Observation using InstantPC kit-based data acquisition and analysis .....  | 42        |
| 3.3.3 Data analysis for FLD chromatogram: .....  | 46        |
| 3.4 Discussion.....  | 47        |
| 3.5 Conclusions.....   | 52        |
| <b>CHAPTER 4 AN AUTOMATED TOOL FOR GLYCOSIMILARITY ASSESSMENT:<br/>TRASTUZUMAB AND BEVACIZUMAB AS CASE STUDIES .....</b> | <b>53</b> |
| 4.1 Introduction.....  | 55        |
| 4.2 Materials and Methods.....   | 58        |
| 4.2.1 Samples .....  | 58        |
| 4.2.2 Sample preparation .....   | 58        |
| 4.2.3 Data acquisition and analysis.....   | 59        |
| 4.2.4 Determination of Glycosimilarity Index.....  | 60        |
| 4.3 Results and Discussion .....   | 64        |
| 4.3.1 Glycan profiling .....   | 64        |
| 4.3.2 LC-FLD composition.....  | 66        |

|   |           |
|---|-----------|
| 4.3.3 Glycosimilarity estimation.....   | 67        |
| 4.3.4 Automated vs Excel-based GI calculator – time, accuracy.....  | 70        |
| 4.4 Conclusions.....  | 71        |
| <b>CHAPTER 5 RAPID ESTIMATION OF SIZE-BASED HETEROGENEITY BY<br/>MACHINE LEARNING ENHANCED DYNAMIC LIGHT SCATTERING .....</b> | <b>73</b> |
| 5.1 Introduction.....   | 75        |
| 5.2 Materials and Methods.....  | 78        |
| 5.2.1 Feed material.....  | 78        |
| 5.2.2 Generation of mAb aggregates .....  | 78        |
| 5.2.3 Size Exclusion Chromatography (SEC).....  | 80        |
| 5.2.4 Dynamic Light Scattering (DLS).....   | 80        |
| 5.2.5 Machine learning algorithms .....   | 80        |
| 5.3 Result and Discussion .....   | 82        |
| 5.3.1 Calibration curve.....  | 82        |
| 5.3.2 Peak deconvolution approach .....   | 83        |
| 5.3.3 Machine learning-based analysis of DLS spectra.....   | 84        |
| 5.4 Conclusion .....  | 90        |
| <b>CHAPTER 6 IN-PROCESS STABILITY: IMPACT OF BUFFERS, INITIAL-<br/>AGGREGATES AND STORAGE TEMPERATURE .....</b>               | <b>91</b> |
| 6.1 Introduction.....   | 93        |
| 6.2 Material and Methods .....  | 95        |
| 6.2.1 Feed Material .....   | 95        |
| 6.2.2 Reagents.....   | 95        |
| 6.2.3 Sample preparation .....  | 96        |
| 6.3 Results.....  | 98        |
| 6.3.1 Quantification of aggregates .....  | 98        |
| 6.3.2 Stability studies at lower initial aggregation level.....   | 99        |
| 6.3.3 Stability studies at medium initial aggregation level.....  | 101       |

|   |            |
|---|------------|
| 6.3.4 Stability studies at high initial aggregation level ..... | 102        |
| 6.4 Discussion .....  | 103        |
| 6.4.1 Effect of buffers .....                                   | 104        |
| 6.4.2 Effect of initial level of aggregates .....               | 104        |
| 6.4.3 Effect of storage temperatures .....                      | 105        |
| 6.4.4 Aggregation kinetics and mechanism .....                  | 106        |
| 6.5 Conclusions.....  | 107        |
| <b>CHAPTER 7 CONCLUSIONS AND SCOPE OF FUTURE WORK.....</b>      | <b>109</b> |
| 7.1 Conclusions.....  | 111        |
| 7.2 Scope of future work.....                                   | 113        |
| <b>REFERENCES.....</b>  | <b>115</b> |
| <b>List of Publications .....</b>                               | <b>136</b> |
| <b>Biodata.....</b>   | <b>137</b> |

## List of Figures

|  |    |
|--|----|
| Figure 1-1: Different structural arrangements of monoclonal antibody N-glycans. A: Intact glycoprotein as well as B: High mannose - , C: Hybrid - , and D: Complex – Asparagine linked oligosaccharide structures. ....  | 5  |
| Figure 1-2: Different stages of the N-glycan analysis workflow and the steps of analysis involved at each stage. The steps involved are automated sample preparation, rapid data acquisition, and fast and automated data analysis. ....   | 7  |
| Figure 2-1: Comparison of the A. Traditional 2-AB dye protocol with the recently developed B. Rapid commercial (Instant-PC) dye protocol along with the required sample preparation steps.....   | 16 |
| Figure 3-1: Sample preparation using Instant-PC labelling kit. Steps involve denaturation, then de-glycosylation followed by rapid labelling, then cleanup using HILIC-SPE technique to get purified labelled glycan.....  | 37 |
| Figure 3-2: Flow diagram illustrating the fundamental structure of the XGBoost algorithm used for spectral analysis in this study. ....  | 40 |
| Figure 3-3: Learning rate vs. RMSE curve for XGBoost model. Learning rate directly affects the model robustness. Best model performance was observed at 0.5 with RMSE of 2.0984..  | 42 |
| Figure 3-4: Comparison of chromatograms obtained with Instant-PC labelled glycans released from commercial trastuzumab sample. A) TIC observed from LC-MS (top) B) Chromatogram observed from LC-FLD (bottom). The glycan pattern of major abundant glycans observed in the LC-FLD chromatogram (bottom) exhibits a comparable retention time difference which has been verified by LC-MS profile (top). One individual peaks, namely FA2G2, and four peak pairs of overlapped peaks for a pair of coeluting glycans of FA1 and A2, FA2 and M5, A2[6]G1 and FA1G1, and isomeric glycans of FA2[6]G1 and FA2[3]G1 were observed in both LC-FLD chromatogram and LC-MS. .... | 44 |
| Figure 3-5: Absolute error between iterative steps for XGBoost model .....   | 46 |
| Figure 3-6: Predicted vs. actual values of glycan using XGBoost algorithm. The root mean square error (RMSE) values that quantifies the average deviation between our predicted values and the observed values for each sample are (A) 0.0587; (B) 0.3270; (C) 1.7551; (D) 0.4739; (E) 0.684976; (F) 1.0409. Here predicted values were quantified by implementing XGBoost algorithm on LC-FLD chromatogram and actual values were quantified using LC-MS analysis. ....   | 48 |

Figure 3-7: Error between the glycan quantification using manual integration, origin-based deconvolution and model-based quantification, from the base case i.e. quantification using LC-MS for 3 different samples. It is visible that the model-based quantification exhibits less error for all cases. ....50

Figure 4-1: Flow diagram illustrating the algorithm developed for tolerance limit estimation. tr: list of retention time points for all glycans; t: Iterating position index; TolFac\_list: List of tolerance factor for each time point of glycan (TolFac); k: position index for iterating over the list of tolerance value; Xmean[t]: Mean of Innovator batches; Std Dev: List of Standard deviations; CV: Empty list of Critical value; UL: Empty list representing upper tolerance limit; LL: Empty list representing lower tolerance limit. ....62

Figure 4-2: Flow diagram illustrating the automated protocol used to estimate the profile similarity score used to estimate the glycosimilarity index. Qin: %points inside tolerance limit; Qout: %points outside tolerance limit; SSQ: Sum of square; Xs: Biosimilar intensity falling outside the tolerance limit; Xmean: Mean of intensities for innovator batches. ....63

Figure 4-3: LC-FLD profile for the innovator sample and its biosimilars: A. Trastuzumab and B. Bevacizumab. Tlavg and Blavg represent the mean glycan composition for trastuzumab and bevacizumab innovators, respectively. TA1 and BA1 correspond to trastuzumab and bevacizumab biosimilar A, batch 1. The same naming sequence is followed by other biosimilars up to TE2 and BE2 (Biosimilar E, batch 2). ....64

Figure 4-4: Compositional quantification of the released glycans analyzed by LC-FLD. A. Trastuzumab (top) B. Bevacizumab (bottom). All the compositions (in percentage area) are quantified using the Xgboost algorithm. Minor variability is observed in all the biosimilars compared to innovators for both the mAbs. Tlavg and Blavg represent the mean glycan composition for trastuzumab and bevacizumab innovators, respectively. TA1 and BA1 correspond to trastuzumab and bevacizumab biosimilar A, batch 1. The same naming sequence is followed for other biosimilars up to TE2 and BE2 (Biosimilar E, batch 2). ....66

Figure 4-5: Representation of profile similarity assessment concept. Illustration of profile similarity assessment, showing upper and lower tolerance limits (dashed orange and red lines, resp.) with a focus on the A2 glycan variant. The calculation of the tolerance limit was done based on the inconsistencies observed across different batches of innovator. Mean and standard deviation across the retention time range define these limits, accounting for batch variability. Biosimilar similarity is assessed by evaluating whether the data falls within the limits and its deviation from the mean. Tlavg: Mean profile for trastuzumab innovator product; UL: Upper limit; LL: Lower limit; TA2: Trastuzumab, Biosimilar A, and batch 2; TD1: Trastuzumab,

|   |    |
|---|----|
| Biosimilar D and batch 1. Black circle: Point within the tolerance limit; Triangular black point: Point outside the tolerance limit. ....   | 68 |
| Figure 4-6: Glycosimilarity Index values for the biosimilars of trastuzumab and bevacizumab. TA and BA correspond to trastuzumab and bevacizumab (Biosimilar A). The same naming sequence is followed for other biosimilars up to TE and BE (Biosimilar E).....   | 70 |
| Figure 5-1: Size range for different protein aggregates and their characterization techniques. The red highlighted box represents the common size range (10 nm-100 nm) in which both size exclusion chromatography and dynamic light scattering work well. ....   | 76 |
| Figure 5-2: SEC chromatographic separations and DLS-PSD of standard proteins and their calibration curves. A) Overlap of chromatograms for five standard marker proteins obtained by SEC, B) Semi-logarithmic calibration curve prepared by plotting the molecular weight v/s normalized elution volume used for oligomer distribution analysis; C) DLS-PSD for five standard marker proteins; D) Plot of mean hydrodynamic diameter v/s molecular weights for the five standard proteins. .... | 81 |
| Figure 5-3: SEC Chromatogram and DLS-PSD for oligomer distribution analysis. A) Elution times of observed oligomers determined by SEC, B) DLS-PSD intensity curves for two samples with the different aggregate compositions containing monomer, dimer, trimer, and tetramer. ....  | 83 |
| Figure 5-4: Predicted % aggregate v/s experimental data for mAb1 test and validation samples. A) Predicted % aggregate using neural network algorithm for test data, B) Predicted % aggregate using support vector regression area correlation graph for test data, C) Predicted % aggregate using neural network algorithm for validation data, D) Predicted % aggregate using support vector regression correlation graph for validation data. ....   | 85 |
| Figure 5-5: Predicted % aggregate v/s experimental data for mAb2 test and validation samples. a) Predicted % aggregate using NN algorithm for test data, b) Predicted % aggregate using SVR area correlation graph for test data, c) Predicted % aggregate using NN algorithm for validation data, d) Predicted % aggregate using SVR correlation graph for validation data. .  | 86 |
| Figure 6-1: Monomer and aggregates quantification using DLS. a) Representation of DLS-PSD intensity curve for a pre-aggregated mAbs sample, b) Quantified compositions of monomer, dimer, trimer, and tetramer using the previously published DLS-ML model (monomer: 79.83%, dimer: 10.45%, trimer: 7.16%, tetramer: 2.56%, aggregate: 20.17%). ....  | 98 |
| Figure 6-2: Impact of various buffers on the degradation of lower level of pre-aggregated mAbs. (a) and (d) Degradation of low-preaggregated mAb (b) and (e) Regression analysis comprising of $1/C_M$ vs. time plot, as the best fit is second order kinetics; (c) and (d) Plots of  |    |

rate constant (k) bar chart and half-life ( $t_{1/2}$ ) line plot for storage at 4°C and 30°C, respectively. .... 100

Figure 6-3: Impact of various buffers on the degradation of medium level of pre-aggregated mAbs. (a) and (d) Degradation of medium pre-aggregated mAbs; (b) and (e) Regression analysis comprising of  $-\ln(C_M/C_{M0})$  vs. time plot, as the best fit is first-order kinetics; (c) and (f) Plot of rate constant (k) bar chart and half-life ( $t_{1/2}$ ) line plot for storage at 4°C and 30°C, respectively. .... 102

Figure 6-4: Impact of various buffers on the degradation of higher level of pre-aggregated mAbs. (a) and (d) Degradation of high- pre-aggregated mAb; (b) and (e) Regression analysis comprising of  $-\ln(C_M/C_{M0})$  vs. time plot, as the best fit is first-order kinetics; (c) and (f) Plot of rate constant (k) bar chart and half-life ( $t_{1/2}$ ) line plot for storage at 4°C and 30°C, respectively. .... 103

Figure 6-5: Graphical representation of fold changes in aggregation half-life under different conditions. The plot offers a comparison of fold reductions due to storage temperatures and initial aggregate levels across citrate, acetate and glycine buffers. Dark blue and orange colour bars represent storage at 30°C vs 4°C, at low (1-5%) and high (15-25%) initial aggregates level respectively; Green and light blue colour bars represent storage at high (15-25%) vs. low (1-5%) initial aggregates level at 30°C and 4°C respectively. .... 105

Figure 6-6: 3D colour surface plots showing the interplay of buffer composition, initial aggregate level, and storage temperature on mAb half-life. Higher half-life values are observed at 4°C (upper plot), with a maximum of 129 days in glycine buffer at low (1–5%) initial aggregates. Lower values are seen at 30°C (lower plot) with a half-life of 3.46 days. .... 107

## List of Tables

|  |    |
|--|----|
| Table 2-1: Data analysis software used for data interpretation of the glycan analysis. ....  | 19 |
| Table 2-2: Analytical tools for stability assessment of monoclonal antibodies (mAbs).....  | 25 |
| Table 2-3: Impact of various stressors on aggregation behaviour and mechanisms behind the observed aggregation behaviour. ....   | 28 |
| Table 3-1: Hyperparameters used in designing XGBoost algorithm for the present dataset. ...  | 41 |
| Table 3-2: List of observed glycoforms derivatized from Instant-PC label and analysed through LC-ESI-MS. ....  | 43 |
| Table 3-3: Comparison of time consumption for N-Glycan analysis using evolving methodologies with the modified model-based quantification approach developed in this study. ....   | 50 |
| Table 3-4: Comparison of estimation of glycan abundance in LC-FLD-ML and LC-MS. ....   | 51 |
| Table 4-1: Impact of gCQA on mAb product quality and their criticality with tolerance limit. ....  | 57 |
| Table 4-2: RT range for N-glycans acquired using LC-FLD analysis, along with their criticality and tolerance limit.....  | 65 |
| Table 5-1: Design and experimental setup of pH, thermal and photo stress for degradation. .  | 79 |
| Table 5-2: Protein standards used to obtain the calibration curve, elution order, and hydrodynamic diameter.....   | 82 |
| Table 5-3: Summary of the test, validation, and regression parameters of experimental SEC and predicted oligomers using DLS-ML model for both the mAbs. ....   | 87 |
| Table 5-4: Mean percentage composition for each of the oligomeric specie (monomer, dimer, trimer, tetramer), at 1mg/mL, 10mg/mL and 21mg/mL for mAb1 using neural network algorithm. ....  | 89 |
| Table 5-5: Comparison of DLS and SEC-HPLC for size estimation of mAbs.....   | 89 |
| Table 6-1: Aggregation of mAbs incubated under various stress conditions. The initial aggregation level comprises of low (1-5%), medium (5-15%) and high (15-25%). The aggregates were generated through forced degradation at 30°C for the specified durations. . | 99 |

## List of Abbreviations

|            |  |
|------------|--|
| mAb        | Monoclonal antibody                                      |
| DLS        | Dynamic Light scattering                                 |
| ML         | Machine learning   |
| MS         | Mass spectroscopy  |
| HMW        | High molecular weight                                    |
| CQA        | Critical quality attribute                               |
| US-FDA     | US Food and Drug Administration                          |
| SEC        | Size exclusion chromatography                            |
| VWD        | Variable wavelength detector                             |
| SVM        | Support vector machine                                   |
| NN         | Neural network   |
| PSD        | Particle size distribution                               |
| SVR        | Support vector regression                                |
| RMSE       | Root means square error                                  |
| HPLC       | High Performance Liquid Chromatography                   |
| BSA        | Bovine serum albumin                                     |
| ADH        | Alcohol dehydrogenase                                    |
| EMA        | European Medicines Agency                                |
| gCQA       | glycan related critical quality attribute                |
| CV         | Critical value   |
| NHC        | N-hydroxysuccinimide carbamate                           |
| SD         | Standard deviation                                       |
| Gal        | Galactose  |
| GlcNAc     | N-acetylglucosamine                                      |
| PD         | pharmacodynamics   |
| PK         | pharmacokinetics   |
| GI         | Glycosimilarity index                                    |
| LC-FLD     | Liquid chromatography- Fluorescence detector             |
| LC-MS      | Liquid chromatography- mass spectrometry                 |
| Instant-PC | Instant Procaine   |
| Qin        | %points inside tolerance limit                           |
| Qout       | %points outside tolerance limit                          |
| SSQ        | Sum of square  |
| Xs         | Biosimilar intensity falling outside the tolerance limit |
| Xmean      | Mean of intensities for innovator batches                |

|          |  |
|----------|--|
| RT       | Retention time                         |
| XGBoost  | Extreme Gradient Boosting              |
| QbD      | Quality by Design                      |
| PAT      | Process Analytical Technology          |
| HTP      | High throughput                        |
| PNGase-F | Peptide N-glycosidase F                |
| CE       | Capillary electrophoresis              |
| HILIC    | Hydrophilic interaction chromatography |
| ESI-MS   | Electrospray ionisation-MS             |
| SPE      | Solid phase extraction                 |
| TIC      | Total ion chromatogram                 |