Spectroscopy as process analytical technology for preparative protein purification

zur Erlangung des akademischen Grades eines
Doktors der Ingenieurswissenschaften (Dr.-Ing.)

von der KIT-Fakultät für Chemieingenieurwesen und Verfahrenstechnik
des
Karlsruher Instituts für Technologie (KIT)
genehmigte

Dissertation

von
MSc ETH Matthias Roman Rüdt
aus Winterthur, Schweiz

Tag der mündlichen Prüfung: 25.07.2018

Erstgutachter: Prof. Dr. Jürgen Hubbuch
Zweitgutachterin: Prof. Dr. Gisela Guthausen

O God, I could be bounded in a
nutshell and count myself a king of
infinite space [...]

Hamlet, WILLIAM SHAKESPEARE

Acknowledgement

My research during the last couple of years would not have been possible without the support of many people and institutions. Here, I would like to express my gratitude for their help.

Prof. Dr. Jürgen Hubbuch offered me a PhD position in his group on a unique and very interesting research project. He always supported my work with great enthusiasm and was open to new ideas and side-projects. It has been a great pleasure to work in the group. I could not have made a better choice.

Prof. Dr. Gisela Guthausen assessed my PhD project as co-referee. I would like to thank her for her time and effort she put into the evaluation and the discussion with me.

My research was funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No 635557. I am thankful for the financial support and for the very interesting NextBioPharmDSP project I could contribute to. The project allowed me to follow my research freely and attend a number of interesting conferences. Furthermore, I would like to thank the project partners at Lek Pharmaceuticals, Merck, Sandoz, BOKU, KI, and NSYS for their collaboration, for the interesting meetings and discussions in Germany, France, Austria and Slovenia.

Dr. Bertrand Guélat offered me an internship at Novartis in 2013 and thus started my excursion into the field of protein purification. I am very thankful for the opportunity and for the chain of events that followed. Furthermore, I would like to thank all of my former colleagues at Novartis in Basel and Mengeš for teaching me a lot and for the good time I had there. My special thanks go to Gorazd Hribar and Edo Prašnikar for taking care of me during my time in Slovenia.

During my thesis, I collaborated with Sebastian Andris, Nina Brestrich, Steffen Grosshans, Stefan Heissler, Josefine Morgenstern, Laura Rolinger, Adrian Sanden, Philipp Vormittag, and Gang Wang. I would like to thank them for the ideas, interesting work, and fruitful discussions we had to-

gether. Till Briskot, Daniel Büchler, Juliane Diehm, Nils Hillebrandt, Marius Meier, Jonas Rogalla, and Robin Schiemer were involved in my research with their Bachelor or Master theses. I hope, they enjoyed their work as much as I did supervising them.

Furthermore, I would like to thank the MAB for taking in a Swiss guy with open arms. Working at MAB has been a lot of fun. I enjoyed sailing, cycling, climbing, and many other activities with the group. The social activities made my PhD a very memorable time and I would be sorry to miss any of these. Thank you!

My parents have contributed in many ways to my PhD. At all stages of my education, they supported me morally, gave me a warm and welcoming home, and contributed financially. They took my decision to do my PhD in Germany in a stride and took every opportunity to visit me here and at the other locations I have been living in the last years. I do not think that this can be taken for granted. I am profoundly thankful for having them as my parents and for their support of my work.

Last but definitely not least, I would like to thank my girlfriend Laura Rolinger for being with me during my PhD. This sometimes included waking up in the morning and—as the first thing—being told my newest idea. I am very grateful for her help, for her taking care of me, the shared adventures, and for the extremely valuable time we spend together.

Matthias Rüdt, Karlsruhe, 12.12.2018

Abstract

In modern pharmacy, biotechnology-based products provide medication for many diseases otherwise hard to treat. The so-called biopharmaceuticals are highly effective but difficult to produce. Both facts are reflected by the biopharmaceuticals' market values. In 2016, biopharmaceuticals contributed eight out of ten of the most valuable pharmaceutical products. The production of biopharmaceuticals relies on their expression in bioreactors by host cells. The resulting fermentation broth is however not pure but consists of a wide range of metabolites, proteins, Deoxyribonucleic Acid (DNA), and Ribonucleic Acid (RNA) for cell maintenance and proliferation. Furthermore, biotechnological production also results in numerous product-related heterogeneities, many of which may be contaminants. This mixture cannot be directly administrated to a patient. Instead, a second stage of biopharmaceutical production, Downstream Processing (DSP), focuses on purifying the product.

DSP of biopharmaceuticals comprises a number of steps for purifying, modifying, and concentrating the product. Frequently applied unit operations are chromatography for purification, batch reactions for chemical modifications, and Cross-Flow Filtration (CFF) for buffer exchange and concentration. Due to uncontrolled variations during fermentation and the DSP process, variability is introduced into the biopharmaceutical product. In general, such variability is unwanted as it may lead to changing efficacy and side-effects in patients. Regulatory agencies and industry have taken a systematic, knowledge-based approach to reducing process variability. Next to Quality by Design (QbD), Process Analytical Technology (PAT) can help to improve process understanding during development and in production by providing rich and timely information on the performed process. Ideally, the selected PAT allows to measure Critical Quality Attributes (CQAs) and Critical Process Parameters (CPPs). Based on the obtained information, improved process control can be realized.

The objective of this thesis was to developed a toolbox of spectroscopic PAT sensors for DSP and apply them to different unit operations of biopharmaceuticals. To achieve sensitivity towards a wide variety of CQAs and CPPs, the sensors were selected to measure based on multiple different physical principles. Sensors were furthermore selected such that all protein structural levels were covered to allow for measuring multiple orthogonal attributes. Selected sensors were Ultraviolet/Visible (UV/Vis) spectroscopy, Fourier-Transform Infrared (FTIR) spectroscopy, Static Light Scattering (SLS), and Dynamic Light Scattering (DLS). All sensors were adapted to flow-through measurements and implemented in- or on-line. By relying on a knowledge-based sensor selection, the sensors were applied to different unit operations and biopharmaceuticals.

The first step in DSP is commonly a chromatographic Protein A capture. Due to the high costs of the Protein A resin, it is economically important to maximize resin utilization. During the load phase, monitoring is however difficult as the chemical composition of the column effluent is complex and continuously changing. UV/Vis spectroscopy demonstrated previously a high sensitivity and selectivity for proteins even in complex mixtures. Thus, UV/Vis spectroscopy in conjunction with Partial-Least Squares (PLS) modeling was applied to monitor the load phase of a monoclonal Antibody (mAb) capture step. This approach allowed to accurately predict the product breakthrough. Based on the predicted product concentration, a process control was realized. The load step was terminated as soon as a threshold concentration was reached.

Since UV/Vis spectroscopy is sensitive (i.e. proteins have relatively high absorption coefficients), the method is also prone to detector saturation at high concentrations. This poses a challenge for application of UV/Vis spectroscopy at high protein concentrations typical during e.g. the elution step in preparative chromatography. To address this problem, Variable Pathlength (VP) UV/Vis spectroscopy in conjunction with PLS modeling was investigated. While former allowed to measure UV/Vis spectra in a significantly extended concentration range, latter allowed to predict single protein concentrations in mixtures. The method was applied to monitor elutions in-line of Cation-Exchange (CEX) chromatography steps at high loading densities. First, the separation of cytochrome c from lysozyme was observed. Second, mAb was separated from High Moleculare Weight Variants (HMWs) and also monitored in-line. The protein concentration predicted from the PLS model were in subsequent runs used to control the fractionation of the run. Product was collected as long as certain purity criteria of the produced pools were met.

UV/Vis spectroscopy is partly so powerful for monitoring DSP, as it measures many protein structural elements simultaneously including the solvatization of aromatic amino acids. The solvatochromic sensitivity of UV/Vis spectroscopy may however also be used for monitoring other chromophores. In a study for monitoring the batch conjugation reaction of a mAb with two different surrogate drugs UV/Vis spectroscopy was selected as PAT. Changes in the solvatization of the drugs were expected during the conjugation reaction since they move from an aqueous into a proteinous environment. This hypothesis was tested at two scales. In a microplate format, in situ spectral measurements were performed with a UV/Vis plate reader. In a lab-scale setup, a Diode Array Detector (DAD) was implemented on-line. Based on the acquired spectra, the conjugation could be monitored at both scales and with both drugs.

While UV/Vis spectroscopy showed sensitivity for the primary, tertiary, and weakly for quaternary protein structure, it did not allow to measure the secondary structure. Furthermore, many chemicals are not absorbing in the accessible UV/Vis spectral range. As an alternative method, FTIR spectroscopy was explored. FTIR spectroscopy shows sensitivity for almost any molecule which makes it an interesting method for process monitoring. Furthermore, FTIR spectroscopy is a frequently used analytical tool for assessing protein secondary structures. To evaluate FTIR spectroscopy as a PAT for chromatography, a setup was developed with in-line Attenuated Total Reflection (ATR) flowcell. With this setup, first the separation of mAb and lysozyme was monitored. The two proteins have significant differences in secondary structure and could thus be selectively quantified by PLS regression. Second, the separation of different Polyethylene Glycole (PEG) lysozyme conjugates was observed. PEG is not active in UV/Vis but could easily be detected in Infrared (IR). Third, a process related impurity was selectively monitored in the flow-through during the load.

Virus-Like Particles (VLPs) are promising new biopharmaceuticals with potential applications to many diseases. To realize their full potential and meet the necessary purity, un-enveloped VLPs need to be disassembled and subsequently reassembled. A multimodal spectroscopic approach was chosen to monitor the reassembly of Hepatitis B core Antigen (HBcAg) VLPs by CFF. A novel experimental setup was established consisting of a commercial CFF unit with custom-made on-line measurement loop and control software. The on-line measurement loop included a UV/Vis spectrometer and a light-scattering photometer. UV/Vis spectroscopy provided information on the protein concentration and the solvatization of aromatic amino acids. Hydrophobic interactions of tyrosines are essential for HBcAg assembly. The change in the measured hydrophobicity by UV/Vis spectroscopy

was thus related to the rate of assembly. DLS and SLS measurements provided information on the particles sizes and concentration in solution. This setup was used for monitoring VLP reassembly of three different constructs each at three different Transmembrane Pressures (TMPs). The results not only allowed to follow the VLP assembly process but also provided evidence on adverse effects of aggregates on VLP assembly.

The last performed study focused on the data analysis of protein chromatograms. Most of the previously performed data analysis of protein chromatograms relied on PLS regression. PLS regression is however a correlative method requiring model calibration prior to application. Calibration as well as model maintenance can however involve significant effort. Therefore, the final part of this thesis focused on developing a calibration-free method for factorizing protein chromatograms. With a combination of second-derivative spectral pretreatment and hard-constraint Multivariate Curve Resolution (MCR), bilinear chromatograms were successfully factorized into concentrations and protein spectra. The method was tested for multiple case studies including the separation of a ternary mixture, the simultaneous factorization of multiple binary chromatograms, and the preparative separation of an Antibody Drug Conjugate (ADC) from the unconjugated mAb. The estimated elution peak shapes corresponded closely to the measured concentration from off-line analytics. The estimated protein spectra allowed to identify the different species based on a protein spectral library. In summary, the method provided a powerful calibration-free method for factorizing protein chromatograms and may in future simplify the analysis of bilinear protein chromatograms.

Zusammenfassung

In der modernen Pharmazie ermöglichen biotechnologische Produkte die Bekämpfung vieler Krankheiten, die ansonsten schwierig zu behandeln sind. Die sogenannten Biopharmazeutika sind häufig sehr effektiv aber kompliziert herzustellen. Beide Aspekte spiegeln sich im hohen Marktwert von Biopharmazeutika wieder. Im Jahr 2016 gehörten acht der zehn umsatzstärksten Medikamente zur Gruppe der Biopharmazeutika. Die Produktion der biotechnologischen Produkte findet in Bioreaktoren mittels gentechnisch veränderter Wirtszellen statt. Die resultierende Fermentationsbrühe enthält aber nicht reines Produkt sondern besteht aus einer komplexen Mischung aus Metaboliten, Proteinen, Desoxyribonukleinsäuren (DNS) und Ribonukleinsäuren (RNS), die von den Wirtszellen zur Homöostase und zur Vermehrung produziert werden. Des Weiteren entstehen während der Kultivierung viele mit dem Zielprodukt verwandte Produktheterogenitäten. Die Fermentationsbrühe kann daher nicht direkt einem Patienten verabreicht werden. Stattdessen folgt ein zweiter Produktionsabschnitt, das sogenannte *Downstream Processing (DSP)*, der sich insbesondere mit der Aufreinigung des Produktes beschäftigt.

Das DSP von Biopharmazeutika umfasst eine Reihe von Schritten zur Aufreinigung, aber auch zur chemische Modifizierung und zur Aufkonzentrierung des Produktes. Häufig eingesetzte Prozessschritte sind Chromatographie, Batchreaktionen und Querstromfiltration. Durch unkontrollierte Einflüsse während der Fermentierung und im DSP entsteht Variabilität im finalen Biopharmazeutikum. Solch eine Variabilität ist generell unerwünscht, da sie sich auf die Effizienz des Biopharmazeutikums in Patienten auswirken oder auch zu unerwarteten Nebenwirkungen führen kann. Die regulatorischen Behörden und pharmazeutische Industrie haben einen systematischen, wissensorientierten Ansatz gewählt, um die Produktvariabilität zu reduzieren. Neben dem Ansatz *Quality by Design (QbD)* können Prozess Analytische Technologien (PAT) helfen das Prozessverständnis während der Entwicklung und in der Produktion zu verbessern. PAT zielt darauf

ab zeitnah zum laufenden Prozess Messungen durchzuführen, die reich an Informationen über den Prozess sind. Idealerweise erlauben die PAT kritische Qualitätsattribute und kritische Prozessparameter zu messen. Mit den gemessenen Daten kann dann eine verbesserte Prozesskontrolle umgesetzt werden.

Das Ziel dieser Arbeit war es, eine vielseitig einsetzbare Auswahl an spektroskopischen Sensoren für das DSP zu evaluieren und einzusetzen, um verschiedene Prozessschritte und Biopharmazeutika zu überwachen. Um eine gute Sensitivität für viele kritische Qualitätsattribute und Prozessparameter zu erreichen, wurden Sensoren ausgewählt, die auf verschiedenen physikalischen Messprinzipien beruhen. Insbesondere wurde beachtet, dass die Sensoren die verschiedenen Strukturebenen von Proteinen abdecken um verschiedene orthogonal Attribute messen zu können. Die gewählten Sensoren waren UV/Vis-Spektroskopie, Fourier-Transform-Infrarotspektroskopie (FTIR-Spektroskopie), sowie statische und dynamische Lichtstreuung. Alle Sensoren wurden für Durchflussmessungen angepasst und *in-* oder *online* eingesetzt. Die verwendeten Sensoren wurden in verschiedene Studien getestet, wobei für die jeweilige Problemstellung die vielversprechendsten Sensoren ausgewählt wurden.

Der erste Schritt im DSP von mAbs und verwandten Produkten ist häufig eine Protein A Affinitätschromatographie. Durch die hohen Kosten von Protein A Chromatographiemedien ist es ökonomisch wichtig, die ausgenutzte Kapazität der Chromatographiesäule zu maximieren. Jedoch ist besonders während der Beladung der Chromatographiesäule die Überwachung des Prozesses schwierig, da die Zusammensetzung des Säuleneffluents sehr komplex ist und sich kontinuierlich ändert. UV/Vis-Spektroskopie hat sich bereits in früheren Studien durch die hohe Sensitivität und Selektivität für Proteine ausgezeichnet. Für die Überwachung der Beladungsphase eines Protein A Affinitätschromatographieschrittes mit einem monoklonalen Antikörper (engl. *monoclonal Antibody – mAb*) wurde deshalb UV/Vis-Spektroskopie kombiniert mit *Partial-Least Squares (PLS)* Regression ausgewählt. Dieser Ansatz ermöglichte die genaue Vorhersage des Produktdurchbruches im Beladungsschritt. Basierend auf der vorhergesagten Proteinkonzentrationen wurde anschließend der Prozess kontrolliert. Die Beladung der Säule wurde unterbrochen, sobald die Produktkonzentration im Durchbruch einen Grenzwert überschritten hatte.

Da UV/Vis-Spektroskopie relativ sensitiv ist (Proteine daher hohe Absorptionskoeffizienten besitzen), ist die Methode bei hohen Konzentrationen auch oft von Detektorsaturierung betroffen. Dies macht es schwierig, UV/Vis-Spektroskopie für hochkonzentrierte Lösungen einzusetzen, wie sie zum Beispiel während der Elution bei präperativer Chromatographie auftre-

ten können. Um dieses Problem zu reduzieren, wurde Variable-Pfadlängen (VP) UV/Vis-Spektroskopie zusammen mit PLS Regression untersucht und eingesetzt. Ersteres erlaubt dabei, den linearen Bereich von UV/Vis-Spektroskopie deutlich zu erweitern. Letzteres ermöglicht die Vorhersage von Einzelproteinkonzentrationen in Mischungen. Diese Methode wurde zur *inline* Überwachung der Elution bei der Kationenaustauschchromatographie bei hohen Ladungsdichten eingesetzt. Zuerst wurde die Auftrennung von Cytochrom c und Lysozym beobachtet. Eine zweite Untersuchung befasste sich mit der Auftrennung von mAb Proteinaggregaten von nativem mAb. In beiden Fällen erlaubte die Methode die selektive Quantifizierung der Proteine. In anschließenden Experimenten wurden die vorhergesagten Proteinkonzentrationen der PLS Modelle dafür ausgenutzt, die Produktsammlung zu kontrollieren. Dafür wurden Reinheitskriterien für den Produktpool bestimmt. Das Produkt wurde dann solange gesammelt, wie die Reinheitskriterien für den Pool eingehalten wurden.

UV/Vis-Spektroskopie ist unter anderem deshalb so nützlich für die Prozessüberwachung im DSP, weil sie viele verschiedene Strukturelemente von Proteinen gleichzeitig messen kann. Dazu gehört auch die Solvatisierung aromatischer Aminosäuren. Die solvatochromische Sensitivität von UV/Vis-Spektroskopie kann aber auch dafür genutzt werden andere Chromophore zu überwachen. In einer Studie zur Überwachung der Batchkonjugierung eines mAbs mit zwei verschiedenen Ersatzwirkstoffen zu *Antibody Drug Conjugates (ADCs)* wurde deshalb auch UV/Vis-Spektroskopie ausgewählt. Während der Konjugationsreaktion findet ein Übergang des Wirkstoffes vom gelösten Zustand in die Proteinumgebung statt. Dieser Übergang wurde mit UV/Vis-Spektroskopie gemessen und zur Überwachung des Fortschrittes der Konjugation genutzt. Die Überwachung wurde in zwei experimentellen Maßstäben getestet. In einer Mikrotiterplatte wurden spektrale Messungen *in situ* mit einem UV/Vis-Plattenlesegerät durchgeführt. Im Labormaßstab wurde die Batchreaktion *on-line* mittels eines Photodiodenzeilendetektors überwacht. Mit den spektralen Daten und einem PLS Modell konnten anschließend in beiden Maßstäben und mit beiden Wirkstoffen der Reaktionsfortschritt überwacht werden.

Während UV/Vis-Spektroskopie die primäre, tertiäre und schwach die quartäre Proteinstruktur messen kann, zeigte sie keine Sensitivität für die Sekundärstruktur von Proteinen. Wichtig ist auch, dass viele Chemikalien nicht in der zugänglichen UV/Vis-Region absorbieren. Als alternative Methode wurde deshalb FTIR-Spektroskopie untersucht. FTIR-Spektroskopie ist eine häufig eingesetzte Methode um die Sekundärstruktur von Proteinen zu messen. Außerdem zeigt FTIR-Spektroskopie eine gewisse Sensitivität für fast alle Moleküle. Um FTIR-Spektroskopie als PAT für Chromatogra-

phie zu testen, wurde ein experimenteller Aufbau entwickelt, bei dem ein FTIR-Spektrometer *in-line* mit *Attenuated Total Reflection (ATR)* Flusszelle an einen präperativen Chromatographen angehängt wurde. Mit diesem Aufbau wurde zuerst die präperative chromatographische Trennung von mAb und Lysozym überwacht. Die zwei Proteine unterscheiden sich stark in der Sekundärstruktur und konnten deshalb selektiv mittels PLS Regression quantifiziert werden. Ein zweites Experiment zielte auf die Überwachung der präperativen Auftrennung von verschiedenen Polyethylenglycol (PEG)-Lysozymkonjugate ab. PEG absorbiert nicht im UV/Vis-Spektrum, aber konnte mit FTIR-Spektroskopie einfach detektiert werden. In einem dritten Experiment wurde eine prozessspezifische Kontaminante während dem Beladungsschritt selektiv überwacht und quantifiziert.

Virusartige Partikel (engl. *Virus-Like Particles – VLPs*) bilden eine vielversprechende Klasse neuartiger Biopharmazeutika mit potentiellen Anwendungen für viele verschiedene Krankheiten. Damit sie ihr volles Potential als Biopharmazeutika erreichen, müssen insbesondere VLPs ohne Hülle dis- und anschließend reassembliert werden. Um die Reassemblierung von *Hepatitis B core Antigen (HBcAg)* VLPs während einer Querstromfiltration zu überwachen, wurde ein multimodaler spektroskopischer Ansatz gewählt. Dazu wurde ein kommerzielles Querstromfiltrationsgerät mit einem *on-line* Messaufbau und einer spezialisierten Kontrollsoftware erweitert. Der *on-line* Messaufbau enthielt dabei ein UV/Vis-Spektrometer und ein Lichtstreumessgerät. Das UV/Vis-Spektrometer maß die Proteinkonzentration und ermöglichte die Überwachung der Solvatisierung aromatischer Aminosäuren. Für die Assemblierung von HBcAg VLPs ist die hydrophobe Interaktion eines Tyrosinrests mit dem benachbarten Homodimer essentiell. Die Veränderung der Solvatisierung der Tyrosinreste spiegelte deshalb die Assemblierungsgeschwindigkeit der VLPs wieder. Dynamische und statische Lichtstreuungsmessungen ermöglichten eine kombinierte Messung aus Proteingröße und -konzentration. Dieser Versuchsaufbau wurde genutzt, um die Reassemblierung von drei modifizierten VLP-Konstrukten bei drei verschiedenen Transmembrandrücken zu überwachen. Die Resultate erlaubten nicht nur der VLP-Assemblierung zu folgen, sondern zeigten auch, dass sich VLP-Aggregate negativ auf die Assemblierung auswirkten.

Die letzte durchgeführte Studie untersuchte einen neuen datenanalytischen Ansatz für die Auswertung von präperativen Proteinchromatogrammmen. Die meisten in dieser Arbeit durchgeführten Datenanalysen für die Evaluierung von Proteinchromatogrammen nutzten PLS Regression. PLS Regression ist jedoch eine korrelative Methode, muss also vor der Anwendung kalibriert werden. Die Modellkalibrierung und der Unterhalt des Modells kann viel Arbeit verursachen. Deshalb wurde im letzten Teil dieser

Thesis eine kalibrierungsfreie Methode entwickelt, um Proteinchromatogramme in Konzentrationen und Spektren zu faktorisieren. Die Methode verwendete dabei eine vorbereitende zweite Ableitung der spektralen Daten sowie das *Multivariate Curve Resolution (MCR)* Verfahren mit zusätzlichen Bedingungen zum Elutionsverhalten von Proteinen. Mit diesem Ansatz konnten verschiedene bilineare Proteinchromatogramme erfolgreich faktorisiert werden. Die Methode wurde für die Faktorisierung eines tertiären Proteinchromatogrammes, mehrerer binärer Chromatogramme und für die Faktorisierung eines ADC-Chromatogrammes verwendet. Die geschätzten Elutionskurven und die Referenzanalytik zeigten dabei sehr ähnliche Verläufe. Durch die geschätzten Spektren konnten Proteine mittels einer Proteinspektrenbibliothek identifiziert werden. Zusammenfassend ermöglicht die Methode die Analyse von Proteinchromatogrammen mit limitiertem Vorwissen und ist deshalb potentiell ein wertvolles Instrument für die Prozessentwicklung und Analytik von Proteinen.

Contents

Nina Brestich[1], Matthias Rüdt[1], Daniel Büchler, Jürgen Hubbuch ([1] contributed equally)

Sebastian Andris[1], Matthias Rüdt[1], Jonas Rogalla, Michaela Wendeler, Jürgen Hubbuch ([1] contributed equally)

6 In-line Fourier-transform Infrared Spectroscopy as a Versatile Process Analytical Technology in Preparative Protein Chromatography 71

Steffen Großhans[1], Matthias Rüdt[1], Adrian Sanden[1], Nina Brestrich, Josefine Morgenstern, Stefan Heissler, Jürgen Hubbuch ([1] contributed equally)

7 Process Monitoring of Virus-Like Particle Reassembly by Diafiltration with UV/Vis Spectroscopy and Light Scattering 89

Spectroscopic PAT for Protein Purification

Matthias Rüdt[1], Philipp Vormittag[1], Nils Hillebrandt, Jürgen Hubbuch ([1] contributed equally)

Matthias Rüdt, Sebastian Andris, Robin Schiemer, Jürgen Hubbuch

1

Introduction

adapted from a review article by Matthias Rüdt[1], Till Briskot[1], and Jürgen Hubbuch[1]

[1] Institute of Engineering in Life Sciences, Section IV: Biomolecular Separation Engineering, Karlsruhe Institute of Technology (KIT), Germany

In 2004, the United States' Food and Drug Administration (FDA) published *Guidance for industry. PAT – A framework for innovative pharmaceutical development, manufacturing and quality assurance* [1]. Within the guidance, FDA promotes the implementation of Process Analytical Technology (PAT) into all unit operations to monitor Critical Quality Attributes (CQAs). PAT is described as being part of process design and furthermore intended to contribute to process control, i.e. to be taken actively into account for process decisions. While being intended for both small molecules and biologics, the implementation into these two domains of pharmaceuticals is advancing at different paces. In the past, PAT was adopted more quickly in the production of small molecules. For an extensive review thereof, the authors defer to [2]. This article will focus on biologics only.

In contrast to most small molecules, biologics are produced in living organisms which are very sensitive to a wide variety of external factors. Most biologics are complex proteins. They do not consist of one chemical entity but a distribution of many species. Already slight process changes can affect the product quality profile [3]. In order to ensure a consistent

product quality and to reduce batch-to-batch variability, PAT for biologics is of great interest [4]–[8]. Furthermore, current trends towards continuous manufacturing may require an improved process control for keeping a steady state over prolonged periods of time. Such a control may be simplified by the possibility to monitor critical quality attributes in real-time [9].

Other advantages of PAT include the simplification of root cause analysis [10] and improvement of process understanding. Eventually, the improved process understanding and real-time monitoring capabilities may lead to the implementation of the concept of real-time release [1]. Thus, the CQA profile of the final product can be guaranteed to lie within acceptable quality limits solely based on real-time measurements and production batches can be released based on this data.

Early approaches to PAT for biologics widely addressed the problem by implementing on-line analytical chromatography. Already before the release of the PAT guidance, on-line High Performance Liquid Chromatography (HPLC) has been used to control column loading and pooling decisions during chromatographic purification steps [11]–[13]. Subsequently, on-line and at-line HPLC was further used for a variety of applications [14]–[16]. Recently, at-line HPLC has been also implemented in the control of continuous chromatography equipment [17]. HPLC provides high resolution of different species. However, it is complex regarding the required equipment, consisting of a device for sampling as well as the chromatograph itself. This may be undesirable in a manufacturing environment as reliability may be an issue. Furthermore, automated sampling and the analytical separation also lead to non-negligible time delays. Depending on the decision time of a unit operation, this may lead to late notice of process deviations or even completely prevent real-time monitoring.

Spectroscopy is a powerful tool for process monitoring [18]. Spectroscopic equipment has similar investement costs ($20k to $200k) as on-line HPLC. Measurement times are fast, typically in the subsecond range up to a few minutes. Furthermore, measurements can often readily be performed in-line. Fast measurement times are especially important for preparative chromatography, the workhorse in current Downstream Processing (DSP). Preparative chromatographic processes are highly non-linear and feature sharp concentration fronts [19]. Thus, CQAs of the effluent such as the mass fraction of impurities are quickly changing. To reliably control such processes, the used monitoring method needs to have short response times. Typical decision times for preparative protein chromatography lie in the range of 30 s to several minutes. In contrast to at-line HPLC, spectroscopy provides signals with limited selectivity for different components. To overcome this limitation, a combination of multivariate measurements

and mathematical tools for Multivariate Data Analysis (MVDA) is generally applied to extract information from spectroscopic measurements.

Following this argumentation, this article is focusing in a first part on the review of two widely used chemometric tools for the analysis of spectroscopic data. Subsequently, the current state-of-art of spectroscopic PAT in DSP is discussed.

1.1 Multivariate Data Analysis for PAT

The implementation of the PAT framework is often accompanied by the application of *multivariate mathematical approaches* [1], also known as chemometrics. In chemometrics, mathematical and statistical tools are used to extract useful chemical information from large amounts of multivariate measurements or raw data [20]. The multivariate nature of spectroscopic data for PAT arises out of necessity, since no univariate process analyzer has significant selectivity to monitor a specific CQA without interferences from other properties [18]. Chemometrics can be used for a wide variety of tasks, including experimental design (DoE) and MVDA [21]. The present article does not aim to give a complete review of all elements in chemometrics, but focuses solely on MVDA. Furthermore, only the two most common MVDA tools in PAT are discussed more closely: Principal Component Analysis (PCA) and Partial-Least Squares (PLS). A more thorough review of chemometric tools is given in the textbook of Bakeev [18].

1.1.1 Multivariate Projection Methods

Multivariate projection (decomposition) to latent structures forms the basis of many approaches in MVDA [22]. According to Kvalheim [23], [24], the Latent Variables (LV) projection of a data matrix $X = (\vec{x}_1, ..., \vec{x}_k)$, with n observations and k variables, can most easily be understood by reference to variable and objective space, as illustrated in Figure 1.1. The former case (Fig. 1.1a) reveals relationships between observations by plotting the observations in a space spanned by the k variables in X. In the object space (Fig. 1.1b), the coordinate system is defined by the n observations. It visualizes information about the relationship between variables [23]. The main goal of latent projection methods is to reduce the dimensions in the variable space by summarizing variables with similar information in LVs. All latent projection methods help getting fundamental insights into complex multivariate data by (1) discovering groupings in the data, (2) data compression, (3) regression, and more [25].The variable decomposition into
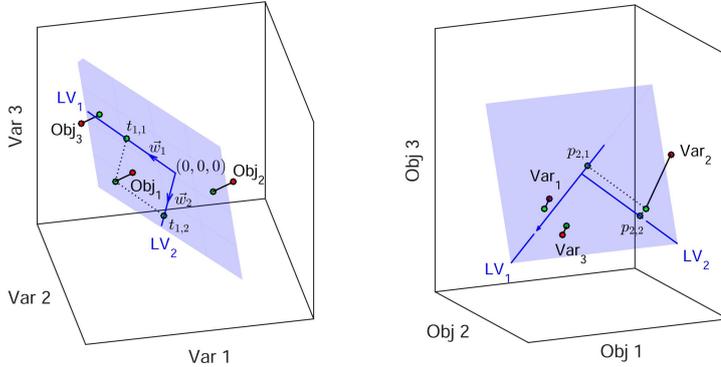
Figure 1.1: Visualization of a data matrix X consisting of three observations with three variables in the variable space (a) and object space (b). Observations in the variable space are projected on a latent structure defined by the weight vectors $\vec{w}_i$, leading to the projection coordinates (scores) $\vec{t}_i$. Projection coordinates (loadings) of the variables in the object space are summarized in the loading vectors $\vec{p}_i$.

LVs can geometrically be interpreted as a projection of the data in the variable and object space on a-dimensional hyperplanes, whereby a represents the number of LVs. Since the projection is performed in both spaces, the maximum number of LVs is $\min(n, k)$. The projection coordinates (scores) of the observations in the variable space on the i-th LV are summarized in the score vector $\vec{t}_i$ and are obtained by projecting the samples on the corresponding weight vector $\vec{w}_i$ [24]. The vectors $\vec{t}_i$ and $\vec{w}_i$ are orthogonal and orthonormal, respectively. Any latent projection method can be derived over the definition of $\vec{w}_i$ [21]. The projection coordinates (loadings) of the variables in the object space are summarized in the loading vector $\vec{p}_i$. The loading vectors $\vec{p}_i$ are not necessarily orthogonal.

1.1.2 Principal Component Analysis

PCA is a common tool in exploratory data analysis and is used for data reduction, simplification, outlier detection, classification, and noise reduction [26]. Data decomposition of a matrix X according to

$$X = TP^T + E_X \tag{1.1}$$

is performed with the objective to explain as much as of the variance in X by a linear combination of a complementary set of scores $T = (\vec{t}_1, ..., \vec{t}_a)$ and loadings $P = (\vec{p}_1, ..., \vec{p}_a)$. In order to differentiate the data decomposition

by PCA from other latent projection methods, the LVs are referred to as principal components (PCs). In PCA, the loadings $\vec{p_i}$ are equal to the weights $\vec{w_i}$ and thus orthonormal. They give a quantitative measure of the part of variance and observed variable shares with the PC [23]. Thus, the whole information regarding the linear relationship between variables is compressed in the loading matrix P. The hidden structure of X concerning the object space can be visualized by loading plots, where the loadings $\vec{p_i}$ are plotted against each other [26]. Variables having similar loading values on a PC are linear dependent (collinear) and are redundant concerning this PC. For mean centered data, as illustrated in Fig. 1.1, collinearity between two variables can graphically be visualized by the cosine of the angle between the two variables in the object space. In the same manner as relationships between variables can be illustrated by loading plots, relationships between observations can be visualized by score plots [26]. Score plots can reveal patterns, clusters, and outliers in the observations (measurements). Usually, two or three PCs are already sufficient to reveal hidden patterns in X by loading and score plots, since the most useful information (variance) in X is explained by the first few PCs. The remaining ones are assumed to comprise predominantly noise [26]. By neglecting these minor PCs, PCA achieves a data simplification and noise reduction in X. Since both scores and loadings are orthogonal, PCA is also able to reduce collinearity in X, which is why it also plays a central role in regression analysis.

1.1.3 Partial Least Square Regression

Linear regression methods like PLS are tools in exploratory data analysis, relating one or more response variables Y with several predictor variables X, by a linear multivariate model

$$Y = XB + E_Y \tag{1.2}$$

where B contains the regression coefficients connecting the predictor variables to the responses. The deviation between model responses and measurements is summarized in the residual matrix E_Y. In the simplest case, when the matrix X is of full rank, multiple linear regression (MLR) can be applied and the regression coefficients can be obtained by the least square solution

$$B = (X^T X)^{-1} X^T Y. \tag{1.3}$$

In most PAT applications, however, the observation to variable ratio is rather low and the X-variables are collinear and noisy. In such cases, prediction abilities of MLR models can be very poor since the estimated regres-

sion coefficients become unstable and can deviate substantially from their expected values [27], [28].

An alternative way to determine the regression coefficients B is by using latent projection methods like principal component regression (PCR) and PLS. In PCR the collinearity problem is solved by (1) decomposing the predictor matrix X to orthogonal PCs and (2) regressing the responses Y on the orthogonal scores T instead of X. The score matrix T is of full rank and allows the prediction of stable regression coefficients. Furthermore, data decomposition prior to regression allows noise reduction and thus the calibration of more robust models. A major drawback of PCR is that data decomposition is performed under the objective to explain as much as possible of the variance in X. However, the variance in X that is relevant for the prediction of Y could be rather small in comparison with the total variance in X. Thus, much of the relevant variance could be lost by PCA [18].

In contrast to PCR, PLS performs a simultaneous decomposition of X and Y with the objective to explain as much as possible of the covariance between the data sets [29]. The decomposition of X and Y can be described by

$$T = XW \tag{1.4}$$

and

$$Y = UC^T + E_C \tag{1.5}$$

where $U = (\vec{u}_1, ..., \vec{u}_a)$ contains the corresponding Y-scores $\vec{u}_i$ on the i-th latent variable, E_C represents the Y-residuals, and $C = (\vec{c}_1, ..., \vec{c}_a)$ denotes the linear transformation defined by the orthogonal Y-loadings $\vec{c}_i$. Since the weight matrix W is determined under the objective of maximizing covariance between X and Y, the scores T are good predictors of the original data X

$$X = TP^T + E_X \tag{1.6}$$

and model also the responses [30]

$$Y = TC^T + E_Y. \tag{1.7}$$

In contrast to PCA, weights $\vec{w}_i$ and loadings $\vec{p}_i$ are not equal. The orthonormal weights can be considered as *tilted* X-loadings since they describe the effective relationship between X and Y. Depending on how strong Y effects W, the weights $\vec{w}_i$ deviate more or less from the loadings $\vec{p}_i$ [31]. The X-loadings are not orthogonal to each other [25]. Comparing Eq. (1.2) with Eq. (1.7) leads to the regression coefficient

$$B = WC^T. \tag{1.8}$$

Since the regression model B is calculated from the orthogonal latent structures W and C, PLS is able to analyze data with strongly collinear, noisy, and numerous X-variables [30].

1.2 Spectroscopy for Process Monitoring in DSP

In the past, spectroscopic methods have been widely used as tools for structural analysis of proteins [32]–[34]. From a biochemical point of view the analysis of proteins can be split into the assessment of primary, secondary, tertiary and quaternary structures. Spectroscopic methods provide information about each of these layers of abstraction within the protein structure (cf. Figure 1.2) [32]. To assess the sequence and total concentration of protein, especially Ultraviolet/Visible (UV/Vis) spectroscopy and Fourier-Transform Infrared (FTIR) spectroscopy are of interest. UV/Vis spectroscopy mainly measures the primary structure, i.e. the content of aromatic amino acids as well as weak spectral shifts due to the solvochromatic effects [32]. The secondary structure of proteins has been frequently measured by vibrational spectroscopy such as FTIR and Raman spectroscopy [33], [35], [36]. The methods allow to measure the vibrational modes of the backbone of polypeptides. The tertiary structure of proteins is accessible over the fluorescence of the aromatic amino acids. The tryptophan emission is solvatochromatic, reacting to changes in the local polarity around tryptophan residues [32], [34]. Thus, structural changes which affect the local environment of tryptophan residues can be detected by fluorescence spectroscopy. Finally, the quaternary structure of proteins, i.e. assembly of multiple subunits or native aggregation of protein monomers, may be assessed over the protein size by light scattering methods including Static Light Scattering (SLS) and Dynamic Light Scattering (DLS) [4], [32].

All of the above mentioned methods are of major interest for process monitoring as each method provides access to orthogonal information about the product. Key aspects of the different methods have been summarized in Table 1.1. In literature, especially UV/Vis absorption and FTIR have been used for a variety of PAT applications (cf. Subsection 1.2.1 and 1.2.2). Literature for fluorescence spectroscopy as well as DLS is less broad. However interesting applications exist (cf. Subsection 1.2.3). In the following sections, the different applications will be discussed.
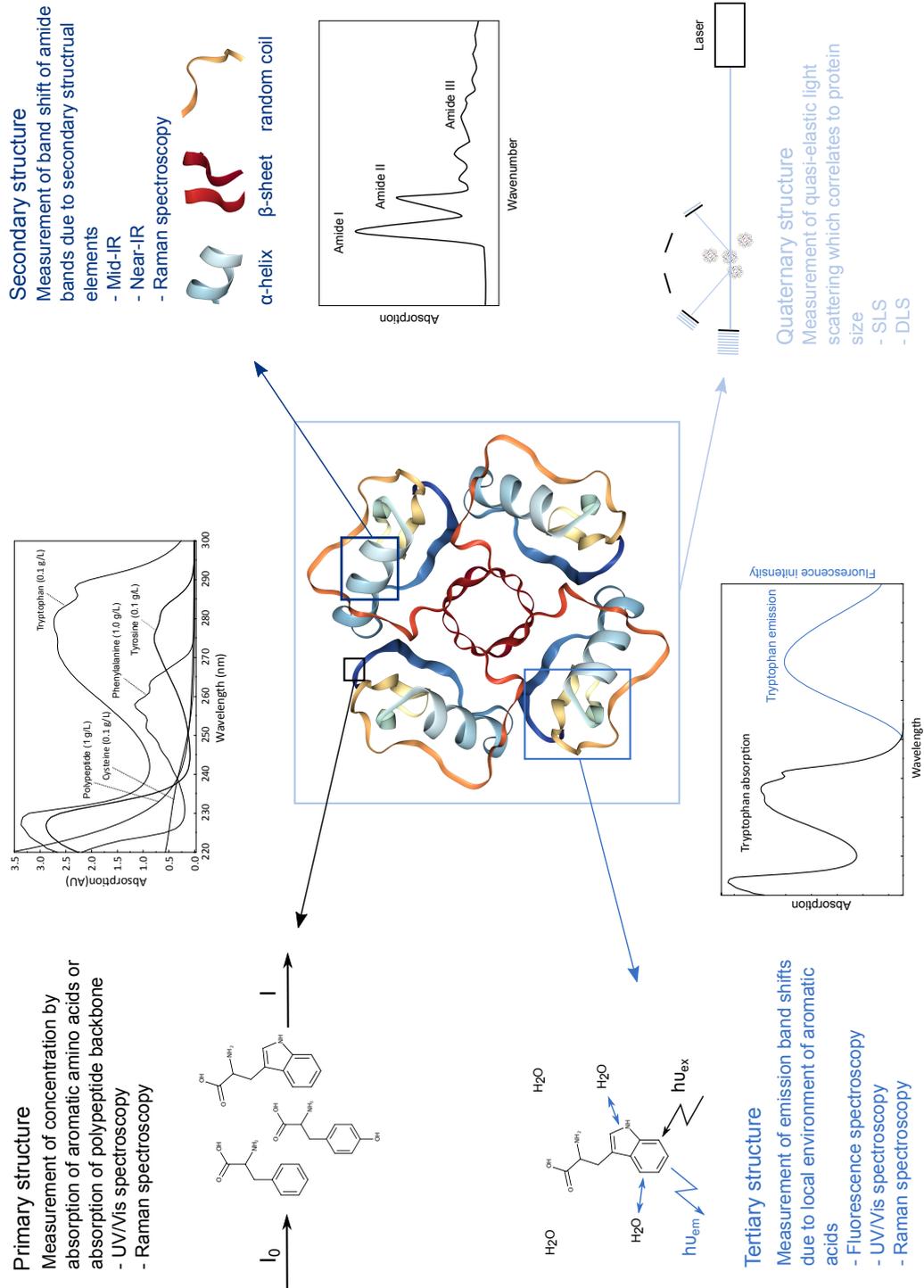
Figure 1.2: (Continued on the following page.)

Figure 1.2: Based on the example of ovomucoid, the four different level of protein structure are illustrated. To each level, suitable spectroscopic methods are listed with a short explaination of what is measured. The lists are not extensive but rather correspond to the most promising methods in the authors eyes. Protein structure retrieved from PDB ID: 1OVO [37], [38]. UV/Vis spectra obtained from [39].

1.2.1 UV/Vis Spectroscopy

UV/Vis spectroscopy measures the absorption of proteins generally in the range between 240 to 340 nm. Mainly due to the content of aromatic amino acids (phenylalanine, tyrosine, and tryptophan) proteins significantly absorb in this region (cf. Figure 1.2, primary structure) [19], [32], [39]. Due to the sensitivity, reproducibility of signals and robustness of the spectrometers, UV/Vis absorption at 280 nm is widely used as a primary detection method of protein concentrations. While current applications mainly rely on univariate UV/Vis measurements, it has been shown that UV/Vis spectra contain a significant amount of information on proteins and may be used for selective quantification even if only minute spectral differences exist [39].

Multivariate UV/Vis spectroscopy in conjunction with PLS modeling for selective protein quantification first appeared in 1994 [40]. Arteaga et al. demonstrated the quantification of the three main bovine caseins by PLS regression on the fourth derivative UV/Vis spectra. The PLS model was calibrated based on designed mixing ratios. In contrast to latter publications which focus on (near)-real-time assays, Areaga et al. intended the proposed method as an off-line analytical assay. In the scope of the publication, the method was not applied to process samples.

The first at-line application for chromatography was only reported in 2011 as a tool to circumvent the analytical bottleneck created by high throughput experimentation [41]. Similar to Arteaga et al., a PLS model was calibrated based on designed mixing ratios of pure protein components. The calibrated PLS model was used to selectively quantify the protein content in elution fractions of multiple co-eluting species from miniaturized and parallelized chromatography experiments. The results were later confirmed by [52]. Subsequently, the method was transferred to an in-line setup with a diode array detector and applied for a selective and real-time quantification of 3 model proteins [42]. It was shown that the deconvoluted signal from the detector could be directly used in a feed-forward controller to trigger product pooling. Experiments were performed in diluted conditions to prevent detector saturation.

Figure 1.3: Summary of different spectroscopic methods of interest with key information on each method.

Spectroscopic method	Wavelength range	Acquisition time	Apparent sensitivity	Measured attribute	Remarks	References
UV/Vis spectroscopy	240–340 nm	0.01–30 s	+	mainly aromatic amino acids	sensitive, quantitative, simple instrumentation strong water bands, for	[10], [32], [39]–[44]
Mid-IR	2.5–25 μm	single scan 0.5–4 s	− −	peptide backbone	high signal-to-noise ratios multiple scans (100–600) are necessary strong water bands, low sensitivity, low	[33], [35], [45]–[49]
Near-IR	0.8–2.5 μm	single scan 0.5–4 s	−	vibrational overtones of peptide backbone	selectivity, simple instrumentation, fiber probes readily available generally low sensitivity,	[18]
Raman spectroscopy	depends on excitation laser	depends on excitation laser	− −	peptide backbone	high selectivity, not infringed by water absorption, fiber probes readily available	[32], [35]
Fluorescence spectroscopy	excitation: 240–300 nm emission: 260–450 nm	0.01–300 s	++	aromatic amino acids and their solvochromatic environment	broad measurement ranges feasible, calibration may be challenging based on time	[32], [34], [50]
DLS	visible light, e.g. 633 nm	0.5–8 min	−	diffusion behaviour of macromolecules	correlation, limited suitability for flow-through	[51]
SLS	visible light, e.g. 633 nm	<1 s	−	light scattering due to particle size	difficult to obtain stable baselines	[32]
Circular dichronism	190–260 nm	<1 s	+	chirality of secondary structure	impractical at high concentrations	[32]

While the above mentioned publications provided accurate predictions of protein concentrations in multi-protein mixtures, they all relied on designed mixing ratios of pure proteins. This may pose major difficulties when calibrating a PLS model for applied examples, e.g. the purification of a monoclonal Antibody (mAb) from its High Moleculare Weight Variants (HMWs). Brestrich et al. addressed this problem by using process based samples for the PLS model calibration [43]. Instead of using pure protein samples to produce designed mixing ratios, chromatographic runs at variable conditions were performed to span a model calibration space. The column effluent of those experiments was fractionated and analyzed by suitable off-line analytics. They applied the newly designed method to different separation problems including the diluted separation of a mAb from its impurities and the at-line measurement of different protein species in human blood fractionation. As a supportive tool, it was applied together with mechanistic modeling for a generic root-cause investigation for model proteins [10]. In summary, UV/Vis spectroscopy is currently among the most promising PAT tools for DSP.

1.2.2 FTIR Spectroscopy

FTIR spectroscopy is frequently applied as a PAT technology for small molecule production [2]. For proteins, FTIR was first established as a tool for assessing the secondary structure [32], [33], [35], [36]. Proteins are detected by the vibration of the polypeptide backbone. Multiple vibrational modes correspond to different detected amide bands (cf. Figure 1.2, primary and secondary structure). The absorption of the amide bands is directly proportional to the amount of polypeptide backbone. The most prominent proteinogenic band, the amide I band, is mainly caused by C=O stretching. Secondary structural elements induce band shifts of the amide bands. This phenomenon can be used to quantify the proportion of different secondary structural elements, e.g. by taking the second derivative or applying Fourier self-deconvolution. Thus, FTIR is a promising candidate for monitoring the overall protein mass as well as the structural integrity of proteins by their secondary structure. The application is however hindered by the strong absorption of water in the same spectral region. It is a non-trivial task to correct for the water absorption. To prevent total extinction in the transmission cell, typical pathlengths need to be very short (approximately $5\,\mu$m), which however also reduces the sensitivity towards proteins. Despite the existing problems, a number of promising applications have been reported.

Publications demonstrated the possibility to selectively detect mAbs, HMWs and Host Cell Proteins (HCPs) [45], [46] with FTIR for biopharmaceutical applications. Experiments were performed in an at-line setup. The approach was later extended to further downstream processing unit operations [47]. Capito et al. demonstrated the use of a calibrated PLS model to selectively quantify mAb, HMW and HCP concentrations of samples drawn from different unit operations. Again experiments were performed at-line. mAb could be quantified down to concentrations of 0.7 g/l while HMW concentrations as low as 1% [w/w] were detected.

During the refolding process of an inclusion body of an autoprotease, FTIR was applied as an in-line PAT tool to monitor the relative content of different secondary structural elements [48]. A time evolution of the relative content of structural elements could be shown during the refolding process. However, the results did not allow prediction of the refolding yield based on the computed content of secondary structural elements.

Recently, an approach was published to monitor the in-column binding behaviour of mAb during a Protein A capture step by FTIR [49]. A microcolumn was packed on top of an Attenuated Total Reflection (ATR) crystal. With a PLS model, the total protein content of resin in contact with the ATR crystal was measured over multiple process steps. The publication showed, that the clean-in-place steps do not seem to be able to reliably remove all bound protein. While being an interesting scientific approach, a transfer to a larger scale system may be difficult. The proposed setup samples the resin very locally, which may not be representative for the overall column. Furthermore, lateral stress had to be applied to generate an increased contact area between resin and ATR crystal. Nevertheless, the approach shows the versatility of FTIR spectroscopy.

1.2.3 Other Spectroscopic PAT Tools

To the best of our knowledge, other spectroscopic methods have only been studied by two articles as PAT technologies for DSP of biologics. Fluorescence spectroscopy was proposed as an at-line PAT tool for a chromatographic purification step of a fusion protein [50]. Here, it was shown that the fluorescence signal could be correlated with the fraction purity from an hydrophobic interaction chromatography step separating misfolds from the product. DLS was used to investigate the unfolding and refolding process of a recombinant fusion protein from an inclusion body and its dependence on a chaotropic agent [51]. Yu et al. could accurately predict the aggregation and folding state compared to reference analytics. The method was however not applied for real-time process monitoring or control.

2

Thesis Outline

2.1 Research Proposal

Biopharmaceuticals provide new treatments for many diseases. Despite significant research and development spendings of pharmaceutical companies, the reproducible and efficient production of biopharmaceuticals remains challenging. This is reflected by changing CQAs over different lots as well as high manufacturing costs. A major goal of regulatory agencies and industry is therefore to improve process understanding and control. Already in 2004, the FDA stated: "As pharmaceutical manufacturing evolves from an art to a science and engineering based activity, application of this enhanced science and engineering knowledge [...] should improve the efficiency and effectiveness of both manufacturing and regulatory decision-making." Next to more widely adopting risk management strategies such as Quality by Design (QbD), the implementation of PAT is discussed.

Monitoring qualtity attribtues in biopharmaceutical processes is generally not straight-forward because proteins are highly complex molecules. In DSP, spectroscopic PAT sensors are of special interest. They often feature fast measurement times, the possibility for in-line implementation, and maintainable costs. Furthermore, a rich body of literature is available on protein structure determination by spectroscopy. Spectroscopic methods have been developed to measure the primary, secondary, tertiary and quaternary structure of proteins. For process monitoring, it is also interesting to measure the different protein structural levels. Each structural level contains some orthogonal information on a protein. Accessing an additional

13

structural level thus increases the selectivity of process monitoring and may provide crucial information on a certain quality attribute. Most spectroscopic methods only cover a fraction of all structural information. In consequence, one type of sensor may not be adequate for measuring all quality attributes. The implementation of whole sensor arrays may however not be feasible because of the linked investment costs. Instead, a conscious selection of different sensors or sensor combinations could be a viable solution. This is especially important since the product portfolio of biopharmaceuticals is broadening. New formats such as conjugated proteins, Virus-Like Particles (VLPs), antibody fragments, nanobodies, and Fc-fusion proteins are emerging, each of which may have different requirements towards a monitoring solution.

The ultimate goal of this thesis is to provide a PAT toolbox with spectroscopic sensors for different DSP unit operations and products. Spectroscopic sensors are selected such that all levels of protein structure are covered. The investigated sensors should permit flow-through measurements to enable in- or on-line data acquisition. Software-wise, a flexible framework with a clear, modular implementation is beneficial for an easy addition and exchange of sensors. Thus, this thesis also includes work on a generic software framework.

Based on literature, UV/Vis, FTIR, and light scattering methods are potentially sensitive to a wide range of quality attributes. These techniques also cover all levels within the protein structure. UV/Vis absorption spectroscopy in conjunction with MVDA has previously been shown to be a powerful PAT tool for chromatography by measuring primary, tertiary, and – with low sensitivity – quarternary structure. The method still needs to be evaluated for complex mixtures typically occurring during capture steps as well as highly concentrated protein solutions frequent during preparative elution steps. FTIR provides insight on the secondary structure of proteins and has been applied for monitoring protein purification at-line. In-line application promises faster measurement times and the possibility to implement closed-loop control. Finally, to provide a sensor with high sensitivity towards quarternary structure, DLS or SLS are evaluated. To reflect the current trends in biopharmaceutical industries, PAT sensors are not only tested on model proteins and mAbs but also on multiple new formats, including Antibody Drug Conjugates (ADCs), PEGylated proteins, and VLPs.

As a final topic of this thesis, Multivariate Curve Resolution (MCR) is studied as a calibration-free method to deconvolute chromatograms. Current approaches to spectroscopic data in protein chromatography widely focus on statistical models such as PLS regression. However, statistical

models require a significant amount of data for calibration. They are furthermore prone to degenerate over time due to effects not attributed by the model, such as sensor fluctuations or due to changes of the contaminant levels. MCR promises to extract information from spectroscopic data without the need for extensive model calibration and can thus simplify the application of many spectroscopic PAT tools.

2.2   Overview of Research Papers

In this section, an overview is given of the research papers written in the scope of this thesis. Chapter 3, 4, and 5 investigated the application of UV/Vis spectroscopy as PAT to different DSP unit operations. In Chapter 6, FTIR spectroscopy was implemented in-line in a preparative chromatography system and tested for different applications. Chapter 7 applies a multi-modal spectroscopic approach to monitoring the assembly of VLPs. Finally, in chapter 8, the calibration-free factorization of bilinear chromatograms by MCR is investigated. In the following, the different papers are listed with a short summary and the respective publication status.

3. **Real-time Monitoring and Control of the Load Phase of a Protein A Capture Step**

Matthias Rüdt[1], Nina Brestrich[1], Laura Rolinger, Jürgen Hubbuch ([1] contributed equally)

This manuscript investigates UV/Vis spectroscopy for monitoring the load phase during a Protein A capture step. By combining in-line UV/Vis spectral measurements and PLS regression, the break-through concentration of mAb was estimated in real-time. Based on the predicted protein concentration, the termination of the load phase was automated.

4. **Selective Protein Quantification for Preparative Chromatography using Variable Pathlength UV/Vis Spectroscopy and Partial Least Squares Regression**

Nina Brestich[1], Matthias Rüdt[1], Daniel Büchler, Jürgen Hubbuch ([1] contributed equally)

In this paper, Variable Pathlength (VP) UV/Vis spectroscopy is tested in multiple case studies for preparative chromatography. To prevent detector saturation at high protein concentrations, the optical pathlength is dynamically reduced in VP UV/Vis spectroscopy. The approach allowed to measure spectra at concentrations typically occurring during the elution phase of preparative chromatography. To retrieve single component concen-

trations from the spectral data, PLS regression was applied. The approach was subsequently used to control product pooling.

Manuscript published in Chemical Engineering Science, 176, 157-164, 2018.

5. **Monitoring of Antibody-Drug Conjugation Reactions with UV/Vis Spectroscopy**

Sebastian Andris[1], Matthias Rüdt[1], Jonas Rogalla, Michaela Wendeler, Jürgen Hubbuch ([1] contributed equally)

This study focuses on establishing a fast approach for monitoring the conjugation reaction during the production of ADCs. The applicability of UV/Vis spectroscopy was demonstrated relating the reaction to changes in the solvatization of chromophores. Based on the UV/Vis spectral data, the progress of the reaction of a mAb with two different surrogate drugs could be observed. Experiments were performed at two scales with two different detectors to investigate the reproducibility and scalability.

Manuscript published in the Journal of Biotechnology, 288, 15-22, 2018.

6. **In-line Fourier-transform Infrared Spectroscopy as a Versatile Process Analytical Technology in Preparative Protein Chromatography**

Steffen Großhans[1], Matthias Rüdt[1], Adrian Sanden[1], Nina Brestrich, Josefine Morgenstern, Stefan Heissler, Jürgen Hubbuch ([1] contributed equally)

FTIR spectroscopy provides orthogonal information compared to the more common monitoring methods in DSP. So far, it has however never been implemented in-line for preparative protein chromatography. Here, a custom-made setup was realized and applied to multiple studies with industrial relevance. Next to monitoring the elution of Polyethylene Glycole (PEG)-modified species, the method was used for selective protein quantification based on differences in the secondary structure. Finally, in-line FTIR was used to selectively quantify a process-related impurity.

Manuscript published in the Journal of Chromatography A, 1547, 37-44, 2018.

7. Process Monitoring of Virus-Like Particle Reassembly by Diafiltration with UV/Vis and Spectroscopy Light Scattering

Matthias Rüdt[1], Philipp Vormittag[1], Nils Hillebrandt, Jürgen Hubbuch ([1] contributed equally)

VLPs are promising new biopharmaceuticals for many applications. To achieve their full potential, they need to be dis- and reassembled in vitro. This study established a new experimental setup with multimodal spectroscopic sensors to enable a holistic monitoring of the assembly process by diafiltration. UV/Vis spectroscopy allowed to measure the protein concentration and monitor the VLP tertiary structure. The VLP quaternary structure was observed by SLS and DLS. Based on the rich information provided by the sensors, aggregate inhibition of VLP assembly was identified as a potential bottleneck of the unit operation. To realize real-time monitoring of the VLP concentration, a PLS model was calibrated on the UV/Vis spectral data.

Manuscript under review.

8. Factorization of Preparative Protein Chromatograms with Hard-Constraint Multivariate Curve Resolution and Second Derivative Pretreatment

Matthias Rüdt, Sebastian Andris, Robin Schiemer, Jürgen Hubbuch

In this study, the application of MCR is investigated for the factorization of preparative protein chromatograms. Currently, PAT for protein chromatography only focuses on calibrated statistical models. This study demonstrates the applicability of hard-constraint MCR for evaluating this type of data. To increase spectral differences between proteins and reduce spectral drifts, MCR was extended to deal with second derivative data. The method was tested in multiple case studies and the results compared to reference data.

Manuscript in press in the Journal of Chromatography A.

<div style="text-align: right; font-size: 3em;">3</div>

Real-time Monitoring and Control of the Load Phase of a Protein A Capture Step

Matthias Rüdt[*,1], Nina Brestrich[*,1], Laura Rolinger[1], Jürgen Hubbuch[1]

[*]   Contributed equally
[1]   Institute of Engineering in Life Sciences, Section IV: Biomolecular Separation Engineering, Karlsruhe Institute of Technology (KIT), Germany

Abstract

The load phase in preparative Protein A capture steps is commonly not controlled in real-time. The load volume is generally based on an off-line quantification of the mAb prior to loading and on a conservative column capacity determined by resin-life time studies. While this results in a reduced productivity in batch mode, the bottleneck of suitable real-time analytics has to be overcome in order to enable continuous mAb purification. In this study, PLS modeling on UV/Vis absorption spectra was applied to quantify mAb in the effluent of a Protein A capture step during the load phase. A PLS model based on several breakthrough curves with variable mAb titers in the harvested cell culture fluid was successfully calibrated. The PLS model predicted the mAb concentrations in the effluent of a validation

experiment with a Root Mean Square Error (RMSE) of 0.06 mg/ml. The information was applied to automatically terminate the load phase, when a product breakthrough of 1.5 mg/ml was reached. In a second part of the study, the sensitivity of the method was further increased by only considering small mAb concentrations in the calibration and by subtracting an impurity background signal. The resulting PLS model exhibited a RMSE of prediction of 0.01 mg/ml and was successfully applied to terminate the load phase, when a product breakthrough of 0.15 mg/ml was achieved. The proposed method has hence potential for the real-time monitoring and control of capture steps at large scale production. This might enhance the resin capacity utilization, eliminate time-consuming off-line analytics, and contribute to the realization of continuous processing.

3.1   Introduction

A capture step is the first unit operation in the protein purification process which is used to bind the target protein from crude Harvested Cell Culture Fluid (HCCF). It increases product concentration as well as purity and prevents proteolytic degradation. Due to its high selectivity, Protein A capture is widely used in current mAb purification platform processes [53]–[57].

A difficulty in Protein A capture is a lack of real-time analytics for mAb quantification in the HCCF and in the column effluent during loading. As both the mAb and impurities contribute to the absorption at 280 nm (A_{280}), single wavelength measurements are not suitable as selective analytics [58]. To determine the mAb titer in the HCCF, elaborate off-line analytics is commonly performed [12], [13]. As mAb titers are influenced by variability in the cell culture, this off-line analytics has to be repeated for every lot in order to adapt the load volume onto the column [12]. While this results in a reduced productivity in batch mode, the bottleneck of suitable real-time analytics has to be overcome to enable continuous mAb purification.

In addition to the mAb titer in the HCCF, the optimal load volume onto the column is also influenced by the resin capacity. Due to leaching and degradation of the Protein A ligands as well as pore and ligand blocking by leftover impurities or product, the capacity of the resin decreases over cylce time [59]. In batch mode, a conservative loading is commonly applied to avoid breakthrough of the expensive product at the cost of productivity. In contrast to that, columns are overloaded in continuous mode to maximize productivity [60]. In this case, the determination of the the percentual product breakthrough is necessary for process control [61].

To perform (near) real-time process monitoring and control, several PAT tools have been developed to enable fast mAb quantification in the cell culture fluid and in the column effluent during loading. For instance, at-line mid-IR spectroscopy in combination with multivariate data analysis has been applied for secreted mAb quantification during a Chinese Hamster Ovary (CHO) cell culture process [62]. Selective mAb quantification in upstream processing was also successfully realized by at-line matrix-assisted laser desorption/ionization mass spectrometry [63]. For the control of the load phase of a two column continuous protein A chromatography process, which was connected to a CHO perfusion culture, at-line analytical chromatography was applied [64]. At-line monitoring however bears the risk of human errors resulting in contamination, time-delays, or missing data.

In order to minimize human impact, automated sampling can be applied. Automated analytical chromatography has been used in upstream processing to monitor the mAb titers [65]–[67]. In downstream processing, this technique was successfully used for mAb quantification in the column effluent during the load phase of Protein A chromatography. As soon as 1 % mAb breakthrough was detected, the load phase was automatically terminated [12]. Automated analytical chromatography is relatively easy to develop and equipment is commercially available. However, the equipment is expensive and the technique error-prone. Besides from the risk of contamination, the time delay between sampling and analytical results bears the risk of late reaction or requires a slow-down of the process.

PAT tools that operate in real-time, such as Ultraviolet (UV)-based methods, overcome these limitations. In a patent application, a UV-based control method for determining binding capacities in Protein A capture was disclosed [68]. The method is based on the calculation of a difference signal between two detectors situated at the column in- and outlet. During the load phase, the post column signal is supposed to stabilize and is referred to as impurity baseline. As soon as the mAb breaks through, there is an increase in the post-column UV signal above the impurity baseline which corresponds to a breakthrough level of the product. Consequently, the method is very suitable for determining column switching times in continuous Protein A capture. It allows for an equal loading in terms of percentual breakthrough regardless of the mAb titer variability in the feed or decreasing column capacities. However, it requires two detectors posing a risk of unequal detector drifts. A further limitation might be displacement effects of contaminants that prevent a stabilized impurity baseline. The technique might also be limited to the equipment of the future patent holder.

Another recently published UV/Vis-based method for monitoring and control in protein chromatography applies UV/Vis absorption spectra in-

stead of single wavelength measurements [42], [43]. Different protein species exhibit distinct variations in their UV absorption spectra. Consequently, PLS technique has been used to correlate absorption spectra with selective protein concentrations. The method was successfully applied for a selective in-line protein quantification and for product purity-based pooling decisions in real-time. However, no load control in Protein A chromatography has been performed so far using this technique.

In this study, PLS models correlating UV/Vis absorption spectra with mAb concentrations were applied for real-time monitoring and control of the load phase in Protein A chromatography. In contrast to previous publications in this field, this application requires the monitoring of one protein in the background of many protein and non protein-based contaminants. For the PLS model calibration, several breakthrough experiments were performed and the corresponding absorption spectra of the effluent were acquired. In order to generate variable mixing ratios of mAb and contaminants for a PLS model training data set, experiments with variable mAb titers in the feed were performed. The column effluent was collected in fractions and analyzed using analytical Protein A chromatography. The recorded absorption spectra were averaged according to the fraction time and correlated with the determined mAb concentrations using PLS technique. The PLS model was eventually applied for a real-time control of the load phase and terminated loading, when 5% or 50% product breakthrough was reached.

3.2   Materials and Methods

3.2.1   Cell Culture Fluid and Buffers

HCCF and mock were obtained from Lek Pharmaceuticals d.d. (Mengeš, Slovenia) and stored at $-80°$C before experimentation. The HCCF and mock were filtered with a cellulose acetate filter with a pore size of $0.22\,\mu$m (Pall, Port Washington, NY, USA) before use. In order to achieve a variable mAb concentration in the feed, the HCCF was diluted with mock.

For all preparative runs, the following buffers were applied: Equilibration with $25\,$mM tris and $0.1\,$M sodium chloride at pH 7.4, wash with $1\,$M tris and $0.5\,$M potassium chloride at pH 7.4, elution with $20\,$mM citric acid at pH 3.6, sanitization with $50\,$mM sodium hydroxide and $1\,$M sodium chloride, and storage with $10\,$mM sodium phosphate, $130\,$mM sodium chloride, 20% ethanol.

For analytical Protein A chromatography, column equilibration was carried out using a buffer with $10\,\mathrm{mM}$ phosphate (from sodium phosphate and potassium phosphate) with $0.65\,\mathrm{M}$ sodium ions (from sodium chloride and potassium chloride) at pH 7.1. Elution was performed with the same buffer, but titrated to pH 2.6 with hydrochloric acid. All buffer components were purchased from VWR, West Chester, USA. The buffers were prepared with Ultrapure Water (PURELAB Ultra, ELGA LabWater, Viola Water Technologies, Saint-Maurice, France), filtrated with a cellulose acetate filter with a pore size of $0.22\,\mu\mathrm{m}$ (Pall), and degassed by sonification.

3.2.2 Chromatographic Instrumentation

All preparative runs were realized with an Akta Pure 25 purification system controlled with Unicorn 6.4.1 (GE Healthcare, Chalfont St Giles, UK). The system was equipped with a sample pump S9, a fraction collector F9-C, a column valve kit (V9-C, for up to 5 columns), a UV-monitor U9-M ($2\,\mathrm{mm}$ pathlength), a conductivity monitor C9, and an I/O-box E9. Additionally, an UltiMate 3000 Diode Array Detector (DAD) equipped with a semi-preparative flow cell ($0.4\,\mathrm{mm}$ optical pathlength) and operated with Chromeleon 6.8 (Thermo Fisher Scientific, Waltham, USA) was connected to the Akta Pure. The DAD was positioned between the conductivity monitor and the fraction collector.

The communication between Unicorn and Chromeleon was implemented analogous to the protocol published in [42]. Shortly, Unicorn triggers the DAD data acquisition by sending a digital signal to a Matlab script (MathWorks, Natick, USA), which communicates with Chromeleon via a Visual Basics for Application Macro (Microsoft, Redmond, USA). If a certain condition such as a defined mAb concentration is fulfilled, the Matlab script sends a signal back to Unicorn to terminate a phase in the chromatographic method.

Reference analysis of collected fractions was performed using a Dionex UltiMate 3000 rapid separation liquid chromatography system (Thermo Fisher Scientific). The system was composed of a HPG-3400RS pump, a WPS-3000 analytical autosampler, a TCC-3000RS column thermostat, and a DAD-3000RS detector.

3.2.3 Chromatography Runs

In order to generate variable mixtures between mAb and impurities for the PLS model calibration and validation, breakthrough experiments with variable mAb titers in the feed were performed. The mAb titers in the different

experiments were 2.7, 2.85, 3, 3.15, and 3.3 mg/ml. For each experiment, a Sartobind 2 ml Protein A membrane (Sartorius, Göttingen, Germany) was first equilibrated for 3 membrane volumes (MVs) and then loaded with 33.15 mg of mAb. At the beginning of the load phase, the DAD was triggered to record absorption spectra between 200-410 nm and the membrane flow-through was collected in 200 µl fractions. After a first wash with equilibration buffer for 4.5 MVs, the membrane was flushed with wash buffer for 5.5 MVs and with equilibration buffer for 4.5 MVs. Elution was carried out for 5 MVs followed by a re-equilibration of 1.5 MVs. Eventually, the column was sanitized for 5 MVs and, between the runs, kept in the storage buffer. The flow rate was 1 ml/min for all phases and experiments.

3.2.4   Analytical Chromatography

As displayed in Figure 3.1, the collected fractions of all runs were examined by analytical Protein A chromatography to obtain the mAb concentrations. For each sample, a 2.1x30 mm POROS prepacked Protein A column (Applied Biosystems, Foster City, USA) was equilibrated with 2.6 column volumes (CVs) of equilibration buffer, flowed by an injection of 20 µl sample. The column was then equilibrated with 0.8 CVs of equilibration buffer and eluted with 1.4 CVs of elution buffer. The flow rate was 2 ml/min for all phases and experiments.

3.2.5   Data Analysis

For the correlation of the absorption spectra with the mAb concentrations, PLS technique was applied using SIMCA (MKS Data Analytics Solutions, Umeå, Sweden). SIMCA applies the NIPALS-algorithm for PLS. Before performing PLS, all spectra were preprocessed by mean centering using SIMCA. PLS finds variation in the spectral data matrix, which is relevant for the correlation with the mAb concentrations and thereby separates information in the matrix from detector noise [22], [69], [70]. In order to achieve this separation, collinearity in the data is reduced by summarizing variables (here wavelengths) with similar information in LVs. This is done in a way such that the content of relevant information for the correlation included in each LV is highest for the first LV and decreases for the following ones. The number of applied LVs in a PLS model is hence a measure of data reduction and only a few LVs are required to obtain the correlation between absorption spectra and mAb concentrations.
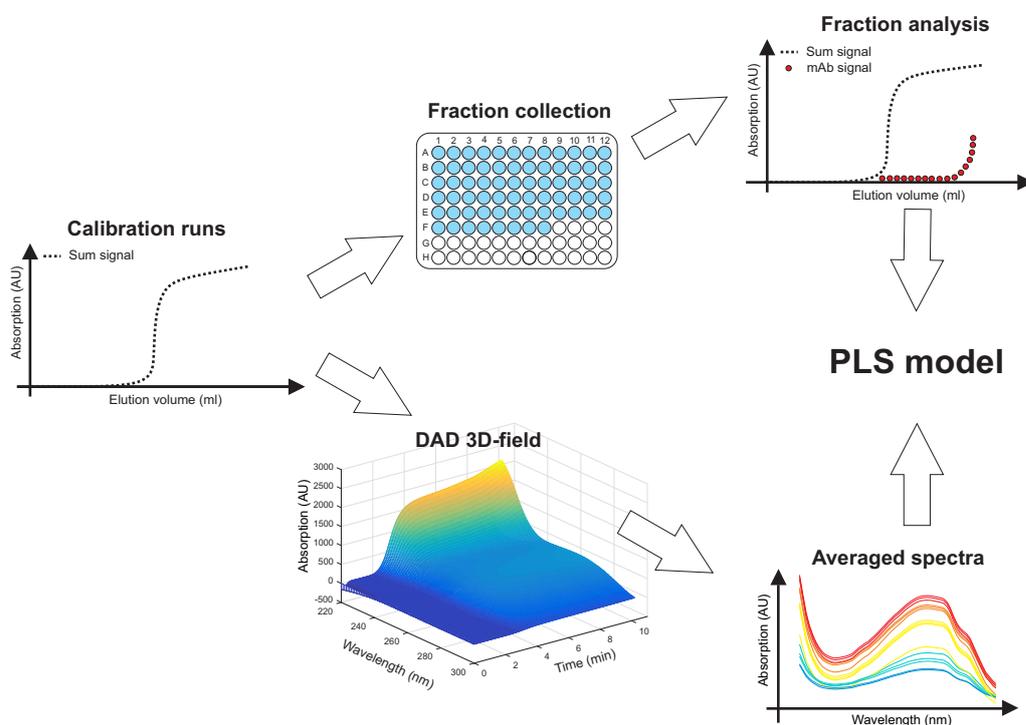
Figure 3.1: Experimental procedure for the PLS model calibration: For each calibration run, $200\,\mu l$ fractions were collected and analyzed by analytical Protein A chromatography to obtain the mAb breakthrough curves. In addition, averaged spectra corresponding to the fraction size were calculated from the time, wavelength, and absorption 3D-field. Averaged spectra and mAb concentrations were eventually correlated using PLS technique.

The number of applied LVs has to be evaluated thoroughly to avoid under- or overfitting of a model. In order to determine a reasonable number of LVs, the root mean square error (RMSE) for the prediction of validation samples is usually determined in dependence on the number of LVs applied in a PLS model. The minimum corresponds to the optimal number of LVs. In this study, cross validation was performed to determine an optimal number of LVs. Therefore, the calibration data was separated into seven groups. One group was then excluded during model calibration and the RMSE for theses samples was calculated subsequently. For every number of LVs, this procedure was performed until each group was excluded. Based on the so obtained number of LVs, completely independent runs were predicted to evaluate the final models.

A first PLS model calibration was based on the results of the runs with the following mAb titers in the feed: 2.7, 2.85, 3.15, and 3.3 mg/ml. The

results of the corresponding spectral acquisitions are time, wavelength and absorption 3D-fields. The 3D-fields were averaged in time according to the fraction duration as displayed in Figure 3.1. The results of theses calculations were stored in an absorption matrix. Afterwards, PLS was carried out to correlate the mAb concentrations of the collected fractions with the the corresponding absorption matrix. For lower protein concentrations, a second PLS model was calibrated. Only samples with mAb concentrations below 0.5 mg/ml were considered in the model calibration. For those samples, a background subtraction was performed. As soon as the change in absorption signal after impurity breakthrough fell under a predefined threshold, an average absorption was calculated for every wavelength. This impurity background was subtracted from the absorption of all following data points.

3.2.6   Real-Time Monitoring and Control

The first calibrated PLS model was subsequently applied for a real-time monitoring of the mAb concentrations in a run with a mAb titer of 3 mg/ml in the feed. While the calibration of the PLS model was performed using averaged spectra, predictions were based on the 3D-fields. This means that the a spectrum at each time point was applied to predict the mAb concentrations. The absorption spectra of the effluent were recorded and translated into mAb concentrations in real-time by the calibrated PLS model. The calculation of the mAb concentrations was executed in Matlab. In a first run, a stop criterion of 1.5 mg/ml mAb concentration (50 % product breakthrough) was set in the Matlab evaluation script. As soon as the termination criterion was reached, a digital signal was send from Matlab to Unicorn and the load phase was terminated. In a second run, the stop criterion to terminate the load phase was set to a target concentration of 0.15 mg/ml (5 % product breakthrough). For this condition, the second PLS model was used.

3.3   Results and Discussion

As described above, the breakthrough of mAb was monitored in real-time by UV/Vis spectroscopy in combination with a PLS model. To calibrate the PLS model, 4 chromatographic runs at mAb concentrations of 2.7, 2.85, 3.15 and 3.3 mg/ml in the feed were performed and analyzed by off-line analytics. The model was eventually confirmed by performing a real-time control of two runs with a mAb titer of 3 mg/ml. The difference in the

mAb titers in the feed ensured variable mixing ratios between product and contaminants. This was done to imitate variability in upstream processing and to span a calibrated design space for the PLS model.

3.3.1 PLS Model Calibration

The results of the model calibration are illustrated by Figure 3.2. It compares the A_{280} (recorded at a pathlength of 0.4 mm and displayed as dashed black line) to the concentrations measured by off-line analytics (blue bars) and the signal calculated by the calibrated PLS model (solid red lines). The number of LVs was set to 4 based on a minimal RMSE of 0.08 mg/ml in the cross validation. The calibrated PLS model was applied to evaluate all 3D-fields. In contrast to model calibration, where averaged spectra were used, the spectral raw data at each time point was translated into concentrations. The estimated concentrations by the PLS model closely follow the measured values by off-line analytics. It is worth noting that no clear plateau of the A_{280} is reached after the breakthrough of media components. Instead, the A_{280} continuous to increase. This may be caused by different impurities being retained differently on the membrane. Indeed, it has previously been shown, that major interactions between HCPs, the stationary phase and mAbs may occure [71], [72]. The advent of mAb breakthrough cannot be clearly distinguished from A_{280} alone. Based on the multivariate spectral data, the PLS model is able to predict protein concentrations, which allows for real-time monitoring and control.

3.3.2 Real-Time Monitoring and Control

For the confirmation of the obtained results, the calibrated PLS model was used to control the load phase of a Protein A capture step in real-time. In a first run, a target breakthrough concentration of 1.5 mg/ml was set, which corresponds to 50 % product breakthrough. Figure 3.3 A shows the A_{280} (dashed black line), the real-time prediction of mAb concentrations (solid red line) and the corresponding off-line analytics (blue bars). The model reached an RMSE for prediction of 0.06 mg/ml compared to the off-line analytics. This approach may be of interest for controlling a continuous chromatography system. In this context, the prediction of lower mAb concentrations is not so crucial.For a possible application in batch chromatography, the sensitivity of the model was further improved. A second PLS model was hence calibrated based on the calibration data set as described in the method section. The recalibration was performed to increase the sensitivity in the given concentration range. It was noticed, that
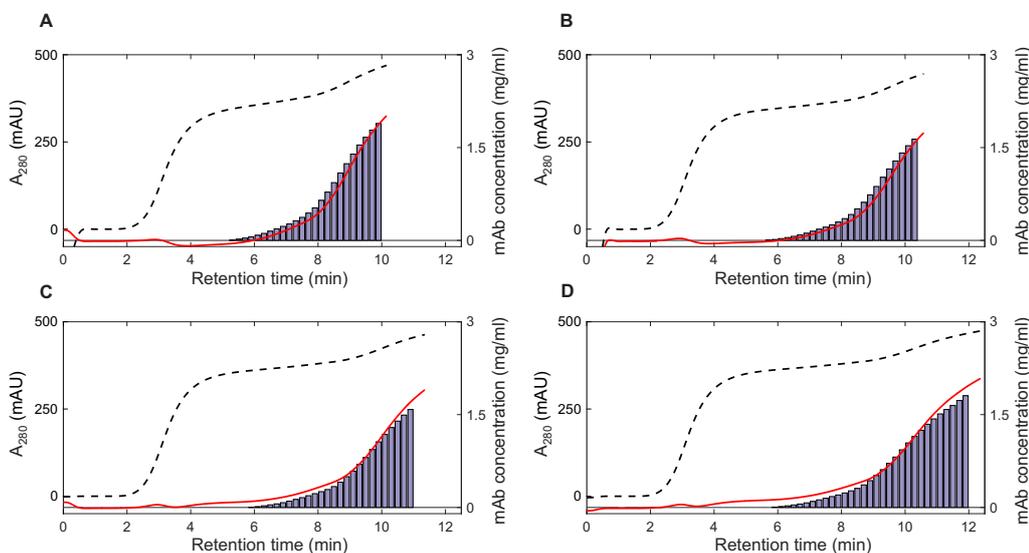
Figure 3.2: Results of the PLS model calibration. The A_{280} (measured at a pathlength of 0.4 mm and displayed as dashed black line) is compared with the results of the off-line analytics for mAb quantification (blue bars). The PLS model prediction is illustrated as red lines. The four runs exhibited variable mAb titers in the feed A: 3.3 mg/ml, B: 3.15 mg/ml, C: 2.85 mg/ml, D: 2.7 mg/ml.
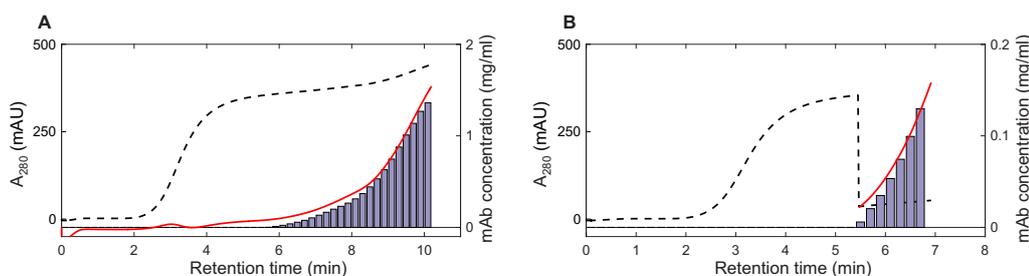
Figure 3.3: Results of the model evaluation by performing a real-time control of the load phase using a mAb titer of 3 mg/ml in the feed. The PLS model prediction (red lines) is compared with the results of the off-line analytics (blue bars) as well as the A_{280} (measured at a pathlength of 0.4 mm and displayed as dashed black line). The load phase was automatically terminated, when a mAb concentration in the effluent of A: 1.5 mg/ml or B: 0.15 mg/ml was reached. The sudden decrease in the A_{280} arises from the background subtraction.

it is difficult to accurately calibrate a PLS model for broad concentration ranges. By reducing the concentration calibration range, smaller RMSE values could be achieved. The model was used predict and stop a load phase in a second run at $0.15\,\mathrm{mg/ml}$, which corresponds to $5\,\%$ product breakthrough. The results of this second run are displayed in Figure 3.3 B. As an impurity background was subtracted to increase the sensitivity of the method, the A_{280} suddenly decreases. The second PLS model reached an RMSE for prediction of $0.01\,\mathrm{mg/ml}$.

During both runs, the respective load phases were successfully terminated close to the intended breakpoints. In Table 3.1, a summary of intended and measured mAb concentrations in the last fraction of both confirmation runs is shown. The Matlab script sent a digital signal to Unicorn and terminated the load phase, when the targeted breakthrough concentration was reached. As the targeted breakthrough set points were concentrations at discrete time points, they are expected to be slightly higher than the concentrations of the last fraction determined by off-line analytics. This was observed for both confirmation runs (cf. Table 3.1). For an easier comparison between model and off-line analytics, a concentration based on an averaged absorption spectrum was calculated for the last fractions of both runs and compared with the corresponding off-line analytics. For the first run, the deviation between prediction and reference was $8.0\,\%$, while for the second run a deviation of $2.3\,\%$ was found. This demonstrates that the described method can be successfully used to control the load phase in a Protein A capture step.

Table 3.1: Results of both confirmation runs: The targeted concentration to terminate loading is compared with the mAb concentration in the last fraction determined by off-line analytics. In addition, a PLS model prediction for the last fraction based on an averaged absorption spectrum is shown for comparison.

c_{target} [mg/ml]	$c_{\text{analytics}}$ [mg/ml]	$c_{\text{mean,PLS}}$ [mg/ml]
1.5	1.36	1.469
0.15	0.129	0.126

3.4   Conclusion and Outlook

A real-time monitoring and control of the load phase in a Protein A capture step was successfully realized in this study. It was demonstrated that PLS

modelling on UV/Vis absorption spectra can be applied to quantify mAb in the effluent during the load phase despite of the background of many protein and non protein-based impurities. Based on the quantification, the load phase was automatically terminated, when a product breakthrough concentration of $1.5\,\mathrm{mg/ml}$ or $0.15\,\mathrm{mg/ml}$ was reached. Consequently, the proposed method has potential for the monitoring and control of capture steps at large scale production. In batch chromatography, the loading volume may be defined dynamically to allow for increased resin capacity utilization while still keeping the product loss small. Additionally, time-consuming off-line determination of the mAb titer in HCCF could be eliminated. The method may also be interesting for controlling column switching times in continuous chromatographic capture steps. Future challenges are especially related to the scale up and robustness of the method. Regarding the latter, especially upstream variations should be calibrated into the PLS model. Research will now focus on the migration of the method to the control of continuous capture steps.

Acknowledgment

4

Selective Protein Quantification for Preparative Chromatography using Variable Pathlength UV/Vis Spectroscopy and Partial-Least Squares Regression

Nina Brestich[*,1], Matthias Rüdt[*,1], Daniel Büchler[1], Jürgen Hubbuch[1]

[*]   Contributed equally

[1]   Institute of Engineering in Life Sciences, Section IV: Biomolecular Separation Engineering, Karlsruhe Institute of Technology (KIT), Germany

Abstract

In preparative protein chromatography, broad dynamic ranges of protein concentrations as well as co-elution of product and impurities are common. Despite being the standard in biopharmaceutical production, monitoring of preparative chromatography is generally limited to surrogate signals, e.g. UV absorbance at 280 nm. To address this problem, VP spectroscopy in conjunction with PLS was used to monitor preparative chromatography. While VP spectroscopy enabled the acquisition of absorbance data for a

broad concentration range, PLS modeling allowed for the differentiation between the protein species. The approach was first implemented for monitoring the separation of lysozyme from cytochrome c at an overall loading density of 92 g/L. The same method was then applied to the polishing step of a mAb at 40 g/L loading density. For PLS model prediction of the mAb monomer and the HMWs, the RMSE was 1.07 g/L and 0.42 g/L respectively. To demonstrate the usability of the approach for in-line control, pooling decisions for both separation problems were subsequently taken based on the computed concentrations or thereof derived purities. In summary, VP spectroscopy in conjunction with PLS modeling is a promising option for in-line monitoring and control of future chromatography steps at large scale.

4.1 Introduction

In current purification processes of biopharmaceuticals, preparative liquid chromatography is key for separating the target product from media components, Deoxyribonucleic Acid (DNA), host cell proteins, and product related impurities [19]. The method is used because of its high separative power while minimizing product loss. Despite being the standard, monitoring of preparative chromatography is generally limited to surrogate signals, e.g. UV absorbance at 280 nm. The PAT initiative of the US FDA however promotes the acquisition of critical quality attributes in (near) real-time [1]. Especially for chromatography, PAT is an active field of research [8], [73].

Preparative chromatographic processes are generally run at high loading densities to realize efficient processes [19]. High loading densities subsequently lead to a broad range of protein concentrations eluting from the column. Thus, monitoring techniques must feature a broad dynamic range to capture the peak concentration as well as lower concentrated impurities in the peak flanks. Furthermore, due to the limited resolution of preparative chromatography, baseline separation between product and impurities is rarely achieved. PAT techniques should therefore selectively quantify the product and the main impurities. Due to the rapid changes in chromatography, the typical decision time within chromatographic processes is limited [73], [74].

In literature, a number of different approaches have been proposed for (near) real-time monitoring of chromatography. On-line monitoring and control of preparative chromatography has been realized by automated sampling and subsequent analysis by high performance liquid chromatography [11], [16], [75]. The broad concentration ranges in preparative liquid chro-

matography can be managed easily by varying injection volumes. Disadvantages comprise time delay due to sample handling and assay times as well as the risk of contamination.

FTIR spectroscopy has been applied for at-line monitoring of downstream processes [62]. Samples were taken at various stages of downstream processing, dried and subsequently measured by FTIR. The method allowed to quantify product, high molecular weight variants (HMWs), and host cell proteins over a broad range of concentrations. Despite showing promising results, at-line measurements bear the risk of introducing operator errors and may increase the risk of contamination. Furthermore, the delay due to sample handling and measuring times may be too long for the typical decision times in chromatography.

Previously, UV/Vis spectroscopy has been proposed as a method for in-line monitoring of chromatographic processes [42], [43], [76]. Its usefulness was shown for multicomponent mixtures of model proteins and for real-life separation problems. Based on UV/Vis spectral data and PLS modeling, process decisions were taken such as the beginning and end of product pooling. Other applications include high throughput process development [41], [77] and coupling with chromatography modeling [10]. While being very fast and accurate, previous applications of in-line monitoring using UV/Vis spectroscopy only took process decisions for separation problems in diluted conditions. The problem of broad concentration ranges occurring in preparative liquid chromatography was not addressed.

To increase the dynamic range of spectroscopic acquisitions, measurement cells have been designed which allow to continuously change the optical pathlength to achieve ideal sensitivity for virtually any analyte concentration [78], [79]. This approach in conjunction with PLS modeling was later successfully applied for monitoring the chemical oxygen demand of wastewater [80]. The methodology has also been further developed by a commercial vendor for protein related applications. With the commercialized product, spectra of proteins were acquired at a broad range of concentrations [81]. Samples were studied for spectral effects of protein-protein interactions in UV/Vis with protein concentrations up to $250\,\text{g/L}$. Recently, an additional product line has been launched for in-line VP measurements. The device is theoretically able to provide measurement results up to absorbances of approximately $80\,\text{AU/mm}$ corresponding to a mAb concentration of almost $600\,\text{g/L}$.

In this publication, we demonstrate in-line monitoring of preparative chromatography runs by UV/Vis VP spectroscopy in conjunction with PLS modeling. While VP spectroscopy allowed to monitor chromatographic processes at almost arbitrary protein concentrations, PLS models selectively

quantified multiple co-eluting species from spectral data. The approach was first implemented for the separation of a mixture of the model proteins lysozyme (lys) and cytochrome c (cyt c) at high loading densities. Former was considered the product, latter the contaminant with high respectively low concentration in the feed. Subsequently, a preparative polishing step of mAb monomer and its HMWs was monitored. To demonstrate the usefulness of this approach for process control, the pooling of products in both separation problems was controlled in-line based on either the predicted concentrations of eluting proteins for the model system or on the calculated purity for the mAb purification.

4.2   Materials and Methods

In both studies, cation exchange chromatography runs with different gradient lengths were executed for PLS model calibration and confirmation. Thereby, different mixing ratios and concentrations of proteins were obtained in order to span a calibration space for the PLS models.

4.2.1   Chromatography Instrumentation

All preparative chromatography experiments were performed on an ÄKTA Pure 25 system equipped with a sample pump S9, a fraction collector F9-C, a column valve kit (V9-C, for up to 5 columns), a UV-monitor U9-M (2 mm pathlength), a conductivity monitor C9, and an I/O-box E9. The system was controlled with Unicorn 6.4.1. (all GE Healthcare, Chalfont St Giles, UK). The column effluent was monitored using a FlowVPE VP UV/Vis spectrometer (C Technologies, Bridgewater, US). It was integrated into the flow path of the ÄKTA system between the conductivity monitor and fraction collector.

The reference analytics of collected fractions was performed using a UltiMate 3000 rapid separation liquid chromatography system (Thermo Fisher Scientific, Waltham, US). The system was composed of a HPG-3400RS pump, a WPS-3000 analytical autosampler, a TCC-3000RS column thermostat, and a DAD-3000RS detector.

4.2.2   VP Spectroscopy

The FlowVPE VP spectrometer uses a mobile optical fiber to change the pathlength L of the flow cell (cf. Figure 4.1). A VP measurement cycle consists of a screening phase and a subsequent acquisition phase. During

the screening phase, optimal pathlengths for the acquisition phase are estimated based on 3 measurements at $L_s \in \{5, 25, 100\}$ µm at 280 nm. During the acquisition phase, absorbances $A_{w,i}$ are measured at five different pathlengths $i \in \{1, \ldots, 5\}$ for each wavelength of interest w. Based on the Lambert-Beer law, a linear correlation is assumed between the absorbance estimate $\tilde{A}_w$ and the pathlength L (Equation (4.1)). By solving the least squares problem in Equation (4.2), slope m_w and intercept b_w of the regression are estimated.

$$\tilde{A}_w(L; b_w, m_w) = b_w + m_w L \qquad (4.1)$$

$$\min_{b_w, m_w} \sum_i \left(A_{w,i} - \tilde{A}_w(L_i; b_w, m_w) \right)^2 \qquad (4.2)$$

As slope spectra are obtained during a whole chromatography run, the result of an experiment is a point set in (time, wavelength, slope), thus a 3-dimensional chromatogram. By equating Equation (4.1) and the Lambert-Beer law and taking the derivative with respect to L, Equation (4.3) is obtained.

$$m_w = C \varepsilon_w^T \qquad (4.3)$$

This shows, that the obtained slope spectra are correlated with the protein concentrations C over the absorbance coefficients ε_w. Since only the slopes are of interest, the fiber offset has no impact on the final result. Generally, for mid-UV spectra measurement cycles take approximately 30 s.

4.2.3 Case Study I: Separation of Cyt c from Lys

Proteins and Buffers

As model system, a protein mixture consisting of lys from hen egg white and cyt c from equine heart was used (both Sigma-Aldrich, St. Louis, US). For the preparative runs, the equilibration buffer was 20 mM sodium phosphate (pH 7). Elution was carried out with 20 mM sodium phosphate and 500 mM sodium chloride (pH 7). For the in-line pooling decision, the sodium chloride concentration was increased to 550 mM to simulate a process disturbance. For reference analytics (analytical Cation-Exchange (CEX)), equilibration was performed with 20 mM Tris (pH 8) and elution was carried out with 20 mM Tris and 700 mM sodium chloride (pH 8). All buffer components were purchased from VWR, West Chester, US. The buffers were prepared with Ultrapure Water (PURELAB Ultra, ELGA LabWater, Viola Water Technologies, Saint-Maurice, France), filtrated with a cellulose acetate filter
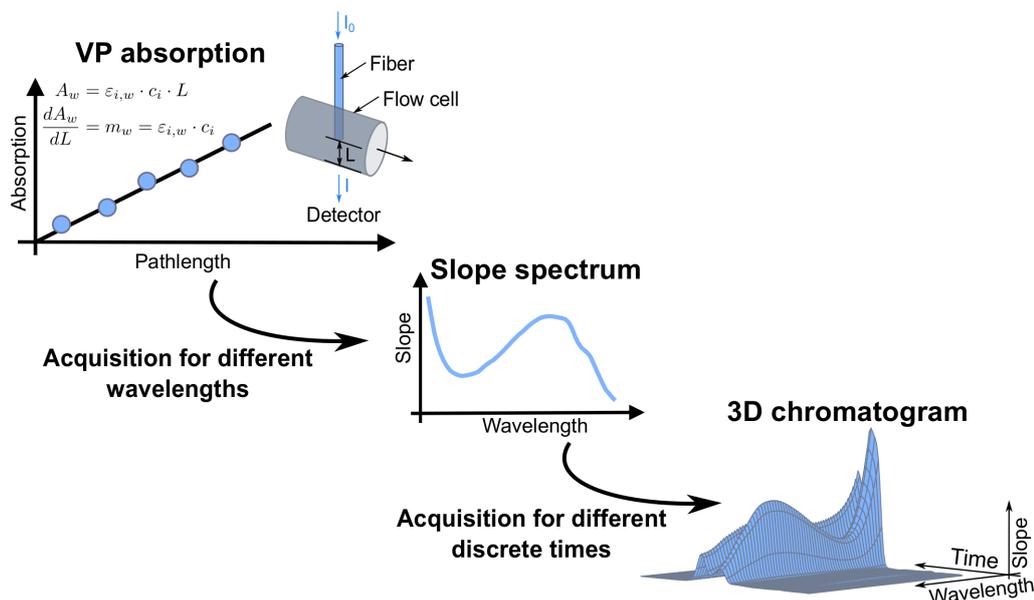
Figure 4.1: VP spectroscopy in chromatography: Each measurement cycle, variable pathlengths L are set by actuating the optical fiber of the FlowVPE. After the screening phase for the identification of the optimal measurement pathlengths, UV absorbance spectra at five different pathlengths are acquired. The slope m_w is calculated for each wavelength giving a slope spectrum. For a whole chromatography run, this results in a 3-dimensional chromatogram in (time, wavelength, slope).

with a pore size of $0.22\,\mu m$ (Pall, Port Washington, US), and degassed by sonification before usage.

Chromatography Runs

A HiTrap $16\,mm \times 25\,mm$ column prepacked with SP Sepharose FF (GE Healthcare) was first equilibrated for $5\,CV$ and then loaded with $418\,mg$ lys and $41.8\,mg$ cyt c. After a wash of $1\,CV$ with equilibration buffer, the proteins were eluted by performing a linear gradient from $0\,\%$ to $100\,\%$ elution buffer. The gradient length was $2\,CV$, $4\,CV$, $6\,CV$, and $8\,CV$. While the results of the runs with gradient lengths $2\,CV$, $4\,CV$, and $8\,CV$ were applied for the PLS model calibration, the run with a gradient length of $6\,CV$ was used to confirm the model. At the beginning of each linear gradient, data acquisition of the FlowVPE was started and slope spectra from $240\,nm$ to $300\,nm$ with $2\,nm$ resolution were obtained. After the linear gradient elution, the column was regenerated for $3\,CV$ with the elution buffer. The flow

rate was kept constant at 0.5 mL/ min for all steps and experiments. During the elution and regeneration, 1000 µL fractions were collected in deep well plates (VWR).

Reference Analytics

The collected fractions were analyzed by analytical cation exchange chromatography. For each injection, the Proswift SCX-1S 4.6 mm × 50 mm column (Thermo Fisher Scientific) was first equilibrated for 2 min, loaded with sample, washed for 0.5 min with equilibration buffer, and eluted with a linear salt gradient from 10 % to 100 % elution buffer in 2 min. The column was subsequently regenerated with 100 % elution buffer for 1 min. The flow rate was constant at 1.5 mL/ min for all steps.

4.2.4 Case Study II: Separation of HMWs from mAb Monomer

Proteins and Buffers

mAb Protein A pool was obtained from Lek Pharmaceuticals d.d. (Mengeš, Slovenia) and stored at −80 °C before experimentation. Because of the low, not detectable HMW content of the mAb Protein A pool, it was partly low pH stressed to reach an overall HMW level of 10 %. For the preparative runs, an equilibration buffer consisting of 20 mM sodium citrate (pH 6) was used. Elution was performed with 20 mM sodium citrate and 500 mM sodium chloride (pH 6). For the in-line pooling decision, the sodium chloride concentration was increased to 550 mM to simulate a process disturbance. For reference analytics (analytical size exclusion chromatography), a buffer with 200 mM potassium phosphate and 250 mM potassium chloride (pH 7) was used. All buffer components were purchased from VWR. All buffers were prepared with Ultrapure Water. Prior to the experiments, the buffers and the feed were filtrated with a cellulose acetate filter with a pore size of 0.22 µm (Pall). The buffers were also degassed by sonification before usage.

Chromatography Runs

The HiTrap 16 mm × 25 mm SP Sepharose FF column was first equilibrated for 5 CV and then loaded with 200 mg mAb (monomer and HMWs). After a wash of 3 CV with equilibration buffer, variable linear gradients from 0 % to 100 % elution buffer were performed. Gradient length was set to 4 CV, 5 CV, 6 CV, and 7 CV. The results of the runs with gradient lengths 4 CV, 6 CV, and 7 CV were used to calibrate the PLS model, while the

run with a gradient length 5 CV was applied to confirm the model. As for the model protein study, data acquisition of the FlowVPE was started at the beginning of the gradient and slope spectra from 240 nm to 340 nm with 2 nm resolution were recorded. After the linear gradient elution, the column was regenerated for 3 CV with the elution buffer. The flow rate was kept constant at 0.5 mL/ min for all steps and experiments. During the elution and regeneration, 1000 µL fractions were collected in deep well plates.

Reference Analytics

As reference analytics, Size-Exclusion Chromatography (SEC) was performed with the collected fractions to determine the concentration of mAb monomer and HMWs. Samples were injected into a 4.6 mm × 150 mm TSKgel SuperSW mAb HTP column (Tosoh, Tokio, Japan). The flow rate was set to 0.3 mL/ min.

4.2.5 Data Analysis

Protein slope spectra were extracted from the 3-dimensional chromatograms while pure protein eluted according to off-line analytics. Each slope spectrum was normalized by dividing by its average slope.

As the total duration of a measurement cycle of the FlowVPE (including the screening for the linear range) varied slightly, the results of the off-line reference analytics were linearly interpolated such that they matched the slope spectra. The slope spectra were then preprocessed by mean centering and correlated with the results of the off-line analytics using PLS1 regression [22], [69], [70]. PLS1 was used as little correlation was observed between the measured concentrations of the different species [29]. The number of latent variables in the corresponding PLS model was based on the minimization of the RMSE of the model prediction in a cross validation.

For both case studies, the concentrations of the different species were subsequently smoothed over time by a Savitzky-Golay filter in Matlab (MathWorks, Inc., Natick, US) [82]. This filtering approach was chosen as the PLS model does not consider time-wise correlation of the data. The Savitzky-Golay filter allows to smooth data over time. For the calibration and validation runs, the Savitzky-Golay filter was used in a symmetric form smoothing the central data point in a given frame (frame size: 11 points, 2nd order polynomial). The Savitzky-Golay filter was not applied for the in-line pooling decision to avoid a time delay. For calculating the mass balances for purity and yield off-line, the concentrations were again smoothed as described above. Smoothing in spectral dimension was not applied as

information crucial to the PLS may be lost in the smoothing process. Furthermore, PLS models already suppress noise in the spectral dimension by mapping spectral data onto latent variables [29].

4.2.6 In-line Monitoring and Control

The pooling of the products in both case studies was controlled in-line based on the predicted concentrations of the PLS model. A .NET assembly provided by the vendor of the FlowVPE triggered a call-back function in Matlab after each spectral measurement. Matlab subsequently computed the slope spectra and the protein concentrations. The communication between Matlab and Unicorn was implemented analogous to the protocol published earlier [42]. Shortly, Matlab triggers a block in a Unicorn method by sending a digital signal when a predefined pooling criterion is fulfilled.

To demonstrate that the method can handle process disturbances, modified elution buffers with increased salt concentrations (550 mM sodium chloride instead of 500 mM sodium chloride in both case studies) were used. For the second case study, the loading density was additionally decreased to 150 mg as it may occur in production when the remainder of a batch is loaded on a column. In the first case study, lys was declared as product. Lys was automatically collected, as soon as its concentration exceeded 2 g/L and the cyt c concentration fell below 1.8 g/L. Pooling was continued as long as the lys concentration remained above 2 g/L. Pooling in the second case study was triggered when the mAb monomer concentration exceeded 2 g/L. Pooling was stopped as soon as the purity of the pool fell below 95 %. For calculating the mass balances, negative concentrations computed by the PLS model were set to 0. The termination of pooling was only allowed after the eluate was collected for three minutes to prevent a fractionation stop due to noise in the prediction.

4.3 Results and Discussion

As described above, linear gradients with variable lengths were performed in both case studies to obtain different mixing ratios and concentrations of the examined proteins. The acquired slope spectra and the results of the fraction analysis of three runs were used to calibrate a PLS model, i.e. to span a calibration space for the model. The fourth run was subsequently applied to test the PLS model in both studies. Before the results of both studies are discussed in detail, typical chromatograms obtained by the combination of chromatography and VP spectroscopy will be presented.

4.3.1 Application of VP Spectroscopy for Chromatography

In VP spectroscopy, absorbance is measured at different pathlengths. Figure 4.2A illustrates a typical chromatogram obtained by VP spectroscopy at 280 nm. The green lines represent the measured absorbance over time at different pathlengths. The corresponding pathlengths are illustrated by the grey lines. In orange, the slope over time is depicted. As shown by Equation 4.3, the slope is linearly dependent on the total protein concentration. It is worth noting that the absorbance values stay almost constant during protein elution, while the pathlengths decrease. During the screening phase, the longest pathlength for the measurement phase is selected such that the absorbance is expected to be 1 AU at 280 nm. In consequence, the set pathlengths are inversely proportional to the slope and thus to the protein concentration (cf. Figure 4.2A).While Figure 4.2A displays a typical chro-
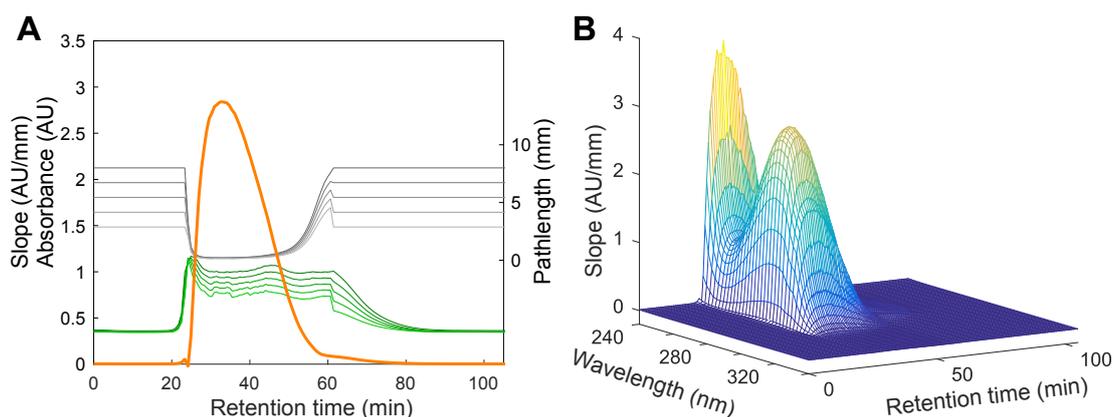
Figure 4.2: Typical chromatograms obtained by VP spectroscopy for one wavelength (A) and multiple wavelengths (B). A: The green lines represent the obtained absorbance values at the different pathlengths (grey lines). At each point in time, the slope (orange line) was determined by a linear regression between five absorbance values and pathlengths. B: If slope spectra are recorded during a chromatography run, a 3-dimensional chromatogram is obtained.

matogram for one wavelength, a chromatogram for multiple wavelengths is presented in Figure 4.2B. Instead of single wavelength slopes, slope spectra were acquired. As the slopes were recorded during a whole chromatography run, this resulted in a 3-dimensional chromatogram in (time, wavelength, slope). The obtained 3-dimensional chromatograms of the chromatogra-

phy experiments were the starting point for the PLS model calibration and confirmation.

4.3.2 Case Study I: Separation of Cyt c from Lys

Cyt c and lys feature significant spectral differences in UV due to the different mass fraction of aromatic amino acids as well as the heme group in cyt c (cf. Figure 4.3A). The two proteins thus allow to test the proposed setup for a simple application. Based on the performed cross validation,
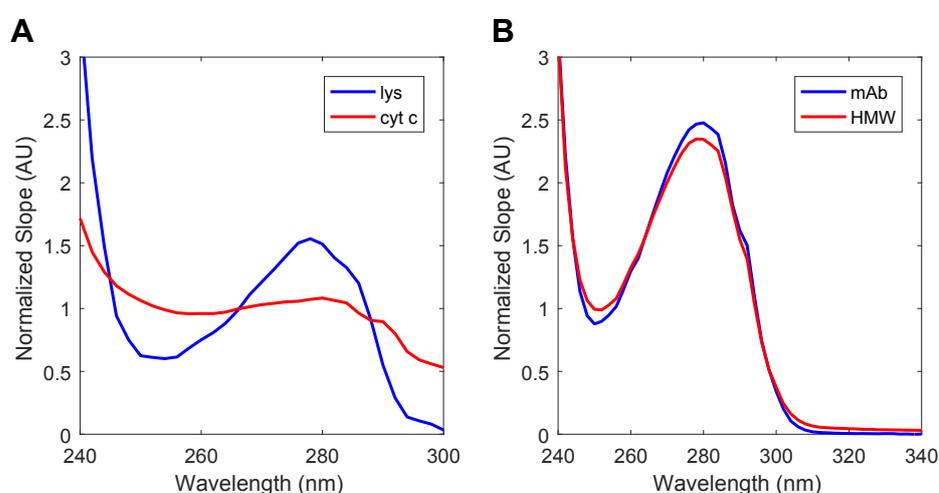

Figure 4.3: Comparison of normalized protein slope spectra for case study I (A) and case study II (B).

PLS1 models with 2, respectively 3 latent variables were selected for lys and cyt c. The resulting PLS model predictions for the three calibration runs (gradient length of 2 CV, 4 CV, and 8 CV) are displayed in Figure 4.4A, B, and D. The plots compare the PLS model prediction for lys (solid blue lines) and cyt c (solid red lines) with the results of the corresponding reference analytics (blue bars for lys and red bars for cyt c). A good agreement was observed between PLS model prediction and reference for all calibration runs. To confirm the model, predictions for a 6 CV gradient run were made, which closely follow the corresponding reference analytics (cf. Figure 4.4C). The different shades of blue and red of the PLS model prediction in Figure 4.4 illustrate the unsmoothed and smoothed data (lighter colors for the unsmoothed and darker colors for the smoothed data). The RMSE of this run was 0.53 g/L for cyt c and 1.11 g/L for lys. The RMSE of the smoothed prediction data with a Savitzky-Golay filter was 0.48 g/L for cyt c and 1.05 g/L for lys. Applying the Savitzky-Golay filter clearly
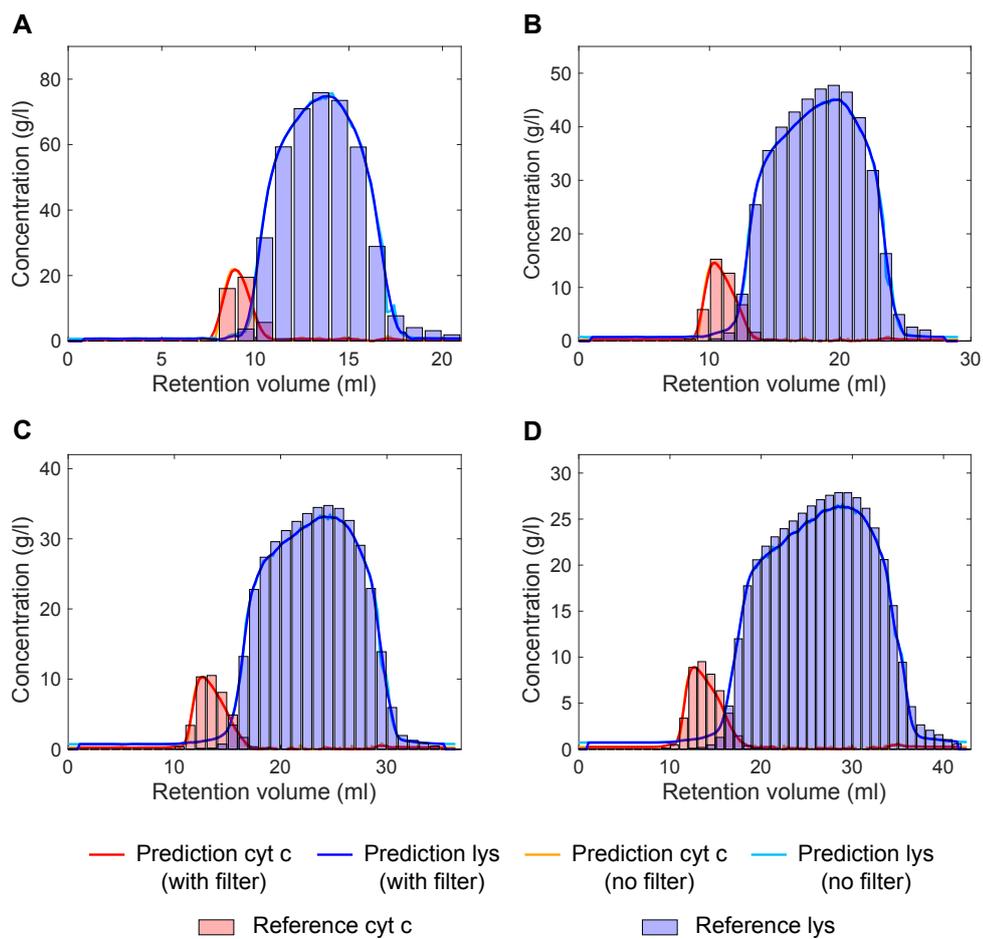
Figure 4.4: Comparison of the PLS model prediction for lys and cyt c with the results of the off-line reference analytics for a gradient length of A: 2 CV, B: 4 CV, C: 6 CV (confirmation run), and D: 8 CV. The different shades of blue and red of the PLS model prediction illustrate the effects of smoothing.

improved the RMSE for both the prediction of lys and cyt c (cf. Table 4.1). The results show that the RMSE of the model predictions is partly due to random noise but also partly systematic. As Savitzky-Golay filtering helps to reduce random noise, the method was selected as a suitable post-run data treatment. The combination of VP spectroscopy and PLS modeling allowed

Table 4.1: RMSE for the prediction in the confirmation runs of the two case studies: The RMSE of unfiltered predictions is compared with the RMSE of the filtered data.

	Case study I		Case study II	
	lys [g/l]	cyt c [g/l]	mAb monomer [g/l]	HMW [g/l]
No filter	1.10	0.53	1.26	0.50
Savitzky-Golay filter	1.04	0.48	1.07	0.42

for a selective quantification of lys and cyt c during chromatography runs with highly loaded columns and lys peak concentrations of up to 80 g/L. This shows that the method is applicable for the typical concentration range of preparative chromatography.

4.3.3 Case Study II: Separation of HMWs from mAb Monomer

In contrast to cyt c and lys, mAb monomer and HMW presumably contain the same mass fraction of aromatic amino acids and disulfide bridges. Spectral differences are therefore either related to changes in tertiary structure or due to light scattering [32]. As a result, the differences in the spectra of mAb monomer and HMW are comparably small (cf. Figure 4.3 B). This makes the quantification by PLS modeling more challenging.

Based on the performed cross validation, a PLS1 model with 4 latent variables was selected for the mAb monomer, while a PLS1 model with 8 latent variables was used for the HMWs. Figure 4.5A, C, and D display the PLS model prediction for the three calibration runs (gradient length of 4 CV, 6 CV, and 7 CV), while Figure 4.5B displays the results of the confirmation run (5 CV gradient). The figures show a comparison between the PLS model prediction for the mAb monomer (solid blue lines) and the HMWs (solid red lines) with the results of the reference analytics (blue bars for mAb monomer and red bars for HMWs). In all four runs, the model prediction matched the reference analytics closely. The results thus demonstrate the applicability of the method for proteins with only slight

differences in the absorption spectra. The RMSE of prediction was 1.26 g/L for the mAb monomer and 0.50 g/L for HMWs. The RMSE could be further reduced by smoothing the data with a Savitzky-Golay filter. This led to RMSE values of 1.07 g/L for mAb monomer and 0.42 g/L for HMWs. Figure 4.4 compares the unsmoothed and the smoothed data for all runs. Applying the Savitzky-Golay filter improved the RMSE for both the prediction of the mAb monomer and HMWs indicating the presence of random noise (cf. Table 4.1).
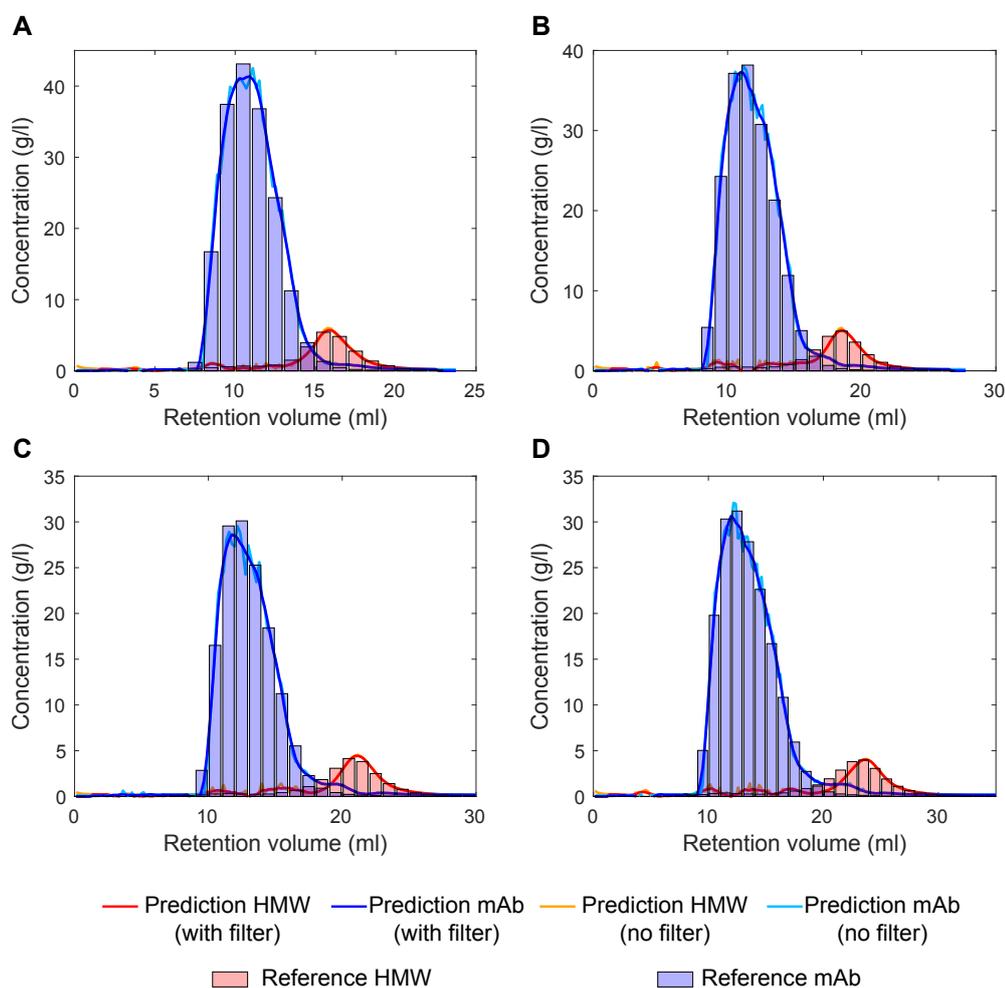

Figure 4.5: Comparison of the PLS model prediction for mAb monomer and HMWs with the results of the off-line reference analytics for a gradient length of A: 4 CV, B: 5 CV (confirmation run), C: 6 CV, and D: 7 CV. The different shades of blue and red of the PLS model prediction illustrate the effects of smoothing.

4.3.4 In-line Pooling Decisions

To show the usefulness of VP spectroscopy in conjunction with PLS modeling, the implemented methods were used for process decisions. For both case studies, the pooling decision for chromatographic runs with process disturbances, i.e. away from the set point, were performed. For the first case study, lys was purified from cyt c in a 6 CV gradient with an increased salt concentration in the elution buffer (550 mM sodium chloride). The start of the pooling was triggered by the PLS model as soon as the concentration of lys exceeded 2 g/L and the cyt c concentration fell bellow 1.8 g/L (cf. Figure 4.6A). For the second case study, a 5 CV gradient was performed also with an increased salt concentration in the elution buffer (550 mM sodium chloride). Furthermore, the loading density of the column was decreased from 40 g/L to 30 g/L. Such a situation may occur in production processes, when the pool from the previous process step cannot be transferred to an integer number of column cycles. Thus, a number of cycles are performed with fully loaded columns. The remainder is then loaded onto a column resulting in a lower loading density, which might change the separation and pool purity. This problem can be solved by applying the described in-line control. The method was applied to trigger pooling as soon as the mAb monomer concentration exceeded 2 g/L. Pooling was stopped, when the purity fell bellow 95 % (cf. Figure 4.6B). The collected pools were subsequently analyzed by off-line analytics. In both case studies, accurate predictions could be made with minor deviations. For the separation of cyt c and lys, a purity of 99.0 % was predicted while off-line analytics measured a 99.7 % purity. For the mAb monomer product pool, the PLS model predicted a purity of 94.4 % compared to 94.2 % measured by off-line analytics. Here, Savitzky-Golay filtering did not lead to an improved prediction because the measurement noise cancelled itself out over time (cf. Table 4.2). The re-

Table 4.2: Results of the in-line pooling decisions: For both case studies, the pool purity was calculated based on the PLS model prediction and compared with the corresponding results from off-line analytics as reference.

	Case study I Purity [%]	Case study II Purity [%]
Reference	99.7	94.2
Prediction (no filter)	99.0	94.4
Prediction (with filter)	99.0	94.4

sults show that the proposed method may be applied for in-line control of
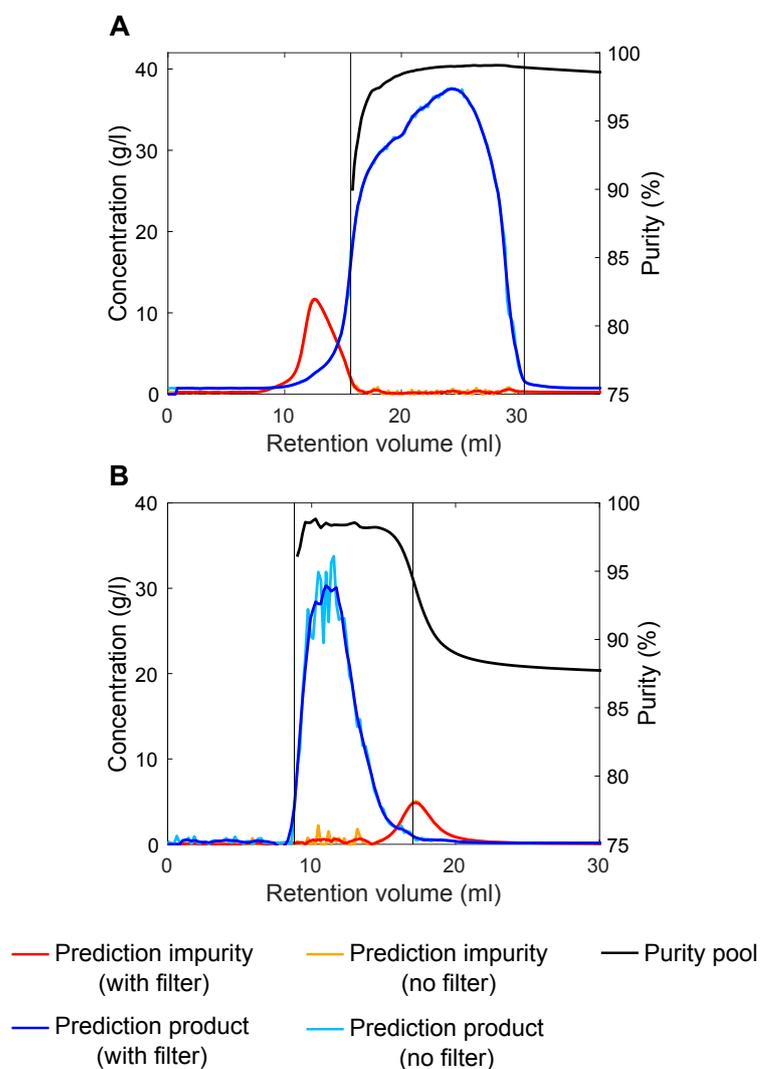
Figure 4.6: Prediction of the PLS model during the in-line control run for case study I (A) and case study II (B). The different shades of blue and red of the PLS model prediction illustrate the effect of smoothing. Based on the unsmoothed data, the pool purity was calculated (black). The thin black vertical lines indicate the start and end of pooling.

a chromatography system. The used residence time of 10 min is above the industrial standard. However, a reduction of the flow rate during elution may be justifiable as the elution phase is relatively short compared to the complete process. It is worth noting that the discussed approach is significantly faster than on-line HPLC PAT methods proposed in literature. In production scale, on-line HPLC was applied with 2.5 to 3 times slower response time (1.3 min resp. 96 s) [16], [75]. Thus, the applicability of VP spectroscopy should be further evaluated on a case-to-case basis.

4.4 Conclusion and Outlook

In-line monitoring of preparative chromatography was successfully realized in this study. It was demonstrated, that the combination of VP UV/Vis spectroscopy and PLS modeling allows for a selective in-line protein quantification in a broad dynamic range of concentrations. The method enabled the monitoring of chromatography runs with highly loaded columns. Product peak concentrations varied between 30 g/L to 80 g/L, while contaminant peak concentrations were only 4 g/L to 20 g/L. Consequently, the proposed method has potential for the in-line monitoring and control of preparative chromatography. It might also be applicable for controlling switching times in continuous chromatography. Future challenges are especially related to the scale up and robustness of the method as well as to the optimization of the measurement time of VP spectroscopy. As each VP measurement relies on accurate mechanical positioning of the optical fiber, an increase in measurement speed is challenging. Additionally, further investigations have to be made to improve understanding and sensitivity for HMW detection.

Acknowledgment

<div style="text-align: right; font-size: 3em;">5</div>

Monitoring of Antibody-drug Conjugation Reactions with UV/Vis Spectroscopy

Sebastian Andris[*,1], Matthias Rüdt[*,1], Jonas Rogalla[1], Michaela Wendeler[2], Jürgen Hubbuch[1]

[*]   Contributed equally
[1]   Institute of Engineering in Life Sciences, Section IV: Biomolecular Separation Engineering, Karlsruhe Institute of Technology (KIT), Germany
[2]   Department of Purification Process Sciences, MedImmune LLC, United States

Abstract

The conjugation reaction of mAbs with small molecule drugs is a central step during production of ADCs. The ability to monitor this step in real-time can be advantageous for process understanding and control. Here, we propose a method based on UV/Vis spectroscopy in conjunction with PLS regression for non-invasive monitoring of conjugation reactions. In experiments, the method was applied to conjugation reactions with two surrogate drugs in microplate format as well as at $20\,\mathrm{mL}$ scale. All calibrated PLS models performed well in cross-validation ($Q^2 > 0.975$ for all models). In

microplate format, the PLS models were furthermore successfully validated with an independent prediction set ($R^2_{\text{pred}} = 0.977$ resp. 0.894). In summary, the proposed method provides a quick and easily implementable tool for reaction monitoring of ADC conjugation reactions and may in the future support the implementation of PAT.

5.1 Introduction

ADCs are among the most promising new formats in the biopharmaceutical industry [83]. More than 60 candidates are currently evaluated in clinical trials. ADCs gain their potential from combining the high selectivity of mAbs with the high cytotoxicity of small-molecule drugs. Next to specificity and cytotoxicity, ADCs also inherit other attributes of both species, such as the absorption bands of both protein and drug and an often increased hydrophobicity compared to unmodified mAbs due to the apolar character of the drugs [84]–[86].

One of the most central steps during ADC production is the conjugation reaction which links the drug to the mAb via a linker. The conjugation reaction may either be site-specific or unspecific, with the currently developed ADCs mainly focusing on site-specific conjugation reactions with well-defined Drug-to-Antibody Ratios (DARs) [87]–[89]. The conjugation yield and DAR are generally measured off-line by analytical Hydrophobic-Interaction Chromatography (HIC) or Reversed-Phase (RP) chromatography, often in combination with mass spectrometry [90]. This is, however, time-demanding and needs manual sample handling. Due to the toxicity of the drugs, analytical chromatography often is performed in a chemically secured environment. If only the DAR is needed without the concentration of each conjugate species, a simple method relying on UV/Vis absorption measurements can be applied [90]. It requires the drug to have an absorption band different from the one of the protein (280 nm). Using the absorption at both maxima and the respective extinction coefficients, the concentrations of protein and drug can be mathematically determined without further analytical methods. The technique has been used for conjugations with different drugs like the maytansinoid DM1 and dipeptide-linked auristatins (e.g. vcMMAE) [91], [92], but is limited to purified conjugates, as residual free drug and other possibly UV-active contaminants have to be removed. As a consequence, this approach as well as analytical chromatography are not very well suited for fast and prompt characterization of ADC conjugation reactions. Therefore, only complex analytical solutions are found so far for the monitoring of these reactions. SEC with a post-column re-

action was proposed for DAR determination of cysteine-conjugated ADCs [93]. Tang et al. present an approach for rapid DAR measurement by fast deglycosylation and LC-MS detection [94].

It would be highly beneficial to establish a fast analytical method for monitoring the progress of conjugation reactions without any sample processing. Ideally, such a method would also provide the means for application as a PAT in accordance with the PAT initiative by the FDA. For this, the applied method needs to be fast, without manual sample handling, and robust [18]. To monitor the process, the method should be sensitive to the progress of protein conjugation reactions. UV/Vis absorption spectroscopy is a rapid, noninvasive, and quantitative method which is widely established in biopharmaceutical manufacturing. It has previously been applied to process monitoring of proteins and small molecules [2], [18], [39], [74]. Hansen et al. showed the potential of UV/Vis spectroscopy to distinguish between different proteins based on their content of aromatic amino acids and their solvatization [39]. The method was later transferred to chromatographic separations by Brestrich et al. [42]. There are some examples of UV/Vis spectroscopy in reaction monitoring applications. Quinn et al. followed a small-molecule reaction in lab scale using fiber-optic UV/Vis spectroscopy [95]. Gurden et al. employed a model based on UV/Vis absorption data to detect and diagnose process variations in a non-protein biochemical conversion reaction [96].

Drugs used in ADCs frequently feature delocalized electron systems thus providing absorption bands in the UV/Vis range [90] besides the ones of the aromatic amino acids of the mAbs. Spectral shifts of UV/Vis absorption may not only be caused by a structural change in the UV/Vis active compounds, they can also occur as a result of changes in the local environment of the chromophores [97], [98], e.g. a change in solvent composition. If the conjugation reaction thus causes a change in the environment of the aromatic amino acids or the drug, it will cause spectral shifts which in turn may help to monitor this type of reaction.

This work investigates a new and easily applicable method for on-line conjugation reaction monitoring. Monitoring was accomplished by a combination of UV/Vis spectroscopy and PLS modeling. Spectra were recorded and analyzed during conjugation reactions in two different scales with different UV/Vis detectors. Based on the results, a method was established for small scale screening in 96-well plates which provides an estimate of the amount of drug conjugated to the antibody by PLS regression. Two different surrogate drugs, 7-Diethylamino-3-(4'-Maleimidylphenyl)-4-Methylcoumarin (CPM) and N-(1-Pyrenyl)maleimide (NPM), were applied in both setups. Additional variability was introduced by changing the con-

centrations of the reactants. The method was then adapted to a lab-scale conjugation reaction with an on-line DAD to show applicability as a PAT tool.

5.2 Materials and Methods

5.2.1 Chemicals

For disulfide reduction, we used Tris(2-Carboxyethyl)phosphine Hydrochloride (TCEP) (Sigma-Aldrich, catalog number C4706). (L)-Dehydroascorbic Acid (DHA) (Sigma Aldrich, catalog number 261556) was used as oxidation agent for re-oxidation of interchain disulfides. As non-toxic substitutes for cytotoxic drugs, NPM and CPM (both Sigma-Aldrich, catalog number C1484 and P7908) were selected. Their structural formulae is are shown in Figure 5.1. Dimethyl Sulfoxide (DMSO) (Sigma Aldrich) was used to dissolve DHA, CPM, and NPM. N-Acetyl Cysteine (NAC) (Sigma Aldrich, catalog number A7250) was applied to quench residual free drug. For buffer preparation, $NaH_2PO_4 \cdot 2\,H_2O$ was obtained from VWR International GmbH. The buffers were titrated to the desired pH with $4\,M$ NaOH (Merck KGaA) and filtered through a $0.2\,\mu m$ cellulose acetate membrane filter (Sartorius AG, Göttingen, Germany).
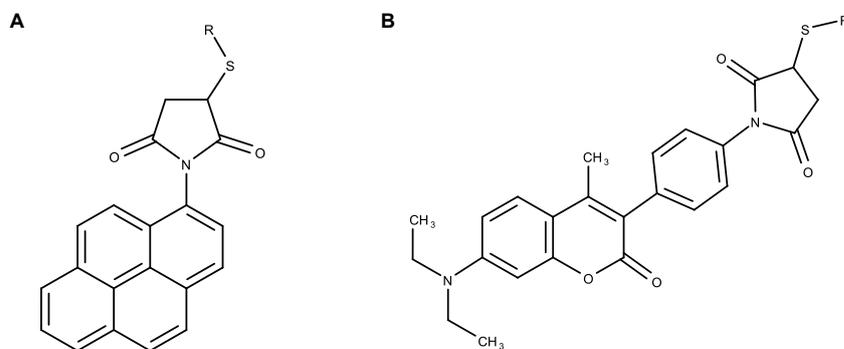

Figure 5.1: The structure of conjugated NPM (Figure 5.1A) and conjugated CPM (Figure 5.1B) are shown. R denotes the protein.

5.2.2 Model System and Conjugation Process

Purified Immunoglobulin G (IgG)1 mAb with two engineered cysteines as conjugation sites was provided at a concentration of $12.4\,mg/mL$ in phosphate buffered solution with $5\,mM$ EDTA, pH 7.2 by MedImmune, LLC.

CPM and NPM were used as non-toxic maleimide-functionalized surrogate drugs and conjugated to the antibody's two engineered cysteines via their maleimide linker. For the conjugation experiments, aliquots of the engineered mAb stock solution were thawed and diluted to the desired concentration (2 mg/mL) with 50 mM sodium phosphate buffer (pH 7.2). The resulting mAb concentrations were determined with a NanoDrop 2000c spectrometer (ThermoFisher Scientific, Waltham, USA).

The following mAb preparation steps (reduction and re-oxidation) were conducted in 50 mL centrifugation tubes (VWR International GmbH). A reduction step was performed to uncap engineered cysteine residues. For this purpose, a 40-fold molar excess of TCEP (over mAb concentration) was added to the mAb solution. After 3 h of incubation at room temperature, the reduced mAb solution was transferred into a dialysis cassette with a 10 kDa molecular weight cut-off to remove TCEP. The dialysis was performed in a volume of 1.7 L of 50 mM sodium phosphate buffer at 5 °C over night (approx. 19 h). The mAb concentration after dialysis was determined with the NanoDrop spectrometer.

For re-oxidation of the interchain disulfide bonds, 20-fold molar excess of DHA (3 mM stock solution in DMSO) was added. The mixture was incubated at room temperature for 4 h. Through addition of the DHA solution, DMSO content was increased to around 8.5 %. To remove potential precipitate before spectroscopic conjugation monitoring, the mAb solution was filtered through a 0.2 µm polyethersulfone syringe filter (VWR International GmbH). The final mAb concentration for the conjugation experiments was set via dilution with 50 mM sodium phosphate buffer containing 10 % of DMSO. Conjugation experiments were executed with mAb concentrations in the range of 1.0 mg/mL to 2.0 mg/mL.

The conjugation reaction in the different experimental setups was initialized by addition of the surrogate drug (NPM or CPM) to the re-oxidized mAb solution. The molar ratio (drug to mAb) was set to 2 for the NPM conjugations and to 3 for CPM. The concentration of the surrogate drug stock solutions was varied (from 2 mM to 6 mM) depending on the mAb concentration to result in a final DMSO content of approximately 10 %. This content of DMSO was maintained to ensure solubility of the hydrophobic surrogate drugs in the water-based solution. The conjugation reaction was quenched by addition of a 12-fold molar excess of NAC (over the applied amount of surrogate drug) to ensure the immediate termination of the conjugation reaction.

5.2.3 High-throughput On-line Monitoring Experiments in Microplate Format

The high-throughput conjugation experiments were conducted in 96-well UV-transparent microplates (UV-STAR, Greiner bio-one GmbH, Frickenhausen, Germany). The reaction was monitored by the acquisition of UV/Vis absorption spectra of the reaction mixture in the range from 250 nm to 450 nm with an Infinite M200 microplate spectrometer (Tecan Group Ltd., Männedorf, Switzerland). To allow for the correlation of UV/Vis absorption data with the progress of the conjugation reaction, spectra had to be recorded while different time points of the reaction were sampled. The used experimental setup is depicted in Figure 5.2. The UV-microplate was divided into *monitoring wells* designated for UV/Vis absorption measurements and *quenching wells* designated for off-line analytics. In the latter, the reaction was quenched at different time points to generate samples for off-line analysis. The six monitoring wells contained 200 µL of liquid and were further separated into two blank wells and four reaction wells. One blank well contained buffer solution, the other one re-oxidized mAb. The remaining monitoring wells were used for reaction monitoring in duplicates at two different conditions. There were 16 quenching wells for each of the two screened conditions, containing 100 µL of the corresponding reaction mixture. In this study, the mAb concentration was varied in the range of 1.0 mg/mL to 2.0 mg/mL while all other process conditions were kept constant for all experiments. This resulted in 6 calibration and 2 prediction runs for NPM and 5 calibration and 2 prediction runs for CPM. The conjugation reaction was started by adding the surrogate drug to the re-oxidized mAb solution in a 50 mL centrifugation tube. After short mixing, aliquots were transferred immediately to the microplate. The reaction in the first quenching well was instantly stopped by addition of NAC solution before placing the microplate into the reader and starting the on-line monitoring procedure. The UV/Vis spectra acquisition was controlled by the software Magellan (Tecan Group Ltd.) according to the following process: Prior to each measurement, the plate was shaken for 15 s (orbital shaking, 1.5 mm amplitude, 335.8 rpm). For the first 22 min or 25 min, single spectra were recorded and after each measurement one well was quenched. At later time points more spectra were acquired between each quenching step, resulting in quenching time intervals of 4 min to up to 10 min. The spectral range for the conjugation reaction with NPM was defined at 250 nm to 390 nm and for CPM at 250 nm to 450 nm (step size 4 nm, 5 reads) to cover the characteristic absorption maxima of the surrogate drugs. The conjugation reaction was monitored over a run time of 50 min. Afterwards, the microplate was
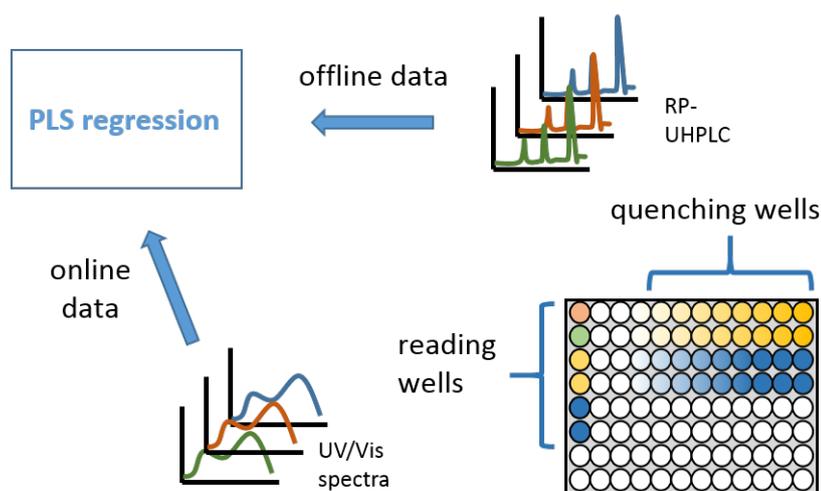
Figure 5.2: Experimental setup used for high-throughput on-line monitoring in microplate format. UV/Vis spectra were recorded during the conjugation reaction using the integrated Tecan plate reader. Reactions in the quenching wells were stopped at different time points and analyzed by RP-UHPLC. On-line and off-line data was used for the generation of PLS regression models.

centrifuged (1789 g, 7 °C) to remove potential precipitate prior to off-line analysis. The supernatants were measured by RP Ultra High Performance Liquid Chromatography (UHPLC).

5.2.4 Lab-scale On-line Monitoring Experiments

Preparation of the mAb was conducted as described above in the conjugation process section. The re-oxidized mAb solution at a concentration of 2 mg/mL was used for the experiments. Here, the acquisition of UV/ Vis spectra was performed with an Ultimate 3000 DAD (Dionex Softron GmbH, Germering, Germany) with a semi-preparative flow cell (volume 0.7 µL, 0.4 mm optical path length) at a spectral resolution of 1 nm. The experimental setup consisted of a 50 mL beaker glass as reaction vessel, a peristaltic pump (Minipuls 3, Gilson, Villiers de Bel, France) with marprene pump tubing, and the DAD. All elements were connected via Polyether Ether Ketone (PEEK) tubing (0.5 mm diameter). By attaching the beaker glass to a thermal shaker (HLC BioTech, Bovenden, Germany), the solution was continuously mixed at 200 rpm and the temperature was kept constant around 23 °C. The reaction mixture was circulated from the reservoir via the peristaltic pump through the DAD flow cell and back into the reservoir.

In- and outlet were placed below the liquid surface. The flow rate was approximately 3 mL/ min which equaled the maximum speed of the peristaltic pump (48 rpm).

Prior to monitoring the reaction, the DAD was equilibrated with sodium phosphate buffer for 2 h and with re-oxidized mAb solution for 15 min. Autozero of the DAD signal was performed either with re-oxidized mAb (NPM experiments) or with sodium phosphate buffer (CPM experiments). After DAD "warm-up", the reactions were started by addition of the surrogate drugs in the molar ratio of 2 for NPM and 3 for CPM. Three runs were performed for each surrogate drug.

The conjugation reactions were monitored over 30 min while UV/Vis spectra were acquired by the DAD every 0.2 s. To reduce noise, the spectra were then averaged over 10 s. The recorded spectral range was 250 nm to 390 nm for NPM experiments and 250 nm to 450 nm for CPM experiments.

Over the runtime of 30 min, 21 samples were taken and transferred to a 96-well microplate for off-line analysis. The wells were previously loaded with NAC stock solution to facilitate immediate quenching of the reaction upon sampling. After termination of the experiment, the microplate was centrifuged (1789 g, 7 °C). The supernatant was analyzed by RP-UHPLC.

5.2.5   RP Chromatography

To assess conjugation results, RP-UHPLC was applied. An Ultimate 3000 system was used, equipped with pump unit, RS autosampler, RS column compartment and diode array detector (Dionex Softron GmbH). Reduction or different sample preparation was not required. An Acquity UPLC Protein BEH C4 column (Waters Corporation, Milford, USA; 300 Å, 1.7 μm, 2.1 mm×50 mm) was run at a flow rate of 0.45 mL/ min. The column oven was heated to 80 °C. Solvent A consisted of 0.1 % Trifluoracetic Acid (TFA) in ultrapure water, solvent B was 0.1 % TFA in acetonitrile. After equilibration and injection at 26 % B, content of B was raised to 30 %. Next, a 4.8 min gradient from 30 % B to 38 % B was used for separation of the conjugate species. Including strip at 95 % B and re-equilibratio,n the runtime was 7 min. UV signals at 280 nm and at the corresponding absorption maximum of the used surrogate drug were recorded (384 nm for CPM and 338 nm for NPM). The resulting chromatograms yielded peak areas of unconjugated, mono-conjugated and di-conjugated mAb, as well as of the remaining free drug. Using the areas at 280 nm and 384 nm or 338 nm, concentrations of the different conjugate species could be calculated with a previously determined calibration curve for the mAb peak area. From these concentrations,

the amount of conjugated drug was calculated to be used as response for PLS modeling.

5.2.6 Data Analysis

All data analysis was performed in Matlab R2016a (The MathWorks). For lab-scale experiments, the spectral band shifts were additionally analyzed by interpolation similar to methods proposed in the literature [32]. First, the spectra were smoothed with a 5th order Savitzky-Golay filter with a 9 point window. Subsequently, the $1\,$nm resolved spectral data was interpolated with a cubic spline to a final resolution of $0.01\,$nm. The wavelength of the maximal absorbance λ_{max} was obtained from the interpolated data.

In the case of microplate experiments, the experiments were split into calibration runs (at mAb concentrations of $1.0\,$mg/mL, $1.5\,$mg/mL, and $2.0\,$mg/mL; NPM 86 samples, CPM 75 samples) and prediction runs (at mAb concentrations of $1.28\,$mg/mL and $1.7\,$mg/mL; NPM 28 and CPM 30 samples). The prediction runs were excluded from model calibration and only used for evaluating the model prediction and calculating Root Mean Square Error of Prediction (RMSEP) values. No prediction runs were performed in case of the lab-scale experiments.

For model calibration, the spectroscopic data was first preprocessed and subsequently fitted with a PLS-1 model by the SIMPLS algorithm [99]. Parameters for preprocessing and model fitting were selected based on an optimization. Preprocessing consisted of multiple steps. First, a baseline was subtracted from each spectrum to reduce possible effects of baseline drifts. For NPM and CPM $390\,$nm, respectively $450\,$nm, were selected as reference wavelength. Subsequently, a Savitzky-Golay filter with a second-order polynomial was applied to the spectra, and, optionally, the first or second derivative was taken [82]. Finally, and only for the lab scale experiments, the spectra were normalized by a 1-norm to further decrease instrumental drifts.

For all models, cross-validation was performed by successively excluding each batch, calibrating a PLS model based on the remaining runs and calculating a residual sum of squares for the excluded batch. All residual sums of squares of the different submodels were summed yielding the Predicted Residual Sum of Squares (PRESS). The PRESS was scaled according to Wold et al. by the number of samples and latent variables used in the PLS model [27]. Based on the scaled PRESS, an optimization was performed using the built-in genetic algorithm of Matlab for integers [100]. The genetic algorithm optimized the window width of the Savitzky-Golay filter, the order of derivative, as well as the number of latent variables for the PLS-1

model. The Root Mean Square Error of Cross Validation (RMSECV) was calculated from the PRESS by dividing through the total number of samples. The Q^2 values were calculated by dividing the PRESS through the summed squares of the response corrected to the mean [27]. The coefficient of determination for the prediction R^2_{pred} was calculated in the same way for the prediction set.

For lab-scale experiments with CPM, the spectral band shift at 394 nm was additionally analyzed by interpolation similar to methods proposed in literature [32]. First, the spectra were smoothed with a 5th order Savitzky-Golay filter with a 9 point window. Subsequently, the 1 nm resolved spectral data was interpolated with a cubic spline to a final resolution of 0.01 nm. The wavelength of the maximal absorbance $\lambda_{\max}$ was obtained from the interpolated data.

5.3 Results

5.3.1 Analysis of UV/Vis Absorption Spectra

In Figure 5.3, the measured spectra of two of the six lab-scale calibration runs are shown (spectra of all experiments, both microplate and lab scale, are shown in the supplementary data; see Figure 5.7). The different spectra are colored according to the reaction progress (blue to red). The autozero for NPM was performed while already flushing the DAD with mAb and, thus, the protein band does not show in the spectra. For comparison, pure component spectra of mAb, NPM, and CPM are supplied in the supplementary material (Figures 5.9, 5.10, and 5.11). In both experiments, a baseline drift is visible at all wavelengths.

NPM features a structured absorption band between 300 nm and 360 nm; CPM a broad band between 330 nm and 450 nm. During NPM conjugation reaction (Figure 5.3), a small bathochromic (red) shift (up to 2 nm) of all NPM bands upon conjugation can be observed. Looking at the bottom graph in Figure 3, a bathochromic shift is also observed for CPM. The maximum around 390 nm is shifted by more than 2 nm. On the right side of Figure 3, the location of the band maxima over time is compared to conjugated drug concentrations from off-line analytics. The two curves show a high degree of correlation for both NPM and CPM (Pearson correlation coefficient > 0.97). This is also true for the remaining lab-scale runs, except for the CPM run 1 which reached a correlation coefficient of 0.92.
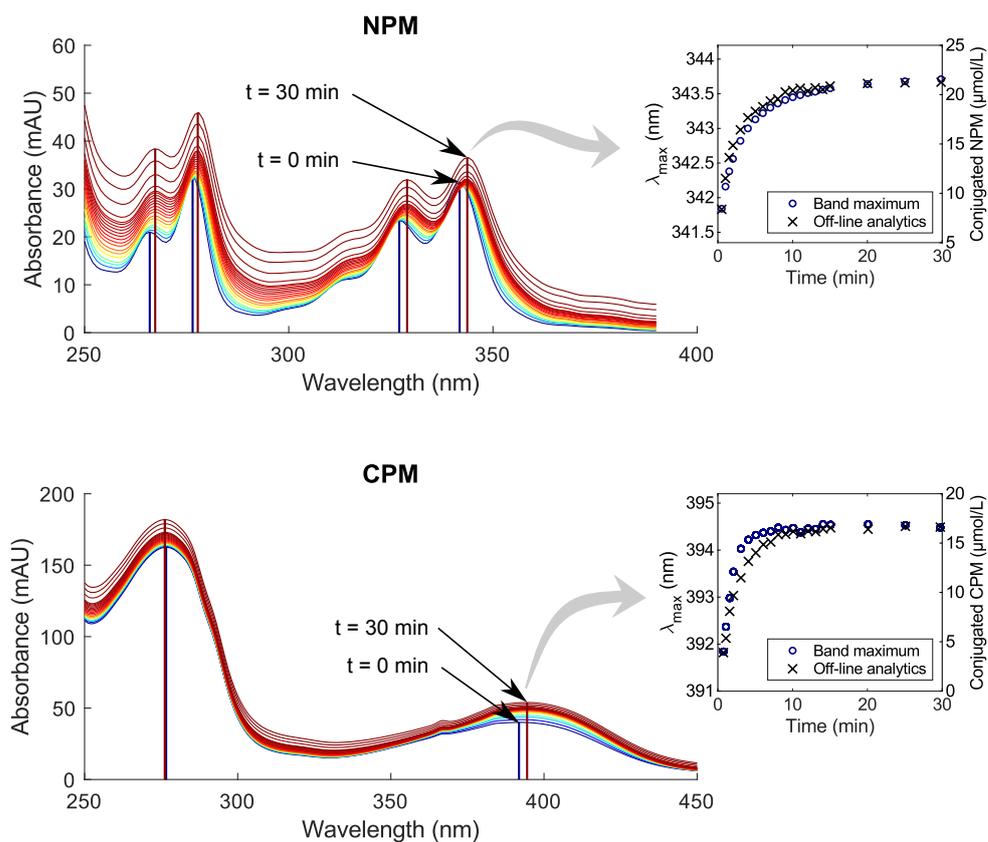
**Figure 5.3:** The raw spectra of two lab-scale experiments for NPM (top) and CPM (bottom) are shown. The spectra are colored according to the reaction progress from blue to red. The location of the band maxima of the first (0 min) and the last spectrum (30 min) are marked by vertical lines. On the right side, the time evolution of the band maxima location is compared to the amount of conjugated drug measured by off-line analytics.

5.3.2 PLS Model Calibration and Validation for Microplate Experiments

For the microplate experiments, the data was split in a calibration set and an independent prediction set. Multiple parameters were set during model calibration (Savitzky-Golay window width, derivative, number of latent variables). As a systematic approach, a numerical optimization was chosen with the scaled PRESS from cross-validation as an objective. Figure 5.4 shows the calibrated model for the NPM and CPM experiments. For all experiments, the measured concentration of conjugated drug first increases quickly and approximates a limit after 10 min to 20 min. For all calibration experiments, the PLS prediction follows the concentrations from off-line analytics. Tabel 5.1 summarizes the optimized parameters. For NPM and CPM, RMSECV values of 0.60 µmol/L ($Q^2 = 0.9856$) and 0.56 µmol/L ($Q^2 = 0.9875$), respectively, were reached. The calibrated PLS
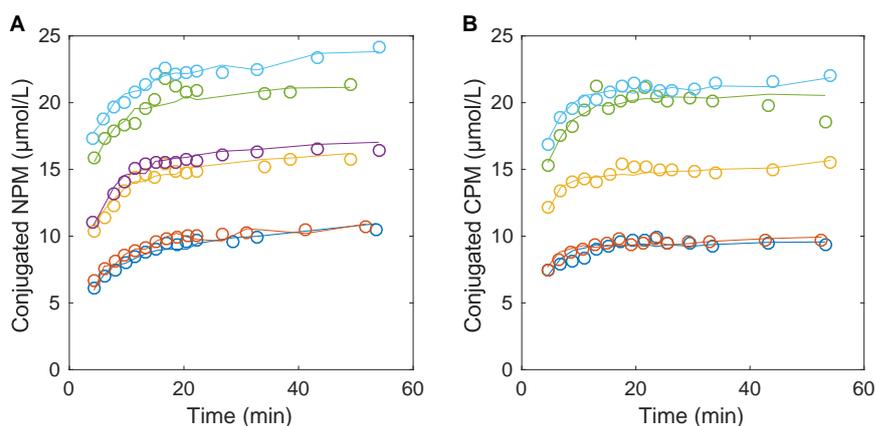
Figure 5.4: PLS model calibration for the microplate experiments is shown for NPM (Figure 5.4A) and CPM (Figure 5.4B). The mAb concentration of the different experiments are: red/blue 1 mg/mL, yellow/violet 1.5 mg/mL and green/cyan 2 mg/mL.

models were then validated by applying them to a prediction set (Figure 5.5). The shape of the conjugated drug concentration is similar to the calibration set and captured by the PLS prediction in all experiments. In case of CPM, the PLS prediction is offset for both experiments to higher concentrations. RMSEPs of 0.57 µmol/L ($R^2_{\text{pred}} = 0.9770$) and 0.90 µmol/L ($R^2_{\text{pred}} = 0.8940$) were reached for NPM and CPM, respectively.

Table 5.1: Summary of optimized parameters for the spectral preprocessing and PLS model as well as the performance of each model in cross-validation and on independent prediction sets.

|	Microplate		Lab-scale	
	NPM	CPM	NPM	CPM
Number of latent variables	6	5	4	2
Window for Savitzky-Golay	17	13	35	71
Derivative	1	0	1	1
Q^2	0.9856	0.9875	0.9792	0.9755
RMSECV (µmol/L)	0.60	0.56	0.56	0.57
R^2_{pred}	0.9770	0.8940	-	-
RMSEP (µmol/L)	0.57	0.90	-	-
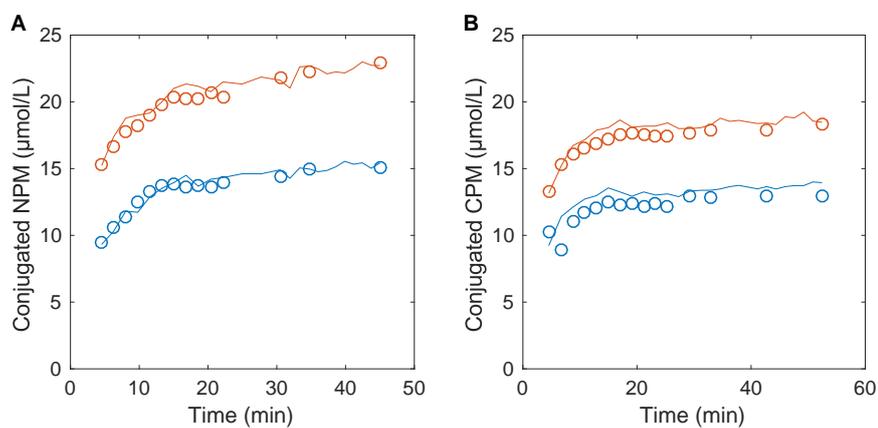
Figure 5.5: PLS model prediction for the microplate experiments is shown for NPM (Figure 5.5A) and CPM (Figure 5.5B). The mAb concentration of the different experiments are: blue $1.3\,\mathrm{mg/mL}$ and red $1.7\,\mathrm{mg/mL}$.

5.3.3 PLS Model Calibration for Experiments at Lab-Scale

PLS model calibration for lab-scale experiments was optimized in the same way as the calibration for experiments in microplates (Table 5.1). Due to material limitations, no experiments were designated for a prediction set. Instead, the PLS models were assessed only by cross-validation. For NPM, an RMSECV of 0.56 µmol/L ($Q^2 = 0.9792$) was reached. For CPM, the RMSECV was 0.57 µmol/L ($Q^2 = 0.9755$). For ADCs, the degree of conjugation is commonly expressed as DAR. By normalizing the conjugated drug concentration by the initial mAb concentration, the DAR was derived and used for plotting (Figure 5.6).

5.4 Discussion

To correlate the progress of conjugation reactions with changes in the UV/Vis absorption spectra, reactions were performed in microplate format as well as in a lab-scale setup while measuring absorption spectra. First, the spectra were interpreted qualitatively to justify the assumption that the conjugation reaction affects the absorption spectra of the protein/drug mixture. Subsequently, the obtained datasets were used to calibrate four PLS models predicting the concentration of conjugated drug for CPM or NPM in the two different setups.

During the conjugation reaction, UV/Vis absorption spectra are expected to change for multiple reasons. While reacting, the drug moves from an aqueous to the proteinaceous environment. Due to solvatochroism, the absorption bands of the drug thus may shift.21 Second, the proximity of the drug to aromatic amino acids can change the local hydrophobicity which in turn impacts the absorption spectra of aromatic amino acids [32], [101], [102]. Finally, maleimide has been reported to generate a relatively weak absorption band around 273 nm [103]. During the conjugation reaction, the double bond in maleimide is reduced and the band at 273 nm is expected to disappear. For the used surrogate drugs (NPM and CPM, cf. Figure 5.1), the maleimide linker is coupled to the chromophores of pyrene and phenylcoumarin. Thus, they may not have the same absorption bands as free maleimide, and the conjugation reaction may also influence the chromophore intramolecularly.

Based on the spectral changes clearly correlated to the reaction progress observed in Figure 5.3, it was concluded, that the conjugation reactions of both NPM and CPM indeed affect the respective absorption spectra. For
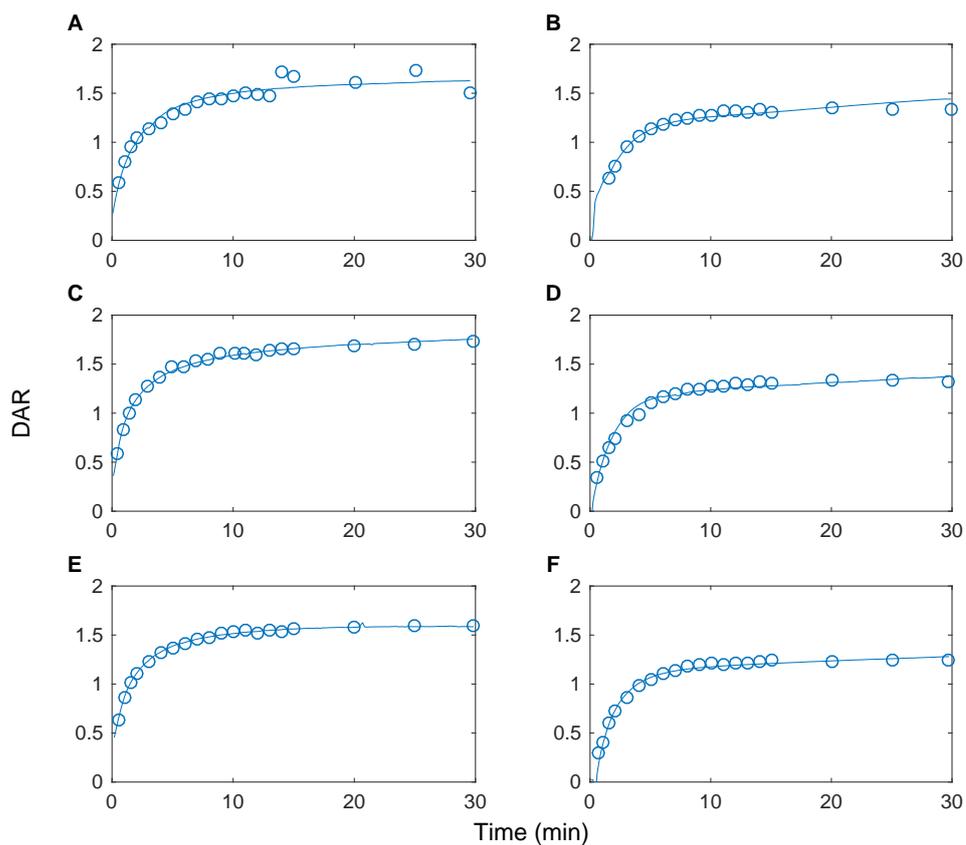
**Figure 5.6:** PLS model calibration for the lab-scale experiments is shown for NPM (left column) and CPM (right column). Each model was calibrated on 3 replicates shown in the different subplots. We decided to use the DAR for plotting as it is frequently used to specify the conjugation degree of ADCs. For calculating the DAR, a constant protein concentration was assumed over the course of the reaction.

further verification, experiments with previously quenched NPM and CPM were conducted, and spectra were recorded over 15 min. Here, no spectral shift was detected, since no reaction was taking place (Figure 5.8). As a consequence of the spectral change caused by the conjugation reaction, predicting the reaction progress from the spectra should be possible. Further data analysis focused on establishing quantitative PLS models for each setup and drug.

For each model, parameters for Savitzky-Golay smoothing and derivative as well as the number of latent variables were optimized. We decided to rely on a numerical optimization with an integer based genetic algorithm to implement a systematic selection of model parameters. For the optimization, the scaled PRESS served as an objective function. In more detail, cross-validation was performed by iteratively excluding a complete run from PLS model calibration. The reasoning was to make cross-validation representative for the prediction of future runs and thereby maximize the predictive power of the PLS model. The so calibrated models were able to predict most of the variation of the measured concentrations based on the spectral data ($Q^2 > 0.9750$).

For the microplate setup, it is worth noting that the calibration data spans a range from 1 mg/mL to 2 mg/mL of mAb with the corresponding surrogate drug concentrations. As the external validation shows, the model is able to predict the reaction course for different concentrations in the calibration space. Interestingly, the RMSEP for NPM lies below the corresponding RMSECV. For CPM, the RMSEP is noticeably higher than the RMSECV. This seems to be related to a slight offset in the PLS prediction (Figure 5.5B). Nevertheless, the results show that it is possible to quantitatively monitor conjugation reactions of NPM and CPM to an IgG1 antibody in the microplate format by UV/Vis spectroscopy. The results furthermore show that the chosen way of model optimization did not cause a strong overfit.

Lab-scale experiments led to RMSECV and Q^2 values similar to those found in the microplate experiments. The smooth prediction of the PLS models indicates that the error of the model is mainly related to systematic errors and not to the measurement noise. For reactions with varying protein concentration, it would be possible to estimate the concentration by a PLS model. The DAD experiments successfully show the ease of implementation of the approach in a lab-scale format. As the DAD measurements are very fast, the approach facilitates real-time monitoring, which may be beneficial for kinetic studies or process monitoring and control.

5.5 Conclusion

In summary, we established a novel spectroscopic PAT approach for monitoring ADC conjugation reactions. In two experimental setups, with two different detectors, the conjugation process of surrogate drugs to a mAb was monitored by UV/Vis absorption spectroscopy and PLS regression. The results show that UV/Vis spectroscopy allows to monitor conjugation reactions in microplates as well as in lab-scale. The method may thus simplify process development by reducing the analytical bottle neck. This may be especially interesting in combination with High-throughput Process Development (HTPD) on liquid-handling stations for ADCs [104], [105]. In lab-scale, the method allows for real-time process monitoring. Due to the flexibility and ease of implementation, the method may be further developed to a PAT approach for conjugation monitoring at commercial scale.

Future steps should focus on testing the method with cytotoxic drugs. While common drugs contain chromophores, the solvatochromic behavior of those drugs is unknown. Furthermore, the position of the engineered cysteines may have a strong impact on the solvent exposure of the drug and, thus, the change in hydrophobicity in the environment of the drug upon conjugation. Other techniques more sensitive to solvatochromism (e.g. fluorescence spectroscopy) or the changing of covalent bonds (e.g. vibrational spectroscopy) could be evaluated. Due to the simplicity of UV/Vis absorption spectroscopy, it is still a reasonable first choice for future studies.

Acknowledgements

5.6 Supplementary Material
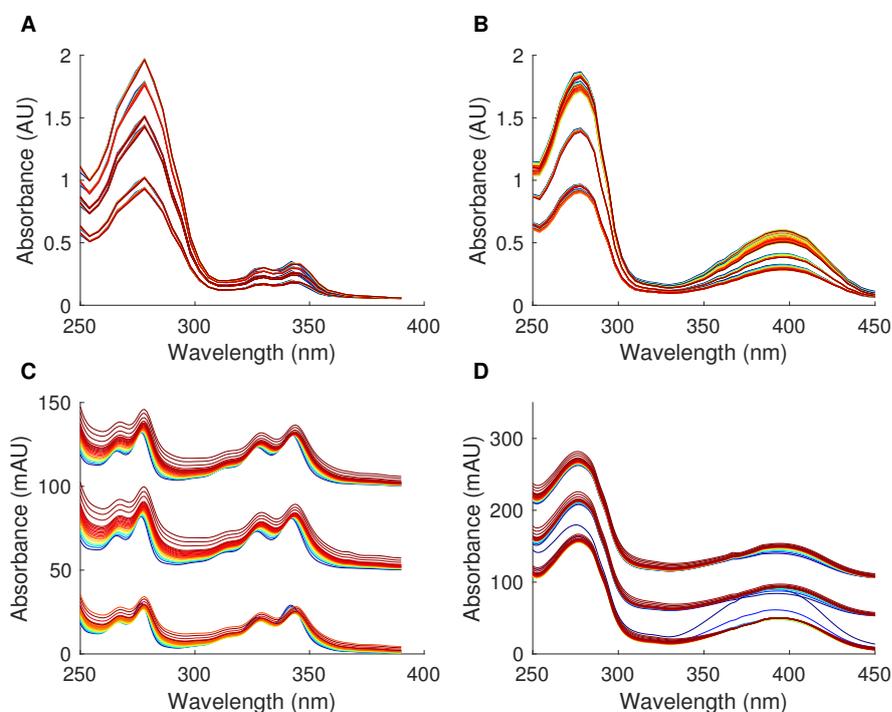
Figure 5.7: Raw spectra of all calibration samples. The spectra are colored according to the reaction progress from blue to red. The microplate experiments are depicted in the top row, while the bottom row shows the spectra recorded in the lab-scale setup. Since the lab-scale experiments were performed at the same nominal mAb concentration, the different runs are artificially offset by 50 mAU. The left column shows experiments with NPM, the right column shows experiments with CPM.
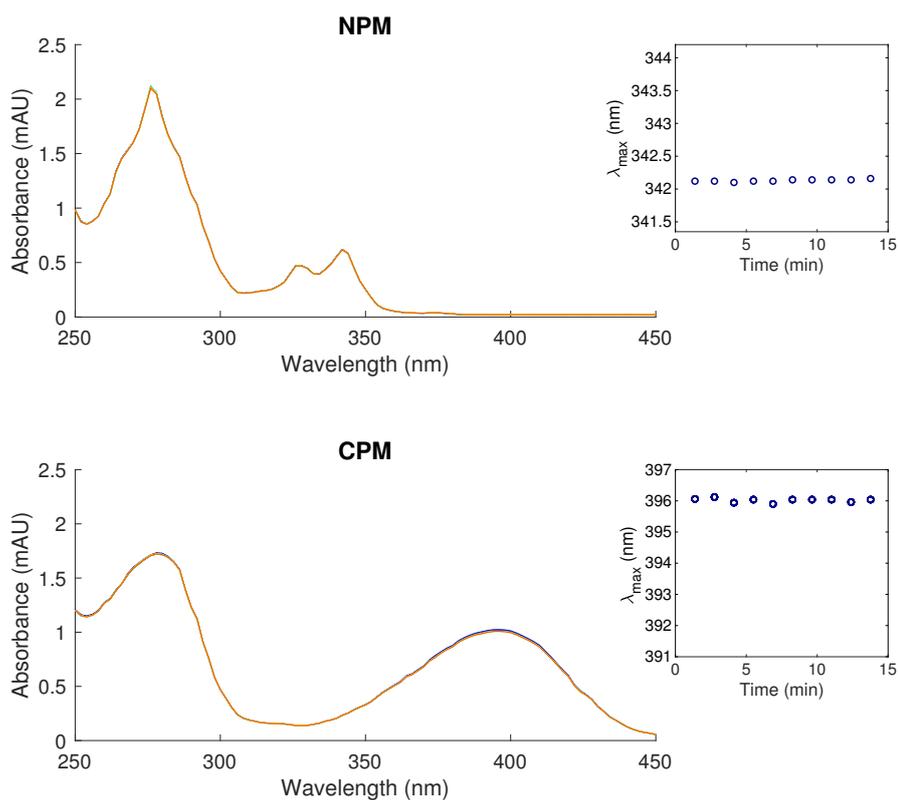
Figure 5.8: Raw spectra of a mixture of mAb and quenched drug recorded over the course of 15 min. The spectra are colored according to "reaction time" from blue to red. mAb, drug and NAC concentrations are the same as in the lab-scale experiments. The surrogate drugs were quenched prior to addition to the mAb solution in order to prevent the conjugation reaction. DMSO content is 10 % as in the other experiments. The evolution of the band maxima of the drugs over time is shown on the right side. No shift in band maxima is observed.
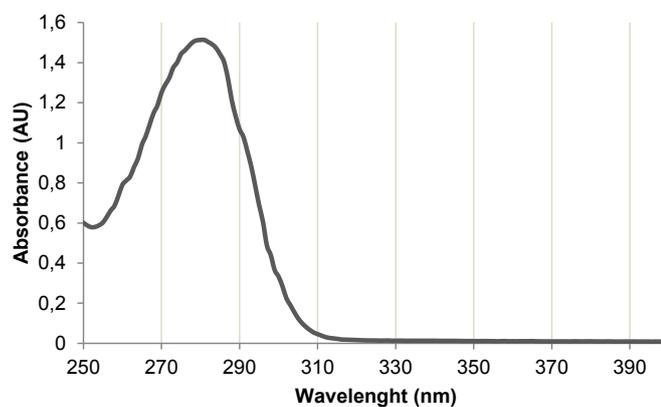
Figure 5.9: Pure component UV/Vis absorbance spectrum of mAb measured in Tecan plate reader M200 Pro at a concentration of $2\,g/L$ (in $50\,mmol/dm^3$ sodium phosphate buffer).
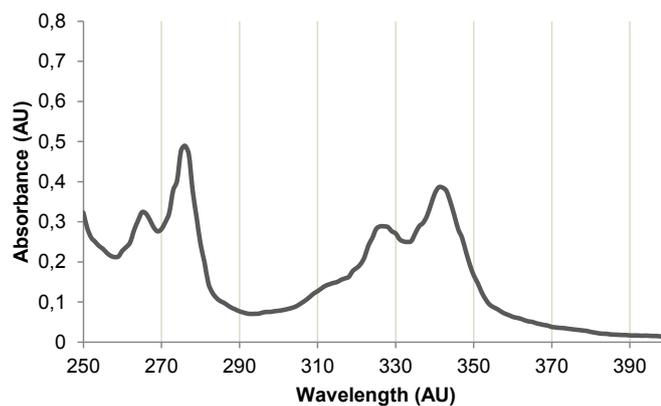
Figure 5.10: Pure component UV/Vis absorbance spectrum of NPM in phosphate buffer containing $10\,\%$ of DMSO measured in Tecan plate reader M200 Pro.
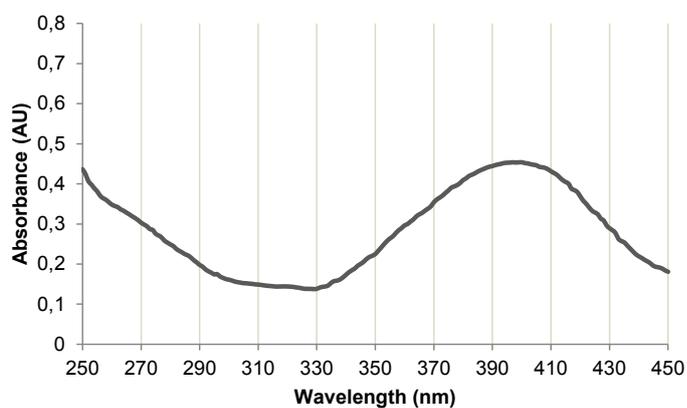
Figure 5.11: Pure component UV/Vis absorbance spectrum of CPM in phosphate buffer containing 10 % of DMSO measured in Tecan plate reader M200 Pro.

6

In-line Fourier-transform Infrared Spectroscopy as a Versatile Process Analytical Technology in Preparative Protein Chromatography

Steffen Großhans[*,1], Matthias Rüdt[*,1], Adrian Sanden[*,1], Nina Brestrich[1], Josefine Morgenstern[1], Stefan Heissler[2], Jürgen Hubbuch[1]

*   Contributed equally
[1]   Institute of Engineering in Life Sciences, Section IV: Biomolecular Separation Engineering, Karlsruhe Institute of Technology (KIT), Germany
[2]   Institute of Functional Interfaces, Group: Chemistry of Oxydic and Organic Interfaces, Karlsruhe Institute of Technology (KIT), Germany

Abstract

FTIR is a well-established spectroscopic method in the analysis of small molecules and protein secondary structure. However, FTIR is not com-

monly applied for in-line monitoring of protein chromatography. Here, the potential of in-line FTIR as a PAT in downstream processing was investigated in 3 case studies addressing the limits of currently applied spectroscopic PAT methods. A first case study exploited the secondary structural differences of mAb and lysozyme to selectively quantify the two proteins with PLS giving RMSECV of 2.42 g/l and 1.67 g/l, respectively. The corresponding Q^2 values are 0.92 and, respectively, 0.99, indicating robust models in the calibration range. Second, a process separating lysozyme and PEGylated lysozyme species was monitored giving an estimate of the PEGylation degree of currently eluting species with RMSECV of 2.35 g/l for lysozyme and 1.24 g/l for PEG with Q^2 of 0.96 and 0.94, respectively. Finally, Triton X-100 was added to a feed of lysozyme as a typical process related impurity. It was shown that the species could be selectively quantified from the FTIR 3D-field without PLS calibration. In summary, the proposed PAT tool has the potential to be used as a versatile option for monitoring protein chromatography. It may help to achieve a more complete implementation of the PAT initiative by mitigating limitations of currently used techniques.

6.1 Introduction

Preparative chromatography of biopharmaceuticals is typically monitored by measuring univariate signals such as pH, conductivity, pressure and UV/Vis absorbance at a given wavelength [19], [106]. Among those, especially single wavelength UV/Vis spectroscopy has been a staple for process monitoring of biopharmaceutical chromatography due to its linear response to protein concentration as well as its broad dynamic range, sensitivity, and robustness. For all the advantages, single wavelength UV/Vis absorption measurements generally do not allow for selective quantification of multiple co-eluting proteins [58].

Even before the PAT initiative by the FDA in 2004 [1], research towards more selective monitoring methods for preparative chromatography was conducted. But the often small differences between biopharmaceutical product and protein as well as non-protein contaminants make this a non-trivial task [8], [74]. As a possible solution, fast at- or on-line analytical methods, such as analytical chromatography, have been established. Discrete samples are taken from the process stream and analyzed on the spot. This approach has been proposed for controlling capture [13], [107], [108] and polishing steps [11], [14]. However, at- or on-line analytical chromatography is equipment-wise complex requiring a sampling module as well as

an analytical chromatography system close to the process stream. Furthermore, the sampling and analysis time may be too long compared to the typical time frame available for taking process decisions.

An alternative approach exploits slight differences in UV/Vis absorption spectra of different components to selectively quantify different species by chemometric methods [74]. The approach yields results quickly enough to allow for real-time process decisions in chromatography [42], [43], [76] and works for minute spectral differences [39]. However, in the commonly measured spectral ranges, UV/Vis spectroscopy lacks sensitivity towards relevant aspects of protein structure, notably the secondary structure [32]. Furthermore, organic compounds are often not UV active (e.g. sugars, polyols, and PEG [109], [110]) or they may obscure the protein signal (e.g. Triton X-100 [111] and benzyl alcohol [32]). Due to the high sensitivity, UV/Vis absorption spectroscopy is also prone to detector saturation [74], [81].

FTIR allows to address several of those short-comings. Similar to UV/Vis spectroscopy, FTIR is a non-destructive, quantitative, and quick method which can be performed in-line [112]–[114]. FTIR measures the vibrational modes of samples and thereby provides a spectroscopic fingerprint for different organic molecules. Proteins absorb in the Infrared (IR) spectral range mainly due to vibrations of the polypeptide backbone [32], [35], [36]. Based on the backbone vibrations, FTIR grants insight into the secondary structure of the measured proteins. In consequence, FTIR is a widely used method for assessing the structural integrity of proteins during protein purification and formulation [32]. Furthermore, FTIR was previously used as an at-line PAT tool in downstream processing of biopharmaceuticals for quantifying product content, HMWs, and HCPs [45], [47].

In this work, in-line FTIR as a PAT tool for preparative protein purification was implemented. An FTIR instrument was coupled to a lab-scale preparative chromatography system to perform the experiments. Three case studies were selected to investigate potential applications of FTIR as PAT tool. First, a mixture of lysozyme and mAb was chosen due to the significant differences in secondary structure of the two proteins. While lysozyme mainly consists of alpha-helices (PDB ID 193L), mAb largely consists of beta-sheets (PDB ID 1HZH). The expected spectral differences can be used to selectively quantify the 2 proteins by PLS regression. 4 linear gradient elutions with varying gradient lengths were performed. Based on the results, a PLS model for each protein was optimized. The error of the PLS model was assessed by cross validation. Second, the preparative separation of PEGylated lysozyme was monitored. In contrast to UV/Vis spectroscopy, PEG gives a distinct signal in IR which can be used for quantification by PLS regression. Again, 4 linear gradient elutions were performed for the

calibration of two PLS models. Finally, the potential to monitor process related impurities using in-line FTIR was demonstrated by adding Triton X-100 to a feed solution of lysozyme. Triton X-100 is employed for virus inactivation in biopharmaceutical production and has to be removed from the product [111], [115]. Based on an off-line calibration curve, mass-balancing of Triton X-100 in the flow-through during product loading was performed.

6.2 Materials and Methods

6.2.1 Experimental Setup

In-line FTIR measurements were performed using a Tensor 27 by Bruker Optics (Ettlingen, Germany) connected to an ÄKTApurifier system by GE Healthcare (Little Chalfort, UK). The chromatography system was equipped with a P-900 pump, a P-960 sample pump, UV-900 UV/Vis cell, and a Frac-950 fraction collector (all GE Healthcare). Unicorn 5.31 (GE Healthcare) was used to control the system. The FTIR was equipped with a liquid nitrogen cooled Mercury Cadmium Telluride (MCT) detector and a BioATR II (Bruker Optics) with a flow-cell insert and a 7 reflection silicon crystal. The instrument was controlled by OPUS 7.2 (Bruker Optics).

In this setup, the effluent stream from the column outlet was diverted through the FTIR instrument and then back into the UV/Vis cell in the ÄKTApurifier system. The flowpath is illustrated in Figure 6.1. The delay volume between the FTIR and the fraction collector was determined gravimetrically. As the flow rate was set in the chromatographic methods, the measurement of the delay volume enables the correlation of spectral data from the FTIR to collected fractions.

The interconnection between OPUS and Unicorn was achieved using a software solution developed in-house consisting of a Matlab (The Mathworks, Natick, MA, United States) script and a VBScript in the built-in visual basic script engine of OPUS. The custom software enables the start of a measurement at a time defined by Unicorn by sending a digital signal through the I/O port of the pump of the ÄKTApurifier System. The signal is captured by a USB-6008 data acquisition device (National Instruments, Austin, Tx, United States) controlled by Matlab which in turn triggers the measurement in OPUS.
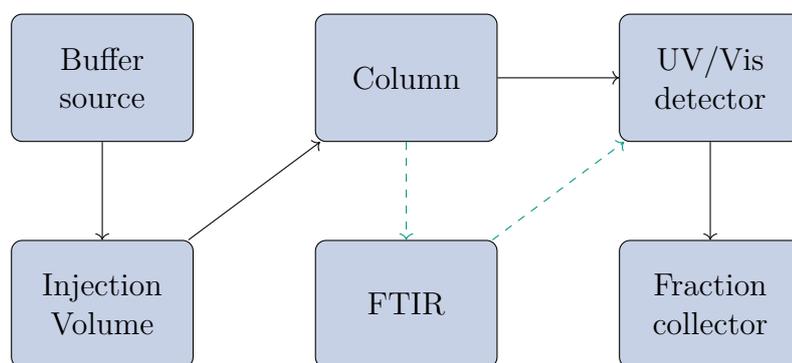
Figure 6.1: Schematic representation of the flow path in the custom chromatography setup, solid lines represent the common flow path in the ÄKTApurifier while the dashed line represents the modification.

6.2.2 Proteins and Buffers

All solutions were prepared using water purified by a PURELAB Ultra water purification system by ELGA Labwater (High Wycombe, United Kingdom). Buffers were filtered using $0.2\,\mu$m filter purchased from Sartorius (Göttingen, Germany) and degassed by sonification before use. All buffers were pH adjusted using 32 % HCl (Merck, Darmstadt, Germany).

Lysozyme was purchased from Hampton Research (Aliso Viejo, CA, United States). mAb was provided by Lek Pharmaceuticals d.d. (Mengeš, Slovenia) as a virus-inactivated Protein A eluate pool.

Preparative CEX chromatography runs in case study I and III were conducted with 50 mM sodium citrate buffer as equilibration buffer and with an added 500 mM NaCl as elution buffer. Both buffers were adjusted to pH 6.0. Sodium citrate tribasic dihydrate was purchased from Sigma-Aldrich (St. Louis, MO, United States), sodium chloride was purchased from Merck. For the CEX chromatography experiments in case study II, 25 mM sodium acetate buffer (pH 5.0) was used as equilibration buffer. As elution buffer, 25 mM sodium acetate buffer with 1 M NaCl (pH 5.0) was used. Sodium acetate trihydrate was purchased from Sigma-Aldrich. Batch-PEGylation of lysozyme was performed in 25 mM sodium phosphate buffer at pH 7.2 using sodium phosphate monobasic dihydrate (Sigma-Aldrich) and di-sodium hydrogen phosphate dihydrate (Merck).

Analytical CEX was carried out at pH 8.0 using 20 mM Tris (Merck) buffer for equilibration and 20 mM Tris buffer with 700 mM NaCl for elution.

PEGylation of Lysozyme

The PEGylation protocol was adapted from [116]. Briefly, activated 5 kDa PEG was purchased as Methoxy-PEG-propionaldehyde (mPEG-aldehyde, Sunbright ME-050 AL) from NOF Corporation (Tokyo, Japan). Sodium cyanoborohydride (NaCNBH$_3$, Sigma Aldrich) was added to the reaction buffer to a concentration of 20 mM as reducing agent. mPEG-aldehyde was added to a molar PEG-to-protein ratio of 6.67. After 3 h, the mixture was diluted volumetrically 7-fold using acetate equilibration buffer and loaded onto the chromatography column.

6.2.3 Preparative Chromatography Experiments

For all chromatography experiments, FTIR spectra were recorded continuously in the chromatography mode of OPUS with a resolution of 2 cm^{-1} in a range from 4000 cm^{-1} to 900 cm^{-1} without averaging multiple scans. In given setup, each measurement took 3.22 s. Background measurements in the beginning of chromatographic runs were taken at the same resolution with 400 scans in equilibration buffer. All experiments were conducted twice, once with protein injection and once with buffer only as a blank run. The FTIR spectra from the blank runs were subsequently subtracted from the protein runs to account for spectral effects by the gradient.

Case Study I: Selective Protein Quantification

For case study I, a HiTrap column by GE Healthcare prepacked with SP Sepharose FF resin (Cross-Validation (CV) 5 ml) was used. The column was loaded to a density of 18.75 g/l, consisting of 12.5 g/l lysozyme and 6.25 g/l monoclonal antibody. The flow rate for all experiments was set to 0.5 ml/min. The column was equilibrated in low salt buffer for 5 CV before injection. The 50 ml sample was injected using a 50 ml superloop from GE Healthcare. Elution was carried out with a linear gradient from 0 % to 100 % high salt buffer with gradient lengths of 1 CV, 2 CV, 3 CV, and 4 CV. After elution a high salt wash of 8 CV was performed for column regeneration. The effluent was collected over the complete injection and elution in 500 µl fractions for off-line analytics.

Case Study II: Separation of PEGylated Lysozyme Species

The experiments with different PEGylated lysozyme species were conducted with Toyopearl Gigacap S-650M resin prepacked in a MiniChrom column (CV 5 ml) by Tosoh (Griesheim, Germany). The column was loaded to a

density of 50 g/l of the heterogeneous batch PEGylation. The sample pump was run at 1 ml/min for loading. For the remaining chromatography run, the flow rate was set to 0.5 ml/ min. The column was first equilibrated for 1 CV, followed by an injection of 57.6 CV of sample solution. Linear gradient elutions from 0 % to 100 % high salt buffer were conducted with gradients of 2 CV, 3 CV, 4 CV, and 5 CV length, followed by 2 CV high salt rinse. The effluent was collected from the beginning of the gradient until the end of the high salt rinse in 500 µl fractions for off-line analytics.

In some of the collected fractions unconjugated lysozyme started to precipitate after elution probably due to the low pH, high salt concentration or low temperature [116], [117]. Fractions and the corresponding spectra showing signs of precipitation were excluded from PLS model calibration.

Case Study III: Process-Related Impurity

For the simulated process-related impurity experiments, a HiTrap column by GE Healthcare prepacked with SP Sepharose FF resin (CV 5 ml) was used. Triton X-100 Biochemica was purchased from AppliChem GmbH (Darmstadt, Germany). The column was loaded with 5 ml of 25 g/l lysozyme and 10 g/l Triton X-100 solution [115]. The elution step was set to 2 CV.

Reference samples were generated by diluting defined amounts of Triton X-100 in equilibration buffer at concentrations from 1.25 g/l to 10 g/l. To generate a calibration curve, the samples were manually applied onto the ATR crystal. FTIR measurements were performed with 400 scans for background and samples.

6.2.4 Analytical CEX Chromatography

As reference analytics for case study I, analytical CEX chromatography was performed using a Dionex UltiMate 3000 liquid chromatography system by Thermo Fisher Scientific (Waltham, MA, United States). The system was composed of a HPG-3400RS pump, a WPS-3000TFC analytical autosampler, a TCC-3000RS column thermostat, and a DAD3000RS detector. The system was controlled by Chromeleon 6.80 (Thermo Fisher Scientific). Fractions from preparative CEX chromatography were analyzed off-line on a Proswift SCX-1S 4.6 mm × 50 mm column by Thermo Fisher Scientific. A flow rate of 1.5 ml/ min was used. For each sample, the column was first equilibrated for 1.8 min with equilibration buffer. Next, 20 µl sample was injected into the system and washed for 0.5 min with equilibration buffer. A linear gradient was performed during the next 2 min from 0 % to 50 % followed by a step to 100 % elution buffer which was maintained for 2 min.

For the experiments in case study II, a Vanquish UHPLC system (Thermo Fisher Scientific) was used. The Vanquish UHPLC System consisted of a Diode Array Detector HL, a Split Sampler FT, a Binary Pump F and a Column Compartment H including a preheater and post-column cooler (all Thermo Fisher Scientific). The same buffers, column, and flow rate were used as for case study I. After injecting $5\,\mu l$ of sample, the column was washed for 0.5 min. Subsequently, a bilinear gradient was performed from 0 % to 50 % elution buffer over 5 min and 50 % to 100 % elution buffer over 1.75 min. After the elution a high salt strip at 100 % was run for 1 min. Calibration was performed by a dilution series of pure lysozyme. Since PEG does not absorb in UV/Vis, solely lysozyme contributes to the absorption signal. Peak identification with respect to the PEGylation degree was conducted using purified samples prepared according to [110]. From the molar concentration of PEGylated lysozyme species, the molar concentration of PEG was calculated.

6.2.5   Data Analysis

All data analysis was performed in Matlab. For case study I and II, the data was first preprocessed and subsequently fitted with PLS-1 models by the SIMPLS algorithm [99]. Preprocessing consisted of linearly interpolating off-line analytics to be on the same time scale as the FTIR spectra. For case study I and II, spectral data above $2000\,\mathrm{cm}^{-1}$ resp. above $3100\,\mathrm{cm}^{-1}$ was discarded. Next, a Savitzky-Golay filter with a second order polynomial was applied on the spectra and optionally the first or second derivative taken [82]. Cross-validation was performed by excluding one chromatography run, calibrating a PLS model on the remaining runs and calculating a residual sum of squares on the excluded run. This procedure was repeated until all runs had been excluded once. All residual sums of squares for the different submodels were subsequently summed giving the PRESS. The PRESS was scaled according to Wold et al. by the number of samples and latent variables used in the PLS model [30]. Based on the scaled PRESS, an optimization was performed using the built-in genetic algorithm of Matlab for integers [100]. The genetic algorithm optimized the window width of the Savitzky-Golay filter, the order of derivative, as well as the number of latent variables for the PLS-1 model. The RMSECV was calculated from the PRESS by dividing by the total number of samples. The Q^2 values were calculated by dividing the PRESS by the summed squares of the response corrected to the mean [30].

For case study III, spectral data was smoothed both in direction of time and wavenumber using a Savitzky-Golay filter with a second order

polynomial and a frame length of 17 and 51 respectively. A linear baseline was calculated and subtracted for each spectrum individually to account for a non-horizontal non-zero baseline. The baseline subtraction was performed on the reference spectra as well as the spectra from the chromatography experiment. Based on the area under the spectrum between wavenumbers $1007\,\mathrm{cm}^{-1}$ to $1170\,\mathrm{cm}^{-1}$, a mass balance for Triton X-100 was calculated from the spectral data of the chromatography run. The volume represented by each spectrum was calculated from the recording time and the volumetric flow rate of the experiment. Triton X-100 masses in each segment were calculated utilizing the calibration curve and summed up over time.

6.3   Results and Discussion

In-line FTIR measurements were applied as a PAT tool for different preparative chromatographic protein separations. In three different case studies, FTIR was used for selective quantification of different species. First, background correction of the FTIR chromatograms is discussed which was necessary for further data processing. In a first case study, the capability of FTIR to measure differences in secondary structure in-line and utilize the differences for selective quantification of mAb and lysozyme was demonstrated. A second case study made use of the absorption of PEG in IR to monitor the PEGylation degree of eluting PEGylated lysozyme species. Finally, the third case study used the selectivity of FTIR to selectively quantify Triton-X 100, a detergent used for viral inactivation.

6.3.1   On Background Subtraction and Spectral Preprocessing

Background subtraction for in-line FTIR measurements is of major importance as water has an absorption band around $1600\,\mathrm{cm}^{-1}$ (cf. Figure 6.2A) which coincides with the most prominent protein band amide I . The spectral processing workflow is illustrated in Figure 6.2 using data from case study I. Specifically the elution of mAb and lysozyme using a $4\,\mathrm{CV}$ gradient is shown. Most of the water absorption can be eliminated by taking a background with the equilibration buffer in the beginning of each chromatographic run. The water band is, however, also influenced by the salt content of the buffer around $1650\,\mathrm{cm}^{-1}$. Salt gradients therefore cause a change in absorption over the run (cf. Figure 6.2A and B). To reduce buffer effects, it is important to find a suitable dynamic background correction. An approach based on reference spectra matrices and chemo-

metric correlations was not implemented due to the overlap of water and protein bands [118]. Instead, an alternative approach was chosen. Based on the retention time, a blank run without protein but including the salt gradient was subtracted from the actual preparative run (cf. Figure 6.2C). The resulting chromatogram provided a smooth baseline over the whole experiment. After baseline correction, additional data preprocessing was performed. The single scan spectra were smoothed by a Savitzky-Golay filter to reduce random noise (cf. Figure 6.2D) and to take derivatives on the spectral data.

6.3.2   Case Study I: Selective Protein Quantification

mAb and lysozyme feature significant differences in secondary structure. While mAb consists largely of beta-sheets (PDB ID 1HZH), lysozyme contains mainly alpha-helices (PDB ID 193L). These differences make the two proteins simple model components to study the performance of in-line FTIR for selectively quantifying proteins. The bands visible between $1200\,cm^{-1}$ to $1700\,cm^{-1}$ in Figure 6.2D are characteristic amide bands associated with the protein backbone [32], [35], [36]. Especially the amide I band is frequently used for assessing the secondary structure of proteins. For PLS calibration, all wavenumbers below $2000\,cm^{-1}$ were taken into account to include all protein bands without interference at the boundary due to the Savitzky-Golay filter.

Based on 4 CEX runs, 2 PLS-1 models were optimized for selective quantification of mAb and lysozyme respectively. The resulting model parameters are listed in table 6.1. Figure 6.3 shows a comparison from off-line analytics and the prediction of PLS models. Both PLS models match peak maxima and peak widths well and are able to discern the two components. For mAb, a root mean square error of cross validation (RMSECV) of $2.42\,g/l$ was reached. For lysozyme, the RMSECV was $1.67\,g/l$. The corresponding Q^2 values were 0.92 and 0.99, respectively. The high Q^2 values show, that a large part of the variation in the off-line concentration measurements could be explained by the PLS model. The differentiation between different proteins may however become more challenging for smaller differences in secondary structure. Interestingly, the combination of Savitzky-Golay filtering and PLS modeling allowed to reduce the measurement noise compared to single-wavelength measurements. As shown by Figure 6.2C and 6.3, the measurement noise in the IR spectra is higher than the noise observed in the PLS prediction. By filtering and projecting the spectra to latent variables, random noise is reduced [30], [82]. Furthermore, $3.23\,s$ measurement time makes FTIR quick enough for monitoring most practical preparative chro-
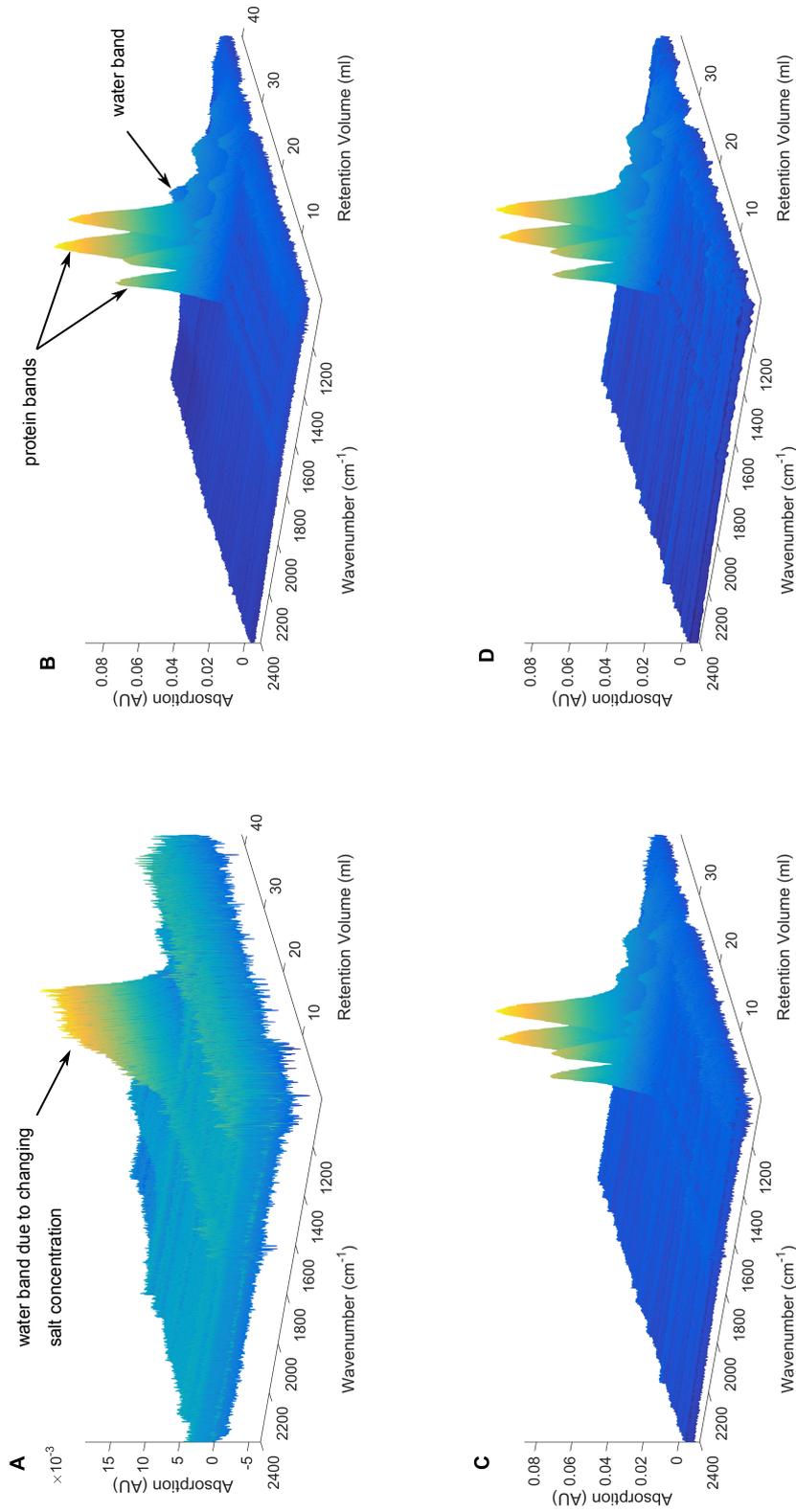
Figure 6.2: (Continued on the following page.)

Figure 6.2: Work flow for data treatment of chromatography spectra illustrated with data from case study I, 4 CV run: background run – salt gradient without protein (A); raw spectra of the run with protein (B); spectral data after the background has been subtracted (C); data after smoothing by Savitzky-Golay algorithm (D).

matography applications in real-time. In-line FTIR spectroscopy allowed to cover high concentration ranges. The predicted concentration of lysozyme during the 1 CV run reaches 112 g/l without any interference from detector saturation. The measurement setup therefore covers all concentrations typically occurring in preparative protein chromatography.

Table 6.1: Model parameters for case study I and II are listed below including the parameters for the Savitzky-Golay filter and the latent variables of the PLS-1 model. Additionally, the RMSECV for each model is listed.

	Case study I		Case study II	
	mAb	lysozyme	lysozyme	PEG
Savitzky-Golay Window	215	21	101	361
Derivative	0	0	2	2
Latent variables	3	7	6	8
RMSECV (g/l)	2.41	1.63	2.35	1.24

In summary, the results show that FTIR in conjunction with PLS modeling can differentiate in-line between proteins based on their secondary structure and has the potential to be applied for real-time monitoring and control of preparative chromatography.

6.3.3 Case Study II: Separation of PEGylated Lysozyme Species

In conventional chromatography systems, the separation of differently PEGylated species cannot be monitored holistically as PEG does not absorb in UV. Contrary to that, PEG produces a number of prominent bands in IR. A strong band around $1090\,\mathrm{cm}^{-1}$ with multiple shoulders is characteristic for $C{-}O$ stretching [119]. Due to symmetric CH_2 stretching, PEG furthermore generates a doublet at $2884\,\mathrm{cm}^{-1}$ and $2922\,\mathrm{cm}^{-1}$. Bands occurring between $1200\,\mathrm{cm}^{-1}$ to $1700\,\mathrm{cm}^{-1}$ are related to the protein backbone with some interference from PEG $C{-}H$ bending.

Figure 6.4 shows a typical chromatographic separation of PEGylated lysozyme species. During the elution, the ratio between PEG and protein
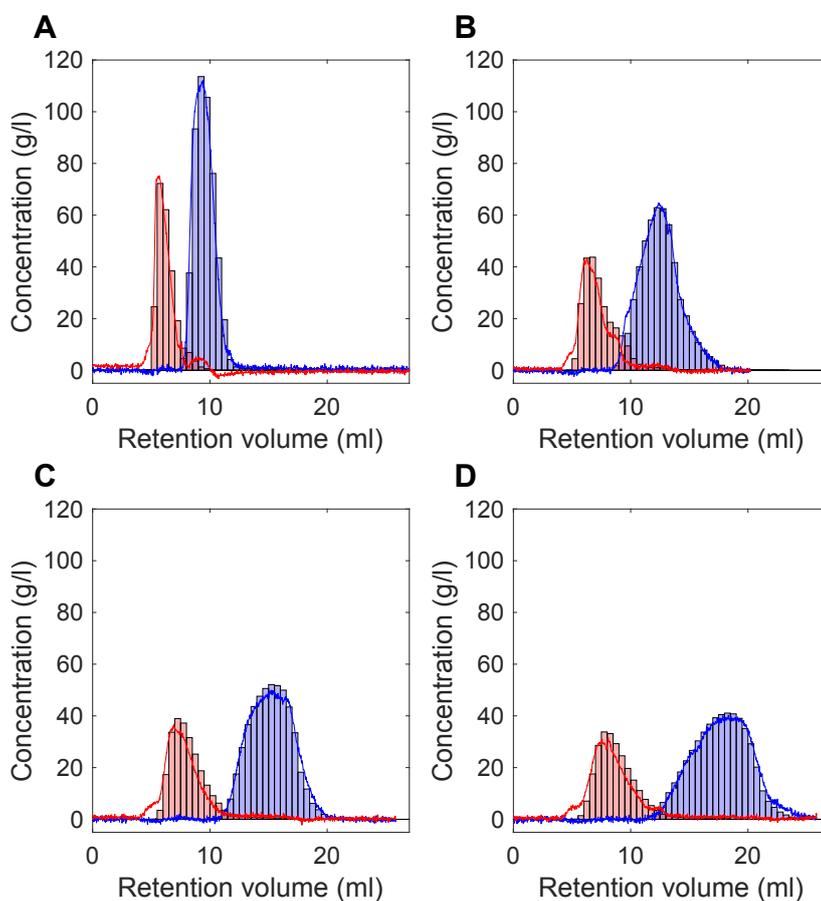
Figure 6.3: 4 chromatographic runs are shown for in-line FTIR measurements and selective quantification of mAb and lysozyme. The red bars and lines refer to the mAb off-line measurement and mAb PLS prediction, respectively. The blue bars and lines refer to the lysozyme off-line measurement and lysozyme PLS prediction, respectively. The different subplots show different gradient lengths: A 1 CV, B 2 CV, C 3 CV, D 4 CV.

bands decreases. First, with a retention volume of 6.8 ml, the absorption of the C−O band at 1090 cm^{-1} (denoted as CO$_1$ in Figure 6.4) exceeds the absorption of amide I band (AI$_1$). For the second peak with a retention volume of 10.3 ml the absorption of the amide I (AI$_2$) is higher than for the C−O stretching band (CO$_2$). The last peak does not show characteristic PEG bands, i.e. consists of unconjugated lysozyme. The order of elution followed a descending degree of PEGylation which is in line with previous publications [110], [120], [121].
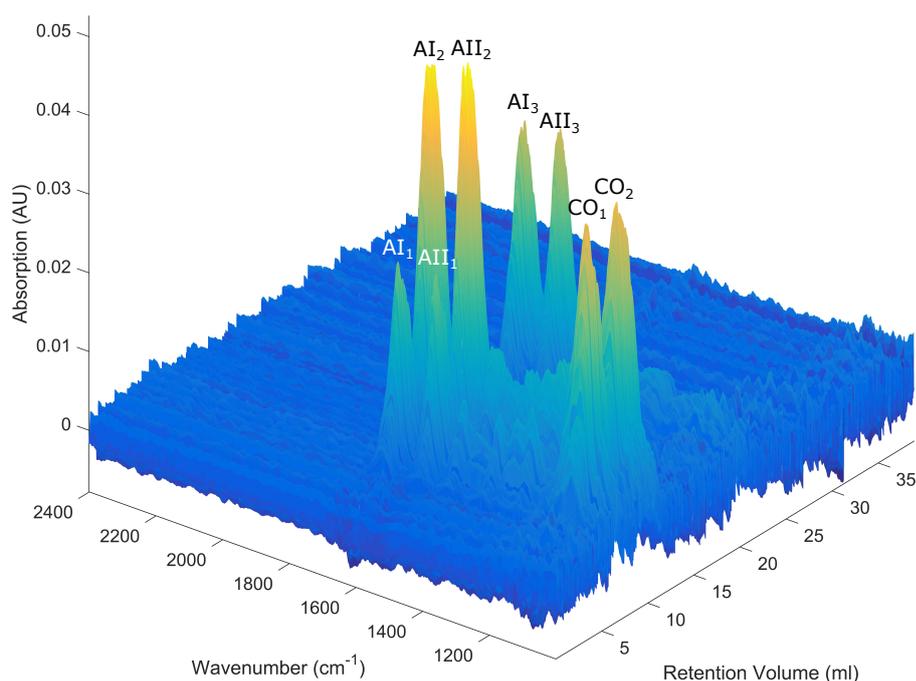
Figure 6.4: Elution of PEGylated lysozyme species from a CEX column with a gradient length of 5 CV. Bands visible between wavenumbers $1200 \, \text{cm}^{-1}$ to $1700 \, \text{cm}^{-1}$ are the characteristic amide bands associated with protein. The major protein bands amide I and amide II are mark as AI and AII, respectively. The band at approximately $1100 \, \text{cm}^{-1}$ is characteristic for PEG ($C-O$ stretching, marked as CO). The subscript numerals refer to the elution order.

Based on the evaluation of IR absorption bands, it was decided to include all wavenumbers from $900 \, \text{cm}^{-1}$ to $3100 \, \text{cm}^{-1}$ into PLS model calibration. Initial PLS calibration on the concentration of the different PEGylated lysozyme species showed that the conjugation did not cause large enough

band shifts to allow for selective quantification of the different PEGylated lysozyme species. Instead, two PLS models were fitted on the total PEG resp. lysozyme concentration independently. PEG concentration was calculated by weighting the off-line lysozyme concentration according to the PEGylation degree. In table 6.1, the optimization results are summarized. Figure 6.5 compares the PLS prediction with off-line analytics. RMSECV values of 1.24 g/l and 2.35 g/l were reached for the PEG and lysozyme concentration, respectively. The corresponding Q^2 values were respectively 0.96 and 0.94 showing that the PLS models predicted the responses well. Based on the PEG and lysozyme concentrations, a molar ratio could be calculated corresponding to the current average PEGylation degree. To simplify visual interpretation, the molar ratio is only plotted if the lysozyme concentration exceeded its RMSECV 3-fold.

The predicted PEG and lysozyme concentrations accurately followed the concentrations measured by off-line analytics. Furthermore, the molar ratio gives a suitable tool for in-line monitoring of the elution of different PEG species. Interestingly, the two PLS models are able to extend their prediction over the calibration range, i.e. to perform a weak extrapolation. This can be seen as the PEG to lysozyme ratio exceeds the value of two, which limits the calibration range spanned by off-line analytics. Higher PEGylated species of lysozyme do however occur and could be measured by the FTIR [110], [122].

In summary, FTIR allows to monitor not only the protein and PEG concentration but also the PEGylation degree during chromatographic separations.

6.3.4 Case Study III: Quantification of a Process Related Impurity

Triton X-100 is used for viral inactivation of biopharmaceuticals if pH treatment has to be circumvented, e.g. for Factor VIII or pH sensitive mAbs [111], [115]. To achieve viral inactivation, Triton X-100 concentration needs to be above a minimal level. Typically, a concentration of $1\,\%$ (w/V) is used. Here, Triton X-100 concentration of a mock virus inactivation batch was monitored during the subsequent load phase onto a chromatographic column. During the chromatographic run, in-line FTIR measurements were perform (cf. Figure 6.6).

In IR, Triton X-100 causes a characteristic band due to $C-O$ stretching at $1090\,cm^{-1}$. By comparison of the blank run and the actual experiment it was concluded that Triton X-100 is not retained on the column and is mainly
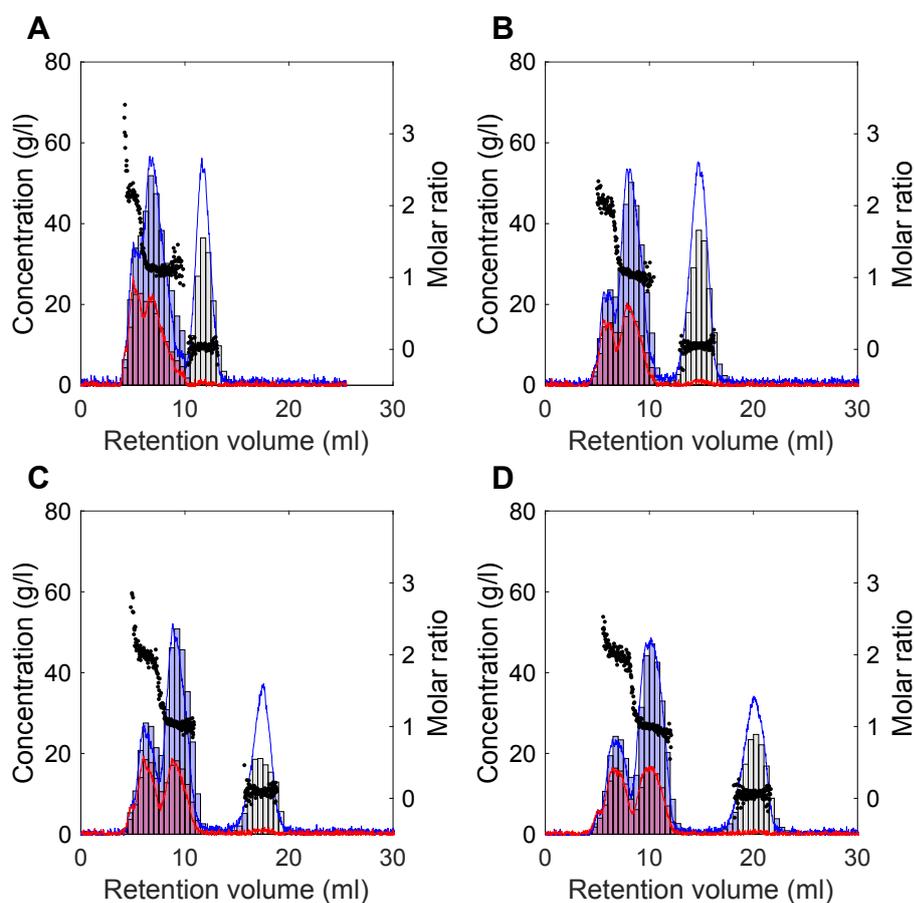
Figure 6.5: 4 chromatographic runs are shown for in-line FTIR measurements and selective quantification of PEG and lysozyme. The red bars and lines refer to the PEG off-line measurement and PEG PLS prediction, respectively. The blue bars and lines refer to the lysozyme off-line measurement and lysozyme PLS prediction, respectively. Grey bars correspond to measured protein concentrations on partially precipitated samples. Black dots show the molar ratio between PEG and lysozyme, i.e. the current mean PEGylation degree. The different subplots show different gradient lengths: A 2 CV, B 3 CV, C 4 CV, D 5 CV.

present in the flow-through. The flow-through occurred between 5.5 ml to 11 ml. As Triton X-100 and protein spectra only weakly interfere with each other, the Triton X-100 content was measured by simply correlating the band area of C−O stretching from $1007\,\mathrm{cm^{-1}}$ to $1170\,\mathrm{cm^{-1}}$ to the Triton X-100 concentration. A linear regression for the calibration curve resulted in a
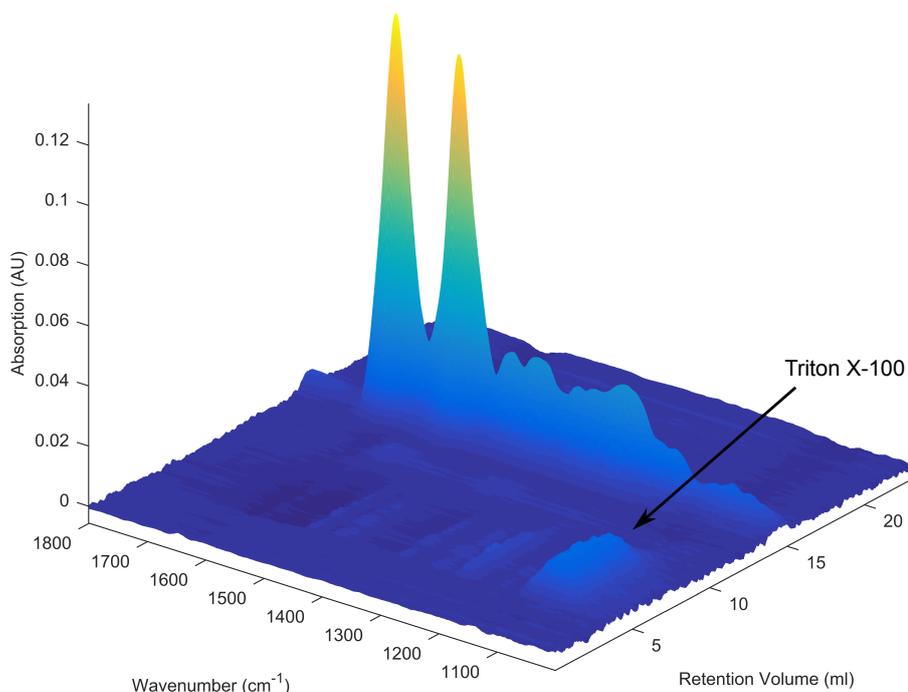
Figure 6.6: Triton X-100 as a process related impurity can be seen in the flow-through of the cation-exchange experiment from 5.5 ml to 11 ml at 1090 cm^{-1}.

$R^2 > 0.9997$. Based on the calibration curve, in-line mass-balancing could be performed. The mass balance for Triton X-100 showed a recovery rate of 94.12 % in the flow-through. This shows that it is possible to selectively quantify Triton X-100 content during the chromatographic load phase.

6.4   Conclusion and Outlook

FTIR spectroscopy was successfully implemented in-line as a PAT tool for biopharmaceutical purification processes. It was demonstrated that FTIR is able to distinguish and selectively quantify proteins in-line based on their secondary structure. Furthermore, FTIR presents a powerful tool for mon-

itoring different chemical components such as PEG or Triton X-100. Based on selective in-line quantification of PEG and protein, PEGylation degrees could be measured in-line. Selective mass balancing was performed on the process-related contaminant Triton X-100. In summary, FTIR provides orthogonal information to the typically measured UV/Vis spectra. It therefore is potentially interesting for monitoring process attributes which have been previously hidden. FTIR may help to achieve a more complete implementation of the PAT initiative.

Future research should be directed towards making the setup more compatible with the production environment. Challenges include the use of detectors without liquid nitrogen cooling and the application of fiber optics for in-line process probes.

Acknowledgment

7

Process Monitoring of Virus-Like Particle Reassembly by Diafiltration with UV/Vis Spectroscopy and Light Scattering

Matthias Rüdt[*,1], Philipp Vormittag[*,1], Nils Hillebrandt[1], Jürgen Hubbuch[1]

*   Contributed equally

1   Institute of Engineering in Life Sciences, Section IV: Biomolecular Separation Engineering, Karlsruhe Institute of Technology (KIT), Germany

Abstract

VLPs have shown great potential as biopharmaceuticals on the market and in clinics. Non-enveloped, in vivo-assembled VLPs are typically dis- and reassembled in vitro to improve particle stability, homogeneity, and immunogenicity. At industrial scale, Cross-Flow Filtration (CFF) is the method of choice for performing reassembly by diafiltration. Here, we developed an experimental CFF setup with on-line measurement loop for the implemen-

tation of PAT. The measurement loop included an UV/Vis spectrometer as well as a light-scattering photometer. These sensors allowed for monitoring protein concentration, protein tertiary structure, and protein quaternary structure. The experimental setup was tested with three Hepatitis B core Antigen (HBcAg) variants. With each variant, three reassembly processes were performed at different Transmembrane Pressures (TMPs). While light scattering provided information on the assembly progress, UV/Vis allowed for monitoring the protein concentration and the rate of VLP assembly based on the microenvironment of Tyrosine-132. Furthermore, the experimental results provided evidence of aggregate-related assembly inhibition and showed that off-line SEC does not provide a complete picture of the particle content. Finally, a PLS model was calibrated to predict VLP concentrations in the process solution. Q^2 values of 0.947 to 0.984 were reached for the three HBcAg variants. In summary, the proposed experimental setup provides a powerful platform for developing and monitoring VLP reassembly steps by CFF.

7.1   Introduction

VLPs are biopharmaceuticals with potential applications against various diseases such as viral and bacterial infections, cancer, Alzheimer's disease, and autoimmune disorders [123]–[127]. They are generally designed to trigger an immune response by presenting antigens on their surface. These antigens are either part of the native viral capsid or introduced artificially. Chimeric VLPs were, for example, constructed based on HBcAg [124], [128], [129], Hepatitis B surface Antigen (HBsAg) [130], GH1-Qβ [131], and Murine Polyomavirus VP1 (MuPyVP1) [123]. VLPs are resilient to most environmental stresses, have great potential to be produced inexpensively, and efficiently elicit potent immune responses due to their repetitive and particulate structure [132], [133].

Similar to viruses, VLPs are assemblies of one or several types of capsid proteins forming a higher-order structure [127]. VLPs are expressed in genetically modified host organisms [125], [127], [134]. Subsequent production-scale purification most frequently consists of precipitation, chromatography, and Ultrafiltration/Diafiltration (UF/DF) [135]. In vivo self-assembled, non-enveloped VLPs are often disassembled and subsequently reassembled to remove impurities from within the capsid [136], [137]. Disassembling and reassembling also leads to increased structural homogeneity, improved overall stability, and enhanced antigenicity [138]–[140]. An overview of a typical VLP production process is given in Figure 7.1.
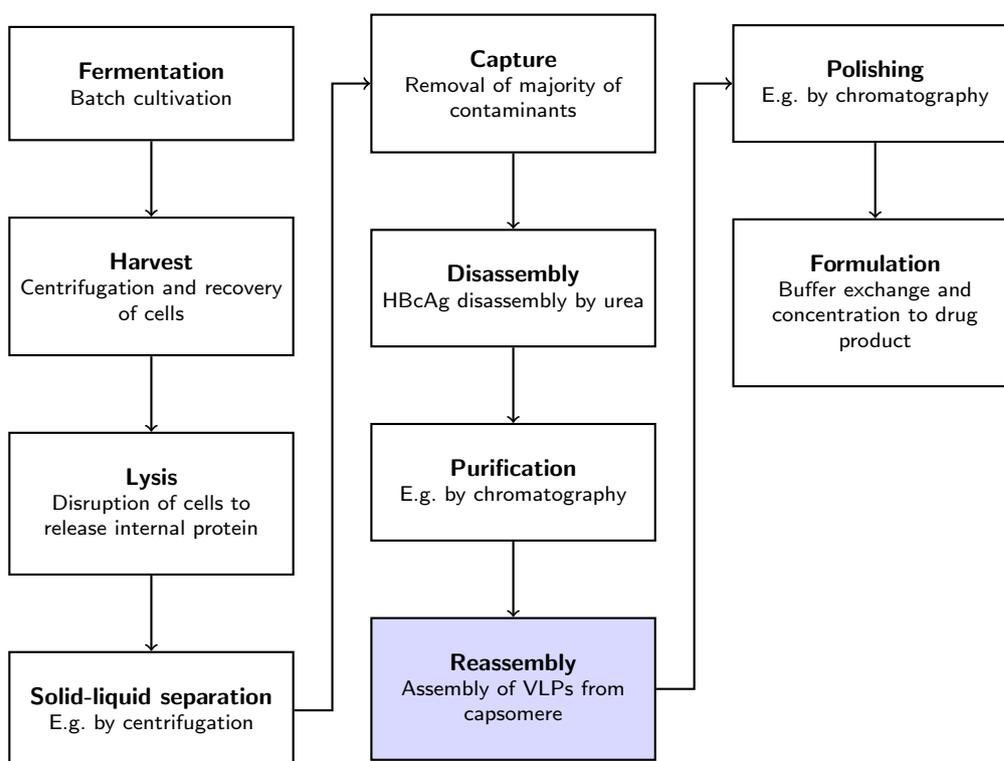
Figure 7.1: Illustration of a typical VLP production process. The downstream processing train may consist of eight or more unit operations. The unit operation investigated here—the VLP reassembly—is marked in blue.

Generally, a change in the quaternary structure of VLPs is induced by altering their physicochemical environment, i.e. the ionic strength of the protein solution, the pH, or the concentration of a reducing agent [139]. At lab scale, dialysis is the most common method for buffer exchanges [138]. Dialysis has, however, some drawbacks such as long processing times and significant buffer consumption [141]. In preparative downstream processes, CFF is more popular because of its simple scalability, a reduced buffer consumption, and reduced processing time [141], [142]. CFF has been successfully applied to VLPs for capture, buffer exchange, and concentration [134], [143], [144]. Compared to dialysis and batch diafiltration, assembly of VLPs by constant volume diafiltration was shown to increase VLP yield [145]. Despite the many advantages, CFF may also cause problems due to protein-membrane interaction [146], [147] which was observed to impact process performance [148]. To reduce these problems, CFF process time has to be minimized while maximizing the process efficiency.

PAT [18], [74], [149] is thus of interest to monitoring the assembly progress. Protein concentration measurements allow to detect protein adsorption to the membrane. Particle size measurements provide information on the assembly progress of the capsid proteins into VLPs. Previous publications have also reported effects of the VLP tertiary structure on UV/Vis and fluorescence absorption spectra [150]–[154]. Following a systematic approach to process monitoring, a combination of PAT sensors should be chosen which allows to monitor protein concentration, protein tertiary structure, and protein size.

In this study, we developed a CFF setup consisting of a commercial lab-scale CFF device with a custom-made on-line measurement loop for process analytical instrumentation. The on-line measurement loop included a light-scattering photometer (DLS and SLS) and a UV/Vis absorption spectrometer. DLS allowed for monitoring the mean hydrodynamic diameter of particles. SLS outputs an aggregated scattered-light intensity influenced by the particle concentrations and the diameters. Finally, UV/Vis spectroscopy provided information on the protein concentration and on changes in the tertiary structure by second derivative spectroscopy [32]. The usefulness of the custom-made setup was tested for monitoring the reassembly of three different chimeric HBcAg variants at three different TMPs.

7.2 Materials and Methods

7.2.1 Experimental Setup

A custom-made setup was developed for the CFF experiments. Figure 7.2 shows the setup as a Piping and Instrumentation Diagram (P&ID). A Kros-Flo KRIIi CFF unit with a modified Polyethersulfone (mPES) hollow fiber membrane module (10 kDa cutoff, 13 cm^2 membrane area) and a 50 mL conical tube retentate reservoir (all Spectrum Labs, Rancho Dominguez, USA) made up the core of the CFF unit. A Topolino magnetic stirrer (IKA Werke, Staufen im Breisgau, DE) ensured homogeneous mixing of the retentate reservoir. A T-piece with injection plug (Fresenius Kabi, Bad Homburg, DE) was inserted into the retentate line as sample port to draw liquid for off-line analytics. The retentate reservoir was modified with two additional PEEK capillaries to supply the on-line measurement loop with liquid from the process.

In the direction of flow, the on-line measurement loop consisted of a Gilson Minipuls 3 peristaltic pump, a 0.7 µm particle retention Minisart glass fiber syringe filter (Sartorius Stedim Biotech, Göttingen, DE), a Ze-
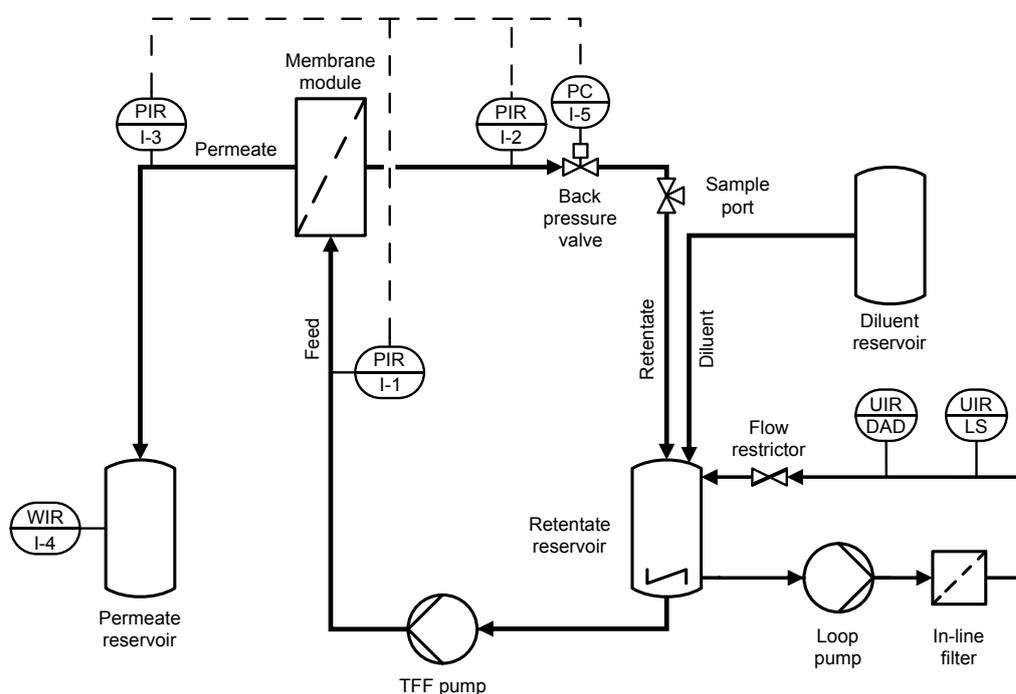
Figure 7.2: Piping and instrumentation diagram of the experimental setup. At the bottom right, the online measurement loop is shown. The remaining piping is required for the CFF. All sensors are connected to a computer for capturing the data centrally. Electronic communication lines are indicated by dashed lines. I-5 is a pinch valve actuated by a closed-loop controller for the TMP. The letters indicate: C control, I indicate, P pressure, R record, U multivariable, W weight, DAD diode array detector, LS light scattering.

tasizer Nano ZSP photometer (Malvern Instruments, Malvern, GB) with a 10 mm pathlength flowcell (Hellma Analytics, Müllheim, DE), an Ultimate DAD-3000 DAD (Dionex Corporation, Sunnyvale, US) with a 0.4 mm pathlength flowcell, and a FR-902 flow restrictor (GE Healthcare, Chalfont St Giles, GB). The pump of the on-line measurement loop was controlled via a NI USB-6008 data acquisition device (National Instruments, Austin, USA).

7.2.2 Proteins, Chemicals, and Buffers

Three chimeric HBcAg constructs , i.e. VLP A, B, and C provided by BioNTech Protein Therapeutics GmbH (Mainz, DE), were used in this study. The HBcAg variants were recombinantly modified in the Major Im-

munodominant Region (MIR) to display three different peptides on their surfaces (see also Figure 7.3).
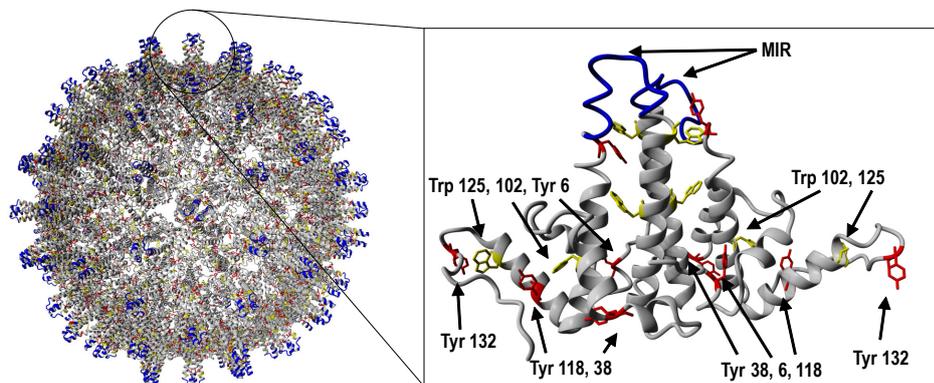
Figure 7.3: An assembled HBcAg VLP is shown on the left side (PDB ID 1QGT, [155]). The right side shows a cartoon of a single homodimer (adapted from PDB ID 4BMG, [156]). The tryptophan (Trp) and tyrosine (Tyr) side chains are depicted as sticks and colored in yellow and red, respectively. Tyrosine and tryptophan side chains located in the base of the molecule are numbered. These residues undergo a change of hydrophobicity in their environment during assembly. The MIR loop, whereto the foreign epitope is inserted, is shown in blue.

All variants were present as homodimer stock solutions in disassembly buffer (3.5 M urea, 50 mM Tris(hydroxymethyl)-aminomethane, pH 9.0) as obtained after purification (see also Figure 7.1). Protein concentration calculations were based on extinction coefficients derived from the primary structure as provided by the ProtParam tool [157] of the Swiss Institute of Bioinformatics. Immediately before each experiment, stock solutions were diluted with disassembly buffer to a protein concentration of 1 g/L (by UV absorbance at 280 nm) and filtered through a 0.2 µm Polyethersulfone (PES) filter (VWR International, Radnor, US). The reassembly buffer was a high-salt buffer at pH 7.0.

For SEC, 50 mM potassium phosphate at pH 7.0 was used as running buffer. If not mentioned otherwise, chemicals were purchased from Merck KGaA (Darmstadt, DE). All buffers and solutions were prepared with ultrapure water (arium pro UV, Sartorius, Göttingen, DE) and filtered through a 0.2 µm pore size Supor filter (Pall, Port Washington, US) immediately before each experiment.

7.2.3 VLP Reassembly Monitoring

The CFF unit and the measurement loop were filled with ultrapure water for pre-experimental preparation. The lamps of the DAD were turned on at least 1 h before starting the experiments. At the end of the equilibration phase, the absorbance signal was zeroed in ultrapure water. Subsequently, the CFF unit and measurement loop were first flushed with disassembly buffer and then changed into 25 mL of protein solution. The CFF pump was set to 70 mL/ min corresponding to a shear rate of approximately $6000\,s^{-1}$ in the hollow fibers. The measurement loop pump (1 mL/ min) and data acquisition were started.

After 5 min, constant TMP diafiltration was initiated by applying a TMP of 0.25 bar, 0.5 bar, or 1 bar with reassembly buffer as diluent. 0.4 mL samples were taken every 0.5 Diafiltration Volumes (DVs) via the sample port. Experiments were stopped after 3 DV except for VLP C for which the runs had to be terminated early due to membrane clogging. After each run, the CFF membrane was cleaned with ultrapure water, a 0.1 M sodium hydroxide solution, and a 15 vol% ethanol solution.

For SEC analysis, samples were centrifuged (Centrifuge 5810R, Eppendorf, Hamburg, DE) at 3220 rcf for 5 min to settle large particles. The supernatant was analyzed with a Sepax SRT SEC-1000 column (Sepax Technologies, Newark, US) on an Ultimate 3000 RS UHPLC system consisting of a Pump HPG-3400RS, an Autosampler WPS-3000TFC, a Column Compartment TCC-3000RS, and a Diode Array Detector DAD-3000 controlled by Chromeleon version 6.8 SR15 (all Thermo Fisher Scientific, Waltham, US). The run duration was 7 min with a flow rate of 0.8 mL/ min and SEC buffer as a mobile phase. 20 µL were injected for each analysis. Samples were analyzed in triplicates.

Off-line DLS analysis was performed using a sample volume of $45\,\mu L$ in a 3x3 mm quartz cuvette (Hellma Analytics, Müllheim, DE) and the same DLS spectrometer as mentioned above. Unfiltered samples were measured three times, each measurement consisting of 12 to 14 10 s runs at 25 °C, and 173° backscatter. Lower and upper limits for data processing were 1 nm and 6000 nm, respectively. The measurements were compared based on the VLP peak diameter in the regularization fit.

The spectrometer was also used for electrophoretic mobility measurements of pooled and formulated samples of each construct. The samples of different TMPs were pooled and dialyzed into a $pH\,7.2$ buffer of 50 mM Tris and 100 mM sodium chloride. Samples were filtered with a 0.2 µL polyethersulfone (PES) filter (VWR International, Radnor, USA) and concentration was adjusted with Vivaspin 20 filters with a 30 kDa pore rating (Sartorius,

Göttingen, DE). 50 µL of sample was inserted into buffer-filled folded disposable capillary cells (DTS1070, Malvern Instruments Ltd., Malvern, UK) using a diffusion barrier technique (Patent WO2012083272A1). Samples were measured in pentaplicates in automatic mode. Each measurement comprised a 120 seconds equilibration and five runs with up to 15 sub runs. The measurements were performed at 60 mV and 25 °C. Zeta potential was calculated by Zetasizer Software version 7.12 (Malvern Instruments Ltd., Malvern, UK) assuming a material refractive index of 1.45, absorption of 0.001, a viscosity of 0.8872 mPa s, a dielectric constant of 78.54, and a Smoluchowski approximation of 1.5 [158].

7.2.4 Data Acquisition and Analysis

During experiments, all integrated sensors communicated with a custom application developed in Matlab (version R2016b, The Mathworks, Natick, US). Next to starting and stopping measurements, the application gathered the sensor signals (3 pressure signals, the permeate weight, z-average, and UV/Vis absorbance spectra). Communication and control were performed through software libraries provided by the different instrument softwares. The signals were displayed on the Graphical User Interface (GUI) and stored on the hard drive with a time stamp. For calculating the permeate volume, the density of the permeate was assumed to be $1\,\text{g/cm}^3$. Data acquisition and analysis of light scattering and UV/Vis measurements were performed as described below.

Light-scattering Measurements

The Zetasizer Nano ZSP was utilized for DLS and SLS measurements using the chromatography flow Standard Operating Procedure (SOP) of the Zetasizer software (version 7.12, Malvern Instruments). The Zetasizer acquires data in a back-scattering geometry at 173°. Each measurement duration was 10 s. While DLS measurements were exported on-line, SLS data was extracted off-line. From the DLS measurement, the z-average was obtained as calculated by the Zetasizer software by the method of cumulants [159]. Viscosity (0.8872 mPa s), refractive indices (protein 1.45; water 1.33) (as provided by the Zetasizer software), temperature (25 °C), and flow rate (1 mL/ min) were assumed to be constant for the calculation of the z-average. The z-average data was subsequently filtered by a moving median over 60 s to remove outliers. The SLS signal was not filtered. The transition from process phase I to process phase II was detected from the scattered-light intensity by the CUSUM algorithm [160], [161]. The transition from

process phase II to process phase III was set at the global maximum of the scattered-light intensity.

UV/Vis Absorption Measurements and Processing

During VLP assembly, UV/Vis spectra were continuously acquired at 1 Hz in the spectral range from 240 nm to 340 nm with a resolution of 1 nm. To gain information on the local environment of aromatic amino acids, the spectral data was filtered by a moving average over 30 s and the second derivatives were computed with a Savitzky-Golay filter [82] of order 5 with a 9-point window [32], [151]. An example spectrum with the subsequent data evaluation is shown in the Supplementary Material 7.6. The resulting second-derivative spectra were interpolated with a cubic spline to a final resolution of 0.01 nm. From the interpolated data, the location of the minimum near 292 nm was used as a measure of tryptophan solvent exposure [32], [162]. The exposure of tyrosine was assessed based on the a/b-ratio as defined by Ragone et al. [101]. Briefly, the vertical distance between trough and peak near 285 nm a was normalized by the trough-peak distance near 294 nm b. The inflection point of the a/b-ratio over time was computed by taking the first derivative with a second-order Savitzky-Golay filter (window width 501 points corresponding to 8.35 min) and finding the minimum.

PLS Model Calibration

PLS model calibration was performed in Matlab (version 2016a). For each VLP, a PLS model was calibrated based on the UV/Vis spectroscopic data in combination with the off-line SEC VLP concentration. Data of all three TMPs were included into one model. PLS model calibration was performed similarly as described previously [163]. The data were first preprocessed and subsequently fitted with a PLS-1 model by the SIMPLS algorithm [99]. For preprocessing, a Savitzky-Golay filter with a second-order polynomial was applied on the spectra and, optionally, the first or second derivative was taken. Cross-validation was performed by iteratively excluding one sample of each CFF run ($\frac{1}{7}$, resp. $\frac{1}{6}$ of the data), calibrating a PLS model on the remaining samples ($\frac{6}{7}$, resp. $\frac{5}{6}$ of the data), and calculating a residual sum of squares on the excluded run. This procedure was repeated until all runs had been excluded once. All residual sums of squares for the different submodels were subsequently accumulated yielding the PRESS. The PRESS was scaled according to Wold et al. by the number of samples and latent variables used in the PLS model [30]. Based on the scaled PRESS, an optimization was performed using the built-in genetic algorithm of Matlab for integers [100].

The genetic algorithm optimized the window width of the Savitzky-Golay filter $5 \leq w \leq 21$, the order of derivative $0 \leq n \leq 2$, as well as the number of latent variables for the PLS-1 model $4 \leq N \leq 14$. The RMSECV was calculated from the PRESS by dividing by the total number of samples. The Q^2 and R^2 values were calculated by dividing the PRESS, respectively the residual sum of squares, by the summed squares of the response corrected to the mean [30].

7.3 Results

In this study, a new UF/DF setup with on-line measurement loop was developed to monitor VLP reassembly steps. In the measurement loop, a UV/Vis spectrometer and a light-scattering photometer were integrated. Furthermore, an application was implemented in Matlab providing a GUI, communication capabilities to the different sensors, as well as a common time base for all performed measurements. This allowed for acquiring and synchronizing measurements in a controlled manner. Within the application, UV/Vis spectra, DLS measurements, pressure, and weight readings were immediately available for processing and display. To demonstrate the advantages of this experimental setup, nine UF/DF runs with three different HBcAg constructs at three different TMPs were performed.

7.3.1 Monitoring of Standard Processes Parameters

During the UF/DF processes, the initial buffer was replaced by reassembly buffer to form HBcAg VLPs from homodimers. In Table 7.1, process data of all runs are summarized (original data plotted in Supplementary Material 7.8). The tables also show that the feed stock purity of VLP A is higher than VLP C and VLP B. At 0.25 bar TMP, VLP A, B, and C showed nearly constant increases in permeate mass over time implicating constant fluxes. The average flux for these three runs was 25.8 L/m²h to 29.1 L/m²h. At 0.5 bar and 1 bar TMP, the average flux was higher for all three VLPs (from 36.3 L/m²h to 48.7 L/m²h). CFF processes at 0.5 bar and 1 bar TMP showed a decreasing flux over time after an initial constant phase (except for VLP B at 0.5 bar). A decrease in flux at constant TMP indicates the formation of a fouling layer on the membrane [164], [165].

Table 7.1: Process data is summarized for all performed runs.

	VLP A		
TMP / bar	0.25	0.5	1
Feed stock purity[a] / %		73.5	
Zeta potential[b] / mV		$-7.9(7)$	
Total run time / min	118	78	75
Mean flux / $(\mathrm{L\,m^{-2}\,h^{-1}})$	30.5	46.9	48.4
Max. VLP conc. / (g/L)	0.248	0.275	0.250
Inflection a/b-ratio / DV	1.5	0.8	0.7
VLP peak diameter[c] / nm	40(6)	46(11)	42(7)

	VLP B		
TMP / bar	0.25	0.5	1
Feed stock purity[a] / %		22.6	
Zeta potential[b] / mV		$-11.8(6)$	
Total run time / min	133	75	79
Mean flux / $(\mathrm{L\,m^{-2}\,h^{-1}})$	26.8	48.7	45.9
Max. VLP conc. / (g/L)	0.126	0.133	0.116
Inflection a/b-ratio / DV	1.5	1.4	0.7
VLP peak diameter[c] / nm	35(5)	40(11)	46(10)

	VLP C		
TMP / bar	0.25	0.5	1
Feed stock purity[a] / %		44.1	
Zeta potential[b] / mV		$-9.5(8)$	
Total run time / min	108	71	70
Mean flux / $(\mathrm{L\,m^{-2}\,h^{-1}})$	27.6	36.3	40.0
Max. VLP conc. / (g/L)	0.134	0.103	0.126
Inflection a/b-ratio / DV	1.6	0.9	0.6
VLP peak diameter[c] / nm	41(12)	48(5)	36(11)

[a] assessed by reversed-phase chromatography as described in the Supplementary Material 7.7.
[b] denotes median and median absolute deviation in parenthesis.
[c] denotes mean and standard deviation of all DLS acquisitions ($n = 36 - 42$) in parenthesis.

7.3.2 Process Monitoring with On-line PAT Sensors

In Figures 7.4, 7.5, and 7.6, the on-line PAT sensor measurements as well as off-line analytics are shown for VLP A, B, and C, respectively. All data were plotted over DV indicating the progress of buffer exchange. Each figure shows the absorbance at 280 nm, off-line VLP concentration measurements by SEC, second-derivative spectral analysis, and light-scattering data. It is important to note that an insufficient scattered-light intensity was recorded for VLP C at 1 bar TMP due to an incorrectly set laser attenuation. The corresponding light-scattering results were excluded. The run could not be repeated because of material constraints.
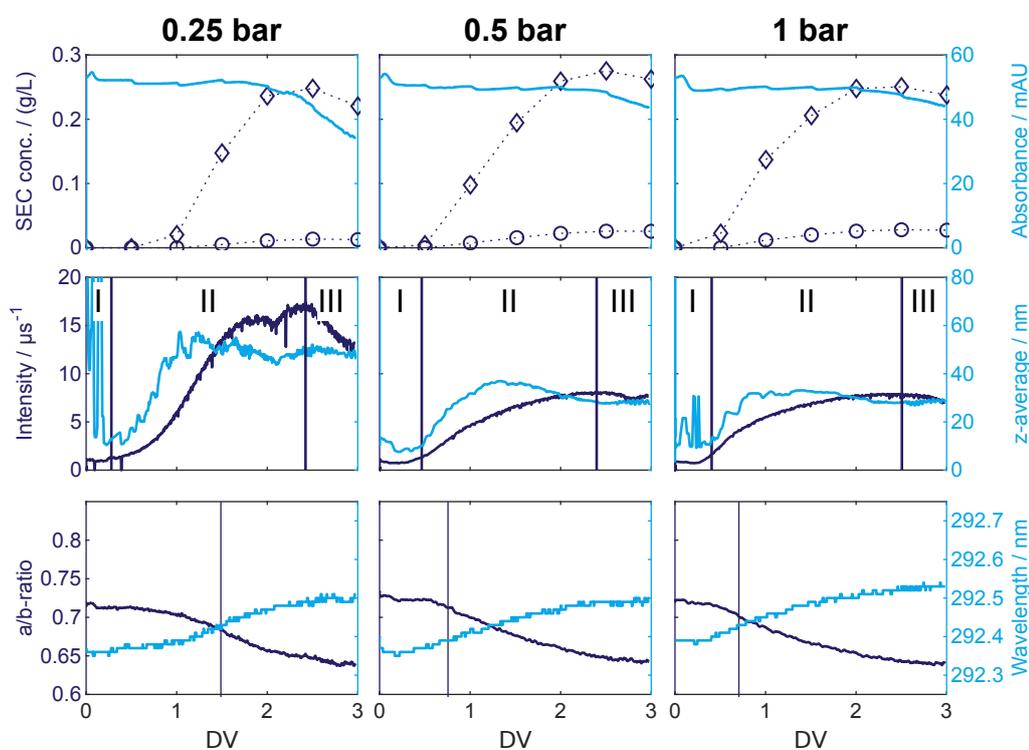

Figure 7.4: The figure displays the on-line sensor measurements as well as off-line analytics against the DV of VLP A. The rows display measurements of different sensors. Top row: Off-line VLP ◇ and aggregate ○ concentration measurements by SEC, UV absorbance at 280 nm −. Middle row: DLS and SLS measurements. Roman numbers indicate the different process phases. Bottom row: Second-derivative spectral analysis for tyrosine (a/b-ratio) and tryptophan (location of the minimum around 292 nm). The inflection point of the a/b-ratio is marked by a vertical bar. The columns correspond to different TMPs. Left column: 0.25 bar, middle column: 0.5 bar, right column: 1 bar. At 0.25 bar TMP the z-average is corrupted with noise early in the process.
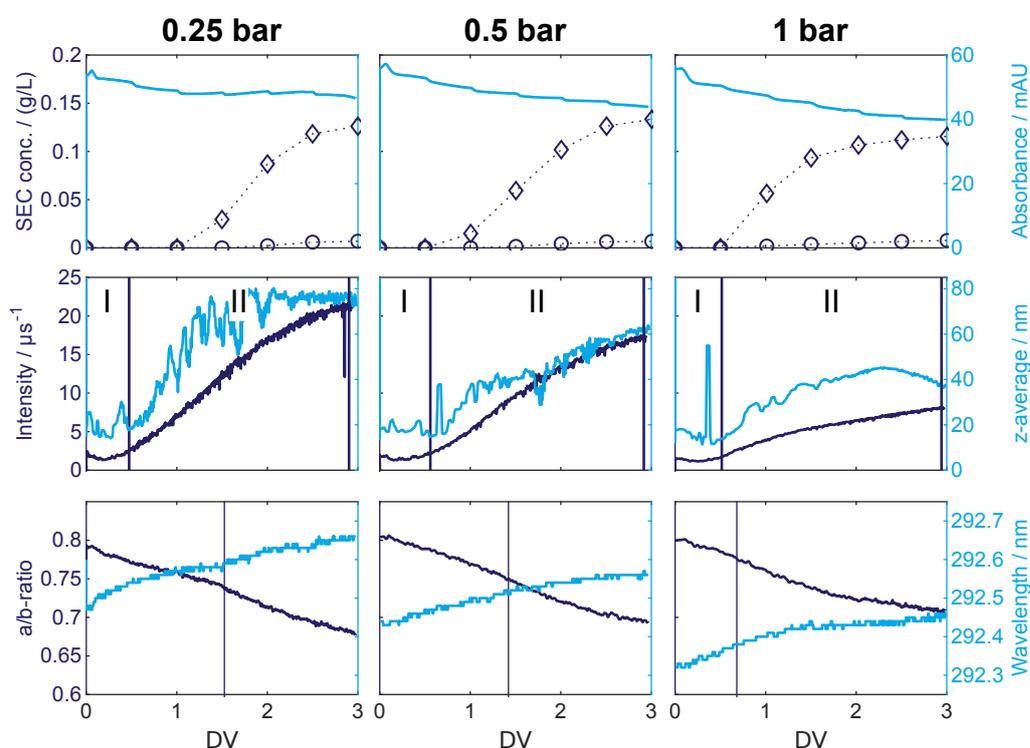
Figure 7.5: The figure displays the on-line sensor measurements as well as off-line analytics against the DV of VLP B. The rows display measurements of different sensors. Top row: Off-line VLP $\diamond$ and aggregate $\circ$ concentration measurements by SEC, UV absorbance at 280 nm $-$. Middle row: DLS and SLS measurements. Roman numbers indicate process phases. Bottom row: Second-derivative spectral analysis for tyrosine (a/b-ratio) and tryptophan (location of the minimum around 292 nm). The inflection point of the a/b-ratio is marked by a vertical bar. The columns correspond to different TMPs. Left column: 0.25 bar, middle column: 0.5 bar, right column: 1 bar.

Off-line SEC was performed in triplicates resulting in standard deviations smaller than 0.011 g/L. In all runs, the off-line VLP concentration first remained at zero followed by an increase to the maximum VLP concentration. Thereafter, the concentration was approximately constant or decreased slightly. Depending on the TMP, off-line VLP concentration started to increase at 0.5 DV to 1.5 DV. The onset occurred at a DV that was lower the higher the TMP. The maximum observed VLP concentration was between 0.248 g/L and 0.275 g/L for VLP A, between 0.116 g/L and 0.133 g/L for VLP B, and between 0.103 g/L and 0.134 g/L for VLP C. The SEC aggregate content was between 5 % to 15 % of the VLP concentration.

UV absorbance at 280 nm decreased in all runs over time. Small step-like decreases were due to sampling for off-line analytics. The drawn sample
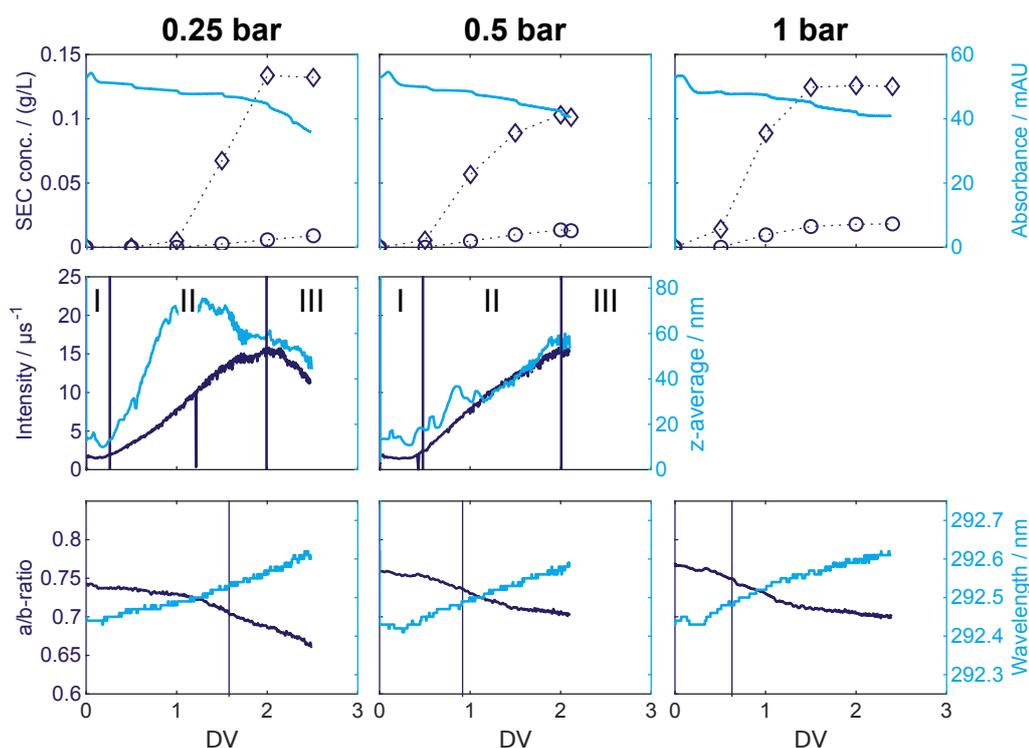
Figure 7.6: The figure displays the on-line sensor measurements as well as off-line analytics against the DV of VLP C. The rows display measurements of different sensors. Top row: Off-line VLP ◇ and aggregate ○ concentration measurements by SEC, UV absorbance at 280 nm –. Middle row: DLS and SLS measurements. Roman numbers indicate process phases. Bottom row: Second-derivative spectral analysis for tyrosine (a/b-ratio) and tryptophan (location of the minimum around 292 nm). The inflection point of the a/b-ratio is marked by a vertical bar. The columns correspond to different TMPs. Left column: 0.25 bar, middle column: 0.5 bar, right column: 1 bar. DLS and SLS measurements at 1 bar were excluded because of an erratically set laser attenuator.

volume was replaced by reassembly buffer resulting in dilution of the process liquid. For VLP A, B, and C, a rapid decrease in the absorbance at 0.25 bar TMP occurred towards the end of the runs, suggesting a loss of protein.

Solvatization of aromatic amino acids and particle formation were observed during CFF by on-line UV/Vis and light-scattering measurements. UV/Vis spectral data were examined by second derivative analysis. From the derived spectra, characteristics were calculated for the solvatization of tryptophan (location of the minimum around 292.5 nm) and tyrosine (a/b-ratio) [32]. For all runs, a shift towards longer wavelengths of the tryptophan minimum was observed, while the a/b-ratio decreased. Both trends indicated an increase in the mean hydrophobicity around tryptophans and tyrosines. Especially for higher TMPs, the characteristics followed a sigmoidal curve shape. The inflection points of the a/b-ratio in all runs were marked by a vertical line and were located either around 0.8 DV or 1.5 DV (see Table 7.1).

DLS measurements were interpreted based on the z-average. In all experiments, an initial phase of relatively constant z-average values below 20 nm was observed. The second phase was characterized by a rapid increase in z-average to around 40 nm for TMPs of 0.5 bar and 1 bar. At a TMP of 0.25 bar, the second phase showed a larger increase of the z-average to 50 nm to 80 nm. The third phase resulted in relatively constant z-averages over time.

SLS measurements are influenced by the particle diameter and concentration. Similar to the z-average, scattered-light intensities started to increase after an initial constant phase. The increase continued even after the z-average reached a plateau and eventually flattened. For VLP A and C at 0.25 bar TMP, scattered-light intensities rapidly decreased towards the end of the runs.

At 0.5 bar and 1 bar, z-averages, scattering intensities, and SEC VLP concentrations of each run started to increase simultaneously within off-line time resolution. Interestingly, for processes at 0.25 bar, the z-averages and scattering intensities increased earlier than VLP and aggregate concentration by SEC. The initial increase in phase two at 0.25 bar ended at high z-averages > 45 nm, not observed in the other processes. In all runs, the inflection point of the a/b-ratio occurred around the steepest increase in the VLP concentration by SEC.

7.3.3 Selective Prediction of VLP Concentration by PLS Modeling

The PLS model calibration results are shown in Figure 7.7 and Table 7.2. Figure 7.8 shows the PLS regression coefficients. All PLS models were fitted to the second derivative of the UV/Vis spectral data with 6 to 9 latent variables. The achieved Q^2 values were 0.984, 0.984, and 0.947 for VLP A, B, and C, respectively.
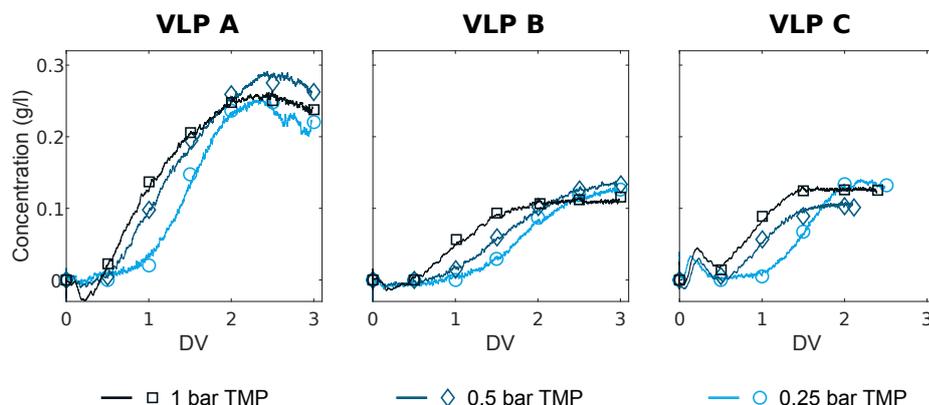

Figure 7.7: A PLS model was fitted to the UV/Vis spectral data for each construct to predict the concentration of assembled VLPs. The concentration predicted by the calibrated PLS model is compared to off-line analytics in the current plot. Each TMP is reflected by a color. The markers show the concentration measured by off-line analytics while the lines correspond to the concentrations predicted by the PLS model.

Table 7.2: Spectral preprocessing parameters, parameters for the PLS model, and the prediction quality of the chemometric models are summarized.

|	VLP A	VLP B	VLP C
No. of samples	21	21	18
No. of cross-validation groups	7	7	6
No. of latent variables	6	9	7
Window Savitzky-Golay filter	7	9	9
Derivative	2	2	2
R^2	0.995	0.997	0.994
Q^2	0.984	0.984	0.947
RMSECV / (g/L)	0.01	0.01	0.01
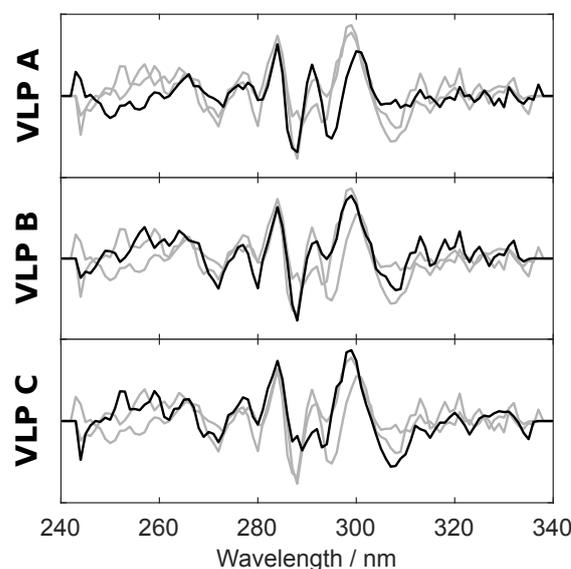
Figure 7.8: Regression coefficients of the three PLS models. Each row corresponds to the regression coefficients of one VLP in black while the other regression coefficients are supplemented in gray.

7.3.4 Analysis of Post-Assembly Samples

Off-line DLS data was measured at the end of all processes. The VLP peak diameter data is shown in Table 7.1. 36 to 42 acquisitions were used to calculate a mean diameter and standard deviation. The mean diameter across all runs was 41 nm nm with a standard deviation of 11 nm. Median zeta potentials and corresponding median absolute deviations were calculated from 25 runs of the 5 measurements for each VLP construct (compare Table 7.1). VLP B was had the most negative zeta potential with $-11.8(6)$ mV, followed by VLP C with $-9.5(8)$ mV, and VLP A with $-7.9(7)$ mV.

7.4 Discussion

7.4.1 On-line Measurement Setup

As shown in Figure 7.2, the experimental setup included a flow restrictor and a filter next to the sensors in the on-line measurement loop. The flow restrictor and filter were added to improve the measurement quality. The flow restrictor set a minimal back pressure in the measurement loop reducing pressure fluctuations and air bubbles. The filter (cut-off 0.7 µm) retained bubbles and large particles adversely affecting light-scattering mea-

surements. The light-scattering measurements depend strongly on the particle diameter d [166]. Thus, large particles, such as air bubbles or large aggregates, can completely obliterate the light scattering of smaller particles in SLS and DLS measurements.

7.4.2 Interpretation of SLS and DLS Measurements

During VLP reassembly, anticipated particles in the process solution were homodimers, VLPs, VLP aggregates, and process-related impurities, all of which contributed to light scattering. Thus, the scattered-light intensity is a sum signal generated by all scattering species. By neglecting any interaction between the particles and assuming Rayleigh scattering, the scattered-light intensity I_R can be described as [166]

$$I_R \propto \sum_i c_i d_i^6 \, , \tag{7.1}$$

where i iterates over all species, c_i is the molar concentration of species i, and d_i is the diameter of species i. Based on this formula, it can be verified that particle agglomeration and concentration leads to increased scattered-light intensities.

The z-average is the intensity-weighted harmonic mean hydrodynamic diameter [167]. Therefore, the z-average is not proportional to the concentration but reflects an apparent mean particle diameter. A small fraction of large particles can still significantly increase the z-average. During reassembly, an increase of scattered-light intensity and z-average was expected because of the formation of VLPs and aggregates.

7.4.3 DLS Measurements in Flow

DLS measures the time correlation of scattered-light intensity. In contrast to the typical DLS measurement setup, the time correlation in the on-line measurement loop was not only influenced by diffusion but also by convective flow [168]. It has been previously demonstrated that the convective flow results in increased estimated diffusion coefficients and thus in reduced particle diameters [169]. The effect was shown to be more pronounced for larger particles. Consequently, underestimation of particle sizes was expected to be more pronounced for aggregates than VLPs than homodimers. No effect on SLS was expected from convective flow.

7.4.4 General Considerations on the VLP Assembly Processes

During the diafiltration process, the disassembly buffer was gradually exchanged by an assembly buffer. The chemical environment of the HBcAg dimers increasingly favored assembly. This is different to the conventional approach in VLP kinetic studies where the composition of the assembly reaction liquid is usually adjusted by rapid dilution [170], [171]. In said studies, assembly equilibrium phases were reached in a few minutes. Given the comparably large time frame of diafiltration experiments (75 min to 135 min), we assume that the VLP concentration was almost exclusively dependent on the buffer composition.

Figure 7.9 illustrates the formation of particles out of HBcAg dimers during a diafiltration process and expected sensor responses. The diafiltration process was split into phases 1 to 3 based on different reactions occurring during each phase.

In phase 1, buffer exchange starts but no assembly occurs, i.e. the VLP concentration remains zero. However, aggregates may form resulting in an increase in scattered-light intensity and z-average, as seen in Figures 7.4, 7.5, and 7.6.

In phase 2, homodimers assemble into VLPs. Native HBcAg VLPs are 30 nm to 34 nm in diameter [172]. VLP concentration increases to its maximum, while the scattered-light intensity and z-average continue to rise. To explain the sensor response more comprehensively, phase 2 was subdivided into two subphases, 2a and 2b. In subphase 2a, z-average and scattered-light intensity both increase. In subphase 2b, scattered-light intensity further increases while z-average remains constant. The increase in scattered-light intensity is caused by the ongoing formation of VLPs and aggregates. Conversely, the z-average stagnates as it is an intensity-weighted harmonic mean. When the z-average is close to the size of VLPs, further assembly has only a small effect on the z-average, while the scattered-light intensity still increases due to the formation of particles.

In phase 3, the VLP concentration no longer increases. Thus, the end of the assembly process is reached. A loss of aggregates is reflected by a decrease in z-average and scattered-light intensity. A decrease in scattered-light intensity and UV absorbance with constant z-average reflects a decrease in overall protein concentration with constant particle size distribution.

Towards the end of some processes (most pronounced for VLP A and C at 0.25 bar), both light-scattering signals decreased combined with a decrease in the UV signal at 280 nm. Thus, the protein concentration de-
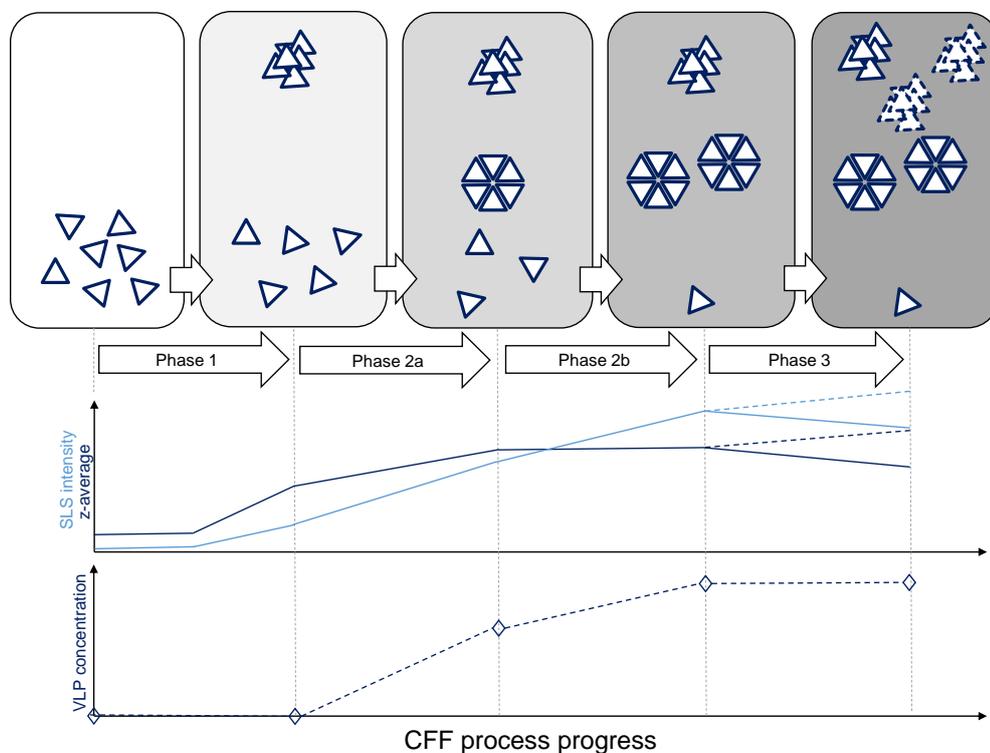
Figure 7.9: Theoretical consideration of particle formation during the assembly process by CFF. Homodimers, aggregates, and VLPs are shown as schematics. The expected development of SLS, z-average, and VLP concentration signals is shown over the CFF process progress subdivided into four phases. In the process, the buffer of a homodimer solution is gradually exchanged by assembly buffer to initiate VLP assembly. In phase 1, few aggregates are formed and no assembly takes place. The formation of aggregates increases the light-scattering signals while the VLP concentration remains at zero. As a consequence of exceeding a critical buffer composition, VLPs start to form in phase 2a, visualized by an increase in VLP concentration. The light-scattering signals continue to increase as a response to particle formation. In phase 2b, assembly continues, indicated by a further increase in VLP concentration and static light scattering. The z-average remains comparably constant as its value is already close to the actual VLP diameter and is thus only marginally influenced by further assembly. In phase 3, the assembly reaction is no longer proceeding. Particles are depleted resulting in a decrease in the light-scattering signals.

creased due to adsorption to the CFF membrane or retention on the measurement loop filter. The elevated salt concentration of the process liquid at this stage of the process may have promoted adsorption of protein to the hollow fiber membrane [146]. In both runs, the z-average started to decrease already earlier than the UV signal at 280 nm around the location of the inflection point of the a/b-ratio while the UV absorbance was still approximately constant. This could indicate a partial disintegration of aggregates. Phase 3 was generally short, as either its onset was close to the final DV or the process was stopped early due to membrane fouling.

The assembly of HBcAg VLPs also induces changes in mean hydrophobicity around aromatic amino acids as capsid assembly relies on hydrophobic interaction forces [155], [173]. Tyrosine-132 is especially important for the assembly [174]. In homodimers, tyrosine-132 is highly solvent-exposed, as shown in Figure 7.3. After VLP assembly, tyrosine-132 is buried in a hydrophobic pocket of the neighboring homodimer. During diafiltration, the solvatization of tyrosine changes because of aggregation as well as VLP assembly. If the mean effect on hydrophobicity by aggregation is small compared to the mean effect caused by assembly, the change over time of the a/b-ratio correlates to the rate of assembly. As a result, its inflection point marks the point of the highest rate of assembly. Similarly, the increase in the wavelength of the tryptophan absorption minimum marks an increase in hydrophobicity around tryptophans. Since the change in the solvent exposure of tryptophans during VLP assembly is less pronounced, the effect is weaker and more biased by aggregation.

7.4.5 CFF for VLP Assembly

VLP A was assembled from the purest dimer stock solution of the three investigated VLPs. The process was thus expected to perform comparably well. This agreed with the experimental results at 0.5 bar and 1 bar TMP. The observed z-averages of 28 nm to 29 nm in phase 3 showed that there was a significant fraction of VLPs. Few large particles were generated while other factors such as the flow reduced the z-average compared to off-line DLS analytics (see Table 7.1). The higher final z-average and an elevated scattered-light intensity at 0.25 bar TMP provided evidence of the formation of large aggregates. The observations made for VLP A were in general also applicable to VLP B and C. Both VLPs were adversely affected at lower TMPs by aggregation reflected by increased z-averages and light-scattering intensities.

A further interesting result of this study was the clustering of the inflection points of the a/b-ratio either around 1.5 DV or around 0.8 DV. An

early inflection point is consistent with early VLP formation. Conversely, a late inflection point correlated to an early increase in aggregates. By keeping in mind that the DV is indicative of the progress of buffer exchange, the conclusion may be drawn that VLP assembly is inhibited by aggregates. Indeed, a similar conclusion was previously proposed for MuPyVP1 VLPs [175]. Ding and coworkers described a competition of capsomere association into aggregates and precursors of MuPyVP1 VLPs.

The results of the diafiltration experiments for all VLPs showed that a low TMP of 0.25 bar lead to an increased aggregation propensity and an increased process time compared to the other conditions. At 0.5 bar and 1 bar TMP, the process time, VLP concentration, and aggregate content depended on the VLP construct and stock purity but were not solely dependent on the TMP. For increased yield and decreased aggregate content, it could be helpful to introduce a further purification step for VLP B and C. In all runs, aggregate concentration by SEC did not reflect the data obtained by light scattering. The reason for this seemed to be that large aggregates were depleted during sample preparation or in the SEC column. As a consequence, light scattering provided a more complete picture of the aggregate content.

Process phase III is characterized by product loss. The process should therefore be terminated at the end of phase II. It is worth noting that the end of phase II is influenced by the VLP construct but seems to be independent of the applied TMP. No plateau or decrease in assembly was observed for VLP B. VLP B was charged strongest, requiring higher ionic strengths to overcome the electrostatic charges of the homodimers during assembly (see Table 7.1). Zeta potentials of VLP A and C are similar. For both, a transition into phase III was observed.

To compare the assembled VLPs with standard characterization methods, we performed DLS measurements on the assembled VLPs. Off-line DLS VLP peak diameters are comparable to that of wild type HBcAg VLPs (typically 30-34 nm [172]) with a mean of 41 nm and a standard deviation of 11 nm. No significant influence of the TMP or construct on the final VLP peak diameter could be observed.

In summary, the analytical measurements of the VLP size and structure confirm the information obtained from the PAT tools.

110

7.4.6 Benefits of Using PAT for Process Development and Production

PAT is currently a frequently investigated approach of increasing the acquired information about unit operations in biopharmaceutical process development and production by timely measurements. Generating information on the process in (near) real time potentially results in a better understanding, faster optimization, and reduced off-line analytical samples [18].

Here, the UV absorbance at 280 nm provided insight into changes in the concentration of protein and other absorbing species in real time. This can be of advantage for assessing the membrane performance (e.g. membrane fouling, pore rating out-of-specification, or membrane damage). A mechanistic understanding is, however, often not possible solely based on a single wavelength. A more in-depth view on the on-going processes during UF/DF could be realized based on the acquired UV/Vis spectra. For HBcAg, tyrosine-132 is especially important for the VLP assembly. The a/b-ratio provides a mechanistic insight into the assembly reaction based on the mean tyrosine solvatization. Furthermore, other UV/Vis chromophores are phenylalanine, tryptophan, and disulfide bridges [32]. These may be affected during the assembly of other VLPs. For example, during the assembly of human papilloma virus-like particles, disulfide bridges are the key to the formation of higher-order structures [176]. Next to means for quantification, the UV/Vis spectrometer implemented in the presented setup thus provides mechanistic process understanding.

Another changing protein attribute, which can be monitored, is the particle size. The significant increase in size has a large impact on the scattering characteristics of the process fluid. The light-scattering spectrometer thus allowed for the detection of the start of the assembly reaction and maximal VLP concentration. Light-scattering spectrometers are universal detectors that are not dependent on the protein primary structure. As a consequence, any VLP assembly reaction can be monitored with this technique. In development and production, light-scattering detectors provide the means for detecting the ideal point to stop CFF or to initiate the next process step. This can improve the product quality (as process phase III is omitted) and allow for process intensification.

Generally, the on-line sensors provide data with high temporal resolution which typically is difficult to achieve with off-line analytics. In consequence, smaller changes in process characteristics (e.g. assembly onset, end of phase II) can be detected. This may be helpful for the further assessment of

111

different processes in development or for detecting deviations or hidden trends in production.

For process monitoring in production, it may be beneficial to retrieve VLP concentrations in real time. A PLS model was thus developed to demonstrate the possibility to monitor VLP concentration on-line by UV/Vis spectroscopy. The model was optimized by a constrained heuristic search algorithm. The minimal number of four latent variables was set to reflect the minimal amount of independent UV-active species (VLP concentration, DNA concentration, urea concentration, and aggregates). Reliable VLP concentration predictions were possible for all three constructs. In production, UV/Vis measurements in conjunction with a PLS model could thus be used for the real-time assessment of the assembly progress and ultimately for process control. Based on the regression coefficients of the PLS model (Figure 7.8), it is clearly visible that the fine structure of the tyrosine and tryptophan absorption is of major importance for the regression. Therefore, the PLS model accesses information similar to that provided by the a/b-ratio and the tryptophan minimum. The differences between the regression coefficients for VLP A, B, and C were attributed to the changing purity of the stock solutions. Provided that no additional chromophores are introduced into the MIR, a universally applicable PLS model for different HBcAg constructs is conceivable. This may be evaluated further in future studies.

7.5   Conclusion and Outlook

In this study, we investigated HBcAg assembly by diafiltration of three different constructs at three different TMPs. We developed an on-line measurement setup consisting of a UV/Vis and a light-scattering sensor (DLS and SLS) with a unified software platform. This setup allowed for monitoring mean particle sizes, hydrophobicity around tyrosine and tryptophan as well as UV/Vis spectra. Based on the UV/Vis spectra, we calibrated three PLS models for predicting VLP concentrations in real-time. Regarding process performance, we observed that processes with hollow fiber modules at 0.25 bar TMP resulted in increased aggregation. In all processes, the maximum rate of assembly occurred around two characteristic DV. This behavior was interpreted as a result of aggregation-related inhibition of VLP assembly, which makes it especially important to prevent aggregation in a VLP assembly process. In summary, the established setup has shown great potential for improving process monitoring, development, and understanding during VLP assembly by diafiltration.

In the future, strategies may have to be developed for process control during VLP reassembly. The proposed setup allowed for monitoring central quality attributes during the process with and without calibrated chemometric models. It is therefore a good starting point for any further research in this direction. From a process development point of view, the current results have not yet shown a reduced process efficiency at the highest TMP. A further increase in TMP may thus be attractive. Alternative membrane options, such as membrane cassettes, could strongly affect the process and may be interesting to evaluate with the setup.

Acknowledgment

7.6 Supplementary Material: Calculation of Local Hydrophobicity around Aromatic Amino Acids

The local hydrophobicity around tryptophan and was assessed by performing a second derivative on the UV/Vis spectra and interpolating the resulting data. An interpolated derivative spectrum is shown in Supplementary Figure 7.10. The spectrum is annotated with the tryptophan minimum, the a-value, and the b-value. The a and b values are used for calculating the a/b-ratio by dividing the former through the latter.

7.7 Supplementary Material: Reversed-phase Chromatography

The purity of the stock solutions was assessed by reverse-phase chromatography based on the absorbance of the eluting species at 280 nm. The stock solutions were analyzed with a Waters Acquity BEH300 C4 1.7 μm column
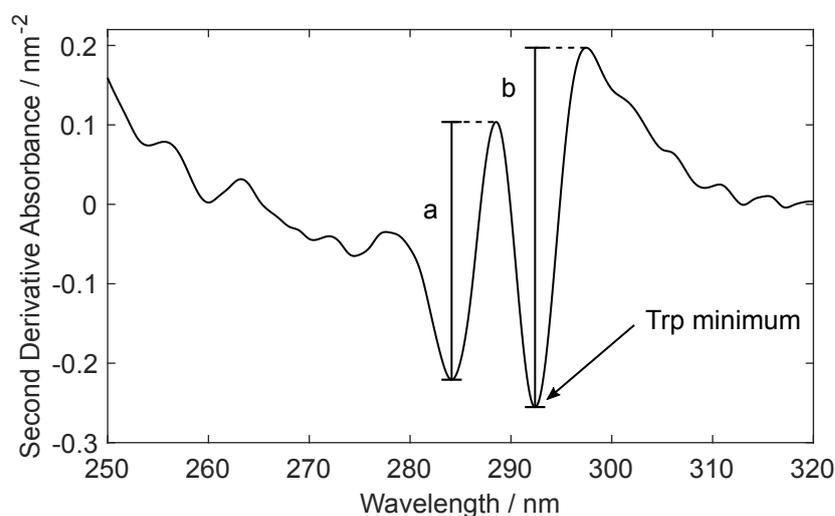
Figure 7.10: An interpolated second derivative spectrum of VLP A is shown. The tryptophan (Trp) minimimum, the a-value, and the b-value are marked.

(Waters Corporation, Milford, US) on an Ultimate 3000 RS UHPLC system consisting of a Pump HPG-3400RS, an Autosampler WPS-3000TFC, a Column Compartment TCC-3000RS, and a Diode Array Detector DAD-3000 controlled by Chromeleon version 6.8 SR15 (all Thermo Fisher Scientific, Waltham, US). The run duration was 6.8 min with a flow rate of 0.45 mL/min at a temperature of 80° with solvent A as 0.1 % TFA in water and solvent B as 0.1 % TFA in acetonitrile. Equilibration was done at 5 % B, and a gradient of 4.7 min was run from 23.5 % to 63.5 % B. The column was stripped with 95 % B for 0.5 min and then reequilibrated at 5 % B for 1.3 min. 2 µL were injected for each analysis. Samples were analyzed in triplicates. The purity of the stock solutions was calculated as the percentage of absorbance at 280 nm of the respective HBcAg construct of the total absorbance of all eluting species.

7.8 Supplementary Material: CFF Process Progress

For interested readers, the permeate mass over time of the different processes is shown in Supplementary Figure 7.11.
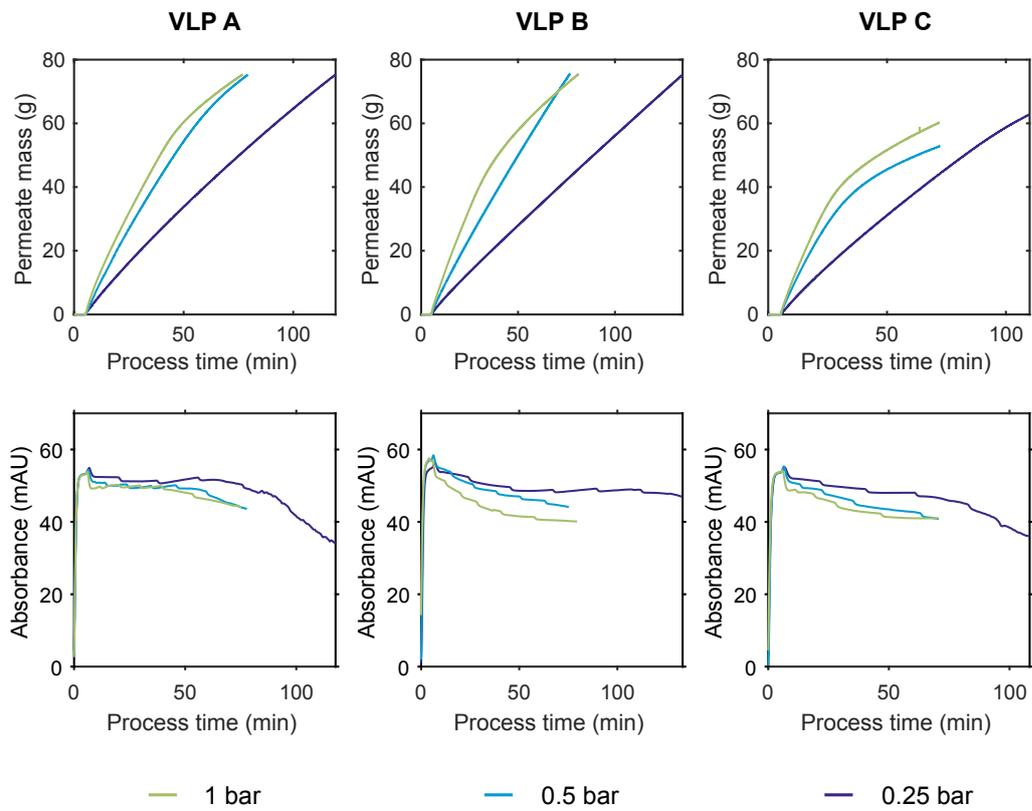
Figure 7.11: In the top row, the permeate mass over process time is shown. The bottom row shows the UV absorbance at 280 nm.

Factorization of Preparative Protein Chromatograms with Hard-Constraint Multivariate Curve Resolution and Second Derivative Pretreatment

Matthias Rüdt[1], Sebastian Andris[1], Robin Schiemer[1], Jürgen Hubbuch[1]

[1] Institute of Engineering in Life Sciences, Section IV: Biomolecular Separation Engineering, Karlsruhe Institute of Technology (KIT), Germany

Abstract

Current biopharmaceutical production heavily relies on chromatography for protein purification. Recently, research has intensified towards finding suitable solutions for monitoring the chromatographic steps by multivariate spectroscopic sensors. Here, hard-constraint MCR was investigated as a calibration-free method for factorizing bilinear preparative protein chromatograms into concentrations and spectra. Protein elutions were assumed to follow Exponentially Modified Gaussian (EMG) curves. In three case studies, MCR was applied to chromatograms of second-derivative UV/Vis

spectra. The three case studies consisted of the separation of a ternary mixture (ribonuclease A, cytochrome c, and lysozyme), multiple binary chromatography runs of cytochrome c and lysozyme, and the separation of an ADC from unconjugated IgG. In all case studies, good estimates of the elution curves were obtained. R^2 values compared to off-line analytics exceeded 0.90. The estimated spectra allowed for protein identification based on a protein spectral library. In summary, MCR was shown to be well able to factorize protein chromatograms without prior calibration. The method may thus substantially simplify analysis of multivariate protein chromatograms with multiple co-eluting species. It may be especially useful in process development.

8.1 Introduction

In modern biopharmaceutical protein purification, preparative chromatography is the method of choice for capturing and polishing steps [19]. Chromatography is popular because it can simultaneously deliver high purity and high yield. To achieve the necessary performance, chromatographic steps need to be well-designed. Already slight process changes can influence the quality profile of the product [3]. The situation is further complicated due to the necessity of complex off-line analytical methods for assessing the quality of biopharmaceuticals.

As a means to improve process monitoring, control and understanding in development and production, PAT has raised a lot of interest [2], [5], [18], [76], [149]. The goal of PAT is to develop and implement sensors which allow for (near) real-time monitoring of quality attributes. Most frequently, on-line and at-line HPLC has been used for different applications including the monitoring of capture and polishing steps [11], [13]–[16], [107].

Recently, spectroscopic approaches in combination with MVDA for the retrieval of overlapping peaks have become more popular [76]. Spectroscopic methods are often non-invasive, fast, and robust [18]. They have been used for the selective *in-situ* quantification of proteins in multi-well plates [39], [41] and selective in-line quantification in preparative chromatography [42]–[44], [76], [163]. These applications have in common that they use spectroscopic data and PLS modeling for selective protein quantification. As PLS regression generates correlative models, a calibration has to be performed prior to application. Furthermore, the model may be susceptible to degeneration and needs to be tested regularly.

As an alternative method for evaluating spectra, MCR has been widely discussed [177]–[179]. MCR maximizes the explained variance of factors

while physically or chemically meaningful constraints are imposed on the behavior of the pure components. Predominantly employed in analytical chemistry, the evolution of MCR is still ongoing, regarding both theory and application [178]. Nevertheless, it has already been used for the resolution of complex chemical mixtures since the 1980s [180], [181]. Since then, different algorithms have been developed for various applications [178].

Regarding the application of MCR for the resolution of protein chromatograms, literature is scarce. Compared to small molecules, different protein spectra exhibit a lower degree of variability [39], which makes resolution more challenging. Additionally, observed 'pure' protein spectra are often combinations of multiple heterogeneities. During a chromatographic elution, these heterogeneities may be separated resulting in a variation of the spectra even for 'pure' components. Vandeginste et al. published a method for three-component curve resolution of proteins in 1985 [182]. More recently, a hybrid-MCR algorithm was shown to be able to determine accurate retention times of simulated SEC chromatograms for up to four co-eluting proteins [183]. Due to the interesting findings of the above study and the efforts necessary for calibrating a statistical model, a further evaluation of MCR for preparative protein chromatography is of interest.

In this study, we investigated the factorization of UV/Vis spectral data from preparative protein chromatography. To increase spectral differences of proteins, second derivative spectral pretreatment was applied. The obtained spectra were analyzed by an EMG-constrained MCR algorithm. The factorization was based on the Pure Component Decomposition (PCD) algorithm originally proposed by Neymeyr et al. [184]. In a first case study, three model proteins (ribonuclease A, cytochrome c, and lysozyme) were separated by CEX. A second case study factorized an augmented data matrix from multiple binary elutions of the model proteins cytochrome c and lysozyme. A third case study monitored the separation of a surrogate ADC from its unconjugated IgG by HIC. In all case studies, the estimated concentration profiles were compared to off-line analytics. The estimated spectra of the three case studies were compared to a protein spectral library.

8.2 Theory

8.2.1 MCR by PCD

Considering a spectroscopic transmission measurement, the absorbance generally follows the Lambert-Beer law. For a multi-wavelength and multi-component case, it reads:

$$A = CS^T + E, \tag{8.1}$$

where $A \in \mathbb{R}^{n \times m}$ is the absorbance matrix, $C \in \mathbb{R}^{n \times o}$ is the concentration matrix, $S \in \mathbb{R}^{m \times o}$ is the spectral matrix and $E \in \mathbb{R}^{n \times m}$ is the residual matrix. n, m and o refer to the number of samples, the number of wavelengths and the number of species, respectively.

The goal of MCR is to retrieve approximate C and S from A under certain constraints such as the chromatographic elution profile. As proposed by Sawall et al. [185], this can be formulated by adapting the PCD algorithm [184] as a minimization problem of the function

$$F(C, S, p) = \left\lVert A - CS^T \right\rVert_F^2 + \gamma \, f_{\text{hard}}(C, S, p). \tag{8.2}$$

The first part on the right-hand side consists of the squared Frobenius matrix norm of the residual matrix E. It thus describes the deviation of the product of the computed matrices C and S from the absorbance data. For a good solution, the Frobenius norm should be close to zero. The second part $f_{\text{hard}}(C, S, p)$ defines an error term for additional hard constraints which are discussed in subsection 8.2.2. p are the parameters for the hard constraints. For the current application, $f_{\text{hard}}(C, S, p)$ was multiplied by a weighting factor $\gamma = 100$ to penalize deviations from the hard constraints strongly [185].

Estimating C and S can be difficult, as both matrices may contain a large number of entries. It was previously proposed to retrieve estimates of C and S by rotating a limited number of factors from an easy to compute matrix factorization scheme such as Singular Value Decomposition (SVD) [184]–[186] or PCA [187]. SVD factorizes the original absorbance matrix into the matrices $U \in \mathbb{R}^{n \times n}$, $\Sigma \in \mathbb{R}^{n \times m}$ and $V \in \mathbb{R}^{m \times m}$ according to

$$A = U\Sigma V^T. \tag{8.3}$$

U and V are orthonormal matrices. Σ is a rectangular diagonal matrix with the singular values s_i on the diagonal. The entries are ordered according to their magnitude, i.e. $s_1 \geq s_2 \geq \cdots \geq 0$. The original matrix A can now be low-rank approximated with only a small number of q singular values $\tilde{\Sigma} = \Sigma(1 : q, 1 : q)$ and singular vectors $\tilde{U} = U(:, 1 : q)$, $\tilde{V} = V(:, 1 : q)$.

The number of included factors needs to be evaluated depending on the experiment. Often, q is equal to the number of species in the mixture o. Importantly, the low-rank approximation by SVD captures the maximum possible amount of variance from A with the given number of factors q.

The concentration matrix C and spectral matrix S can now be approximated as a rotation of the singular vectors by $T \in \mathbb{R}^{o \times q}$.

$$\tilde{A} = \tilde{U}\tilde{\Sigma}\tilde{V}^T = \underbrace{\tilde{U}\tilde{\Sigma}T^{-1}}_{=C}\underbrace{T\tilde{V}^T}_{=S^T} \tag{8.4}$$

T^{-1} denotes the matrix inverse. If $o \neq q$, T^{-1} is replaced by pseudo inverse T^+. Neymeyr et al. proofed that a perfect reconstruction of C and S in Equation (8.4) is possible in the absence of noise [184]. The objective function is now reformulated to

$$G(T, p) = F(\tilde{U}\tilde{\Sigma}T^{-1}, \tilde{V}T^T, p). \tag{8.5}$$

Through the low-rank approximation of A, the matrix factorization problem is thus simplified to estimating $o \times q$ rotational parameters and p.

8.2.2 Formulation of the EMG Hard Constraint

It is worth noting that Equation 8.1 and the Frobenius norm in Equation (8.2) and (8.5) do not take into account any time correlation of the concentration. Thus, any intended time correlation needs to be captured by $f_{\text{hard}}(C, S, p)$. In chromatography, the elution of different components is often empirically described as EMG curves [19]. An EMG describes a Gaussian peak convoluted with a continuously stirred tank reactor. It is selected as a hard constraint on the columns of C. A similar approach was recently taken by Arase et al. who factorized analytical chromatograms of small molecules by MCR with a bidirectional EMG constraint [188]. In this work, the EMG computation $c(t; h, \mu, \sigma, \tau)$ proposed by Kalambet et al. is used [189].

$$c(t; h, \mu, \sigma, \tau)$$

$$= \begin{cases} h \cdot \frac{\sigma}{\tau} \cdot \sqrt{\frac{\pi}{2}} \cdot \exp\left(\frac{\mu-t}{\tau} + \frac{\sigma^2}{2\tau^2}\right) \cdot \text{erfc}(z), & \text{if } z \leq 0, \\ h \cdot \frac{\sigma}{\tau} \cdot \sqrt{\frac{\pi}{2}} \cdot \exp\left(-\frac{(\mu-t)^2}{2\sigma^2}\right) \cdot \text{erfcx}(z), & \text{if } 0 < z \leq 6.71 \cdot 10^7, \\ h \cdot \frac{\exp\left(-\frac{(\mu-\tau)^2}{2\sigma^2}\right)}{1 + \frac{(\mu-t)\cdot\tau}{\sigma^2}}, & \text{else,} \end{cases} \tag{8.6}$$

$$z = \frac{1}{\sqrt{2}} \left(\frac{\mu - t}{\sigma} + \frac{\sigma}{\tau} \right). \tag{8.7}$$

Here, t refers to the time. h is a scaling factor of the EMG. μ and σ denote the mean value and standard deviation of a Gaussian peak before convolution. τ is the decay constant of the continuously stirred tank reactor. Additionally, fronting can be implemented by reflecting t at μ for negative τ, i.e. $\hat{t} = 2\mu - t$ and $c(\hat{t}; \mu, \sigma, -\tau)$ if $\tau < 0$.

For each species, an EMG peak shape is now included as a hard constraint in $f_{\text{hard}}(C, A, p)$ and evaluated at every measured time point t_i of the absorbance matrix.

$$f_{\text{hard}}(C, A, p) = \sum_{i=1}^{n} \sum_{j=1}^{o} \left(C_{ij} - c\left(t_i; p(:, j) \right) \right)^2, \tag{8.8}$$

where p is the parameter matrix containing $4 \times o$ entries. As the EMG is positive for $h > 0$, a constraint on $C \geq 0$ is implicitly set. Due to the application to second derivative spectra, the spectral matrix S is not ≥ 0 but may also have negative entries. As a result, no constraint on the positivity of the spectral matrix must be set.

The objective function $G(T, p)$ can now be solved with a deterministic numerical solver. We used a quasi-Newton approach as implemented in MATLAB (version 2016a, The Mathworks, Naticks, US). For our purposes, the optimization is split into multiple substages. First, only p is released for optimization. Next, T is optimized for the estimated p. After convergence, the EMG scaling factors h are multiplied into the rotational matrix T. Finally, all remaining parameters are released for optimization until convergence is achieved. The staged approach helps to prevent the solver from diverging.

8.3 Materials and Methods

8.3.1 Proteins and Buffers

In Table 8.1, the proteins used in this paper and their respective manufacturer are listed. All protein solutions and buffers were produced with Ultrapure Water (PURELAB Ultra, ELGA LabWater, Veolia Water Technologies, Saint-Maurice, France). After thorough mixing, the buffers were pH adjusted with HCl, filtrated with cellulose acetate filters with a pore size of $0.2\,\mu\text{m}$ (Sartorius, Göttingen, Germany) and degassed by sonification.

Table 8.1: All proteins used for this study are listed with their respective manufacturer.

Protein	Manufacturer
Ribonuclease A from bovine pancreas	Sigma Aldrich[a]
Cytochrome c from bovine heart	Sigma Aldrich
Lysozyme from chicken egg	Sigma Aldrich
IgG1	MedImmune[b]
IgG2	Lek Pharmaceuticals[c]
Ovomucoid	Sigma Aldrich
Bovine serum albumin	Sigma Aldrich
apo-Transferrin human	Sigma Aldrich
Myoglobin from equine skeletal muscle	Sigma Aldrich
Glucose oxidase from aspergillus niger	Sigma Aldrich

[a] St. Louis, US; [b] Gaithersburg, US; [c] Ljubljana, SL

8.3.2 Preparative Chromatographic Instrumentation

The preparative chromatographic runs were performed using a custom-made experimental setup consisting of a conventional liquid chromatography system and a DAD. The liquid chromatography system was an ÄKTA purifier 10 equipped with pump P-900, sample pump P-960, UV monitor UV-900 (10 mm optical path length), conductivity monitor C-900, pH monitor pH-900, autosampler A-905, and fraction collector Frac-950. The liquid chromatography system was controlled with UNICORN 5.31 (all GE Healthcare, Chalfont St. Giles, UK). In order to obtain in-line UV/Vis absorption spectra, an UltiMate DAD-3000 was added to the flow path after the column. The DAD was equipped with a semi-preparative flow cell (0.4 mm optical pathlength) except for the ADC separation where an analytical flow cell (10 mm optical path length) was used. The DAD was controlled with Chromeleon 6.80 (all Thermo Fisher Scientific, Waltham, US). The data acquisition of the DAD was triggered by custom-made software written in MATLAB and Visual Basic for Applications (VBA, Microscoft, Redmond, US). A detailed description can be found in [42].

8.3.3 Analytical Chromatographic Instrumentation

As reference analytics, analytical chromatography was performed with the collected fractions, using a Dionex UltiMate 3000 liquid chromatography system. The system was composed of a HPG-3400RS pump, a WPS-

3000TFC analytical autosampler, a TCC-3000RS column thermostat, and a DAD-3000RS detector. The system was controlled by Chromeleon 6.80 (all Thermo Fisher Scientific).

8.3.4 Preparative CEX Chromatography

Five CEX runs were performed with a 1 ml MediaScout MiniChrom column (Atoll, Weingarten, Germany) with dimensions $5\,\text{mm} \times 50\,\text{mm}$ prepacked with SP Sepharose FF (GE Healthcare). First, the column was equilibrated (20 mM sodium phosphate [Sigma Aldrich], pH 7.0) and then loaded with 500 mg of each protein used in the run (injection volume 100 µL). Elution was performed with a linear gradient from 0 % to 100 % elution buffer (20 mM sodium phosphate, 500 mM sodium chloride [Merck, Darmstadt, Germany], pH 7.0). During all runs, the flow rate was 0.2 mL/min and 200 µL fractions were collected. Spectra were acquired in the range from 240 nm to 310 nm. Four runs were executed with a two-component mixture of cytochrome c and lysozyme. Gradients were run in 1 CV, 3 CV, 5 CV, and 7 CV. Additionally, a 3 CV run with a three-component system consisting of lysozyme, cytochrome c, and ribonuclease A was carried out.

Analytical Chromatography

The fractions from preparative CEX chromatography were analyzed by analytical CEX chromatography on a Proswift SCX-1S $4.6\,\text{mm} \times 50\,\text{mm}$ column (Thermo Fisher Scientific). A flow rate of 1.5 mL/ min was used during the whole run. For each sample, the column was first equilibrated for 2.5 min with load buffer (20 mM TRIS [Merck, Darmstadt, Germany], pH 8.0). Next, 20 µL of sample was injected into the system and washed for 0.5 min with load buffer. A bilinear gradient was performed during the next 4 min with 0 % to 10 % (2 min) and 10 % to 100 % elution buffer (20 mM TRIS, 700 mM sodium chloride [Merck], pH 8.0). Finally, the column was stripped for 0.5 min with 100 % elution buffer.

8.3.5 Preparative HIC of a surrogate ADC

The load for the preparative HIC step was produced by the conjugation reaction of a surrogate drug (7-Diethylamino-3-(4'-maleimidylphenyl)-4-methylcoumarin) with an IgG1. The resulting surrogate ADC had similar characteristics regarding structure and hydrophobicity to normal ADCs however lacked their toxicity. The load was prepared by mixing IgG 1 with surrogate ADC to a final concentration of 2 g/L for each component.

124

A 1 mL Toyoscreen 650M Phenyl column was purchased from Tosoh (Tokyo, Japan). For the preparative chromatographic run, the flow rate was set to 0.2 mL/ min. The column was equilibrated for 5 mL with 25 mM sodium phosphate and 1 M ammonium sulfate at pH 7.0. 100 µL of the load were injected and washed for 2 mL. Subsequently, a 15 mL linear gradient was performed with the elution buffer (18.75 mM phosphate, pH 7.0, 25 % (V/V) 2-propanol) from 20 % to 70 %. The column was stripped with 8 mL elution buffer. During the whole chromatographic separation, spectra were acquired in the range from 250 nm to 450 nm. The eluent was collected in 200 µL fractions in 96-well plates.

Analytical Chromatography

Analytics were performed by reversed-phase chromatography to quantify the ADCs as well as the unmodified IgG1. Reduction or different sample preparation was not required. An Acquity UPLC Protein BEH C4 column (Waters Corporation, Milford, USA; 300 Å, 1.7 µm, 2.1 mm × 50 mm) was run at a flow rate of 0.45 mL/ min. The column oven was heated to 80 °C. Solvent A consisted of 0.1 % TFA in ultrapure water. Solvent B was 0.1 % TFA in acetonitrile. After equilibration and injection at 26 % B, the fraction of B was raised to 30 %. Next, a 4.8 min gradient from 30 % B to 38 % B was used for separation of the conjugate species. The resulting chromatograms yielded peak areas of unconjugated, mono-conjugated and di-conjugated mAb. For the current application all conjugated species were summed.

8.3.6 UV/Vis Spectral Library

For the spectral library, all proteins in Table 8.1 except the IgG1 and IgG2 were dissolved at 2.5 g/L in 20 mM sodium phosphate buffer at pH 7.0. The IgG2 was provided as a virus inactivated solution from a Protein A purification step. It was diluted in phosphate buffer to 2.5 g/L. The IgG1 was not included in the spectral library.

Each entry in the spectral library was generated by injecting the protein solutions with the autosampler and a 100 µL sample loop into the chromatography system at a flow rate of 0.2 mL/ min. No column was attached to the system. The samples were pumped through the DAD resulting in chromatograms with EMG peak shapes due to the system dispersion. To obtain spectra normalized by mass, the chromatograms were integrated over time for each wavelength λ_i in MATLAB with a trapezoidal integra-

tion scheme, multiplied by the flow rate u and normalized by the injected mass m and optical pathlength l.

$$\varepsilon_{ref,\lambda_i} = \frac{u}{m \cdot l} \int A_{\lambda_i}(t)dt \tag{8.9}$$

8.3.7 Data Analysis

All data analysis was performed in MATLAB on a personal computer equipped with a Core i5-4440 CPU at 3.10 GHz (Intel, Santa Clara, US). The optimization problem was implemented as described in section 8.2. Second derivatives were taken of the spectroscopic data with a second order Savitzky-Golay filter [82] with a 7 point window width. The resulting absorbance matrix A was used for MCR.

8.4 Results and Discussion

In this publication, the factorization of multivariate UV/Vis data from preparative protein chromatography by MCR was tested. Instead of using the absorbance matrix directly for MCR, spectra were first derived twice. This was done for two reasons: First, taking second derivatives of spectral data helps to remove baseline offsets and measurement drifts [22]. Second, it is also a popular technique in protein analytics to enhance the UV/Vis fine structure. Generally, protein UV/Vis spectra are relatively uniform with comparably little variation (see Figure 8.1). Taking the second derivative enhances spectral differences of proteins [32], [162]. Contrary to the original spectra, derived spectra contain positive as well as negative bands. Thus, no positivity constraint was set on the spectral matrix S. The positivity of the concentration was enforced by the EMGs. This approach was evaluated in three case studies.

8.4.1 Analysis of a Three-Component Protein Chromatogram

Three model proteins (ribonuclease A, cytochrome c, and lysozyme) were eluted from a CEX column with a 3 CV linear gradient. In Figure 8.2, the resulting absorbance at 280 nm is shown. The normalized protein concentrations were color-coded into the absorbance trace. In the same figure, the time-evolution of the original and derived spectra is depicted. Compared to the original spectra, the second derivative spectra allow a distinction of the
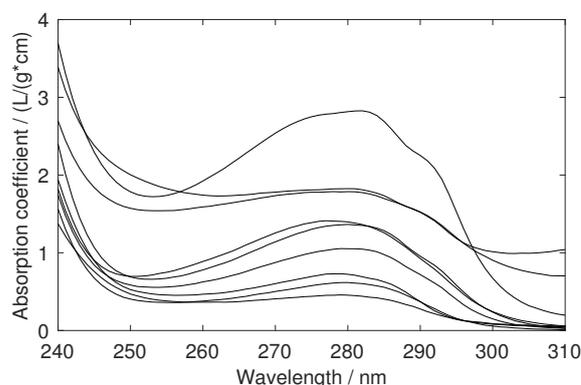
Figure 8.1: Protein spectra from a spectral library are shown. The protein spectra are relatively uniform with an absorption maximum around 280 nm. Differences are visible on the shoulder of the absorption bands and in the through-to-peak distance between 250 nm and 280 nm.

different components based on spectral features. Furthermore, the observed background drift could be reduced.
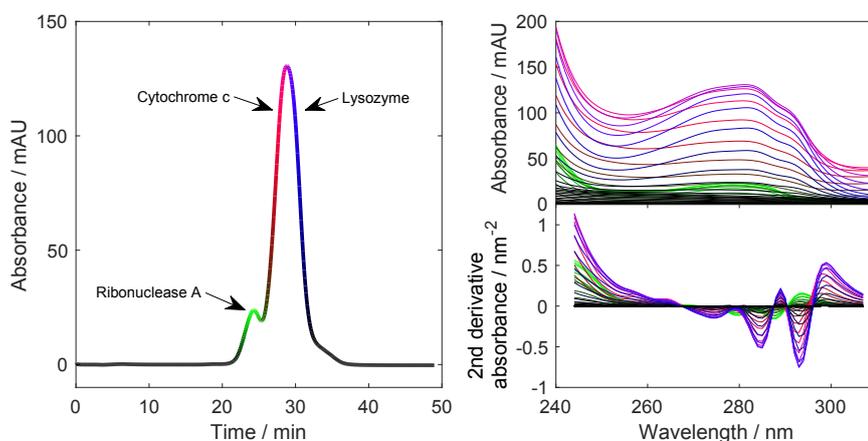

Figure 8.2: Spectral changes during elution are illustrated for case study I. On the left side, the absorbance at 280 nm is shown. The absorbance trace is color-coded with the normalized concentrations of ribonuclease A (green), cytochrome c (red), and lysozyme (blue). The spectra in corresponding colors are shown on the right side (top: original spectra, bottom: second derivative spectra).

The second-derivative absorbance matrix A was subsequently analyzed by SVD. In Figure 8.3, the singular values Σ as well as the first four left and right singular vectors (U and V) are shown. The singular values showed an approximate exponential decay over the first five points and flattened out

for latter entries. The left and right singular vectors one, two, and three only seemed to contain little noise. However, the fourth left singular vector was offset from zero over the whole elution, i.e. the fourth singular vectors contain the baseline offset. The fourth right singular vector showed signs of noise with high fluctuations between subsequent wavelengths. Based on these observations, it was decided to use the first three singular vectors for MCR. For the deterministic optimization of the objective function, initial
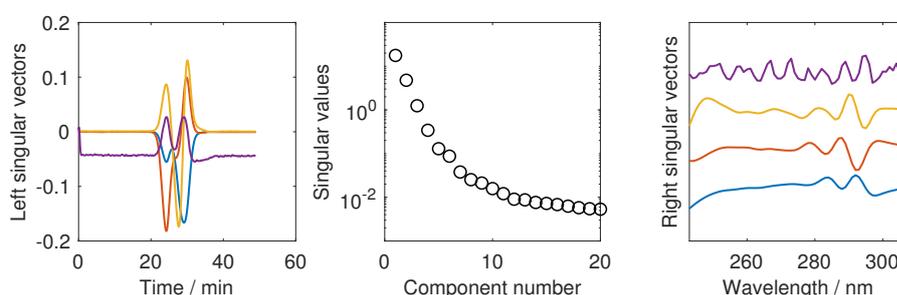

Figure 8.3: SVD of the UV/Vis spectral data of the first case study. The plots show the first four left singular vectors (left), the singular values (middle), and the first four right singular vectors (right). The right singular vectors are offset to simplify interpretation. The vectors are colored according to their column number. Blue: first singular vector, red: second singular vector, yellow: third singular vector, violet: fourth singular vector. It is worth noting, that the extremes of the left singular vectors occurred during the elution of the proteins.

values were set for T as well as p. Figure 8.3 shows that the first singular vector followed the total protein concentration while vector two and three contained information on the time evolution of the spectral differences of the proteins. Consequently, the extremes of the vectors coincided with the concentration maxima of the different components. Based on this argumentation, the initial MCR parameters were set based on the SVD. The initial mean values μ_0 for the EMGs were selected based on the location of the extremes of the left singular vectors. For the convergence of the algorithm, it was of major importance to provide good initial values of the peak location. The initial rotational matrix T_0 was established by inspecting the contribution of the left different singular vectors at the different μ_0. If the left singular vector contributed positively at μ_0, it was added and otherwise subtracted. To normalize the magnitude of the contributions, each entry

was multiplied by the singular value. For the first case study, this resulted in following rotational matrix:

$$T_0 = \begin{pmatrix} -s_1 & -s_2 & s_3 \\ -s_1 & -s_2 & -s_3 \\ -s_1 & s_2 & s_3 \end{pmatrix}. \tag{8.10}$$

The initial standard deviations σ_0 and decay constants τ_0 were set for all proteins to the values 10 and 1, respectively. σ_0 was selected to be in the range of the peak widths observed in U. τ_0 was selected to initially yield an almost symmetric peak. With this initial set of parameters, the optimization converged in less than 30 s.

In Figure 8.4, the optimized MCR results are shown. The estimated maximal concentration location from MCR coincided well with the results from off-line analytics. The good overall agreement between MCR and off-line analytics was also reflected by the high R^2 values. Based on normalized peak areas, values of 0.94, 0.93, and 0.92 were reached for ribonuclease A, cytochrome c and lysozyme, respectively. Differences in the peak shape were visible especially regarding peak tailing. As similar differences occurred for all eluted proteins, the additional tailing in off-line analytics was explained by the system dispersion between detector and fractionator. For
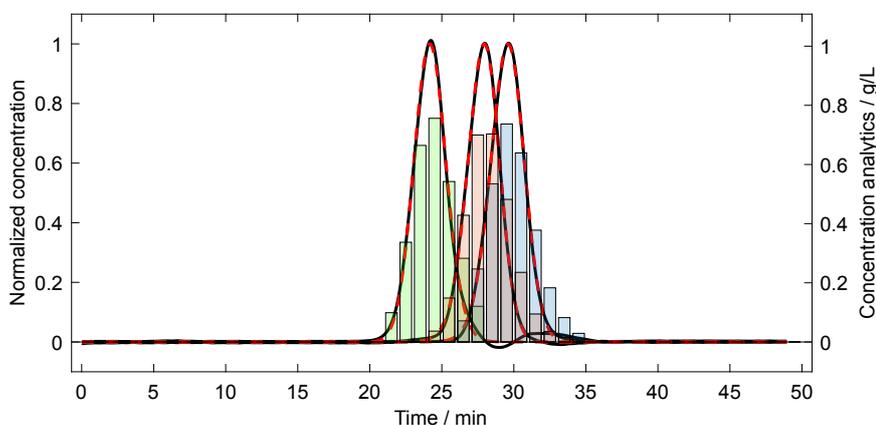

Figure 8.4: Chromatogram of the first case study as retrieved by MCR and compared to off-line analytics. The dashed red lines show the normalized concentration estimate from the hard model. The solid black lines correspond to the rotated left singular vectors. The bars show the measured concentration by off-line analytics. Yellow: ribonuclease A, red: cytochrome c, blue: lysozyme.

a single three-component run, the combination of MCR with an EMG hard

constraint and second derivative spectra provided a good estimation of the elution profile of the different protein components without prior calibration.

8.4.2 Simultaneous Application to Multiple Chromatograms

Next, the PCD algorithm was tested for factorizing multiple binary chromatograms simultaneously. To this end, the absorbance matrices from the individual chromatography runs were concatenated column-wise resulting in $A_{super} \in \mathbb{R}^{\bar{n} \times m}$ with $\bar{n} = \sum_i n_i$ and n_i being the number of measurements per run. For all subsequent analysis, A_{super} was used.

Similar to the evaluation of the ternary protein elution, A_{super} was first analyzed by SVD (Figure 8.5). As expected for a binary mixture, the first two singular values were significantly larger than the following. This was also reflected by the shape of the singular vectors. The third left and right singular vectors already contained a significant contribution of baseline drift and noise. Thus, MCR was performed based on two singular vectors. The initial rotational matrix was defined in the same manner as described above. As each chromatography run was described by two EMGs and a total of four runs were performed a total of eight sets of EMG parameters were necessary. Initial parameter assignment followed the same reasoning as described for the ternary mixture. After initialization, the optimization converged in
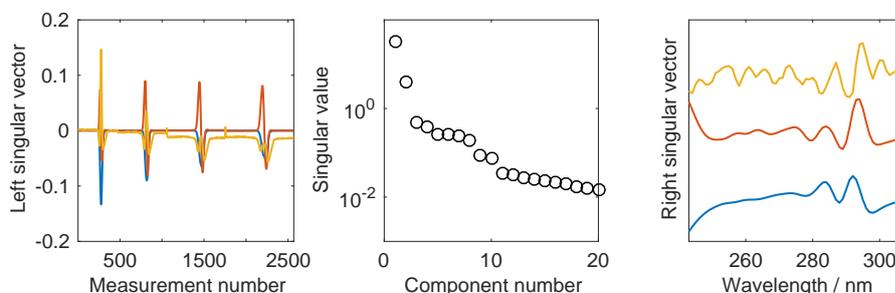

Figure 8.5: SVD of the UV/Vis spectral data of the second case study. The plots show the first three left singular vectors (left), the singular values (middle), and the first three right singular vectors (right). The right singular vectors are offset to simplify interpretation. The vectors are colored according to their column number. Blue: first singular vector, red: second singular vector, yellow: third singular vector.

a matter of minutes to the final solution (Figure 8.6). The peak-maxima locations were again accurately determined by MCR. Similar to the separation of the ternary mixture, some deviations could be observed in the

peak height and tailing. This was again attributed to system dispersion. Interestingly, the differences between off-line analytics and MCR estimation were more pronounced for steeper elution gradients (see Figure 8.6A and D). This supported the assumption that the differences were caused by system dispersion. The steeper gradients resulted in quicker changes in protein concentrations which in turn were more affected by mixing and diffusive peak broadening. Despite these deviations, good estimations were obtained for the elution of cytochrome c and lysozyme with R^2 values of 0.93 and 0.91, respectively. Between the concentrations by the rotated singular vectors and the hard model, only minor differences occurred. Thus, the method could be extended to the case of multiple chromatographic runs while still obtaining a stable convergence of the algorithm.
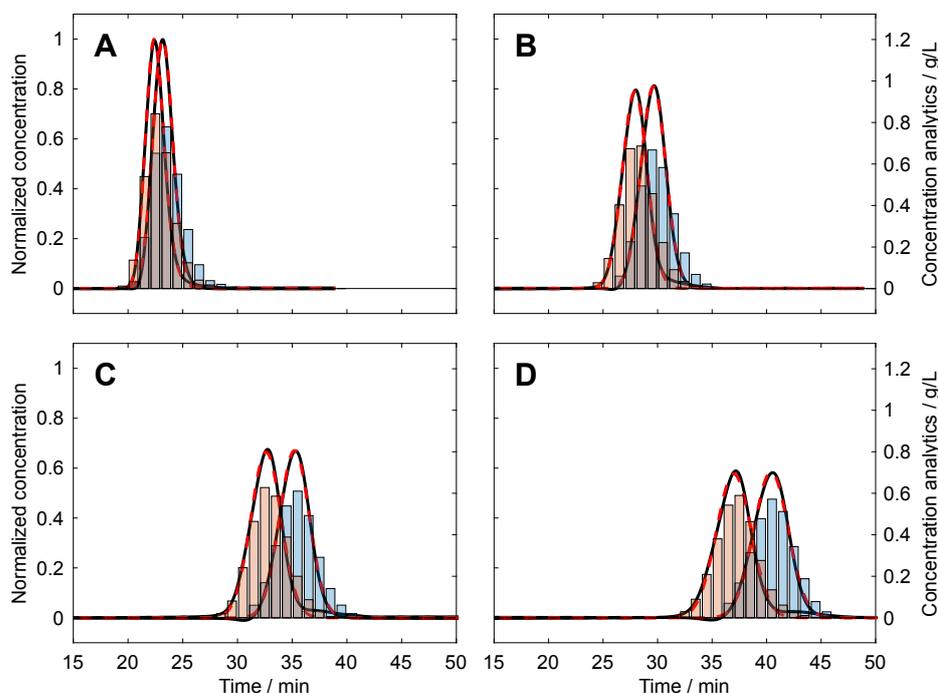

Figure 8.6: Chromatograms of the second case study as retrieved by MCR and compared to off-line analytics. The four plots show the different runs with varied gradient length. A: 1 CV, B: 3 CV, C: 5 CV, D: 7 CV. The dashed red lines show the normalized concentration estimates from the hard model. The solid black lines correspond to the rotated left singular vectors. The bars show the measured concentration by off-line analytics. Red: cytochrome c, blue: lysozyme.

8.4.3 Application of MCR to an ADC Purification Step

In the third case study, an ADC conjugation reaction mixture was loaded onto a HIC column. This purification step aimed to deplete chemical reactants and separate conjugated from native IgG1. Due to the reaction chemicals, the loaded mixture was relatively complex. Additionally, the protein concentration during elution was lower compared to the previous case studies. This increased the perceived noise level and baseline drift. To simplify the analysis of the chromatogram, the evaluation focused on the main elution peak of native and conjugated IgG1.

In Figure 8.7, the results of an SVD are shown. The first two singular values were noticeably larger than the following. Interestingly, the second left singular vector already contained some baseline drift. Interestingly, the baseline drift became stronger for the third left singular vector. The second right singular vector was not influenced by noise and contained strong spectral bands around 375 nm. These bands are typical for the used surrogate drug. The third right singular vector was noticeably deteriorated by noise. Based on these observations, two components were included into the MCR optimization.Optimization of the third case study converged in
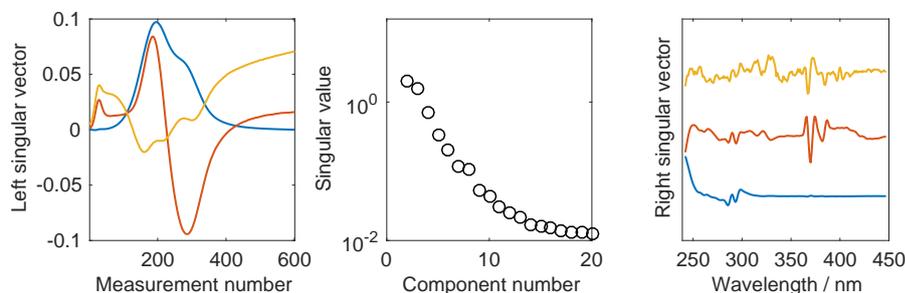

Figure 8.7: SVD of the UV/Vis spectral data of the third case study. The plots show the first three left singular vectors (left), the singular values (middle), and the first three right singular vectors (right). The right singular vectors are offset to simplify interpretation. The vectors are colored according to their column number. Blue: first singular vector, red: second singular vector, yellow: third singular vector.

less than a minute. The resulting chromatogram is shown in Figure 8.8. Similar to the previous case studies, the location of the concentration maxima corresponded well to the off-line analytics. Slight differences could be observed in tailing and fronting. The good results were confirmed by the R^2 values of 0.99 and 0.97 for the native IgG and the ADC, respectively.

The R^2 was again calculated based on the normalized areas. The better agreement between off-line analytics and MCR results were attributed to the long elution gradient which reduced the effects of system dispersion between detector and fractionator as well as possibly the bigger spectral differences between the IgG and the ADC. Interestingly, the differences between the rotated singular vectors and the hard model were bigger in this case. This was explained by the observed baseline drift included in the second singular vectors which again is related to the matrix factorization. SVD captured on each additional dimension as much variation as possible. The information is however not necessarily useful for the estimation of the elution profile. Thus, other matrix factorization approaches may outperform SVD. Nevertheless, the used PCD algorithm provided also in the last case study promising results.
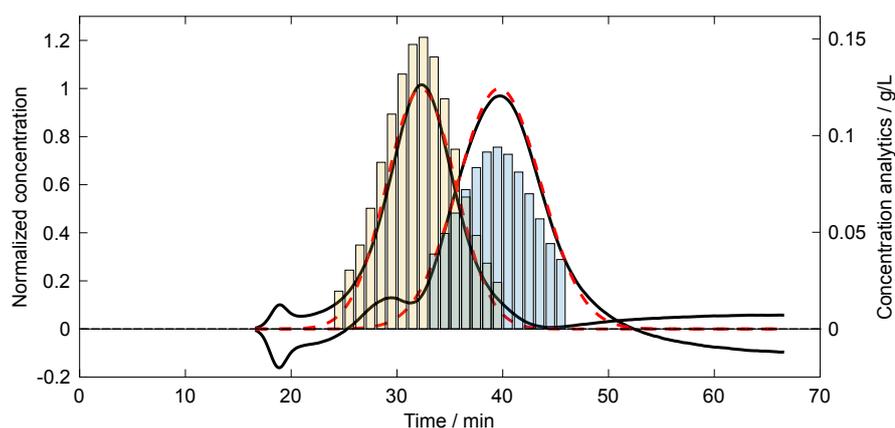

Figure 8.8: Chromatograms of the third case study as retrieved by MCR and compared to off-line analytics. The dashed red lines show the normalized concentration estimate from the hard model. The solid black lines correspond to the rotated left singular vectors. The bars show the measured concentration by off-line analytics. Yellow: native IgG1, blue: ADC.

8.4.4 Protein Identification Based on the Estimated Spectra

To assess how accurate the MCR algorithm estimated data in spectral dimension, the previously estimated spectra were compared to the second derivatives of the spectral library shown in Figure 8.1. Prior to the comparison, all spectra were normalized by standard normal variate transformation to remove any concentration related information. In Figure 8.9, all

spectra were projected onto a plane by PCA. Estimated spectra were projected into the vicinity of the corresponding reference spectra. The results were even more pronounced when directly comparing Euclidean distances between the second derivative spectra. For case study 1, the distances between the reference and estimated spectra of ribonuclease A, cytochrome c, lysozyme were 0.5, 0.6, and 0.3, respectively. All other distances were ≥ 1.8. For the second case study, the Euclidean distances were 0.8 for cytochrome c and 0.3 for lysozyme with all other distances being ≥ 2. For the third case study, only the estimated spectrum from 240 nm to 310 nm of the unconjugated IgG1 was used. The ADC could not be evaluated in this manner, as the drug contributed to the absorption in the protein spectral range and thus biased an identification. The Euclidean distance from the IgG1 was smallest to the IgG2 with 2.1. All other distances were ≥ 2.5. The bigger difference was explained by the structural differences of IgG1 to IgG2 next to the error intoduced by the factorization by MCR. The results show, that the estimated second derivative spectra of the MCR algorithm are close to the spectra of the pure components and may even be used to draw conclusions on the generating protein.
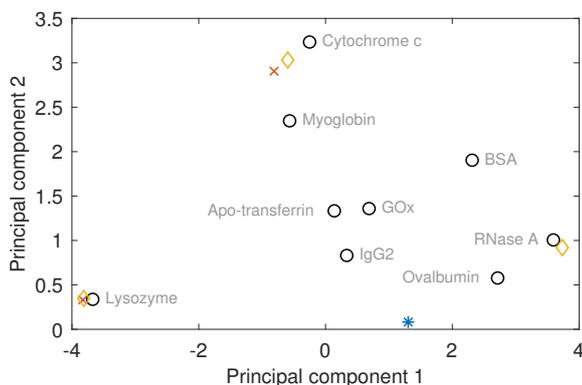
Figure 8.9: Score plot based on a PCA of the spectral library. The spectra from the protein library are marked by black circles. The retrieved spectra from MCR are projected onto the plane. The position of the spectra of the case studies are marked by diamonds (first case study), crosses (second case study), and an asterisk (third case study).

8.5   Conclusion and Outlook

Here, the application of MCR with hard model constraints on preparative protein chromatographic data was tested. The results show that MCR was

well capable of factorizing chromatograms even though protein spectra are subject only to small spectral variation. Differences in peak shape and location of the estimated elution profiles remained small. The matrix factorization of the protein chromatograms could be directly used for protein identification. In summary, MCR seems to be a suitable tool for evaluating protein chromatograms if the eluting species are spectroscopically different. For UV/Vis spectroscopy, mainly the amount of aromatic amino acids, the local environment of aromatic amino acids, and disulfide bridging affect the protein spectra in the investigated spectral range [32]. The proposed method may be especially useful for applications in process development as it is readily applicable without prior calibration.

While the current algorithm is limited to EMGs, other curve shapes could be implemented in a similar manner to also address different elution behavior. Furthermore, MCR is not limited to UV/Vis spectroscopy. Other PAT sensors may benefit from its application as long as they follow a bilinear relation. These occur for many (process) analytical technologies including IR spectroscopy, Raman spectroscopy, and on-/at-/off-line HPLC. In consequence, a wide variety of applications in biopharmaceutical purification are conceivable and may be explored in future.

Acknowledgment

9

Conclusion

In this thesis, a toolbox of spectroscopic PAT sensors for monitoring bio-pharmaceutical DSP operations was developed. All sensors were implemented in- or on-line and each provided unique information on the product and contaminants in the process stream. During the thesis, UV/Vis spectroscopy (Chapters 3, 4, 5, 7 and 8), FTIR spectroscopy (Chapter 6), and light scattering detectors (Chapter 7) were investigated. With these sensors, a broad range of quality attributes, concentrations and protein structural elements were successfully analyzed in (near) real-time. By using a single or multiple sensors in parallel, the toolbox allowed for versatile PAT.

UV/Vis spectroscopy provided high selectivity and sensitivity for proteins coupled with fast measurement times. The method analyzed protein primary, tertiary, and quaternary structure. Its high selectivity and quick measurement times were leveraged in Chapter 3 to monitor the breakthrough during a Protein A capture step in the complex background of HCCF. Based on absorbance spectra in conjunction with PLS modeling, the breakthrough concentration could be accurately predicted. The predicted concentrations were then utilized to control the chromatography system in real-time.

At high concentrations, the application of VP UV/Vis spectroscopy was investigated to prevent detector saturation (Chapter 4). By dynamically adjusting the optical pathlength, protein solutions at concentrations higher than 75 g/L could be measured. PLS modeling was applied to extract protein concentrations from the measured UV/Vis spectra. Based on this approach, elutions from preparative chromatography steps with model proteins as well as a mAb and HMW were monitored selectively. The

predicted concentrations could be used to control pooling decisions of a chromatography system.

By exploiting the sensitivity of UV/Vis spectroscopy to the solvatization of chromophores, the method was further developed to monitor ADC conjugation reactions (Chapter 5). The conjugation reactions of a mAb with two surrogate drugs were observed in-situ in a microplate format and on-line in a lab-scale stirred vessel. From the spectral data, concentrations were predicted by PLS regression. The results showed that the conjugation reactions could be reliably monitored in (near) real-time, an approach which could in future allow for kinetic studies of the ADC conjugation reaction.

In Chapter 6, FTIR spectroscopy was employed in-line to monitor preparative protein chromatography. In contrast to standard UV/Vis spectroscopy, ATR FTIR was not limited by detector saturation at high protein concentrations. The measurement method provided information on the protein secondary structure and the overall protein concentration. Due to the sensitivity of FTIR spectroscopy to any polar covalent bond, it was not only possible to monitor proteins but also the elution of PEGylated protein species and the selective quantification of a process-related impurity.

In Chapter 7, a multimodal spectroscopic approach was chosen for monitoring the reassembly of HBcAg VLPs from homodimers in a CFF process. Light scattering methods contributed information on the molecular weight and the hydrodynamic radius of the particles in solution. These two resulting values allowed to monitor the assembly of VLPs and aggregation directly. UV/Vis spectroscopy measured the protein concentration and changes in tertiary structure around the aromatic amino acids tryptophan and tyrosine. Due to the hydrophobic interaction forces involved in VLP assembly, the changes in the microenvironment of the two aromatic amino acids correlated to the rate of assembly. The rich information from the sensors provided evidence that the VLP assembly process was inhibited by aggregates.

Finally, work on the chemometric data evaluation of preparative chromatograms was performed in Chapter 8. While all other studies in this thesis partly relied on calibrated statistical models, an alternative approach based on MCR was evaluated. Chromatograms were analyzed by combining second-derivative spectroscopy and hard-constraint MCR. This approach factorized preparative protein chromatograms without any information from off-line analytics. The estimated concentration profiles corresponded closely to off-line analytics. The estimated spectra could be used to identify different proteins from a spectral library.

10

Outlook

In the last decade, research on PAT in DSP intensified. While other methods are investigated, a major focus still lies on on-line liquid chromatographic methods. There are multiple reasons for the popularity of liquid chromatography as PAT. Most importantly, liquid chromatography provides superior selectivity for proteins and features a remarkable dynamic range of the analyte concentrations. Analytical liquid chromatography is furthermore the method of choice for measuring many quality attributes in biopharmaceutical industry and thus well established. Hence, future applications of liquid chromatography methods will keep emerging. Problematic are however the long anaylsis times and technical complexity.

If necessary, the selectivity of liquid chromatography can be further enhanced by coupling the system to mass spectroscopy. With modern mass spectrometers reaching acquisition rates above 1 Hz, this is potentially a very powerful method for process monitoring providing unparalleled selectivity. Future applications in DSP may be related to monitoring glycosylation patterns and similar protein modifications as they are otherwise hard to detect. Liquid chromatography coupled to a mass spectrometery is however currently technically complex and very expensive. Without a strong drop in equipment prices, the routine application to all unit operations may generate too high costs.

The application of on-line biochemical assays, microfluidic devices, plasmon surface resonance, quartz crystal microbalances, and similar methods is also a regularly discussed topic in PAT for biopharmaceutical production. The methods are potentially very selective. They are however often

not plug and play and may indeed be very difficult to implement in a GMP environment.

Given the high robustness and frequent application of spectroscopy in other fields as PAT, it will also be established in DSP as one of the most potent tools for process monitoring. The frequent application of spectroscopy for protein structural analysis shows that the methods are also potentially interesting for process monitoring. In other process fields, especially Near-Infrared (NIR) and Raman spectroscopy have been widely applied. Both methods are amendable to fiber optics. The measurement equipment is very robust and for NIR comparably cheap. Due to the low absorption coefficients in NIR, the methodology is not as prone to detector saturation as UV/Vis spectroscopy or mid IR measurements. NIR spectroscopy however has a relatively low selectivity. Already with mid IR, selective protein quantification is difficult. The broad bands in NIR may further complicate analysis. Raman spectroscopy is frequently used for protein structural analysis because of the information content on the primary, secondary and tertiary structure. In upstream, it has shown a high selectivity even in complex cell culture browth. Major drawbacks are the low scattering efficiencies of proteins and the resulting long measurement times. Especially for chromatography, fast measurement times are essential and need to be achieved for efficient process monitoring.

Regarding chemometric data evaluation, the development in DSP stands very much at the beginning. Little research has been performed on soft sensors or coupling measurement results from multiple sensors. Especially chromatography has a strong mechanistic modeling theory already in place which may be coupled with spectroscopic measurements.

Finally, a major challenge in future relates to a flexible implementation of PAT tools into different unit operations. Currently, disposable and single-use technologies are gaining market shares especially during clinical phases [190]. At the same time, the product portfolio of biologics is broadening. New formats such as antibody fragments, nanobodies, conjugated proteins and vaccines, and Fc-fusion proteins are emerging [191]. Depending on the unit operation and biopharmaceutical product, different sensors or sensor combinations may be of interest. Ideally, detectors could therefore be exchanged with little effort. Such a flexible approach to PAT however requires standardized communication between different components e.g. through OPC Foundation's *OPC unified architecture (OPC UA)*. Here, the support of the equipment manufacturer as well as dedicated sensor manufacturer is key [18]. By providing a flexible platform which allows to combine different manufacturing equipment with a range of sensor

technologies, a versatile approach towards future PAT challenges could be implemented.

References

[1] FDA. (2004). Guidance for industry. PAT – A framework for innovative pharmaceutical development, manufacturing, and quality assurance.

[2] 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. Oreilly, 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. Ogrady, 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 and Development*, vol. 19, no. 1, pp. 3–62, 2015.

[3] M. Schiestl, T. Stangler, C. Torella, T. Čepeljnik, H. Toll, and R. Grau, „Acceptable changes in quality attributes of glycosylated biopharmaceuticals", *Nature Biotechnology*, vol. 29, pp. 310–312, 2011.

[4] S. Flatman, I. Alam, J. Gerard, and N. Mussa, „Process analytics for purification of monoclonal antibodies", *Journal of Chromatography B*, vol. 848, no. 1, pp. 79–87, 2007.

[5] J. Glassey, K. V. Gernaey, C. Clemens, T. W. Schulz, R. Oliveira, G. Striedner, and C.-F. Mandenius, „Process analytical technology (PAT) for biopharmaceuticals", *Biotechnology Journal*, vol. 6, no. 4, pp. 369–377, 2011.

[6] S. Challa and R. Potumarthi, „Chemometrics-based process analytical technology (PAT) tools: Applications and adaptation in pharmaceutical and biopharmaceutical industries", *Applied Biochemistry and Biotechnology*, vol. 169, no. 1, pp. 66–76, 2013.

[7] S. M. Mercier, B. Diepenbroek, R. H. Wijffels, and M. Streefland, „Multivariate PAT solutions for biopharmaceutical cultivation: Current progress and limitations", *Trends in Biotechnology*, vol. 32, no. 6, pp. 329–336, 2014.

[8] A. S. Rathore and G. Kapoor, „Application of process analytical technology for downstream purification of biotherapeutics", *Journal of Chemical Technology and Biotechnology*, vol. 90, no. 2, pp. 228–236, 2015.

[9] K. B. Konstantinov and C. L. Cooney, „White Paper on Continuous Bioprocessing May 20–21 2014 Continuous Manufacturing Symposium", *Journal of Pharmaceutical Sciences*, vol. 104, no. 3, pp. 813–820, 03/2015.

[10] N. Brestrich, T. Hahn, and Hubbuch, „Application of spectral deconvolution and inverse mechanistic modelling as a tool for root cause investigation in protein chromatography", *Journal of Chromatography A*, vol. 1437, pp. 158–167, 2016.

[11] R. L. Fahrner, P. M. Lester, G. S. Blank, and D. H. Reifsnyder, „Real-time control of purified product collection during chromatography of recombinant human insulin-like growth factor-I using an on-line assay", *Journal of Chromatography A*, vol. 827, no. 1, pp. 37–43, 1998.

[12] R. L. Fahrner and G. S. Blank, „Real-time control of antibody loading during protein A affinity chromatography using an on-line assay", *Journal of Chromatography A*, vol. 849, no. 1, pp. 191–196, 1999.

[13] R. L. Fahrner and G. S. Blank, „Real-time monitoring of recombinant antibody breakthrough during Protein A affinity chromatography", *Biotechnology and Applied Biochemistry*, vol. 29, no. 2, pp. 109–112, 1999.

[14] A. S. Rathore, M. Yu, S. Yeboah, and A. Sharma, „Case study and application of process analytical technology (PAT) towards bioprocessing: Use of on-line high-performance liquid chromatography (HPLC) for making real-time pooling decisions for process chromatography", *Biotechnology and Bioengineering*, vol. 100, no. 2, pp. 306–316, 2008.

[15] A. S. Rathore, R. Wood, A. Sharma, and S. Dermawan, „Case study and application of process analytical technology (PAT) towards bioprocessing II: Use of ultra-performance liquid chromatography for making real-time pooling decisions for process chromatography",

Biotechnology and Bioengineering, vol. 101, no. 6, pp. 1366–1374, 2008.

[16] O. Kaltenbrunner, Y. Lu, A. Sharma, K. Lawson, and T. Tressel, „Risk-benefit evaluation of on-line high-performance liquid chromatography analysis for pooling decisions in large-scale chromatography",

[17] M. Krättli, F. Steinebach, and M. Morbidelli, „Online control of the twin-column countercurrent solvent gradient process for biochromatography", *Journal of Chromatography A*, vol. 1293, pp. 51–59, 2013.

[18] K. A. Bakeev, *Process analytical technology: Spectroscopic tools and implementation strategies for the chemical and pharmaceutical industries*, 2nd Editio. John Wiley & Sons, 2010, p. 557.

[19] G. Carta and A. Jungbauer, *Protein chromatography: Process development and scale-up.* John Wiley & Sons, 2010.

[20] B. R. Kowalski, „Chemometrics: Views and propositions", *Journal of Chemical Information and Modeling*, vol. 15, no. 4, pp. 201–203, 11/1975.

[21] T. Rajalahti and O. M. Kvalheim, „Multivariate data analysis in pharmaceutics: A tutorial review", *International Journal of Pharmaceutics*, vol. 417, no. 1-2, pp. 280–290, 2011.

[22] L. Eriksson, E. Johansson, N. Kettaneh-Wold, J. Trygg, C. Wikström, and S. Wold, *Multi- and megavariate data analysis.* Umetrics Academy, 2006.

[23] O. M. Kvalheim, „Interpretation of direct latent-variable projection methods and their aims and use in the analysis of multicomponent spectroscopic and chromatographic data", *Chemometrics and Intelligent Laboratory Systems*, vol. 4, no. 1, pp. 11–25, 1988.

[24] O. M. Kvalheim and T. V. Karstang, „Interpretation of latent-variable regression models", *Chemometrics and Intelligent Laboratory Systems*, vol. 7, no. 1, pp. 39–51, 1989.

[25] S. Wold, L. Eriksson, J. Trygg, and N. Kettaneh, „The PLS method – partial least squares projections to latent structures – and its applications in industrial RDP (research, development, and production)", Umeå University, Tech. Rep., 2004.

[26] S. Wold, K. Esbensen, and P. Geladi, „Principal component analysis", *Chemometrics and Intelligent Laboratory Systems*, vol. 2, no. 1-3, pp. 37–52, 08/1987.

[27]   S. Wold, A. Ruhe, H. Wold, and W. Dunn, „The collinearity prob-
       lem in linear regression. The partial least squares (PLS) approach
       to generalized inverses", *SIAM Journal on Scientific and Statistical
       Computing*, vol. 5, no. 3, pp. 735–743, 1984.

[28]   H. Martens and T. Naes, *Multivariate calibration.* John Wiley &
       Sons, 1991, p. 438.

[29]   S. Wold, J. Trygg, A. Berglund, and H. Antti, „Some recent develop-
       ments in PLS modeling", *Chemometrics and Intelligent Laboratory
       Systems*, vol. 58, no. 2, pp. 131–150, 2001.

[30]   S. Wold, M. Sjöström, and L. Eriksson, „PLS-regression: A basic tool
       of chemometrics", *Chemometrics and Intelligent Laboratory Systems*,
       vol. 58, no. 2, pp. 109–130, 2001.

[31]   W. Kessler, *Einführung in die multivariante Datenanalyse*, 1st ed.
       Weinheim: Wiley-VCH Verlag, 2007, pp. 183–210.

[32]   W. Jiskoot and D. Crommelin, Eds., *Methods for structural analysis
       of protein pharmaceuticals.* American Association of Pharmaceutical
       Scientists, 2005.

[33]   A. Barth, „Infrared spectroscopy of proteins", *Biochimica et Bio-
       physica Acta*, vol. 1767, no. 9, pp. 1073–1101, 2007.

[34]   J. R. Lakowicz, *Principles of fluorescence spectroscopy principles of
       fluorescence spectroscopy*, J. R. Lakowicz, Ed. Springer, Boston, US,
       2006.

[35]   R. Schweitzer-Stenner, „Advances in vibrational spectroscopy as a
       sensitive probe of peptide and protein structure: A critical review",
       Vibrational Spectroscopy, vol. 42, no. 1, pp. 98–117, 2006.

[36]   E. Goormaghtigh, V. Cabiaux, and J.-M. Ruysschaert, „Determina-
       tion of soluble and membrane protein structure by Fourier trans-
       form infrared spectroscopy: I. Assignment of model compounds",
       in *Physicochemical methods in the study of biomembranes*, vol. 23,
       Springer, 1994, pp. 328–357.

[37]   E. Papamokos, E. Weber, W. Bode, R. Huber, M. W. Empie, I. Kato,
       and M. Laskowski Jr, „Crystallographic refinement of Japanese quail
       ovomucoid, a Kazal-type inhibitor, and model building studies of
       complexes with serine proteases.", eng, *Journal of Molecular Biology*,
       vol. 158, no. 3, pp. 515–537, 07/1982.

[38] A. S. Rose, A. R. Bradley, Y. Valasatava, J. M. Duarte, A. Prlić, and P. W. Rose, „Web-based molecular graphics for large complexes", in *Proceedings of the 21st International Conference on Web3D Technology*, ser. Web3D '16, Anaheim, California: ACM, 2016, pp. 185–186.

[39] S. K. Hansen, B. Jamali, and J. Hubbuch, „Selective high throughput protein quantification based on UV absorption spectra", *Biotechnology and Bioengineering*, vol. 110, no. 2, pp. 448–460, 2013.

[40] G. E. Arteaga, Y. Horimoto, E. Li-Chan, and S. Nakai, „Partial least-squares regression of fourth-derivative ultraviolet absorbance spectra predicts composition of protein mixtures: Application to bovine caseins", *Journal of Agricultural and Food Chemistry*, vol. 42, no. 9, pp. 1938–1942, 1994.

[41] S. K. Hansen, E. Skibsted, A. Staby, and J. Hubbuch, „A label-free methodology for selective protein quantification by means of absorption measurements", *Biotechnology and Bioengineering*, vol. 108, no. 11, pp. 2661–2669, 2011.

[42] N. Brestrich, T. Briskot, A. Osberghaus, and J. Hubbuch, „A tool for selective inline quantification of co-eluting proteins in chromatography using spectral analysis and partial least squares regression", *Biotechnology and Bioengineering*, vol. 111, no. 7, pp. 1365–1373, 2014.

[43] N. Brestrich, A. Sanden, A. Kraft, K. McCann, J. Bertolini, and J. Hubbuch, „Advances in inline quantification of co-eluting proteins in chromatography: Process-data-based model calibration and application towards real-life separation issues", *Biotechnology and Bioengineering*, vol. 112, pp. 1406–1416, 2015.

[44] N. Brestrich, M. Rüdt, D. Büchler, and J. Hubbuch, „Selective protein quantification for preparative chromatography using variable pathlength UV/Vis spectroscopy and partial least squares regression", *Chemical Engineering Science*, vol. 176, pp. 157–164, 02/2018.

[45] F. Capito, R. Skudas, H. Kolmar, and B. Stanislawski, „Host cell protein quantification by fourier transform mid infrared spectroscopy (FT-MIR)", *Biotechnology and Bioengineering*, vol. 110, no. 1, pp. 252–259, 2013.

[46] F. Capito, R. Skudas, H. Kolmar, and C. Hunzinger, „Mid-infrared spectroscopy-based antibody aggregate quantification in cell culture fluids", *Biotechnology Journal*, vol. 8, no. 8, pp. 912–917, 2013.

147

[47] F. Capito, R. Skudas, H. Kolmar, and C. Hunzinger, „At-line mid infrared spectroscopy for monitoring downstream processing unit operations", *Process Biochemistry*, vol. 50, no. 6, pp. 997–1005, 2015.

[48] C. Walther, S. Mayer, A. Jungbauer, and A. Dürauer, „Getting ready for PAT: Scale up and inline monitoring of protein refolding of Npro fusion proteins", *Process Biochemistry*, vol. 49, no. 7, pp. 1113–1121, 2014.

[49] M. Boulet-Audet, S. G. Kazarian, and B. Byrne, „In-column ATR-FTIR spectroscopy to monitor affinity chromatography purification of monoclonal antibodies", *Scientific Reports*, vol. 6, no. 30526, pp. 1–13, 2016.

[50] A. S. Rathore, X. Li, W. Bartkowski, A. Sharma, and Y. Lu, „Case study and application of process analytical technology (PAT) towards bioprocessing: Use of tryptophan fluorescence as at-line tool for making pooling decisions for process chromatography", *Biotechnology Progress*, vol. 25, no. 5, pp. 1433–1439, 2009.

[51] Z. Yu, J. C. Reid, and Y. P. Yang, „Utilizing dynamic light scattering as a process analytical technology for protein folding and aggregation monitoring in vaccine manufacturing", *Journal of Pharmaceutical Sciences*, vol. 102, no. 12, pp. 4284–4290, 2013.

[52] M.-H. Kamga, H. Woo Lee, J. Liu, and S. Yoon, „Rapid quantification of protein mixture in chromatographic separation using multi-wavelength UV spectra", *Biotechnology Progress*, vol. 29, no. 3, pp. 664–671, 2013.

[53] R. Hahn, K. Shimahara, F. Steindl, and A. Jungbauer, „Comparison of protein A affinity sorbents III. Life time study", *Journal of Chromatography A*, vol. 1102, no. 1–2, pp. 224–231, 2006.

[54] 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.

[55] A. A. Shukla and J. Thömmes, „Recent advances in large-scale production of monoclonal antibodies and related proteins", *Trends in Biotechnology*, vol. 28, no. 5, pp. 253–261, 2010.

[56] R. D. R. Tarrant, M. L. Velez-Suberbie, A. S. Tait, C. M. Smales, and D. G. Bracewell, „Host cell protein adsorption characteristics during protein A chromatography", *Biotechnology Progress*, vol. 28, no. 4, pp. 1037–1044, 2012.

[57] M. Tsukamoto, H. Watanabe, A. Ooishi, and S. Honda, „Engineered protein A ligands, derived from a histidine-scanning library, facilitate the affinity purification of IgG under mild acidic conditions“, *Journal Biological Engineering*, vol. 8, no. 1, pp. 1–9, 2014.

[58] M. Gupta, Ed., *Methods for affinity-based separations of enzymes and proteins*. Birkhäuser Basel, 2013.

[59] C. Jiang, J. Liu, M. Rubacha, and A. A. Shukla, „A mechanistic study of protein A chromatography resin lifetime“, *Journal of Chromatography A*, vol. 1216, no. 31, pp. 5849–5855, 2009.

[60] M. Angarita, T. Müller-Späth, D. Baur, R. Lievrouw, G. Lissens, and M. Morbidelli, „Twin-column capture SMB: A novel cyclic process for protein A affinity chromatography“, *Journal of Chromatography A*, vol. 1389, pp. 85–95, 2015.

[61] V. Warikoo, R. Godawat, K. Brower, S. Jain, D. Cummings, E. Simons, T. Johnson, J. Walther, M. Yu, B. Wright, J. McLarty, K. P. Karey, C. Hwang, W. Zhou, F. Riske, and K. Konstantinov, „Integrated continuous production of recombinant therapeutic proteins“, *Biotechnology and Bioengineering*, vol. 109, no. 12, pp. 3018–3029, 2012.

[62] F. Capito, A. Zimmer, and R. Skudas, „Mid-infrared spectroscopy-based analysis of mammalian cell culture parameters“, *Biotechnology Progress*, vol. 31, no. 2, pp. 578–584, 2015.

[63] R. F. Steinhoff, D. J. Karst, F. Steinebach, M. R. Kopp, G. W. Schmidt, A. Stettler, J. Krismer, M. Soos, M. Pabst, A. Hierlemann, M. Morbidelli, and R. Zenobi, „Microarray-based MALDI-TOF mass spectrometry enables monitoring of monoclonal antibody production in batch and perfusion cell cultures“, *Methods*, vol. 104, pp. 33–40, 2015.

[64] D. Karst, F. Steinebach, and M. Morbidelli, „Integrated continuous processing for the manufacture of monoclonal antibodies“, in *Integrated Continuous Biomanufacturing II, ECI Symposium Series*, 2015.

[65] H. A. Chase, „Rapid chromatographic monitoring of bioprocesses“, *Biosensors*, vol. 2, no. 5, pp. 269–286, 1986.

[66] S. S. Ozturk, J. C. Thrift, J. D. Blackie, and D. Naveh, „Real-time monitoring of protein secretion in mammalian cell fermentation: Measurement of monoclonal antibodies using a computer-controlled HPLC system (BioCad/RPM)", *Biotechnology and Bioengineering*, vol. 48, no. 3, pp. 201–206, 1995.

[67] S. K. Paliwal, T. K. Nadler, D. I. C. Wang, and F. E. Regnier, „Automated process monitoring of monoclonal antibody production", *Analytical Chemistry*, vol. 65, no. 23, pp. 3363–3367, 1993.

[68] P. Bängtsson, E. Estrada, K. Lacki, and H. Skoglar, *A method in a chromatography system*, EP Patent App. EP20,100,792,416, 2012.

[69] H. Martens and T. Næs, *Multivariate calibration*. Wiley, 1989.

[70] A. Höskuldsson, „PLS regression methods", *Journal of Chemometrics*, vol. 2, pp. 211–228, 1988.

[71] A. A. Shukla and P. Hinckley, „Host cell protein clearance during protein A chromatography: Development of an improved column wash step", *Biotechnology Progress*, vol. 24, no. 5, pp. 1115–1121, 2008.

[72] N. Aboulaich, W. K. Chung, J. H. Thompson, C. Larkin, D. Robbins, and M. Zhu, „A novel approach to monitor clearance of host cell proteins associated with monoclonal antibodies", *Biotechnology Progress*, vol. 30, no. 5, pp. 1114–1124, 2014.

[73] 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.

[74] M. Rüdt, T. Briskot, and J. Hubbuch, „Advances in downstream processing of biologics – Spectroscopy: An emerging process analytical technology", *Journal of Chromatography A*, vol. 1490, pp. 2–9, 2017.

[75] A. S. Rathore, L. Parr, S. Dermawan, K. Lawson, and Y. Lu, „Large scale demonstration of a process analytical technology application in bioprocessing: Use of on-line high performance liquid chromatography for making real time pooling decisions for process chromatography", *Biotechnology Progress*, vol. 26, no. 2, pp. 448–457, 2010.

[76] M. Rüdt, N. Brestrich, L. Rolinger, and J. Hubbuch, „Real-time monitoring and control of the load phase of a protein A capture step", *Biotechnology and Bioengineering*, vol. 114, no. 2, pp. 368–373, 2017.

[77] P. Baumann, T. Huuk, T. Hahn, A. Osberghaus, and J. Hubbuch, „Deconvolution of high-throughput multicomponent isotherms using multivariate data analysis of protein spectra", *Engineering in Life Sciences*, vol. 16, no. 2, pp. 194–201, 2016.

[78] R. L. Doucen, J. P. Houdeau, C. Cousin, and V. Menoux, „Variable path-length, low-temperature cells for absorption spectroscopy", *Journal of Physics E: Scientific Instruments*, vol. 18, no. 3, p. 199, 1985.

[79] P. Flowers and S. Callender, „Variable path length transmittance cell for ultraviolet, visible, and infrared spectroscopy and spectro-electrochemistry", *Analytical Chemistry*, vol. 68, no. 1, pp. 199–202, 1996.

[80] B. Chen, H. Wu, and S. F. Y. Li, „Development of variable path-length UV-vis spectroscopy combined with partial-least-squares regression for wastewater chemical oxygen demand (COD) monitoring", *Talanta*, vol. 120, pp. 325–330, 2014.

[81] S. V. Thakkar, K. M. Allegre, S. B. Joshi, D. B. Volkin, and C. R. Middaugh, „An application of ultraviolet spectroscopy to study interactions in proteins solutions at high concentrations", *Journal of Pharmaceutical Sciences*, vol. 101, pp. 3051–3061, 2012.

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

[83] A. Beck, T. Wurch, C. Bailly, and N. Corvaia, „Strategies and challenges for the next generation of therapeutic antibodies", *Nature Reviews Immunology*, vol. 10, p. 345, 2010.

[84] Y. Xu, G. Jiang, C. Tran, X. Li, T. H. Heibeck, M. R. Masikat, Q. Cai, A. R. Steiner, A. K. Sato, T. J. Hallam, *et al.*, „RP-HPLC DAR characterization of site-specific antibody drug conjugates produced in a cell-free expression system", *Organic Process Research & Development*, vol. 20, no. 6, pp. 1034–1043, 2016.

[85] R. P. Lyon, T. D. Bovee, S. O. Doronina, P. J. Burke, J. H. Hunter, H. D. Neff-LaFord, M. Jonas, M. E. Anderson, J. R. Setter, and P. D. Senter, „Reducing hydrophobicity of homogeneous antibody-drug conjugates improves pharmacokinetics and therapeutic index", *Nature Biotechnology*, vol. 33, no. 7, p. 733, 2015.

[86] N. Jain, S. W. Smith, S. Ghone, and B. Tomczuk, „Current ADC linker chemistry", *Pharmaceutical Research*, vol. 32, no. 11, pp. 3526–3540, 2015.

[87] J. R. Junutula and H.-P. Gerber, *Next-generation antibody-drug conjugates (ADCs) for cancer therapy*. ACS Publications, 2016.

[88] D. Schumacher, C. P. Hackenberger, H. Leonhardt, and J. Helma, „Current status: Site-specific antibody drug conjugates", *Journal of Clinical Immunology*, vol. 36, no. 1, pp. 100–107, 2016.

[89] P. Agarwal and C. R. Bertozzi, „Site-specific antibody-drug conjugates: The nexus of bioorthogonal chemistry, protein engineering, and drug development", *Bioconjugate Chemistry*, vol. 26, no. 2, pp. 176–192, 2015.

[90] A. Wakankar, Y. Chen, Y. Gokarn, and F. S. Jacobson, „Analytical methods for physicochemical characterization of antibody drug conjugates", *mAbs*, vol. 3, no. 2, pp. 161–172, 2011.

[91] R. V. Chari, B. A. Martell, J. L. Gross, S. B. Cook, S. A. Shah, W. A. Blättler, S. J. McKenzie, and V. S. Goldmacher, „Immunoconjugates containing novel maytansinoids: Promising anticancer drugs", *Cancer Research*, vol. 52, no. 1, pp. 127–131, 1992.

[92] K. J. Hamblett, P. D. Senter, D. F. Chace, M. M. Sun, J. Lenox, C. G. Cerveny, K. M. Kissler, S. X. Bernhardt, A. K. Kopcha, R. F. Zabinski, *et al.*, „Effects of drug loading on the antitumor activity of a monoclonal antibody drug conjugate", *Clinical Cancer Research*, vol. 10, no. 20, pp. 7063–7070, 2004.

[93] K. Furuki and T. Toyo'oka, „Determination of thiol-to-protein ratio and drug-to-antibody ratio by in-line size exclusion chromatography with post-column reaction", *Analytical Biochemistry*, vol. 527, pp. 33–44, 2017.

[94] Y. Tang, F. Tang, Y. Yang, L. Zhao, H. Zhou, J. Dong, and W. Huang, „Real-time analysis on drug-antibody ratio of antibody-drug conjugates for synthesis, process optimization, and quality control", *Scientific Reports*, vol. 7, no. 1, p. 7763, 2017.

[95] A. C. Quinn, P. J. Gemperline, B. Baker, M. Zhu, and D. S. Walker, „Fiber-optic UV/visible composition monitoring for process control of batch reactions", *Chemometrics and Intelligent Laboratory Systems*, vol. 45, no. 1-2, pp. 199–214, 1999.

[96] S. P. Gurden, J. A. Westerhuis, and A. K. Smilde, „Monitoring of batch processes using spectroscopy", *AIChE Journal*, vol. 48, no. 10, pp. 2283–2297, 2002.

[97] P. Suppan, „Invited review solvatochromic shifts: The influence of the medium on the energy of electronic states", *Journal of Photochemistry and Photobiology A: Chemistry*, vol. 50, no. 3, pp. 293–330, 1990.

[98] C. Reichardt and T. Welton, *Solvents and solvent effects in organic chemistry*. John Wiley & Sons, 2011.

[99] S. de Jong, „SIMPLS: An alternative approach to partial least squares regression", *Chemometrics and Intelligent Laboratory Systems*, vol. 18, no. 3, pp. 251–263, 1993.

[100] K. Deep, K. P. Singh, M. Kansal, and C. Mohan, „A real coded genetic algorithm for solving integer and mixed integer optimization problems", *Applied Mathematics and Computation*, vol. 212, no. 2, pp. 505–518, 2009.

[101] R. Ragone, G. Colonna, C. Balestrieri, L. Servillo, and G. Irace, „Determination of tyrosine exposure in proteins by second-derivative spectroscopy", *Biochemistry*, vol. 23, no. 8, pp. 1871–1875, 04/1984.

[102] H. Mach and C. R. Middaugh, „Simultaneous monitoring of the environment of tryptophan, tyrosine, and phenylalanine residues in proteins by near-ultraviolet second-derivative spectroscopy", *Analytical Biochemistry*, vol. 222, no. 2, pp. 323–331, 1994.

[103] L. M. Tedaldi, A. E. Aliev, and J. R. Baker, „[2+ 2] Photocycloadditions of thiomaleimides", *Chemical Communications*, vol. 48, no. 39, pp. 4725–4727, 2012.

[104] S. Andris, M. Wendeler, X. Wang, and J. Hubbuch, „Multi-step high-throughput conjugation platform for the development of antibody-drug conjugates", *Journal of Biotechnology*, vol. 278, pp. 48–55, 2018.

[105] K. C. Catcott, M. A. McShea, C. U. Bialucha, K. L. Miller, S. W. Hicks, P. Saxena, T. G. Gesner, M. Woldegiorgis, M. E. Lewis, C. Bai, M. S. Fleming, S. A. Ettenberg, H. K. Erickson, and N. C. Yoder, „Microscale screening of antibody libraries as maytansinoid antibody-drug conjugates", *mAbs*, vol. 8, no. 3, pp. 513–523, 2016.

[106] H. Schmidt-Traub, *Preparative chromatography of fine chemicals and pharmaceutical agents*. Wiley, 2005.

[107]   D. J. Karst, F. Steinebach, M. Soos, and M. Morbidelli, „Process performance and product quality in an integrated continuous antibody production process", *Biotechnology and Bioengineering*, vol. 114, no. 2, pp. 298–307, 2017.

[108]   A. Williams, E. K. Read, C. D. Agarabi, S. Lute, and K. A. Brorson, „Automated 2D-HPLC method for characterization of protein aggregation with in-line fraction collection device", *Journal of Chromatography B*, vol. 1046, pp. 122–130, 2017.

[109]   D. A. Parkins and U. T. Lashmar, „The formulation of biopharmaceutical products", *Pharmaceutical Science and Technology Today*, vol. 3, no. 4, pp. 129–137, 2000.

[110]   J. Morgenstern, M. Busch, P. Baumann, and J. Hubbuch, „Quantification of PEGylated proteases with varying degree of conjugation in mixtures: An analytical protocol combining protein precipitation and capillary gel electrophoresis", English, *Journal of Chromatography A*, vol. 1462, pp. 153–164, 2016.

[111]   A. Strancar, P. Raspor, H. Schwinn, R. Schutz, and D. Josic, „Extraction of triton-X-100 and its determination in virus-inactivated human plasma by the solvent detergent method", *Journal of Chromatography A*, vol. 658, no. 2, pp. 475–481, 1994.

[112]   E. E. Remsen and J. J. Freeman, „A size-exclusion chromatography/FT-IR (SEC/FT-IR) technique for improved FTIR spectroscopy of proteins in D_2O solution", *Applied Spectroscopy*, vol. 45, no. 5, pp. 868–873, 1991.

[113]   G. W. Somsen, C. Gooijer, and U. A. T. Brinkman, „Liquid chromatography-Fourier-transform infrared spectrometry", *Journal of Chromatography A*, vol. 856, no. 1-2, pp. 213–242, 1999.

[114]   J. Kuligowski, G. Quintás, M. de la Guardia, and B. Lendl, „Analytical potential of mid-infrared detection in capillary electrophoresis and liquid chromatography: A review", *Analytica Chimica Acta*, vol. 679, no. 1-2, pp. 31–42, 2010.

[115]   P. L. Roberts, „Virus elimination during the purification of monoclonal antibodies by column chromatography and additional steps", *Biotechnology Progress*, vol. 30, no. 6, pp. 1341–1347, 2014.

[116]   J. Morgenstern, P. Baumann, C. Brunner, and J. Hubbuch, „Effect of PEG molecular weight and PEGylation degree on the physical stability of PEGylated lysozyme", *International Journal of Pharmaceutics*, vol. 519, no. 1, pp. 408–417, 2017.

[117] K. Baumgartner, L. Galm, J. Nötzold, H. Sigloch, J. Morgenstern, K. Schleining, S. Suhm, S. A. Oelmeier, and J. Hubbuch, „Determination of protein phase diagrams by microbatch experiments: Exploring the influence of precipitants and pH", *International Journal of Pharmaceutics*, vol. 479, no. 1, pp. 28–40, 2015.

[118] G. Quintás, B. Lendl, S. Garrigues, and M. de la Guardia, „Univariate method for background correction in liquid chromatography-Fourier transform infrared spectrometry", *Journal of Chromatography A*, vol. 1190, no. 1-2, pp. 102–109, 2008.

[119] E. Pretsch, P. Bühlmann, C. Affolter, E. Pretsch, P. Bhuhlmann, and C. Affolter, *Structure determination of organic compounds.* Springer, 2009, vol. 13.

[120] J. E. Seely and C. W. Richey, „Use of ion-exchange chromatography and hydrophobic interaction chromatography in the preparation and recovery of polyethylene glycol-linked proteins", *Journal of Chromatography A*, vol. 908, no. 1, pp. 235–241, 2001.

[121] S. Yamamoto, S. Fujii, N. Yoshimoto, and P. Akbarzadehlaleh, „Effects of protein conformational changes on separation performance in electrostatic interaction chromatography: Unfolded proteins and PEGylated proteins", *Journal of Biotechnology*, vol. 132, no. 2, pp. 196–201, 2007.

[122] S. K. Hansen, B. Maiser, and J. Hubbuch, „Rapid quantification of protein-polyethylene glycol conjugates by multivariate evaluation of chromatographic data", *Journal of Chromatography A*, vol. 1257, pp. 41–47, 2012.

[123] A. P. Middelberg, T. Rivera-Hernandez, N. Wibowo, L. H. Lua, Y. Fan, G. Magor, C. Chang, Y. P. Chuan, M. F. Good, and M. R. Batzloff, „A microbial platform for rapid and low-cost virus-like particle and capsomere vaccines", *Vaccine*, vol. 29, no. 41, pp. 7154–7162, 2011.

[124] T. Klamp, J. Schumacher, G. Huber, C. Kühne, U. Meissner, A. Selmi, T. Hiller, S. Kreiter, J. Markl, Ö. Türeci, and U. Sahin, „Highly specific auto-antibodies against claudin-18 isoform 2 induced by a chimeric HBcAg virus-like particle vaccine kill tumor cells and inhibit the growth of lung metastases", *Cancer Research*, vol. 71, no. 2, pp. 516–527, 2011.

[125] N. Kushnir, S. J. Streatfield, and V. Yusibov, „Virus-like particles as a highly efficient vaccine platform: Diversity of targets and production systems and advances in clinical development", *Vaccine*, vol. 31, no. 1, pp. 58–83, 2012.

[126] M. F. Bachmann and P. Whitehead, „Active immunotherapy for chronic diseases", *Vaccine*, vol. 31, no. 14, pp. 1777–1784, 2013.

[127] L. H. Lua, N. K. Connors, F. Sainsbury, Y. P. Chuan, N. Wibowo, and A. P. Middelberg, „Bioengineering virus-like particles as vaccines", *Biotechnology and Bioengineering*, vol. 111, no. 3, pp. 425–440, 2014.

[128] D. C. Whitacre, B. O. Lee, and D. R. Milich, „Use of hepadnavirus core proteins as vaccine platforms", *Expert Review of Vaccines*, vol. 8, no. 11, pp. 1565–1573, 2009.

[129] U. Arora, P. Tyagi, S. Swaminathan, and N. Khanna, „Chimeric Hepatitis B core antigen virus-like particles displaying the envelope domain III of dengue virus type 2", *Journal of Nanobiotechnology*, vol. 10, no. 1, p. 30, 2012.

[130] D. C. Kaslow and S. Biernaux, „RTS,S: Toward a first landmark on the Malaria Vaccine Technology Roadmap", *Vaccine*, vol. 33, no. 52, pp. 7425–7432, 2015, Malaria Vaccines 2015.

[131] J. G. Low, L. S. Lee, E. E. Ooi, K. Ethirajulu, P. Yeo, A. Matter, J. E. Connolly, D. A. Skibinski, P. Saudan, M. Bachmann, B. J. Hanson, Q. Lu, S. Maurer-Stroh, S. Lim, and V. Novotny-Diermayr, „Safety and immunogenicity of a virus-like particle pandemic influenza A (H1N1) 2009 vaccine: Results from a double-blinded, randomized Phase I clinical trial in healthy Asian volunteers", *Vaccine*, vol. 32, no. 39, pp. 5041–5048, 2014.

[132] Y. P. Chuan, N. Wibowo, L. H. Lua, and A. P. Middelberg, „The economics of virus-like particle and capsomere vaccines", *Biochemical Engineering Journal*, vol. 90, pp. 255–263, 2014.

[133] O. S. Kumru, S. B. Joshi, D. E. Smith, C. R. Middaugh, T. Prusik, and D. B. Volkin, „Vaccine instability in the cold chain: Mechanisms, analysis and formulation strategies", *Biologicals*, vol. 42, no. 5, pp. 237–259, 2014.

[134] T. Vicente, A. Roldão, C. Peixoto, M. J. T. Carrondo, and P. M. Alves, „Large-scale production and purification of VLP-based vaccines", *Journal of Invertebrate Pathology*, vol. 107, no. Supplement, S42–S48, 2011.

[135] C. Ladd Effio and J. Hubbuch, „Next generation vaccines and vectors: Designing downstream processes for recombinant protein-based virus-like particles", *Biotechnology Journal*, vol. 10, no. 5, pp. 715–727, 2015.

[136] Y. Ren, S. M. Wong, and L. Y. Lim, „In vitro-reassembled plant virus-like particles for loading of polyacids", *Journal of General Virology*, vol. 87, no. 9, pp. 2749–2754, 2006.

[137] A. Link, F. Zabel, Y. Schnetzler, A. Titz, F. Brombacher, and M. F. Bachmann, „Innate Immunity Mediates Follicular Transport of Particulate but Not Soluble Protein Antigen", *The Journal of Immunology*, vol. 188, no. 8, pp. 3724–3733, 2012.

[138] H. Mach, D. B. Volkin, R. D. Troutman, B. Wang, Z. Luo, K. U. Jansen, and L. Shi, „Disassembly and reassembly of yeast-derived recombinant human papillomavirus virus-like particles (HPV VLPs)", *Journal of Pharmaceutical Sciences*, vol. 95, no. 10, pp. 2195–2206, 10/2006.

[139] Q. Zhao, M. J. Allen, Y. Wang, B. Wang, N. Wang, L. Shi, and R. D. Sitrin, „Disassembly and reassembly improves morphology and thermal stability of human papillomavirus type 16 virus-like particles", *Nanomedicine: Nanotechnology, Biology, and Medicine*, vol. 8, no. 7, pp. 1182–1189, 2012.

[140] Q. Zhao, Y. Modis, K. High, V. Towne, Y. Meng, Y. Wang, J. Alexandroff, M. Brown, B. Carragher, C. S. Potter, D. Abraham, D. Wohlpart, M. Kosinski, M. W. Washabaugh, and R. D. Sitrin, „Disassembly and reassembly of human papillomavirus virus-like particles produces more virion-like antibody reactivity", *Virology Journal*, vol. 9, no. 1, p. 52, 2012.

[141] R. T. Kurnik, A. W. Yu, G. S. Blank, A. R. Burton, D. Smith, A. M. Athalye, and R. Van Reis, „Buffer exchange using size exclusion chromatography, countercurrent dialysis, and tangential flow filtration: Models, development, and industrial application", *Biotechnology and Bioengineering*, vol. 45, no. 2, pp. 149–157, 1995.

[142] M. W. Jornitz and T. H. Meltzer, Eds., *Filtration and purification in the biopharmaceutical industry*, 2nd. CRC Press, 2008.

[143] B. J. Russell, J. O. Velez, J. J. Laven, A. J. Johnson, G. J. J. Chang, and B. W. Johnson, „A comparison of concentration methods applied to non-infectious flavivirus recombinant antigens for use in di-

agnostic serological assays", *Journal of Virological Methods*, vol. 145, no. 1, pp. 62–70, 2007.

[144] T. Vicente, S. Burri, S. Wellnitz, K. Walsh, S. Rothe, and J. Liderfelt, „Fully aseptic single-use cross flow filtration system for clarification and concentration of cytomegalovirus-like particles", *Engineering in Life Sciences*, vol. 14, no. 3, pp. 318–326, 2014.

[145] M. W. Liew, Y. P. Chuan, and A. P. Middelberg, „Reactive diafiltration for assembly and formulation of virus-like particles", *Biochemical Engineering Journal*, vol. 68, pp. 120–128, 2012.

[146] J. Hanemaaijer, T. Robbertsen, T. van den Boomgaard, and J. Gunnink, „Fouling of ultrafiltration membranes. The role of protein adsorption and salt precipitation", *Journal of Membrane Science*, vol. 40, no. 2, pp. 199–217, 1989.

[147] M. K. Ko, J. J. Pellegrino, R. Nassimbene, and P. Marko, „Characterization of the adsorption-fouling layer using globular proteins on ultrafiltration membranes", *Journal of Membrane Science*, vol. 76, no. 2, pp. 101–120, 1993.

[148] C. Peixoto, M. F. Q. Sousa, A. C. Silva, M. J. T. Carrondo, and P. M. Alves, „Downstream processing of triple layered rotavirus like particles", *Journal of Biotechnology*, vol. 127, no. 3, pp. 452–461, 2007.

[149] P. Roch and C.-F. Mandenius, „On-line monitoring of downstream bioprocesses", *Current Opinion in Chemical Engineering*, vol. 14, pp. 112–120, 2016.

[150] L. Hu, J. M. Trefethen, Y. Zeng, L. Yee, S. Ohtake, D. Lechuga-Ballesteros, K. L. Warfield, M. J. Aman, S. Shulenin, R. Unfer, S. G. Enterlein, V. Truong-Le, D. B. Volkin, S. B. Joshi, and C. R. Middaugh, „Biophysical Characterization and Conformational Stability of Ebola and Marburg Virus-Like Particles", *Journal of Pharmaceutical Sciences*, vol. 100, no. 12, pp. 5156–5173, 12/2011.

[151] S. F. Ausar, T. R. Foubert, M. H. Hudson, T. S. Vedvick, and C. R. Middaugh, „conformational stability and disassembly of norwalk virus-like particles: Effect of ph and temperature", *Journal of Biological Chemistry*, vol. 281, no. 28, pp. 19 478–19 488, 2006.

[152] S. J. Hanslip, N. R. Zaccai, A. P. Middelberg, and R. J. Falconer, „Assembly of Human Papillomavirus Type-16 Virus-Like Particles: Multifactorial Study of Assembly and Competing Aggregation“, *Biotechnology progress*, vol. 22, no. 2, pp. 554–560, 2006.

[153] B. Rajendar, V. Sivakumar, R. Sriraman, M. Raheem, R. Lingala, and R. V. Matur, „A simple and rapid method to monitor the disassembly and reassembly of virus-like particles“, *Analytical Biochemistry*, vol. 440, no. 1, pp. 15–17, 2013.

[154] M. Fang, W. Diao, B. Dong, H. Wei, J. Liu, L. Hua, M. Zhang, S. Guo, Y. Xiao, Y. Yu, L. Wang, and M. Wan, „Detection of the Assembly and Disassembly of PCV2b Virus-Like Particles Using Fluorescence Spectroscopy Analysis“, *Intervirology*, vol. 58, no. 5, pp. 318–323, 2016.

[155] S. Wynne, R. Crowther, and A. Leslie, „The Crystal Structure of the Human Hepatitis B Virus Capsid“, *Molecular Cell*, vol. 3, no. 6, pp. 771–780, 1999.

[156] C. G. Alexander, M. C. Jürgens, D. A. Shepherd, S. M. V. Freund, A. E. Ashcroft, and N. Ferguson, „Thermodynamic origins of protein folding, allostery, and capsid formation in the human hepatitis B virus core protein“, *Proceedings of the National Academy of Sciences*, vol. 110, no. 30, E2782–E2791, 2013.

[157] E. Gasteiger, C. Hoogland, A. Gattiker, M. R. Wilkins, R. D. Appel, A. Bairoch, *et al.*, „Protein identification and analysis tools on the ExPASy server“, in *The proteomics protocols handbook*, J. M. Walker, Ed., Springer, 2005, pp. 571–607.

[158] M. Smoluchowski, *Handbuch der Elektrizität und des Magnetismus.* Leipzig: Barth, 1921, vol. 2: Stationäre Ströme.

[159] D. E. Koppel, „Analysis of Macromolecular Polydispersity in Intensity Correlation Spectroscopy: The Method of Cumulants“, *The Journal of Chemical Physics*, vol. 57, no. 11, pp. 4814–4820, 12/1972.

[160] E. S. Page, „Continuous Inspection Schemes“, *Biometrika*, vol. 41, no. 1, pp. 100–115, 1954.

[161] O. A. Grigg, V. T. Farewell, and D. J. Spiegelhalter, „Use of risk-adjusted CUSUM and RSPRTcharts for monitoring in medical contexts“, *Stat. Methods Med. Res.*, vol. 12, no. 2, pp. 147–170, 04/2003.

[162] H. Mach and C. R. Middaugh, „Simultaneous monitoring of the environment of tryptophan, tyrosine, and phenylalanine residues in proteins by near-ultraviolet second-derivative spectroscopy.", *Analytical Biochemistry*, vol. 222, pp. 323–331, 1994.

[163] S. Großhans, M. Rüdt, A. Sanden, N. Brestrich, J. Morgenstern, S. Heissler, and J. Hubbuch, „In-line Fourier-transform infrared spectroscopy as a versatile process analytical technology for preparative protein chromatography", *Journal of Chromatography A*, vol. 1547, pp. 37–44, 04/2018.

[164] G. van den Berg and C. Smolders, „Flux decline in ultrafiltration processes", *Desalination*, vol. 77, pp. 101–133, 1990, Proceedings of the Symposium on Membrane Technology.

[165] I. H. Huisman, P. Prádanos, and A. Hernández, „The effect of protein–protein and protein–membrane interactions on membrane fouling in ultrafiltration", *Journal of Membrane Science*, vol. 179, no. 1, pp. 79–90, 2000.

[166] C. F. Bohren and D. R. Huffman, *Absorption and scattering of light by small particles*. Wiley-VCH, 2004.

[167] J. C. Thomas, „The determination of log normal particle size distributions by dynamic light scattering", *Journal of colloid and interface science*, vol. 117, no. 1, pp. 187–192, 1987.

[168] B. J. Berne and R. Pecora, *Dynamic light scattering: with applications to chemistry, biology, and physics*. Courier Corporation, 2000.

[169] A. B. Leung, K. I. Suh, and R. R. Ansari, „Particle-size and velocity measurements in flowing conditions using dynamic light scattering.", *Applied optics*, vol. 45, no. 10, pp. 2186–90, 2006.

[170] A. Zlotnick, J. M. Johnson, P. W. Wingfield, S. J. Stahl, and D. Endres, „A Theoretical Model Successfully Identifies Features of Hepatitis B Virus Capsid Assembly", *Biochemistry*, vol. 38, no. 44, pp. 14 644–14 652, 1999, PMID: 10545189.

[171] S. Mukherjee, M. V. Thorsteinsson, L. B. Johnston, P. A. DePhillips, and A. Zlotnick, „A Quantitative Description of In Vitro Assembly of Human Papillomavirus 16 Virus-Like Particles", *Journal of Molecular Biology*, vol. 381, no. 1, pp. 229–237, 2008.

[172] R. Crowther, N. Kiselev, B. Böttcher, J. Berriman, G. Borisova, V. Ose, and P. Pumpens, „Three-dimensional structure of hepatitis B virus core particles determined by electron cryomicroscopy", *Cell*, vol. 77, no. 6, pp. 943–950, 1994.

[173] B. Venkatakrishnan and A. Zlotnick, „The Structural Biology of Hepatitis B Virus: Form and Function", *Annual Review of Virology*, vol. 3, no. 1, pp. 429–451, 09/2016.

[174] C. R. Bourne, S. P. Katen, M. R. Fulz, C. Packianathan, and A. Zlotnick, „A mutant hepatitis b virus core protein mimics inhibitors of icosahedral capsid self-assembly", *Biochemistry*, vol. 48, no. 8, pp. 1736–1742, 2009.

[175] Y. Ding, Y. P. Chuan, L. He, and A. P. Middelberg, „Modeling the competition between aggregation and self-assembly during virus-like particle processing", *Biotechnology and Bioengineering*, vol. 107, no. 3, pp. 550–560, 10/2010.

[176] M. Li, P. Beard, P. A. Estes, M. K. Lyon, and R. L. Garcea, „Inter-capsomeric Disulfide Bonds in Papillomavirus Assembly and Disassembly", *J. Virol.*, vol. 72, no. 3, pp. 2160–2167, 1998.

[177] A. de Juan and R. Tauler, „Multivariate curve resolution (MCR) from 2000: Progress in concepts and applications", *Critical Reviews in Analytical Chemistry*, vol. 36, no. 3-4, pp. 163–176, 2006.

[178] C. Ruckebusch and L. Blanchet, „Multivariate curve resolution: A review of advanced and tailored applications and challenges", *Analytica Chimica Acta*, vol. 765, pp. 28–36, 2013.

[179] H. Parastar and R. Tauler, „Multivariate curve resolution of hyphen-ated and multidimensional chromatographic measurements: A new insight to address current chromatographic challenges", *Analytical Chemistry*, vol. 86, no. 1, pp. 286–297, 2014.

[180] O. S. Borgen and B. R. Kowalski, „An extension of the multivari-ate component-resolution method to three components", *Analytica Chimica Acta*, vol. 174, pp. 1–26, 1985.

[181] O. S. Borgen, N. Davidsen, Z. Mingyang, and Ø. Øyen, „The mul-tivariate n-component resolution problem with minimum assump-tions", *Microchimica Acta*, vol. 89, no. 1, pp. 63–73, 1986.

[182] B. Vandeginste, R. Essers, T. Bosman, J. Reijnen, and G. Kate-man, „Three-component curve resolution in liquid chromatography with multiwavelength diode array detection", *Analytical Chemistry*, vol. 57, no. 6, pp. 971–985, 1985.

[183] F. Dismer, S. Hansen, S. A. Oelmeier, and J. Hubbuch, „Accurate re-tention time determination of co-eluting proteins in analytical chro-matography by means of spectral data.", *Biotechnology and Bioengi-neering*, vol. 110, no. 3, pp. 683–93, 03/2013.

161

[184] K. Neymeyr, M. Sawall, and D. Hess, „Pure component spectral recovery and constrained matrix factorizations: Concepts and applications", *Journal of Chemometrics*, vol. 24, no. 2, pp. 67–74, 2010.

[185] M. Sawall, A. Börner, C. Kubis, D. Selent, R. Ludwig, and K. Neymeyr, „Model-free multivariate curve resolution combined with model-based kinetics: Algorithm and applications", *Journal of Chemometrics*, vol. 26, no. 10, pp. 538–548, 2012.

[186] W. H. Lawton and E. A. Sylvestre, „Self modeling curve resolution", *Technometrics*, vol. 13, no. 3, pp. 617–633, 1971.

[187] R. Tauler, „Application of non-linear optimization methods to the estimation of multivariate curve resolution solutions and of their feasible band boundaries in the investigation of two chemical and environmental simulated data sets", *Analytica Chimica Acta*, vol. 595, pp. 289–298, 2007.

[188] S. Arase, K. Horie, T. Kato, A. Noda, Y. Mito, M. Takahashi, and T. Yanagisawa, „Intelligent peak deconvolution through in-depth study of the data matrix from liquid chromatography coupled with a photodiode array detector applied to pharmaceutical analysis", *Journal of Chromatography A*, vol. 1469, pp. 35–47, 2016.

[189] Y. Kalambet, Y. Kozmin, K. Mikhailova, I. Nagaev, and P. Tikhonov, „Reconstruction of chromatographic peaks using the exponentially modified Gaussian function", *Journal of Chemometrics*, vol. 25, no. 7, pp. 352–356, 2011.

[190] A. A. Shukla and U. Gottschalk, „Single-use disposable technologies for biopharmaceutical manufacturing", *Trends in Biotechnology*, vol. 31, no. 3, pp. 147–154, 2013.

[191] G. R. Bolton, B. N. Violand, R. S. Wright, S. Sun, K. M. Sunasara, K. Watson, J. L. Coffman, C. Gallo, and R. Godavarti, „Addressing the challenges in downstream processing today and tomorrow", *BioPharm International*, vol. 24, no. 4, s8–s15, 2011.

Abbreviations

ADC Antibody Drug Conjugate viii, 14, 17, 49–51, 62, 63, 65, 118, 119, 123–125, 132–134, 138

ATR Attenuated Total Reflection vii, 12, 138

CEX Cation-Exchange vi, 35, 75, 77, 80, 84, 119, 124, 126

CFF Cross-Flow Filtration v, vii, xviii, 89–92, 95, 97, 98, 103, 109, 111, 114, 138

CHO Chinese Hamster Ovary 21

CPM 7-Diethylamino-3-(4'-Maleimidylphenyl)-4-Methylcoumarin 51–54, 56–64

CPP Critical Process Parameter v, vi

CQA Critical Quality Attribute v, vi, 1–3, 13

CV Cross-Validation 76

DAD Diode Array Detector vii, 23, 24, 52, 55, 56, 58, 64, 93, 95, 123, 125

DAR Drug-to-Antibody Ratio 50, 51, 62, 63

DHA (L)-Dehydroascorbic Acid 52, 53

DLS Dynamic Light Scattering vi, viii, 7, 10, 12, 14, 18, 92, 95, 96, 98, 101–103, 105, 106, 109, 110, 112

DMSO Dimethyl Sulfoxide 52, 53

DNA Deoxyribonucleic Acid v, 32, 112

DSP Downstream Processing v–vii, 2, 3, 11–14, 16, 17, 137, 139, 140

DV Diafiltration Volume 95, 100–102, 109, 110, 112

EMG Exponentially Modified Gaussian 117, 119, 121, 122, 125, 126, 128–130, 135

FDA Food and Drug Administration 1, 13, 32, 51

FTIR Fourier-Transform Infrared vi, vii, 7, 11, 12, 14, 16, 17, 33, 71–74, 76–80, 82, 83, 85–88, 137, 138

GUI Graphical User Interface 96, 98

HBcAg Hepatitis B core Antigen vii, 90, 92–94, 98, 107, 109–112, 114, 138

HBsAg Hepatitis B surface Antigen 90

HCCF Harvested Cell Culture Fluid 20, 22, 30, 137

HCP Host Cell Protein 12, 27, 73

HIC Hydrophobic-Interaction Chromatography 50, 119, 124, 132

HMW High Moleculare Weight Variant vi, 11, 12, 32–34, 37, 38, 43, 44, 47, 73, 137

HPLC High Performance Liquid Chromatography 2, 118, 135

HTPD High-throughput Process Development 65

IgG Immunoglobulin G 52, 64, 118, 119, 123–125, 132–135

IR Infrared vii, 73, 79, 80, 82, 84, 85, 140

LV Latent Variables 3–5, 24, 25, 27

mAb monoclonal Antibody vi–viii, 11, 12, 14, 16, 17, 19–30, 32–34, 37–39, 43–45, 47, 49–58, 60–62, 64, 65, 72, 73, 75, 79, 80, 82, 83, 85, 88, 125, 137, 138

MCR Multivariate Curve Resolution viii, 14–16, 18, 117–121, 126, 128–135, 138

MIR Major Immunodominant Region 93, 112

SEC Size-Exclusion Chromatography 38, 50, 90, 94, 95, 97, 100–103, 110, 119

SLS Static Light Scattering vi, viii, 7, 14, 18, 92, 96, 101–103, 106, 112

SOP Standard Operating Procedure 96

SVD Singular Value Decomposition 120, 121, 127, 128, 130, 132, 133

TCEP Tris(2-Carboxyethyl)phosphine Hydrochloride 52, 53

TFA Trifluoracetic Acid 56, 125

TMP Transmembrane Pressure viii, 90, 92, 95, 97–103, 109, 110, 112, 113

UF/DF Ultrafiltration/Diafiltration 90, 98, 111

UHPLC Ultra High Performance Liquid Chromatography 55, 56, 78, 95, 114

UV Ultraviolet 21–23, 31, 32, 35, 36, 41, 73, 82, 94, 100–102, 107, 109, 111, 112

UV/Vis Ultraviolet/Visible vi, vii, 7, 9–11, 14, 16–19, 21, 22, 26, 30, 33, 34, 47, 49–51, 54–56, 62, 64, 65, 72–74, 78, 88, 90, 92, 96–98, 103, 104, 111–113, 117, 119, 123, 126, 128, 130, 132, 135, 137, 138, 140

VLP Virus-Like Particle vii, viii, 14, 16, 18, 89–93, 95, 97–107, 109–113, 138

VP Variable Pathlength vi, 16, 31–34, 36, 39, 40, 43, 45, 47, 137