PAT AND AUTOMATION STRATEGIES FOR THE PROCESS DEVELOPMENT OF PEGYLATED PROTEINS

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
Adrian Simon Johannes Marius Sanden, M.Sc.
aus Mannheim, Deutschland

Tag der mündlichen Prüfung:	10.08.2020
Erstgutachter:	Prof. Dr. Jürgen Hubbuch
Zweitgutachterin:	Prof. Dr. Gisela Guthausen

Well, sometimes science is more art than science, Morty. A lot of people don't get that.

Rick Sanchez

The only difference between messing around and science is writing it down.

Adam Savage

Danksagung

Diese Arbeit wäre ohne die Beiträge einer großen Anzahl Menschen nicht in dieser Form realisierbar gewesen. Diesen möchte ich im Folgenden, nicht zwangsweise nach Wichtigkeit geordnet, meinen Dank aussprechen. Namen von Einzelpersonen nenne ich nur, wenn diese Personen sowieso in der Öffentlichkeit stehen, dies stellt keine Wertung da.

Prof. Dr. Jürgen Hubbuch für die Möglichkeit nach meiner Masterarbeit in der Arbeitsgruppe zu bleiben und weiter zu forschen. Sein Fokus auf das große Ganze und sein Gespür für interessante Themen waren immer hilfreich für mich. Ich habe die Freiheiten, welche die Arbeit an der Universität und sein Führungsstil bieten immer genossen. Seine Vorlesung im Bachelorstudium ist auch mitverantwortlich für mein ursprüngliches Interesse an Biopharmazeutischen Prozessen das sich in meinem von ihm vermittelten Praktikum bei CSL Behring verfestigt hat.

Prof. Dr. Gisela Guthausen danke ich für die investierte Zeit für die Bewertung meiner Arbeit als Korreferentin und die Verteidigung.

Während meiner Arbeit habe ich direkt mit Nina Brestrich, Steffen Großhans, Sandra Haas, Stefan Heissler, Josefine Morgenstern, Matthias Rüdt und Susanna Suhm zusammengearbeitet. Vielen Dank für die gute Zusammenarbeit und vor allem den Spaß im Labor und im Büro. Mein Büro mit Matthias und Laura war erfreulicherweise kein Büro mit Kollegen, sondern mit Freunden. Die gesamte Arbeitsgruppe am MAB hat sich in meiner Zeit vor allem durch den guten Zusammenhalt unter den Mitarbeitern ausgezeichnet. Danke für die Parties und alle anderen Freizeitaktivitäten! Und Danke, dass wir uns immer gegenseitig geholfen haben, fachlich und privat.

Die Gruppe von Bioingenieuren und Ehrenbioingenieuren und Assoziierten, mit denen ich den Großteil meines Studiums verbracht habe, hat auch ihren Beitrag geleistet. Danke für die gemeinsame Zeit, für fachliche Diskussionen zum Lernen oder in der Freizeit, weil es Spaß macht und für

alles andere. Im Zusammenhang dazu natürlich auch die Skifahrergruppe, mit der es deutliche Überlappungen gibt.

Thomas hat mir die Möglichkeit gegeben ein tolles neues Hobby auszuüben das mir oft geholfen hat meinen Kopf wieder frei zu bekommen. Ich habe sehr viel gelernt und immer Spaß dabei gehabt. Mein Dank gilt auch allen anderen Zwei- und Vierbeinern bei den Pferdefreunden Malsch. Meine vierbeinigen Freunde Lanthino, Lifestyle, Borax, Contador, Dark, Tango, Tassilo und Dorie und alle anderen meiner neuen befellten Freunde waren immer als Ausgleich zur Arbeit da. Danke an die Erwachsenen im Stall, für Ausritte im Wald, und für eure Geduld mir als Späteinsteiger immer etwas beizubringen, an die Jugendlichen dafür, dass ihr mich nicht einfach wie irgendeinen Erwachsenen behandelt habt und ich auch von euch viel lernen durfte.

Meine Eltern haben mich während meines gesamten Studiums auf jede erdenkliche Art unterstützt wodurch ich mir um Grundbedürfnisse zum Glück nie Sorgen machen musste. Danke, dass ihr mir immer geholfen habt wenn ich um Hilfe gefragt habe und mich trotzdem habt machen lassen, wenn ich das wollte. Dasselbe gilt natürlich für meine Schwester. Über meine Eltern hatte ich auch das Glück in San Diego eine Art zweites Zuhause zu finden. Alle die Dort leben haben immer dafür gesorgt, dass ein Besuch sich immer wie nach Hause kommen angefühlt hat. Meine Großeltern, Onkel, Tanten und Cousinen haben mir bei Gelegenheit auch unter die Arme gegriffen.

Zwei Informatiker und gute Freunde dürfen hier natürlich nicht fehlen, danke für die aktive und passive Hilfe bei allem was ich mir so vorgenommen habe.

Meiner Freundin möchte ich dafür danken, dass sie immer für mich da war, auch wenn die Arbeit mich mal mehr gestresst hat. Und für die Freiheit die sie mir gelassen hat Dinge anzugehen, wie ich es für richtig gehalten habe. Danke auch, dass du nie die Lust daran verloren hast mich in Tetris zu besiegen.

Abstract

Modern medicine keeps adopting an increasing amount of high molecular weight drugs like proteins that originate from biological systems. These biopharmaceutical drugs include therapies for several severe and fatal illnesses. Conditions targeted include multiple forms of cancer and conditions such as diabetes and hemophilia that usually require life-long medication. This is also reflected in the market share of such biopharmaceutical drugs. Out of the ten best selling pharmaceuticals in 2018, 8 are large biological molecules of which 6 are monoclonal antibodies (mAbs). Their combined sales exceeded 80 billion USD. [1] Relatively large bio-molecules such as mAbs, clotting factors or insulin can not be chemically synthesized in a sensible way. Therefore these molecules are produced by biological systems that more closely resemble their native origins. Mammalian or bacterial cells are genetically modified to produce large quantities of the target molecule in a cell culture fermenter. As a side-effect of the production in living organisms, a complex impurity profile has to be removed during the so called Downstream Processing (DSP) after clarification of the Harvested Cell Culture Fluid (HCCF). In combination with tedious identification of drug candidates and extensive clinical trials to get them approved by regulatory bodies, large investments are necessary to research, develop and produce these kinds of drugs.

The DSP is designed to separate unwanted product-related and process-related impurities to ensure consistent quality, efficacy and minimal side effects of the produced drugs. While most potential sources of process variations like temperature, pH and buffer composition can be controlled very well, the use of living organisms always introduces variability into the starting material for the purification process. Common unit operations in DSP include Cross-Flow Filtration (CFF), batch reactions and chromatographic separations. The ability to determine the system state and control the process accordingly is important to handle the variability in the starting material. Regulatory bodies including the Food and Drug Administration

(FDA) and European Medicines Agency (EMA) prefer knowledge based approaches for process design. In such a process following Quality by Design (QbD) principles, Process Analytical Technology (PAT) tools are utilized to measure Critical Quality Attributes (CQAs) and Critical Process Parameters (CPPs) to ensure the target Quality Target Product Profile (QTPP) will be met at the end of the process. This is especially true for very advanced protein based drugs where the parent protein is modified by covalently attaching Polyethylene Glycole (PEG) to it after production and purification. Despite the PEG molecule being relatively inert in biological systems there are clear benefits to attaching it to an existing protein drug. By being attached covalently, the hydrodynamic radius is increased, renal clearance is decreased and long term stability is improved among other benefits. Acquiring these benefits however comes at the price of a much more complicated production process. After the parent protein itself is produced and purified, the conjugation reaction itself takes place followed by a second set of purification steps to only produce the drug molecule with the right amount of PEG molecules attached to it. The reaction itself is generally run in batch mode and yields a complex mixture of proteins with different amounts of PEG attached as well as isoforms of each degree of PEGylation.

The aim of this thesis was to implement Fourier-Transform Infrared (FTIR) spectroscopy based PAT methods and automation strategies for the process development of PEGylated proteins. In order to utilize FTIR spectroscopy as a PAT tool a custom experimental setup was developed to enable near real time measurements of FTIR spectra in the effluent of a solid-liquid chromatography step. To this end, the spectrophotometer was integrated into the flow path of a lab scale chromatography system with a flow-through measurement accessory and custom software developed to trigger the measurements automatically. In addition to the MATLAB and OPUS-Script based software, additional communication hardware from National Instruments was utilized to facilitate the communication between MATLAB and the chromatography device. Additionally, the correction of spectral data recorded during ion exchange chromatography steps was evaluated. Water itself is very abundant in comparison to the target molecules, and even the salt used for gradient elutions is still multiple orders of magnitude more concentrated. Therefore water itself and changes in its absorbance caused by changing salt concentrations have to be accounted for to successfully extract information on the analytes from the spectral data. For the first implementation a full background run was subtracted from the data to account for changes in spectra caused by the rising salt concentration during elution. The automation setup for

reaction monitoring was also based on hardware from a lab scale chromatography device. Specifically, the fraction collector and sample pump were re-purposed and fitted with 3D-printed adapters to circulate the reaction solution and automatically collect samples into micro titer plates for later analysis.

After the successful implementation of FTIR as an in-line sensor, multiple case studies were conducted to evaluate its capabilities. The first two case studies were based around Multivariate Data Analysis (MVDA) in combination with spectral data. Lysozyme and a monoclonal antibody (mAb) were loaded onto a chromatography column and separated by a salt gradient elution. A Partial-Least Squares (PLS) regression model was calibrated based on reference analytics and used to estimate the concentration of the two species from the spectral data. The second case study was performed in a similar manner. Lysozyme was conjugated with PEG in random batch PEGylation and then separated by ion-exchange chromatography. Utilizing a similar PLS regression model, the concentration of the individual species with a varying degree of PEGylation was estimated in the column effluent. In both cases, the estimated concentrations based on the calibrated PLS model correlated closely to the offline analytics for all tested conditions. The general idea behind the first case study has been implemented before, while the second case study is only feasible with FTIR because PEG is not active in UV. To showcase the versatility of FTIR, a process related impurity in the form of Triton X-100 was monitored in the flow-through of a chromatography step with a simple linear regression instead of a multivariate regression model.

To expand on the second case study, an integrated ion-exchange chromatography step after on-column PEGylation was monitored using FTIR spectroscopy without multivariate regression by directly extracting information from the absorbance spectra. The parent protein was loaded onto the column and the PEGylation reaction conducted in place. The degree of PEGylation was then estimated in the effluent based solely on the ratio of the extinction coefficients of the protein and the PEG reagent, eliminating the need for labor-intensive calibration of a regression model. A more versatile approach to background correction employing Extended Multiplicative Signal Correction (EMSC) was also applied in this study. In comparison to the offline analytics, it could be shown that the degree of PEGylation of the reacted protein eluting from the column could be estimated with little prior knowledge. The initial work needed to obtain the extinction coefficients was much lower than the previous method of performing multiple experiments and corresponding offline analytics to calibrate a PLS regression model.

The PEGylation reaction in batch mode was the focus of another study. Automatic circulation of the solution through the FTIR device was combined with automatic sampling into 96 well plates by modifying the flow path of an ÄKTApurifier System with custom 3D-printed adapter parts. It was shown that the automatically acquired samples from batches run with different conditions can be used to calibrate a kinetic reaction model with small confidence intervals. In addition, the possibility to utilize the spectral data recorded during the batch reaction to estimate the state of the reaction was also investigated.

Zusammenfassung

In der modernen Medizin werden vermehrt Medikamente eingesetzt die auf Molekülen mit hohem Molekulargewicht basieren die ihren Ursprung in biologischen Systemen haben. Diese biopharmazeutischen Medikamente werden unter anderem für verschiedene schwerwiegende und tödliche Erkrankungen eingesetzt. Darunter befinden sich verschiedene Formen von Krebs und chronische Erkrankungen wie Diabetes und Hämophilie die in der Regel lebenslange Medikation erfordern. Dies zeigt sich auch im Marktanteil dieser Medikamente. Acht der zehn meistverkauften Medikamente 2018 waren große biologische Moleküle wovon wiederum sechs monoklonale Antikörper (mAb) waren. Insgesamt erzielten diese zehn Medikamente einen Umsatz von mehr als 80 Milliarden USD. [1] Diese relativ großen Moleküle wie Antikörper, Gerinnungsfaktoren oder Insulin können nicht sinnvoll auf traditionelle Art chemisch synthetisiert werden. Um dieses Problem zu umgehen, werden stattdessen biologische Systeme zur Produktion verwendet, um die natürliche Synthese zu imitieren. Dazu werden gentechnisch veränderte Bakterien oder Säugerzellen in großen Fermentern produziert um die Zielprodukte zu exprimieren. Als Nebeneffekt der Produktion in komplexen biologischen Systemen wie lebenden Zellen entsteht allerdings ein sehr komplexes Profil an Verunreinigungen die im sogenannten Downstream Processing (DSP) aus der geklärten Zellkulturflüssigkeit (HCCF) entfernt werden müssen. In Kombination mit der unabhängig davon arbeitsintensiven Identifikation von Kandidatenmolekülen und langwierige klinischen Studien als Voraussetzung zur Zulassung durch die Behörden sind dadurch sehr große finanzielle Investitionen notwendig um derartige Medikamente zu erforschen, entwickeln und zu produzieren.

Der DSP wird so entworfen, dass ungewünschte produktbezogene und prozessbezogene Verunreinigungen abgetrennt werden um eine konsistente Qualität, Wirksamkeit und geringe Nebenwirkungen des Produktes zu gewährleisten. Während einige potentielle Quellen für Prozessvarianz

wie Temperatur, pH Wert und Pufferzusammensetzung relativ einfach zu kontrollieren sind, führt die Verwendung von lebenden Organismen in der Produktion immer zu variablen Ausgangsmaterialien im Aufreinigungsprozess. Der Aufreinigungsprozess besteht in der Regel aus einer Aneinanderreihung von Standardschritten wie Querstromfiltration (CFF), chemischen Reaktionen und Chromatographieschritten. Die Fähigkeit den Status des Prozesses zu bestimmen und den Prozess entsprechend zu steuern ist wichtig um die Variation im Ausgangsmaterial auffangen zu können. Regulierungsbehörden wie die Food and Drug Administration (FDA) und European Medicines Agency (EMA) bevorzugen solche wissensbasierten Ansätze für die Prozessentwicklung. Solche Prozesse folgen den Prinzipien von intrinsischer Produktqualität oder Quality by Design (QbD) und verwenden Prozessanalytische Technologien (PAT) um die kritischen Qualitätsattribute und kritischen Prozessparameter Critical Process Parameter (CPP) zu messen, um das Erreichen des gewünschten Produktprofils (QTPP) am Ende des Prozesses sicherzustellen. Dies trifft in besonderem Maße auf spezielle Formate von Proteinbasierten Medikamenten zu bei denen zusätzlich noch ein weiteres Molekül wie Polyethylenglykol (PEG) kovalent an das produzierte Protein gebunden wird nachdem es aufgereinigt wurde. Obwohl PEG selbst sich in biologischen Systemen relativ inert verhält kann es deutliche Vorteile haben ein therapeutisches Protein mit einem kovalent gebundenen PEG zu versehen. Der hydrodynamische Radius wird größer, die Ausscheidung durch die Niere wird verlangsamt und die Stabilität erhöht. Diese Vorteile stehen allerdings am Ende eines verkomplizierten Aufreinigungsprozesses. Nachdem das therapeutische Molekül produziert und aufgereinigt wurde wird die Konjugationsreaktion gefolgt von einer weiteren Reihe von Aufreinigungsschritten durchgeführt um das Protein mit der richtigen Anzahl konjugierter PEG Moleküle zu erhalten. Die Reaktion wird in der Regel in einem Rührkesselreaktor durchgeführt und es entstehen Proteine mit verschiedenen PEGylierungsgraden und Isoformen der jeweiligen Spezies.

Die Zielsetzung dieser Arbeit war es Fourier-Transform Infrared (FTIR) Spektroskopie als Prozessanalytische Technologie (PAT) und Automatisierungsstrategien für die Prozessentwicklung von PEGylierten Proteinen zu implementieren. Um FTIR Spektroskopie als PAT Methode verwendbar zu machen wurde zunächst ein experimenteller Aufbau entwickelt um direkt am Säulenausgang einer Chromatographieanlage Spektraldaten in Beinahe-Echtzeit als in-line Messung aufzunehmen. Dazu wurde das Spektrometer mithilfe einer speziellen Durchflusszelle in den Flussweg integriert und Software entwickelt, um die Messungen automatisiert zu starten. Neben MATLAB und OPUS-Skript basierter software wurde dazu

ein externes Kommunikationsgerät von National Instruments verwendet um den Informationsaustausch mit der Chromatographieanlage zu ermöglichen. Zusätzlich wurden Methoden zur Korrektur der Spektraldaten aus Ionentauschchromatographieexperimenten untersucht. Da Wasser in solchen Experimenten in großen Konzentrationen vorliegt hat das aufgeben eines Salzgradienten einen signifikanten Einfluss auf das Spektrum. Obwohl die Salzkonzentration deutlich geringer ist als die des Wassers ist sie mehrere Größenordnungen höher als die der Analyte in Lösung. Als erste Implementierung zur Hintergrundkorrektur wurden Spektraldaten eines Blindlaufs abgezogen um den Einfluss der sich ändernden Salzkonzentration auszugleichen. Der experimentelle Aufbau für die Automatisierung der Reaktionsexperimente basierte ebenfalls auf einer Chromatographieanlage. Dabei wurden der Fraktionierer und die Probenaufgabepumpe mit 3D-gedruckten Adaptern versehen um die Reaktionslösung zu zirkulieren und die automatisierte Probennahme im Mikrotitermaßstab zu ermöglichen.

Nach der erfolgreichen Implementierung der FTIR Sepktroskopie als in-line Messung wurden mehrere Fallstudien durchgeführt. Die ersten beiden Studien basieren auf der Anwendung von Multivariater Datenanalyse (MVDA) in Kombination mit Spektraldaten. Lysozym und ein Monoklonaler Antikörper (mAb) wurden durch Gradientenelution auf einer Chromatographiesäule getrennt. Mithilfe der Spektraldaten und Offlineanalytik wurde ein PLS Regressionsmodell kalibriert um die Konzentration der beiden Spezies im Prozessstrom basierend auf den Spektraldaten abzuschätzen. Die zweite Fallstudie wurde nach demselben Prinzip durchgeführt. Lysozym wurde in einem Rührkesselreaktor mit PEG konjugiert und durch Ionentauschchromatographie aufgereinigt. Ebenfalls basierend auf einem PLS Regressionsmodell wurde die Konzentration der unterschiedlich PEGylierten Spezies am Säulenausgang abgeschätzt. In beiden Studien korrelierten die Modellantworten mit den Ergebnissen der Referenzanalytik und zeigten nur geringe Abweichungen. Die Grundidee hinter der ersten Fallstudie wurde schon in früheren Arbeiten umgesetzt wobei die zweite Studie speziell von FTIR profitiert, da PEG nicht UV-aktiv ist. Um die Vielseitigkeit von FTIR als PAT Methode zu zeigen wurde in einer weiteren Fallstudie die Konzentration einer Prozessbezogenen Verunreinigung in Form von Triton X-100 im Durchfluss während der Beladung einer Chromatographiesäule überwacht, indem eine lineare Regression auf die Spektraldaten angewendet wurde anstatt ein multivariates Regressionsmodell zu kalibrieren.

In thematischer Fortführung der zweiten Fallstudie wurde die integrierte Festphasen-PEGylierung und Aufreinigung durch dieselbe Chromato-

graphiesäule mit FTIR ohne die Verwendung multivariater Regressionsmethoden mithilfe der Absorptionsspektren überwacht. Das zu PEGylierende Protein wurde auf der Säule gebunden und dann im gebundenen Zustand mit dem Konjugationsreagenz in Kontakt gebracht. Der PEGylierungsgrad der eluierenden Spezies wurde dann auf Basis des Verhältnisses der Extinktionskoeffizenten des Proteins und des PEG-Reagenz abgeschätzt. Dieses Vorgehen ist erheblich weniger arbeitsintensiv als das Durchführen der Experimente um ein Regressionsmodell zu kalibrieren. Des Weiteren wurde in dieser Studie ein vielseitigerer Ansatz zur Hintergrundkorrektur basierend auf Extended Multiplicative Signal Correction (EMSC) verwendet. Im Vergleich mit der Referenzanalytik konnte gezeigt werden, dass der PEGylierungsgrad mit vergleichsweise geringem Vorwissen abgeschätzt werden konnte. Der Aufwand um die Extinktionskoeffizienten zu bestimmen war dabei deutlich niedriger, als der experimentelle Aufwand um Regressionsmodelle wie in den vorhergehenden Studien zu kalibrieren.

Die PEGylierungsreaktion in Form eines Chargenprozesses wurde in einer weiteren Studie betrachtet. Die automatische Zirkulation der Prozesslösung durch das FTIR Spektrometer wurde in Kombination mit der automatisierten Probennahme in Mikrotiterplatten realisiert. Um dies zu erreichen wurde der Flusspfad einer Chromatographieanlage mit selbst entworfenen 3D-gedruckten Elementen verändert. Dabei wurde gezeigt, dass die mit verschiedenen Prozessbedingungen erzeugten Proben verwendet werden können um ein Kinetikmodell mit geringen Konfidenzintervallen zu kalibrieren. Des weiteren wurde untersucht inwieweit sich die Spektraldaten verwenden lassen um den Aktuellen Prozessfortschritt abzuschätzen.

Contents

Danksagung **iii**

Abstract **v**

Zusammenfassung **ix**

Contents **xiii**

1 Introduction **1**
1.1 Process Analytical Technology 3
1.2 Multivariate Data Analysis 4
1.2.1 Partial Least Square Regression 4
1.3 Spectroscopy 5
1.3.1 FTIR Spectroscopy 6
1.4 Protein Conjugation 12
1.4.1 Production of PEGylated Proteins 13
1.4.2 Conjugation reactions 13

2 Thesis Outline **17**
2.1 Research Proposal 17
2.2 Overview and Author Statement 20

3 In-line Fourier-Transform Infrared Spectroscopy as a Versatile Process Analytical Technology for Preparative Protein Chromatography **25**
Steffen Großhans[*,1], Matthias Rüdt[*,1], Adrian Sanden[*,1], Nina Brestrich[1], Josefine Morgenstern[1], Stefan Heissler[2], Jürgen Hubbuch[1] ([*] contributed equally)
3.1 Introduction 26
3.2 Materials and Methods 28
3.2.1 Experimental Setup 28

3.2.2 Proteins and Buffers 29
3.2.3 Preparative Chromatography Experiments 30
3.2.4 Analytical CEX Chromatography 31
3.2.5 Data analysis 32
3.3 Results and Discussion 33
3.3.1 Background Subtraction and Spectral Preprocessing 33
3.3.2 Case Study I: Selective Protein Quantification 36
3.3.3 Case Study II: Separation of PEGylated Lysozyme Species 38
3.3.4 Case Study III: Quantification of a Process Related Impurity 40
3.4 Conclusion and Outlook 41

4 Fourier- Transform Infrared Spectroscopy as a Process Analytical Technology for Near Real Time In-line Estimation of the Degree of PEGylation in Chromatography 43

Adrian Sanden, Susanna Suhm, Matthias Rüdt, Jürgen Hubbuch

4.1 Introduction 44
4.2 Materials and Methods 45
4.2.1 Experimental Setup 45
4.2.2 Proteins and Buffers 46
4.2.3 On-column-PEGylation of lysozyme 46
4.2.4 Analytical CEX Chromatography 47
4.2.5 Spectral data correction by EMSC-ALS 47
4.3 Results and Discussion 50
4.3.1 On-column-PEGylation 50
4.3.2 Spectral data correction by EMSC-ALS 51
4.3.3 Correction of spectral fraction with the ratio of extinction coefficients and offset 53
4.3.4 Comparison to off-line analytics 53
4.4 Conclusion and Outlook 54

5 Modifying an ÄKTApurifier System for the Automated Acquisition of Samples for Kinetic Modeling of Batch Reactions 57

Adrian Sanden, Sandra Haas, Jürgen Hubbuch

5.1 Introduction 58
5.2 Materials and Methods 59
5.2.1 Experimental Setup 59
5.2.2 Proteins and Buffers 60
5.2.3 PEGylation of Lysozyme 62

5.2.4 Analytical CEX Chromatography 62
5.2.5 Data Analysis and Kinetic Reaction Model 62
5.3 Results and Discussion 63
5.3.1 Kinetic Modelling of the PEGylation Reaction . . . 63
5.4 Conclusion and Outlook 64

6 ATR-FTIR for *in situ* Reaction Monitoring of Protein PEGylation in Batch Mode 67

Adrian Sanden, Sandra Haas, Jürgen Hubbuch

6.1 Introduction 68
6.2 Materials and Methods 69
6.2.1 Experimental Setup 69
6.2.2 Batch PEGylation experiments 70
6.2.3 Reference Analytics 71
6.2.4 Data Analysis 72
6.3 Results and Discussion 72
6.3.1 PLS Regression without Baseline Correction 73
6.3.2 PLS Regression with Baseline Correction 77
6.3.3 Varying the external calibration set 81
6.4 Conclusion 82
6.5 Outlook 83

7 General Discussion and Conclusion 85

8 Outlook 89

References 91

Abbreviations 107

1

Introduction

Biopharmaceuticals are a class of medicines that is based on large biological molecules, as opposed to traditional small molecule drugs. They are not synthesized in a traditional chemical way, but by microorganisms or other cells fermented at large scale and purified from the HCCF. The largest group of commercial biopharmaceuticals by sales are the monoclonal antibodies. [2] In late 2014, 47 mAb products were approved by the FDA and EMA [3], with 68 monoclonal antibodies approved between 2014 and 2018 alone [4, 5], making this class of therapeutics one of the fastest growing segments of the market. [6] Most mAbs have multiple target diseases with at least one cancer target for most of them. Other important biopharmaceuticals include Insulin, coagulation factors, vaccines and growth hormones, many of which target potentially fatal diseases. [6, 7] Between all of these, biopharmaceuticals amounted for total sales of USD 228 billion globally in 2016. [8] With new highly expensive personalized medicines such as Novartis's Kymriah entering the market, this number is bound to keep rising for the forseeable future. [9]

A typical production process for a biopharmaceutical consists of fermentation, harvest, clarification, purification, virus inactivation, formulation and final filling steps. While the exact unit operations, their order and number are dependent on the target molecule, a rough outline of such a process is shown in figure 1.1, using the mAb platform process as an example. [10] The separation between upstream processing and downstream processing is made based on the presence of cells in the process stream, so the downstream side of the process starts after the HCCF has been separated from the cells. During the DSP two groups of impurities

have to be removed from the target product. Product-related impurities include aggregates, fragments, isoforms and other species that are made up of parts of the target product. They are challenging to remove because they may exhibit very similar physicochemical properties as the product itself. Process-related impurities are made up of all molecules that are needed to run the process but must not be present in the final product. These include excipients, buffer substances, detergents and many more.

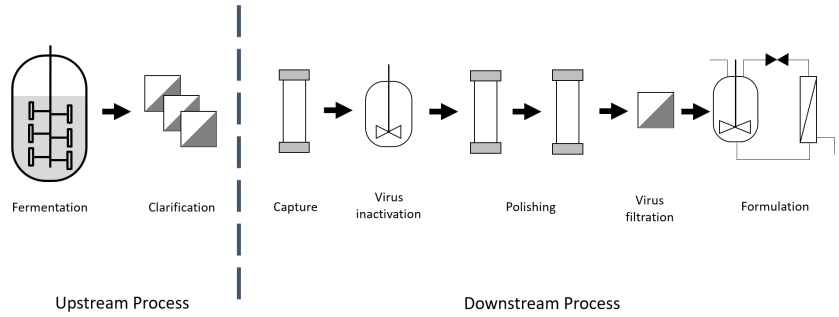

Figure 1.1: The outline of a typical biopharmaceutical production process for a mAb. For different modalities the process may look different, but the general steps of fermentation, purification and formulation are universally applicable.

Unfortunately, these unit operations can practically not be run one after another by just routing the fluid stream from one to the next. Most biopharmaceutical processes are still run in batch mode, which means that each unit operation is conducted somewhat independently. [11] This means that the input material for every step is fed into it, processed, and all of the output collected in a holding tank to ensure homogeneous feed material for the following unit operation. Subsequent steps are then run in the same fashion. [11] Switching from batch mode to continuous manufacturing is an ongoing effort. On the one hand this means that unit operations that are inherently run as batch steps such as bind-and-elute chromatography need to be replaced by multi column setups such as Simulated Moving Bed Chromatography (SMB) or Periodic Counter-Current Chromatography (PCC). [12, 13] On the other hand, the definition of what constitutes a single batch is not trivial when dealing with continuous processes operated in a steady state. [11]

The mAb process is called a platform process because of the outstandingly efficient protein A affinity chromatography step. It is generally used as the first step after clarification and capable of achieving purities of more than 98 % in a single unit operation. [14] To ensure that the entire process

can run as intended, it is ciritical that the conditions at the end of one process step match the desired conditions to ensure that the next process step can proceed as intended. Apart from ensuring that the process can run as intended, it is also necessary to ensure that the product will meet the required specifications, also called the QTPP. The QTPP is the set of CQAs that makes up the entire definition of the products target properties to ensure safety and efficacy. In order to assess whether the process proceeding as designed, analytical methods need to be used to extract information on the state of the process. To this end a plethora of methods has been employed. Two main groups can be distinguished when it comes to process analytics: in-line measurements and off-line measurements. In-line sensors such as pH and conductivity probes and a multitude of spectroscopic sensors offer the advantage that they can provide measurements in near real-time. These sensors are the method of choice when it comes to implementing PAT methods in a process.

1.1 Process Analytical Technology

PAT is a blanket term for multidisciplinary technologies that are used in production processes to gather information on the state of the process. As such, it is closely linked to the definition of CQAs and the implementation of QbD. [15] While the general approach has been around for quite some time, the PAT initiative by the FDA in 2004 focused and intensified the efforts made by the pharmaceutical industry and academia to research and implement these methods. For a long time, the product titers in upstream cell culture were incresing steadily. Some monoclonal antibodies are now even approaching the practical limits of solubility. [16] High titers and high volumes, both driven by the demand for more final product create the need for efficient and stable processes with as little wasted active pharmaceutical ingredient (API) as possible. Processes that implement both QbD and PAT can meet both of these requirements.

Apart from molecules that are used as biopharmaceuticals in the state that they are produced in by the recombinant cells, there is also a class of biopharmaceuticals that is modified by covalent attachment of additional molecules. One of these molecules is PEG and at least a dozen molecules that have been modified by covalent attachment of PEG are available on the market right now. [17] While PEG is mostly inert on its own in the human body, it has several benefits for the drugs it is attached to. PEGylated molecules experience reduced renal clearance which increases their circulation time, higher stability, reduced immunogenicity and increased

activity. [18] The PEGylation process itself is always carried out with the previously produced and purified therapeutic protein to reduce validation and separation challenges. Consequently, in order to produce a PEGylated therapeutic protein, a process to purify the protein itself and to separate the desired PEGylated species have to be developed. [19]

1.2 Multivariate Data Analysis

The ability and desire to conduct MVDA on more and more problems has been strongly linked with the rise of faster, smaller and more affordable computers. [20] MVDA applies multivariate mathematical methods to process data. The extraction of chemical information by such data-driven means is also referred to as chemometrics. [21] In biopharmaceutical production processes, multivariate data is most often present in the form of spectral data. Spectral data is preferred over single wavelength measurements as univariate measurement methods can often not capture all valuable information from a process. [15] An analytical tool that utilizes a multivariate model to extract information from the data obtained by a measurement device in the process is also referred to as a soft sensor. [22, 23]

1.2.1 Partial Least Square Regression

A powerful and widely used tool for MVDA is PLS regression. Especially when it comes to spectroscopic data PLS regression has manifested itself as the standard method of choice. [24] The first version of this algorithm was developed for economics data and published in 1974 and 1975. [25, 26] PLS regression is related to Principal Component Analysis (PCA) with the significant difference being that the principal components of the $X - data$ are defined under utilization of the correlating $Y - data$, i.e. the first latent variables are selected in such a way that they explain the greatest amount of variance in Y with the following variables chosen orthogonally to the previous ones and so that they explain the most of the remaining variance. An abstract representation of a PLS regression is shown in Figure 1.2.

The regression matrix B stores the calibrated model, its amount of columns is the amount of latent variables. The mathematical limit for the amount of latent variables is the number of observations, however in practice the number should be kept as small as possible to avoid overfitting. In short, the amount of latent variables has to be chosen in such a way that as much of the actual information in the data and as little of the noise as

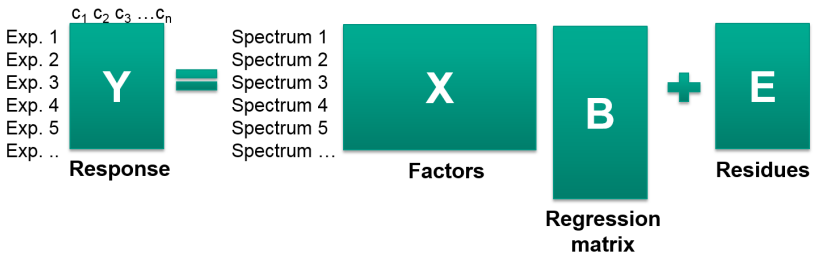

Figure 1.2: Illustration of the correlation between the original data and the data captured by the PLS regression model. The response and factors are needed to calibrate the model, resulting in the regression matrix and residual data. Responses for additional factors can then be estimated with the regression matrix. Here, the columns of the response matrix are denoted as concentrations and the rows of the factors matrix as spectra to indicate a common configuration for the usage of PLS regression in PAT.

possible is captured in the regression matrix. [24] After calibration, the regression matrix can be used to estimate an additional set of responses from an additional set of factors. It should be noted that this calculation can be performed on any data that has the adequate dimensions – it is the users responsibility to determine wether the data is of the same kind as the calibration data and the model's capabilities for potential extrapolation.

1.3 Spectroscopy

All absorption spectroscopy methods that are used for concentration related analysis rely on the assumption that the Beer-Lambert law [27] is applicable. For spectroscopic applications, it is commonly expressed in terms of concentration as in equation 1.1 where the Absorbance A_λ is expressed as a function of the extinction coefficient at the respective wavelength ϵ_λ, the concentration c and the path length d. A_λ is the logarithmic attenuation of the intensity of the light passing through the sample as shown in equation 1.2. As such, the limit of detection is defined by the ability of the detector to distinguish changes in intensity from background noise, which makes measurements non-linear for highly concentrated solutions, i.e. solutions where very little light actually makes it to the detector. The same is true for very dilute solutions, where I and I_0 are so similar that a difference can barely be measured. The concentration at this point is called limit of detection (LOD). In practice however, the limit of quantitation (LOQ) is usually more relevant. [28] Naturally, the extinction

coefficient c is only valid in the linear range of absorption. Additionally, an absorption coefficient has to be determined for each wavelength individually. Most spectrophotometers either use a monochromatic lightsource or a polychromatic lightsource in conjunction with a monochromator or a Diode Array Detector (DAD) to measure the absorbance of a sample at an individual wavelength.

$$A_\lambda = \epsilon_\lambda c d \tag{1.1}$$
$$A_\lambda = -log(I/I_0) \tag{1.2}$$

Different spectroscopic methods can be categorized by the wavelengths of light that they utilize and if the measurement mode is based on absorption, scattering or excitation and emission. Depending on the specific properties of the analytes at hand, the adequate spectroscopic method should be chosen. Two common spectroscopic methods for biomolecules are Ultraviolet/Visible (UV/Vis) and FTIR spectroscopy. One factor in making such a choice may be that certain molecules only absorb in one spectroscopic method but not the other, or do not absorb enough light to be easily detectable with available instruments. One example of this is PEG, which does not absorb in UV/Vis. [29] While not a biomolecule itself, it is often used as a conjugate to proteins or as an excipient in drug formulations. Some methods produce spectral data that contains so much specific information that they can even be used to elucidate the chemical structure of an analyte utilizing reference tables such as the collection curated by Pretsch, Bühlmann and Badertscher. [30] Apart from the choice of spectroscopic method multiple practical considerations have to be made when it comes to in-process spectroscopy. Some spectrophotometers being sensible to vibration, inability to be safely operated in explosive atmospheres without specific safety measures or the need to be flushed with dry air are just some of the challenges that can occur.

1.3.1 FTIR Spectroscopy

FTIR is a spectroscopy method that uses an interferrogram of polychromatic infrared light as the light source. A globar, a semiconductor that is brought to a glowing temperature using electricity, is used as the emitter of infrared light. The light is routed through an interferometer, an optical instrument to create interference in a controlled manner. The Michelson Interferometer is an easy to understand implementation of such an apparatus as shown in Figure 1.3. [31] In the beam-splitter, the light beam is

split into two parts, one of which is directed towards a stationary and one towards a moving mirror. The two beams are then reunited and directed through the sample and towards the detector, while the movement of the mirror moves the two parts of the original beam out of phase and therefore causes interference.

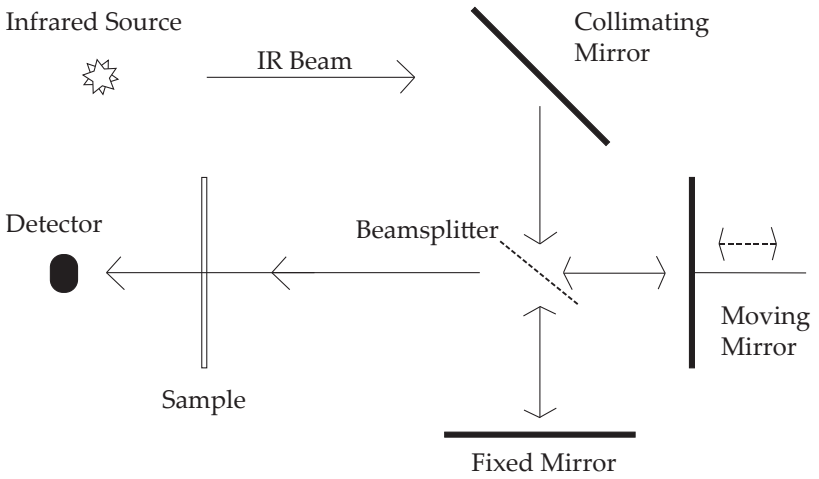

Figure 1.3: The schematic structure of a Michelson Interferometer with one fixed length and one variable length light path, adapted from [32].

The light that passes through the sample can then interact with the molecules that are present in the sample. The interferogram that is then recorded by the detector is a signal in the displacement-intensity domain. As opposed to other spectroscopic methods, the full intensity of the light source less the absorbance caused by the sample can be recorded to the detector, without being obstructed by a monochromator. This is called the *Jacquinot's Advantage* and goes hand in hand with *Fellgett's Advantage* which explains the better signal-to-noise ratio (SNR) of FT spectrometers over dispersive spectrometers due to the full bandwith of light being detected simultaneously, enabling the averaging of multiple scans of the mirror into one spectrum. [33] The Fourier Transform is then performed on the recorded signal to convert it from the displacement-intensity domain to the wavelength-intensity domain. [34] The interferogram $I(x)$ (with x being the position of the mirror, see Figure 1.3) and spectrum $A(\tilde{\nu})$ ($\tilde{\nu}$ being the wavenumber, the inverse of the wavelength as shown in equation 1.3) are linked by the Fourier Transform and its inverse as shown in equations 1.4 and 1.5. [35]

$$\tilde{\nu} = \frac{1}{\lambda} \tag{1.3}$$

$$I(x) = \int_{-\infty}^{\infty} A(\tilde{\nu}) \exp(-2\pi i \tilde{\nu} x) d\tilde{\nu} = F^{-1}\{A(\tilde{\nu})\} \tag{1.4}$$

$$A(\tilde{\nu}) = \int_{-\infty}^{\infty} I(x) \exp(2\pi i \tilde{\nu} x) dx \quad = F\{I(x)\} \tag{1.5}$$

In equations 1.4 and 1.5 $F^{-1}\{\}$ is the inverse Fourier Transform and $F\{\}$ is the Fourier Transform, i is the imaginary unit. In practice, values that x can attain are limited by the possible pathlength along which the mirror can travel. This limits the dimensions a recorded interferogram can have making it also impossible to record a real continuous interferogram. Therefore, the Fourier pair is replaced by a discretization with a finite number of points of the interferogram N with a distance Δx between them. [36] In the differential equations, x is replaced by $n\Delta x$, $\tilde{\nu}$ by $k\Delta\tilde{\nu}$ and summarized over all supporting points to get to the discrete fourier transformation as shown in equations 1.7 and 1.8. [36] Today, most calculations of discrete Fourier Transforms are calculated using Fast Fourier Transformation (FFT), a class of algorithms based on the one proposed by J. Cooley and J. Tukey in 1965 [37] and are usually handled by the measurement software in a way transparent to the user. The distance between calculated points is calculated according to equation 1.6.

$$\Delta\tilde{\nu} = \frac{1}{N\Delta x} \tag{1.6}$$

$$A(k\Delta\tilde{\nu}) = \sum_{n=0}^{N-1} I(n\Delta x) \exp(\frac{2\pi i n k}{N}) \Delta x \tag{1.7}$$

$$I(n\Delta x) = \sum_{k=0}^{N-1} A(k\Delta\tilde{\nu}) \exp(\frac{-2\pi i n k}{N}) \Delta\tilde{\nu} \tag{1.8}$$

Summarizing, to obtain an absorbance spectrum, interferograms of both a background and a sample are recorded and transformed to single beam spectra by the Fourier Transformation. The ratio of the two results in a transmission spectrum which can then be converted to an absorbance spectrum using the negative decadic logarithm. This process is illustrated in Figure 1.4. Most modern spectrophotometers make this entire process mostly transparent to the user.

FTIR spectroscopy can be implemented in three main modes of measurement, i.e. transmission, reflectance and Attenuated Total Reflectance

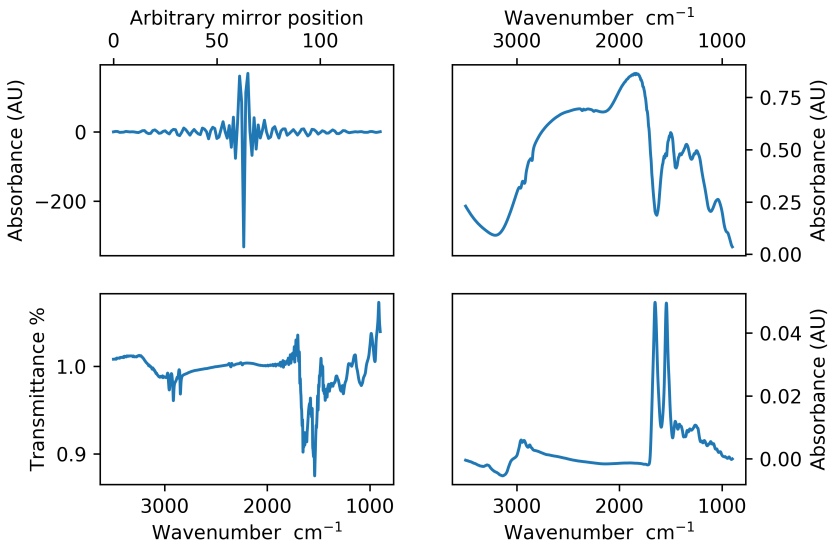

Figure 1.4: Illustration of the correlation between the interferogram and the final absorbance spectrum. The interferogram (top left) gets transformed to single chanel spectra (top right), the ratio between the sample spectrum and the background spectrum is the transmittance (bottom left) and the negative decadic logarithm of the transmittance is the absorbance (bottom right). The two most prominent bands in the absorbance spectrum are the commonly mentioned Amide I and Amide II bands that are very characteristic for protein spectra.

(ATR) with the latter being the method of choice in this work. In transmission mode, the sample transmits the light, in reflectance mode the sample reflects the light and ATR makes use of attenuated total internal reflection. The light that should only be reflected within the crystal also penetrates the optically less dense medium as an evanescent wave. This effect is named *Goos-Hänchen effect* after the first scientists who described it in detail. [38, 39] Figure 1.5 illustrates the path of the light through the ATR crystal and the sample. Here d_p is the penetration depth of the evanescent wave into the thinner medium, the sample, with the refractive index n_2 from the ATR crystal with the refractive index n_1. Θ is the angle at which the beam enters the ATR crystal. Equation 1.9 shows the dependence of the penetration depth on the refractive indices of the crystal and the sample as well as the wavelength of light. This also makes it clear that the sample always has to be less optically dense than the utilized crystal, explaining the need for crystals with a high refraction index such as diamonds and silicon.

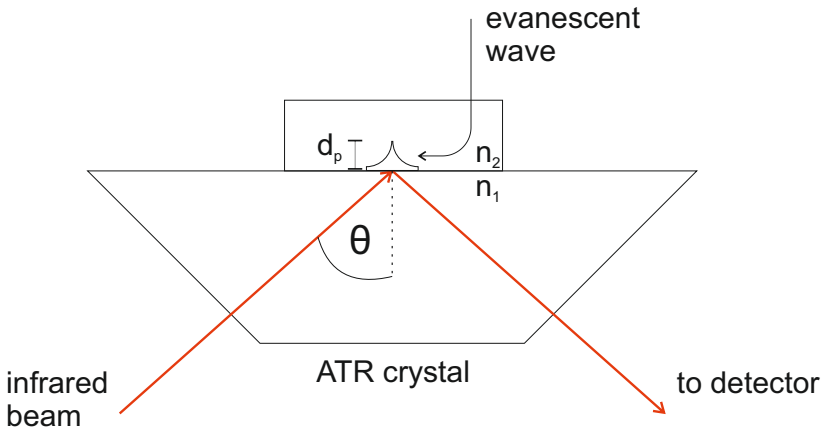

Figure 1.5: The path of the light through the ATR crystal and the evanescent wave. Adapted from [40].

$$d_p = \frac{\lambda}{2\pi n_1 \sqrt{(\sin^2(\Theta) - (\frac{n_2}{n_1})^2)}} \tag{1.9}$$

As the light travels through the crystal into the sample, back into the crystal and then towards the detector, ATR measurements allow for experimental setups that can not be easily achieved in transmission mode. If a homogeneous process stream is the point of measurement, a probe with the ATR crystal at its tip can be inserted into it, almost regardless of volumetric flow rates or dimensions of the pipe. For a transmission measurement, a small part of the process stream would need to be diverted through a transmission cell.

Generally, any heteroatomic bond absorbs infrared light and the vast majority of molecules encountered in biopharmaceutical production processes are infrared active. [41] The most prominent absorption bands of proteins are located between 1700 cm^{-1} to 1200 cm^{-1}. As many of the bonds that are present in these large molecules have overlapping absorption spectra, and the measurement resolution is limited by size constraints of the instrumentation, observed absorption bands consist of absorption caused by multiple structural elements. An example of this effect is shown in Figure 1.6 for the Amide I and Amide II bands. Calculating the derivative of the spectral data can help making distinct features of the absorption spectrum more prominent if the resolution of the spectrophotometer is not high enough to resolve individual spectral features.

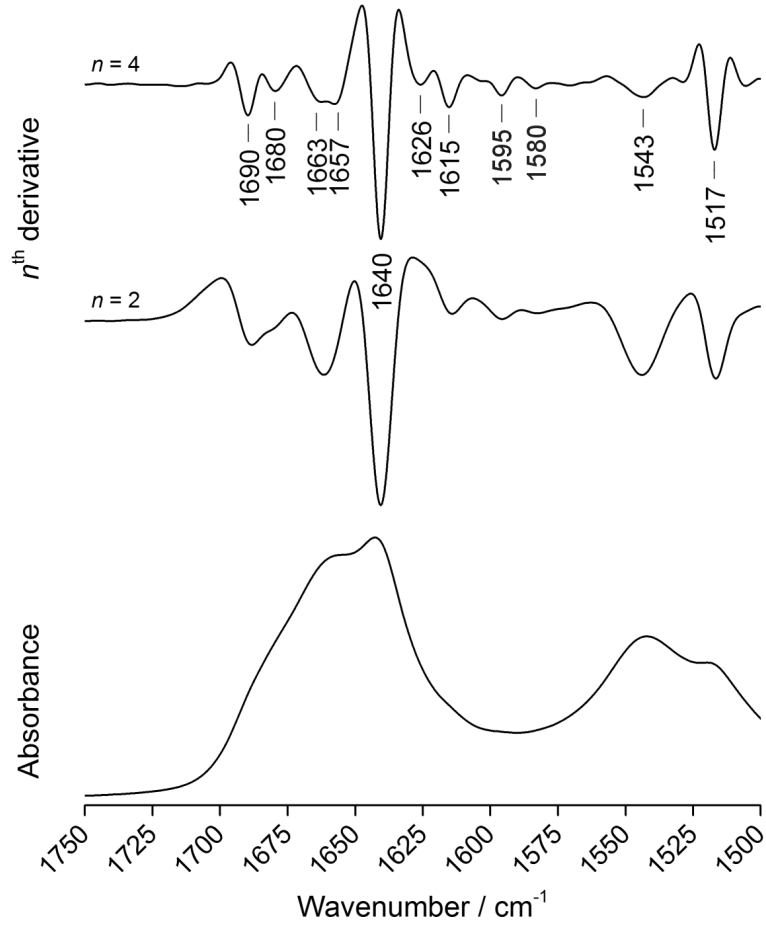

Figure 1.6: Schematic representation of the structural elements that make up the Amide I and Amide II band. Adapted from [42]. While there are multiple structural elements present in protein molecules that absorb in the range from 1750 cm^{-1} to 1500 cm^{-1}, the two bands are relatively smooth. [43, 44]

Another challenge in the application of FTIR spectroscopy for biopharmaceutical processes is the presence of water. In relation to the other molecules, it is present abundantly. Even though the absorption coefficient of water is not particularly high, at a concentration of approximately 55 mol/dm^3 it causes a significant background absorption. As parts of this absorption overlap with regions where proteins absorb as well [41], the need for proper background correction becomes apparent. The main mode of background correction applied in FTIR is background subtraction. This means that a reference spectrum of a sample containing only

buffer is subtracted from every measurement, just as blanking is done for many other spectroscopic techniques. Additional methods to account for changes in the background absorption during an experiment are discussed in later chapters.

1.4 Protein Conjugation

Proteins can be modified by chemical reactions after they have been produced and purified. The most prominent examples are the attachment of a drug molecule to a monoclonal antibody to create an Antibody Drug Conjugate (ADC) or the attachment of PEG to a therapeutic protein. Cytotoxic agents are commonly attached in ADCs to be used in cancer therapy. The advantage of this method over classical chemotherapy is that the antibody directs the cytotoxic drug directly to the target, minimizing its effect on the rest of the body. [45] In contrast to the conjugation molecules in ADCs, PEG is almost completely inert in biological systems. Between 1990 and 2018, 18 PEGylated biopharmaceuticals for human use have been approved. The main reasons for PEGylation in these medicines are reduced immunogenicity, decreased renal clearance and decreased receptor-mediated clearance. [18, 46] Typically, conjugated proteins are manufactured by first producing the unmodified parent protein in a pure form followed by the conjugation reaction itself. PEGylation reagents utilizing different chemistries are commercially available. They fall into three categories: acylating agents, alkylating agents and thiol-reactive agents. The agents used in this work fall into the first two categories. [17] After the conjugation reaction, the desired species of conjugated protein has to be separated out of the mixture, if the reaction does not happen to produce only one species to begin with. [19] The challenges posed by separating the different species also include that it may be necessary to not only separate species with a certain degree of PEGylation from the others but also certain positional isomers. In some cases, the efficacy of these isomers can vary significantly. [46] Apart from the conjugation site, this can also depend on the chain length of the attached PEG molecule. Both factors influence how the protein interacts with its surroundings, which not only influences its efficacy but also potential purification steps such as bind and elute chromatography. [47] Apart from purification steps after the reaction, the conjugation and purification can also be conducted as an integrated process step. Several strategies can be employed to achieve this including immediate separation from the reaction in solution or binding

the protein to chromatographic media before the reaction takes place. [48, 49]

1.4.1 Production of PEGylated Proteins

In most cases, the production of a PEGylated protein begins by first producing an purifying the parent protein in a process as outlined in Figure 1.1. In addition to the loss of target protein that is caused by this first production process, additional losses will occur during the second step which is the reaction and purification of the PEGylated product.

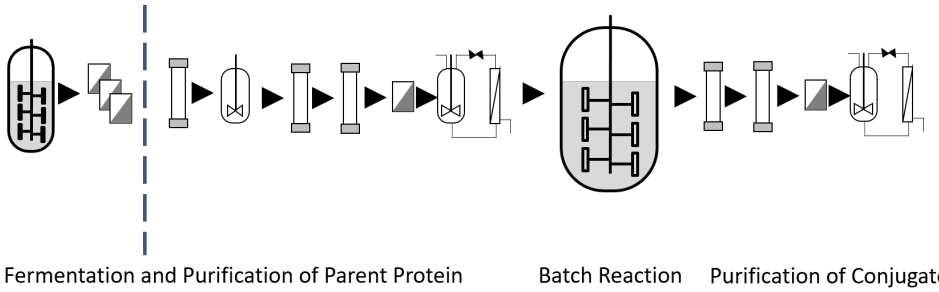

Figure 1.7: The outline of a typical biopharmaceutical production process for a conjugated protein. The parent protein is first produced in a typical process for a biopharmaceutical molecule followed by a reaction step and a second set of purification steps.

Figure 1.7 illustrates an exemplary process for the production of a conjugated protein. As the average yield of a biopharmaceutical production process was about 70 % in 2014, it becomes clear why the production of conjugates is so expensive. [50] Even though the conjugation reaction has been reported to achieve yields of around 90 %, there are still significant losses in this second part of the process. If the yield of the target species is high, but the conversion is low, there would still be significant loss if the unreacted protein can not be recovered from the reaction. [51] The same principles of yield also apply for the second set of purifications which means that the more steps are needed and the less efficient they are, the more product is lost.

1.4.2 Conjugation reactions

The chemical reaction to form protein species with a higher degree of PEGylation are typically multi stage reactions independent of their specific reaction chemistry. Equations 1.10 through 1.12 show the principle of the

multi step reaction that is needed to produce proteins with a higher degree of PEGylation (DOP).

$$\text{Protein} + \text{PEG} \longrightarrow \text{Mono}-\text{PEG}-\text{Protein} \tag{1.10}$$
$$\text{Mono}-\text{PEG}-\text{Protein} + \text{PEG} \longrightarrow \text{Di}-\text{PEG}-\text{Protein} \tag{1.11}$$
$$\text{Di}-\text{PEG}-\text{Protein} + \text{PEG} \longrightarrow \text{Tri}-\text{PEG}-\text{Protein} \tag{1.12}$$

Closer inspection of these equations makes clear why it may be necessary to study the kinetics of a given chemical reaction. If the reaction rates of the second and third reaction are greater than zero and under the assumption that Mono-PEG-Protein is the target product, it becomes clear that we start losing target product from the moment on we start producing any of it. As the formation of every PEGylated species with a DOP higher than one is dependent on all prior reactions, a kinetic model is expressed as a set of differential equations.

$$r_1 = PEG_{reacting} * Protein * k_1 \tag{1.13}$$
$$r_2 = PEG_{reacting} * \textit{Mono-PEG-Protein} * k_2 \tag{1.14}$$
$$r_3 = PEG_{reacting} * \textit{Di-PEG-Protein} * k_3 \tag{1.15}$$
$$dx(Protein) = -r_1 \tag{1.16}$$
$$dx(\textit{Mono-PEG-Protein}) = r_1 - r_2 \tag{1.17}$$
$$dx(\textit{Di-PEG-Protein}) = r_2 - r_3 \tag{1.18}$$
$$dx(\textit{Tri-PEG-Protein}) = r_3 \tag{1.19}$$

Equations 1.13 through 1.19 show the set of differential equations that represent the three consecutive reactions introduced earlier in their simplest form. This is under the assumption that the PEG reagent does not degrade over time and that it is not possible for a protein molecule to react with two molecules of the PEG reagent simultaneously. In these equations, k_n are the rate constants and $PEG_{reacting}$, $Protein$ et cetera are concentrations. Here, k_n are lumped rate constants for all reaction sites, which may not be an accurate representation of reality. Figure 1.8 shows one possible solution of this set of differential equations with arbitrarily chosen initial conditions and rate constants. This illustrates why determining the kinetics of a reaction can be of significant use. Assuming that Mono-PEG-Protein, represented by the solid red line, is the target compound, it is clear that the reaction should be stopped relatively early at a low conversion of the parent protein because the consecutive reactions

lead to product loss. In cases where the protein with a single attached molecule of PEG is the target, unreacted native protein may be recycled but molecules with a higher degree of PEGylation have to be discarded.

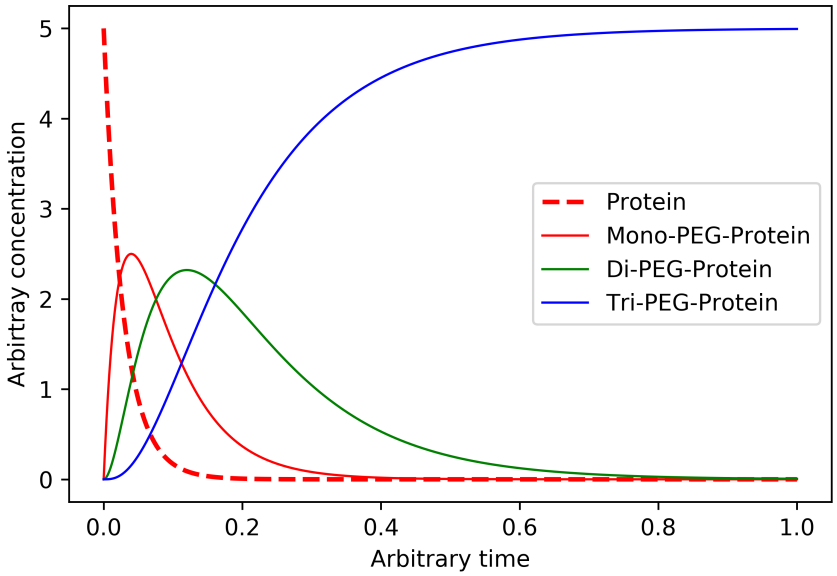

Figure 1.8: One possible solution for the system of differential equations with arbitrary values for initial conditions and rate constants. In this case, where the Tri-PEG-Protein is the species with the highest DOP possible, all parent protein molecules are eventually converted to this species as this result was calculated with a large excess of PEG reagent.

In order to obtain the values of the parameters for a real reaction system, a time resolved measurement of the concentrations of the individual reaction products is necessary. That means that samples have to be taken and either analyzed in a timely manner or the reaction has to be stopped in the samples to analyze them at a later point in time. Then, an adequate reaction model that defines the set of differential equations has to be chosen. For this choice, the reaction mechanism hast to be taken into account. Extensive literature is available on this topic and it should be noted that choosing a model that represents reality accurately is often not feasible. In many cases, it is either impossible or very difficult to analyze isoforms of certain species. In such cases lumping reaction rates is the method of choice to simplify analysis and modeling. Once the model has been chosen and the experimental data obtained, the set of differential equations is fit to the experimental data. Specifically this means that the

set of differential equations is repeatedly solved with varying values for the unknown parameters, followed by a comparison of the model response to the experimental data. This is repeated until the error between model response and experiment is minimized. The simplest method is to try as many values for each parameter as possible or employ more advanced methods such as a genetic algorithm to make the process of finding a local minimum much quicker. Measures such as confidence intervals can be used to evaluate how certain it mathematically is that the value of a parameter is close to the true value and the coefficient of determination to determine how close the model response is to the experimental data. [52] However it should be noted that a small error between the model and the experimental data is not a sufficient condition to decide if the correct model was chosen.

2

Thesis Outline

2.1 Research Proposal

Many fatal and chronic diseases can be treated with biopharmaceutical drugs. In the regulatory landscape, which requires extensive and expensive clinical trials, developing new medicines is a significant undertaking. In addition, the global framework of patent laws requires an early application for a patent on a new product. This in turn rewards achieving a short time to market with a maximized time to recoup the original investment. Despite ongoing efforts, the process development for biopharmaceuticals still takes a significant amount of time and effort, and even then, the commercialized production is not trivial. The regulatory body of the United States of America, the FDA, encouraged a more widespread use of the principals of QbD and PAT with the start of their PAT initiative in 2004. Developing new processes to incorporate the QbD and PAT principles from the beginning ensures that they run reliably from the beginning. The process knowledge gained from designing quality into the entire process is advantageous to obtain regulatory approval of the commercial scale process. The same strategies that are used to implement PAT in processes at commercial scale can also be used to speed up process development.

Biopharmaceuticals are large biomolecules produced by living organisms such as bacteria or mammalian cells. The product itself and the profile of the impurities they need to be separated from are complex. To monitor and control the separation of the target molecules from the impurities, spectroscopic sensors have been applied as PAT tools for decades. The

most established spectroscopy method is UV/Vis spectroscopy. While it is a fast and accurate method to determine protein concentration and has also been utilized for distinguishing different species based on minute spectral differences, certain applications require a higher level of information. Some new modalities such as PEGylated proteins need to be separated based on their DOP, which is not trivial as PEG is not absorbing in UV/Vis light. It is also desirable to access additional CQAs in greater detail directly in the process stream. Instead of only reporting on the presence of a change in process conditions, spectroscopic techniques such as FTIR can also report on their identity based on prior knowledge. Identifying the identity of an undesired species right as it passes through the process can help find the root cause of a deviation and its possible ramifications more quickly.

This thesis aims to implement PAT based on FTIR and automation techniques in the process development of PEGylated protein species. The possibility to leverage the distinct absorption exhibited by PEG and protein respectively should be investigated. The FTIR measurements should be implemented to allow for seamless in-line measurements and a general focus is set on automation and simplification of experiments. This implementation should include the possibility for automatic alignment of spectral data and the corresponding analytics. As a reference implementation, FTIR should be used as a PAT sensor in combination with statistical regression models. This method has been shown to work well with other spectroscopic techniques. Based on literature, FTIR should also enable direct usage of spectral features in addition to statistical regression modeling. Using specific spectral features should also pave the way to FTIR based PAT sensors that rely on less prior knowledge and may thus be more robust. It should be investigated if such a PAT sensor based only on external prior knowledge can be utilized to extract information from a purification process.

In addition to the spectral analysis during DSP, the process to produce the PEGylated proteins itself should also be investigated. Specifically if automation of the sampling from the batch solution can be used to enable consistent and accurate sampling even over long periods of time. As this process usually begins with the purified unmodified protein, it is desirable to control the process closely to waste as little raw material as possible. The conjugation reactions can sometimes be very slow and need to be monitored over their entirety, automation is key to avoid bottle necks due to working hours. Established techniques utilizing pipetting stations are run in micro titer plates in small volumes, effectively running many batch reactions in parallel under the assumption that all wells are equal. The

possibility of monitoring the batch reaction with spectroscopic methods should also be investigated. Another option to be evaluated is performing the PEGylation reaction on a chromatographic column to integrate the reaction with the purification of the desired products. The combination of kinetic analysis of the chemical reaction and the PAT supervision of the downstream process could be used to run the entire process in an optimal state.

2.2 Overview and Author Statement

In this chapter, a short overview of the research papers written in the scope of this thesis is given. Chapter 3 investigates the applicability of in-line FTIR to multiple separation problems. The separations of a mAb from an impurity and different PEGylated protein species from each other are investigated using MVDA. The removal of a process related impurity in the breakthrough of a chromatographic separation is also investigated. Chapter 4 elaborates on the separation of protein species with different degrees of PEGylation and investigates a method to estimate the degree of PEGylation in the effluent of a chromatographic separation in near real-time. In this case, the conjugation reaction and separation are performed in an integrated step on a single chromatography column. A more versatile method for background subtraction, EMSC, is applied to the FTIR data in this paper. Chapter 5 establishes an automated method for the generation of samples to investigate the kinetics of batch PEGylation by modifying existing ÄKTApurifier hardware for automated sampling. This setup is further utilized in Chapter 6 to automatically acquire spectral data while the reaction is running and explore the PAT capabilities for estimating the state of a batch reaction in near real-time. Fundamentals for PAT and conjugation reactions that were applied in this work have been laid by the thesis of Benjamin Maiser (2013), Nina Brestrich (2016), Josefine Morgenstern (2017) and Matthias Rüdt (2018). In the following, the resulting papers are listed with a short summary and their publication status. In the first of the following chapters, first authorship was shared (contributed equally) among colleagues and me. This was undertaken to elevate the quality of the publication and reduce the individual workload that would have been necessitated by time constraints and technical limitations imposed by the experimental work.

3. **In-line Fourier-Transform Infrared Spectroscopy as a Versatile Process Analytical Technology for Preparative Protein Chromatography**
   Steffen Großhans*, Matthias Rüdt*, Adrian Sanden*, Nina Brestrich, Josefine Morgenstern, Stefan Heissler, Jürgen Hubbuch
   (* contributed equally)

This manuscript investigates FTIR spectroscopy as a process analytical technology. The implementation of FTIR as an in-line sensor and the

correlation of spectral data with off-line analytics is established with a custom-made experimental setup and applied to two case studies. Both the separation of a mAb from an impurity and the separation of proteins with different degrees of PEGylation are analyzed using PLS regression. A third case study showcases the selective quantification of a process related impurity in the flow-through of a chromatography step.

The original challenge that led to this work was the implementation of automated in-line acquisition of FTIR spectra in the flow path of a lab scale chromatography device. This was achieved by utilizing a multi-purpose USB I/O device that enables the communication of the process state of the chromatography system to a Matlab script running on the same computer. In turn, I implemented the communication between the control software for the FTIR device and the Matlab script to close the connection between the chromatography system and the FTIR device. This kind of communication has been established for UV/Vis but had not been implemented for FTIR before. [53, 54] To facilitate this work, I implemented a method to control the FTIR device. As there are severe limitations in the OPUS control software for the FTIR device, unconventional methods for control where used. In OPUS, an internal scripting engine based on the Visual Basic (VB) programming language is available to automate tasks. A routine that repeatedly checks for changes in a file on the computer was used to check when the respective change was written to the file by the Matlab script. Once this change was detected, the measurement was started automatically. Trivially, as shown in earlier work, the volume of the flow path was determined gravimetrically to align the spectral data with collected fractions. During the initial setting up of the instruments it also became clear that spectral correction to account for changes caused by the salt gradient during elution had to be employed. To achieve this, I implemented a method to conduct blank experiments without a load phase to collect spectral data only containing changes caused by these gradients and a method to subtract these from the main experimental data. This was crucial for the project to extract as much information about the analytes from the spectral data as possible.

Three case studies were designed to showcase the versatile capabilities of in-line FTIR measurements. Estimating the concentrations during the elution from a chromatography column in a similar manner as already shown with UV/Vis was chosen as the first case study. A monoclonal antibody and lysozyme were loaded onto the chromatography column together and eluted by gradient elutions of different lengths. To show the species could be identified individually, a PLS regression model was fitted on the spectral data and reference analytics.

The separation of PEGylated lysozyme species produced by batch PEGylation was monitored in the second case study. Subsequent to the batch PEGylation process, the reaction solution was diluted and loaded onto a chromatography column. It was known from previous work by Morgenstern et al. [55] that species with a higher degree of PEGylation elute at lower ionic strengths in cation-exchange chromatography. The different species were eluted by gradients in a similar fashion to the first case study and the spectral data correlated to reference analytics. The reference analytics were validated with samples of known composition that were analyzed with Maldi.

For the final case study, I investigated the detection of a process related impurity in the flow through of a chromatography step. Triton X-100 is regularly used as an excipient in biopharmaceutical production processes which needs to be removed before the final formulation of the product. Triton X-100 was added to the load solution in a concentration that could occur in an industrial process. It was then shown that the excipient can be quantified in the flow through during the loading step by means of a basic linear regression. Individual extracted spectra could also be used to confirm the identity of the compound by comparison to a reference spectrum. The main outcome relevant to this thesis was the implementation of the automation and showcasing the benefits over UV/Vis. Matthias Rüdt focused on improving the automatic determination of smoothing and regression parameters. Steffen Großhans worked on the PEGylation reaction and offline analytics. All non-equally contributing authors were consulted regarding minor details about the final manuscript and for troubleshooting challenges with the equipment.

Manuscript published in Journal of Chromatography A, 1547, 37-44, 2018.
DOI: 10.1016/j.chroma.2018.03.005

4. Fourier- Transform Infrared Spectroscopy as a Process Analytical Technology for Near Real Time In-line Estimation of the Degree of PEGylation in Chromatography

Adrian Sanden, Susanna Suhm, Matthias Rüdt, Jürgen Hubbuch

This manuscript investigates FTIR spectroscopy as a process analytical technology, specifically to estimate the DOP in-line without the use of statistical regression modeling. The spectroscopic data is pre-processed

with EMSC followed by the extraction of specific spectral features that are associated with protein and PEG respectively. With the use of the ratio of their extinction coefficients, the DOP can be estimated in the column effluent in near real-time. The PEGylation reaction is also performed on the chromatography column itself by loading the protein onto the column and filling it with the reagent solution. This strategy has been shown to favor lower average DOP of the reaction products by introducing steric hindrance as parts of the proteins surface are inaccessible due to being in contact with the surface of the adsorber material.

Manuscript published in Journal of Chromatography A, 1608C, 37-44, 2019.
DOI: 10.1016/j.chroma.2019.460410

5. Modifying an ÄKTApurifier System for the Automated Acquisition of Samples for Kinetic Modeling of Batch Reactions
Adrian Sanden, Sandra Haas, Jürgen Hubbuch

In this manuscript, the development of an automated sampling setup for the supervision of batch reactions is investigated. An ÄKTApurifier System is modified with custom 3D-printed parts to adapt the flow path. The control software for the chromatography system is used in conjunction with custom software to enable the automatic circulation and sampling without significant manual interference. With the proposed setup, samples from a batch reaction container are automatically taken at predefined time intervals and deposited into a micro titer plate containing a stop solution for the reaction. This enables the direct analysis of the generated samples. A kinetic reaction model was fit to the data and the goodness of fit was evaluated using the confidence intervals calculated from the *Jacobian matrix*. The goal of this study was to investigate if the proposed automatic sampling setup can be used to produce data of adequate quality for model calibration. With the small confidence intervals and the overall fit of the model it could be shown that this setup enables sampling and model calibration from a lab scale batch reaction.

Manuscript published in SLAS Technology, 2019.
DOI: 10.1177/2472630319891976

6. ATR-FTIR for *in situ* Reaction Monitoring of Protein PEGylation in Batch Mode

Adrian Sanden, Sandra Haas, Jürgen Hubbuch

In this chapter, the setup for the automated acquisition of spectral data introduced in Chapter 5 is evaluated further. The method to record the spectral data is elaborated upon and the spectral data itself is processed and analyzed to evaluate the feasibility of estimating the process state using the spectral data. It is shown that the data preprocessing can have a significant influence on the results of the regression model calibration, and great attention needs to be paid to avoid overfitting the model to the data. The purpose of this study was to evaluate if it is possible to estimate the state of the chemical reaction by calibrating a PLS regression model on spectral data recorded *in situ* in the batch reaction. Different pre-processing methods are compared and the model output is evaluated utilizing an external test set. Possible reasons for the failure of this attempt and potential improvements to achieve this goal will also be discussed.

Unpublished

3

In-line Fourier-Transform Infrared Spectroscopy as a Versatile Process Analytical Technology for Preparative Protein Chromatography

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

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

[2] Institute of Functional Interfaces, Karlsruhe Institute of Technology, Hermann-von-Helmholtz-Platz 1, Eggenstein-Leopoldshafen, Germany

([*] Contributed equally)

Abstract

Fourier-Transform Infrared Spectroscopy (FTIR) is a well-established spectroscopic method in the analysis of small molecules and protein secondary

structure. However, FTIR is not commonly applied for in-line monitoring of protein chromatography. Here, the potential of in-line FTIR as a Process Analytical Technology (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 monoclonal antibodies (mAbs) and lysozyme to selectively quantify the two proteins with Partial Least Squares Regression (PLS) giving Root Mean Square Errors of Cross Validation (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.

3.1 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 [56, 57]. 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 [21], 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 nontrivial task [59, 60]. 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 [61–63]

and polishing steps [64, 65]. 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 [60]. The approach yields results quickly enough to allow for real-time process decisions in chromatography [54, 66, 67] and works for minute spectral differences [68]. However, in the commonly measured spectral ranges, UV/Vis spectroscopy lacks sensitivity towards relevant aspects of protein structure, notably the secondary structure [69]. Furthermore, organic compounds are often not UV active (e.g. sugars, polyols, and Polyethylene Glycol [PEG][70, 71]) or they may obscure the protein signal (e.g. Triton X-100 [72] and benzyl alcohol [69]). Due to the high sensitivity, UV/Vis absorption spectroscopy is also prone to detector saturation [60, 73].

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 [74–76]. FTIR measures the vibrational modes of samples and thereby provides a spectroscopic fingerprint for different organic molecules. Proteins absorb in the IR spectral range mainly due to vibrations of the polypeptide backbone [69, 77, 78]. 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 [69]. Furthermore, FTIR was previously used as an at-line PAT tool in downstream processing of biopharmaceuticals for quantifying product content, High Molecular Weight species (HMW), and Host Cell Proteins (HCP) [79, 80].

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. Four 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, four 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 [72, 81]. Based on an off-line calibration curve, mass-balancing of Triton X-100 in the flow-through during product loading was performed.

3.2 Materials and Methods

3.2.1 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 3.1. 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, Natrick, 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.

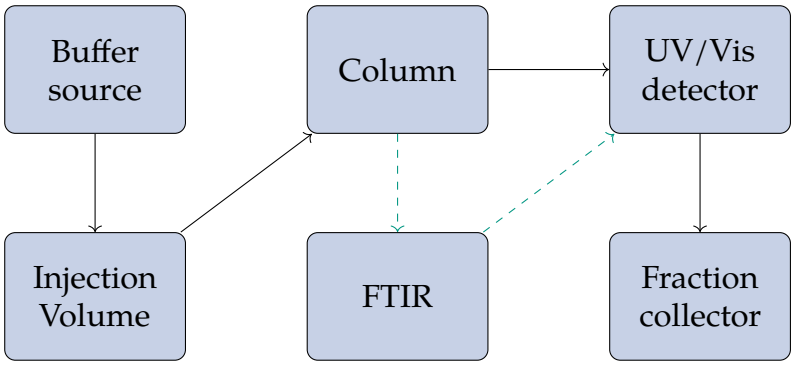

Figure 3.1: Schematic representation of the flow path in the custom chromatography setup, solid lines represent the common flow path in the ÄKTApurifier while the dashed line represents the modification.

3.2.2 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 µm 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 cation-exchange chromatography 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 [82]. 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.

3.2.3 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 (Column Volume [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 offline 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 offline 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 [82, 83]. 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 [81]. 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.

3.2.4 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 $\times$ 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 µl 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 [71]. From the molar concentration of PEGylated lysozyme species, the molar concentration of PEG was calculated.

3.2.5 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 [84]. 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 cm^{-1} resp. above 3100 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 [85]. 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 Predictive 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 [86]. Based on the scaled PRESS, an optimization was performed

using the built-in genetic algorithm of Matlab for integers [87]. 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 [86].

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.

3.3 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.

3.3.1 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 3.2A) which coincides with the most prominent protein band amide I .

The spectral processing workflow is illustrated in figure 3.2 using data from case study I. Specifically the elution of mAb and lysozyme using a 4 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 cm^{-1}. Salt gradients therefore cause a change in absorption over the run (cf. figure 3.2A and B). To reduce buffer effects, it is important to find a suitable dynamic background correction. An approach based on reference spectra matrices and chemometric correlations was not implemented due to the overlap of water and protein bands [88]. 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 3.2C). 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 3.2D) and to take derivatives on the spectral data.

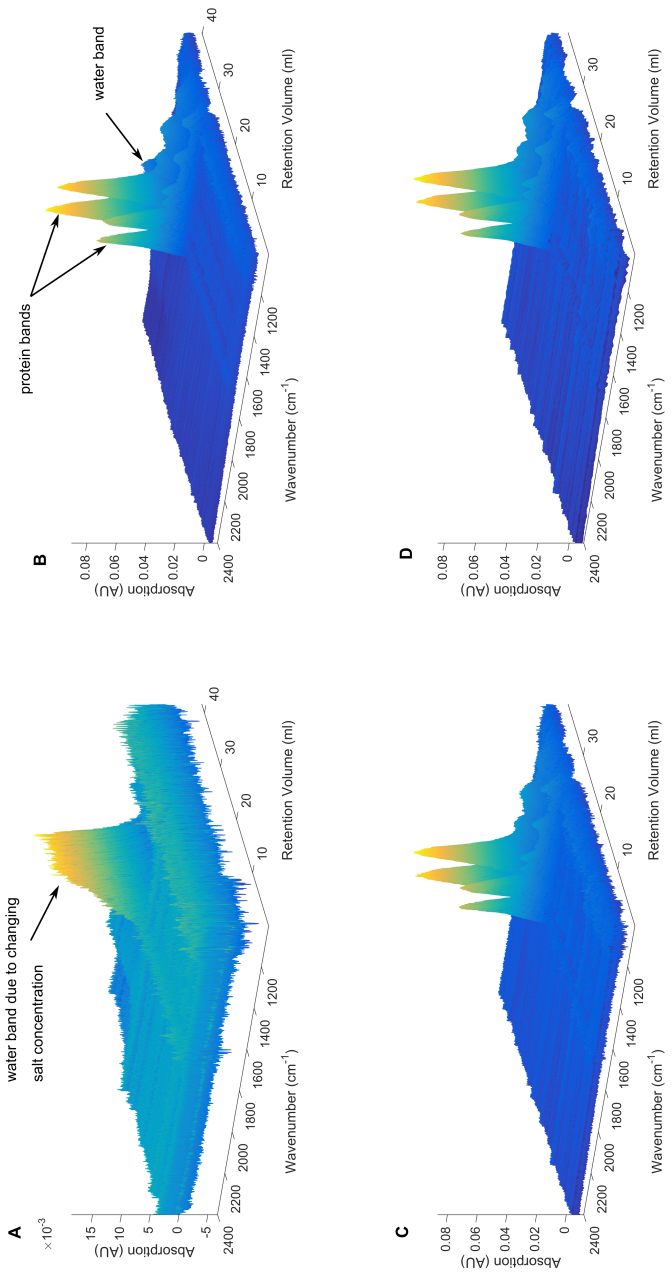

Figure 3.2: Work flow for data treatment of chromatography spectra illustrated with data from case study I, 4 CV run: background run – salt gradient without protein (A); raw spectra of the run with protein (B); spectral data after the background has been subtracted (C); data after smoothing by Savitzky-Golay algorithm (D).

3.3.2 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 mainly contains 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 3.2D are characteristic amide bands associated with the protein backbone [69, 77, 78]. 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 four CEX runs, 2 PLS-1 models were optimized for selective quantification of mAb and lysozyme respectively. The resulting model parameters are listed in table 3.1. Figure 3.3 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 3.2C and 3.3, 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 [85, 86]. Furthermore, 3.23 s measurement time makes FTIR quick enough for monitoring most practical preparative chromatography 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.

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.

Table 3.1: 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

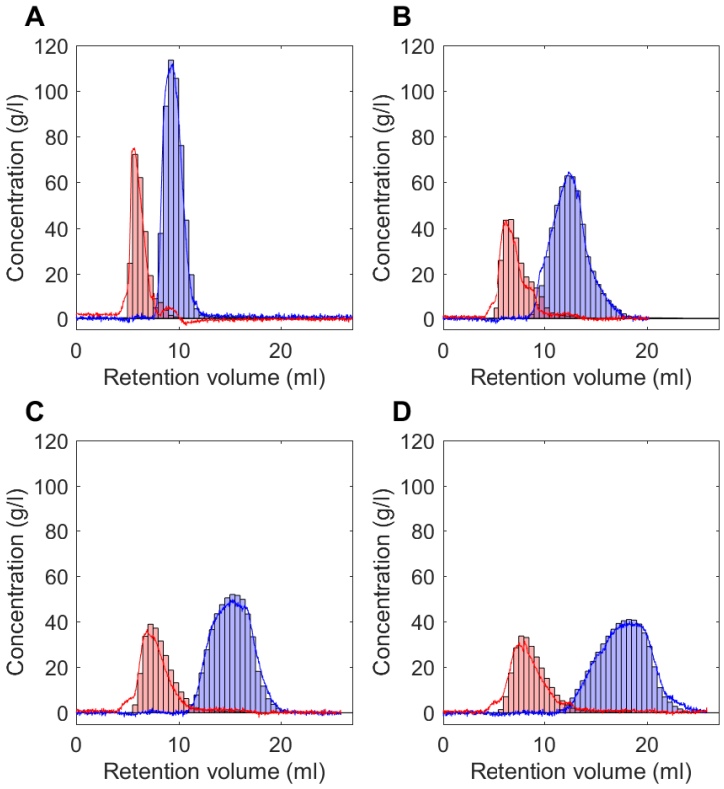

Figure 3.3: Four chromatographic runs are shown for in-line FTIR measurements and selective quantification of mAb and lysozyme. The red bars and lines refer to the mAb off-line measurement and mAb PLS prediction, respectively. The blue bars and lines refer to the lysozyme off-line measurement and lysozyme PLS prediction, respectively. The different subplots show different gradient lengths: A 1 CV, B 2 CV, C 3 CV, D 4 CV.

3.3.3 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 cm^{-1} with multiple shoulders is characteristic for C–O stretching [89]. Due to symmetric CH_2 stretching, PEG furthermore generates a doublet at 2884 cm^{-1} and 2922 cm^{-1}. Bands occurring between 1200 cm^{-1} to 1700 cm^{-1} are related to the protein backbone with some interference from PEG C–H bending.

Figure 3.4 shows a typical chromatographic separation of PEGylated lysozyme species. During the elution, the ratio between PEG and protein 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 3.4) 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 [71, 90, 91].

Based on the evaluation of IR absorption bands, it was decided to include all wavenumbers from 900 cm^{-1} to 3100 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 3.1, the optimization results are summarized. Figure 3.5 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 [71, 92].

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

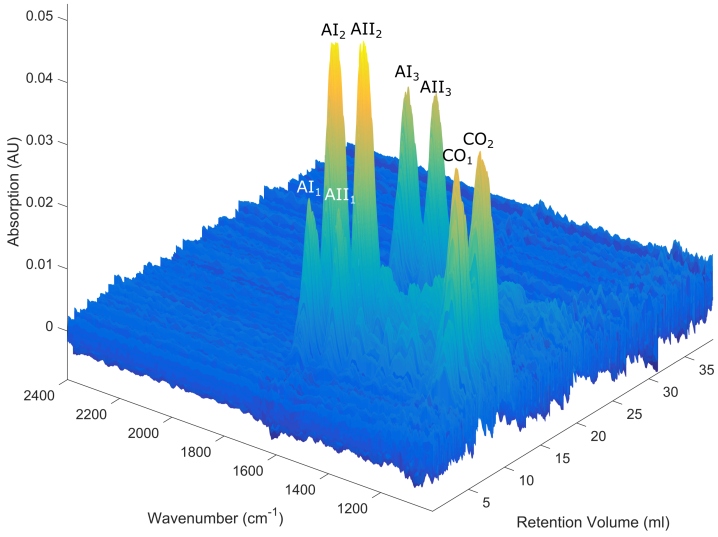

Figure 3.4: Elution of PEGylated lysozyme species from a CEX column with a gradient length of 5 CV. Bands visible between wavenumbers 1200 cm^{-1} to 1700 cm^{-1} are the characteristic amide bands associated with protein. The major protein bands amide I and amide II are mark as AI and AII, respectively. The band at approximately 1100 cm^{-1} is characteristic for PEG (C–O stretching, marked as CO). The subscript numerals refer to the elution order.

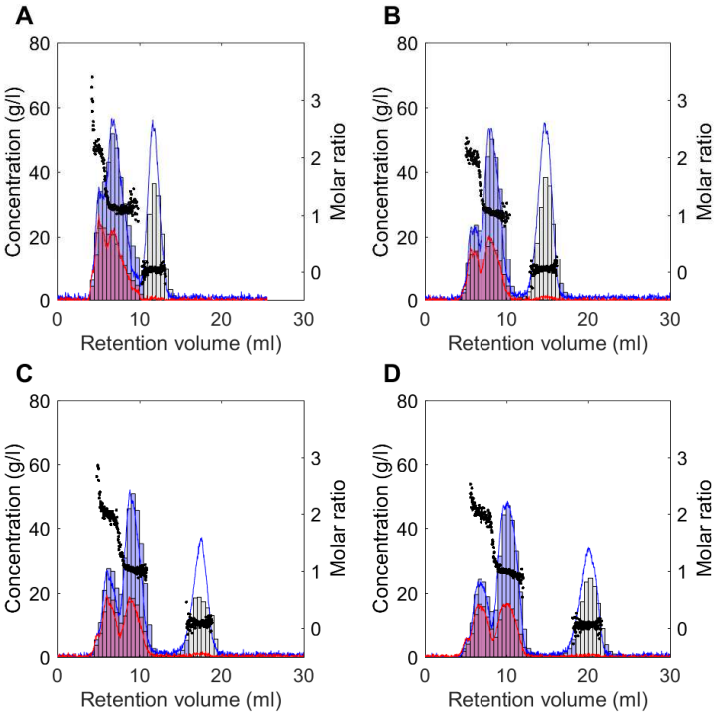

Figure 3.5: Four chromatographic runs are shown for in-line FTIR measurements and selective quantification of PEG and lysozyme. The red bars and lines refer to the PEG off-line measurement and PEG PLS prediction, respectively. The blue bars and lines refer to the lysozyme off-line measurement and lysozyme 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.

3.3.4 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 [72, 81]. 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 3.6).

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 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 cm^{-1} to 1170 cm^{-1} to the Triton X-100 concentration. A linear regression for the calibration curve resulted in a $R^2 > 0.9997$. 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.

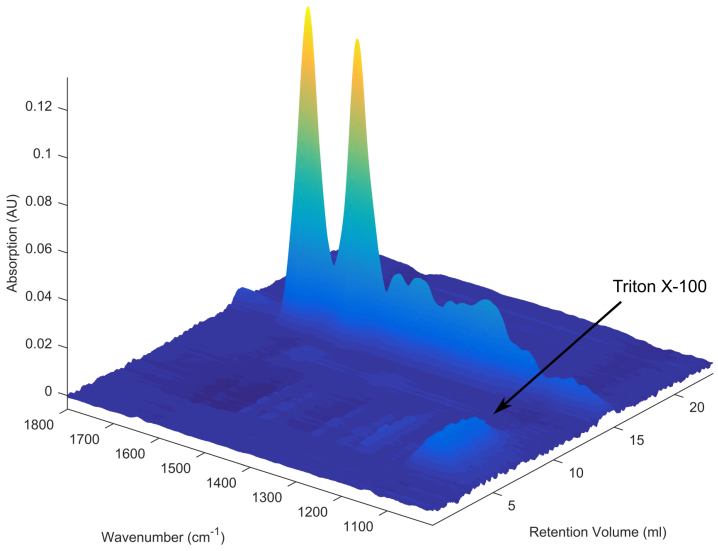

Figure 3.6: Triton X-100 as a process related impurity can be seen in the flow-through of the cation-exchange experiment from 5.5 ml to 11 ml at 1090 cm^{-1}.

3.4 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 monitoring 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

This work has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 635557. We are thankful for the mAb protein A pool which we received from Lek Pharmaceuticals, d.d. We would also like to thank Daniel Büchler for his help conducting the experiments.

4

Fourier- Transform Infrared Spectroscopy as a Process Analytical Technology for Near Real Time In-line Estimation of the Degree of PEGylation in Chromatography

Adrian Sanden[1], Susanna Suhm[1], Matthias Rüdt[1], Jürgen Hubbuch[1]

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

Abstract

PEGylation of biological macromolecules is a well-established strategy to increase circulation half-life, decrease renal clearance and improve biocompatibility. PEGylation is a process in which PEG is covalently attached to a target molecule. The production of PEGylated biopharmaceuticals is usually executed by first producing and purifying the base molecule followed

by the PEGylation reaction and purification of the modified molecule. Most PEGylated pharmaceuticals are produced by random PEGylation in batch mode and need to be purified as mainly the mono-PEGylated form is the desired drug product. In this work we propose a method to estimate the degree of PEGylation (DOP) of modified protein eluting from a chromatography column in near real-time. EMSC is used in conjunction with Alternating Least-Squares (ALS) to alleviate the influence of a salt gradient during ion exchange chromatography (IEX) on the spectral data. To convert the raw data obtained from spectral data to the actual DOP additional information obtained from off-line measurements is utilized. Once the signal correction is applied to in-line spectral data the DOP can be estimated without further use of off-line analytics. As the prerequisites for the application of this method are relatively easy to obtain it may also find use to speed up process development.

4.1 Introduction

A covalent attachment of PEG to biopharmaceuticals is known to enhance their pharmaceutical and pharmacological properties. The main advantages include reduced immunogenicity, increased half-life and greater physical stability. While any active compound can theoretically be PEGylated, macromolecules such as peptide- and protein-based drugs can especially benefit from this modification due to the reduced renal clearance. [93, 94] Despite reports on anti-PEG antibodies causing an immune response in some patients, the clinical efficacy of PEGylation is significant. In some cases the drug substance modified by PEGylation would be of no clinical use without it. [95] In 2018, more than 15 PEGylated drugs have been successfully approved, spanning a variety of different PEG chain lengths, linkers, chemistries and other properties. They include both large biologics and small molecules. [17, 96, 97] PEGylated biopharmaceuticals are usually produced by processing the purified drug substance further and introducing the PEGylation process as an additional step, typically as a batch-reaction. As the batch-reaction is neither site-specific nor specific regarding the achieved degree of PEGylation, additional purification steps follow the PEGylation process. [98] Current processes for the batch-PEGylation of biopharmaceuticals are run based on prior knowledge. [99, 100] Therefore, the batch-PEGylation is run for a pre-defined time under pre-defined conditions and off-line analytics are utilized to release a finished batch. As opposed to batch-PEGylation, solid-phase PEGylation favors reaction products with a lower degree of PEGylation. Even though

this usually comes at the cost of total conversion, it may still be a favorable process. [101, 102]

Since the FDA compiled their view on PAT into their Guidance for Industry in 2004 [21], research and development of new PAT tools has been focused and amplified in academia and industry. For many process steps such as chromatographic separations it has already been shown that it is possible to monitor and control them via spectroscopic methods. [54, 67, 103] Currently, UV/Vis spectroscopy is among the most prevalent techniques, nevertheless an array of other techniques is available. [60] From a process-monitoring point of view, chromatographic separations pose the advantage that the effluent is a time-resolved signal consisting mainly of information related to individual protein species which only overlaps in small parts if the separation works adequately. However, many of the aforementioned techniques require considerable calibration, especially those involving PLS regression models. Additionally, it is not trivial to measure the concentration of PEG in-line in the column effluent, as it is non-ionic and does not absorb UV light at commonly used wavelengths. [104]

In this work, we present an approach to estimate the DOP of currently eluting species from a chromatography column in near real-time without the need for a previously calibrated model. FTIR is used in in-line mode at the column outlet and the spectral data is then processed to calculate the DOP. Specifically, PEGylated species obtained from on-column-PEGylation are utilized as an example. EMSC-ALS is evaluated as a means to correct the spectral data regarding the influence of the salt gradient used for elution. The salt gradient causes absorbance in the same spectral region as protein as well as a distortion of the entire spectrum and thereby hinders data evaluation.

It should be noted that the background spectra required for any FTIR measurements and the background data that is utilized for the EMSC are not identical. The former is used to remove the influence of atmospheric effects and bulk water while the latter contains the information regarding the changing chemical composition of the solute without the analytes.

4.2 Materials and Methods

4.2.1 Experimental Setup

On-column-PEGylation experiments and batch-PEGylation separation were conducted on an ÄktaPure chromatography system and fraction col-

lector, controlled by Unicorn 6.4.1 (all GE Healthcare, Uppsala, Sweden). To enable in-line FTIR measurements, the outlet of the column is attached to a Tensor 27 FTIR instrument with BioATR II flow cell controlled by OPUS 7 (all Bruker Optics, Ettlingen, Germany). A CryoMCT detector cooled by a Stirling engine is used to detect the IR signal. To ensure a steady state, the Stirling engine was turned on at least 30 min before any measurements take place. The BioATR II flow cell was tempered to 22 °C, reflecting the same temperature as the air conditioning in the entire laboratory. FTIR spectra were recorded using the chromatography mode in OPUS, repeatedly recording single spectra from 3500 cm^{-1} to 900 cm^{-1} at a resolution of 4 cm^{-1} with the scanner velocity set to 160 Hz. Individual spectra are recorded by averaging 64 scans, adding up to 4.45 s each. The background is collected while low salt buffer is pumped through the system, averaging 768 scans. Data above 2500 cm^{-1} and below 980 cm^{-1} was discarded before further processing.

4.2.2 Proteins and Buffers

All solutions were prepared with purified water from a PURELAB Ultra water purification system (ELGA Labwater, High Wycombe, United Kingdom). Buffers were filtered using 0.2 µm filters from Sartorius (Göttingen, Germany) and chromatography buffers additionally degassed in a sonification bath before use. All pH adjustments were done using 32 % or less HCl solution. All experiments were conducted using chicken egg white lysozyme from Hampton Research (Aliso Viejo, CA, United States). Activated 5 kDa PEG was purchased from NOF Corporation (Tokyo, Japan) in the form of Methoxy-PEG-NHS (MPEG-NHS, Sunbright ME-050HS).

On-column-PEGylation was run in sodium citrate ($C_6H_5Na_3O_7 \cdot 2\,H_2O$, Sigma Aldrich, St. Louis, United States) buffer with a concentration of 25 mM at pH 7.5, respectively, with the high salt buffer for elution differing only by containing an additional 1 M NaCl (Merck, Darmstadt, Germany).

Off-line analytics using analytical cation-exchange chromatography were carried out using 20 mM Tris at pH 8.0 with an additional 700 mM NaCl for the elution buffer (both Merck, Darmstadt, Germany).

4.2.3 On-column-PEGylation of lysozyme

The on-column-PEGylation was conducted on a 5 mL HiTrap SP FF column from GE Healthcare by loading 250 mg lysozyme in low salt buffer from a sample loop. After washing the unbound protein off the column, the PEGylation reagent was loaded onto the column. The PEG-NHS solution

was prepared by dissolving 1.078 g of PEG-NHS in 5 mL low salt buffer right before loading takes place. After dissolving the PEG-NHS, it was filtered through a 0.2 µm Polyethersulfone (PES) syringe filter (Pall, Port Washington, USA). Once the PEG-NHS was loaded onto the column, the flow through the column was stopped to allow the reaction to occur over the course of 4 h. To remove excess PEG from the column, it was washed for 8 CV before the elution took place by running a gradient to 60 % high salt over 5.5 CV, followed by a high salt strip for 3 CV. The background spectra for the FTIR measurements were collected during the incubation period at the elution flowrate of 1 mL/min by bypassing the column and thus running low salt buffer through the measuring flow cell.

4.2.4 Analytical CEX Chromatography

Analytical chromatography was conducted with the method previously published by our group. [103] All analytics were carried out using a Vanquish UHPLC system (Thermo Fisher Scientific, Waltham, MA, USA). The Vanquish UHPLC System consisted of a Diode Array Detector HL, Binary Pump F, Column Compartment H with integrated preheater and post-column cooler as well as a Split Sampler FT. Control and evaluation of results was done in Chromeleon 7.2 SR4. 5 µl of sample were injected from each sample and the column was washed for 0.5 min to flush out unbound compounds. 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. Subsequently a high salt strip at 100 % was run for 1 min. The quantification of all species was based on a dilution series of pure lysozyme, with the peak identification being based on previous works by our group. [71, 103]

4.2.5 Spectral data correction by EMSC-ALS

Matlab (The Mathworks, Natick, MA, United States) was used for the data preprocessing by EMSC-ALS. All other data manipulation was done in Python 3.7 (Python Software Foundation, Delaware, United States). The EMSC is based on work of Martens et al. [105], combined with using asymmetric least squares (ALS) as proposed by Boelens et al. [106] The chemical background is represented by the coefficients obtained by applying principal component analysis (PCA) to the spectral data of a salt gradient chromatography run without any analytes present. Specifically, a chemical background in the shape of the elution described for the on-column-PEGylation experiment is utilized and is represented by the

first four PCA components. By applying EMSC-ALS the influence of the salt gradient on the 3D-field of spectral data is reduced. Applying ALS penalizes negative bands in the process, as negative absorbance values should not occur in spectral data. The background correction is executed with a polynomial of second order and an asymmetry factor of 0.1. As this method does not represent a simple subtraction, the characteristics of the influence of the salt gradient captured by the EMSC can be used to correct gradients of any shape and length in the same buffer system.

Estimation of the degree of PEGylation

The absorbance data over time corresponding to protein and PEG is extracted from the corrected spectral data. Protein is represented by the Amide II band at $1540\,\mathrm{cm}^{-1}$ [107] and PEG by the band located at $1089\,\mathrm{cm}^{-1}$ [108]. The absorbance of the PEG trace is divided by the protein trace to calculate the fraction between these two. Using the Savitzky-Golay algorithm [109], the fraction is smoothed in direction of time with a window length of 13 time steps and a first order polynomial. Based on the UV signal from the chromatography system, the values of the fraction corresponding to points in time where the UV absorption values are smaller than twice the average of the baseline corrected UV signal are discarded.

Correction of spectral fraction with the ratio of extinction coefficients

The Beer-Lambert law states that the extinction is equal to an extinction coefficient multiplied by the concentration and the path length under the assumption that one is dealing with dilute solutions in the range of linear absorption. Equation 4.1 is the Beer-Lambert law for an arbitrary wavenumber with the absorbance A (AU), the absorption coefficient ϵ (AU/mol), the concentration c (mol/mL) and the pathlength d (micrometer). [27]

$$A_\lambda = \epsilon_\lambda c d \tag{4.1}$$

$$A_{1540\,\mathrm{cm}^{-1}} = \epsilon_{1540\,\mathrm{cm}^{-1}} c_{Protein} d \tag{4.2}$$

$$A_{1089\,\mathrm{cm}^{-1}} = \epsilon_{1089\,\mathrm{cm}^{-1}} c_{PEG} d \tag{4.3}$$

$$\frac{A_{1089\,\mathrm{cm}^{-1}}}{A_{1540\,\mathrm{cm}^{-1}}} = \frac{\epsilon_{1089\,\mathrm{cm}^{-1}}}{\epsilon_{1540\,\mathrm{cm}^{-1}}} \frac{c_{PEG}}{c_{Protein}} \frac{d}{d} \tag{4.4}$$

$$DOP = \frac{c_{PEG}}{c_{Protein}} \tag{4.5}$$

$$F_{DOP} \stackrel{!}{=} \frac{\epsilon_{1540\,\mathrm{cm}^{-1}}}{\epsilon_{1089\,\mathrm{cm}^{-1}}} \tag{4.6}$$

$$F_{DOP} \frac{A_{1089\,\mathrm{cm}^{-1}}}{A_{1540\,\mathrm{cm}^{-1}}} = DOP \tag{4.7}$$

By assuming that the pathlength d, in this specific case the penetration depth of the evanescent wave, is constant, it can be eliminated. In reality, the penetration depth does depend on the wavenumber and the refractive index of the sample. [110] If it is further assumed that there is no unreacted PEG reagent in solution, all absorption can be attributed to PEGylated protein or native protein and that only PEG or protein absorb at either wavenumber. Therefore, the degree of PEGylation (DOP) can be defined as in equation 4.5. Additionally, we define a correction factor F_{DOP} as shown in equation 4.6. Substituting 4.5 and 4.6 in 4.4 leads to a definition of the DOP dependent on the factor F_{DOP} and the absorbance as shown in equation 4.7. Spectroscopically inclined readers will notice that the assumption of only PEG or protein absorbing at the respective wavenumbers is not valid. While the absorbance of PEG at 1540 cm^{-1} is very limited, this is not the case for protein at 1089 cm^{-1}. To calculate the estimated DOP, the part of the raw fraction that is caused by the absorption of protein has to be subtracted first. This offset was determined by running a bind and elute experiment in the same way as the on-column-PEGylation without the actual PEGylation reaction, by only binding and eluting 125 mg lysozyme and determining the fraction of the elution peak after applying the same spectral correction as for the PEGylation experiment (data not shown). The absorption coefficients of PEG at 1089 cm^{-1} and lysozyme at 1540 cm^{-1} were obtained by preparing a dilution series of both compounds in low salt buffer and recording spectra on the FTIR instrument. Concentration ranges were chosen from 2.5 mg/mL to 40 mg/mL for PEG and lysozyme.

4.3 Results and Discussion

Spectral data collected from in-line measurements during the elution of an on-column-PEGylation experiment was utilized as a PAT tool.

4.3.1 On-column-PEGylation

Figure 4.1 shows the UV signal and the time-resolved absorbance of the FTIR bands associated with PEG and protein plotted over the retention time. The green and red lines are the FTIR traces at 1540 cm^{-1} and 1089 cm^{-1} respectively, the pink line UV at 280 nm and the blue line the raw fraction between the PEG and protein traces. The shoulder in the UV signal starting at a retention time of around 400 s is more prominent in UV than in IR as the UV detector is more sensitive. In contrast, the two main peaks are saturated in the UV signal, but not in IR, as seen around 800 s and 1300 s respectively.

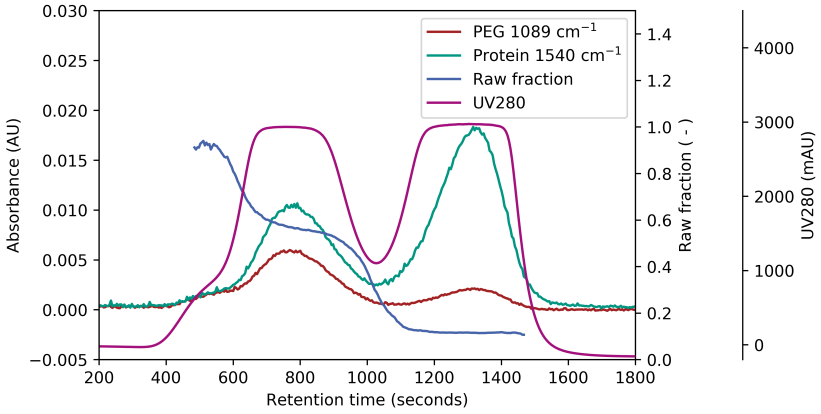

Figure 4.1: On-column-PEGylation: UV at 280 nm, FTIR absorbance at 1540 cm^{-1} and 1089 cm^{-1} as well as the raw fraction of the PEG- and protein-associated bands.

Off-line analytics

The result of the off-line UHPLC analytics is shown in figure 4.2. As expected from previous work, the species with a higher DOP elute prior to those with a lower DOP and the native protein. All elution profiles are approximately Gaussian shaped which is expected for chromatographic separations. The elution profiles are also consistent with the UV sum signal while it is not saturated. The UV280 trace is additionally shown in

figure 4.2 to allow for easier comparison to the offline data. The UV signal reaches saturation at about 3000 mAU for the signal at 280 nm, which is expected for the utilized measuring cell. The DOP can then be determined by calculating the proportion of each species in relation to the total protein concentration in each fraction as shown in equation 4.8.

$$DOP_{offline} = \frac{c_{di-PEGylated}}{c_{total}} * 2 + \frac{c_{mono-PEGylated}}{c_{total}} \tag{4.8}$$

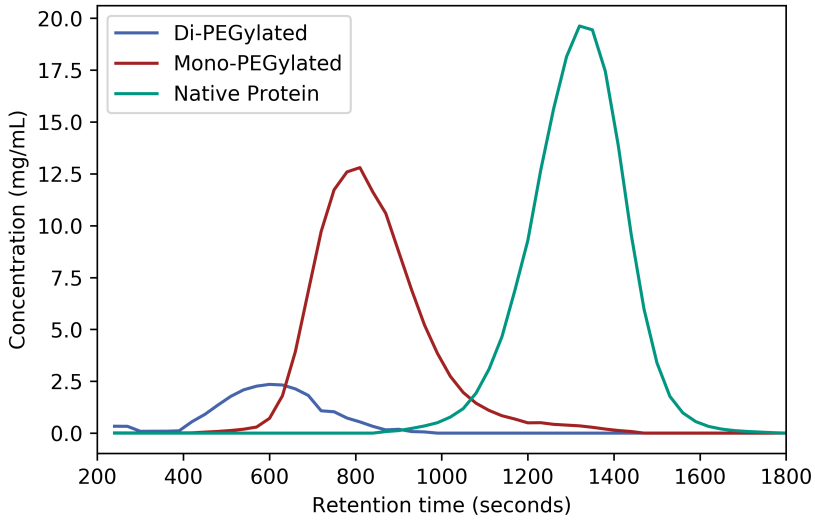

Figure 4.2: On-column-PEGylation: Concentrations for PEGylated and native lysozyme during the elution from the column after on-column-PEGylation. The dashed line shows the UV280 trace from the chromatography system.

4.3.2 Spectral data correction by EMSC-ALS

Figures 4.3 and 4.4 show the raw recorded FTIR 3D-field and the spectral data after correction by EMSC-ALS respectively. The most prominent influence of the salt gradient on the spectral data is visible at around 1640 cm^{-1}, where the linear gradient followed by a high salt strip can be directly recognized in the spectral data. Additionally the entire spectrum is distorted as the salt concentration changes, which is most clearly visible during the high salt step around 3000 s. Both of these effects are accounted for by the spectral correction method, leaving spectral data with very little influence from the changing buffer composition. The entire correction of the spectral data is only based on coefficients of the first four components of

the PCA of the background run. Therefore, this method is more generally applicable than the simple subtraction we used in our previous work.

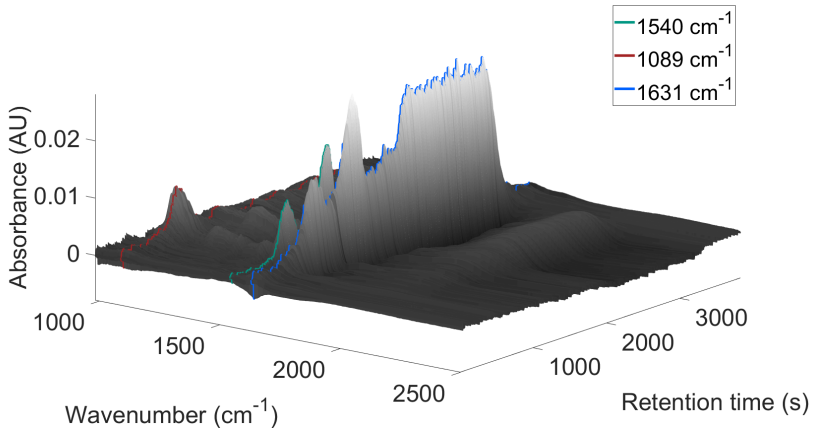

Figure 4.3: Raw spectral data collected during the elution of the on-column-PEGylation experiment. The highlighted traces show the positions of the bands used for the calculation of the DOP, 1540 cm^{-1} in green and 1089 cm^{-1} in red. Furthermore, the blue line highlights the wavenumber with the most prominent influence of the salt gradient at 1631 cm^{-1}.

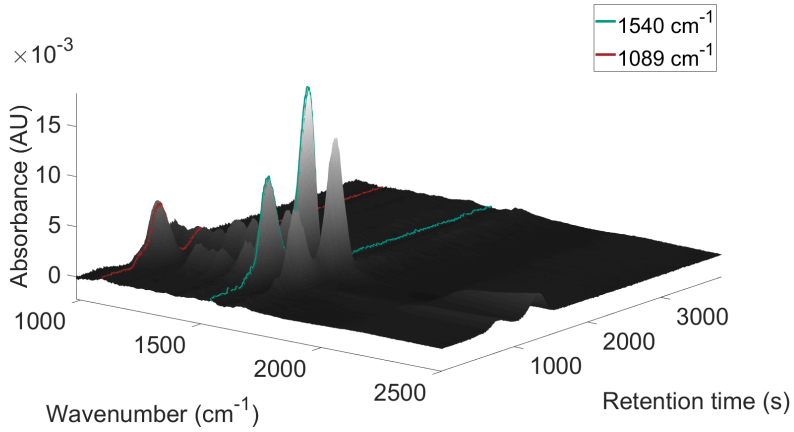

Figure 4.4: Spectral data collected during the elution of the on-column-PEGylation experiment, corrected by EMSC-ALS. The highlighted traces show the positions of the bands used for the calculation of the DOP, 1540 cm^{-1} in green and 1089 cm^{-1} in red.

4.3.3 Correction of spectral fraction with the ratio of extinction coefficients and offset

The extinction coefficients were determined to be 8333.33 AU/mol for PEG and 19 087.61 AU/mol for lysozyme with coefficients of determination of $R^2 = 0.998$ and $R^2 = 0.967$ respectively. From this, the correction factor was calculated to be $F_{DOP} = 2.29$. Additionally, the offset between the raw fraction and the true degree of PEGylation when native protein is eluting from the column was determined to be 0.048. With this data, the raw fraction from the spectral data is corrected to resemble the actual numerical value of the DOP more closely. As the fraction is smoothed using a Savitzky-Golay algorithm with a window length of 13 timesteps, this introduces a delay of about 58 s when the method is implemented as shown here. This delay is caused by the amount of data points needed to apply the smoothing algorithm.

4.3.4 Comparison to off-line analytics

In figure 4.1 the raw fraction of spectral data (blue trace) can be evaluated in relation to the UV signal (purple trace) and the IR signals (red and green traces). Before 500 s and after 1550 s, the raw fraction is very noisy as the absolute values of the IR signals are small and close together. The stabilization of the fraction coincides with the rise of the UV signal as well as the rise of the IR signals. Therefore the UV signal can be utilized to assess whether actual information can be gained from evaluating the IR signal. To compare the estimated DOP from the spectral data to the off-line analytics the IR signal was only utilized for all points in time where the UV signal rose above twice its baseline value, as shown in figure 4.5. Comparing figures 4.2 and 4.5 shows that overlapping elution profiles of different species lead to non integer values as result for the DOP which was also expected. Overall, the estimated DOP from the spectral data is in accordance with the DOP calculated from the off-line data. While there is still a discrepancy between the offset calculated from the injection of lysozyme and the offset during the on-column-PEGylation experiment the proposed method can still be utilized to accelerate process development. Despite not being able to estimate the exact numerical value of the DOP, it is accurate enough to serve as a basis for rapid process optimization without the need for tedious off-line analytics.

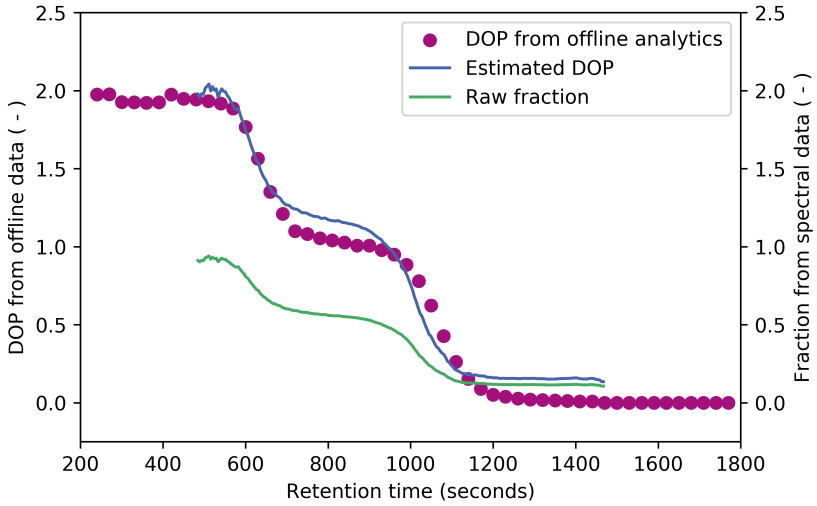

Figure 4.5: On-column-PEGylation: Comparison of the degree of PEGylation calculated from off-line analytics and the estimated DOP based on the spectral data as well as the raw fraction for comparison.

4.4 Conclusion and Outlook

An experimental setup for recording FTIR spectral data in-line in the effluent of a chromatography experiment was successfully utilized to estimate the degree of PEGylation of eluting protein species in near real time. To our knowledge, a combination of EMSC and ALS was applied to FTIR spectra from ion exchange chromatography effluent for the first time and it was shown that this technique can be used to eliminate the influence of the salt gradient. The spectral data was further processed utilizing certain consideration derived from the Beer-Lambert law *ab initio* enabling the estimation of the degree of PEGylation. Further efforts should be directed towards implementing the method with a lower time delay, for example by leveraging novel technologies that lead to a lower level of noise and investigating methods to reduce the detector drift to attempt to reduce the influence of the offset between the estimate and the true value of the DOP. If the elution should be run at higher flow rates, it may also be necessary to reduce the measurement times, which could for example be achieved by decreasing the spectral resolution. Additional effort could also be directed at exploring the generalization of the proposed method. While there is good evidence that the background correction is robust in regards to gradient lengths and shapes deviating from the background run, the entire windows of applicability is yet to be explored. As both PEG

and protein are essentially polymers whose absorption is mostly caused by the building blocks of their backbone, the applicability of the mass based absorption coefficients to different chain lengths of PEG and other proteins may be worth investigating. More specifically, it may be interesting to investigate if absorption coefficients based on the backbone structures alone could be utilized to generalize the method to any combination of protein and PEG-reagent chain length.

Acknowledgment

I would like to thank Susanna Suhm for her invaluable assistance with the experimental work.

5

Modifying an ÄKTApurifier System for the Automated Acquisition of Samples for Kinetic Modeling of Batch Reactions

Adrian Sanden[1], Sandra Haas[1], Jürgen Hubbuch[1]

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

Abstract

Recording the data necessary to assess the kinetics of a reaction can be labor intensive. In this technology brief, we show a method to automate this task by utilizing parts of an ÄKTApurifier chromatography system to automatically take samples from a reaction vessel at predefined time-intervals into 96-well plates and also enable correlating the samples with in-line spectral data of the reaction solution. Automatic sampling can reduce experimental bottlenecks by enabling over-night reactions or a higher degree of parallelization. To demonstrate the feasibility of the method, we performed batch-PEGylation of lysozyme with varying conditions by changing the molar excess of the PEG reagent. We used analytical cation exchange chromatography to analyze the samples taken during the batch

reaction, determining the concentrations of the individual species present at each time step. Subsequently, we fitted a kinetic model on this data. Fitting the model to four different reaction conditions simultaneously yielded a regression coefficient of $R^2 = 0.871$.

5.1 Introduction

The acquisition of experimental data to assess reaction kinetics can be tedious work, especially if sampling is done manually or for long periods of time. Manual sampling is limited by working hours and liquid handling stations are by volume constraints. Examples for repurposed hardware to facilitate this kind of task can also be found in literature.[111–113] Direct coupling of the reaction vessel to GC/MS [114] and thermal conductivity detectors [115] was reported for gaseous samples, along with manual sampling of the solid residue for FTIR and XRD analyses.[114] Manual sampling by pipette can also be an option for reactions in the liquid phase in open systems.[116] Switching a valve to redirect circulating fluid to another instrument for analysis reduces the required manual labor.[117] For closed systems, use of an injection plug for sampling liquids by syringe was also reported.[118] Some reactions are performed under pressure, which allows for sample removal by sampling line, using the pressure difference to the environment as a driving force.[119, 120] Reactions that cause a color change can be followed in 96-well plates in a spectroscopic plate reader.[121] Instead of sampling a bulk reaction repeatedly, starting the reaction at different times in 96-well plates on a liquid handling station also yields time-resolved data.[101] Generally the problem to solve is acquiring samples in a time-resolved manner and immediately analyzing these samples or stopping the reaction in the sample for later analysis. Here we introduce a modification to an ÄKTApurifier system that enables the automatic acquisition of samples from a batch reaction vessel by modifying the flowpath of the fractionator using custom 3D-printed parts. We chose the PEGylation reaction of a model protein as a proof-of-concept example, as the PEGylation of proteins is highly relevant for the biopharmaceutical industry.[17, 122] Additionally, the possibility to acquire spectral data correlating to the reaction is outlined using the Aux signals from the chromatography system in conjunction with a USB-6008 device from National Instruments. Monitoring of chemical reactions utilizing in situ spectroscopy has been utilized before [123, 124] and could be a useful tool for process analytical technologies, where even minis-

cule changes in the absorption profile of a batch reaction can be used to determine its progress.[125]

5.2 Materials and Methods

5.2.1 Experimental Setup

We conducted all batch-PEGylation experiments in a standard 50 mL reaction vessel. To enable the automatic circulation of the reaction solution through the Tensor 27 FTIR instrument with the BioATRCell II flow cell controlled by OPUS 7 (all Bruker Optics, Ettlingen, Germany), we attached the reaction vessel to a P-960 sample pump and P-950 fraction collector (both GE Healthcare, Little Chalfort, United Kingdom) with a custom-made 3D-printed adapter, adapted from an existing project.[126] The reaction vessel was connected onto the adapter using the threads for the lid. Using the adapter, we attached the reaction vessel to the autosampler's waste port, so that the reaction solution circulated back into the vessel when the autosampler was in the waste position. Additionally, we fitted the autosampler with an additional 3D-printed adapter to hold a magnetic stirrer (IKA Topolino, IKA-Werke GmbH, Staufen, Germany) next to the reaction vessel without interfering with the movements of the sampling arm. Both adapters were printed using an Ultimaker Original+ (Ultimaker, Geldermalsen, Netherlands) 3D-printer using polylactic acid (PLA) filament. Thus, the sample pump aspirated liquid from the reaction vessel, routed it through the FTIR instrument and then into the autosampler. Depending on the current position of the autosampler arm, the liquid dropped back either into the reaction vessel through the adapter or into a 96-well plate on the sampling platform. The sample pump and autosampler were controlled by Unicorn (GE Healthcare). We included a schematic representation of the experimental setup in 5.1 and a photo in 5.2. The delay volume in Unicorn can either be defined as the total volume between the reaction vessel and the outlet of the fractionator so that the sampling time correlates to the reaction time in the batch vessel, or to the delay from the spectrophotometer. We determined the delay volumes by weighing the water required to fill the tubing. The pump was calibrated using buffer and we assumed the density to be constant. FTIR spectra were recorded using the chromatography mode in OPUS, repeatedly recording single spectra at a resolution of 2/cm. To correlate offline analytics to recorded spectra, we also determined the delay volume

from the measuring chamber to the fractionator gravimetrically. For all experiments, we set the thermostat for the BioATRCell II flow cell to 22 °C.

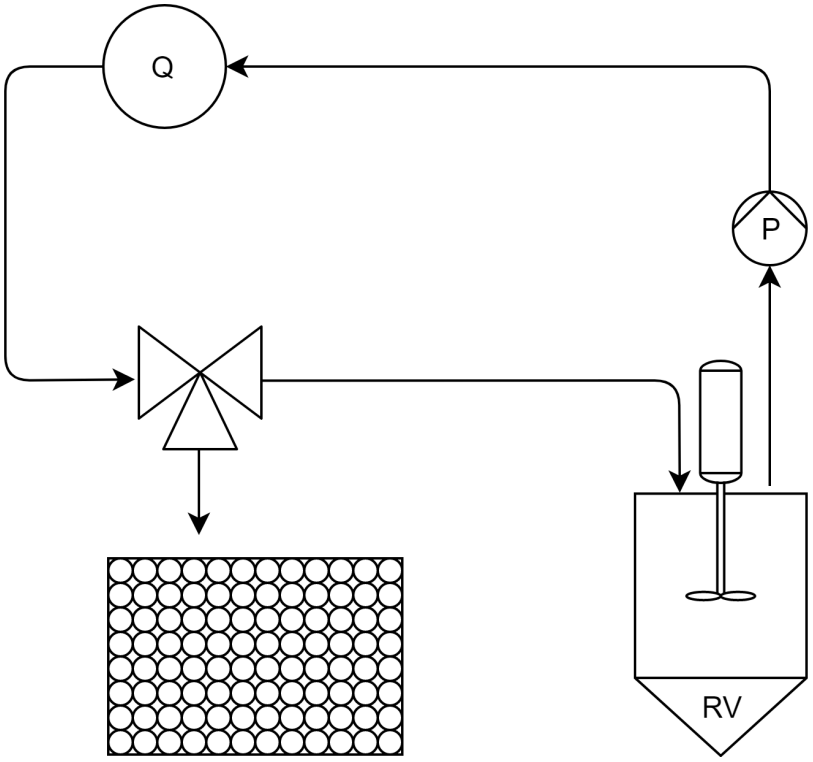

Figure 5.1: Schematic representation of the experimental setup. The reaction happens in the reaction vessel and the liquid is circulated through the attached spectroscopy device. The fractionator is represented as an automatic three-way valve, measurement point Q is the in-line spectroscopy instrument, P is the pump and RV the reaction vessel.

5.2.2 Proteins and Buffers

Purified water from a PURELAB Ultra water purification system (ELGA Labwater, High Wycombe, United Kingdom) was the base for all solutions. Before use, we filtered all buffers using 0.2 µm filters from Sartorius (Göttingen, Germany) and additionally degassed chromatography buffers in a sonification bath for 30 minutes before use. All pH adjustments were done using 32 % HCl.

For all experiments, we used lysozyme from Hampton Research (Aliso Viejo, CA, United States), sodium cyanoborohydride, lysine and sodium citrate were purchased from Sigma-Aldrich (St. Louis, MO, United States).

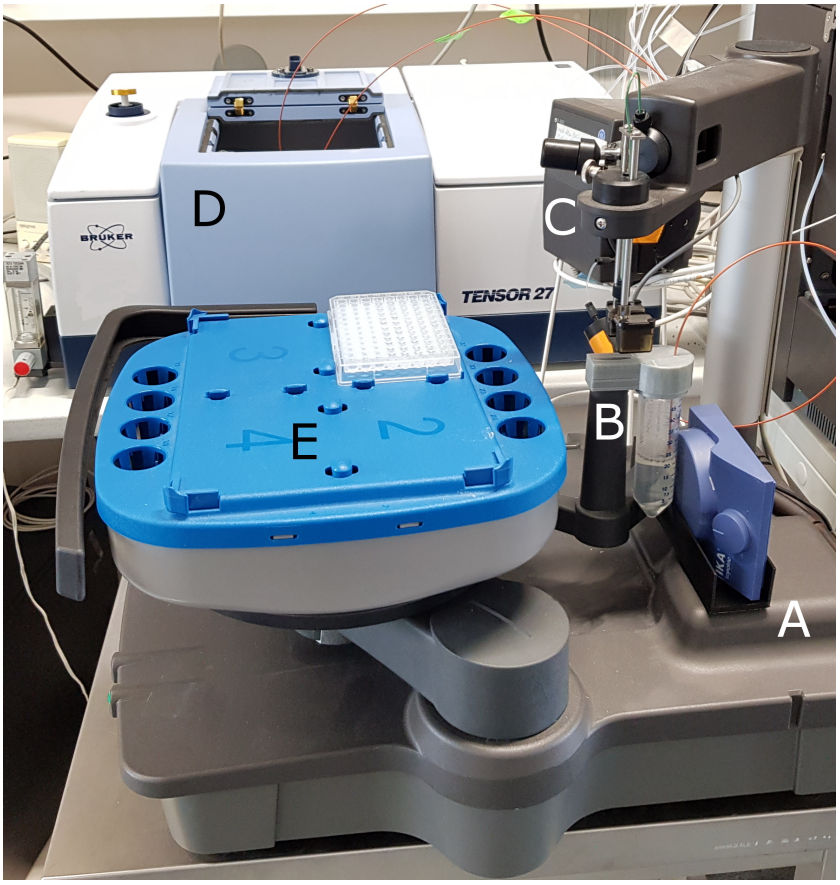

Figure 5.2: Experimental setup for the supervision of the batch PEGylation experiments. A: Adapter with magnetic stirrer; B: Adapter for circulation with reaction vessel; C: Pump; D: FTIR instrument and E: Fraction collector. In this photo the autosampler is in the waste position circulating the reaction solution back into the vessel through the adapter.

We ran all PEGylation experiments in 25 mM sodium citrate buffer at pH 6.8 with a concentration of cyanoborohydride of 25 mM. We obtained the activated 5 kDa PEG from NOF Corporation (Tokyo, Japan) in the form of methoxy-PEG-propionaldehyde (mPEG-aldehyde, SUNBRIGHT ME-050 AL).

We carried out offline analytics using analytical cation-exchange chromatography with 20 mM Tris at pH 8.0 as the low-salt buffer and with an additional 700 mM NaCl for the elution buffer (both Merck, Darmstadt, Germany).

5.2.3 PEGylation of Lysozyme

We adapted the PEGylation protocol from work previously published by our group.[127] Briefly, we dissolved activated 5 kDa PEG and lysozyme in citrate buffer without sodium cyanoborohydride. Immediately before the begin of the experiment, we filtered the solution through a 0.45 µm syringe filter and added it to 50 mM sodium cyanoborohydride in citrate buffer, to achieve a final concentration of 25 mM cyanoborohydride. We set the mPEG-aldehyde concentrations to nominal molar PEG-to-protein ratios of 2,4, 6 and 8, with the target initial concentration of lysozyme set to 5 mg/mL for all experiments. All batch-PEGylation experiments ran for 320 min. The entire laboratory was air conditioned at 22 °C. The reaction solution circulated at 1 mL/min and we set the system up to collect 200 µL samples every 4 min automatically. Samples were collected by the fraction collector into a 96-well plate containing 100 µL of a 600 mM solution of lysine to stop the reaction according to Ottow et al.[128]

5.2.4 Analytical CEX Chromatography

Analytical chromatography was used to analyze reactions in a similar way before[100] and we conducted the method previously published by our group.[103] We ran all analytics using a Vanquish UHPLC system with a ProSwift SCX-1S 4.6 mm by 50 mm column (both Thermo Fisher Scientific, Waltham, MA, USA). The Vanquish UHPLC System consisted of a diode array detector (DAD) HL, binary pump F, column compartment H with integrated preheater and post-column cooler as well as a split sampler FT. Prior to all analytics, the column was equilibrated with low-salt buffer. Five µL of sample were injected from each time step of the reaction and the column was washed for half a minute with low-salt buffer to flush out unbound compounds. Subsequently, a bilinear gradient ran 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 of one minute at 100 % followed. Finally, the column was re-equilibrated with low-salt buffer for two minutes before the next injection. We based the quantification of all species on a dilution series of pure lysozyme, while relying on the previously published peak identification.[55, 103]

5.2.5 Data Analysis and Kinetic Reaction Model

We used MATLAB R2019b (The MathWorks, Natick, MA, United States) for all data analysis to fit the data from the offline analytics to a kinetic

reaction model adapted from Moosmann et al.[100] Specifically, we used the model including the inactivation reaction of the PEG reagent, omitting the reaction from di-PEGylated to tri-PEGylated lysozyme. The system of differential equations was solved numerically with the ode45 solver and lsqnonlin to minimize the difference to the experimental data. The Jacobian matrix calculated with lsqnonlin was used to calculate the 95 % confidence intervals with nlparci. Based on irregularities in the pump rate, we excluded certain samples, which resulted in undefined dilution with stop solution. The definitions of reaction rates and the set of differential equations for the kinetic model lead to Equations (1) - (7). The initial conditions for the determination of the rate constants were the measured initial concentrations of protein and the set point for the PEGylation reagent.

$$r_1 = PEG_{reacting} * Protein * k_1 \tag{5.1}$$
$$r_2 = PEG_{reacting} * \textit{Mono-PEG-Protein} * k_2 \tag{5.2}$$
$$r_3 = PEG_{reacting} * k_3 \tag{5.3}$$
$$dx(Protein) = -r_1 \tag{5.4}$$
$$dx(\textit{Mono-PEG-Protein}) = r_1 - r_2 \tag{5.5}$$
$$dx(\textit{Di-PEG-Protein}) = r_2 \tag{5.6}$$
$$dx(PEG_{reacting}) = -r_1 - r_2 - r_3 \tag{5.7}$$

5.3 Results and Discussion

5.3.1 Kinetic Modelling of the PEGylation Reaction

We fitted the data for the PEGylation reaction to the reaction model shown in Equations (1) – (7), yielding a regression coefficient of $R^2 = 0.871$ across the four data sets. Table 5.1 shows the calculated rate constants. The 95 % confidence intervals correspond to 4.3 %, 6.1 % and 15.6 % of the parameter value respectively. As previously reported for the utilized kinetic model, k_1 is larger than k_2. A possible explanation for this observation may be that the di-PEGylated species originates from the mono-PEGylated species and thus access to the surface of the lysozyme molecule is sterically hindered by the attached PEG molecule.[47] Rate constant k_1 being larger than rate constant k_2 is in line with previously published findings as well. Both are orders of magnitude larger than k_3. We summarized these results in 5.3.

Table 5.1: Model parameters for the kinetic reaction model

Rate constant	Value and 95 % confidence interval
k_1	$(2.846 \pm 0.122)\,\mathrm{L/(mol\,min)}$
k_2	$(2.055 \pm 0.126)\,\mathrm{L/(mol\,min)}$
k_3	$(0.0032 \pm 0.0005)/\mathrm{min}$

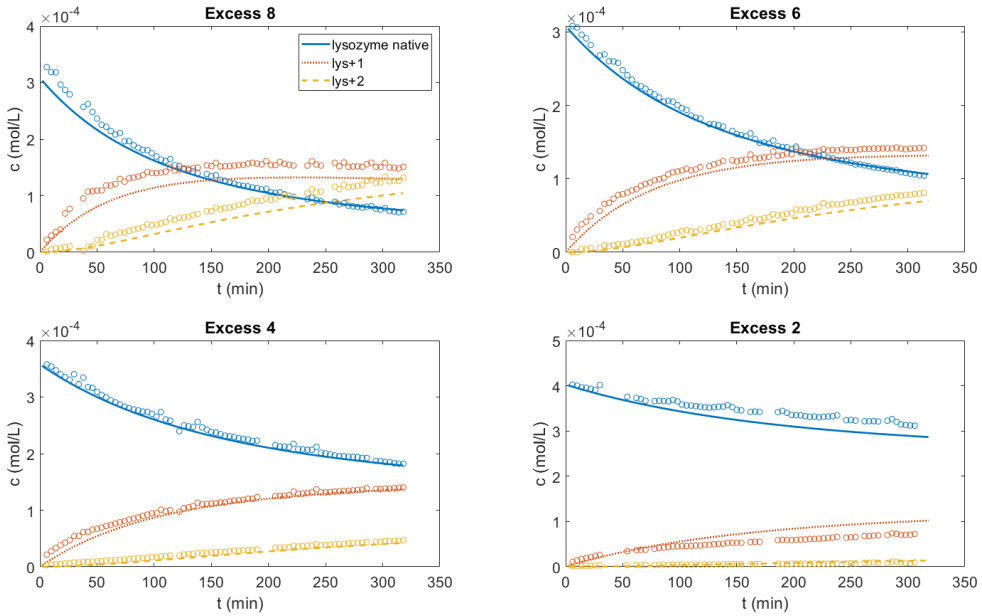

Figure 5.3: Results of the kinetic model. Shown in a comparison between the model response (solid, dotted and dashed lines) and offline data (hollow circles) without excluded data points for the four calibration runs. The lines for lys+1 and lys+2 refer to mono- and di-PEGylated lysozyme, respectively.

5.4 Conclusion and Outlook

We were able to show that the proposed experimental setup enables the automation of sample acquisition from a batch reaction process. Depending on the limitations of pump accuracy and total reaction time, most reactions can be monitored using this setup, if the stop solution is stable at room temperature. Depending on the reaction at hand, it might also be of interest to change the instrument method in the chromatography setup to have the sampling frequency change over the course of the reaction under the assumption that the initial time steps will contain the most rapid changes in conditions. With modifications to the reaction vessel, it would also be possible to monitor heterogeneously catalyzed reactions by placing

the catalyst in the reaction vessel. However, proper mixing of the solution needs to be ensured in this case. If the reaction at hand utilizes reactants that are more stable in solution, preparing the reaction solution from stock solutions might decrease errors associated with weighing small quantities.

Acknowledgment

I would like to thank Sandra Haas for her assistance with the experimental work. I would also like to thank Sebastian Andris for his insights into implementing the reaction kinetics.

6

ATR-FTIR for *in situ* Reaction Monitoring of Protein PEGylation in Batch Mode

Adrian Sanden[1], Sandra Haas[1], Jürgen Hubbuch[1]

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

Abstract

Chemical batch reactions are usually run based on prior knowledge in the form of kinetic data. In essence, the precursors for the desired product are brought into contact with each other at known conditions and the reaction takes its course over a predetermined amount of time. After this time the desired product composition should be reached according to the kinetic model. Here, a PAT method based on FTIR spectroscopy is proposed to estimate the reaction progress in near real-time to ensure the reaction ran as desired. To this end, and automation setup based on an ÄKTApurifier chromatography system, an FTIR instrument and custom 3D-printed parts is proposed to correlate spectral data to reference analytics and kinetic data. Four separate batch reactions were conducted with varying excesses of the PEGylation reagent. A PLS regression model was fit on the spectral data from three of these experiments, using the fourth as an external validation

set. It was evaluated if the spectral data obtained in the batch reaction solution was suitable to estimate the state of the reaction.

6.1 Introduction

To produce a PEGylated protein, the first step is to produce the host protein. This is carried out in the same way any other biopharmaceutical protein would be produced, yielding the purified target protein as raw material for the PEGylation process. Most commonly, the PEGylation itself is carried out as a batch reaction process where the protein and conjugation reagent are brought into contact in solution. [98] By nature, the commercially available PEGylation reagents will not lead to a single species as the reaction product, as there are multiple reaction sites in most proteins and thus random PEGylation in a batch reaction leads to a mixture of reaction products. [95, 100] This makes it also obvious why it is desirable to remove other contaminants before the PEGylation reaction – any Host Cell Protein (HCP) will also react into multiple different species that have to be removed afterwards. One of the most common reaction chemistries targets lysine residues which account for about 10 % of a typical protein. [19]

To control the outcome of the reaction, several strategies can be employed. Modeling the kinetics of the reaction itself can be employed to determine how long the reaction needs to run to achieve the desired outcome based on the initial conditions. [99, 100, 129] This method can also be combined with solid-phase PEGylation. By binding the parent protein to a solid support such as chromatographic resin before the reaction influences the kinetics and distribution of reaction products. [48, 101] Integrating a process with the separation of the target species from the reactant continuously can be employed to hinder the reaction of mono-PEGylated species to higher degrees of PEGylation. [49, 130] Instead of steering the process by influencing the technical side, direct influence on the chemistry can also lead to the desired results. The pH of the reaction can influence the specific site favored by the reaction [19] and certain chemistries are more specific to certain sites on their own. [95] The greatest site specificity can be achieved by genetic insertion of unnatural amino acids with moieties that don't occur anywhere else and utilizing them for the conjugation reaction. However, such an approach has to be carefully validated to ensure the safety and efficacy of the resulting genetically modified parent protein. [95]

All of these strategies have in common that they rely on prior knowledge to determine their endpoint. The desired initial conditions are set and the reaction is run for a predetermined amount of time. A method to determine the state of the reaction in situ would thus be desirable. It has been previously shown that ATR-IR, UV/Vis and Raman spectroscopy can be utilized to monitor reactions *in-situ*. [123, 124, 131] It has also been shown that a small peak shift in UV/Vis spectroscopy can be utilized to monitor an antibody-drug conjugation reaction. [125]

Here, we utilize a modified ÄKTApurifier chromatography to explore the feasibility of monitoring the PEGylation reaction of a protein with *in-situ* ATR-FTIR. Custom 3D-printed parts were used to modify the flowpath of the fractionator and the sample pump. In combination with custom software and an external communication device attached to the chromatography control system this enabled the automated acquisition of spectral data and reference samples.

6.2 Materials and Methods

6.2.1 Experimental Setup

Lysozyme was modified by batch-PEGylation in a standard 50 mL reaction vessel attached to a fraction collector and sample pump from an ÄKTA-purifier System (GE Healthcare, Little Chalfort, United Kingdom) with a modified flow path to enable circulating the reaction solution and automatic sampling. The flow path was modified using custom designed and 3D-printed parts that were printed in-house on an Ultimaker Original+ (Ultimaker, Geldermalsen, Netherlands). Both 3D-printed parts are shown in Figure 6.1. A Bruker Tensor 27 FTIR instrument (Bruker Optics, Ettlingen, Germany) was added to the flowpath to record spectral data during the reaction. The FTIR instrument was fitted with an LN-MCT detector. The flow path is shown in the diagram in Figure 6.2. A photograph of how the entire setup fits together can be found in Figure 5.2 in the previous chapter. As the fractionator and sample pump were controlled by a method in the Unicorn control software, the same automation methods for the spectral acquisition could be applied as in previously published work. [48, 103, 129]

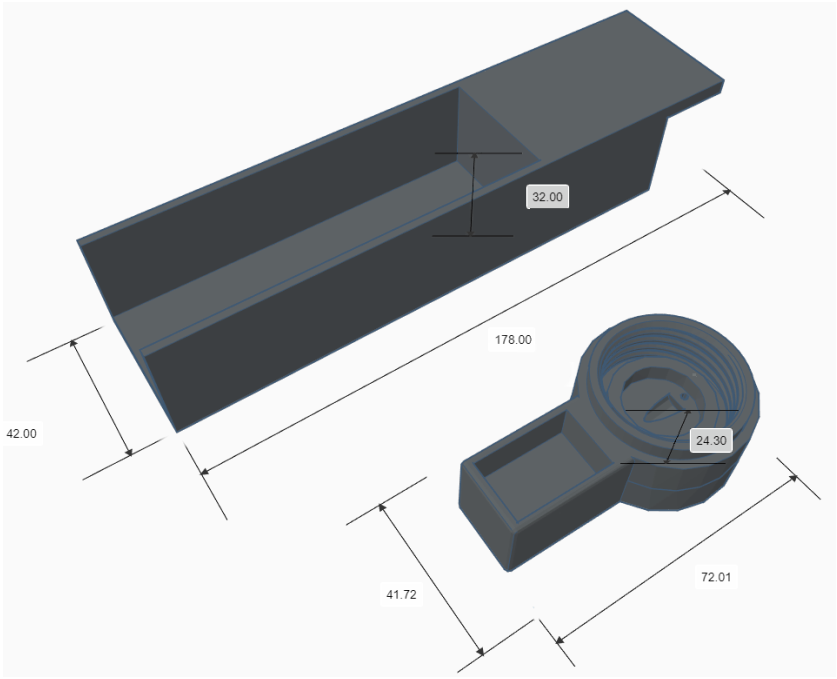

Figure 6.1: The 3D-printed parts that were used to adapt the flow path of the chromatography device. The upper part holds the magnetic stirrer in the large slot on the left hand side and can be attached to the arm that holds the waste port without interfering with the movement of the fractionator. The lower part screws onto a reaction tube with the threads on the right, the slot on the left is used to attach it to the waste port of the fractionator. All measures are given in millimeters.

6.2.2 Batch PEGylation experiments

Lysozyme from Hampton Research (Aliso Viejo, CA, United States) was PEGylated in 25 mM sodium citrate buffer at pH 6.8 with an additional 25 mM sodium cyanoborohydride with activated 5 kDa PEG in the form of methoxy-PEG-propionaldehyde (mPEG-aldehyde SUNBRIGHT ME-050 AL from NOF Corporation, Tokyo, Japan). The reaction chemistry creates a secondary amine between the lysine residue of the protein and the PEG reagent via a Schiff base. To introduce variance into the data, four different experiments were conducted with a varying excess of PEGylation reagent in relation to the protein in solution. The molar excess was set to 2, 4, 6 and 8. All buffers were filtered using 0.2 µm PES filters from Sartorius (Göttingen, Germany). A 96-well plate containing 100 µL of a 600 mM solution of lysine was placed on the fractionator. Lysine was used to stop the reaction according to Ottow et al. [128] Samples were taken

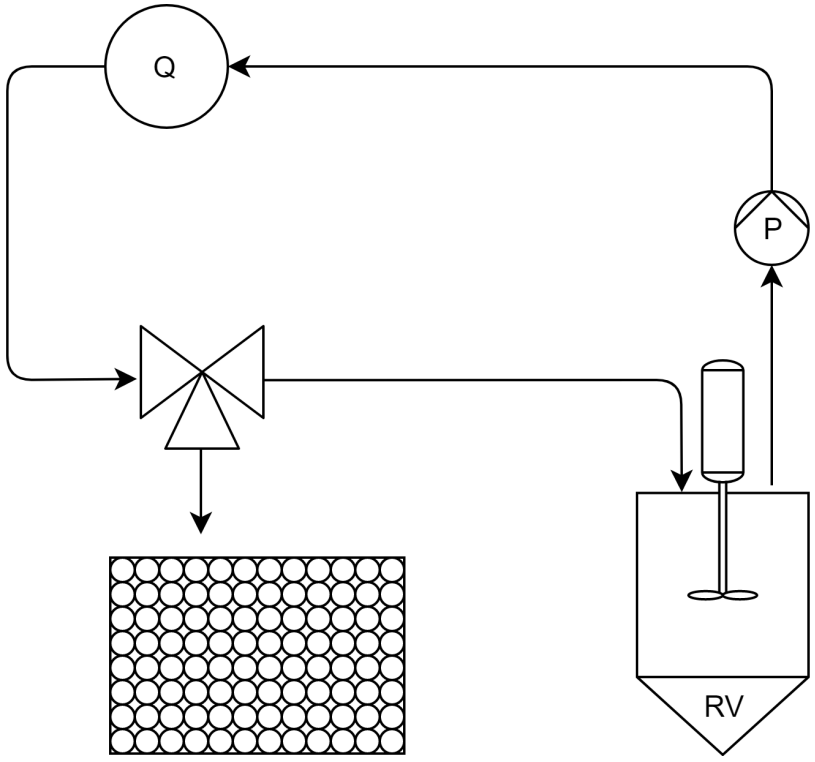

Figure 6.2: Schematic representation of the experimental setup. The reaction happens in the reaction vessel and the liquid is circulated through the attached spectroscopy device. The fractionator is represented as an automatic three-way valve, measurement point Q is the in-line spectroscopy instrument, P is the pump and RV the reaction vessel.

automatically every 4 min starting 2 min after the start of the reaction. FTIR spectra were acquired every 1.57 s at a resolution of 2 cm^{-1}. To record the individual spectra back to back, the chromatography mode in OPUS was used. The detector was filled with liquid nitrogen at least 30 min before any measurements were taken to allow for thermal equilibrium between the detector and the reservoir of liquid nitrogen.

6.2.3 Reference Analytics

Analytical Cation-Exchange (CEX) chromatography was carried out as reference analytics on a Vanquish UHPLC system with a ProSwift SCX-1S 4.6 mm by 50 mm column (both Thermo Fisher Scientific, Waltham, MA, USA). The UHPLC Sytem was fitted with a diode array detector (DAD) HL, binary pump F, column compartment H and split sampler

FT. A 20 mM Tris buffer at pH 8.0 was used as the low-salt buffer and an additional 700 mM NaCl for the elution buffer (both Merck, Darmstadt, Germany). All buffers were filtered using 0.2 µm PES filters from Sartorius (Göttingen, Germany) and degassed in a sonification bath for 30 minutes before use. The analytical routine consisted of low-salt buffer equilibration, a bilinear gradient 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 of one minute at 100 % and a re-equilibration of 2 min were performed before the next injection. A dilution series of pure lysozyme was used for calibration in combination with the peak assignment as previously published. [55, 103]

6.2.4 Data Analysis

All data analysis was performed using Python 3.7. Based on observed irregularities in the pump rate that led to an undefined dilution of the sample with stop solution, certain samples were excluded. The rest of the data was interpolated using the *pchip* interpolation method in the Python Pandas package to fill in these gaps and then expanded to the length of the spectral data using *interp1d*. This was done to have a data point of concentration assigned to every recorded spectrum. The total concentration of reacted PEG reagent was calculated by adding the molar concentration of mono-PEGylated lysozyme to two times the molar concentration of di-PEGylated lysozyme as shown in equation 6.1.

$$PEG_{reacted} = \textit{Mono-PEG-Lysozyme} + 2 * \textit{Di-PEG-Lysozyme} \tag{6.1}$$

The spectra were smoothed with a Savitzky-Golay filter and baseline corrected with the *peakutils* package. The concentration estimates were also smoothed with the same method. A window length of 151 and polynomial order of 3 was set for all smoothing operations.

6.3 Results and Discussion

In the following, two different methods of data analysis are compared. The first method uses the spectral data after smoothing it in spectral direction using a Savitzky-Golay filter. The second part shows the same general data anlysis, however the spectral data is also baseline corrected to eliminate detector drift and baseline offsets from it.

6.3.1 PLS Regression without Baseline Correction

The first version of data analysis was conducted without performing a baseline correction on the spectral data. The data was split into three calibration experiments and one external validation set. Figure 6.3 shows the spectral data of the three calibration sets. The bands between 1800 cm^{-1} to 1600 cm^{-1} are the Amide I and Amide II bands associated with protein. The band at about 1100 cm^{-1} is associated with PEG. These bands and the rest of the signal remains approximately constant over time as is expected because all of these components remain in the bulk solution. Figure 6.4 shows the results of the calibrated model. The model was calibrated on the experiments with a PEG excess of 2, 6 and 8 and the blue trace shows the models response to the training data. The following two figures, 6.5 and 6.6 show the spectral data and model response to the chosen external test set. The spectral data differs only slightly from the spectral data of the other experiments at first glance, however the absolute values differ due to the different concentrations of PEG reagent used. Figure 6.6 indicates that the model response (blue trace) is in good agreement with the experimental data (yellow trace). The purple trace shows the absorbance at 1540 cm^{-1}. As the model response follows this trace very closely there is a strong suspicion that the model is over fitted. One specific feature of the model response and the spectral trace that should be pointed out is the upward spikes such as the one at around spectrum 8000. These are caused by inconsistencies in the pump rate that could be observed in the pressure measurements of the chromatography system. The model response following this feature in the spectral data as well indicates that the model is not really fit to individual features that change in the spectral data over time.

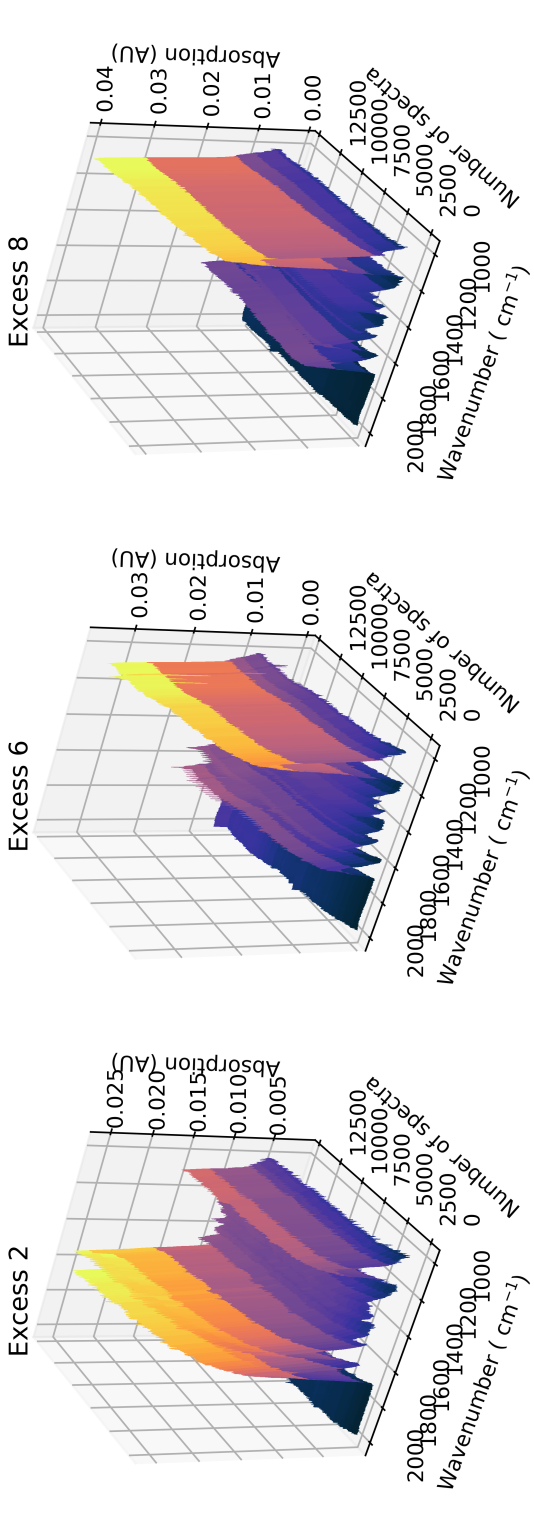

Figure 6.3: Spectral data of the calibration runs. The Amide I and Amide II bands are visible between 1800 cm^{-1} to 1600 cm^{-1} and the band associated with PEG at about 1100 cm^{-1}.

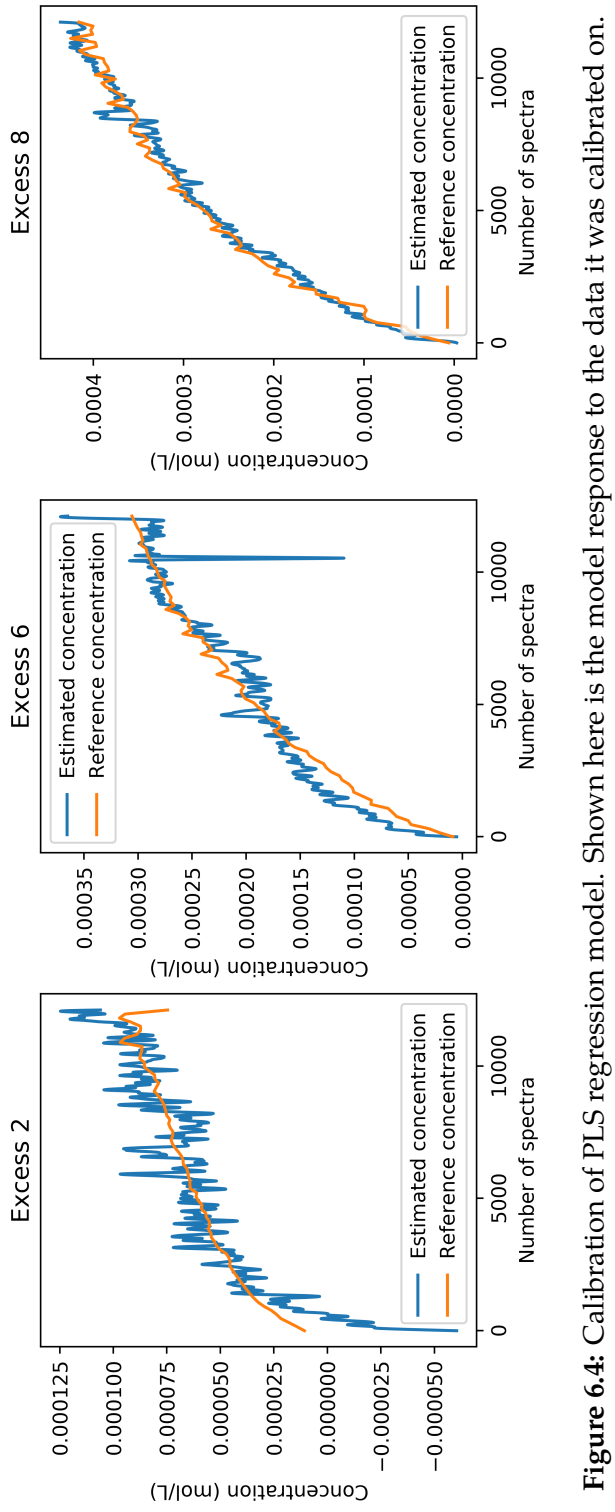

Figure 6.4: Calibration of PLS regression model. Shown here is the model response to the data it was calibrated on.

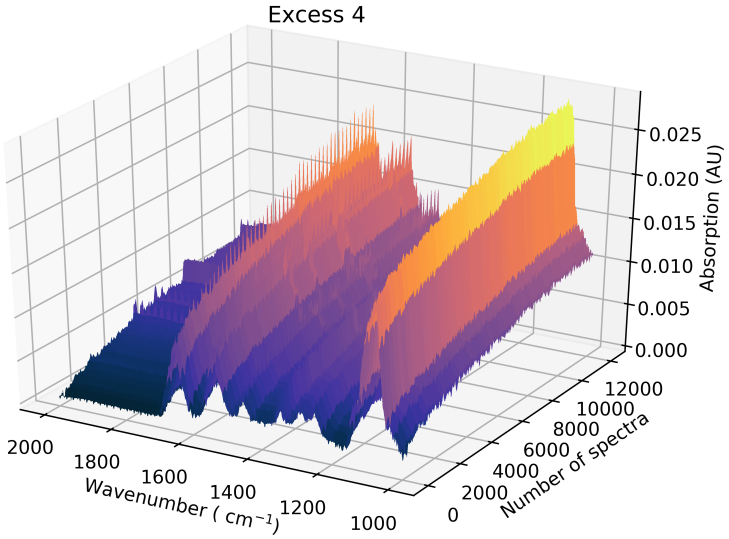

Figure 6.5: Spectral data of the validation run. The Amide I and Amide II bands are visible between 1800 cm^{-1} to 1600 cm^{-1} and the band associated with PEG at about 1100 cm^{-1}.

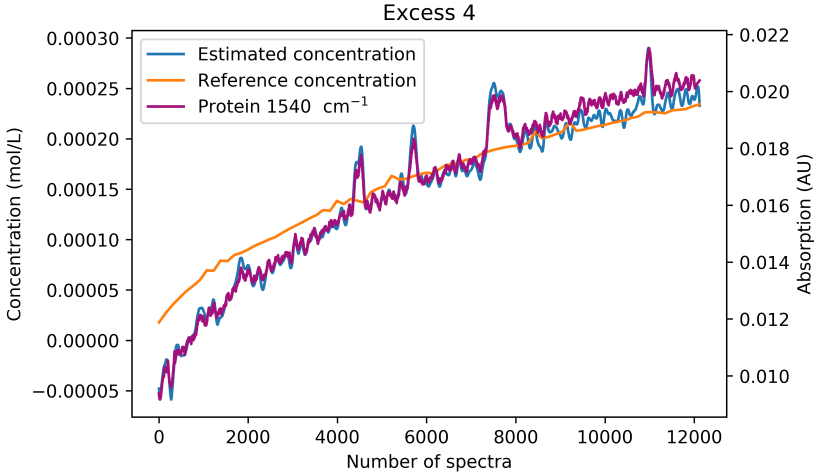

Figure 6.6: PLS regression model validation data of the model that was calibrated without baseline correction.

6.3.2 PLS Regression with Baseline Correction

From Figure 6.3 and Figure 6.5 it can be observed that there is baseline offset present in the spectral data. Both a small offset within each individual spectrum as well as a larger drift over time is present. In this second set of data baseline correction was applied to every individual spectrum before the evaluation of the PLS regression. Figure 6.7 shows the calibration data and 6.9 shows the validation data. The baseline correction significantly changes the appearance of the spectral data, both the offset in every individual spectrum and over time are not present anymore. While Figure 6.8 indicates that the calibration data is still in decent agreement with the model response, the same can not be said for the external validation set. As shown in Figure 6.10 there is a significant offset and difference in slope between the model response and the experimental data. Additionally, as pointed out for the previous set of data, the model response closely follows the absorbance at 1540 cm^{-1}, indicating that the PLS regression model can not actually estimate the state of the reaction in this case either. The same correlation between model response and inconsistent pump rate that was pointed out in the previous section can still be observed after baseline correction, albeit less pronounced.

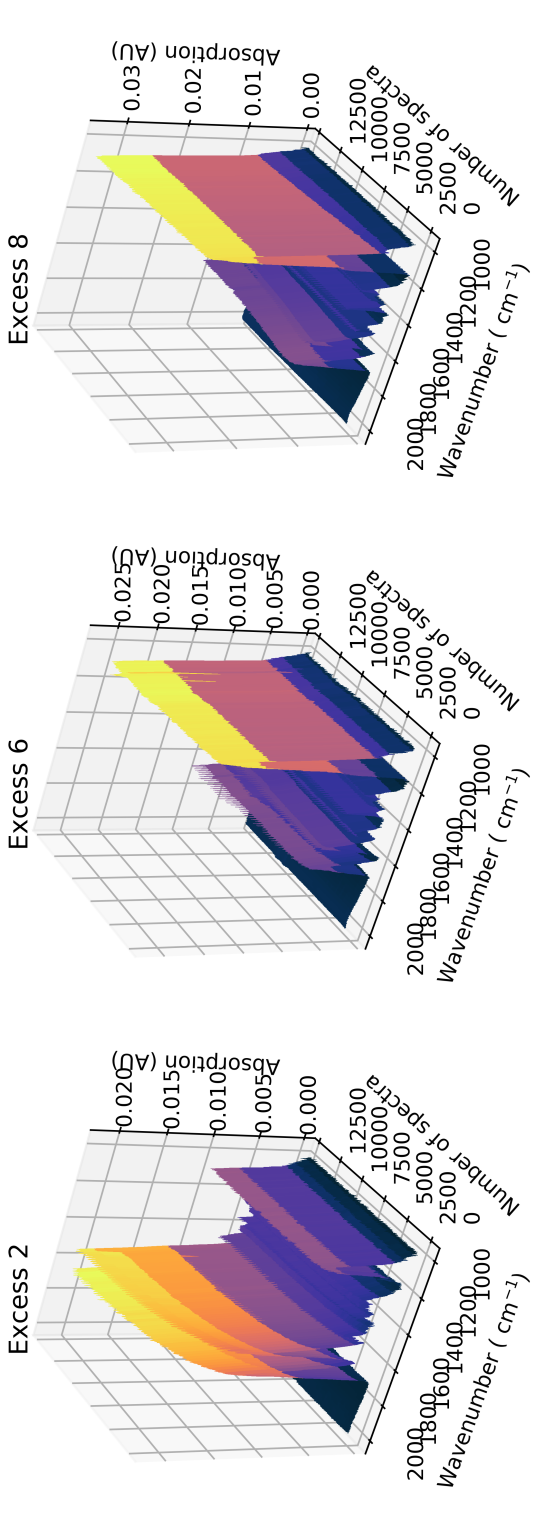

Figure 6.7: Spectral data of the calibration runs after baseline correction. The Amide I and Amide II bands are visible between 1800 cm^{-1} to 1600 cm^{-1} and the band associated with PEG at about 1100 cm^{-1}.

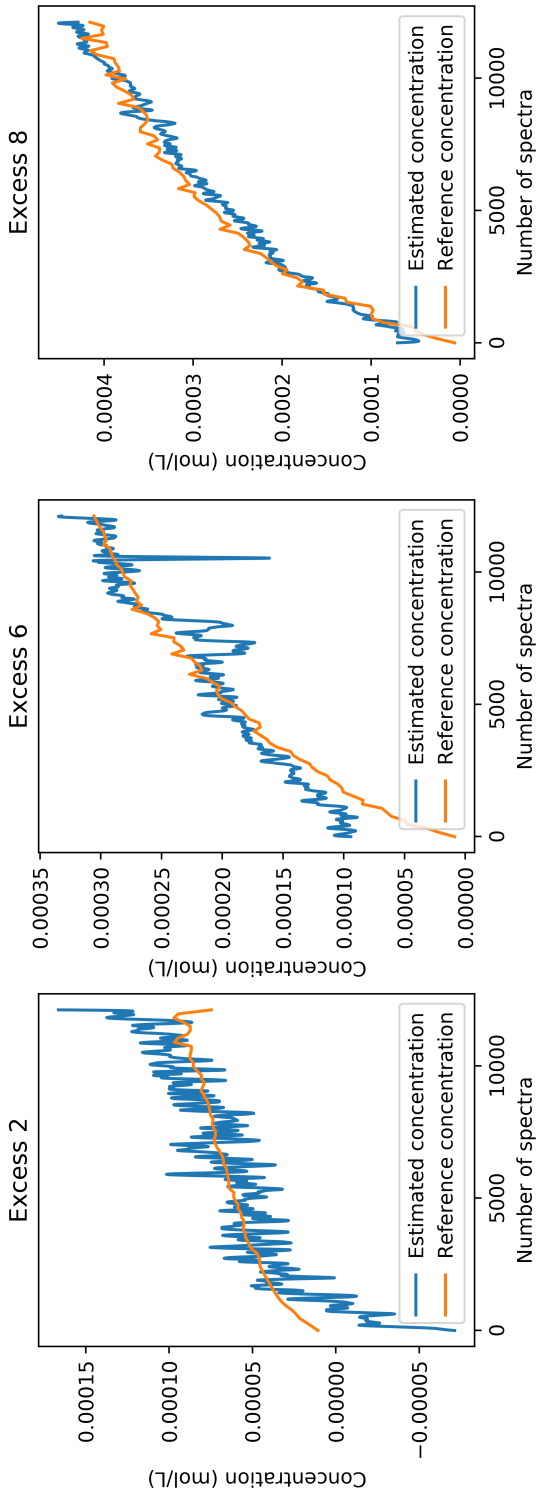

Figure 6.8: Calibration of PLS regression model on the spectral data that had baseline correction applied to it. Shown here is the model response to the data it was calibrated on.

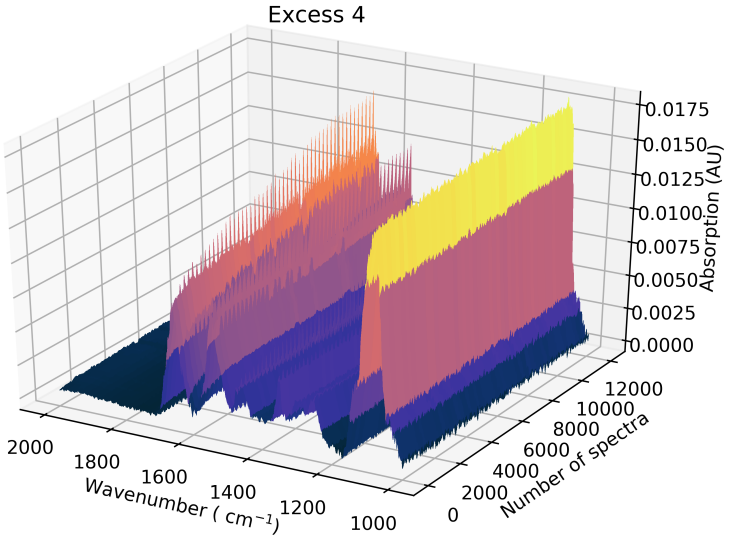

Figure 6.9: Spectral data of the validation run after baseline correction. The Amide I and Amide II bands are visible between 1800 cm^{-1} to 1600 cm^{-1} and the band associated with PEG at about 1100 cm^{-1}.

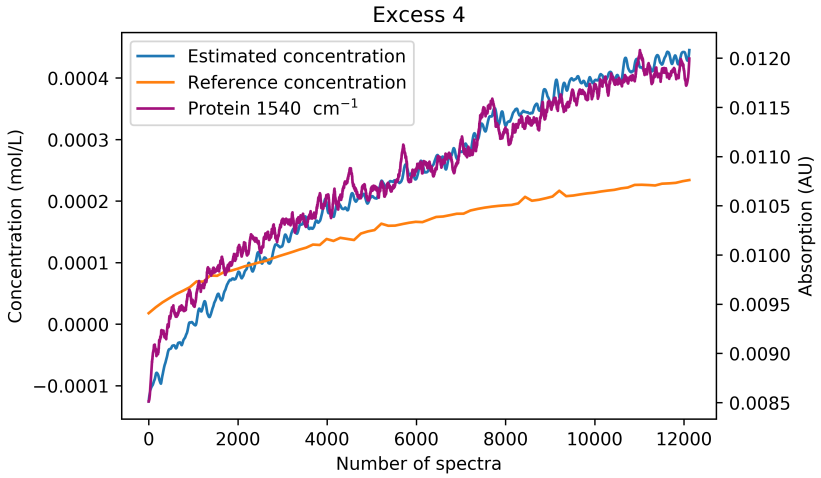

Figure 6.10: PLS regression model validation data for the data that had baseline correction applied to it.

6.3.3 Varying the external calibration set

In addition to model calibration with and without baseline correction, the experiment chosen for external validation was also varied. Based on the same full set of original data, Figure 6.11 shows the model response for the experiment with an excess of 2 and Figure 6.12 for an excess of 6 respectively. It is clear that by changing the validation set, the model performance becomes significantly worse. This indicates that the model was not only overfit in the previously shown cases, but even overfitting could only happen in one edge case. The model response itself still seems to only follow the absorption data itself, without a significant correlation to the reference data.

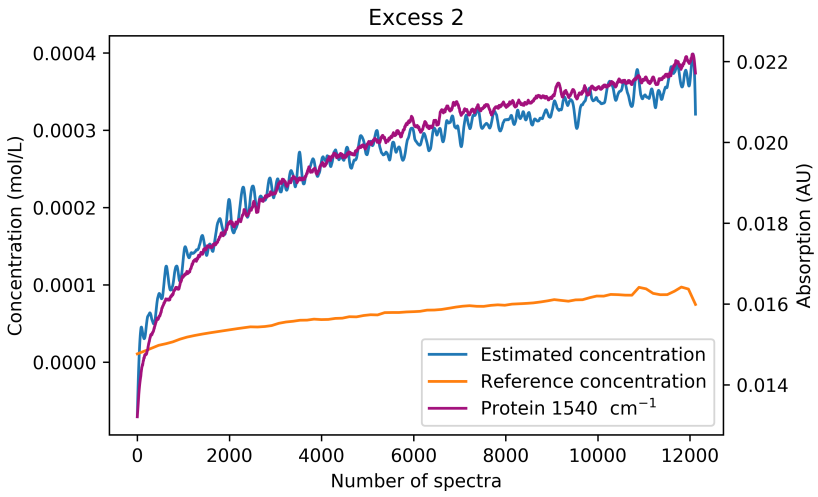

Figure 6.11: PLS regression model validation data with the experiment with PEG excess 2 as the validation set.

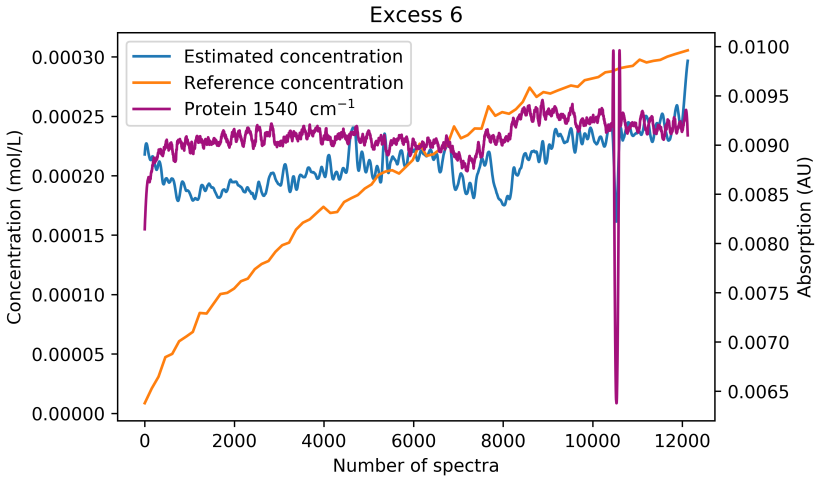

Figure 6.12: PLS regression model validation data with the experiment with PEG excess 6 as the validation set.

6.4 Conclusion

It is not feasible to monitor this specific reaction system *in situ* using this FTIR spectroscopy setup. There are multiple reasons that can be identified or alleged from the recorded data. Generally it is not trivial to predict whether it is possible to calibrate a PLS regression model on a given data set or not. [24] Generally, a successfully calibrated model relies on changes in spectral data. Even for different proteins with very similar spectra it can be possible to achieve very good results with this kind of PLS regression models and plenty examples for this can be found in literature. [54, 66, 68] For the specific case evaluated in this study, it has been previously shown that a PLS model can be calibrated on time resolved chromatography data. This means that the same protein with differing numbers of PEG molecules attached are different enough to calibrate a model when the data is time-resolved. Part of the reason for this is likely that the different species have different ratios of the values of the absorption bands associated with PEG and protein, but the model response is also accurate where the species overlap during elution. [48] Here, as all species including the PEG reagent are present in the bulk solution simultaneously, this ratio can not be utilized by the model. This indicates that the change in the spectrum of a protein molecule is either not large enough to be picked up by model calibration or that the total number of molecules that exhibit a change is too small. In addition to these spectral changes the chemical reaction itself creates a new bond that absorbs in IR. The chemical reaction adds the PEG molecule to

a lysine residue of a protein molecule via a Schiff base transition state and forms a secondary amine bond. [100] This secondary amine bond has several wavenumbers at which it absorbs in FTIR. One option to utilize changes in specific spectral changes is to extract the absorption data over time from the spectra and tracking the changes directly. Utilizing such a method to estimate the degree of PEGylation in the column effluent of a chromatography system has been shown to be feasible before. [48] The secondary amine absorbs between about 3600 cm^{-1} to 2700 cm^{-1} as well at about 1600 cm^{-1} and 1200 cm^{-1} to 400 cm^{-1}. [30] In aqueous samples, the 3600 cm^{-1} to 2700 cm^{-1} range is generally not usable as the detector is saturated by water absorption. As in previous studies, the spectral data in this range was discarded before any further processing was attempted. Comparing the remaining absorption ranges with the spectral data at hand shows that these ranges overlap with the protein absorption spectrum itself. For the batch reaction experiment, the concentration of protein was about 0.0003 M, which is about the same amount of new secondary amine bonds that form during the reaction. Lysozyme, the protein that was used for this study, has a primary chain length of 129 amino acids and a molecular weight of about 14 kDa. [132] This means that the absorption of the Amide I band caused by the polypeptide backbone that is in the range of about 1700 cm^{-1} to 1600 cm^{-1} [30] is about two orders of magnitude larger than the absorption that would be caused by the bond formed during the reaction. While the correlation between the molar amount of absorbing bonds does not have a direct linear correlation to the absorption, this is still a probable cause for it not being possible to have a PLS regression model utilize this spectral change.

6.5 Outlook

There are several possibilities to evaluate to try making *in situ* monitoring of a PEGylation reaction in aqueous solution feasible. One possible point of leverage could be the spectrometer itself. A different device with improved resolution, sensitivity and SNR could lead to spectral data from which more information could be extracted. Conducting the experiments at overall higher concentrations may also help make changes in the spectral data more prominent. In terms of data handling, it may be feasible to treat the spectrum of protein and PEG reagent as background data to avoid obscuring the absorption of newly forming bonds. However this is only promising if the absorption of the bonds is significantly larger than the noise of the measurements. Noise in the measurements is also

one point that may have potential to be improved upon. All experiments were conducted in an air conditioned laboratory and the measuring cell itself was temperature controlled with a water bath thermostat. Adding temperature control to the batch reaction vessel may yield better results, as the spectral data is very sensitive to temperature changes. There are reports in literature where a very similar reaction was monitored with off-line FTIR, but the information given about the model building process is very limited and no rigorous cross validation was performed to ensure there was no over-fitting. [133] While there is no certainty that any of these improvements will lead to success, there is plenty of literature available that indicate the general feasibility of *in situ* reaction monitoring. [134, 135]

7

General Discussion and Conclusion

In this thesis, an experimental setup for in-line spectroscopic measurements utilizing FTIR spectroscopy for in-line measurements in chromatography and automated batch reaction experiments was established. In order to facilitate the spectroscopic measurements a custom software and hardware combination setup was developed. The spectrometer was integrated into the flow path of the respective experiments to enable near real-time acquisition of spectral data. The automation also enables the alignment of spectral data and reference analytics with very little effort, provided that the flow rates and delay volumes are well defined. Chapters 3 and 4 utilize the developed setup for chromatographic experiments and showcase its versatility in multiple case studies. Two of these case studies investigate the separation of PEGylated protein species, showing the potential to use in-line FTIR as PAT. The setup for automated sampling from a batch reaction is used in Chapter 5 to investigate the kinetics of the PEGylation reaction. Here it could be shown that it was possible to collect samples from a batch reaction for kinetic modeling with minimal manual labor. Finally, Chapter 6 investigates the possibility to monitor the batch reaction *in situ* by means of FTIR spectroscopy.

FTIR spectroscopy was combined with PLS regression and linear regression for the case studies in Chapter 3. It was shown that PLS regression models can be calibrated with the spectral data and used to estimate the concentration of co-eluting protein species in the column effluent. This was also enabled in part by the applied background correction method.

Spectral data from a blank run was subtracted from the experimental data to account for the influence of the gradient elution on the FTIR spectra. The separation of lysozyme and a mAb indicated that different protein species can be distinguished by such a PLS regression model. Both retention times and concentrations were estimated accurately. In a similar fashion, the separation of PEGylated protein species produced by batch PEGylation was investigated. While the aim of the first case study was shown to be achievable with UV/Vis spectroscopy before, PEGylated proteins are all isoforms of the same protein with a PEG molecule attached to them which does not absorb in UV/Vis. Again the retention times and concentrations were shown to be estimated accurately by the calibrated regression model.

As proteins and PEG both absorb in FTIR and have very distinct and only slightly overlapping bands, it is evident that the absorption at the respective wavelengths would have a unique proportion depending on the degree of PEGylation (DOP). The DOP is the amount of PEG molecules attached to a single protein molecule. This was utilized in Chapter 4. An integrated conjugation and purification process was supervised using the in-line FTIR setup with a more advanced background correction method based on EMSC. Employing only the absorption coefficients of protein and PEG, the DOP was estimated in the column effluent. It could be shown that the estimated DOP is close to the one determined by reference analytics with the overall benefit that obtaining extinction coefficients of the reactants is significantly less effort than performing multiple experiments to calibrate a regression model. Thus, this method relies on less prior knowledge than a calibrated statistical regression model.

The batch conjugation process to produce PEGylated proteins was also investigated. Automated sampling based on a modified chromatography device was utilized to investigate the reaction kinetics. It could be shown that a kinetic reaction model could be calibrated based on the acquired reference analytics while requiring minimal manual intervention. It could be shown that the sampling is consistent over a long period of time and that inconsistencies can be excluded based on the pressure measurement of the pump. The calibrated model had small confidence intervals and was in good agreement with the reference data. This project was enabled by in-house rapid prototyping and manufacturing utilizing 3D-printers.

Based on the same experimental setup and data, the *in situ* supervision of the conjugation reaction by FTIR spectroscopy was investigated. To this end, the spectral data that was acquired in the same experiments that were used for the kinetic reaction modeling was used to investigate its usage to estimate the state of the reaction. It was found that the experimental setup with the chosen data preparation and model building techniques

is not capable of estimating the reaction state. Different data processing techniques were compared and several possible routes of improvement were proposed.

While not directly correlated to the original aim of this work, during the first case studies the quantification of a process related impurity in the flow-through of a chromatography experiment was also investigated. It could be shown that a basic linear regression model can be used to quantify the substance while FTIR comes with the advantage that the spectral data could be compared to reference spectra, confirming the identity of the detected species simultaneously.

8

Outlook

Many FTIR spectrometers like the one used in this work are not very suitable for use on a process floor as they are susceptible to vibrations that may occur in production. To mitigate this, the use of newer technologies such as Quantum Cascade Lasers as emitters for FTIR instruments should be further investigated. [136] Alternative techniques like Near-Infrared (NIR) or Raman spectroscopy should also be investigated in regards to their ability to achieve the same results with more robust instruments more suited for production use. In addition to the more robust devices both NIR and Raman are also usable with fiber optics which also provides a direct benefit with the ability to have the spectrometer located physically distant from the point of measurement.

As already touched upon earlier, there is extensive material available to elucidate the possible chemical structure of a molecule based on its absorption spectrum. Conversely, if structural elements of an analyte are known, potential areas of absorption can be deducted. When it comes to PEGylated proteins in FTIR it may be possible to extract more information from the spectral data after the process is completed by combining this knowledge with additional knowledge about the PEGylation chemistry and chromatographic separations. As the number of attached PEG molecules must always be an integer and chromatographic separations usually lead to peaks that can be described as exponentially modified gauss shapes, a lot of information could be estimated even without precise knowledge of the absorption coefficients of the individual species. The same strategy could then also be applied to the separation of proteins that are conjugated to other molecules such as an ADC.

As a stretch goal, the work already published on specific spectral features correlated to structural features of protein molecules such as α-helices and β-sheets could be expanded upon to calculate entire IR spectra *in silico*. This would enable the generation of PAT models to distinguish different proteins in in-line spectral data. This would also go hand-in-hand with the general trend of moving away from wet lab experiments and towards digitalization.

It also still remains challenging to move a calibrated PAT method from the laboratory to production scale. [20] That means that for process scale applications, the model has to be either calibrated based on the actual process hardware, or model calibration has to be made independent from the hardware the process is run on. This can either mean utilizing models that inherently do not depend on the specific hardware or to utilize the same hardware for development and full scale production. This would also facilitate scale up efforts as the hardware would be seamlessly usable to develop the scaled up process.

While there has been increasing efforts towards PAT research in recent years, there are still many challenges to overcome. Not only is transferring a calibrated model still difficult, rigorous model validation techniques as well as data fusion still have yet to see widespread use. Data fusion refers to combining the data from multiple sources into one set of data to make it more consistent and accurate. Especially data fusion in combination with soft sensor approaches, combining data from multiple sensors for model building, shows great potential for improved PAT models. [137] In recent years, many new frameworks and software packages for working with advanced machine learning tools such as an Artifical Neural Network (ANN) have emerged. In combination with rigorous validation, these techniques offer multiple advantages over the current standard techniques.

References

[1] A. Philippidis, „Top 15 Best-Selling Drugs of 2018: Sales for most treatments grow year-over-year despite concerns over rising prices“, *Genetic Engineering & Biotechnology News*, vol. 39, no. 4, pp. 16–17, 2019 (cit. on pp. v, ix).

[2] J. M. Carton and W. R. Strohl, „Protein therapeutics (introduction to biopharmaceuticals)“, in *Introduction to Biological and Small Molecule Drug Research and Development*, Elsevier, 2013, pp. 127–159 (cit. on p. 1).

[3] D. M. Ecker, S. D. Jones, and H. L. Levine, „The therapeutic monoclonal antibody market“, in *MAbs*, Taylor & Francis, vol. 7, 2015, pp. 9–14 (cit. on p. 1).

[4] N. K. Tripathi and A. Shrivastava, „Recent Developments in Bioprocessing of Recombinant Proteins: Expression Hosts and Process Development“, *Frontiers in Bioengineering and Biotechnology*, vol. 7, p. 420, 2019 (cit. on p. 1).

[5] G. Walsh, „Biopharmaceutical benchmarks 2018“, *Nat. Biotechnol*, vol. 36, pp. 1136–1145, 2018 (cit. on p. 1).

[6] A. L. Grilo and A. Mantalaris, „The increasingly human and profitable monoclonal antibody market“, *Trends in biotechnology*, vol. 37, no. 1, pp. 9–16, 2019 (cit. on p. 1).

[7] S. Plotkin, J. M. Robinson, G. Cunningham, R. Iqbal, and S. Larsen, „The complexity and cost of vaccine manufacturing–an overview“, *Vaccine*, vol. 35, no. 33, pp. 4064–4071, 2017 (cit. on p. 1).

[8] E. Moorkens, N. Meuwissen, I. Huys, P. Declerck, A. G. Vulto, and S. Simoens, „The market of biopharmaceutical medicines: a snapshot of a diverse industrial landscape“, *Frontiers in pharmacology*, vol. 8, p. 314, 2017 (cit. on p. 1).

[9] E. Silverman, „Kymriah: A Sign of More Difficult Decisions To Come.“, *Managed care (Langhorne, Pa.)*, vol. 27, no. 5, p. 17, 2018 (cit. on p. 1).

[10] D. J. A. Crommelin, R. D. Sindelar, and B. Meibohm, *Pharmaceutical biotechnology: fundamentals and applications*. Springer, 2013 (cit. on p. 1).

[11] S. L. Lee, T. F. O'Connor, X. Yang, C. N. Cruz, S. Chatterjee, R. D. Madurawe, C. M. Moore, X. Y. Lawrence, and J. Woodcock, „Modernizing pharmaceutical manufacturing: from batch to continuous production“, *Journal of Pharmaceutical Innovation*, vol. 10, no. 3, pp. 191–199, 2015 (cit. on p. 2).

[12] J. Rantanen and J. Khinast, „The future of pharmaceutical manufacturing sciences“, *Journal of pharmaceutical sciences*, vol. 104, no. 11, pp. 3612–3638, 2015 (cit. on p. 2).

[13] F. Steinebach, T. Müller-Späth, and M. Morbidelli, „Continuous counter-current chromatography for capture and polishing steps in biopharmaceutical production“, *Biotechnology journal*, vol. 11, no. 9, pp. 1126–1141, 2016 (cit. on p. 2).

[14] G. Guiochon and L. A. Beaver, „Separation science is the key to successful biopharmaceuticals“, *Journal of Chromatography a*, vol. 1218, no. 49, pp. 8836–8858, 2011 (cit. on p. 2).

[15] 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 (cit. on pp. 3, 4).

[16] A. C. Fisher, M.-H. Kamga, C. Agarabi, K. Brorson, S. L. Lee, and S. Yoon, „The current scientific and regulatory landscape in advancing integrated continuous biopharmaceutical manufacturing“, *Trends in biotechnology*, vol. 37, no. 3, pp. 253–267, 2019 (cit. on p. 3).

[17] P. L. Turecek, M. J. Bossard, F. Schoetens, and I. A. Ivens, „PEGylation of biopharmaceuticals: a review of chemistry and nonclinical safety information of approved drugs“, *Journal of pharmaceutical sciences*, vol. 105, no. 2, pp. 460–475, 2016 (cit. on pp. 3, 12, 44, 58).

[18] S. N. Alconcel, A. S. Baas, and H. D. Maynard, „FDA-approved poly (ethylene glycol)–protein conjugate drugs“, *Polymer Chemistry*, vol. 2, no. 7, pp. 1442–1448, 2011 (cit. on pp. 4, 12).

[19] C. J. Fee and J. M. Van Alstine, „PEG-proteins: Reaction engineering and separation issues“, *Chemical engineering science*, vol. 61, no. 3, pp. 924–939, 2006 (cit. on pp. 4, 12, 68).

[20] C. E. Miller, „The "How" of Multivariate Analysis (MVA) in the Pharmaceutical Industry: A Holistic Approach", in *Multivariate Analysis in the Pharmaceutical Industry*, Elsevier, 2018, pp. 93–124 (cit. on pp. 4, 90).

[21] Food and Drug Administration of America, „Guidance for industry. PAT – A framework for innovative pharmaceutical development, manufacturing, and quality assurance", 2004 (cit. on pp. 4, 26, 45).

[22] B. Lin, B. Recke, J. K. Knudsen, and S. B. Jørgensen, „A systematic approach for soft sensor development", *Computers & chemical engineering*, vol. 31, no. 5-6, pp. 419–425, 2007 (cit. on p. 4).

[23] C.-F. Mandenius and R. Gustavsson, „Mini-review: soft sensors as means for PAT in the manufacture of bio-therapeutics", *Journal of Chemical Technology & Biotechnology*, vol. 90, no. 2, pp. 215–227, 2015 (cit. on p. 4).

[24] W. Kessler, *Multivariate Datenanalyse: Für die Pharma, Bio- und Prozessanalytik*. Wiley-VCH, 2006 (cit. on pp. 4, 5, 82).

[25] H. Wold, „Causal flows with latent variables: partings of the ways in the light of NIPALS modelling", *European economic review*, vol. 5, no. 1, pp. 67–86, 1974 (cit. on p. 4).

[26] H. Wold, „Path models with latent variables: The NIPALS approach", in *Quantitative sociology*, Elsevier, 1975, pp. 307–357 (cit. on p. 4).

[27] D. Swinehart, „The beer-lambert law", *Journal of chemical education*, vol. 39, no. 7, p. 333, 1962 (cit. on pp. 5, 48).

[28] V. Thomsen, D. Schatzlein, and D. Mercuro, „Limits of detection in spectroscopy", *Spectroscopy*, vol. 18, no. 12, pp. 112–114, 2003 (cit. on p. 5).

[29] J. C. Zillies, K. Zwiorek, G. Winter, and C. Coester, „Method for quantifying the PEGylation of gelatin nanoparticle drug carrier systems using asymmetrical flow field-flow fractionation and refractive index detection", *Analytical chemistry*, vol. 79, no. 12, pp. 4574–4580, 2007 (cit. on p. 6).

[30] E. Pretsch, P. Bühlmann, and M. Badertscher, *Spektroskopische Daten zur Strukturaufklärung organischer Verbindungen*. Springer-Verlag, 2013 (cit. on pp. 6, 83).

[31] A. A. Michelson and E. W. Morley, „LVIII. On the relative motion of the earth and the luminiferous Æther“, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 24, no. 151, pp. 449–463, 1887 (cit. on p. 6).

[32] B. C. Smith, *Fundamentals of Fourier transform infrared spectroscopy*. CRC press, 2011 (cit. on p. 7).

[33] P. R. Griffiths, H. J. Sloane, and R. W. Hannah, „Interferometers vs monochromators: separating the optical and digital advantages“, *Applied spectroscopy*, vol. 31, no. 6, pp. 485–495, 1977 (cit. on p. 7).

[34] P. Atkins and J. D. Paula, „Atkins' physical chemistry“, in *Chemistry*, 2009, pp. 783–827 (cit. on p. 7).

[35] J. Kauppinen and J. Partanen, *Fourier Transforms in Spectroscopy*. 2001, p. 271 (cit. on p. 7).

[36] K.-H. Mantel, „Fourier-Transform Infrarotspektroskopie an Proteinen“, 2005 (cit. on p. 8).

[37] J. Cooley and J. Tukey, „R66-50 An Algorithm for the Machine Calculation of Complex Fourier Series“, *IEEE Transactions on Electronic Computers*, vol. EC-15, no. 4, pp. 297–301, 1966 (cit. on p. 8).

[38] F. Goos and H. Hänchen, „Über das Eindringen des totalreflektierten Lichtes in das dünnere Medium“, *Annalen der Physik*, vol. 435, no. 5, pp. 383–392, 1943 (cit. on p. 9).

[39] F. Goos and H. Hänchen, „Ein neuer und fundamentaler Versuch zur Totalreflexion“, *Annalen der Physik*, vol. 436, no. 7-8, pp. 333–346, 1947 (cit. on p. 9).

[40] W. Urbaniak-Domagala, *Advanced Aspects of Spectroscopy*, 2012 (cit. on p. 10).

[41] F. Lottspeich and J. Engels, *Bioanalytik*. Heidelberg: Springer Spektrum, 2012 (cit. on pp. 10, 11).

[42] M. Baldassarre, C. Li, N. Eremina, E. Goormaghtigh, and A. Barth, „Simultaneous fitting of absorption spectra and their second derivatives for an improved analysis of protein infrared spectra“, *Molecules*, vol. 20, no. 7, pp. 12 599–12 622, 2015 (cit. on p. 11).

[43] E. Goormaghtigh, V. Cabiaux, and J.-M. Ruysschaert, „Determination of soluble and membrane protein structure by Fourier transform infrared spectroscopy“, in *Physicochemical methods in the study of biomembranes*, Springer, 1994, pp. 405–450 (cit. on p. 11).

[44] J. Güldenhaupt, *ATR-FTIR-spektroskopische Untersuchungen von membrangebundenem Ras*. 2011 (cit. on p. 11).

[45] A. Thomas, B. A. Teicher, and R. Hassan, „Antibody–drug conjugates for cancer therapy“, *The Lancet Oncology*, vol. 17, no. 6, e254–e262, 2016 (cit. on p. 12).

[46] M. J. Bossard and M. J. Vicent, „PEGylated proteins: A rational design for mitigating clearance mechanisms and altering biodistribution“, in *Polymer-Protein Conjugates*, Elsevier, 2020, pp. 23–40 (cit. on p. 12).

[47] J. Morgenstern, G. Wang, P. Baumann, and J. Hubbuch, „Model-Based Investigation on the Mass Transfer and Adsorption Mechanisms of Mono-Pegylated Lysozyme in Ion-Exchange Chromatography“, *Biotechnol J*, vol. 12, no. 9, 2017 (cit. on pp. 12, 63).

[48] A. Sanden, S. Suhm, M. Rüdt, and J. Hubbuch, „Fourier-transform infrared spectroscopy as a process analytical technology for near real time in-line estimation of the degree of PEGylation in chromatography“, *Journal of Chromatography A*, vol. 1608, p. 460 410, 2019 (cit. on pp. 13, 68, 69, 82, 83).

[49] N. Ulmer, D. Pfister, and M. Morbidelli, „Reactive separation processes for the production of PEGylated proteins“, *Current Opinion in Colloid & Interface Science*, vol. 31, pp. 86–91, 2017 (cit. on pp. 13, 68).

[50] R. A. Rader and E. S. Langer, „Biopharmaceutical Manufacturing: Historical and Future Trends in Titers, Yields, and Efficiency in Commercial-Scale Bioprocessing.“, *BioProcessing Journal*, vol. 13, no. 4, 2014 (cit. on p. 13).

[51] D. Pfister and M. Morbidelli, „Integrated process for high conversion and high yield protein PEGylation“, *Biotechnology and bioengineering*, vol. 113, no. 8, pp. 1711–1718, 2016 (cit. on p. 13).

[52] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical recipes 3rd edition: The art of scientific computing*. Cambridge university press, 2007 (cit. on p. 16).

[53] 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 (cit. on p. 21).

[54] 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, no. 7, pp. 1406–1416, 2015 (cit. on pp. 21, 27, 45, 82).

[55] 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“, *J Chromatogr A*, vol. 1462, pp. 153–64, 2016 (cit. on pp. 22, 62, 72).

[56] G. Carta and A. Jungbauer, *Protein chromatography: Process development and scale-up*. John Wiley & Sons, 2010 (cit. on p. 26).

[57] H. Schmidt-Traub, Ed., *Preparative Chromatography of Fine Chemicals and Pharmaceutical Agents*. Wiley, 2005 (cit. on p. 26).

[58] M. N. Gupta, Ed., *Methods for affinity-based separations of enzymes and proteins*, 1st ed., ser. Methods and Tools in Biosciences and Medicine. Birkhäuser, 2002 (cit. on p. 26).

[59] A. S. Rathore and G. Kapoor, „Application of process analytical technology for downstream purification of biotherapeutics“, *J Chem Technol Biotechnol*, vol. 90, no. 2, pp. 228–236, 02/2015 (cit. on p. 26).

[60] 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 (cit. on pp. 26, 27, 45).

[61] R. L. Fahrner and G. S. Blank, „Real-time control of antibody loading during protein A affinity chromatography using an on-line assay“, *J Chromatogr A*, vol. 849, no. 1, pp. 191–196, 1999 (cit. on p. 26).

[62] 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 (cit. on p. 26).

[63] 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 (cit. on p. 26).

[64] 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“, *J Chromatogr A*, vol. 827, no. 1, pp. 37–43, 1998 (cit. on p. 27).

[65] 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“, *Biotechnol Bioeng*, vol. 100, no. 2, pp. 306–316, 2008 (cit. on p. 27).

[66] 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 (cit. on pp. 27, 82).

[67] 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 (cit. on pp. 27, 45).

[68] S. K. Hansen, B. Jamali, and J. Hubbuch, „Selective high throughput protein quantification based on UV absorption spectra“, *Biotechnol Bioeng*, vol. 110, no. 2, pp. 448–460, 2013 (cit. on pp. 27, 82).

[69] W. Jiskoot and D. Crommelin, *Methods for structural analysis of protein pharmaceuticals*. Springer Science & Business Media, 2005, vol. 3 (cit. on pp. 27, 36).

[70] 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 (cit. on p. 27).

[71] 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 (cit. on pp. 27, 32, 38, 39, 47).

[72] 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 (cit. on pp. 27, 28, 40).

[73] 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, no. 9, pp. 3051–3061, 09/2012 (cit. on p. 27).

[74] 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 2 O Solution“, *Applied spectroscopy*, vol. 45, no. 5, pp. 868–873, 1991 (cit. on p. 27).

[75] 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 (cit. on p. 27).

[76] 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 (cit. on p. 27).

[77] E. Goormaghtigh, V. Cabiaux, and J.-M. Ruysschaert, „Determination of soluble and membrane protein structure by Fourier transform infrared spectroscopy: I. Assignment of model compounds“, in *Physicochemical methods in the study of biomembranes*, vol. 23, Springer, 1994, pp. 328–357 (cit. on pp. 27, 36).

[78] 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 (cit. on pp. 27, 36).

[79] F. Capito, R. Skudas, H. Kolmar, and B. Stanislawski, „Host cell protein quantification by fourier transform mid infrared spectroscopy (FT-MIR)“, *Biotechnol Bioeng*, vol. 110, no. 1, pp. 252–259, 2013 (cit. on p. 27).

[80] F. Capito, R. Skudas, H. Kolmar, and C. Hunzinger, „At-line mid infrared spectroscopy for monitoring downstream processing unit operations“, *Process Biochem*, vol. 50, no. 6, pp. 997–1005, 2015 (cit. on p. 27).

[81] 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 (cit. on pp. 28, 31, 40).

[82] 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 (cit. on pp. 30, 31).

[83] 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 (cit. on p. 31).

[84] 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 (cit. on p. 32).

[85] A. Savitzky and M. J. E. Golay, „Smoothing and differentiation of data by simplified least squares procedures“, *Anal Chem*, vol. 36, no. 8, pp. 1627–1639, 1964 (cit. on pp. 32, 36).

[86] 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 (cit. on pp. 32, 33, 36).

[87] 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 (cit. on p. 33).

[88] G. Quintás, B. Lendl, S. Garrigues, and M. de la Guardia, „Univariate method for background correction in liquid chromatography-Fourier transform infrared spectrometry“, *J Chrom A*, vol. 1190, no. 1-2, pp. 102–109, 2008 (cit. on p. 34).

[89] E. Pretsch, P. Bühlmann, C. Affolter, E. Pretsch, P. Bhuhlmann, and C. Affolter, *Structure determination of organic compounds*. Springer, 2009, vol. 13 (cit. on p. 38).

[90] 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 (cit. on p. 38).

[91] 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 (cit. on p. 38).

[92] 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 (cit. on p. 39).

[93] J. S. Kang, P. P. DeLuca, and K. Lee Choon, „Emerging PEGylated drugs“, *Expert Opinion on Emerging Drugs*, vol. 14, no. 2, pp. 363–380, 2009, PMID: 19453284 (cit. on p. 44).

[94] M. Hamidi, A. Azadi, and P. Rafiei, „Pharmacokinetic consequences of pegylation“, *Drug delivery*, vol. 13, no. 6, pp. 399–409, 2006 (cit. on p. 44).

[95] S. Awwad, C. Ginn, and S. Brocchini, „The case for protein PEGylation“, in *Engineering of Biomaterials for Drug Delivery Systems*, Elsevier, 2018, pp. 27–49 (cit. on pp. 44, 68).

[96] B. G. de la Torre and F. Albericio, „The pharmaceutical industry in 2018. An analysis of FDA drug approvals from the perspective of molecules“, *Molecules*, 2019 (cit. on p. 44).

[97] E. J. Park, J. Choi, K. C. Lee, and D. H. Na, „Emerging PEGylated non-biologic drugs“, *Expert Opinion on Emerging Drugs*, vol. 0, no. 0, pp. 1–13, 2019, PMID: 30957581 (cit. on p. 44).

[98] D. Pfister and M. Morbidelli, „Process for protein PEGylation“, *Journal of Controlled Release*, vol. 180, pp. 134–149, 2014 (cit. on pp. 44, 68).

[99] D. Pfister, E. Bourgeaux, and M. Morbidelli, „Kinetic modeling of protein PEGylation“, *Chemical Engineering Science*, vol. 137, pp. 816–827, 2015 (cit. on pp. 44, 68).

[100] A. Moosmann, J. Blath, R. Lindner, E. Müller, and H. Böttinger, „Aldehyde PEGylation kinetics: a standard protein versus a pharmaceutically relevant single chain variable fragment“, *Bioconjugate chemistry*, vol. 22, no. 8, pp. 1545–1558, 2011 (cit. on pp. 44, 62, 63, 68, 83).

[101] B. Maiser, K. Baumgartner, F. Dismer, and J. Hubbuch, „Effect of lysozyme solid-phase PEGylation on reaction kinetics and isoform distribution“, *Journal of Chromatography B*, vol. 1002, pp. 313–318, 2015 (cit. on pp. 45, 58, 68).

[102] D. Pfister, O. Ingold, and M. Morbidelli, „Model-based development of an on-column PEGylation process“, *Reaction Chemistry & Engineering*, vol. 1, no. 2, pp. 204–217, 2016 (cit. on p. 45).

[103] 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, 2018 (cit. on pp. 45, 47, 62, 69, 72).

[104] O. Buzzi, S. Yuan, and B. Routley, „Development and validation of a new near-infrared sensor to measure polyethylene glycol (PEG) concentration in water“, *Sensors*, vol. 17, no. 6, p. 1354, 2017 (cit. on p. 45).

[105] H. Martens, J. P. Nielsen, and S. B. Engelsen, „Light scattering and light absorbance separated by extended multiplicative signal correction. Application to near-infrared transmission analysis of powder mixtures“, *Analytical Chemistry*, vol. 75, no. 3, pp. 394–404, 2003 (cit. on p. 47).

[106] H. F. Boelens, R. J. Dijkstra, P. H. Eilers, F. Fitzpatrick, and J. A. Westerhuis, „New background correction method for liquid chromatography with diode array detection, infrared spectroscopic detection and Raman spectroscopic detection“, *Journal of chromatography A*, vol. 1057, no. 1-2, pp. 21–30, 2004 (cit. on p. 47).

[107] A. Schwaighofer, M. Montemurro, S. Freitag, C. Kristament, M. J. Culzoni, and B. Lendl, „Beyond Fourier Transform Infrared Spectroscopy: External cavity quantum cascade laser-based mid-infrared transmission spectroscopy of proteins in the Amide I and Amide II Region“, *Analytical chemistry*, vol. 90, no. 11, pp. 7072–7079, 2018 (cit. on p. 48).

[108] I. M. Deygen and E. V. Kudryashova, „New versatile approach for analysis of PEG content in conjugates and complexes with biomacromolecules based on FTIR spectroscopy.“, *Colloids and Surfaces B: Biointerfaces*, vol. 141, pp. 36–43, 2016 (cit. on p. 48).

[109] A. Savitzky and M. J. Golay, „Smoothing and differentiation of data by simplified least squares procedures.“, *Analytical chemistry*, vol. 36, no. 8, pp. 1627–1639, 1964 (cit. on p. 48).

[110] S. G. Kazarian and K. A. Chan, „ATR-FTIR spectroscopic imaging: recent advances and applications to biological systems“, *Analyst*, vol. 138, no. 7, pp. 1940–1951, 2013 (cit. on p. 49).

[111] R. Chung, D. Yu, V. T. Thai, A. F. Jones, G. K. Veits, J. Read de Alaniz, and J. E. Hein, „Tandem reaction progress analysis as a means for dissecting catalytic reactions: application to the Aza-Piancatelli rearrangement“, *ACS Catalysis*, vol. 5, no. 8, pp. 4579–4585, 2015 (cit. on p. 58).

[112] J. S. Kwak, W. Zhang, D. Tsoy, H. N. Hunter, D. Mallik, and M. G. Organ, „A Multiconfiguration Valve for Uninterrupted Sampling from Heterogeneous Slurries: An Application to Flow Chemistry“, *Organic Process Research & Development*, vol. 21, no. 7, pp. 1051–1058, 2017 (cit. on p. 58).

[113] K. Somerville, M. Tilley, G. Li, D. Mallik, and M. G. Organ, „A flow reactor with inline analytics: design and implementation“, *Organic Process Research & Development*, vol. 18, no. 11, pp. 1315–1320, 2014 (cit. on p. 58).

[114] L. Sandoval Rangel, J. Rivera de la Rosa, C. J. Lucio Ortiz, and M. J. Castaldi, „Pyrolysis of urea and guanidinium salts to be used as ammonia precursors for selective catalytic reduction of NOx“, *Journal of Analytical and Applied Pyrolysis*, vol. 113, pp. 564–574, 2015 (cit. on p. 58).

[115] G. L. Dimas-Rivera, J. R. de la Rosa, C. J. Lucio-Ortiz, J. A. De Los Reyes Heredia, V. G. Gonzalez, and T. Hernandez, „Desorption of Furfural from Bimetallic Pt-Fe Oxides/Alumina Catalysts“, *Materials (Basel)*, vol. 7, no. 1, pp. 527–541, 2014 (cit. on p. 58).

[116] M. G. Trevisan, C. M. Garcia, U. Schuchardt, and R. J. Poppi, „Evolving factor analysis-based method for correcting monitoring delay in different batch runs for use with PLS: On-line monitoring of a transesterification reaction by ATR-FTIR“, *Talanta*, vol. 74, no. 4, pp. 971–6, 2008 (cit. on p. 58).

[117] D. E. Fitzpatrick, C. Battilocchio, and S. V. Ley, „A Novel Internet-Based Reaction Monitoring, Control and Autonomous Self-Optimization Platform for Chemical Synthesis“, *Organic Process Research & Development*, vol. 20, no. 2, pp. 386–394, 2015 (cit. on p. 58).

[118] M. Rüdt, P. Vormittag, N. Hillebrandt, and J. Hubbuch, „Process monitoring of virus-like particle reassembly by diafiltration with UV/Vis spectroscopy and light scattering“, *Biotechnology and bioengineering*, vol. 116, no. 6, pp. 1366–1379, 2019 (cit. on p. 58).

[119] Q. Jing and X. LÜ, „Kinetics of Non-catalyzed Decomposition of Glucose in High-temperature Liquid Water“, *Chinese Journal of Chemical Engineering*, vol. 16, no. 6, pp. 890–894, 2008 (cit. on p. 58).

[120] S. P. Bawane and S. B. Sawant, „Reaction kinetics of the liquid-phase hydrogenation of benzonitrile to benzylamine using Raney nickel catalyst“, *Chemical Engineering Journal*, vol. 103, no. 1-3, pp. 13–19, 2004 (cit. on p. 58).

[121] E. Cambon, R. Piamtongkam, F. Bordes, S. Duquesne, S. Laguerre, J.-M. Nicaud, and A. Marty, „A new Yarrowia lipolytica expression system: An efficient tool for rapid and reliable kinetic analysis of improved enzymes“, *Enzyme and Microbial Technology*, vol. 47, no. 3, pp. 91–96, 2010 (cit. on p. 58).

[122] D. de Carvalho Carneiro, L. da Cunha, and T. Suzana, „Conjugative Post-Translational Modifications for Pharmacological Improvement of Therapeutic Proteins“, *Current Proteomics*, vol. 15, no. 1, pp. 3–12, 2018 (cit. on p. 58).

[123] K. Grabow and U. Bentrup, „Homogeneous catalytic processes monitored by combined in Situ ATR-IR, UV–Vis, and Raman spectroscopy“, *ACS Catalysis*, vol. 4, no. 7, pp. 2153–2164, 2014 (cit. on pp. 58, 69).

[124] S. Šašic, T. Amari, H. W. Siesler, and Y. Ozaki, „Polycondensation Reaction of Bis (hydroxyethylterephthalate)—Self Modeling Curve Resolution Analysis of On-Line ATR/FT-IR Spectra“, *Applied Spectroscopy*, vol. 55, no. 9, pp. 1181–1191, 2001 (cit. on pp. 58, 69).

[125] S. Andris, M. Rudt, J. Rogalla, M. Wendeler, and J. Hubbuch, „Monitoring of antibody-drug conjugation reactions with UV/Vis spectroscopy“, *J Biotechnol*, vol. 288, pp. 15–22, 2018 (cit. on pp. 59, 69).

[126] Gringer, „Cyclonic Water Filter, https://www.thingiverse.com/thing:1657498“, (cit. on p. 59).

[127] J. Morgenstern, P. Baumann, C. Brunner, and J. Hubbuch, „Effect of PEG molecular weight and PEGylation degree on the physical stability of PEGylated lysozyme“, *Int J Pharm*, vol. 519, no. 1-2, pp. 408–417, 2017 (cit. on p. 62).

[128] K. E. Ottow, T. Lund-Olesen, T. L. Maury, M. F. Hansen, and T. J. Hobley, „A magnetic adsorbent-based process for semi-continuous PEGylation of proteins“, *Biotechnology journal*, vol. 6, no. 4, pp. 396–409, 2011 (cit. on pp. 62, 70).

[129] A. Sanden, S. Haas, and J. Hubbuch, „Modifying an ÄKTApurifier System for the Automated Acquisition of Samples for Kinetic Modeling of Batch Reactions“, *SLAS TECHNOLOGY: Translating Life Sciences Innovation*, p. 2 472 630 319 891 976, 2019 (cit. on pp. 68, 69).

[130] J. H. Santos, C. M. Mendonça, A. R. Silva, R. P. Oliveira, A. Pessoa, J. A. Coutinho, S. P. Ventura, and C. O. Rangel-Yagui, „An integrated process combining the reaction and purification of PEGylated proteins“, *Green Chemistry*, vol. 21, no. 23, pp. 6407–6418, 2019 (cit. on p. 68).

[131] U. Wolf, R. Leiberich, and J. Seeba, „Application of infrared ATR spectroscopy to in situ reaction monitoring“, *Catalysis today*, vol. 49, no. 4, pp. 411–418, 1999 (cit. on p. 69).

[132] R. E. Canfield, „The amino acid sequence of egg white lysozyme“, *Journal of Biological Chemistry*, vol. 238, no. 8, pp. 2698–2707, 1963 (cit. on p. 83).

[133] V. Hebbi, D. Kumar, and A. S. Rathore, „Process Analytical Technology implementation for peptide manufacturing: Cleavage reaction of recombinant Lethal Toxin Neutralizing Factor (rLTNF) concatemer as a case study“, *Analytical Chemistry*, 2020 (cit. on p. 84).

[134] F. S. Cardoso, G. E. Mickle, M. A. da Silva, P. T. Baraldi, and F. B. Ferreira, „Application of In Situ FTIR for the Preparation of 17-α-Estradiol via Mitsunobu Reaction“, *Organic Process Research & Development*, vol. 20, no. 2, pp. 306–311, 2016 (cit. on p. 84).

[135] R. Chung and J. E. Hein, „The More, The Better: Simultaneous In Situ Reaction Monitoring Provides Rapid Mechanistic and Kinetic Insight“, *Topics in Catalysis*, vol. 60, no. 8, pp. 594–608, 2017 (cit. on p. 84).

[136] D. T. Childs, R. A. Hogg, D. G. Revin, I. U. Rehman, J. W. Cockburn, and S. J. Matcher, „Sensitivity advantage of QCL tunable-laser mid-infrared spectroscopy over FTIR spectroscopy“, *Applied Spectroscopy Reviews*, vol. 50, no. 10, pp. 822–839, 2015 (cit. on p. 89).

[137] L. Rolinger, M. Rüdt, and J. Hubbuch, „A critical review of recent trends, and a future perspective of optical spectroscopy as PAT in biopharmaceutical downstream processing“, *Analytical and Bioanalytical Chemistry*, pp. 1–18, 2020 (cit. on p. 90).

Abbreviations

ADC Antibody Drug Conjugate 12, 89

ALS Alternating Least-Squares 44

ANN Artifical Neural Network 90

API active pharmaceutical ingredient 3

ATR Attenuated Total Reflectance 8–10

CEX Cation-Exchange 71

CFF Cross-Flow Filtration v, x

CPP Critical Process Parameter x

CPPs Critical Process Parameters vi

CQAs Critical Quality Attributes vi, 3, 18

DAD Diode Array Detector 6

DOP degree of PEGylation 14, 15, 18, 22, 23, 45, 50, 51, 53, 86

DSP Downstream Processing v, ix, 1

EMA European Medicines Agency vi, x, 1

EMSC Extended Multiplicative Signal Correction vii, xii, 20, 23, 44, 47, 86

FDA Food and Drug Administration v, x, 1, 3, 17

FTIR Fourier-Transform Infrared vi–viii, x–xii, 6, 8, 11, 18, 20–22, 45, 50, 54, 67, 69, 71, 82–87, 89

HCCF Harvested Cell Culture Fluid v, ix, 1

HCP Host Cell Protein 68

LOD limit of detection 5

LOQ limit of quantitation 5

mAb monoclonal antibody vii, ix, xi, 1, 2, 20, 21, 86

mAbs monoclonal antibodies v

MVDA Multivariate Data Analysis vii, xi, 4, 20

NIR Near-Infrared 89

PAT Process Analytical Technology vi, x, xi, 3, 5, 17–20, 45, 50, 67, 85, 90

PCA Principal Component Analysis 4

PCC Periodic Counter-Current Chromatography 2

PEG Polyethylene Glycole vi, vii, x, xi, 3, 6, 12, 15, 18, 23, 43–45, 54, 70, 82, 83, 86, 89

PES Polyethersulfone 47, 70, 72

PLS Partial-Least Squares vii, xi, 4, 5, 21, 24, 67, 77, 82, 83, 85, 86

QbD Quality by Design vi, x, 3, 17

QTPP Quality Target Product Profile vi, x, 3

SMB Simulated Moving Bed Chromatography 2

SNR signal-to-noise ratio 7, 83

UV/Vis Ultraviolet/Visible 6, 18, 21, 22, 69, 86

VB Visual Basic 21