Department of Chemistry, 2025

Data-Driven Biopharmaceutical Manufacturing:

The Role of Process Analytical Technology and Chemometrics

Rafael Machleid

UMEÅ UNIVERSITY

ISBN: 978-91-8070-841-8 (print)
ISBN: 978-91-8070-842-5 (pdf)
Electronic version available at: http://umu.diva-portal.org/
Printed by: Scandinavian Print Group, Skarpnäck, 2025

To Maria

Table of Contents

Abstract . . . ii
Zusammenfassung . . . iii
Sammanfattning . . . iv
List of Publications . . . v
Author Contributions . . . vi
Abbreviations . . . vii

1 Introduction and Scope . . . **1**

2 Background . . . **5**
2.1 Biopharmaceuticals . . . 5
2.2 Biopharmaceutical Manufacturing . . . 6
2.2.1 Upstream Bioprocessing . . . 7
2.2.2 Downstream Bioprocessing . . . 9
2.3 Process Monitoring and Control in Bioprocessing . . . 9
2.3.1 Process Analytical Technology (PAT) . . . 10
2.3.2 Raman Spectroscopy . . . 12
2.3.3 Bio-Capacitance . . . 14
2.4 Chemometrics in Bioprocessing . . . 15
2.4.1 Data Generation and Design of Experiments (DOE) . . . 16
2.4.2 Multivariate Data Analysis (MVDA) . . . 17
2.4.3 Multivariate Calibration . . . 21

3 Results and Discussion . . . **25**
3.1 Paper I: Feasibility and performance of cross-clone Raman calibration models in CHO cultivation . . . 25
3.2 Paper II: Real-time VCC monitoring and forecasting in HEK-cell-based rAAV vector production using capacitance spectroscopy . . . 27
3.3 Paper III: Advancing Raman calibration: Automated data generation, monitoring, and control in multi-parallel perfusion mini bioreactors . . . 29
3.4 Paper IV: Time-adjusted performance evaluation (TAPE) of predictive multivariate models for bioprocess data . . . 32

4 Conclusion and Future Perspective . . . **35**

Acknowledgements . . . **37**

Bibliography . . . **39**

Abstract

Biopharmaceuticals have transformed modern medicine over recent decades due to their efficacy in treating various severe diseases such as cancers, autoimmune disorders, and genetic conditions. While representing a wide class of different treatments, biopharmaceuticals are commonly produced through genetically engineered living organisms. In practice, the therapeutics of interest are typically manufactured through cell culture in bioreactor systems. However, this process is complex as it relies on fragile biological systems that need to be monitored and tightly controlled. To achieve such monitoring and control, process analytical technology (PAT) is needed. Spectroscopic methods, such as Raman and bio-capacitance spectroscopy, have been presented as potential PAT candidates due to their real-time, non-invasive measurement capabilities of various critical cell culture parameters (e.g., glucose and cell concentration). However, the use of spectroscopic sensors as PAT tools greatly depends on robust multivariate calibration models. These models are required to translate spectral data into actual process parameter values.

This thesis addresses fundamental challenges in calibration modeling for PAT implementation in biopharmaceutical manufacturing. Specifically, the use of Raman and bio-capacitance spectroscopy as PAT tools in upstream cell culture is investigated. We explore how biological variation impacts the transferability and robustness of Raman-based monitoring models in **Paper I**. In **Paper II**, we extend beyond monitoring by demonstrating how the combination of classical chemometric calibration models and simplified mechanistic models can yield accurate forecasts during cell culture, effectively developing a predictive decision support system. **Paper III** explores calibration data generation by presenting an automated workflow using a miniature-scale high-throughput bioreactor system. Its usefulness is further demonstrated by developing and deploying calibration models for monitoring and control in perfusion culture. Finally, **Paper IV** explores a novel validation framework for calibration models that tests the specificity and robustness of developed models.

Zusammenfassung

Translated from English to German using DeepL (www.deepl.com) with minor personal corrections.

Biopharmazeutika haben die moderne Medizin in den letzten Jahrzehnten aufgrund ihrer Wirksamkeit bei der Behandlung verschiedener schwerer Krankheiten wie Krebs, Autoimmun- und genetischen Erkrankungen verändert. Obwohl sie eine breite Palette verschiedener Behandlungsmethoden darstellen, werden Biopharmazeutika in der Regel durch gentechnisch veränderte lebende Organismen hergestellt. In der Praxis werden die gewünschten Therapeutika typischerweise durch Zellkulturen in Bioreaktorsystemen produziert. Dieser Prozess ist jedoch komplex, da er auf empfindlichen biologischen Systemen basiert, die überwacht und streng kontrolliert werden müssen. Um eine solche Überwachung und Kontrolle zu gewährleisten, ist die Prozessanalyse-Technologie (PAT) erforderlich. Spektroskopische Verfahren wie Raman- und Biokapazitätsspektroskopie werden aufgrund ihrer Echtzeit- und nicht-invasiven Messprinzipien verschiedener kritischer Zellkulturparameter (z. B. Glukose- und Zellkonzentration) als potenzielle PAT-Kandidaten gehandelt. Der Einsatz spektroskopischer Sensoren als PAT-tools hängt jedoch in hohem Maße von robusten multivariaten Kalibrierungsmodellen ab. Diese Modelle sind erforderlich, um Spektraldaten in tatsächliche Prozessparameterwerte umzuwandeln.

Diese Arbeit befasst sich mit grundlegenden Herausforderungen der Kalibrierungsmodellierung für PAT-Implementierung in der biopharmazeutischen Produktion. Insbesondere wird der Einsatz von Raman- und Biokapazitätsspektroskopie als PAT-tools in der Upstream Zellkultur untersucht. In Artikel I untersuchen wir, wie sich biologische Variationen auf die Übertragbarkeit und Robustheit von Raman-basierten Überwachungsmodellen auswirken. In Artikel II gehen wir über die reine Prozessüberwachung hinaus und zeigen, wie die Kombination aus klassischen chemometrischen Kalibrierungsmodellen und vereinfachten mechanistischen Modellen zu genauen Vorhersagen während der Zellkultur führen kann, wodurch ein prädiktives Entscheidungsunterstützungssystem entwickelt wird. Artikel III untersucht die Generierung von Kalibrierungsdaten, indem es einen automatisierten Arbeitsablauf unter Verwendung eines miniaturisierten Hochdurchsatz-Bioreaktorsystems vorstellt. Die Nützlichkeit wird weiterhin durch die Entwicklung und den Einsatz von Kalibrierungsmodellen für die Überwachung und Steuerung in der Perfusionskultur demonstriert. Schließlich untersucht Artikel IV einen neuartige Validierungsmethode für Kalibrierungsmodelle, welche die Spezifität und Robustheit der entwickelten Modelle testet.

Sammanfattning

Translated from English to Swedish using DeepL (www.deepl.com).

Biofarmaceutiska läkemedel har förändrat den moderna medicinen under de senaste decennierna tack vare sin effektivitet vid behandling av olika allvarliga sjukdomar såsom cancer, autoimmuna sjukdomar och genetiska tillstånd. Biofarmaceutiska läkemedel representerar en bred klass av olika behandlingar, men tillverkas vanligtvis genom genetiskt modifierade levande organismer. I praktiken tillverkas de aktuella läkemedlen vanligtvis genom cellodling i bioreaktorsystem. Denna process är dock komplex eftersom den är beroende av ömtåliga biologiska system som måste övervakas och kontrolleras noggrant. För att uppnå sådan övervakning och kontroll krävs processanalytisk teknik (PAT). Spektroskopiska metoder, såsom Raman- och biokapacitansspektroskopi, har presenterats som potentiella PAT-kandidater på grund av deras förmåga att i realtid och icke-invasivt mäta olika kritiska cellodlingsparametrar (t.ex. glukos- och cellkoncentration). Användningen av spektroskopiska sensorer som PAT-verktyg är dock i hög grad beroende av robusta multivariata kalibreringsmodeller. Dessa modeller krävs för att översätta spektraldata till faktiska processparametervärden.

Denna avhandling behandlar grundläggande utmaningar inom kalibreringsmodellering för PAT-implementering inom biofarmaceutisk tillverkning. Specifikt undersöks användningen av Raman- och biokapacitansspektroskopi som PAT-verktyg i uppströms cellodling. I artikel I undersöker vi hur biologisk variation påverkar överförbarheten och robustheten hos Raman-baserade övervakningsmodeller. I artikel II går vi bortom övervakning genom att visa hur kombinationen av klassiska kemometriska kalibreringsmodeller och förenklade mekanistiska modeller kan ge noggranna prognoser under cellodling, vilket effektivt utvecklar ett prediktivt beslutsstödssystem. Artikel III utforskar generering av kalibreringsdata genom att presentera ett automatiserat arbetsflöde med hjälp av ett miniatyrskaligt bioreaktorsystem med hög genomströmning. Dess användbarhet demonstreras ytterligare genom utveckling och implementering av kalibreringsmodeller för övervakning och kontroll i perfusionsodling. Slutligen utforskar artikel IV ett nytt valideringsramverk för kalibreringsmodeller som testar specificiteten och robustheten hos utvecklade modeller.

List of Publications

This thesis is based on the following publications/manuscripts, referred to as **Paper I**, **II**, **III**, and **IV**:

Paper I *Feasibility and performance of cross-clone Raman calibration models in CHO cultivation.*
Rafael Machleid, Marek Hoehse, Steffi Scholze, Kleanthis Mazarakis, David Nilsson, Erik Johansson, Christoph Zehe, Johan Trygg, Christian Grimm, Izabella Surowiec.
Biotechnolgy Journal, **2024**; 19(1):e2300289.
DOI: https://doi.org/10.1002/biot.202300289.

Paper II *Real-Time VCC Monitoring and Forecasting in HEK-Cell-Based rAAV Vector Production Using Capacitance Spectroscopy.*
Rafael Machleid, Suneetha Nunna, Ajith George, Jonas Austerjost, Magda Tomala, Izabella Surowiec.
Engineering in Life Sciences, **2025**; 25:e70004.
DOI: https://doi.org/10.1002/elsc.70004.

Paper III *Advancing Raman Calibration: Automated Data Generation, Monitoring, and Control in Multi-Parallel Perfusion Mini Bioreactors.*
Rafael Machleid[†], Maximilian Goertz[†], Christian Grimm, Johan Trygg, Sven Mathias, Izabella Surowiec.
Manuscript submitted: Biotechnology Journal.

Paper IV *Time-Adjusted Performance Evaluation (TAPE) of Predictive Multivariate Models for Bioprocess Data.*
Andreas Eriksson, **Rafael Machleid**, Anne Richelle, Johan Trygg, Henrik Antti, Izabella Surowiec, Olivier Cloarec, Pär Jonsson.
Manuscript submitted: Journal of Analytical Science and Technology.

[†] These authors contributed equally to this work and are recognized as co–first authors.

Author Contributions

Paper I Lead role in writing original draft, reviewing, editing, visualization, validation, investigation, formal analysis, data curation, and conceptualizing the methodology along with the co-authors.

Paper II Lead role in writing original draft, reviewing, editing, visualization, validation, investigation, formal analysis, data curation, and conceptualizing the methodology along with the co-authors.

Paper III Lead role in writing original draft, reviewing, editing, and visualization. Shared lead role in validation, investigation, formal analysis, and conceptualizing the methodology.

Paper IV Supporting role in formal analysis, conceptualization, investigation, methodology, data curation, and reviewing original draft.

Abbreviations

BHK	Baby hamster kidney
CAGR	Compound annual growth rate
CHO	Chinese hamster ovary
COVID-19	Coronavirus disease 2019
CPP	Critical process parameter
CQA	Critical quality attribute
CSPR	Cell-specific perfusion rate
DNA	Deoxyribonucleic acid
DO	Dissolved oxygen
DOE	Design of experiments
DSP	Downstream processing
FDA	U.S. Food and Drug Administration
FMEA	Failure mode and effects analysis
HEK	Human embryonic kidney
HPLC	High-performance liquid chromatography
HT	High-throughput
ICH	International Council for Harmonisation
IR	Importance ratio
LOGO	Leave-one-group-out
mAb	Monoclonal antibody
MAE	Mean absolute error
MSPC	Multivariate statistical process control
MVDA	Multivariate data analysis
NIR	Near-infrared
NS0	Mouse myeloma cell line
OFAT	One-factor-at-a-time
OPLS	Orthogonal projections to latent structures

OPLS-DA	Orthogonal projections to latent structures - discriminant analysis
PAT	Process analytical technology
PCA	Principal component analysis
PID	Proportional–integral–derivative
PLS	Projection to latent structures
PTM	Post-translational modification
Q^2	Coefficient of prediction
QbD	Quality-by-design
R^2	Coefficient of determination
rAAV	Recombinant adeno-associated virus
RMSE	Root mean square error
RMSECV	Root mean square error of cross-validation
RMSEE	Root mean square error of estimation
RMSEP	Root mean square error of prediction
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
SMA	Spinal muscular atrophy
SMN1	Survival of motor neuron 1
SNV	Standard normal variate
SR	Selectivity ratio
TAPE	Time-adjusted performance evaluation
TCC	Total cell concentration
TTT	Time-till-transfection
USP	Upstream processing
UV	Ultraviolet
VCC	Viable cell concentration
VCV	Viable cell volume
VIP	Variable importance in projection
WFI	Water for injection

1 Introduction and Scope

Biopharmaceuticals have transformed modern medicine over recent decades, and are now representing the majority of top-selling drugs globally. The transition from small molecules to complex biological products has, in some cases, enabled treatment of previously incurable diseases. This rise was further highlighted during the COVID-19 pandemic where biopharmaceuticals proved essential as part of global mitigation strategies. The impact of biopharmaceuticals on today's healthcare systems is further reflected by their global market size. In 2024, the global market reached $422 billion, which is expected to grow at a projected compound annual growth rate (CAGR) of 12.5% through 2034, corresponding to a market of $1,369 billion [1]. While mAbs and hormones represent rather established biopharmaceutical therapeutics, new modalities such as cell and gene therapies have recently started entering the market.

However, there are major challenges the biopharmaceutical industry must face in the coming years. Expiring patents of many blockbuster biopharmaceuticals pave the way for competitive biosimilars, leading to eroding market shares and substantial cost pressure across the industry. These challenges converge with aging populations, which on one hand increase drug demand, but on the other hand put healthcare systems under considerable financial pressure. This paradox of surging demand and constrained budget threatens not only patients' access to critical treatments but also places additional economic pressure on manufacturers. Moreover, the development of biopharmaceuticals poses significant inherent risks, as it requires massive upfront investments despite high failure rates that persist even in late-stage clinical trials. The need to reduce costs while adhering to the highest quality and regulatory requirements demands a significant boost in biopharmaceutical manufacturing efficiency.

Achieving such efficiency gains is particularly challenging because biopharmaceutical manufacturing is inherently complex, as it relies on fragile genetically engineered biological systems and sophisticated purification steps to isolate the biological product. In response to these constraints, Quality-by-design (QbD) [2] has emerged as a systematic approach to enhance development and manufacturing efficiency while ensuring product quality by embedding quality into the manufacturing process through systematic design. This represents a paradigm shift in biopharmaceutical manufacturing, which traditionally primarily relied on end-product testing strategies (Quality-by-testing) for product release. As opposed to QbD, this approach bears significantly higher risks of producing waste, as product quality is determined mostly after the fact. The ultimate goal of QbD is to ensure consistent product quality through increased process understanding, which in turn can boost overall process efficiency.

Central to QbD is process analytical technology (PAT) [3], often considered the practical QbD-enabler. PAT represents a framework of hardware and software tools that aim to generate process knowledge and enable real-time monitoring and control through timely process mea-

surements. Spectroscopic techniques, such as Raman spectroscopy and multi-frequency biocapacitance, have gained particular interest as potential PAT candidates due to their non-invasive, real-time measurement capabilities.These have been demonstrated for real-time monitoring of key performance indicators, as well as critical process parameters (CPPs) and critical quality attributes (CQAs) in upstream (USP) and downstream processing (DSP) of biopharmaceuticals.

The successful implementation of PAT-based monitoring is highly dependent on robust calibration models. These models are needed to translate multivariate spectral data into actual process parameter values. However, they also introduce certain obstacles that must be addressed for successful PAT implementation. Calibration model development relies on extensive calibration data that captures usual process dynamics and relevant process variations to ensure model robustness. Generating such data in large-scale processes is impractical and can even be infeasible due to the high costs involved, particularly in upstream cell culture. Therefore, small-scale systems are required, which enable calibration data generation workflows at significantly lower costs. However, these systems introduce other challenges regarding model transferability to larger scales.

Model transfer represents another frequently faced challenge in calibration modeling. Given that calibration data generation is cumbersome, there is a strong incentive to reuse historical data, for example, from early process development experiments or other processes. As such, studies have investigated the transferability of calibration models between process scales, cell lines, cell platforms, or even different spectrometers and probes. Opposed to these, Raman calibration model transfer in highly dynamic cell culture environments with large biological variation is not yet well understood.

In addition, a major issue with PAT-based calibration models is their limited specificity and robustness. While these models seek to establish a relationship between spectral data and the process parameter of interest, they cannot distinguish between causal effects and mere correlations in analyzer readings. This is particularly challenging for cell cultivations, in which analytes are typically strongly correlated with each other and with elapsed process time. These correlations create confounding structures, which interfere with models' specificity and undermine their robustness. Weak models can exploit such confounding trends, which is why, ironically, many literature sources claim successful Raman-based monitoring of Raman-inactive analytes, such as monoatomic ions. While such flawed models may perform well in tightly controlled processes, they fail when correlation structures shift due to unexpected process deviations or intentional changes. While experimental methods and systematic calibration data generation frameworks can help to break such correlations, conventional model validation strategies fail to detect models that exploit these effects. This creates substantial risks for PAT-based calibration model deployment, particularly in commercial-stage biopharmaceutical manufacturing.

This thesis addresses critical gaps in calibration modeling for PAT implementation in biopharmaceutical manufacturing. The research focuses specifically on Raman and bio-capacitance spectroscopy as PAT tools in upstream bioprocessing, using Chinese hamster ovary (CHO) and human embryonic kidney (HEK) cells for mAb and recombinant adeno-associated virus (rAAV) vector production, respectively. Four fundamental research questions guide this work: (1) How does biological variation restrict PAT implementation? (2) How can we transition from mere process monitoring to forecasting? (3) How can calibration data be efficiently generated? (4) How can we prevent flawed models from being deployed in commercial biopharmaceutical manufacturing?

In **Paper I**, the impact of clonal-level biological variation on the transferability of Raman calibration models in CHO fed-batch cultures is investigated. While glucose and lactate models demonstrated robust transferability across clones with diverse productivity and growth profiles, titer models were limited in that regard due to a substantial productivity-dependent bias. These findings are especially relevant for applications with large biological variations, such as cell line development. Extending beyond monitoring capabilities, **Paper II** develops viable cell concentration (VCC) forecasting strategies using bio-capacitance sensors in HEK-based rAAV vector production. The study successfully deploys time-till-transfection forecasting models, enabling precise timing of culture transfection to enable optimal rAAV vector yields and consistent product quality. **Paper III** addresses the data generation challenges identified in the previous studies by establishing an automated workflow for robust calibration data generation. The practical utility of this approach is demonstrated through successful Raman-based monitoring and control in miniature-scale CHO perfusion culture. Finally, **Paper IV** introduces a novel validation framework, which helps to identify calibration models that exploit time-dependent confounding rather than relying on genuine analyte-specificity.

These four studies address fundamental challenges of robust PAT implementation in modern-day biopharmaceutical manufacturing using data-driven, reliable chemometric models.

2 Background

2.1 Biopharmaceuticals

Biopharmaceuticals are therapeutics derived from living organisms using biotechnology methods, such as recombinant DNA technology, controlled gene expression, and hybridoma techniques [4]. They encompass a variety of different classes including monoclonal antibodies (mAbs), recombinant proteins, vaccines, nucleic acid-based therapeutics, and products from cell and gene therapies. As such, they differentiate from classical biologics, which are blood and plasma-derived products such as clotting factors, immunoglobulins, and albumin. As opposed to biopharmaceuticals, these can be directly obtained from human blood rather than relying on genetic engineering [5].

Since recombinant human insulin's market approval in 1982 [6], biopharmaceuticals have gained immense popularity, revolutionizing the landscape of modern medicine. This is especially true, considering that for some diseases, such as Type 1 diabetes, they serve as the only viable therapeutic option on the market [7]. The efficacy and functionality of biopharmaceuticals are based on their molecular structure-function relationship, which enables versatile binding capabilities coupled with precise target recognition [8, 4].

As such, biopharmaceuticals can target specific disease mechanisms. For example, anti-amyloid mAbs used in Alzheimer's disease treatment target amyloid beta plaques, which have been shown in clinical trials to modestly slow cognitive decline [9]. Another example is Zolgensma, a gene therapy drug that delivers a functional copy of the SMN1 gene directly to patients' cells using a rAAV vector [10]. Thus, it targets the underlying cause of spinal muscular atrophy (SMA). For many diseases, biopharmaceuticals represent the greatest hope for future cures, moving beyond the mere management of symptoms.

In 2024, biopharmaceuticals accounted for ~37% of global pharmaceutical sales [11, 12]. This significant market share is mostly associated with the high treatment costs, considering that they contribute only a small fraction to total drug prescription volumes [13]. For example, individual treatments can range from tens of thousands of dollars per year [14] to more than $2 million for one-time gene therapy treatments such as Zolgensma [15]. These high costs are driven by complex development processes that often span more than a decade [16], require substantial upfront investments, and carry a substantial risk of failure in clinical trials [17, 18]. In addition, regulatory requirements and complex manufacturing processes add to overall costs. The latter will be discussed in more detail in the following section.

2.2 Biopharmaceutical Manufacturing

From a high-level perspective, biopharmaceuticals are produced by the so-called expression systems. These are living organisms genetically engineered to *express* a molecule of interest, often a therapeutic protein. For example, the production of recombinant human insulin was based on a prokaryotic expression system, using *E. coli* bacteria [19]. Bacteria-based expression systems generally exhibit rapid growth and high volumetric yield, leading to relatively inexpensive therapeutics. However, the products also lack essential human-like molecular features, which are critical for biological activity and immunogenicity [20]. Eukaryotic yeast-based expression systems can introduce such molecular features, for example glycosylations, through post-translational modifications (PTMs). However, the resulting patterns often differ from the human profiles [21].

In contrast, mammalian-based expression systems produce therapeutics with human-like PTMs, making them the go-to choice for modern biopharmaceutical production [22]. In recent years, most of the new biopharmaceuticals entering the US and EU markets were produced through mammalian expression systems (72% of all approvals from 2020-2022) [23]. The most widely used mammalian cell types are CHO cells, particularly used for mAb production. Other commonly used mammalian cell types include NS0 mouse myeloma, baby hamster kidney (BHK) cells, and HEK cells [23]. During the COVID-19 pandemic, HEK cells became a key platform for vaccine production due to their high transfection efficiency and ability to produce human-like PTMs [24]. As such, they supported the generation of viral vectors [25] and the expression of recombinant SARS-CoV-2 spike proteins [26].

Although biopharmaceuticals produced in mammalian expression systems provide significant benefits compared to other systems, these advantages come with substantial costs arising from their complex manufacturing processes. While the expression step (also known as production) is at the core, it is embedded within a chain of multiple critical unit operations, forming the entire biopharmaceutical manufacturing process (Figure 1). At a high level, the process can be divided into upstream and downstream bioprocessing operations, which will be covered separately in more detail.

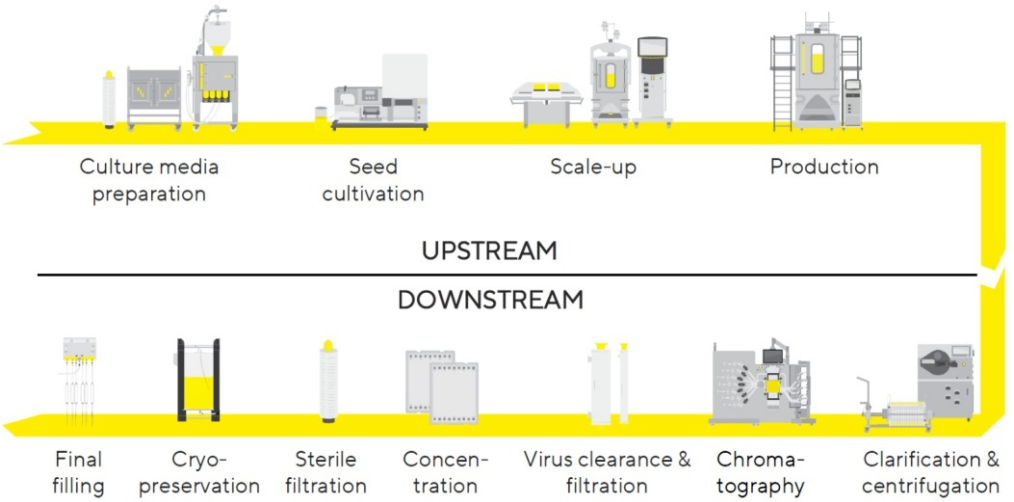

Figure 1: Illustration of mAb production unit-operations arranged according to the typical product manufacturing flow, divided into upstream (upper part) and downstream (lower part) operations. Reproduced from [27].

2.2.1 Upstream Bioprocessing

The upstream bioprocess operation has the production step at its core and its final step. The previous steps are all aimed at preparing the expression system and complementary parts for this essential unit operation. This section provides a general overview of upstream bioprocessing for mAb and viral vector production through mammalian cells. Notably, the exact steps can vary depending on the therapeutic product, expression system, and manufacturing strategy.

The first step in upstream bioprocessing is usually the culture media preparation. In this step, raw materials, such as amino acids, sugars, vitamins, salts, and buffers, are weighed and dissolved in WFI (water for injection), according to stringent recipes to generate the required culture medium. Subsequently, pH and osmolality are adjusted to match optimal cell culture conditions for following upstream bioprocessing steps. Lastly, the culture medium is sterilized, usually through sterile filtration. Culture media preparation is a critical step as the media composition directly influences the cell growth, productivity, and product quality of the expressed therapeutic.

Seed cultivation and scale-up describe the stepwise expansion of the cryopreserved working cell bank from a vial to larger bioreactors scales, in a step-wise approach. Here, cells are cultured using the culture medium previously prepared (or variations thereof). The goal is to make cells adapt to the culture conditions and medium, and reach high cell concentrations, while maintaining sufficient viability. This process is repeated and expanded until the required inoculum volume for the production bioreactor is reached, completing the seed train.

The production step entails the final cell cultivation, with the aim of producing the therapeutic of interest in large quantities. This operation is typically performed in large-scale bioreactors with volumes up to thousands of liters. To ensure high productivity, high cell concentrations and viability are essential. These can only be achieved if optimal cell culture conditions are established in the production bioreactor. In this context, the CPPs define factors that are critical for the success of the cultivation process. These generally encompass temperature, pH, dissolved oxygen (DO), osmolality, agitation, nutrient concentrations, feeding strategy, by-product levels, and others. Strict control strategies are required to keep the CPPs within limits otherwise CQA deviations become likely. The CQAs quantify the quality profile of the product and generally comprise PTMs, impurities, aggregates, fragments, charge variants, and for viral vectors, parameters such as full/empty capsid ratio.

The production step exists in three different operation modes: batch, fed-batch, and perfusion. In the simplest mode, namely batch mode, cell culture is not replenished with fresh nutrients, causing a quick nutrient depletion, by-product accumulation, and relatively quick culture death. In contrast, the fed-batch mode uses regular feeding steps to replenish nutrients. Thus, a longer-lasting cell culture can ultimately be achieved with higher productivity. Still, the accumulation of toxic by-products, such as lactate, ammonia, and formate, remains a major limiting factor [28].

The increasing demand for biopharmaceuticals, combined with competitive and regulatory pressure, drives the need for substantial manufacturing optimizations. Among the promising initiatives are process intensification [29] and continuous manufacturing [30], both of which aim to minimize facility footprints, optimize media consumption, and overall increase process efficiency. In the production step, continuous perfusion culture has been identified as an essential improvement [31]. In perfusion, the accumulation of inhibitory and toxic by-products is substantially reduced by continuously adding fresh culture medium to the bioreactor while removing spent media, ideally reaching steady-state conditions. Thus, higher cell concentrations can be reached and maintained for longer periods, significantly increasing volumetric productivity [32]. Here, cell-retention devices are critical in ensuring that cells are retained in the bioreactor while the product is harvested in a continuous manner.

This thesis addresses upstream production-step monitoring challenges across multiple platforms: CHO-based mAb production in fed-batch (**Paper I**) and perfusion (**Paper III**) culture, and HEK293-based rAAV vector production in batch culture (**Paper II**).

2.2.2 Downstream Bioprocessing

This section provides a concise overview of downstream bioprocessing [33] for the sake of completeness, although it does not represent a central focus of this thesis.

As shown in Figure 1, the first downstream unit operation, clarification & centrifugation, removes cells and cellular debris from the product-containing broth obtained in the upstream production step. For mAbs, which are secreted into the culture medium, this step directly yields clarified supernatant. For viral vectors, additional lysis steps may be required to release the intracellular product into the culture medium before clarification. Deviations from this general workflow occur in perfusion culture, where the product is already harvested during the continuous cultivation, or in processes where the cells themselves are the product (e.g., cell therapies).

All subsequent operations aim to increase the purity of the product by separating it from aggregates, fragments, host-cell proteins, and other cell culture by-products and components. Both mAb and viral vector purification rely on multiple chromatography steps, though viral vector purification faces the additional challenge of separating full from empty capsids. Considerable emphasis is placed on making the product safe for patient administration, comprising concentration and sterilization steps. For mAbs, virus clearance & filtration and sterilization steps are crucial, while for viral vectors maintaining capsid integrity with adequate purity is needed. Finally, the product is formulated and filled into vials, syringes, or other containers.

2.3 Process Monitoring and Control in Bioprocessing

Around the world, industries are changing due to the digital transformation [34]. In the manufacturing industry, it has even started to reverse decades of off-shoring, as companies regain cost competitiveness by relocating production to technologically advanced sites [35]. The biopharmaceutical industry, with its complex processes and stringent quality requirements, stands to benefit particularly from this digital transformation. Specifically, efficiency gains are urgently needed considering that the biopharmaceutical industry must face major economic pressures in the coming years. While on one hand patent cliffs and biosimilars entering the market reduce revenues [36], on the other hand the aging population puts healthcare budgets under considerable pressure [37], requiring governments to take harsher stances in price negotiations. These converge with immense development costs and generally long time-to-market for biopharmaceuticals, requiring large upfront investments despite high probability of never reaching market approval [17]. In addition, the manufacturing process itself is inherently complex and subject to highest quality and regulatory standards, making optimizations and changes in general difficult. In response to these constraints, the QbD initiative was introduced by the ICH as part of the ICH Q8 guideline [2].

The QbD framework represents a paradigm shift within pharmaceutical production: instead of testing the quality of a product after it has been produced (Quality-by-Testing), it builds quality into the manufacturing process itself, by design. Practically speaking, QbD seeks to identify CPPs and their relationship with CQAs, and thereby establish a robust design space where process variations can be managed without compromising product quality. This methodology allows greater flexibility during manufacturing, effectively reducing regulatory burden in the long-run. Importantly, QbD shifts quality ownership from regulatory oversight to industry expertise [38] - requiring manufacturers to build extensive process knowledge as early as during product and process development and incorporate it into the manufacturing process.

However, successful QbD implementation depends critically on comprehensive process monitoring and control capabilities. This is problematic considering that state-of-the-art biomanufacturing remains highly dependent on manual operator interventions. To this day, the production step often still requires regular manual sampling to determine the state of the culture, a practice that entails considerable drawbacks. Besides being labor-intensive and costly, manual sampling substantially increases the risk of contamination and potential batch loss. Moreover, off-line sample analyses compound these challenges: bio-analyzers provide rapid results but with limited information, while high-performance separation techniques provide an extensive characterization but require sample preparation and often take hours to complete. As a result, operators are mostly left in the dark about the real-time culture state, including specific CPPs and CQAs, making timely interventions, in case of process deviations, nearly impossible. This underscores the necessity for PAT in biopharmaceutical manufacturing.

2.3.1 Process Analytical Technology (PAT)

The PAT framework [3] was launched by the U.S. Food and Drug Administration (FDA) in 2004. It is defined as

> *"a system for designing, analyzing, and controlling manufacturing through timely measurements (i.e., during processing) of critical quality and performance attributes of raw and in-process materials and processes, with the goal of ensuring final product quality." [3]*

In this sense, PAT can be regarded as the practical enabler of QbD. PAT encompasses a variety of tools intended to increase process knowledge, support continuous improvement, and enable risk management. These tools include hardware components (process analyzers and sensors), software (data acquisition, analysis, and control tools), and methodologies (design tools such as design of experiments (DOE), and continuous improvement and knowledge management frameworks) [3]. Further, PAT is applicable throughout the entire biopharmaceutical product lifecycle, from initial screening studies using DOE to final CPP/CQA monitoring and control in commercial manufacturing using process analyzers. The benefits of applying the PAT framework lie in improved process understanding by identifying the relationship between CPPs and

CQAs, thereby helping to optimize processes and establish robust conditions for production operations. In addition, manufacturing efficiency can be improved by detecting deviations as early as possible, mitigating out-of-specification events and waste. Together, these capabilities aim to maintain consistent product quality and assure patient safety. Ultimately, real-time release testing becomes feasible through PAT, reducing the dependence on end-product testing strategies [3, 39, 40, 41].

Although the classical pharmaceutical industry has increasingly adopted PAT control strategies [42, 43], the complexity of biological systems has made their implementation in the biopharmaceutical industry exceptionally difficult. Biopharmaceutical manufacturing requires advanced analyzers capable of differentiating between the biological product and highly similar by-products. Such methods exist and are usually used as part of end-product testing strategies (e.g., HPLC), however, they are rather time-intensive and costly, and may require (manual) sample preparation [44, 45]. These limit their use for real-time on-line/in-line applications. Secondly, the on-line/in-line integration of process analyzers into bioprocess systems poses challenges concerning sterility and technological scalability. Most process analyzers rely on reusable probes that must be physically inserted into the system and sterilized between uses, creating a fundamental mismatch with the industry's increasing adoption of single-use technologies [46]. **Papers I**, **II**, and **III** address these integration challenges through automated at-line and in-line deployment of spectroscopic sensors with bioprocess control systems, demonstrating practical solutions for real-time data acquisition, process monitoring and control.

Spectroscopic methods, such as Raman, NIR (near-infrared), and UV, have recently attracted increased consideration as viable PAT candidates in bioprocessing [47]. These enable (near) real-time measurements of multiple analytes simultaneously and in a non-invasive manner. Nonetheless, their indirect measurement principle presents distinct challenges, such as limited specificity and sensitivity. In addition, quantitative assessments using these methods rely on complex statistical modeling which presents considerable hurdles [48], as will be discussed later.

In contrast, bio-capacitance sensors offer high specificity and sensitivity to monitor viable cell volume (VCV) and have been established in bioprocessing for decades [49]. More recently, bio-capacitance frequency scanning has emerged as an advanced variant of this technique, enabling the acquisition of additional cell physiological information by scanning permittivity across multiple frequencies. However, its implementation does also rely on multivariate calibration models.

Although these advanced technologies hold considerable promise, their effective implementation as PAT tools into biopharmaceutical manufacturing still poses significant challenges. In the following sections, Raman spectroscopy and bio-capacitance sensors will be discussed in more detail.

2.3.2 Raman Spectroscopy

Raman spectroscopy is a spectroscopic method based on the inelastic scattering of photons when interacting with matter. It is named after the Nobel prize-winning Indian physicist C.V. Raman, who discovered the *Raman effect* in 1928 [50]. The Raman effect describes the change in energy photons undergo when interacting with molecules, which is rooted in the excitation of vibrational modes. The vibrational modes describe certain ways atoms in a molecule can move relative to each other while the molecule's center of mass is retained. These movements can be categorized into either stretching (symmetric and asymmetric) or bending (scissoring, rocking, wagging, and twisting) modes. While changes in bond length are defined as stretching modes, movements involving changes in bond angles are defined as bending modes. The intensity of the Raman signal strongly depends on how much a vibration changes the polarizability of the molecule, with symmetric vibrations typically producing stronger signals than asymmetric ones.

When photons interact with a molecule, the molecule transitions from a ground state (or lower vibrational energy state) into an excited virtual energy state, a state that is very short-lived. The molecule relaxes immediately, leading to the return into a lower vibrational energy state, emitting a photon. In case the molecule returns into the initial vibration energy state, termed elastic or Rayleigh scattering, the emitted photon has exactly the same energy as the initial one. Rayleigh scattering represents the predominant scattering form. Stokes-Raman scattering describes the case in which the molecule returns into a vibrational energy state that is higher in energy than before the excitation. Here, a photon of lower energy is emitted, which is observed at only a tiny fraction of all scattering events (approximately 0.0001%) [51]. Even more rare is the case in which the molecule returns into a vibrational energy state lower than it started from, referred to as Anti-Stokes-Raman scattering. Due to its higher intensity, Stokes-Raman scattering is the signal typically measured in practical Raman spectroscopy applications. The energy difference between incoming and emitted photons, typically expressed in wavenumbers, is defined as the Raman shift. The Raman shift and the respective measured intensity represent the basis of the Raman spectrum.

In bioprocessing, Raman spectroscopy has recently attracted a great deal of interest, mostly for the purpose of on-line/in-line CPP and CQA monitoring and control. In upstream operations specifically, it is frequently used for quantitative assessment of analyte concentrations, such as glucose, lactate, amino acids, ammonium, and product titer. In addition, monitoring of other cell culture parameters such as VCC, total cell concentration (TCC), cell viability, osmolality, pH, DO, and even mononuclear ions (sodium and potassium ions) is regularly reported. Although these might not provide a specific Raman signal, tracking could still be possible due to correlations with other Raman-active analytes in the culture medium [52, 53]. However, this approach must be treated with caution, as analyte correlations can negatively affect model specificity, as discussed later.

Recently, Raman-based glucose feeding strategies have been reported for continuous CHO perfusion culture [54]. The authors demonstrated successful in-line glucose feedback control around 4 g/L and 1.5 g/L set-points for several days with a variability of ±0.4 g/L. In this case, the Raman probe was integrated into the cell-free harvest line after the cell retention device (permeate line), mostly avoiding the fluorescence background interference typically observed with Raman bioreactor measurements. Considering this advantage, Raman integrations into the harvest-line have seen increased interest lately: Studies have demonstrated successful Raman-based monitoring for various analytes, including mAb titer, glucose, lactate, glutamine, and glutamate [54, 55], and even scalability and model transferability from 10 L to 500 L bioreactor scales [56]. While the approach raises questions about the comparability of bioreactor and harvest line analyte concentrations due to sieving [57], it also prevents measurement of cell-related parameters, such as VCC and TCC. These parameters have been measured in CHO fed-batch culture through in-line Raman probe integration directly into the bioreactor [58, 59, 60]. Furthermore, Raman-based VCC feedback control in CHO perfusion culture has been presented recently [61]. Although there are multiple studies showing the ability of Raman to detect and differentiate CQA-related changes in biopharmaceuticals [62, 63], examples of actual real-time monitoring in cell culture remain more scarce [64, 65]. In addition, multiple studies have demonstrated the successful application of Raman spectroscopy in DSP [44, 66].

The benefits of using Raman spectroscopy in bioprocessing are evident, still several limitations must be taken into account. As discussed before, Raman spectroscopy exhibits low sensitivity because of the inherently low probability of Stokes-Raman scattering, which makes the detection of analytes difficult. Another major limitation is fluorescence interference: many components within cell culture media fluoresce under laser excitation at wavelengths relevant to Raman spectroscopy, which can mask the Raman signal. This often causes baseline shifts, which need to be corrected through spectral preprocessing. Furthermore, the inherent complexity of cell culture media introduces additional complications. Suspensions of cells, nutrients, metabolites, and by-products, lead to spectral overlap and potentially nonlinear relationships between signal and analyte concentrations. Therefore, multivariate calibration models are needed for quantitative assessment, which introduces further obstacles related to model development, validation, and transferability, as will be discussed later.

2.3.3 Bio-Capacitance

Bio-Capacitance sensors determine the dielectric properties of samples by applying an AC voltage across two electrodes and measuring the resulting current. The measured capacitance depends on both the sample's permittivity and the probe geometry. By normalizing for geometry, the permittivity of the sample can be obtained. This technology is of particular interest for cell-based processes because of the polarizability of viable cells. Unlike dead cells, viable cells have intact membranes that maintain selective permeability. These intact membranes act as dielectric barriers, allowing viable cells to be electrically polarized, whereas dead cells lose this property due to membrane damage. The cellular polarizability is dependent on the applied frequency and directly reflected in the measured permittivity of a cell suspension. Specifically, permittivity is primarily a function of the number of viable cells and their volume, in other words, the VCV. As such, permittivity measurements can be utilized to determine the VCV in cell culture.

However, the parameter that is typically more important than VCV is the VCC, the concentration of viable cells independent of their volume. To derive the VCC from VCV, accurate cell size/ volume measurements are needed. This represents a major obstacle, given that cell sizes often change during mammalian cell culture [67] and, therefore, regular (often manual) adjustments are required. Multi-frequency bio-capacitance, also called bio-capacitance scanning or bio-capacitance spectroscopy, has been shown to circumvent this issue [68]. These sensors measure across a range of different frequencies (e.g., from 50 kHz to 20000 kHz), opposed to the classical single-frequency approach. Thereby a permittivity spectrum is obtained rather than just a single point per measurement. This spectrum is called the β-dispersion curve (Figure 2). The β-dispersion curve [69] reflects more properties of a cell suspension than just the VCV. These include information related to cell size, through the critical frequency, and the cell size distribution, through the Cole-Cole alpha parameter. Utilizing this additional information captured by the β-dispersion curve makes it possible to estimate VCC, even without access to cell size/volume measurements. However, the shift from single- to multi-frequency data requires the use of multivariate data analysis (MVDA) techniques, such as multivariate calibration, which poses other significant challenges. These will be discussed later in this thesis.

In recent years, bio-capacitance-based sensors have seen increased adoption as on-line or in-line PAT tools in biomanufacturing [70]. A recent study demonstrated the transferability of single-frequency univariate models for VCV measurement in CHO culture from 50 L to 2000 L bioreactors, while also showing limitations in VCC measurement due to cell size changes, especially during the death phase of the culture [71]. These limitations were alleviated in another study through a multi-frequency approach, resulting in effective on-line VCC monitoring in CHO culture throughout all phases [72]. In addition, bio-capacitance multi-frequency scanning has been demonstrated for VCC monitoring in viral production, such as rAAV vectors produc-

tion in Sf9 cells [73] and microcarrier-based lentivirus production by Vero cells [74]. Besides VCV/VCC monitoring, there are many other examples of bio-capacitance sensor usage in bioprocessing [70, 49].

While **Paper I** and **III** demonstrate use of Raman spectroscopy for monitoring in CHO culture, **Paper II** presents bio-capacitance-based measurements in HEK culture.

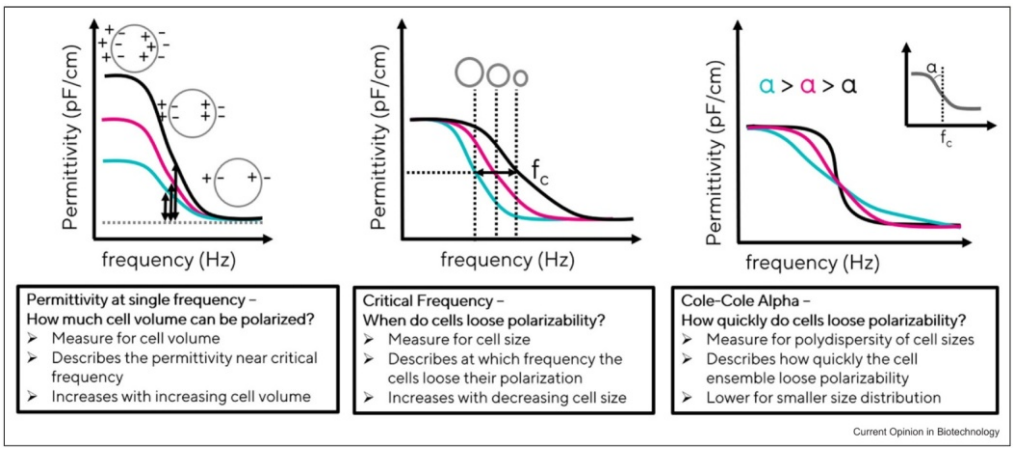

Figure 2: The β-dispersion curve illustrating frequency-dependent permittivity changes in cell suspensions. Left: permittivity at single frequency indicates viable cell volume. Center: critical frequency correlates with cell size. Right: Cole-Cole alpha parameter reflects cell size distribution. Reproduced from [49].

2.4 Chemometrics in Bioprocessing

Chemometrics is the discipline that applies mathematical, statistical, and computational methods to design experiments and analyze data with the goal of extracting meaningful process information. Chemometrics originally focused on chemical applications, hence its name, however, now it is widely used in diverse fields such as medicine, biology, and bioprocessing.

While PAT establishes the framework for real-time process monitoring and control, chemometrics offers the mathematical and statistical tools necessary to translate complex, multivariate data into meaningful process parameter values for monitoring, control and optimization. Therefore, chemometrics is often considered the backbone of PAT implementation, even though the FDA's PAT guidelines use more generalized wording, referring to multivariate tools for experimental design, data acquisition and analysis [3].

2.4.1 Data Generation and Design of Experiments (DOE)

DOE [75] describes a framework for generating experimental data in a statistically sound manner. Its goal is to maximize the information obtained from a set of experiments while minimizing the experimental effort required to achieve the objective. In addition to its economic benefits, DOE is the primary way to achieve QbD in the biopharmaceutical industry [2], as it facilitates systematic identification of process parameter and quality attribute relationships. Specifically, DOE helps to establish which process variables significantly affect the CQAs, thereby distinguishing CPPs from non-critical factors and providing the experimental foundation for design space estimations [76]. In the bioprocess context, the design space describes a multidimensional region where process parameters have been demonstrated to consistently generate product meeting CQA specifications. This provides manufacturing flexibility by allowing process variations within proven boundaries without requiring regulatory approval.

For example, *Brunner et al.* used a central composite design to investigate the effect of various parameters (pH, pO_2, and pCO_2) on CHO cell culture performance and specific CQAs [77]. The study identified notable interaction effects between pH and pCO_2 on productivity and mAb charge variant distribution. These relationships would have remained undetected using one-factor-at-a-time (OFAT) approaches. Recently, *Schwarz et al.* demonstrated the optimization of culture media components, specifically amino acids and sugar, for CHO perfusion cultivation, based on a mixture design [78]. Thus, an optimized medium formulation was established that, among other benefits, resulted in reduced lactate and ammonium production, while yielding a targeted mAb glycosylation profile.

In the context of bioprocess design space estimation, *Abu-Absi et al.* described a comprehensive methodology for design space mapping for mAb cell culture processes [79]. The authors utilized failure modes and effects analysis (FMEA) to identify CPPs, followed by a two-stage DOE approach comprising a screening design and a (central composite) optimization design. This enabled classification of process parameters as non-key, key, or critical in studies across shake flask, seed bioreactor, and production bioreactor. Furthermore, *Li et al.* established a systematic approach for scale-down model development for the characterization of commercial cell culture processes [80]. They demonstrated effective replacement of DOE studies at manufacturing scale through scale-down models, thereby reducing experimental efforts and costs while maintaining process consistency across scales.

Beyond process characterization, optimization, and design space estimation, DOE is pivotal in generating calibration data for the robust implementation of advanced analyzers according to the PAT framework. The calibration data required to develop PAT-based models must cover the design space comprehensively, while simultaneously enabling specificity to the analytes of interest. The latter is particularly challenging in cell culture processes due to the inherent cor-

relation between different analytes and specific analyte-time trends, which lead to confounding structures in the calibration data. These confounding structures can significantly impact the specificity and robustness of PAT-based models, which are not able to distinguish between causal effects and mere correlations in analyzer readings. Weak models can exploit such confounding patterns, and may obtain high predictivity, despite being fundamentally flawed. In the past different approaches have been investigated to break the dependence of spectroscopic models on such confounding effects. *Rowland-Jones et al.* utilized the IV-optimal algorithm to select samples for PAT-based multivariate modeling from a library of historical cell culture samples, representative of the design space [81]. This approach successfully reduced analyte correlations in the calibration data for Raman-based monitoring models. Another approach sought to reduce correlations through DOE-based spiking of lactate, mAb, and ammonium into the culture medium, combined with variable seeding VCCs and glucose feeding rates [82]. Thereby, cultivations with different trajectories were obtained, effectively breaking correlation between analytes and increasing the specificity and robustness of PAT-based models.

The advent of miniature-scale high-throughput automated bioreactor systems with integrated at-line spectrometers and bio-analyzers has further advanced calibration data generation strategies. These automated systems enable execution of experimental designs that systematically alter cell culture conditions (e.g., pH, dO_2, seeding VCC, glucose setpoint), producing calibration data covering a wide range of process variability across the design space. In addition confounding effects from analyte and time correlations can be effectively reduced. Further, these systems have been used for automated spiking of samples taken from the bioreactor, decreasing unwanted correlations even more and increasing calibration ranges [83, 53].

Papers I and **III** leverage such high-throughput platforms to systematically acquire data that span diverse biological variations and process conditions. Thus, workflows are established which enable DOE-based studies and systematic calibration data generation, exemplifying the importance of DOE methodology in robust PAT implementation by linking experimental design, process understanding, and real-time monitoring capabilities.

2.4.2 Multivariate Data Analysis (MVDA)

Modern bioprocessing generates an ever-increasing amount of process data. However, data alone do not constitute information or generate process knowledge unless they are properly contextualized and interpreted. Systematic data analysis is required to transform data into meaningful parameters that support process understanding and decision making.

A univariate approach to data analysis can easily become overwhelming, considering the number of different sensors and variables involved in bioprocessing. But more importantly, it can miss the big picture. When variables are analyzed in isolation, correlations between them are

overlooked, and important patterns and trends essential for process understanding may remain hidden. MVDA has a clear advantage in this regard as it considers all variables simultaneously, which provides a holistic view of process dynamics and reveals relationships that might otherwise be missed.

Today, MVDA is an established technique in bioprocessing, particularly used as a tool for process monitoring and control. In particular, multivariate statistical process control (MSPC) describes the use of MVDA techniques to develop statistical representations of ideal process trajectories and their acceptable ranges from relevant process data, called golden-batch modeling. Once established, a golden-batch model can be used to monitor future batches in real-time, using multivariate batch control charts to reveal process deviations resulting from changes in the underlying correlation structure. Coupled with operator alarms and potentially even feedback control loops, MSPC represents an efficient way to avoid and correct for process deviations. Beyond monitoring and control, MVDA is commonly utilized for root cause analysis. *Kirdar et al.* demonstrated this by analyzing small-scale (2 L), pilot-scale (2,000 L), and commercial-scale (15,000 L) CHO cell culture data to investigate scale-up issues affecting CQAs and other performance parameters. Through MVDA tools, they identified that basal medium powder, elevated culture metabolism (lactate, osmolality), and increased mid-culture pCO_2 were key factors leading to atypical product quality attributes at commercial scale due to a metabolic cascade effect [84]. This analysis enabled confirmation and correction of the root cause with only three pilot runs, showing considerable time and resource savings in contrast to conventional approaches. Furthermore, MVDA is also extensively used for predictive modeling, with regression models forecasting final product quality and process outcomes based on early-stage process data (e.g., soft-sensing) [85]. Combined with on-line model deployment, this approach enables proactive decision-making and process adjustments during the process. In addition, MVDA techniques are foundational in multivariate calibration, as will be discussed in detail later.

One prominent MVDA technique is principal component analysis (PCA) [86]. PCA is an unsupervised linear dimensionality reduction method that seeks to reduce the dimensionality of large datasets while retaining as much relevant variation as possible. Dimensionality reduction in PCA is possible as variables are often correlated with each other due to the same influence of external effects. Practically, PCA works by transforming the original correlated variables (e.g., temperature, pH, DO, etc.) into a smaller set of orthogonal variables called principal components (PCs). Each PC is a linear combination of the original variables and captures a portion of distinct variation in the data. As PCA focuses on correlations between variables, trends and clusters in the data are easily identified, and outliers can be detected. Therefore, PCA has become an essential part of exploratory data analysis, for example for spectral data, often representing the very first step in the MVDA workflow.

Another frequently used MVDA technique is projection to latent structures (PLS) [87], which is a supervised linear regression method. Similarly to PCA, it fits orthogonal components to the original data, but with a different goal. While PCA establishes orthogonal components representing the maximum variation in the original data (the X-block), PLS seeks to establish orthogonal components maximizing the covariance between the X-block and the Y-block (the responses, or output variables). Maximizing covariance describes a balancing act between 1) maximizing the extracted variation in the X-block, and 2) maximizing the correlation between the X- and Y-block LVs. This approach results in fitting LVs that describe the original data as well as possible, while also predicting the responses as well as possible.

A popular variation of PLS is orthogonal projections to latent structures (OPLS) [88]. OPLS separates systematic variation in the X-block into two parts: predictive variation that is correlated with Y, and orthogonal variation that is unrelated to Y. The orthogonal variation still represents structured information rather than random noise, but it is clearly distinguished from the predictive part. As a result, predictive and orthogonal components can be analyzed independently. OPLS models are therefore generally simpler to interpret than PLS models, while offering equivalent predictive capacity when models with the same number of latent variables are compared [88]. In strictly regulated environments such as the biopharmaceutical industry, model interpretability is paramount. This explains why black-box techniques like artificial neural networks (ANNs), which can capture more complex non-linear relationships, have not been widely adopted there yet, while linear techniques like PCA, PLS, and OPLS continue to dominate [89].

Before any modeling can be performed, the original multivariate data needs to be preprocessed. Preprocessing usually includes mean-centering, scaling, and transformation. Mean-centering subtracts the mean of each variable from the respective variable, ensuring a mean of zero for all variables. This focuses the analysis on how each observation deviates from the variables' mean, not on the absolute magnitude. Further, scaling is needed to balance the influence of each variable on the analysis. Without scaling, large-range variables (e.g., glucose 0-10 g/L) dominate the analysis, while the influence of smaller-range but potentially critical variables (e.g., pH 6.8-7.2) is limited. A widely applied scaling technique is unit variance scaling, which applies mean-centering first, followed by dividing each variable by its standard deviation, resulting in a mean of 0 and a variance of 1 for each variable. The scaling method required depends on the specific data type. Spectral data, for example, is usually only mean-centered and not scaled to unit variance to avoid noise amplification [90]. Furthermore, transformation may be required if variables are not normally distributed, or if their distribution is highly skewed, for example, logarithmic or negative logarithmic transformation. The specific preprocessing steps required highly depend on the data. The specialized approaches required for spectral data will be discussed later.

Beyond preprocessing, rigorous model validation is essential to ensure that models are robust and generalize to new data. Specifically, validation procedures are required that avoid data leakage during model training and validation, and thereby prevent overly optimistic model performances. Cross-validation is a commonly used resampling validation technique [91]. In k-fold cross-validation specifically, the dataset is divided into k groups (folds). The model is then trained k times, where each time k-1 folds are used for training and the remaining fold for validation. This procedure ensures that every observation is used for validation exactly once, and only when being excluded from the corresponding training set. In bioprocessing, leave-one-group-out (LOGO) cross-validation is commonly applied to ensure that data from the same batch are not split across multiple groups/folds, avoiding data leakage between training and validation sets. However, as will be discussed later, even rigorous LOGO cross-validation strategies cannot prevent all forms of systematic model bias.

The performance of multivariate models is typically quantified using the coefficient of determination (R^2), which is a measure for how well the model fits the data, and the coefficient of prediction (Q^2) for cross-validated model predictivity [92]. For regression models, the root mean square error (RMSE) provides another frequently used prediction accuracy metric, notably expressed in the same unit as the predicted parameter. There are three variants of RMSE that are commonly used: RMSE of estimation (RMSEE) for the training/calibration set, RMSE of cross-validation (RMSECV) for cross-validation, and RMSE of prediction (RMSEP) for the independent test set. The RMSE is calculated as:

$$\mathrm{RMSE} = \sqrt{\sum_{i=1}^{N} \frac{(y_i - \hat{y}_i)^2}{N - P}} \tag{1}$$

where N denotes the number of samples, P the parameter estimates, y_i the measured reference value, and $\hat{y}_i$ the predicted value. Lower RMSE values are favorable as they indicate better predictive performance. After model training (RMSEE) and validation (RMSECV), the prediction accuracy of the model on unseen data is evaluated using the test set (RMSEP), which provides an unbiased assessment of how the model will perform when deployed in production. This final evaluation step is essential, as cross-validation metrics are susceptible to overoptimistic results, given that it is often used for hyperparameter optimization (e.g., number of latent variables in PLS/OPLS). The test set, however, remains completely isolated from model development, and thus serves as the ultimate benchmark for assessing model generalization and performance.

2.4.3 Multivariate Calibration

An essential MVDA technique for spectroscopic applications is multivariate calibration, which relates high-dimensional spectral data to process parameters through regression. In biopharmaceutical manufacturing, spectroscopic measurements are often related to CPPs and CQAs through PLS or OPLS calibration models. These models are essential for translating spectroscopic readings into actual process parameter values, and therefore highly relevant for process monitoring and control as part of the QbD and PAT frameworks.

Typical calibration model development is based on multiple subsequent steps. Initially, the appropriate calibration data needs to be generated. As previously discussed, this process usually involves DOE to adequately cover the process variability and reduce confounding effects. When executing the experimental design, respective spectroscopic data (e.g., Raman spectra) in combination with the response data (e.g., CPP or CQA off-line analyzer data) need to be collected. In the second step, the acquired data are analyzed.

Before performing multivariate calibration, it is essential to preprocess Raman spectral data in order to eliminate artifacts and enhance chemical information. Typical preprocessing procedures include removing cosmic ray and fluorescence background, smoothing for noise reduction, and normalizing spectra to improve comparability across measurements. Baseline correction techniques like wavelet transform [93] and asymmetric least squares (AsLS) [94] are frequently used for background removal. Smoothing filters like Savitzky-Golay (SG) smoothing [95], moving average smoothing, or median filtering are often utilized to lower spectral noise. Peak area and peak intensity normalization, and standard normal variate (SNV) represent typical normalization techniques, aiming to improve spectral comparability. Further, derivatives, such as first-order, second-order, or SG-derivatives, are frequently employed, as they can simultaneously remove background effects while enhancing spectral features through smoothing. Identifying optimal preprocessing is nontrivial, given that most filters require hyperparameter optimization, while further the order of applying filters plays a critical role. Therefore, robust optimization and validation pipelines are required.

After preprocessing, calibration modeling is initiated. In that step, usually PLS or OPLS are used to regress spectroscopic data onto the responses of interest. Great emphasis is placed on the model validation aspect to ensure robust calibration models with high predictive ability. The model validation procedure requires splitting the initial data into a training set and a test set. The model is then fitted using the training set and cross-validation procedures [96]. However, during multivariate calibration of spectral data, distinct challenges in terms of model specificity and robustness arise, which can be greatly impacted by confounding effects, as discussed before. Weak calibration models may exploit confounding structures to achieve high predictivity during model validation, despite lacking genuine analyte-specificity. Standard validation tech-

niques, including Q^2 from LOGO cross-validation, fail to detect this, producing misleadingly high performance metric values for fundamentally flawed models. A prominent example of this is Raman-based modeling of Raman-inactive analytes, such as sodium or potassium ions, which can achieve high predictivity by exploiting correlations with Raman-active species [52]. In reality, however, such models are fragile and function only as long as the underlying correlation structure does not change, either through unexpected process deviations or intentional changes.

Paper IV addresses these validation gaps through a novel validation approach, called TAPE, which decomposes model performance into time-dependent (between-timepoint) and time-independent (within-timepoint) parts. TAPE detects models that obtain their predictivity from time-dependent correlations with confounders, rather than being truly analyte-specific.

As part of model training, variable selection is often performed. Variable selection describes the process of selecting specific variables from the X-block while removing non-relevant or noisy ones. This process can be knowledge-driven, data-driven, or both. Knowledge-driven variable selection uses prior information about the process/system. For example, specific Raman spectral bands which are directly associated with specific vibrational modes of the analyte of interest may be selected, while regions known to be noisy or irrelevant may be excluded [97]. Data-driven variable selection, in contrast, relies on algorithms that select variables based on statistical metrics. Typically, an initial model is fitted using all available variables, from which statistical metrics are derived to guide further algorithmic selection [98].

However, variable selection in calibration modeling has drawbacks. The removal of variables can lead to loss of valuable chemical information encoded in these spectral region, which increases the risk of model overfit and reduced generalization to new samples. A reduced variable set also increases the chances of model failure from instrumental artifacts, given that the averaging effect across multiple correlated variables is substantially reduced. This can further limit the transferability of models across instruments. Moreover, selected variables may become highly specific to training conditions, compromising the robustness that full-spectrum approaches may maintain through redundant information across spectral regions. Therefore, variable selection needs to be treated with caution, often requiring special attention during model validation [98].

Despite these limitations, variable selection is a frequently used tool in calibration modeling. Variable importance in projection (VIP) [99, 100] is a widely used method in PLS and OPLS modeling. VIP assigns scores to variables based on their contribution to explaining Y-block variance. A general rule of thumb states that variables with VIP scores > 1 are retained for subsequent modeling. The selectivity ratio (SR) [101] represents another popular data-driven method, which calculates the ratio between explained and residual variance for each X-block variable in the target projection. Again, a cut-off threshold is required, which is typically determined using an F-test [102]. Although there are numerous methods for selecting variables,

no single approach is universally optimal, and performance of each method may vary on a case-by-case basis.

Paper I introduces a novel variable selection method for Raman calibration modeling, the importance ratio (IR) method [103]. The IR method calculates the importance ratio of the absolute predictive OPLS loading to the jackknife standard error of cross-validation for each variable. Variables with high importance ratios have strong correlation to the Y-response (large absolute predictive loading) or high confidence in their predictive loading (small jackknife standard error), while low ratios indicate weak correlation (small absolute predictive loading) or low confidence (large jackknife standard error). The study systematically compares VIP, SR, and IR methods for Raman calibration models, demonstrating that IR's cross-validation-based approach is superior in excluding spectral regions associated with confounding analytes. Hence, in this case, the model specificity was improved by reducing the influence of confounded analytes.

With appropriate preprocessing, variable selection, and validation strategies in place, calibration models can be deployed for routine process monitoring. However, deployment marks the start of the model lifecycle rather than its conclusion, as continuous maintenance is necessary to guarantee continued performance. Process drifts or modifications, such as raw material variations, changes in media formulation, or feeding strategies, are common reasons for degrading prediction ability over time. Therefore, the performance of calibration models needs to be monitored regularly, which can involve multiple complementary strategies, either individually or in combination. First, prediction errors can be tracked over time by comparing new predictions against reference measurements. Second, standard multivariate outlier detection tools, such as DModX (distance to model in X-space) and Hotelling's T^2, can be used tp identify unusual spectral variations or samples that fall outside the model's calibration domain. Third, periodic model validation studies to analyze a subset of samples by comparing model prediction with reference measurements can be implemented. In case the performance metric values exceed predefined thresholds, corrective actions are required to ensure good model performance in the future. Such action might include model recalibration with new data or even complete model rebuilding.

The practical implementation of spectral calibration models in biopharmaceutical manufacturing faces several fundamental challenges. As discussed before, the generation of proper calibration data at manufacturing scale can be impractical and even infeasible, highlighting the benefit of representative small-scale solutions. Although such systems enable streamlined calibration data generation at significantly lower costs, they introduce other challenges related to scalability and model transferability. Model transferability represents a critical hurdle in multivariate calibration, which is relevant in various ways. Given the impracticability of generating calibration data at manufacturing scale and the burden of performing experiments in

general, the reuse of calibration data across different scales or applications has seen increased attention in recent years. For example, *Berry et al.* have investigated the transferability of Raman monitoring models in CHO fed-batch culture. The authors found that models for glucose, lactate, and osmolality transferred well across 5, 200 and 2000 L scales, while glutamate, ammonium, VCC, and TCC required combination-scale models to achieve good predictivity in the 2000 L manufacturing scale. The results demonstrated that scale dependencies exist, although they are specific for each analyte/parameter [104]. Recently, another study demonstrated the transferability of Raman models in perfusion culture, with the Raman probe integrated in the cell-free harvest line. This approach enabled successful model transfer from 2 and 10 L to 50 and even 500 L bioreactor scales for various analytes [56].

Beyond scale transfer, the transferability of calibration models between different cell lines of the same platform is often investigated [105, 106]. This is particularly interesting for companies that rely on platform processes for their biopharmaceutical production. For example, *Webster et al.* investigated the transfer of Raman calibration models trained on two cell lines of the CHOK1SV GS-KO^{TM} platform to a third one of the same platform. Again, the results were analyte-dependent: while glucose, lactate, ammonium, VCC and TCC models retained high predictivity, models for glutamate and titer showed considerable predictivity loss when tested on the third cell line [105]. For bio-capacitance-based calibration models, successful scale and cell line transfer for VCC measurements in CHO culture has also been demonstrated, although one-point calibration for off-set correction was still required [107].

The aforementioned challenges motivated the four studies comprising this thesis. Together, these works address critical gaps in the implementation of robust calibration models across different bioprocess applications. **Paper I** investigates the impact of clonal-level biological variations on model transferability in CHO fed-batch culture. **Paper II** demonstrates in-line deployment of bio-capacitance-based monitoring with predictive forecasting for rAAV vector production in HEK batch culture. **Paper III** establishes an automated workflow for efficient calibration data generation in miniature-scale perfusion culture. **Paper IV** introduces a validation framework to detect models that exploit time-dependent confounding rather than being genuinely analyte-specific. The following chapter presents each study in detail.

3 Results and Discussion

3.1 Paper I: Feasibility and performance of cross-clone Raman calibration models in CHO cultivation

Raman spectroscopy has become an established PAT tool for real-time monitoring in biopharmaceutical manufacturing. While usage of Raman spectroscopy in commercial production with stable cell lines is well-documented, deployment in highly dynamic environments such as cell line development has not been investigated yet. In cell line development substantial biological variations are common, with individual clones exhibiting diverse productivity levels, growth rates, and metabolic profiles. The impact of these on Raman calibration model performance and transferability across clones are still unclear. Paper I addresses this knowledge gap by evaluating the transferability of Raman calibration models across eleven single-cell line CHO clones with large biological variations.

Aim The main aim of Paper I was to evaluate the transferability of Raman-based OPLS calibration models across clones of the same CHO cell line, with large diversity in productivity levels, growth rates, and metabolic profiles. The specific aims included: (1) developing calibration models for glucose, lactate, and titer across multiple clones; (2) assessing whether developed calibration models exhibit any clone-related bias; (3) investigating whether variable selection methods (VIP, SR, and IR) can mitigate clone-related bias; and (4) assessing how productivity bias contributes to the clone-related bias of the titer model.

Results and Discussion Eleven CHO-DG44 clones producing the same IgG antibody were cultivated in Ambr® 250 mL bioreactors over 14 days in fed-batch mode. The clones exhibited substantial biological variation (Paper I, Figure 1), with final titer values ranging from below 2.5 g/L (low producers) to above 2.5 g/L (high producers). A PCA of Raman spectral data revealed similar trajectories during early cultivation (days 0-3), while spectral variation increased considerably with process maturity. This reflected an increasing impact of biological variation on the culture media and Raman spectra with time (Paper I, Figure 1; SI Figure S6).

Initial OPLS calibration models based on the Raman spectral range from 450 to 1800 cm^{-1} demonstrated good cross-validated predictivity. Glucose and lactate models achieved RMSECV values of 0.17 g/L and 0.13 g/L with $R^2CV > 0.97$, while the titer model showed RMSECV and R^2CV of 0.29 g/L and 0.94 respectively. However, the RMSEE/RMSECV ratio for the titer model (0.73) indicated potential overfitting or systematic bias (Paper I, Table 1).

To specifically assess the models for clone-related bias, clone-wise LOGO cross-validation was used, in which each clone formed a unique cross-validation group. Subsequently, a linear regression analysis was performed on the observed versus predicted values for each clone individually.

This analysis revealed substantial differences among the investigated analytes. Glucose and lactate models demonstrated linear regression slopes close to the ideal value of 1.0 for all clones (glucose: 0.91 to 1.09 and lactate: 0.89 to 1.01), with only one exception. This indicated that glucose and lactate models exhibited no systematic clone-related bias, with the possibility of transfer between different clones. In contrast, the titer model showed pronounced clone-related bias with slopes ranging from 0.71 to 1.20. Notably, low-producing clones generally exhibited lower slopes than high-producing clones (Paper I, Figure 2C; SI Tables S5), indicating that the productivity level related to the observed effect. The analysis revealed that the aggregate performance metrics ($R^2CV = 0.94$) were masked by clone-specific prediction accuracy. Besides productivity, this bias was also correlated with the TCC profiles. This suggested that cell density variations might influence Raman signal intensities through changes in the optical sampling volume and thereby contribute to the bias.

To investigate whether a lack in mAb specificity caused the clone-related bias, three variable selection methods were applied to the initial titer model: VIP, SR, and the IR method. Only the IR method, which incorporates cross-validation-based confidence estimates, accurately detected and excluded the Raman band at 855 cm^{-1}, which is strongly associated with the lactate carboxyl stretching modes (Paper I, Figure 3). Further analysis revealed that by exclusion of this Raman band, the specificity of the titer model was substantially increased, while the influence of lactate-related confounding was reduced (Paper I, Figure 4). Despite these improvements, the clone-related bias of the titer model persisted, with productivity-dependent patterns remaining evident.

Following this, the contributions of different productivity levels to the clone-related bias were assessed by developing separate high- and low-producer titer models. Their prediction performance was evaluated on the respective opposite productivity group. This worst-case scenario represented model extrapolation beyond training productivity ranges and exemplified the risk of not sufficiently accounting for broad productivity ranges during model training. The results showed that high-producer models systematically overestimated the titer for low-producing clones (RMSEP = 0.49 g/L), while low-producer models substantially underestimated the titer for high-producing clones (RMSEP = 0.90 g/L, Paper 1, Figure 5; SI Tables S7-S8). These RMSEP values represented considerable increases compared to the cross-validated error of the initial titer model containing all clones (RMSECV = 0.29 g/L). This demonstrated that clonal productivity was a major contributor to the observed clone-related bias of the titer model.

Conclusion This study establishes important boundaries for Raman calibration model transferability in the presence of biological variations on a clonal level. Glucose and lactate models demonstrated robust transferability across diverse CHO clones, showing no clone-related bias and maintaining high predictivity regardless of clonal productivity or growth characteristics. Therefore, the implementation of Raman-based monitoring in cell line development settings

where biological variation is inherent should be feasible for these analytes. Further, the results suggest that data from early cell line development could be utilized for calibration modeling in manufacturing applications.

Conversely, titer models exhibit significant clone-related bias, strongly associated with clonal productivity. While the IR variable selection methods helped to increase the titer models' specificity, it could not eliminate the observed bias. For industrial implementation, this suggests that generic titer models must incorporate cell culture data with diverse productivity levels. Generic calibration model approaches have been demonstrated in different contexts [104, 105, 108], with this study extending the understanding to biological variations within a single CHO cell line platform.

3.2 Paper II: Real-time VCC monitoring and forecasting in HEK-cell-based rAAV vector production using capacitance spectroscopy

Abstract rAAV vectors have emerged as a leading tool in gene therapy, with HEK293 cell platforms representing the favorable expression system. The effective production of rAAV vectors relies critically on the transfection step, which must be timed precisely to occur at specific VCCs. The traditional VCC monitoring strategy, based on manual sampling and off-line analysis, considerably limits the availability of real-time process data and thereby reduces the chances of timely transfection. Bio-capacitance-based sensors represent an alternative approach, providing continuous, non-invasive VCC measurements during cell culture. Although this technique has been demonstrated in various applications, its implementation for real-time VCC monitoring and forecasting in HEK-based rAAV vector production remains unexplored. Paper II addresses this gap by developing and deploying bio-capacitance-based VCC models directly for rAAV-8 vector production. In addition, novel forecasting models are deployed to enable optimal transfection timing.

Aim The main aim of Paper II was to develop and deploy bio-capacitance-based models for real-time VCC monitoring and forecasting in HEK293-based rAAV vector production. The specific aims included: (1) comparing single-frequency and multi-frequency (OPLS) modeling approaches for VCC prediction; (2) deploying models in-line to provide real-time VCC predictions to the bioprocess control system; (3) developing a Time-Till-Transfection (TTT) forecasting model to predict optimal transfection timing; and (4) validating both monitoring and forecasting capabilities on two independent rAAV-8 vector production batches.

Results and Discussion Six independent rAAV-8 vector production batches with a transfection target VCC of 1.5×10^6 cells/mL were performed in 10 L bioreactors. During the cell cultivation, bio-capacitance data were recorded across 25 discrete frequencies (50-20,000 kHz). Further, regular sampling followed by off-line analysis was performed to generate VCC reference

values for subsequent calibration modeling. With the data in hand, single-frequency (580 kHz) and multi-frequency (OPLS) calibration models were developed using four batches as training data. The models demonstrated comparable cross-validated performance with R^2CV values of 0.95 and 0.96 and RMSECV values of 0.16 $\times$ 10^6 cells/mL (both) respectively (Paper II, Figure 2A). This similar performance was attributed to minimal cell diameter changes (<2 µm) throughout the rAAV vector process, which was confirmed by off-line measurements (Paper II, SI Figure S1A). While multi-frequency approaches tend to outperform single-frequency models in processes with substantial cell size changes through access to the full β-dispersion curve, in this case, the single-frequency approach proved sufficient.

With the goal of real-time VCC monitoring in mind, both models were deployed in-line and tested on two independent batches. The results showed robust performance with RMSEP values of 0.11 to 0.17 x 10^6 cells/mL for the OPLS model and 0.15 to 0.17 $\times$ 10^6 cells/mL for the SF model (Paper II, Figures 2B-D). Notably, these prediction errors were comparable to the inherent measurement uncertainty of the off-line reference itself.

After real-time VCC monitoring was established, the Time-Till-Transfection (TTT) forecaster was developed to address precise transfection timing prediction. This is critical in rAAV vector production as reagent preparation must be completed within constrained time windows prior to transfection. Therefore, mere real-time monitoring of VCC is not sufficient for a reliable transfection protocol. The TTT forecaster employed a novel stacked-model architecture: continuous VCC predictions from the OPLS model served as input to an exponential growth model with a 15-hours sliding regression window, enabling real-time forecasting of the time remaining until the transfection target was reached.

The TTT forecaster was deployed in-line for the two independent test batches. Initially, the forecasting model over-estimated the remaining time until transfection due to the lag phase growth dynamics. However, predictions stabilized within 24 hours post-inoculation. From 24 hours and onward, the forecasting accuracy remained within ±3 hours of the actual transfection time until the target VCC was reached (Paper II, Figure 3). This prediction error of ±3-hours corresponded to a deviation of ±0.09 $\times$ 10^6 cells/mL in the target VCC, assuming a growth rate of 0.03 $\times$ 10^6 cells/(mL·h). This accuracy was well within acceptable limits for successful transfection timing. Hence, the TTT forecaster provided accurate transfection timing predictions well in advance, enabling timely reagent preparation and assuring process repeatability.

Conclusion This study demonstrates successful in-line deployment of bio-capacitance-based VCC monitoring and forecasting for HEK293-based rAAV vector production. Single- and multi-frequency models achieved similar and sufficient prediction accuracy, comparable to the reference method uncertainty itself. The single-frequency model proved accurate enough for

this process due to minimal cell size changes. The stacked-model architecture of the TTT forecaster, combining OPLS with exponential growth models, enabled accurate transfection timing predictions one full day in advance. The developed approach served as a predictive decision support system for transfection initiation, addressing a critical operational challenge in transient rAAV vector production.

3.3 Paper III: Advancing Raman calibration: Automated data generation, monitoring, and control in multi-parallel perfusion mini bioreactors

Abstract Perfusion is considered a key process intensification strategy for biopharmaceutical manufacturing due to offering higher volumetric productivity and continuous steady-state operation compared to conventional fed-batch culture. However, the dynamic nature of perfusion processes, which entails continuous medium exchange, cell retention, and elevated cell densities, presents significant challenges for PAT implementations. Raman spectroscopy has been presented as a useful PAT tool recently, providing real-time, non-invasive analyte monitoring capabilities. However, its implementation requires robust calibration models that account for the unique operational complexity of perfusion culture. A critical bottleneck in developing these models is the generation of sufficient calibration data covering adequate process variability. This has proven to be impractical at production scale due to high costs and operational constraints. Paper III addresses this challenge by demonstrating an automated data generation workflow for Raman calibration modeling in miniature-scale perfusion culture. Further, it demonstrates the development of generic models across multiple cell lines, mAb products, and process conditions, and validates Raman-based bleed-rate control by comparison to the conventional approach.

Aim The main aim of this study was to develop and validate an automated workflow for Raman calibration data generation in perfusion culture using a high-throughput miniature bioreactor system. The specific aims included: (1) establishing an automated calibration data generation workflow; (2) developing generic OPLS calibration models for glucose, lactate, titer, glutamate, and VCC spanning multiple CHO cell lines, mAb products, and process parameters; (3) validating model performance on independent test batches; and (4) demonstrating Raman-based bleed-rate control by comparing its performance to conventional cell counter-based approaches.

Results and Discussion Fifteen perfusion cultivations were performed in the Ambr® 250 HT Perfusion system with three CHO-DG44 cell lines producing three distinct mAb products. The culture conditions were systematically varied to ensure sufficient process variations, with target VCCs ranging from 25 to 100 $\times$ 10^6 cells/mL, CSPRs from 10 to 50 pL/cell/day, and additional changes in batch phase media, cross-flow rates, and temperature shifts (Paper III, Table 1). Varying the culture conditions led to substantial differences in nutrient, metabolite,

cell count, and titer profiles, as intended (Paper III, Figure 2).

In the first step, a PCA of the preprocessed Raman spectra was performed (Paper III, Figure 3). The analysis of the scores plot revealed distinct clustering patterns: In early cultivation (days 0-3), the individual culture trajectories clustered according to the three batch phase media. Following perfusion initiation and medium exchange (days 3-5), these media-specific clusters converged, indicating that the exchange to the perfusion medium was accomplished. After that, spectral variation was mostly driven by target VCC and CSPR rather than the specific cell lines. Interestingly, fast-growing cultures displayed characteristic U- or V-shaped trajectories similar to the ones found in fed-batch culture [53], while slow-growing cultures showed more irregular patterns. Notably, target VCC emerged as a dominant factor influencing spectral trajectories. This was demonstrated by comparing cultivations with identical CSPR but different VCC targets, which showed clear trajectory separation as VCCs diverged (Paper III, SI Figure S3).

Subsequently, OPLS calibration models were developed for five individual analytes. Here, LOGO cross-validation was used to avoid data leakage during model training and validation. The cross-validated model performance varied by analyte: the glucose model achieved excellent predictivity (RMSECV = 0.39 g/L, R^2CV = 0.98); lactate's showed good performance (RMSECV = 0.20 g/L, R^2CV = 0.88); titer's demonstrated moderate performance (RMSECV = 0.16 g/L, R^2CV = 0.76); glutamate's showed good predictivity (RMSECV = 0.37 mmol/L, R^2CV = 0.80); and VCC's achieved good performance (RMSECV = 5.6×10^6 cells/mL, R^2CV = 0.92) (Paper III, Table 2, Figure 4). The relatively low predictivity of the titer model was attributed to the inherently low mAb concentration range reached. Low analyte concentrations are a general hurdle for calibration modeling in perfusion culture - a result of continuous medium exchange and product removal.

The model validation in an independent culture demonstrated that the cross-validated performance translated well to new data, confirming robustness and generalization capabilities (Paper III, Table 2, Figure 5). A performance comparison of our models with ones presented in literature, for both cell-containing (bioreactor samples) and cell-free (permeate samples) Raman calibration models in perfusion, showed strong alignment for glucose, lactate, and titer. However, glutamate and VCC models performed slightly worse than some reports, which was explained by the generic nature of our models that span multiple cell lines and products versus the single-platform literature examples (Paper III, Table 3).

To further demonstrate practical applicability, Raman-based VCC predictions were deployed as input for the bleed-rate control system in another cultivation. In parallel, a second culture was performed under equal process conditions, but with the conventional cell counter-based approach as input for bleed-rate control. Both cultivations operated at VCC targets of

40×10^6 cells/mL (days 0-19), followed by a target of 60×10^6 cells/mL from day 19 and onward. The Ambr® bleed-rate control algorithm continuously adjusted the bleed based on the respective VCC inputs.

The results showed that Raman-based control achieved a mean VCC of $(38.2 \pm 7.19) \times 10^6$ cells/mL at the 40 million target (mean absolute error (MAE) = 5.97×10^6 cells/mL) and $(58.37 \pm 6.42) \times 10^6$ cells/mL at the 60 million target (MAE = 5.13×10^6 cells/mL). The cell counter-based control showed slightly tighter performance at the lower target ($(40.88 \pm 4.59) \times 10^6$ cells/mL, MAE = 3.90×10^6 cells/mL), likely due to more accurate VCC measurements (Paper III, Table 4, Figure 6). However, the higher target was not reached for the cell counter-based control cultivation due to contamination. Compared to literature, the single available example demonstrating Raman-based bleed-rate control in CHO culture [61] reported tighter VCC control around a 50×10^6 cells/mL target ($(50.67 \pm 2.67) \times 10^6$ cells/mL). However, their approach utilized continuous measurements via Raman immersion probes in 3 L bioreactors, enabling near-real-time adjustments through PID-based feedback control, as opposed to our 12 hour at-line sampling approach. Additionally, the literature example developed models on a single cell line, in contrast to our generic multi-cell-line approach. Given these differences, the achieved control performance is considered satisfactory and demonstrates the feasibility of Raman-based bleed-rate control in miniature perfusion culture.

Conclusion This study establishes an automated workflow for Raman calibration data generation in miniature-scale perfusion culture. As such, it represents the first example of automated multi-parallel calibration data generation for perfusion culture without the need for multiple Raman probes. Generic Raman models spanning three cell lines, three mAb products, and diverse culture parameters demonstrated acceptable predictivity on independent test data and enabled successful Raman-based bleed-rate control.

3.4 Paper IV: Time-adjusted performance evaluation (TAPE) of predictive multivariate models for bioprocess data

Abstract Multivariate calibration models are essential in modern-day biopharmaceutical manufacturing, as they translate raw sensor data into actual process-related information. Specifically, such models are used to relate spectroscopic data into key performance indicators, CPPs or CQAs, enabling real-time process monitoring and control strategies. Before any calibration models are deployed for use in commercial production, they need to be properly validated. However, conventional validation metrics, like the cross-validated coefficient of prediction (Q^2), consistently miss a crucial problem: models that achieve high predictivity by taking advantage of time-dependent correlations instead of capturing true analyte-specific relationships. This is particularly apparent in cell culture processes, where analytes are inherently correlated with each other and elapsed process time, due to metabolic pathways. These effects lead to confounding structures, which weak models can exploit to achieve overly optimistic validation metric values, despite lacking specificity for the target analyte. Therefore, apparently robust models may fail when confounding structures change, which bears substantial risks for commercial biopharmaceutical manufacturing. This study introduces a novel validation framework, called Time-Adjusted Performance Evaluation (TAPE). TAPE decomposes (OPLS) model performance into time-dependent and time-decoupled parts, which helps to detect models that obtain predictivity from time-dependent confounding rather than genuine analyte-specificity.

Aim The main aim of this study was to develop a validation framework capable of detecting models deriving their predictivity from confounding effects, which conventional validation methods fail to identify. The specific aims included: (1) establishing a methodology to decompose predictive performance into time-dependent ($Q^2_{between}$) and time-decoupled (Q^2_{within}) parts through per-timepoint centering; and (2) demonstrating TAPE's capability of detecting flawed models lacking analyte-specificity for multiple data sources (Raman spectroscopy, metabolomics, transcriptomics).

Results and Discussion The TAPE framework is based on the principle that genuine analyte-specific calibration models should maintain predictivity even after removing time-dependent variation. In practice, TAPE calculates two complementary performance metrics: $Q^2_{between}$, quantifying model predictive ability only using confounding time-dependent variation and Q^2_{within}, quantifying predictivity after removing systematic time-driven artifacts through per-timepoint centering. Ideally, calibration models with genuine analyte-specificity should maintain high Q^2_{within}, whereas models exploiting time-dependent correlations should demonstrate substantial reduction in Q^2_{within} despite high $Q^2_{between}$ values.

TAPE was applied to Raman calibration models from two distinct CHO fed-batch cultures, covering calibration models for glucose, lactate, K^+, and NH_4^+ analytes. While glucose and lac-

tate molecules are known for strong Raman activity, NH_4^+ has very weak Raman signals, and K^+ is not Raman-active at all. The conventional LOGO cross-validation resulted in good to excellent prediction performance for all models ($Q^2 \geq 0.70$), with only NH_4^+ showing moderate predictivity ($Q^2 = 0.54$) in one data set. Therefore, the results of the conventional validation method proved no indication of potential issues.

However, TAPE analysis revealed striking differences between different analytes (Paper IV, Figure 3 and SI Figure S4). The Raman calibration models for glucose and lactate demonstrated Q^2_{within} values of 0.95 and 0.89 in the first data set, and 0.89 and 0.63 in the second, respectively. On the other hand, the K^+ and NH_4^+ calibration models resulted in substantially lower Q^2_{within} values, with 0.28 and 0.31 in the first data set, and 0.41 and 0.03 in the second, respectively. The combination of low Q^2_{within} and high Q^2_{between} values confirmed that K^+ and NH_4^+ models exploited time-dependent correlations. These results demonstrated that only analyte models with strong Raman activity retained their predictivity after removing time-dependent confounding artifacts.

TAPE was further used for validation of predictive models based on metabolomic and transcriptomic cell culture data. The results showed that all models with good to moderate predictivity according to the conventional validation method, were deemed invalid by TAPE, except for the transcriptomic-based VCC model. According to these results, deriving accurate relationships between omics data and cellular biochemistry or CQAs through predictive modeling proved to be rather challenging.

Conclusion This study establishes TAPE as an essential validation framework for multivariate calibration models in bioprocess applications. The novel framework addresses a critical problem in conventional validation methodologies, which fail to detect models exploiting time-dependent confounding leading to apparent good model predictivity. TAPE exposes these vulnerabilities during model development, enabling informed decisions about robust model deployment, required constraints, or need for alternative measurement approaches. Thereby, TAPE provides a critical methodological advancement for multivariate calibration model validation, ensuring robust PAT implementation in QbD-driven biopharmaceutical manufacturing.

4 Conclusion and Future Perspective

This thesis investigated several critical aspects of multivariate calibration for robust PAT implementation in biopharmaceutical manufacturing. We found that strong biological variations can restrict Raman-based monitoring in CHO fed-batch culture. Even though these results were analyte-dependent, they highlight that Raman calibration models are susceptible to model bias, stressing the importance of proper validation strategies. However, developing generic Raman models for use in applications with large biological variation is possible, contingent on the fact that models cover sufficient process variations during training (**Papers I and III**). Further, we demonstrated how PAT-based real-time monitoring models can be utilized to achieve accurate process forecasting. Combining standard chemometric modeling techniques (OPLS) with mechanistic models proved to be a simple yet effective strategy, yielding a predictive decision support system for data-driven process handling (**Paper II**). Moreover, we presented workflows for streamlined automated calibration data generation through small-scale high-throughput bioreactor systems with PAT integrations. These systems allow for calibration data generation at substantially lower costs, while also giving the opportunity for systematic process parameter changes (e.g., through DOE), ultimately enabling more specific and robust calibration models (**Paper III**). Lastly, we demonstrated that conventional validation frameworks fail to detect calibration models that exploit time-dependent confounding structures, thereby reaching apparently high predictivity despite lacking genuine analyte-specificity. A novel model validation framework is presented that can identify such flawed models and prevent them from being deployed in commercial biopharmaceutical manufacturing (**Paper IV**).

This thesis offers considerable industrial implications. While **Paper I** demonstrated that the IR variable selection method could reduce the influence of confounding effects on model predictions, it also revealed a strong productivity-dependent bias in Raman-based titer models. Although this may not directly affect Raman models deployed in commercial production, it greatly affects applications with larger biological variation, such as in cell line development. Future studies could investigate the mechanistic basis of the observed productivity-dependent bias. Until clarified, generic titer model development for such dynamic environments must account for broad productivity levels, covering at least the ranges expected in the specific application. Alternatively, a hierarchical modeling approach could be investigated, classifying clones into specific productivity groups first (e.g., through OPLS-DA), and then performing quantitative titer assessment using appropriate group-specific models.

The TTT forecaster developed in **Paper II** represents a unique value proposition for industrial applications. The advent of progressively complex biopharmaceutical manufacturing processes involving time-critical steps - such as perfusion initiation, transfection, infection, product harvest, or cell lysis - necessitates capabilities extending beyond simple process monitoring. The TTT forecaster demonstrates how classical chemometric modeling techniques coupled with sim-

plified mechanistic models can enable accurate timeline estimation for future events. Notably, the forecaster is sensor-agnostic, meaning the approach can be transferred to any (continuous) VCC source, beyond bio-capacitance-based OPLS models. Follow-up studies could investigate transferring this stacked-model architecture to more sophisticated mechanistic or hybrid models, potentially enabling forecasts of nutrients, products, and by-products, and informing new process strategies. For rAAV vector production specifically, fully automated transfection could be implemented, contingent on automated transfection reagent generation or close-by storage capabilities.

The automated calibration data generation workflow presented in **Paper III** establishes the foundation for robust calibration model development in perfusion culture. It provides practical guidelines for which specific process parameters need to be considered during Raman calibration data generation. This workflow can be utilized in the future for DOE-based studies and data generation experiments including diverse PAT analyzers. Thus, effective coverage of the design space and reduced confounding effects at substantially lower costs can be achieved. Follow-up studies could further explore the transferability of the developed calibration models from miniature to production scale.

Paper IV potentially carries the strongest implications for future PAT implementation in biopharmaceutical manufacturing. Beyond the potential for developing models with higher analyte-specificity, the TAPE framework could influence regulatory PAT method validation requirements. A potential strategy involves establishing industry-standard TAPE thresholds for specific PAT analyzers and cell culture analytes/parameters, mitigating the risk of flawed models entering the commercial manufacturing stage. Future extensions could investigate model validation against non-temporal confounders (e.g., cell lines, clones, culture media).

This thesis demonstrates that robust PAT implementation in biopharmaceutical manufacturing is not merely a matter of model deployment, but rather of proper data generation, model development, and validation strategies. The systematic investigation of model specificity, confounding effects, calibration data generation, and validation vulnerabilities revealed that the primary limitations are methodological gaps in multivariate calibration model development and validation. While more advanced modeling techniques and sensor fusion will undoubtedly play important roles in future bioprocess control strategies, this thesis establishes that rigorous chemometric frameworks are necessary to ensure that models are genuinely specific, properly validated, and appropriately constrained if needed. The provided studies and their methodologies address these foundational challenges, and thereby contribute to successful PAT implementation in modern-day biopharmaceutical manufacturing.

Acknowledgements

First, I would like to thank **Johan** for the opportunity to conduct my PhD research in his group. Thank you for your support throughout this time, your availability for questions, your pragmatic approach to solving issues, and your commitment to enabling travel even during times of cost savings! A special thanks to **Izabella** for supervising my work from start to finish, always being available with the highest levels of effort and passion to ensure our joint success! I truly appreciate your role in this thesis, but even more so, I appreciate the supportive, hard-working, and loyal person that you are!

I would like to thank all the collaborators from Sartorius and UmU who have made the individual projects a success: **Max, Pär, Andreas, Marek, Erik, Steffi, Kleanthis, Christoph, Christian, David, Suneetha, Ajith, Magda, Jonas, Sven, Anne, Henrik, and Olivier**. Without you, none of this would have been possible! Great thanks also to Sartorius' PAT department, especially **Kristina, Eva, Lennard**, for your trust, support, and the time we spent together, especially the *fun* lab sessions. A special thanks also to **Edvin** for all the help you provided, especially with everything related to getting this thesis submitted! Thanks to **Kate** and **David** for taking good care of me at UmU!

Ich will auch meiner Familie meinen tiefsten Dank für die mentale, finanzielle, und geistliche Unterstützung im Gebet während meines gesamten Studiums aussprechen. **Mama, Papa, Oma, Melanie, Carina und Liana**, danke, dass ihr immer für mich da wart und seid! Auch meinen Freunden, **Daniel, Steffen, Lukas** und allen Anderen, bin ich dankbar für die guten Zeiten und Ablenkung abseits von Uni und Arbeit!

Finally, I would like to thank **Maria**, my ex-girlfriend, now wife, for the last five years. Thank you for putting up with my weird humor, your consistent support, always believing in me, being an awesome, loving, and forgiving partner, and making sure that our small family always has what it needs. Sorry for the last weeks of writing. To many more years - I love you!

And lastly, thank you to **Jesus**, my creator and savior, the greatest scientist of all time, for all the grace, forgiveness, and blessings I experience through you.

Bibliography

[1] Towards Healthcare, "Biopharmaceuticals market size, growth business strategies," Towards Healthcare Research, Market Analysis Report, 2025, accessed: 23 Oct 2025. [Online]. Available: https://www.towardshealthcare.com/insights/biopharmaceuticals-market-is-rising-rapidly

[2] International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH), "Ich harmonised tripartite guideline Q8(R2): Pharmaceutical development," ICH, Tech. Rep., 2009, accessed: 27 Jul 2025. [Online]. Available: https://database.ich.org/sites/default/files/Q8%28R2%29%20Guideline.pdf

[3] U.S. Food and Drug Administration (FDA), "Guidance for industry: PAT - a framework for innovative pharmaceutical development, manufacturing, and quality assurance," FDA, Tech. Rep., 2004, accessed: 23 Oct 2025. [Online]. Available: https://www.fda.gov/media/71012/download

[4] R.-M. Lu, Y.-C. Hwang, I.-J. Liu, C.-C. Lee, H.-Z. Tsai, H.-J. Li, and H.-C. Wu, "Development of therapeutic antibodies for the treatment of diseases," *Journal of Biomedical Science*, vol. 27, p. 1, 2020.

[5] S. Watanabe, S. Yasuda, T. Kurosawa, M. Tada, T. Ishibashi, and K. Hattori, "Pharmacokinetics of biopharmaceuticals: Their critical role in molecular design," *Pharmaceuticals*, vol. 16, no. 5, p. 679, 2023.

[6] U.S. Food and Drug Administration (FDA), "100 years of insulin," 2022, accessed: 03 Oct 2025. [Online]. Available: https://www.fda.gov/about-fda/fda-history-exhibits/100-years-insulin

[7] F. Boscari and A. Avogaro, "Current treatment options and challenges in patients with type 1 diabetes: Pharmacological, technical advances and future perspectives," *Reviews in Endocrine and Metabolic Disorders*, vol. 22, pp. 217–240, 2021.

[8] M. L. Chiu, D. R. Goulet, A. Teplyakov, and G. L. Gilliland, "Antibody structure and function: The basis for engineering therapeutics," *Antibodies*, vol. 8, no. 4, p. 55, 2019.

[9] C. H. van Dyck, C. J. Swanson, P. Aisen, R. J. Bateman, C. Chen, M. Gee, M. Kanekiyo, D. Li, L. Reyderman, S. Cohen, L. Froelich, S. Katayama, M. Sabbagh, B. Vellas, D. Watson, S. Dhadda, M. Irizarry, L. D. Kramer, and T. Iwatsubo, "Lecanemab in early alzheimer's disease," *The New England Journal of Medicine*, vol. 388, no. 1, pp. 9–21, 2023.

[10] J. R. Mendell, S. Al-Zaidy, R. Shell, W. D. Arnold, L. R. Rodino-Klapac, T. W. Prior, L. Lowes, L. Alfano, K. Berry, K. Church, J. T. Kissel, S. Nagendran, J. L'Italien, D. M.

Sproule, C. Wells, J. A. Cardenas, M. D. Heitzer, A. Kaspar, S. Corcoran, L. Braun, S. Likhite, C. Miranda, K. Meyer, K. D. Foust, A. H. M. Burghes, and B. K. Kaspar, "Single-dose gene-replacement therapy for spinal muscular atrophy," *The New England Journal of Medicine*, vol. 377, no. 18, pp. 1713–1722, 2017.

[11] Fortune Business Insights, "Biopharmaceuticals market size, share and forecast 2024-2032," Fortune Business Insights, Market Analysis Report, 2024, accessed: 23 Oct 2025. [Online]. Available: https://www.fortunebusinessinsights.com/biopharmaceuticals-market-106928

[12] Towards Healthcare, "Global pharmaceutical market size report 2024-2034," Towards Healthcare Research, Market Analysis Report, 2024, accessed: 23 Oct 2025. [Online]. Available: https://www.towardshealthcare.com/insights/pharmaceutical-market-sizing

[13] F. D. Makurvet, "Biologics vs. small molecules: Drug costs and patient access," *Medicine in Drug Discovery*, vol. 9, p. 100075, 2021.

[14] Institute for Clinical and Economic Review, "Targeted immune modulators for rheumatoid arthritis: Effectiveness and value," Institute for Clinical and Economic Review, Evidence Report, 2017, accessed: 23 Oct 2025. [Online]. Available: https://icer.org/wp-content/uploads/2020/10/NE_CEPAC_RA_Evidence_Report_FINAL_040717.pdf

[15] R. Stein, "At $2.125 million, new gene therapy is the most expensive drug ever," NPR, 2019, accessed: 03 Oct 2025. [Online]. Available: https://www.npr.org/sections/health-shots/2019/05/24/725404168/at-2-125-million-new-gene-therapy-is-the-most-expensive-drug-ever

[16] J. A. DiMasi, R. W. Hansen, and H. G. Grabowski, "The price of innovation: new estimates of drug development costs," *Journal of Health Economics*, vol. 22, no. 2, pp. 151–185, 2003.

[17] J. A. DiMasi and H. G. Grabowski, "The cost of biopharmaceutical R&D: is biotech different?" *Managerial and decision Economics*, vol. 28, no. 4-5, pp. 469–479, 2007.

[18] O. J. Wouters, M. McKee, and J. Luyten, "Estimated research and development investment needed to bring a new medicine to market, 2009-2018," *JAMA*, vol. 323, no. 9, pp. 844–853, 2020.

[19] D. V. Goeddel, D. G. Kleid, F. Bolivar, H. L. Heyneker, D. G. Yansura, R. Crea, T. Hirose, A. Kraszewski, K. Itakura, and A. D. Riggs, "Expression in escherichia coli of chemically synthesized genes for human insulin." *Proceedings of the National Academy of Sciences*, vol. 76, no. 1, pp. 106–110, 1979.

[20] Z.-X. Zhang, F.-T. Nong, Y.-Z. Wang, C.-X. Yan, Y. Gu, P. Song, and X.-M. Sun, "Strategies for efficient production of recombinant proteins in *Escherichia coli*: alleviating

the host burden and enhancing protein activity," *Microbial Cell Factories*, vol. 21, no. 1, p. 191, 2022.

[21] G. D. Barone, A. Emmerstorfer-Augustin, and A. Biundo, "Industrial production of proteins with Pichia pastoris - Komagataella phaffii," *Biomolecules*, vol. 13, no. 3, p. 441, 2023.

[22] K. Dammen-Brower *et al.*, "Strategies for glycoengineering therapeutic proteins," *Frontiers in Chemistry*, vol. 10, p. 863118, 2022.

[23] G. Walsh and E. Walsh, "Biopharmaceutical benchmarks 2022," *Nature Biotechnology*, vol. 40, no. 12, pp. 1722–1760, 2022.

[24] E. Tan, C. S. H. Chin, Z. F. S. Lim, and S. K. Ng, "HEK293 cell line as a platform to produce recombinant proteins and viral vectors," *Frontiers in Bioengineering and Biotechnology*, vol. Volume 9 - 2021, 2021. [Online]. Available: https://www.frontiersin.org/journals/bioengineering-and-biotechnology/articles/10.3389/fbioe.2021.796991

[25] S. A. Mendonça, R. Lorincz, P. Boucher, and D. T. Curiel, "Adenoviral vector vaccine platforms in the SARS-CoV-2 pandemic," *npj Vaccines*, vol. 6, p. 97, 2021.

[26] J. M. Schaub, F. L. Kearns, and et al., "Expression and characterization of SARS-CoV-2 spike proteins," *Nature Protocols*, vol. 16, no. 12, pp. 5339–5376, 2021.

[27] Sartorius AG, "Investor presentation Q1 2024," Available at: https://www.sartorius.com/download/45888/ir-presentation-en-web-sag-data.pdf (archived at https://web.archive.org/web/20250919125847/https://www.sartorius.com/download/45888/ir-presentation-en-web-sag-data.pdf), 2024, accessed: 19 Sep 2025.

[28] A. J. Rish, J. K. Drennen, and C. A. Anderson, "Metabolic trends of chinese hamster ovary cells in biopharmaceutical production under batch and fed-batch conditions," *Biotechnology Progress*, vol. 38, no. 1, p. e3220, 2022.

[29] J. Schaub, A. Ankenbauer, T. Habicher, M. Löffler, N. Maguire, D. Monteil, S. Püngel, L. Stepper, F. Stiefel, J. Thoma, A. Unsöld, J. Walther, C. Wayne, and T. Wucherpfennig, "Process intensification in biopharmaceutical process development and production – an industrial perspective," *Physical Sciences Reviews*, vol. 9, no. 9, pp. 2989–3041, 2024.

[30] U.S. Food and Drug Administration (FDA), "Q13: Continuous manufacturing of drug substances and drug products," FDA, ICH Harmonised Guideline, 2023, accessed: 23 Oct 2025. [Online]. Available: https://www.fda.gov/media/165775/download

[31] J.-M. Bielser, M. Wolf, J. Souquet, H. Broly, and M. Morbidelli, "Perfusion mammalian cell culture for recombinant protein manufacturing – a critical review," *Biotechnology Advances*, vol. 36, no. 4, pp. 1328–1340, 2018.

[32] S. Xu, J. Gavin, R. Jiang, and H. Chen, "Bioreactor productivity and media cost comparison for different intensified cell culture processes," *Biotechnology Progress*, vol. 33, no. 4, pp. 867–878, 2017.

[33] A. A. Shukla, B. Hubbard, T. Tressel, S. Guhan, and D. Low, "Downstream processing of monoclonal antibodies - application of platform approaches," *Journal of Chromatography B*, vol. 848, no. 1, pp. 28–39, 2007.

[34] H. Lasi, P. Fettke, H.-G. Kemper, T. Feld, and M. Hoffmann, "Industry 4.0," *Business & Information Systems Engineering*, vol. 6, no. 4, pp. 239–242, 2014.

[35] J. Bilbao-Ubillos, V. Camino-Beldarrain, G. Intxaurburu-Clemente, and E. Velasco-Balmaseda, "Industry 4.0 and potential for reshoring: A typology of technology profiles of manufacturing firms," *Computers in Industry*, vol. 148, p. 103904, 2023.

[36] M. Kesik-Brodacka, "Progress in biopharmaceutical development," *Biotechnology and applied biochemistry*, vol. 65, no. 3, pp. 306–322, 2018.

[37] Organisation for Economic Co-operation and Development (OECD), "The future of health systems," 2024, accessed: 03 Oct 2025. [Online]. Available: https://www.oecd.org/en/topics/policy-issues/the-future-of-health-systems.html

[38] A. S. Rathore and H. Winkle, "Quality by design for biopharmaceuticals," *Nature Biotechnology*, vol. 27, no. 1, pp. 26–34, 2009.

[39] A. S. Rathore, R. Bhambure, and V. Ghare, "Process analytical technology (PAT) for biopharmaceutical products," *Analytical and Bioanalytical Chemistry*, vol. 398, no. 1, pp. 137–154, 2010.

[40] B. Scott and A. Wilcock, "Process analytical technology in the pharmaceutical industry: a toolkit for continuous improvement," *PDA Journal of Pharmaceutical Science and Technology*, vol. 60, no. 1, pp. 17–53, 2006.

[41] A. S. Rathore, "Roadmap for implementation of quality by design (QbD) for biotechnology products," *Trends in Biotechnology*, vol. 27, no. 9, pp. 546–553, 2009.

[42] S. Laske, A. Paudel, O. Scheibelhofer, S. Sacher, T. Hoermann, J. Khinast, A. Kelly, J. Rantannen, O. Korhonen, F. Stauffer, F. De Leersnyder, T. De Beer, J. Mantanus, P.-F. Chavez, B. Thoorens, P. Ghiotti, M. Schubert, P. Tajarobi, G. Haeffler, S. Lakio, M. Fransson, A. Sparen, S. Abrahmsen-Alami, S. Folestad, A. Funke, I. Backx, B. Kavsek, F. Kjell, M. Michaelis, T. Page, J. Palmer, A. Schaepman, S. Sekulic, S. Hammond, B. Braun, and B. Colegrove, "A review of PAT strategies in secondary solid oral dosage manufacturing of small molecules," *Journal of Pharmaceutical Sciences*, vol. 106, no. 3, pp. 667–712, 2017.

[43] L. L. Simon, H. Pataki, G. Marosi, F. Meemken, K. Hungerbühler, A. Baiker, S. Tummala, B. Glennon, M. Kuentz, G. Steele, H. J. M. Kramer, J. W. Rydzak, Z. Chen, J. Morris, F. Kjell, R. Singh, R. Gani, K. V. Gernaey, M. Louhi-Kultanen, J. O'Reilly, N. Sandler, O. Antikainen, J. Yliruusi, P. Frohberg, J. Ulrich, R. D. Braatz, T. Leyssens, M. von Stosch, R. Oliveira, R. B. H. Tan, H. Wu, M. Khan, D. O'Grady, A. Pandey, R. Westra, E. Delle-Case, D. Pape, D. Angelosante, Y. Maret, O. Steiger, M. Lenner, K. Abbou-Oucherif, Z. K. Nagy, J. D. Litster, V. K. Kamaraju, and M.-S. Chiu, "Assessment of recent process analytical technology (PAT) trends: A multiauthor review," *Organic Process Research & Development*, vol. 19, no. 1, pp. 3–62, 2015.

[44] L. Rolinger, M. Rüdt, and J. Hubbuch, "A critical review of recent trends, and a future perspective of optical spectroscopy as PAT in biopharmaceutical downstream processing," *Analytical and bioanalytical chemistry*, vol. 412, no. 9, pp. 2047–2064, 2020.

[45] G. Mahdi, A. Alamdary, H. Kaghazian, N. Ahmadi, R. Mardani, and A. Gholami, "Biopharmaceutical analysis by HPLC: Practices and challenges," *Chonnam Medical Journal*, vol. 61, no. 1, p. 19, 2025.

[46] A. A. Shukla and U. Gottschalk, *Trends in biotechnology*, vol. 31, no. 3, pp. 147–154, 2013.

[47] J. Claßen, F. Aupert, K. F. Reardon, G. Scheper, D. Schügerl, and T. R. Scheper, "Spectroscopic sensors for in-line bioprocess monitoring in research and pharmaceutical industrial application," *Analytical and Bioanalytical Chemistry*, vol. 409, no. 3, pp. 651–666, 2017.

[48] A. Mishra, M. Aghaee, I. M. Tamer, and H. Budman, "Spectroscopic advances in real time monitoring of pharmaceutical bioprocesses: A review of vibrational and fluorescence techniques," *Spectroscopy Journal*, vol. 3, no. 2, p. 12, 2025.

[49] I. Surowiec and J. Scholz, "Capacitance sensors in cell-based bioprocesses: online monitoring of biomass and more," *Current Opinion in Biotechnology*, vol. 83, p. 102979, 2023.

[50] C. V. Raman and K. S. Krishnan, "A new type of secondary radiation," *Nature*, vol. 121, no. 3048, pp. 501–502, 1928.

[51] T. Vankeirsbilck, A. Vercauteren, W. Baeyens, G. Van der Weken, F. Verpoort, G. Vergote, and J. Remon, "Applications of Raman spectroscopy in pharmaceutical analysis," *Trends in Analytical Chemistry*, vol. 21, no. 12, pp. 869–877, 2002.

[52] H. Tanemura, R. Kitamura, Y. Yamada, M. Hoshino, H. Kakihara, and K. Nonaka, "Comprehensive modeling of cell culture profile using Raman spectroscopy and machine learning," *Scientific Reports*, vol. 13, no. 1, p. 21805, 2023.

[53] R. C. Rowland-Jones, A. Graf, A. Woodhams, P. Diaz-Fernandez, S. Warr, R. Soeldner, G. Finka, and M. Hoehse, "Spectroscopy integration to miniature bioreactors and large scale production bioreactors - increasing current capabilities and model transfer," *Biotechnology Progress*, vol. 37, no. 1, p. e3074, 2021.

[54] A. Graf, J. Lemke, M. Schulze, R. Söldner, K. Rebner, M. Hoehse, and J. Matuszczyk, "A novel approach for non-invasive continuous in-line control of perfusion cell cultivations by Raman spectroscopy," *Frontiers in bioengineering and biotechnology*, vol. 10, p. 719614, 2022.

[55] D. Yilmaz, H. Mehdizadeh, D. Navarro, A. Shehzad, M. O'Connor, and P. McCormick, "Application of Raman spectroscopy in monoclonal antibody producing continuous systems for downstream process intensification," *Biotechnology Progress*, vol. 36, no. 3, p. e2947, 2020.

[56] B. Wan, M. Patel, G. Zhou, M. Olma, M. Bieri, M. Mueller, E. Appiah-Amponsah, B. Patel, and K. Jayapal, "Robust platform for inline Raman monitoring and control of perfusion cell culture," *Biotechnology and Bioengineering*, vol. 121, no. 5, pp. 1688–1701, 2024.

[57] M. Pappenreiter, H. Schwarz, B. Sissolak, A. Jungbauer, and V. Chotteau, "Product sieving of mab and its high molecular weight species in different modes of ATF and TFF perfusion cell cultures," *Journal of Chemical Technology & Biotechnology*, vol. 98, no. 7, pp. 1658–1672, 2023.

[58] N. R. Abu-Absi, B. M. Kenty, M. E. Cuellar, M. C. Borys, S. Sakhamuri, D. J. Strachan, M. C. Hausladen, and Z. J. Li, "Real time monitoring of multiple parameters in mammalian cell culture bioreactors using an in-line Raman spectroscopy probe," *Biotechnology and bioengineering*, vol. 108, no. 5, pp. 1215–1221, 2011.

[59] J. Whelan, S. Craven, and B. Glennon, "In situ Raman spectroscopy for simultaneous monitoring of multiple process parameters in mammalian cell culture bioreactors," *Biotechnology progress*, vol. 28, no. 5, pp. 1355–1362, 2012.

[60] B. Berry, J. Moretto, T. Matthews, J. Smelko, and K. Wiltberger, "Cross-scale predictive modeling of cho cell culture growth and metabolites using r aman spectroscopy and multivariate analysis," *Biotechnology progress*, vol. 31, no. 2, pp. 566–577, 2015.

[61] G. Chen, J. Hu, Y. Qin, and W. Zhou, "Viable cell density on-line auto-control in perfusion cell culture aided by in-situ Raman spectroscopy," *Biochemical Engineering Journal*, vol. 172, p. 108063, 2021.

[62] B. S. McAvan, L. A. Bowsher, T. Powell, J. F. O'Hara, M. Spitali, R. Goodacre, and A. J. Doig, "Raman spectroscopy to monitor post-translational modifications and degradation

in monoclonal antibody therapeutics," *Analytical Chemistry*, vol. 92, no. 15, pp. 10 381–10 389, 2020.

[63] M. Duan, J. Liu, L. Wang, J. Du, and J. Zhang, "Novel identification of mabs by Raman spectroscopy," *Engineering Reports*, vol. 6, no. 12, p. e12954, 2024.

[64] H. Schwarz, M. E. Mäkinen, A. Castan, and V. Chotteau, "Monitoring of amino acids and antibody n-glycosylation in high cell density perfusion culture based on Raman spectroscopy," *Biochemical Engineering Journal*, vol. 182, p. 108426, 2022.

[65] M. Y. Li, B. Ebel, C. Paris, F. Chauchard, E. Guedon, and A. Marc, "Real-time monitoring of antibody glycosylation site occupancy by in situ Raman spectroscopy during bioreactor CHO cell cultures," *Biotechnology Progress*, vol. 34, no. 2, pp. 486–493, 2018.

[66] K. A. Esmonde-White, M. Cuellar, and I. R. Lewis, "The role of Raman spectroscopy in biopharmaceuticals from development to manufacturing," *Analytical and Bioanalytical Chemistry*, vol. 414, no. 2, pp. 969–991, 2022.

[67] X. Pan, C. Dalm, R. H. Wijffels, and D. E. Martens, "Metabolic characterization of a CHO cell size increase phase in fed-batch cultures," *Applied Microbiology and Biotechnology*, vol. 101, pp. 8101–8113, 2017.

[68] J. Lemke, R. Söldner, and J. Austerjost, "Online deployment of an O-PLS model for dielectric spectroscopy-based inline monitoring of viable cell concentrations in chinese hamster ovary cell perfusion cultivations," *Engineering in Life Sciences*, vol. 23, p. e2200053, 2023.

[69] B. J. Downey, L. J. Graham, J. F. Breit, and N. K. Glutting, "A novel approach for using dielectric spectroscopy to predict viable cell volume (VCV) in early process development," *Biotechnology Progress*, vol. 30, no. 2, pp. 479–487, 2014.

[70] A. Bergin, J. Carvell, and M. Butler, "Applications of bio-capacitance to cell culture manufacturing," *Biotechnology Advances*, vol. 61, p. 108048, 2022.

[71] S. Metze, S. Ruhl, G. Greller *et al.*, "Monitoring online biomass with a capacitance sensor during scale-up of industrially relevant CHO cell culture fed-batch processes in single-use bioreactors," *Bioprocess and Biosystems Engineering*, vol. 43, no. 2, pp. 193–205, 2020.

[72] S. Metze, S. Blioch, J. Matuszczyk *et al.*, "Multivariate data analysis of capacitance frequency scanning for online monitoring of viable cell concentrations in small-scale bioreactors," *Analytical and Bioanalytical Chemistry*, vol. 412, no. 9, pp. 2089–2102, 2020.

[73] D. A. Pais, C. Brown, A. Neuman, K. Mathur, I. A. Isidro, P. M. Alves, and P. G. Slade, "Dielectric spectroscopy to improve the production of raav used in gene therapy," *Processes*, vol. 8, no. 11, p. 1456, 2020.

[74] J. P. Lomont and J. P. Smith, "In situ process analytical technology for real time viable cell density and cell viability during live-virus vaccine production," *International Journal of Pharmaceutics*, vol. 649, p. 123630, 2024.

[75] D. C. Montgomery, *Design and Analysis of Experiments*, 9th ed. Hoboken, NJ: John Wiley & Sons, 2017.

[76] C.-F. Mandenius and A. Brundin, "Bioprocess optimization using design-of-experiments methodology," *Biotechnology Progress*, vol. 24, no. 6, pp. 1191–1203, 2008.

[77] M. Brunner, J. Fricke, P. Kroll, and C. Herwig, "Investigation of the interactions of critical scale-up parameters (pH, pO2 and pCO2) on CHO batch performance and critical quality attributes," *Bioprocess and biosystems engineering*, vol. 40, no. 2, pp. 251–263, 2017.

[78] H. Schwarz, K. Lee, A. Castan, and V. Chotteau, "Optimization of medium with perfusion microbioreactors for high density CHO cell cultures at very low renewal rate aided by design of experiments," *Biotechnology and Bioengineering*, vol. 120, no. 9, pp. 2523–2541, 2023.

[79] S. F. Abu-Absi, L. Yang, P. Thompson, C. Jiang, S. Kandula, B. Schilling, and A. A. Shukla, "Defining process design space for monoclonal antibody cell culture," *Biotechnology and Bioengineering*, vol. 106, no. 6, pp. 894–905, 2010.

[80] F. Li, Y. Hashimura, R. Pendleton, J. Harms, E. Collins, and B. Lee, "A systematic approach for scale-down model development and characterization of commercial cell culture processes," *Biotechnology Progress*, vol. 22, no. 3, pp. 696–703, 2006.

[81] R. C. Rowland-Jones, F. van den Berg, A. J. Racher, E. B. Martin, and C. Jaques, "Comparison of spectroscopy technologies for improved monitoring of cell culture processes in miniature bioreactors," *Biotechnology Progress*, vol. 33, no. 2, pp. 337–346, 2017.

[82] R. C. Rowland-Jones and C. Jaques, "At-line Raman spectroscopy and design of experiments for robust monitoring and control of miniature bioreactor cultures," *Biotechnology progress*, vol. 35, no. 2, p. e2740, 2019.

[83] A. Graf, A. Woodhams, M. Nelson, D. D. Richardson, S. M. Short, M. Brower, and M. Hoehse, "Automated data generation for Raman spectroscopy calibrations in multi-parallel mini bioreactors," *Sensors*, vol. 22, no. 9, p. 3397, 2022.

[84] A. O. Kirdar, K. D. Green, and A. S. Rathore, "Application of multivariate data analysis for identification and successful resolution of a root cause for a bioprocessing application," *Biotechnology progress*, vol. 24, no. 3, pp. 720–726, 2008.

[85] H. Le, S. Kabbur, L. Pollastrini, Z. Sun, K. Mills, K. Johnson, G. Karypis, and W.-S. Hu, "Multivariate analysis of cell culture bioprocess data - lactate consumption as process indicator," *Journal of Biotechnology*, vol. 162, no. 2, pp. 210–223, 2012.

[86] I. T. Jolliffe, *Principal Component Analysis*, 2nd ed. New York: Springer, 2002. [Online]. Available: https://link.springer.com/book/10.1007/b98835

[87] P. Geladi and B. R. Kowalski, "Partial least-squares regression: a tutorial," *Analytica Chimica Acta*, vol. 185, pp. 1–17, 1986.

[88] J. Trygg and S. Wold, "Orthogonal projections to latent structures (O-PLS)," *Journal of Chemometrics*, vol. 16, no. 3, pp. 119–128, 2002.

[89] A. S. Rathore, S. Nikita, G. Thakur, and S. Mishra, "Artificial intelligence and machine learning applications in biopharmaceutical manufacturing," *Trends in Biotechnology*, vol. 41, no. 4, pp. 497–510, 2023.

[90] J. Engel, J. Gerretzen, E. Szymańska, J. J. Jansen, G. Downey, L. Blanchet, and L. M. Buydens, "Breaking with trends in pre-processing?" *Trends in Analytical Chemistry*, vol. 50, pp. 96–106, 2013.

[91] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed., ser. Springer Series in Statistics. New York, NY: Springer, 2009.

[92] L. Eriksson, E. Johansson, N. Kettaneh-Wold, J. Trygg, C. Wikström, and S. Wold, *Multi- and Megavariate Data Analysis: Part I - Basic Principles and Applications*, 2nd ed. Umeå, Sweden: Umetrics Academy, 2006.

[93] Y. Hu, T. Jiang, A. Shen, W. Li, X. Wang, and J. Hu, "A background elimination method based on wavelet transform for Raman spectra," *Chemometrics and Intelligent Laboratory Systems*, vol. 85, no. 1, pp. 94–101, 2007.

[94] P. H. C. Eilers and H. F. M. Boelens, "Baseline correction with asymmetric least squares smoothing," Leiden University Medical Centre Report, Tech. Rep., 2005.

[95] A. Savitzky and M. J. E. Golay, "Smoothing and differentiation of data by simplified least squares procedures." *Analytical Chemistry*, vol. 36, no. 8, pp. 1627–1639, 1964.

[96] F. Westad and F. Marini, "Validation of chemometric models – a tutorial," *Analytica Chimica Acta*, vol. 893, pp. 14–24, 2015.

[97] R. M. Santos, J.-M. Kessler, P. Salou, J. C. Menezes, and A. Peinado, "Monitoring mab cultivations with in-situ Raman spectroscopy: The influence of spectral selectivity on calibration models and industrial use as reliable PAT tool," *Biotechnology progress*, vol. 34, no. 3, pp. 659–670, 2018.

[98] T. Mehmood, S. Sæbø, and K. H. Liland, "Comparison of variable selection methods in partial least squares regression," *Journal of Chemometrics*, vol. 34, no. 6, p. e3226, 2020.

[99] S. Wold, E. Johansson, and M. Cocchi, "PLS - partial least squares projections to latent structures," in *3D QSAR in Drug Design: Theory, Methods and Applications*, H. Kubinyi, Ed. Leiden: ESCOM Science Publishers, 1993, pp. 523–550.

[100] B. Galindo-Prieto, L. Eriksson, and J. Trygg, "Variable influence on projection (VIP) for orthogonal projections to latent structures (OPLS)," *Journal of Chemometrics*, vol. 28, no. 8, pp. 623–632, 2014.

[101] T. Rajalahti, R. Arneberg, F. S. Berven, K.-M. Myhr, R. J. Ulvik, and O. M. Kvalheim, "Biomarker discovery in mass spectral profiles by means of selectivity ratio plot," *Chemometrics and Intelligent Laboratory Systems*, vol. 95, no. 1, pp. 35–48, 2009.

[102] T. Rajalahti, R. Arneberg, A. C. Kroksveen, M. Berle, K.-M. Myhr, and O. M. Kvalheim, "Discriminating variable test and selectivity ratio plot: Quantitative tools for interpretation and variable (biomarker) selection in complex spectral or chromatographic profiles," *Analytical Chemistry*, vol. 81, no. 7, pp. 2581–2590, 2009.

[103] E. Johansson and K. Mazarakis, "Computer-implemented method, computer program product and system for data analysis," Patent, 2021, EP4036919, Sartorius Stedim Data Analytics AB.

[104] B. Berry, J. Moretto, T. Matthews, J. Smelko, and K. Wiltberger, "Cross-scale predictive modeling of CHO cell culture growth and metabolites using Raman spectroscopy and multivariate analysis," *Biotechnology Progress*, vol. 31, pp. 566–577, 2015.

[105] T. A. Webster, B. C. Hadley, W. Hilliard, C. Jaques, and C. Mason, "Development of generic Raman models for a GS-KO™ CHO platform process," *Biotechnology Progress*, vol. 34, no. 3, pp. 730–737, 2018.

[106] H. Mehdizadeh, D. Lauri, K. M. Karry, M. Moshgbar, R. Procopio-Melino, and D. Drapeau, "Generic Raman-based calibration models enabling real-time monitoring of cell culture bioreactors," *Biotechnology progress*, vol. 31, no. 4, pp. 1004–1013, 2015.

[107] V. Konakovsky, A. C. Yagtu, C. Clemens, M. M. Müller, M. Berger, S. Schlatter, and C. Herwig, "Universal capacitance model for real-time biomass in cell culture," *Sensors*, vol. 15, no. 9, pp. 22 128–22 150, 2015.

[108] S. André, S. Lagresle, A. Da Sliva, P. Heimendinger, Z. Hannas, É. Calvosa, and L. Duponchel, "Developing global regression models for metabolite concentration prediction regardless of cell line," *Biotechnology and Bioengineering*, vol. 114, no. 11, pp. 2550–2559, 2017.