Development of process analytical technologies for chromatography based protein purification

zur Erlangung des akademischen Grades eines
DOKTORS DER INGENIEURWISSENSCHAFTEN (Dr.-Ing.)

der Fakultät für Chemieingenieurwesen und Verfahrenstechnik des
Karlsruher Instituts für Technologie (KIT)
vorgelegte

genehmigte
DISSERTATION

von
Dipl.-Ing. Nina Brestrich
aus Neunkirchen (Saar)

Referent: Prof. Dr. Jürgen Hubbuch
Korreferent: Prof. Dr. Matthias Franzreb
Tag der mündlichen Prüfung: 15.07.2016

"The important thing in science
is not so much to obtain new facts
as to discover new ways of thinking
about them."

William Bragg

Danksagung

Während der letzten Jahre haben mich viele Menschen begleitet und unterstützt, denen ich von ganzem Herzen danken möchte.

Prof. Dr. Jürgen Hubbuch danke ich für die Möglichkeit, meine Doktorarbeit über eine spannende und zuvor kaum untersuchte Thematik zu schreiben. Sein entgegengebrachtes Vertrauen und seine inspirierende Art, Ergebnisse in einem größeren Zusammenhang zu sehen, haben sehr zum Gelingen dieser Arbeit beigetragen.

Prof. Dr. Matthias Franzreb danke ich für die Übernahme des Korreferats sowie für seine interessante Vorlesung über mechanistische Chromatographiemodellierung, die während meines Studiums mein Interesse an der Thematik geweckt hat.

Dem Ministerium für Wissenschaft, Forschung und Kunst des Landes Baden-Württemberg danke ich für die finanzielle Unterstützung meiner Arbeit.

Weiterhin möchte ich mich bei meinen Projektpartnern Rentschler, Lek und CSL für die Unterstützung mit Proteinen und Daten bedanken. Ohne sie wäre es nicht möglich gewesen, eine Anwendbarkeit der in der Arbeit entwickelten Methodik für reale Beispiele aufzuzeigen.

Allen Mitarbeitern des Instituts möchte ich für die herzliche Aufnahme in die Gruppe und die gemeinsame Zeit danken. Der große Zusammenhalt der Gruppe, die gute Stimmung im Labor sowie die zahlreichen Diskussionen in Seminaren haben sehr zum Gelingen dieser Arbeit beigetragen.

Besonders bedanke ich mich bei Matthias Rüdt und Steffen Großhans für die sehr produktive und erfolgreiche Projektkooperation sowie die gemeinsame Betreuung studentischer Arbeiten. Tobias Hahn danke ich für die Unterstützung bei der mechanistischen Modellierung mit der von ihm entwickelten Software ChromX, ohne die ein Teilprojekt meiner Arbeit gar nicht möglich gewesen wäre.

Auch danke ich Matthias Rüdt, Tobias Hahn und Thiemo Huuk für die freundschaftliche Atmosphäre im Büro und die spannenden fachlichen und außerfachlichen Diskussionen.

Mein großer Dank gilt ebenso den Studenten Till Briskot, Sebastian Andris, Axel Kraft, Laura Rolinger, Daniel Büchler und Adrian Sanden, die mit großer Motivation und großem Zeiteinsatz sehr zum Gelingen einzelner Teilprojekte beigetragen haben.

Ganz besonders möchte ich mich bei meinen Eltern Martha und Lutz bedanken, die mich immer moralisch und finanziell unterstützt haben. Sie haben mir oft den Rücken

freigehalten, so dass ich mich meinen Interessen widmen konnte. Ohne ihre Unterstützung wäre diese Arbeit wahrscheinlich nie entstanden.

Zum Schluss gilt mein ganz besonderer Dank meinem Freund Kai, der mir in den letzten Jahren immer beigestanden hat. Er hat zahlreiche schlaflose Nächte, in denen ich über ungelöste Probleme nachgedacht habe, sowie etliche Wochenendarbeit gelassen akzeptiert und mich moralisch unterstützt. Ohne die gemeinsame Zeit wäre diese Arbeit für mich unendlich viel schwerer gewesen.

Karlsruhe, den 3.5.2016

Nina Brestrich

Zusammenfassung

Therapeutische Proteine, wie zum Beispiel monoklonale Antikörper, Insulin oder menschliches Wachstumshormon, stellen derzeit einen der am schnellsten wachsenden Bereiche des internationalen Pharmazeutikamarktes dar. Die Herstellung dieser Medikamente ist herausfordernd, da es sich um große, biologische Moleküle mit einer komplexen Struktur handelt. Die Entwicklungs- und Produktionskosten für therapeutische Proteine sind daher hoch, was auch zu den hohen Preisen für diese Medikamente beiträgt. In den letzten Jahren ist allerdings ein zunehmender Preisdruck auf die Prozessentwicklungs- und Produktionskosten entstanden. Die Gründe hierfür liegen an dem Aufkommen der ersten Generika, den sogenannten *Biosimilars*, sowie an zunehmenden Einsparungen im öffentlichen Gesundheitswesen. Dies hat zur Erforschung neuer, kostengünstigerer Technologien geführt. Zusätzlich stellen internationale Gesundheitsbehörden verstärkte Anforderungen an das Prozess- und Produktverständnis sowie an die Reproduzierbarkeit von Prozessen und Produktqualität durch den Einsatz von Technologien zur Echtzeitüberwachung und -kontrolle.

In diesem Zusammenhang wird die Entwicklung von prozessanalysierenden Technologien (PAT) sowie deren Integration in die entsprechenden Herstellungsprozesse sowohl von industrieller als auch wissenschaftlicher Seite verstärkt diskutiert. Besonders die Evaluierung von kontinuierlichen Verfahren, welche die Prozesseffizienz weiter steigern sollen, hat zu einer verstärkten Forschungsarbeit an PAT geführt. Der vermehrte Einsatz von PAT in der Herstellung therapeutischer Proteine könnte den Automatisierungsgrad erhöhen, die Reproduzierbarkeit steigern sowie Personalkosten einsparen. Die Integration von PAT im sogenannten *Upstream Processing*, der Synthese der Moleküle in Bakterien- oder Säugerzellen, ist bereits weit fortgeschritten. Im sogenannten *Downstream Processing*, also der Aufreinigung der therapeutischen Proteine, gibt es dagegen derzeit kaum Technologien zur Echtzeitüberwachung und -kontrolle. Das wichtigste Verfahren im *Downstream Processing* ist Chromatographie, da hiermit eine hohe Auflösung zwischen dem Produkt und Kontaminanten erreicht werden kann. Somit kann die erforderliche hohe Produktreinheit bei gleichzeitig hohen Ausbeuten erzielt werden. Aus diesem Grund war die Entwicklung von geeigneten PAT für chromatographische Verfahren zur Proteinaufreinigung das Ziel dieser Doktorarbeit.

Die größte Herausforderung, die mit der Entwicklung der entsprechenden Technologien angegangen wurde, ist Prozessvariabilität. Verglichen mit chemischen Verfahren weisen biotechnologische Prozesse eine erhöhtes Maß an Variabilität auf. Das führt oft zu variablen Produkt- und Kontaminantenkonzentrationen während des gesamten Prozesses. Zusätzlich wird der Freiheitsgrad durch andere unkontrollierbare Einflüsse, wie zum Beispiel das Altern von Säulen, das zu Kapazitäts- oder Effizienzeinbußen führt, erhöht. Da

es kaum geeignete Technologien zur Echtzeitquantifizierung von Proteinen gibt, stellt die Prozessvariabilität eine große Herausforderung für die Kontrolle der Beladungs- und Elutionsphase in chromatographischen Verfahren dar. Während der Elutionsphase kann diese Variabilität zu abweichenden Peakkonzentrationen und zu einer variablen Auflösung zwischen dem Produkt und Kontaminanten führen. Ein für die Prozesskontrolle dieser Phase interessanter Parameter ist die Fraktionierung des Produkts. Der derzeitige industrielle Standard zur Steuerung der Fraktionierung sind Prozessentscheidungen basierend auf univariaten Absorptionsmessungen bei 280 nm. Da jedoch sowohl das Produkt als auch Kontaminanten bei dieser Wellenlänge absorbieren und die Elutionsprofile variabel sein können, ist diese Methodik nur bei vollständig aufgetrennten Peaks reproduzierbar. Dies trifft jedoch nur in den seltensten Fällen auf Verfahren im Produktionsmaßstab zu.

Ein für die Prozesskontrolle der Beladungsphase interessanter Parameter ist der Zeitpunkt, an dem die Säule mit dem Produkt saturiert ist und sich das Produkt im Effluenten ansammelt. Zu diesem Zeitpunkt sollte die Phase beendet werden. Dies ist besonders kompliziert für sogenannte *Capture Steps*, die den Anfang einer chromatographischen Aufreinigungskette darstellen. Die zahlreichen Kontaminanten überdecken hierbei die Produktabsorption bei 280 nm. Da die Produktkonzentration jedoch variabel sein kann, wird das Beladungsvolumen in der Regel basierend auf einer vor dem Verfahrensschritt durchgeführten Laboranalytik zur Produktquantifizierung in der Zufuhr festgelegt. Um einen Durchbruch des wertvollen Produkts im Fall von Säulenalterung zu verhindern, werden meist konservative Werte für die Säulenkapazität angenommen.

Obwohl einige analytische Verfahren zur Echtzeitüberwachung und -kontrolle für chromatographische Prozesse bereits untersucht wurden, stellen die vorgeschlagenen Lösungen entweder ein zusätzliches Kontaminationsrisiko dar oder weisen zu lange Analysezeiten auf. Der Hauptfokus dieser Doktorarbeit lag folglich darauf, das Problem der Prozessvariabilität durch Entwicklung geeigneter analytischer Technologien zu lösen. Diese Technologien sollten nicht nur in der Lage sein wichtige Qualitäts- und Leistungsattribute aufzuzeichnen, sondern den Prozess auch so zu manipulieren, dass diese innerhalb eines vorher definierten Bereiches bleiben. Die entsprechenden Technologien sollten vorzugsweise in Echtzeit operieren und auch für industriell relevante Trennprobleme anwendbar sein. Ein weiteres Ziel dieser Doktorarbeit bestand deshalb darin, auch die Ursachen von Variabilität zu identifizieren, die zu abweichenden Chromatogrammen führen.

Die vorliegende Arbeit setzt sich aus fünf Publikationen/Manuskripten zusammen, die sich mit unterschiedlichen Aspekten der Entwicklung von PAT für die chromatographische Proteinaufreinigung beschäftigen. In der ersten Veröffentlichung wurde die Anwendbarkeit von UV-Spektroskopie zur Echtzeitüberwachung und -kontrolle chromatographischer Verfahren untersucht. Unterschiedliche Proteinspezies weisen eindeutige Variationen in ihren UV-Absorptionsspektren auf, da sie eine unterschiedliche Anzahl und ein unterschiedliches Verhältnis der aromatischen Aminosäurenreste enthalten. Dies ermöglicht die Generierung multivariater Regressionsmodelle (hier: Partial Least Squares regression – PLS-Modelle), welche die Absorptionsspektren mit Einzelproteinkonzentrationen korrelieren. Basierend auf dieser Methodik, wurde eine Technologie zur Prozessanalyse und -kontrolle entwickelt, die selektiv Proteine im Prozessstrom quantifizieren kann. Sie setzt sich aus einem Diodenarray-Detektor (DAD) zur Aufzeichnung von UV-Absorptionsspektren, einem PLS-Modell, sowie Schnittstellen zwischen verschiedenen Software-

paketen, die eine Datenauswertung und Prozesskontrolle in Echtzeit ermöglichen, zusammen. Nach erfolgreicher Kalibrierung und Validierung des entsprechenden PLS-Modells, wurde die entwickelte Technologie erfolgreich zur Echtzeitanalyse von Proteinkonzentrationen während einer chromatographischen Auftrennung verwendet. Die auf diese Weise erzielte Dekonvolution des Elutionspeaks in die Einzelsignale der ko-eluierenden Proteine wurde dazu verwendet, die Elutionsphase in Echtzeit zu kontrollieren. Hierbei wurde eine Fraktionierung des Produkts basierend auf zuvor festgelegten Reinheitskriterien erreicht.

Die zweite Veröffentlichung und das dritte Manuskript beschreiben die Weiterentwicklung der Technologie, um deren Anwendbarkeit für industriell relevante Trennprobleme zu ermöglichen. Ein Kriterium in diesem Zusammenhang war eine Methodik zur Kalibrierung der benötigten PLS-Modelle, die nur das Ausgangsmaterial der chromatographischen Auftrennung und nicht die Reinkomponenten benötigt. Aus diesem Grund wurde in der zweiten Veröffentlichung eine auf Prozessdaten basierende Methodik zur Kalibrierung der entsprechenden PLS-Modelle entwickelt. Hierfür wurden UV-Absorptionsspektren während der gesamten chromatographischen Trennung aufgezeichnet. Gleichzeitig wurde der Säuleneffluent in viele Fraktionen aufgetrennt, die anschließend mit einer Standardreferenzanalytik ausgewertet wurden, um die spezifischen Proteinkonzentrationen zu bestimmen. Um die Absorptionsspektren schließlich mit den Ergebnissen der Fraktionsanalytik in einem PLS-Modell zu korrelieren, wurden diese entsprechend der Fraktionsgröße zeitlich gemittelt. Durch Anwendung dieser auf Prozessdaten basierenden Kalibrierungsmethode konnte die entwickelte Technologie zur Prozessanalyse erfolgreich zur selektiven Quantifizierung von ko-eluierenden monoklonalen Antikörpermonomeren, -aggregaten und -fragmenten verwendet werdet. Zusätzlich konnte auch eine Anwendbarkeit zur selektiven Quantifizierung von Serumproteinen in einem chromatographischen Aufreinigungsschritt des Cohn-Prozesses gezeigt werden.

Eine weitere Anforderung für die industrielle Anwendbarkeit war die Verwendbarkeit der Technologie unter Bedingungen, bei denen die volle Säulenkapazität ausgeschöpft wird. Dies führt in der Regel dazu, dass ein hoch konzentrierter Produktpeak vorliegt, der Kontaminanten in den Flanken enthält, die in deutlich niedrigeren Konzentrationen vorliegen. Folglich liegt ein enormer Konzentrationsunterschied zwischen den einzelnen Proteinen vor. Das dritte Manuskript dieser Doktorarbeit beschäftigt sich daher damit, wie mit diesen hohen Konzentrationsunterschieden umgegangen werden kann. Hierfür wurde ein Instrument in den experimentellen Aufbau integriert, das nicht wie der DAD bei einer festen Pfadlänge Spektren akquiriert, sondern Messungen mit variabler Pfadlänge durchführt. Hierbei wird der lineare Bereich der Abhängigkeit der Absorption von der Pfadlänge ermittelt und die Steigung der Funktion in diesem Bereich berechnet. Anstelle von Absorption wurde diese Steigung für jede Wellenlänge erfasst und mit den entsprechenden Proteinkonzentrationen in einem PLS-Modell korreliert. Die so generierten Modelle ermöglichten eine Dekonvolution von Peaks, die sich aus hoch konzentrierten Produkten mit ko-eluierenden, niedrig konzentrierten Kontaminanten zusammensetzten.

In dem vierten Manuskript wurde die Anwendbarkeit der entwickelten Technologie für die Kontrolle der Beladungsphase in einem *Capture Step* erforscht. Es konnte gezeigt werden, dass PLS-Modelle, die auf UV-Absorptionsspektren basieren, auch zur selektiven Quantifizierung von monoklonalen Antikörpern im Hintergrund von vielen proteinogenen und nichtproteinogenen Verunreinigungen verwendet werden können. Die entwickel-

te Technologie zur Prozessanalyse ermöglichte eine Detektion des Produktdurchbruchs während der Beladungsphase eines Protein A *Catpure Steps*. Sobald ein zuvor definierter Schwellenwert an Produktdurchbruch erreicht wurde, beendete die entwickelte Technologie zur Prozessanalyse und -kontrolle automatisch die Beladungsphase.

Schließlich beschäftigt sich die fünfte Veröffentlichung mit der Ermittlung von Ursachen für Variabilität. Es konnte gezeigt werden, dass inverse mechanistische Modellierung dazu verwendet werden kann, die Ursachen von abweichenden Chromatogrammen zu identifizieren. Bei der inversen mechanistischen Modellierung wurden Parameter, die Abweichungen in Chromatogrammen hervorgerufen haben könnten, solange systematisch variiert, bis Simulation und experimentell ermitteltes Chromatogramm übereinstimmten. Da dieser Ansatz die Konzentrationen aller beteiligten Spezies erforderte, wurde zuvor ein Dekonvolution der Elutionspeaks basierend auf PLS-Modellierung mit UV-Absorptionsspektren durchgeführt. Hierdurch konnten Abweichungen zwischen Soll- und Istwert für die Flussrate sowie für die Natriumionenkonzentration in Binde- und Elutionspuffer ermittelt werden.

Folglich konnte diese Doktorarbeit nicht nur dazu beitragen das Problem von Prozessvariabilität in der chromatographischen Aufreinigung von Proteinen zu lösen, sondern zeigt auch Möglichkeiten auf, wie Ursachen für Variabilität ermittelt werden können. Die entwickelte Technologie könnte die Echtzeitüberwachung und -kontrolle der Beladungs- und Elutionsphase in künftigen Chromatographieprozessen ermöglichen. Dies könnte zu einer allgemeinen Prozessintensivierung und zur Verwirklichung kontinuierlicher Herstellungsverfahren sowie der Echtzeitfreigabe von Chargen beitragen. Der vorgeschlagene Ansatz zur Fehlerursachendiagnostik, der auf mechanistischer Modellierung basiert, könnte weiterhin ein verbessertes Prozessverständnis liefern und die Anzahl an verworfenen Lots minimieren. Dies wäre auch im Einklang mit den verstärkten Anforderungen der Gesundheitsbehörden an Prozessverständnis und Reproduzierbarkeit durch Kontrolle von Prozessvariabilität.

Summary

Therapeutic proteins, such as monoclonal antibodies, insulin, or human growth hormone, are currently one of the fastest-growing sector in the international pharmaceutical market. The manufacturing of these drugs is very challenging due to their size, complex structure, and biological origin. As a consequence, process development and manufacturing costs for therapeutic proteins are large, which contributes to the high prices for these drugs. However in the last years, there has been a downward pressure upon process development and manufacturing costs. Reasons for this prize pressure comprise a larger competition due to the emerge of biosimilars, the biological equivalents of chemical generics, and due to public healthcare budget cuts. This resulted in the evaluation of new, more cost-efficient technologies. In addition, health authorities request a larger degree of process and product understanding as well as an application of real-time process monitoring and control to ensure process consistency and product quality.

In this context, the development of Process Analytical Technologies (PAT) as well as their integration into the corresponding manufacturing processes is an increasingly discussed topic, both in industry and academia. Especially the evaluation of continuous processing to increase process efficiency has been a main driver for research into PAT. The use of PAT in protein manufacturing could increase the degree of automation, guarantee reproducibility, and decrease personnel costs. While the integration of PAT is already well-advanced in upstream processing, there is still a lack of suitable technologies for real-time process monitoring and control in downstream processing. The workhorse in the downstream processing of therapeutic proteins is chromatography. It achieves a high resolution between product and contaminants, resulting in the required product purity combined with high yields. The development of suitable PAT for chromatography based protein purification was hence the main objective of this thesis.

The major challenge which was addressed with this development is process variability. Compared to their chemical equivalents, biological processes exhibit a large degree of variability. This often results in variable concentrations of the product and contaminants throughout the whole process. In addition, the degree of freedom is increased by other uncontrollable influences such as column capacity or efficiency loss due to aging. Combined with a lack of real-time analytics for selective protein quantification, this variability represents a challenge for the control of the load and the elution phase in chromatography. During the elution phase, variability can lead to deviating peak concentrations and variable resolution between product and contaminants. An interesting attribute for the control of this phase is the pooling of the product fraction, which is also referred to as peak-cutting. The current industrial standard is a peak-cutting based on univariate UV absorption measurements at 280 nm. As both the contaminants and the product con-

tribute to UV absorption at 280 nm and as the elution profiles are subjected to variability, this control strategy is only reproducible for well separated peaks. This is however seldom the case at large scale production. An interesting attribute to be controlled during the column loading is the termination of the load phase as soon as the resin is saturated with the product and the product starts to accumulate in the column effluent. This is especially complicated in capture steps at the beginning of the purification chain, where many contaminants obscure the product UV absorption at 280 nm. Because the product titer is subjected to variability, the load volume onto the column is hence determined based on offline product quantification in the feed prior to column loading. To prevent product breakthrough in the case of column aging, conservative values for the column capacity are commonly applied.

Although some investigations into analytical tools for real-time process monitoring and control in chromatography have been performed, the proposed solutions pose an additional risk of contamination or have too long analysis times. The major focus of this thesis was hence to address the problem of process variability by developing analytical technologies for chromatography. These technologies should not only be able to monitor quality and performance attributes, but also manipulate the process in a way such that these attributes are kept in a certain predefined range. The corresponding technologies should preferably operate in real-time and should be applicable for real-life separation issues. Another goal of this thesis was the investigation of the causes for variability that lead to the corresponding deviations in chromatograms.

This thesis consists of five publications/manuscripts, that focus on different aspects of the development of PAT for chromatography based protein purification. In the first publication, the possibility of applying UV spectroscopy for the monitoring and control of chromatography was investigated. Different protein species exhibit distinct variations in their UV absorption spectra due to a different number and ratio of the aromatic amino acid residues. This enables the generation of multivariate regression models (here: Partial Least Squares regression – PLS models) correlating the absorption spectra with selective protein concentrations. Based on this technique, a PAT tool for selective inline quantification of proteins was developed. It consists of a diode array detector (DAD) to acquire the absorption spectra, a PLS model, and customized interfaces between different types of software to enable real-time data evaluation and process control. After a calibration and validation of the corresponding PLS model, the PAT tool was successfully applied to monitor selective protein concentrations in a chromatographic separation in real-time. The so obtained deconvolution of the elution peak into the profiles of all contributing proteins was then applied for the control of the elution phase. The peaks were cut in real-time using product-purity criterions.

In the second publication and the third manuscript, the PAT tool was further developed to enable its applicability for real-life separation issues. One prerequisite in this context was the calibration of the corresponding PLS models using feedstocks or in-process materials, which is dealt with in the second publication. This was addressed by developing a process-data-based calibration method for PLS models. For this kind of calibration, the UV absorption spectra of a chromatographic separation were acquired. At the same time, fractions of the column effluent were collected and analyzed by standard reference analytics to obtain the selective protein concentrations. To correlate the

spectra with the results of the fraction analysis using the PLS technique, they were averaged in time according to the fraction size. Applying this process-data-based calibration method, the PAT tool was successfully used to monitor the selective concentrations of co-eluting monoclonal antibody monomer, fragments, and aggregates. Additionally, an applicability for the monitoring of co-eluting serum proteins in a chromatographic step of the Cohn process could be demonstrated.

Another requirement for the real-life applicability of the PAT tool was its usability for conditions were the columns are completely loaded. The result of such a preparative chromatographic separation is commonly a highly concentrated product peak with contaminants in the flanks exhibiting much smaller concentrations. Hence, there is a large concentration difference of the involved species which has to be dealt with. The third manuscript of this thesis describes the solution for this issue by applying variable pathlength (VP) spectroscopy. Instead of measuring the UV absorption spectra at one fixed pathlength using a DAD, a novel device for VP measurements was integrated into the PAT tool. The device acquires the slope in the linear range of the dependency of the absorption from the pathlength. Instead of applying the absorption, the slopes at every wavelength were acquired and correlated with the protein concentrations using PLS technique. These models allowed a peak deconvolution for separations, where the elution peaks consisted of high and low concentrated species.

The fourth manuscript explored the possibility of applying the PAT tool for the control of a load phase in a capture step. It was demonstrated, that PLS modelling with UV absorption spectra is also applicable for the quantification of monoclonal antibody in the background of many protein and non protein-based impurities. The PAT tool allowed the detection of the product breakthrough during the load phase of a Protein A capture step. As soon as a previously defined breakthrough level of the antibody was reached, the PAT tool automatically terminated the load phase.

Eventually, the fifth publication addressed the investigation of causes for variability. It was demonstrated that inverse mechanistic chromatography modelling can be applied to identify the actual causes for deviations in chromatograms. In inverse mechanistic modelling, parameters that could have caused deviations were varied systematically until the simulated chromatogram matched the recorded one. As this approach required selective concentrations of all involved protein species, a peak deconvolution based on PLS modelling with UV absorption spectra was applied prior to mechanistic modelling. By this way, deviations between the set points and the actual values of the flow rate and the sodium-ion concentration in loading and elution buffer were successfully identified.

Consequently, this thesis not only contributes to solve the issues of process variability in protein chromatography, but also shows a possibility how causes for variability could be investigated. The developed technologies might enable real-time monitoring and control of the load and elution phase in future chromatography steps. This might contribute to an overall process intensification and to the realization of continuous processing as well as real-time release. The proposed approach for root cause investigation based on inverse mechanistic modelling might further provide a better process understanding and minimize lot rejections. This is in compliance with the health authorities' requirements of increased process understanding and improved reproducibility by control of variability.

Contents

1 Introduction

Therapeutic proteins, such as monoclonal antibodies, insulin, or human growth hormone, are currently one of the fastest-growing sector in the international pharmaceuticals market (Knäblein, 2008; Knezevic and Griffiths, 2011; Otto et al., 2014). Since 2012, Humira, a monoclonal antibody, has been the world's best selling pharmaceutical product with a annual sales of 12.5 billions US dollars in 2014 (Genetic Engineering and Biotechnology News, 2013; Statistica.com, 2015). However, therapeutic proteins are very challenging to manufacture due to their size, complex structure, and biological origin. Figure 1.1 illustrates the size and complexity of therapeutic proteins in contrast to small, chemical pharmaceuticals by comparing the molecules with means of transport. While Aspirin consists of 21 atoms (Wheatley, 1964), a therapeutic protein can contain between 777 (insulin) (Rosenberg, 1974) and 20-25,000 atoms (monoclonal antibody) (Binns, 2010; Otto et al., 2014). In addition, protein manufacturing is limited to unit operations with mild conditions in order to maintain the biological function of the target molecule (Crommelin et al., 2003). The final protein drug product to be administrated to the patient is a complex mixture due to the biological origin. It consists of the target molecule, proteins related to the target molecule such as aggregates, charge variants, and glycovariants as well as proteins related to the production host (host cell proteins). The thresholds for

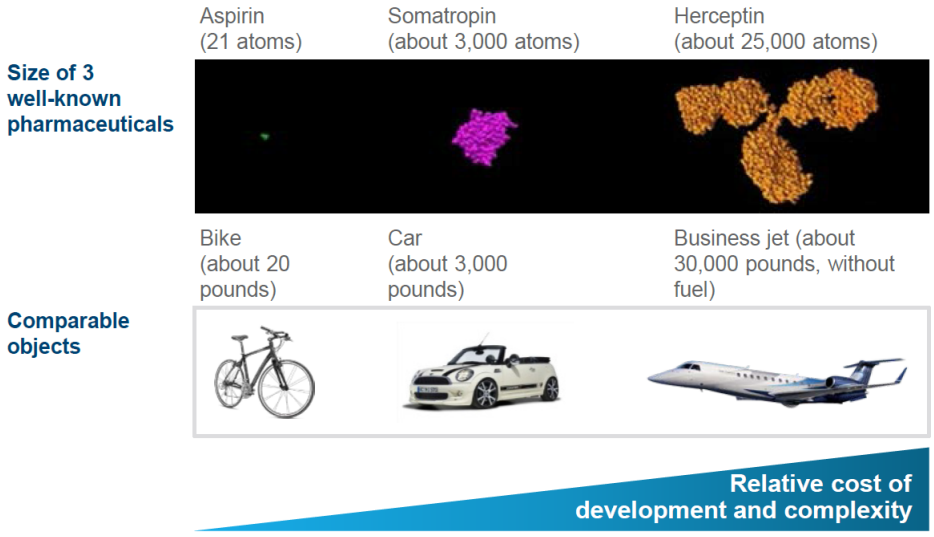


Figure 1.1: Comparison of size and complexity of therapeutic proteins and chemical pharmaceuticals: While Aspirin consists of 21 atoms, somatropin contains 3,000 and Herceptin, a monoclonal antibody, about 25,000 atoms. Complexity and size of the molecules is illustrated by comparing them with means of transport (Note: The molecules are on scale; the means of transport not) (Otto et al., 2014).

all impurities as well as the product variants profiles have to be well-defined to avoid harm to the patient and to guarantee efficacy. This is complicated by the fact, that these attributes are sensitive to changes in the manufacturing process (Schiestl et al., 2011). In order to meet the quality requirements, unit operations achieving high resolution between product and contaminants such as chromatography are required. Chromatography steps are however elaborate to develop and have high material costs (Przybycien et al., 2004; Rege et al., 2006). The degree of real-time analytics and process automation for chromatography and other unit operations in protein manufacturing is low as protein mixtures are challenging to characterize. Thus, protein manufacturing has a large requirement of well-trained employees. Other limitations include low yields and long processing times.

As a consequence, process development and manufacturing costs for therapeutic proteins are large, which contributes to the high prices for these drugs. But a part of the costs is also caused by a lack of generics (Harbour, 2007). The novelty of therapeutic proteins and the possibility to address previously unmet medical needs currently allow the biopharmaceutical companies to claim high prices for their products (Danzon and Furukawa, 2006; McCamish and Woollett, 2011; Otto et al., 2014). The daily treatment costs with a therapeutic protein are on average 20 times more than those of a traditional small molecule product (Emerton, 2013), despite of similar costs for process development and for bringing the drugs to the market (Trusheim et al., 2010). This results in a large burden for the public healthcare systems.

However in the last years, several patents of blockbuster therapeutic proteins have expired and more will follow in the near future. A blockbuster is a drug with annual sales of more than 1 billion US dollars (European Commission, 2008). When the patents of a group of blockbusters are expiring in the same time period, this usually results in an sudden drop of global sales which is referred to as a patent cliff (Fernández and Martínez-Hurtado, 2012). Figure 1.2 illustrates the 2012-2019 patent cliff for therapeutic proteins with annual sales between 1.7 to 12.5 billion US dollars in 2014. In the last three years, the patents of seven protein blockbusters have expired in Europe. At the same time, technical advances such as highly productive cell lines (Browne and Al-Rubeai, 2007), disposable bioreactors (Zheng, 2010), and high capacity chromatography resins (McGlaughlin, 2010) have enabled a more economic manufacturing of therapeutic proteins. The technical progress combined with the patent cliff, new regulations of health authorities for biological generics, and public healthcare budget cuts have lead to the emerge of so called biosimilars (Fernández and Martínez-Hurtado, 2012). This group of drugs is comparable with chemical generics.

In 2006, the first biosimilar product was approved in Europe (somatropin/human growth hormone) followed by several others such as erythropoietin (hormone for red blood cell production) in 2007, and filgrastim (hormone for stimulating the bone marrow to produce granulocytes and stem cells) in 2008 (Sekhon and Saluja, 2011). In 2013, the first biosimilar monoclonal antibody was approved (McKeage, 2014) and in 2014, a follow-on product for insulin glargin entered the market (DeVries et al., 2015). The rise of biosimilars has induced a larger competition and price pressure in the biopharmaceutical market (Tsuruta et al., 2015). As a consequence, there is a downward pressure upon manufacturing costs (Broly et al., 2010; Walsh, 2010). New, more cost-effective technologies are evaluated. Single-use technology might reduce sanitization times, lower

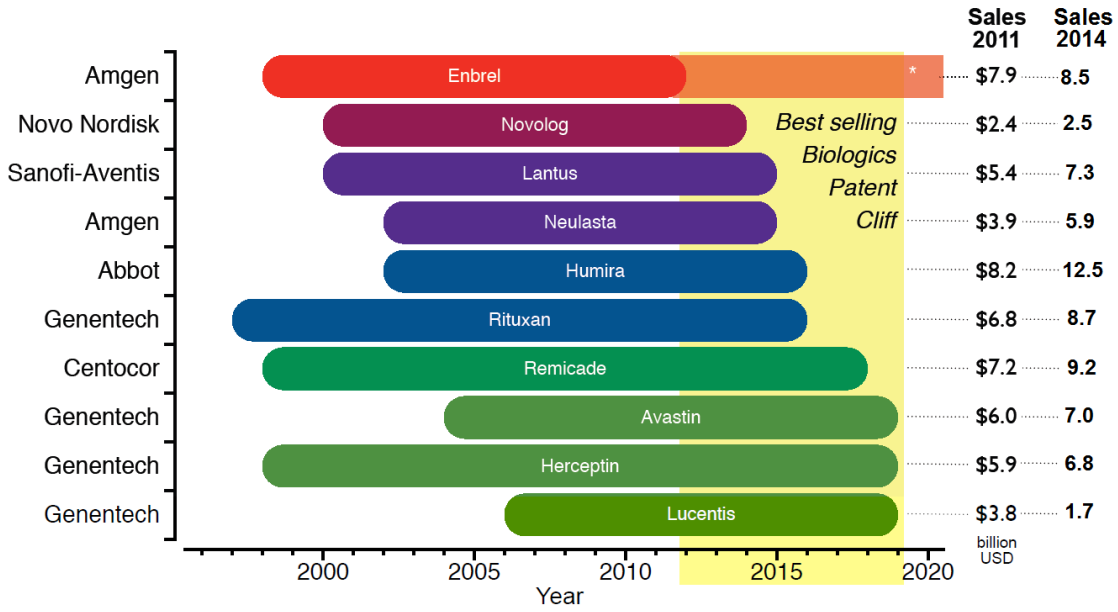


Figure 1.2: Patent expiration for the top ten selling biologics (all therapeutic proteins). The 2012-2019 patent cliff is highlighted in yellow. The bars represent the period of patent protection. Company names are shown on the left hand side, while global sales for 2011 are displayed on the right hand side (Fernández and Martínez-Hurtado, 2012). *The patent for Enbrel expired in 2015 in Europe, but protection in the US will last until 2028 (Statistica.com, 2015). Annual sales data for 2014 was added from (Statistica.com, 2015; Roche.com, 2014; Novonordisk.com, 2014).

footprints, and enable multi-product plants (Kuczewski et al., 2011). In downstream processing, alternatives to chromatography such as crystallization, precipitation, membrane chromatography, and aqueous two phase extraction were explored (Oelmeier et al., 2013; Przybycien et al., 2004; Rosa et al., 2007; Sim et al., 2012; Thömmes and Etzel, 2007). In upstream processing, continuous perfusion culture at high cell densities has been demonstrated (Warikoo et al., 2012) and triggered the evaluation of continuous downstream processing. Integrated continuous manufacturing of therapeutic proteins is currently discussed in numerous conferences and publications (Cramer and Holstein, 2011; Godawat et al., 2015; Jungbauer, 2013; Klutz et al., 2015; Xenopoulos, 2015). A shift from batch to continuous processing could increase productivity, lower the footprint, and decrease overall costs (Jungbauer, 2013; Warikoo et al., 2012).

Besides from development of new equipment and production techniques, the application of Process Analytical Technologies (PAT) has become of great interest (Challa and Potumarthi, 2013; Glassey et al., 2011; Mercier et al., 2014; Rathore and Kapoor, 2015; Read et al., 2010). Efficient analytical tools are an absolute requirement for a successful process development and manufacturing of therapeutic proteins. They are essential to determine the product quality as well as the process consistency and efficiency (Flatman et al., 2007). As the implementation of real-time monitoring and control strategies is a critical aspect for the success of continuous processing, the evaluation of continuous processing has given rise to an increasing research into PAT (Konstantinov and Cooney, 2015; Rathore et al., 2015). The use of PAT in protein manufacturing could increase the degree of automation, guarantee reproducibility, and decrease personnel costs. The

current state-of-the art of analytical tools and process control in protein manufacturing will be discussed in the following section.

1.1 Analytical Tools and Process Control in Protein Manufacturing

Therapeutic proteins are very complex products and challenging to characterize. The thresholds for all impurities as well as the product identity and activity are so called Critical Quality Attributes (CQAs) and have to be well-controlled to avoid harm to the patient. Even slight changes in the manufacturing process can influence the CQAs (Schiestl et al., 2011). Thus, health authorities require a careful investigation of the relationship between the process and the product. Parameters in the process influencing CQAs are called Critical Process Parameters (CPPs). By understanding the relationships between CPPs and CQAs, the process can be designed such that the desired product quality is achieved in the presence of process variability. This concept is referred to as Quality by Design (QbD). (ICH, 2008)

In order to further ensure a consistent product quality, the US Food and Drug Administration (FDA) published a PAT concept, which promotes the monitoring and control of both CQAs and CPPs. The PAT initiative aims to replace offline laboratory measurements by real-time or near real-time measurements. This helps to manage batch-to-batch variability and to develop a process, in which all variability is managed by the process itself. Near real-time process monitoring can be achieved by atline analysis, the performance of fast assays in close proximity to the process stream. A more sophisticated approach is the use of online monitoring, the automated diversion and measurement of samples from the production stream. Depending on the sampling and analysis time, data can be obtained in real-time or near real-time. A third option is the application of inline monitoring, the real-time acquisition and evaluation of data with sensors that are directly integrated into the production stream. (FDA, 2004)

Although the PAT initiative advocates the use of (near) real-time monitoring of CQAs and CPPs, the industrial application of this technique at large scale is currently limited to simple, univariate parameters such as pH, conductivity, or absorption at 280 nm. Central parameters and quality attributes like purity and yield or product identity and activity are usually determined offline. The major reason for this is the complexity of the product as well as the number of process and product related contaminants. Process related contaminants comprise Host Cell Proteins (HCPs), DNA, viruses, endotoxins, extractables and leachables from resins and filters, process buffers, and detergents. Further, product related impurities such as High Molecular Weight species (HMWs), Low Molecular Weight species (LMWs), and heterogeneities of the product due to chemical modifications need to be identified and quantified (Liu et al., 2010). Common analytics to detect these impurities and to prove product identity and activity are, among others, High Performance Liquid Chromatography (HPLC), Isoelectric Focusing, Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE), Capilary Electrophoresis, Mass Spectrometry (MS), quantitative Real Time Polymerase Chain Reaction (qPCR), Enzyme-linked Immunosorbent Assay (ELISA), and Peptide Mapping (Chirino and Mire-

Sluis, 2004; Flatman et al., 2007). Table 1.1 illustrates the variety of properties that have to be analyzed during therapeutic protein production (here: monoclonal antibodies) and the corresponding offline assays for this purpose (Flatman et al., 2007). This clearly shows the complexity to realize (near) real-time monitoring and control of protein manufacturing.

Major disadvantages of offline analysis comprise the time input and personnel deployment as well as the risk of a late reaction on process variability or production errors. The latter might even lead to the discarding of a lot. To overcome this analytical bottleneck and to enable the (near) real-time monitoring and control of CQAs and CPPs, several PAT tools have been developed in the past decades. One approach is the development of time-optimized assays for atline analysis. Especially in upstream processing, where the process is slow and the assay time is not limiting, atline analysis might be suitable. For instance, the quantification of monoclonal antibodies and several impurities has been realized by atline matrix-assisted laser desorption/ionization MS (Steinhoff et al., 2015). The quantification of monoclonal antibodies during a Chinese Hamster Ovary (CHO) cell culture process was achieved by atline mid-IR spectroscopy in combination with Multivariate Data Analysis (MVDA) (Capito et al., 2015b).

Atline spectroscopic methods in combination with MVDA have also been used for the monitoring of downstream processing. A flocculation process for removing contaminants such as cell debris, host cell proteins, and nucleic acid could be monitored using near-infrared spectroscopy (Yeung et al., 2000). During an ultrafiltration-based separation of a ternary protein mixture, atline monitoring of the protein concentrations in the retentate and permeate was realized by fluorescence spectroscopy (Elshereef et al., 2010). The same spectroscopic method (without application of MVDA) has also been applied to monitor the titers of co-eluting misfolded and correctly folded proteins in order to control a chromatographic process (Rathore et al., 2009). A recent advancement in atline spectroscopy is the use of atline mid-infrared spectroscopy for the quantification of HCPs, aggregates, and product in common unit operations for downstream processing of monoclonal antibodies (Capito et al., 2013b,c,a, 2015a). The technique enabled the quantification down to 700 ng HCPs/ml and down to 1% [w/w] aggregates.

Besides from spectroscopic methods, PAT tools based on atline analytical chromatography have been developed to perform process monitoring and control. For instance, atline HPLC has been used to analyze the product stream in a continuous Multicolumn Countercurrent Solvent Gradient Purification (MCSGP) process to gain a feed-back for a proportional-integral-derivative controller regulating the peak cutting of the upcoming cycle (Krättli et al., 2013). For the control of the load phase of a two column continuous Protein A chromatography process, which was connected to a CHO perfusion culture, atline analytical chromatography was applied (Karst et al., 2015). Disadvantages of all atline techniques include the personnel costs for technicians as well as the risk of contamination and missing or delayed data.

To minimize human impact, automated sampling techniques for online analysis have been developed. The first publications in this field can be found for upstream processing, most likely because in this context the minimization of assay times is not as crucial as for downstream processing due to slower process times. For instance, the antibody titer in cell culture has been monitored by automated analytical protein A chromatography

Table 1.1: Offline assays commonly applied for the production of monoclonal antibodies (Flatman et al., 2007).

Product characteristic	Analysis properties	Test method platform
Physical and chemical characteristics	Purity	Electrophoresis Reverse phase HPLC Size exclusion HPLC
	Integrity/molecular weight	Electrophoresis Mass spectrometry Size exclusion HPLC Light scatter
	Identity	Isoelectric focussing Peptide mapping Ion exchange HPLC
Potency/activity	Antigen binding	Immunoassay
	Biological methods	Cell proliferation Complement mediated cytotoxicity Reporter-gene assays
Product related impurities	Aggregation/fragments	Electrophoresis Size exclusion HPLC
Process related impurities	Host cell proteins	Immunoassay
	Host cell DNA	DNA hybridisation qPCR DNA binding-threshold Fluorescent-picogreen
	Leached Protein A ligands	Immunoassay
	Cell culture medium proteins	Immunoassay
	Viruses	qPCR Electron microscopy In vivo/in vitro assays
	Microorganisms	Bioburden Endotoxin test
	Others: leachates and extractables, cell culture medium components, purification reagents and chemicals	Various e.g. reversed phase HPLC, ion chromatography, gas chromatography-MS

(Chase, 1986; Ozturk et al., 1995; Paliwal et al., 1993). In downstream processing, online analytical chromatography has been applied for a selective quantification of the product and contaminants during chromatographic processes. The resulting concentration profiles of the contributing species were obtained in near real-time and allowed the control of the peak-cutting in chromatographic processes (Fahrner et al., 1998; Kaltenbrunner et al., 2012; Rathore et al., 2008a,b, 2010b). The same technique was also applied to monitor the concentration of a monoclonal antibody in the column effluent during the load phase of a protein A chromatography step. The load phase was automatically stopped by the PAT tool, when 1% product breakthrough was detected (Fahrner and Blank, 1999a). Automated sampling is relatively easy to develop and the equipment is commercially available. Disadvantages are however a risk of contamination and the high equipment prices. A further limitation of this technique for downstream processing is the rather long response time of the analysis. Depending on the elution flow rate of the preparative separation step in question, the time delay might require a slow-down of the process, resulting in a decreased productivity.

PAT tools that use integrated sensors to acquire and evaluate data in real-time overcome these limitations. The requirement for "real-time" in this context is to achieve a result with an analytical tool in a time frame that is substantially shorter than the occurring changes over time. This time frame is generally longer for upstream than for downstream processing and might explain the larger number of publications in this field for the former. In upstream processing, several PAT tools using integrated fluorescence, near-infrared, and mid-infrared spectroscopic sensors and MVDA for real-time process monitoring and control have been reported (Arnold et al., 2002; Boehl, 2003; Haack et al., 2004; Jung et al., 2002; Kara et al., 2010; Mazarevica et al., 2004; Navrátil et al., 2005; Tamburini et al., 2003).

Integrated spectroscopic sensors in downstream processing are predominately UV/Vis- or fluorescence-based. A possible reason might be that infrared spectroscopy commonly requires longer measurement times to achieve a good signal to noise ratio for proteins. Especially for chromatography this could be a limiting factor, as the concentration changes of the contributing species occur rather fast in this separation technique. An industrial standard in chromatography is the monitoring of the column effluent using the UV absorption at 280 nm. This method is often applied to control the peak-cutting, but is limited to separations with well separated product and contaminant peaks. As the target protein and most process- and product-related contaminants contribute to UV absorption at 280 nm, univariate UV absorption measurements do not allow a differentiation between co-eluting species. In order to still enable process control, UV absorption at 280 nm in combination with a mathematical model has been proposed (Mendhe et al., 2015). The authors have successfully established a model based on the retention time of a characteristic peak eluting prior to the product peak. This model allowed the control of the peak-cutting, but requiring a pre-peak as reporter, it is not applicable for all separation issues.

Another PAT tool based on UV absorption at 280 nm was designed for the control of the load phase in chromatographic separation steps (Bängtsson et al., 2012). It is especially suitable to determine dynamic binding capacities in capture steps. These steps are at the beginning of the protein purification chain and separate the target protein from

a major part of the contaminants. During column loading, the UV absorption at 280 nm of the contaminants obscures the product absorption. In order to still determine the product breakthrough and thus the dynamic binding capacity of columns, the proposed method is based on the calculation of a difference signal between two detectors situated at the column in- and outlet. During the load phase, the post column signal is supposed to stabilize and is referred to as impurity baseline. As soon as the product breaks through, there is an increase in the post-column UV signal above the impurity baseline which corresponds to the breakthrough level of the product. A disadvantage of the method is however that it needs two detectors. As both detectors have to be equally calibrated, there is a risk of unequal detector drifts. A further limitation might be displacement effects of contaminants that prevent a stabilized impurity baseline.

1.2 Theoretical Foundation

A novel approach for real-time monitoring and control of chromatography is proposed in this thesis. Instead of the UV absorption at 280 nm, multivariate UV absorption measurements in combination with MVDA are applied. A deeper insight into protein UV absorption is given in section 1.2.1, while fundamentals of MVDA are discussed in section 1.2.2. In addition to MVDA and spectroscopy, mechanistic chromatography modelling is applied to obtain a fundamental understanding of the physical phenomena during a chromatographic separation and to use this understanding as a PAT tool. The applied mechanistic model is described in more detail in section 1.2.3.

1.2.1 Intrinsic Protein UV Absorption

More than sixty years ago, it was proposed that changes in the UV absorption spectrum of a protein are correlated with its structure (Beaven and Holiday, 1951). Several chromophores contribute to the intrinsic protein UV absorption, as depicted in Figure 1.3 A. Due to their delocalized Pi-electrons, the peptide bond and the aromatic amino acid residues tryptophan, tyrosine, and phenylalanine mainly contribute to the intrinsic protein absorption. The peptide bond mainly absorbs in the region between 180-240 nm and can provide information about the secondary structure of a protein (Rosenheck and Doty, 1961). Although alpha helix, beta-sheet, and random coil peptides show distinct spectral differences in this range, the interpretation of protein absorption spectra in this area is rather complicated and is rarely used for secondary structure elucidation. The reason for that is the overlapping of the peptide bond absorption with the bands of the aromatic residues, with the bands of thiol (cysteine) and thioether (methionine) groups, and, below 200 nm, with the bands of common solution components such as inorganic ions and dissolved oxygen (Wetlaufer, 1962).

In the range of 250-300 nm, mainly the aromatic amino acid residues contribute to UV absorption. The strongest absorbing residue is tryptophan with an absorption maximum at 280 nm and a shoulder near 292 nm, which is responsible for the frequently observed shoulder in protein spectra. Tyrosine absorbs about 4 times less than tryptophan with a maximum at 276 nm. Compared to the two other residues, phenylalanine absorbs weakly in the range of 240-270 nm (about 30 times less than tryptophan) and is only visible

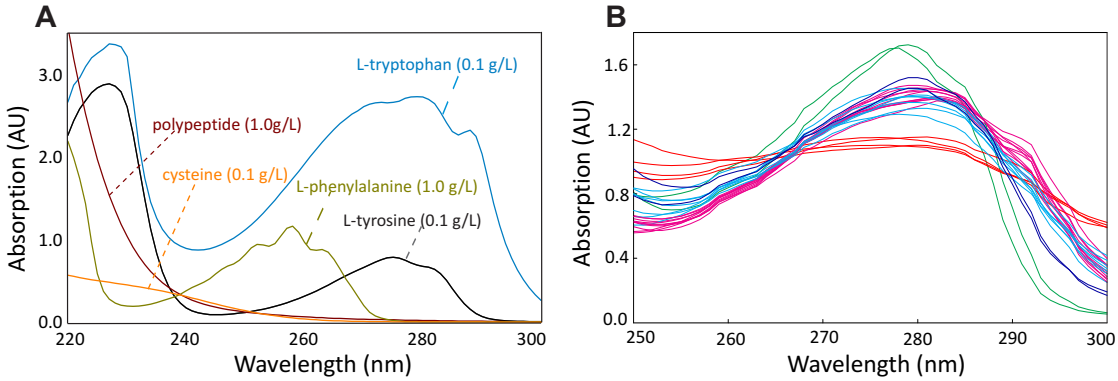


Figure 1.3: A: Spectra of the components mainly contributing to protein UV absorption spectra. B: UV absorption spectra of 24 proteins with different tryptophan/tyrosine ratio. Green: 0.0, Violet: 0-0.1, Blue: 0.26-0.47, Pink:0.48-4.00. The red-colored spectra are from proteins with a heme group. The spectra were recorded in protein solutions of 1.0 g/L and are normalized to equal total intensity for a better visual comparison. Figures and data taken from (Hansen et al., 2013).

as small unevennesses in some absorption spectra (Crommelin et al., 2003). Although cysteine does not absorb in the range of 250-300 nm, the disulfid bonds between cysteine residues can contribute weakly to the absorption spectra in this area.

As tryptophan and tyrosine are the main originators of the absorption between 250-300 nm, Hansen et al. recorded the absorption spectra of various proteins with different tryptophan to tyrosine ratio (Hansen et al., 2013). Their results are displayed in Figure 1.3 B and the spectra are colored according to the tryptophan to tyrosine ratio. Green: 0.0, Violet: 0-0.1, Blue: 0.26-0.47, Pink:0.48-4.00. The spectra were recorded in protein solutions of 1.0 g/L and are normalized to equal total intensity for a better visual comparison. With exemption of the red-colored spectra from proteins containing a heme group, the main difference in the spectra is related to the tryptophan to tyrosine ratio. As tryptophan is the only aromatic residue that absorbs beyond 295 nm, proteins without this residue (green spectra) are narrower. The higher the tryptophan to tyrosine ratio, the more UV absorption between 295-300 nm can be observed (cf. pink and blue spectra) (Hansen et al., 2013). In addition, the typical shoulder at 292 nm becomes visible for those proteins.

Besides from the tryptophan to tyrosine ratio in a protein and the presence of a heme group, the UV protein absorption is also influenced by the environment of the chromophores. Any change in the solvent that leads to conformational changes of a protein and/or changes in the local environment of the chromophores can effect the UV absorption spectrum (Crommelin et al., 2003). To visualize the environmental influence Hansen et al. compared the spectrum of L-tryptophan dissolved in water with the spectrum of a tryptophan rich protein in water (cf. Figure 1.4 A). Compared to the absorption spectrum of the pure amino acid, the protein spectrum is red shifted. This bathochromic shift is caused by the different polarity of the environment. The environmental polarity is higher for a dissolved pure tryptophan molecule compared to a folded protein where the tryptophan residues are predominately situated in the hydrophobic core. The same effect can also be found by comparing Figure1.3 A and B. While there are minima of all three amino acid residues between 230-240 nm, this minimum is red shifted for proteins (Hansen et al., 2016). To further demonstrate how even small conformational changes

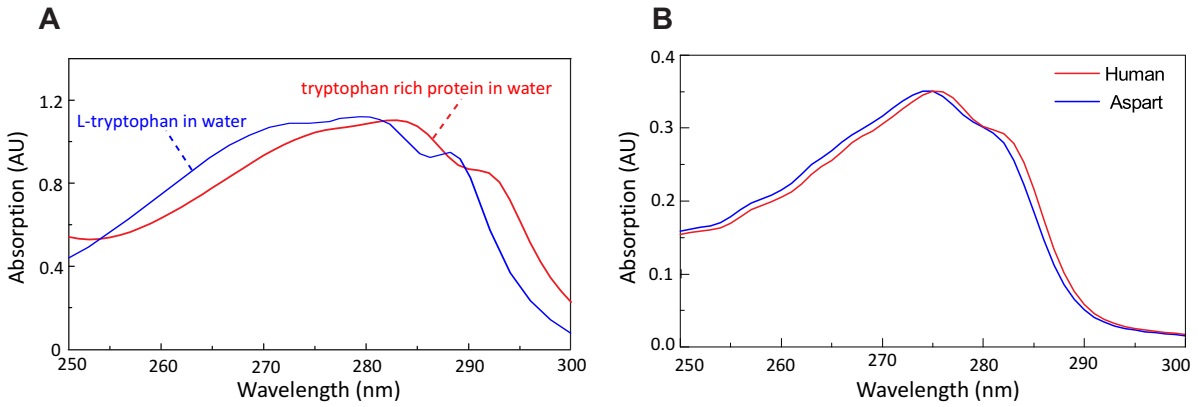


Figure 1.4: A: UV absorption spectra of L-tryptophan and a tryptophan rich protein. The spectra have been normalized to equal total intensity. B: UV absorption spectra of insulin detemir, insulin aspart and human insulin. The spectra have been normalized to equal total intensity. Figures taken from (Hansen et al., 2016).

can influence the resulting UV absorption spectrum, Figure 1.4 B illustrates the spectra of insulin aspart and human insulin. Insulin aspart differs from human insulin in only one amino acid, which leads to a small change in the tertiary structure. This structural change leads to a different environment for some chromophores resulting in a blue shift (Hansen et al., 2016).

As conformational changes leading to a different environment of the chromophores can influence a protein UV absorption spectrum, protein aggregation might alter the shape of the resulting aggregate sum spectrum as well. Especially during non-native aggregation, which is accompanied by conformational changes (Chi et al., 2003), a shift in the spectrum might be possible. Figure 1.5 illustrates the enrichment of aggregates in the rear peak flank during a cation exchange chromatography step as well as the change of the sum absorption spectrum during the run. The observed spectral change is correlated with an increase in the aggregate level. On the one hand, this might be caused by the conformational changes due to aggregation. On the other hand, an absorption beyond 320 nm indicates light scattering effects. Scattering of incident light on solute protein multimers can prevent the light from reaching the detector and results in an artifical increment of the absorption values (Crommelin et al., 2003). Especially for larger aggregates this effect might overrule the conformation related spectral changes.

Besides from the polarity of the solvent, its pH, its content of ions, and its temperature might also influence the protein UV absorption. The majority of pH related effects is associated with conformational changes due to the pH shift. However in addition, the hydroxyl group of the tyrosine residue can be ionized, resulting in red shift of the absorption spectrum for this amino acid residue (Balestrieri et al., 1978). As for the pH, the temperature can influence the protein conformation and thus the resulting UV absorption spectrum. In addition, the temperature influences the Boltzmann distribution among vibrational and rotational engergy levels of a solute molecule and temperature dependent interaction between molecules such as hydrogen bonding (Ito, 1960). While the different distribution of the vibrational and rotational levels leads to sharpening or broadening of the UV absorption region, the increase and decrease of hydrogen bonds in the solvent can induce bathochromic or hypsochromic effects. Eventually, cations of salts

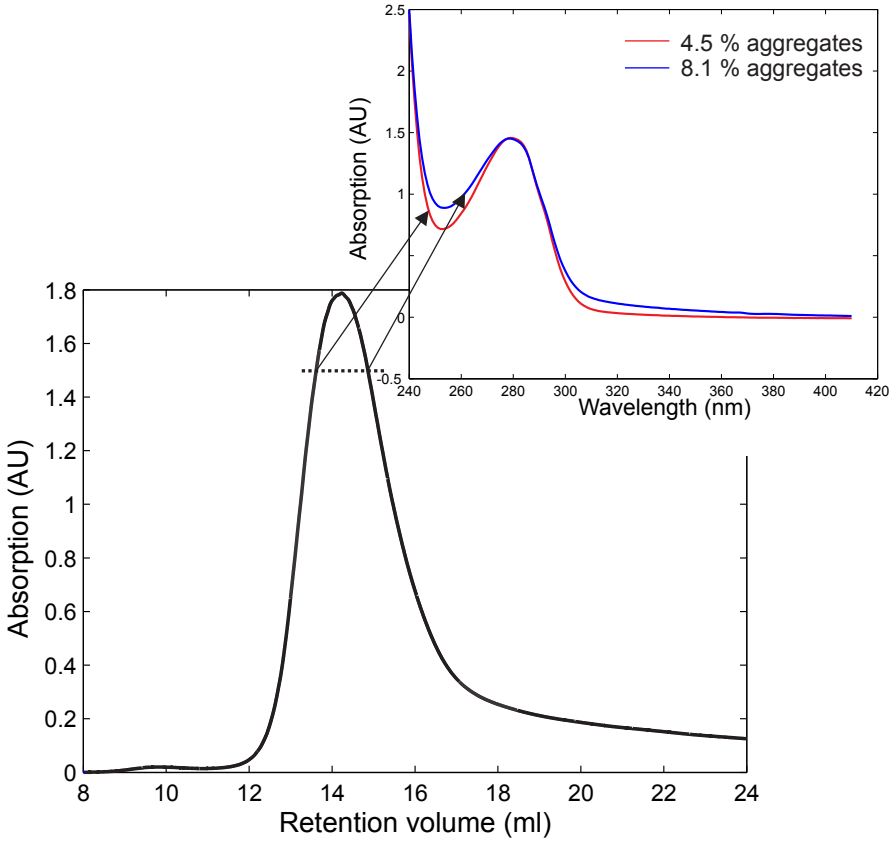


Figure 1.5: Change in the UV absorption spectrum due to an increasing aggregate level. Absorption beyond 320 nm indicates scattering effects. All absorption measurements were performed using a Diode Array Detector with a pathlength of 0.4 mm and then recalculated for a pathlength of 10 mm.

dissolved in the solute can influence the UV absorption. Small cations such as sodium-ions can form complexes with the Pi-electrons of the three aromatic amino acid residues, causing shifts in the absorption bands (Crommelin et al., 2003).

The described changes in protein UV absorption spectra due to a different number and ratio of the aromatic residues or due to conformational changes indicate the feasibility to selectively quantify proteins in multicomponent mixtures using UV spectroscopy (Hansen et al., 2011, 2013). This requires a multivariate regression to correlate the matrix of protein absorption spectra $\mathbf{X}$ (factors) with the matrix of selective protein concentrations $\mathbf{Y}$ (response). The simplest mathematical method to correlate $\mathbf{X}$ with one concentration vector $\mathbf{y}$ is Multilinear Regression (MLR). If $\mathbf{X}$ is however not of full rank (contains collinearities), the regression coefficients of MLR become unstable. Unstable in this context means that they are influenced by measurement noise to a large extend (Næs and Mevik, 2001). As a result, the predictions with this kind of model are error prone, if the newly acquired spectral data does not coincidentally contain the same measurement noise. A common statistical method to remove collinearities from large data sets is MVDA, which will be discussed in the following section.

1.2.2 Multivariate Data Analysis

The application of spectroscopic methods results in large data sets which are usually difficult to interpret. Measuring the UV absorption between 240-300 nm of 30 samples exemplarily results in a 30x61 dimensional matrix. In order to analyze and interpret these high dimensional data sets, mathematical methods are required, which can recognize structures and extract information. These mathematical methods are commonly referred to as MVDA and aim to reduce the dimension of variables in a data set without loosing information. This is achieved by separating information from detector noise and by removing collinearity. The following sections discuss two tools of MVDA: Principal Component Analysis (PCA) and Partial Least Squares regression (PLS).

Principal Component Analysis

PCA first became popular in economic sciences in the 1930s, when Harold Hostelling started to apply this mathematical method for practical case studies (Hotelling, 1933). Its application in chemistry was introduced by Edmund Malinowski (Malinowski, 2002) and Bruce Kowalski (Kowalski et al., 1986) in the 1960s and has become an established method since then. In chemistry it is often, together with the application of other MVDA tools, referred to as chemometrics. The main goal of PCA is to reduce the dimension (number of variables) of a data set by summarizing correlated variables in linear combinations, which are called principal components (PCs). This is done in a way such that the calculated PCs approximate the variance in the data in a least square sense.

Figure 1.6 illustrates the fundamental principle of PCA using a two dimensional example. A plane is spanned by the variables/coordinates x1 and x2. These two variables were measured for ten objects, that are displayed in the plane. The ten objects form two groups, which is the main information in the data. As can be seen in the scatter plots, the variables x1 and x2 are correlated with each other. The left side of Figure 1.6 displays the first step in PCA, the calculation of the first PC, which is a linear combination of x1 and x2 and reflects the direction of largest variance. The right side illustrates the following step in PCA, the calculation of the second PC, which is a linear combination of both variables as well and points into the direction of the remaining, orthogonal variance. It goes through the center of gravity of the data set, which means that $\mathbf{X}$ is mean-centered. Mean-centering is not a requirement for PCA, but allows a better interpretation of the data (above or below average). The two PCs form a new coordinate system, which demonstrates that the PCA effects a transformation of the axes. By projecting the objects onto the new coordinates, the new coordinate values can be obtained. The perpendiculars of the projection are minimized during the calculation of the PCs. The main information in the data (presence of two groups) is already described by the first PC. The second PC describes only measurement noise and can be omitted to reduce the dimension, meaning that the data is projected from the plane onto a straight line. (Kessler, 2006b)

If m variables are measured for n objects ($\mathbf{X} \in \mathbb{R}^{nxm}$), the PCA transforms the objects from an n-dimensional to a k-dimensional space. The PCs form the loading matrix $\mathbf{P} \in \mathbb{R}^{mxk}$ and the new coordinate values on the calculated PCs are summarized in the score matrix $\mathbf{T} \in \mathbb{R}^{nxk}$. The first row of $\mathbf{P^T}$ contains the first PC, explaining the largest part of the variance of the original data. Since it points into the direction of largest

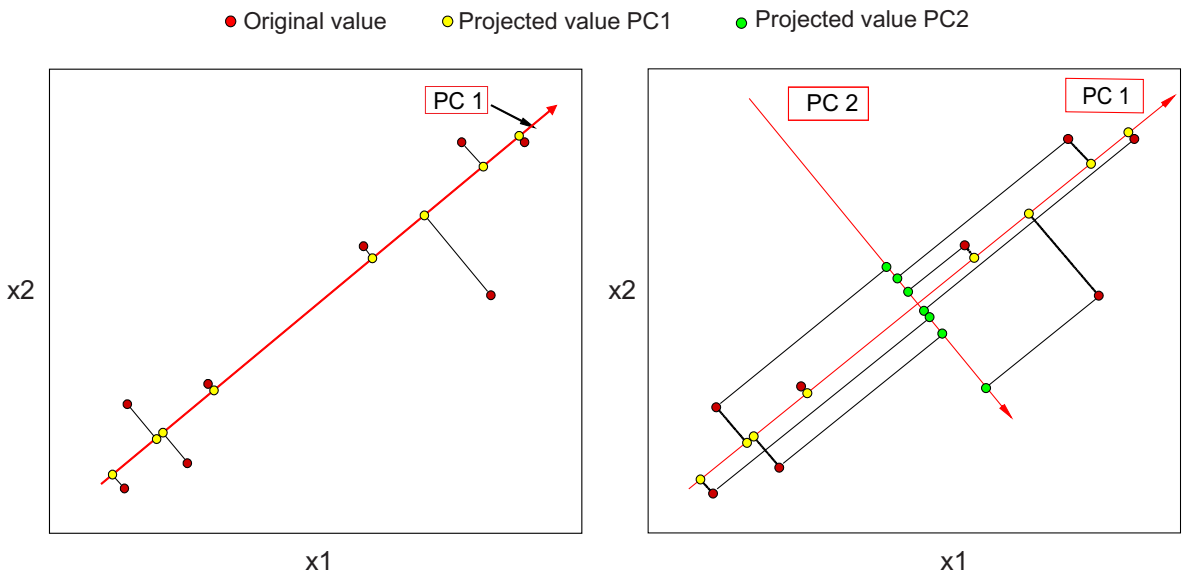


Figure 1.6: The geometrical meaning of the PCs. Left side: The first PC points into the direction of largest variance in the data. Right side: The second PC points into the direction of the remaining, orthogonal variance and goes through the center of gravity of a data set. By projecting the data onto the PCs, the new coordinate values can be obtained. The main information of the data set (presence of two groups) is already described by the first PC. The second PC describes only measurement noise and can be left out to condense the data. Figure taken from (Kessler, 2006b).

variance, a variable with a large scale and variance has a high influence on the direction of the first PC. As a consequence, scaling is required, if variables with different scales are applied. The second PC forms the second row of $\mathbf{P^T}$ and explains the second largest, orthogonal part of the variance and so on. If no data reduction is performed, the number of PCs k equals the original number of variables m. The higher PCs explain however only a marginal part of the variance of the original data and can usually be omitted, such that the data is projected from an m-dimensional to a k-dimensional subspace. The omitted part of the data is then included in the residual matrix $\mathbf{E} \in \mathbb{R}^{nxm}$. This results in the following equation for the factorization of $\mathbf{X}$:

$$\mathbf{X} = \mathbf{TP^T} + \mathbf{E}. \tag{1.1}$$

The dimension reduction separates noise from information and simplifies the data visualization. It also allows the recognition of groups, patterns, and/or outliers of the measured objects as well as the identification of correlations between variables. In addition, qualitative information about non-measurable variables, that are expressed by the PCs, can be gained (Eriksson et al., 2006). For instance, a PCA with UV absorption spectra of different single-component protein solutions results in a group formation, which allows a qualitative conclusion to the tryptophan to tyrosine ratio (Hansen et al., 2013). In order to not only obtain a qualitative but also quantitative information, PLS can be performed. This tool of MVDA is discussed in the following section.

Partial Least Squares Regression

The conduction of PCA with spectral data before the performance of MLR solves the collinearity problem. Instead of correlating $\mathbf{X}$ with $\mathbf{y}$, the condensed score matrix $\mathbf{T}$ can be correlated with $\mathbf{y}$. This means that the objects are described by the score values of PC1 to PCk. The number of applied PCs k is selected such that the PCs explain together a major part of the variance and that the omitted variance is only detector noise. As $\mathbf{T} \in \mathbb{R}^{nxk}$ is of full rank, stable regression coefficients and thus reliable models are to be expected. This method is referred to as Principal Component Regression (PCR). (Kessler, 2006b)

A limitation of PCR is that the chosen PCs are calculated and selected in a way that the relevant variance of $\mathbf{X}$ is explained. However, this does not necessarily mean that these PCs explain the information in $\mathbf{X}$ which is relevant for the correlation with $\mathbf{y}$. In contrast to that, PLS finds components from $\mathbf{X}$ that are also relevant for the correlation with $\mathbf{y}$. These components are called latent variables (LVs). This is why in some cases PLS has been shown to be more efficient than PCR (Martens and Næs, 1989). Due to the high quality of the regression models and the simplicity of its implementation, PLS has become the most popular among the multivariate regressions methods (Lavine, 2000).

There are two different approaches for PLS. The simpler one correlates $\mathbf{X}$ with a vector $\mathbf{y}$ and is usually referred to as PLS1. The more complex approach is called PLS2 and correlates $\mathbf{X}$ with a matrix of different responses $\mathbf{Y}$. As the PLS1 is a special case of the PLS2 approach, the latter is dealt with in the following explanation. The spectral matrix $\mathbf{X} \in \mathbb{R}^{nxm}$ is factorized into the score matrix $\mathbf{T} \in \mathbb{R}^{nxk}$, the loading matrix $\mathbf{P} \in \mathbb{R}^{mxk}$ and a residual matrix $\mathbf{E} \in \mathbb{R}^{nxm}$:

$$\mathbf{X} = \mathbf{T}\mathbf{P}^{\mathbf{T}} + \mathbf{E}. \tag{1.2}$$

For a vectors that form the response matrix $\mathbf{Y} \in \mathbb{R}^{nxa}$, $\mathbf{Y}$ is factorized into the score matrix $\mathbf{U} \in \mathbb{R}^{nxk}$, the loading matrix $\mathbf{Q} \in \mathbb{R}^{axk}$, and a residual matrix $\mathbf{F} \in \mathbb{R}^{nxa}$:

$$\mathbf{Y} = \mathbf{U}\mathbf{Q}^{\mathbf{T}} + \mathbf{F}. \tag{1.3}$$

In contrast to PCA, the k latent variables are not orthogonal. The factorization of $\mathbf{X}$ and $\mathbf{Y}$ is also not performed independently. Instead of finding the major variation in $\mathbf{X}$ and $\mathbf{Y}$, PLS looks for a direction of variance in both, which results in maximum correlation between the t-scores and u-scores. This means that the loadings are rotated in comparison to PCA in order to maximize the covariance between t-scores and u-scores. The covariance between a vector $\mathbf{t}$ with the mean $\bar{t}$ and a vector $\mathbf{u}$ with the mean $\bar{u}$ is defined as

$$Covar(\mathbf{t}, \mathbf{u}) = \frac{\sum_{i=1}^{n}(t_i - \bar{t})(u_i - \bar{u})}{(n-1)}. \tag{1.4}$$

(Kessler, 2006a)

Figure 1.7 illustrates the rotation of the LVs compared to the PCs of a PCA for a two-dimensional example with $\mathbf{X}$ and $\mathbf{Y} \in \mathbb{R}^{9x2}$. The green lines in Figure 1.7 A and B display the calculated PC1s, if a PCA is performed separately for $\mathbf{X}$ and $\mathbf{Y}$. The scores **u1** (PC1 scores for the $\mathbf{Y}$ PCA) are plotted with respect to the scores **t1** (PC1

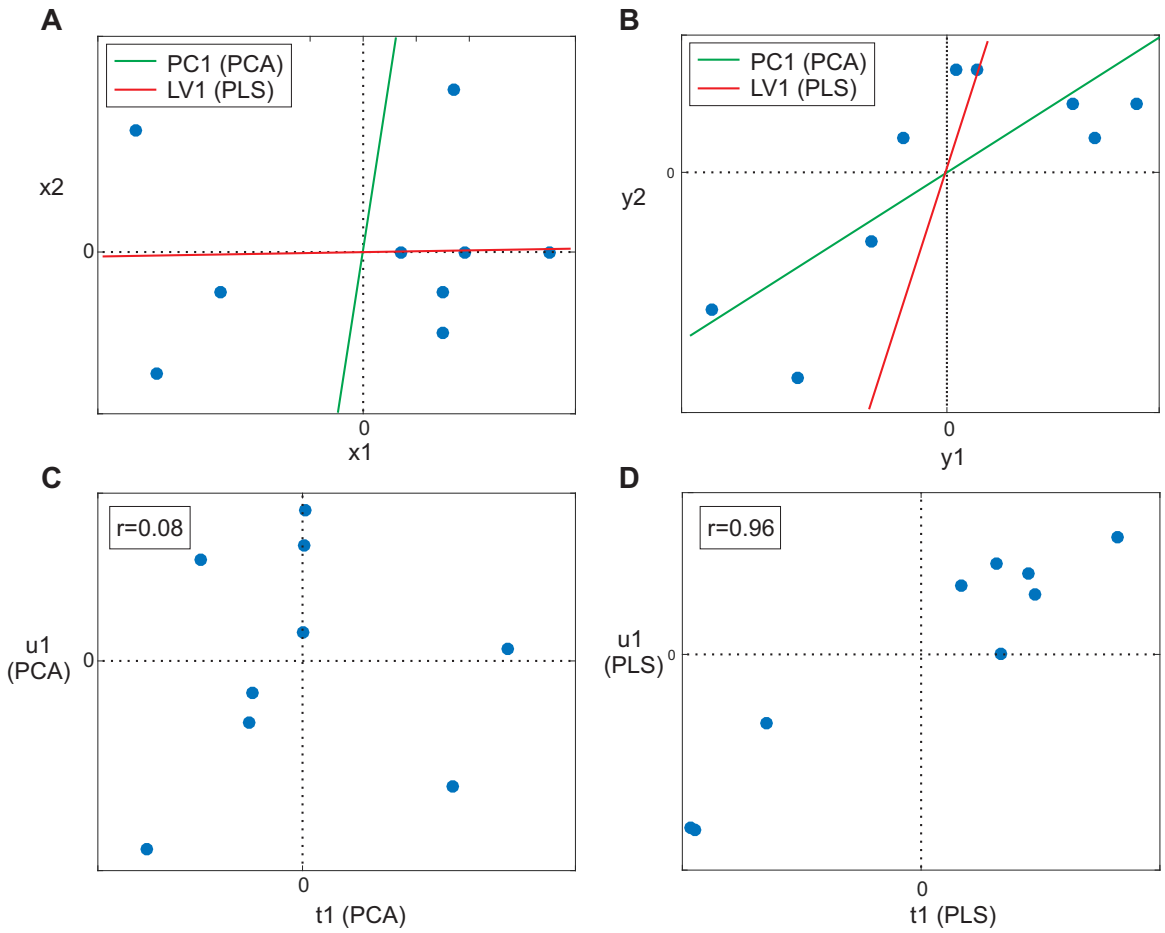


Figure 1.7: The geometrical meaning of the LVs: A: Direction of PC1 (PCA) and LV1 (PLS) for the **X**-data. B: Direction of PC1 (PCA) and LV1 (PLS) for the **Y**-data. C: Correlation between the scores u1 and t1 for two independently performed PCAs (r=0.07). D: Correlation between the scores u1 and t1 for PLS (r=0.96). The data for the performance of PCA/PLS was taken from (University of Copenhagen, 2015).

scores for the **X** PCA) in Figure 1.7 C. No notable correlation between the scores can be observed, which is reflected in a correlation coefficient of 0.08. A correlation coefficient is a normalized covariance, which can be calculated by dividing Equation 1.4 by the product of the standard deviations of **t** and **u**. If PLS instead of two independent PCAs is performed, the directions of the components are rotated (cf. red lines in Figure 1.7 A and B). If the scores on these first LVs are plotted against each other (cf. Figure 1.7 D), a clear correlation can be seen with a correlation coefficient of 0.96. The example was taken from (University of Copenhagen, 2015).

Because of the inner relation **I** between t-scores and u-scores with **U**=**T** **I**, the matrices **Y** and **X** can be correlated via a matrix of regression coefficients $\mathbf{B} \in \mathbb{R}^{mxa}$ using

$$\mathbf{Y} = \mathbf{XB} + \mathbf{G}, \tag{1.5}$$

where $\mathbf{G} \in \mathbb{R}^{nxa}$ is the resulting residual matrix. The regression matrix **B** can be determined by a set of calibration samples, with known concentrations **Y**. After measuring the spectra **X**, PLS can be performed to obtain **B**, which is referred to as model calibration.

The samples applied in the PLS model calibration should be representative and cover the data space as completely and evenly as possible. The higher the number of LVs used in a PLS model calibration, the more precise is the established model for the prediction of $\mathbf{Y}$ in the calibration samples. Thus, the root mean square error (RMSE) between predicted responses $\mathbf{Y_{pred}}$ and the reference $\mathbf{Y_{ref}}$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_{pred} - y_{ref})^2}{n}} \tag{1.6}$$

decreases with increasing number of LVs during calibration. However, the use of too many latent variables can result in overfitting. Detector noise is fitted into the model and the RMSE increases for the prediction of samples with different compositions than the calibration samples. To determine an optimal number of LVs, external or cross validation needs to be performed.

For external validation, calibrated PLS models with different numbers of LVs can be applied to predict the $\mathbf{Y}$ of several validation samples. These samples should exhibit a different composition to that of the calibration samples and ideally be situated in-between the calibration samples in the data space. The RMSE for these samples can then be calculated and plotted as a function of the number of LVs applied in the model calibration to find the optimum number of LVs. If it is not possible to identify or generate representative validation samples, internal cross validation can be carried out. In cross validation, some samples are excluded during model calibration and the RMSE for these samples is calculated subsequently. For every number of LVs, this procedure is performed until each sample was excluded. As for the external validation, the optimal number of LVs corresponds to the minimum in a plot of the RMSE with respect to the number of LVs applied in the model. After PLS model calibration and validation, predictions can be performed. This means that newly acquired $\mathbf{X_{new}}$ data is multiplied by $\mathbf{B}$ to obtain the responses $\mathbf{Y_{new}}$. By this way, time-consuming reference analytics for obtaining $\mathbf{Y_{new}}$ can be omitted.

The PLS technique was successfully applied by Hansen et al. to correlate protein mid-UV absorption spectra with selective protein concentration (Hansen et al., 2011, 2013). For the PLS model calibration and validation, mixtures of pure protein components were generated systematically by applying a design of experiments. In this thesis, this method is further developed as a PAT tool for real-time monitoring and control in chromatography.

PLS models represent empirical models, which are obtained by fitting a calibration data set. Hence, they are limited to the calibrated data space. In contrast to empirical models, mechanistic models are based on physico-chemical laws and can be extrapolated. Instead of delivering a matrix of regression coefficients, the calibration of mechanistic models results in model parameters, which are based on a fundamental understanding of a physical or (bio-)chemical process (Eriksson et al., 2006). In this thesis, a mechanistic model was applied to simulate the physical phenomena during protein separation by cation-exchange chromatography. The fundamental understanding of these phenomena is subsequently used as a PAT tool. This applied mechanistic chromatography model is described in more detail in the following section.

1.2.3 Mechanistic Chromatography Modelling

Several physical phenomena occur during an ion-exchange based chromatographic separation of proteins, which need to be considered for mechanistic modelling. Figure 1.8 illustrates the corresponding transport and adsorption phenomena. The mobile phase, a buffer solution, and the protein mixture to be separated are forced through the column by convection. The column is a packed bed of porous resin particles, which have a sugar or polymer based matrix with ligands and which are referred to as stationary phase. While flowing into the direction of the column outlet, a protein band is subjected to several effects causing peak broadening. One effect is axial dispersion due to eddy diffusion and concentration driven diffusion. In addition, mass transfer limitation results in peak broadening effects. As the stationary phase consists of porous resin particles, film diffusion and pore diffusion take place. Eventually, protein molecules bind to the ligands and the adsorption/desorption kinetics results in peak spreading as well. (Schmidt-Traub, 2005)

In order to obtain a separation, the ligands of the stationary phase are selected such that some proteins of the mixture are retained more strongly than others. Commonly, some of the proteins are first of all adsorbed to the stationary phase, while weakly binding species flow through the column. In a second step, the mobile phase is gradually or stepwise modified to desorb the proteins from the the resin. Due to different interactions with the ligands, a separation of the components is obtained. This mode of chromatography is referred to as bind and elute mode. An alternative to this mode is isocratic elution, where the composition of the mobile phase is not modified. This mode is however seldom applied for cation-exchange based protein purification.

In order to describe the previously mentioned phenomena mathematically, several assumptions need to be made (Schmidt-Traub, 2005):

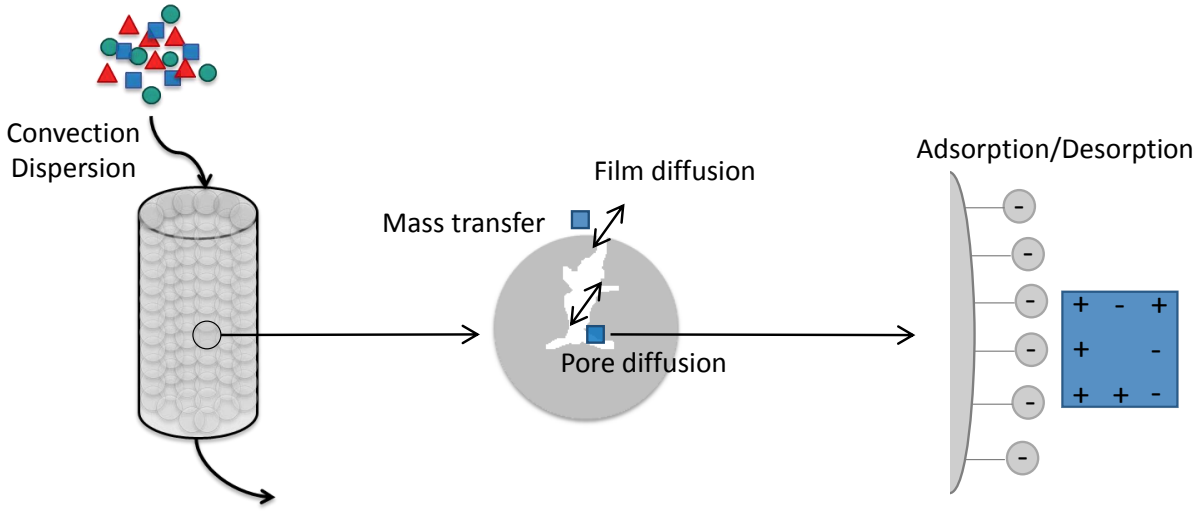


Figure 1.8: Transport and adsorption phenomena during ion-exchange based protein chromatography: Transport phenomena on the column level are convection and dispersion. On the particle level, mass transfer limitation because of film and pore diffusion takes places. On the resin surface, the proteins undergo adsorption/desorption phenomena.

- Size-exclusion effects are lumped into an effective mass transfer coefficient k_{eff}.
- The chromatography experiments are isothermal.
- The mobile phase does not interfere with the stationary phase. Its influence on the adsorption and desorption process is accounted for in the isotherm.
- Fluid density and viscosity are assumed to be constant.
- The size distribution of the resin particles is neglected. All particles are assumed to have the median diameter.
- Only axial distributions in the column are taken into account.
- Convection inside the particles is neglected and a stagnant liquid inside the pores and around the particles is assumed.

Considering those simplifications, various mechanistic models exist that describe the physical phenomena during the separation more or less exactly. The simplest model is the *ideal equilibrium model*, which only considers convection and a thermodynamic adsorption equilibrium. Peak broadening effects are neglected. In contrast, the most detailed chromatography model is called *general rate model* and describes all mentioned transport phenomena as well as adsorption/desorption kinetics. However, computational time increases with model complexity so that simplifications should be made whenever legitimate. In this thesis, a so called *transport dispersive model* was applied. It consists of a balance equation for the liquid and for the stationary phase as well as of an isotherm equation to account for a thermodynamic equilibrium. A simplification in comparison to the *general rate model* is, that all band broadening effects besides from axial dispersion are lumped into one model parameter, the effective mass transfer coefficient k_{eff}. As ion exchange interactions are fast and comparably long ranging, lumping film transfer, pore diffusion, and adsorption/desorption kinetics into one coefficient is legitimate.

Because of the condensing of film and pore diffusion, the protein concentration in the stagnant liquid film around the resin particles equals the protein concentration inside the particle pores. The mobile phase is hence divided into an interstitial volume with concentration c_i of species i and the pore volume with concentration $c_{p,i}$. The concentration change of a species i in the interstitial volume due to convection with interstitial velocity u_{int}, axial dispersion with a coefficient D_{ax}, and lumped mass transfer with a coefficient k_{eff} is described as

$$\underbrace{\frac{\partial c_i}{\partial t}(x,t)}_{\substack{change\,of\\ bulk\,concentration}} = \underbrace{-u_{int}(t)\frac{\partial c_i}{\partial x}(x,t)}_{\substack{convective\\ transport}} + \underbrace{D_{ax}\frac{\partial^2 c_i}{\partial x^2}(x,t)}_{\substack{axial\\ dispersion}} \underbrace{-\frac{1-\varepsilon_b}{\varepsilon_b}\frac{3}{r_p}k_{eff,i}\cdot\left(c_i(x,t)-c_{p,i}(x,t)\right)}_{\substack{mass\\ transfer}} \quad \forall_i, \tag{1.7}$$

where $(1-\varepsilon_b)\cdot 3/r_p$ is the volume based specific surface of the particle bulk with porosity ε_b and particle radius r_p. The species i diffusing into or out of the pores either contribute to a change of the pore concentration $c_{p,i}$ or to a change of the resin load q_i. Thus, the balance equation for the stationary phase with particle porosity ε_p can be described as

$$\underbrace{\frac{3}{r_p}k_{eff,i}\left(c_i\left(x,t\right)-c_{p,i}\left(x,t\right)\right)}_{\substack{mass\\ transfer}} = \underbrace{\varepsilon_p\frac{\partial c_{p,i}}{\partial t}\left(x,t\right)}_{\substack{change\,of\\ pore\,concentration}} + \underbrace{\left(1-\varepsilon_p\right)\frac{\partial q_i}{\partial t}\left(x,t\right)}_{\substack{change\,of\\ resin\,load}} \quad \forall_i. \tag{1.8}$$

(Guiochon et al., 1994; Schmidt-Traub, 2005)

The adsorption/desorption equilibrium is described by the *stoichiometric displacement isotherm* (SDM - stoichiometric displacement model). The SDM is able to model multi-point binding of proteins, as displayed in Figure 1.8. When a protein i with characteristic charge ν_i adsorbs to the resin, ν_i counter ions are freed (index: *salt*). From the law of mass action, the adsorption equilibrium for the protein i with constant $k_{eq,i}$ can be derived as

$$k_{eq,i} = \frac{c_{p,salt}^{\nu_i}\left(x,t\right) q_i\left(x,t\right)}{q_{salt}^{\nu_i}\left(x,t\right) c_{p,i}\left(x,t\right)} \quad \forall_{i\neq salt}. \tag{1.9}$$

Assuming a total ionic capacity of the stationary phase Λ, the bound salt ions can be obtained from

$$q_{salt}\left(x,t\right) = \Lambda - \sum_i \nu_i q_i\left(x,t\right). \tag{1.10}$$

(Velayudhan and Horvath, 1988)

The SDM does not account for steric shielding of ligands by the adsorbed proteins such as considered in the steric mass action isotherm (Brooks and Cramer, 1992). However, shielding effects are negligible for diluted conditions as applied in this thesis, because sufficient binding sides are available.

The resulting system of differential equations was solved numerically using the in-house developed software ChromX (Hahn et al., 2015). The numerical solution was not the focus of this thesis. Summarily, ChromX uses finite elements for space discretization and a fractional step θ-scheme for time discretization. The discrete system of equation is eventually solved with fixed point iteration. As for empirical models, a calibration and validation of the mechanistic model needs to be carried out. Calibration experiments comprise the injection of different tracer solutions to determine the porosities ε_b and ε_p as well as the axial dispersion coefficient D_{ax}. The total ionic capacity of the stationary phase Λ is commonly determined by titration experiments. This is followed by protein experiments under binding and non-binding conditions to obtain the effective mass transfer coefficients $k_{eff,i}$ as well as the isotherm parameters $k_{eq,i}$ and ν_i for all involved species. The calibration procedure is described in more detail in section 6.3. Before the mechanistic model can be applied for predictions, validation experiments should be performed to determine whether extrapolation is possible or if a recalibration is required. Recalibration is necessary, if not all relevant physical phenomena were included in the calibration data. Another reason might be that the selected model was not suitable for the experimental setup and that the applied equations need to be adapted.

1.3 Research Proposal

The rise of biosimilars, the biological equivalents to chemical generics, has induced a larger competition and price pressure in the biopharmaceutical market. Besides from time-to-market and product safety as well as efficacy, development and manufacturing costs are crucial for the success of biosimilars. This has resulted in the development of new technologies such as disposables, alternative unit operations, continuous processing, and PAT. The latter could increase the degree of automation and decrease personnel costs. Especially the evaluation of continuous processing has been a main driver for research into PAT. In addition, health authorities request a larger degree of process and product understanding (QbD-concept) combined with an intensive application of real-time process monitoring and control to guarantee process consistency and product quality (PAT-concept).

Despite of the development of disruptive technologies such as precipitation or aqueous two phase systems, chromatography is still the workhorse in downstream processing of therapeutic proteins. It delivers the required product purity combined with high yields and will most likely continue to be the most important purification technique for therapeutic proteins. Even in the newly investigated continuous processes, the widest applied unit operation is chromatography. For this reason, the development of PAT for chromatography based protein purification was the research objective for this thesis. The major challenge in protein manufacturing to be addressed by this objective was process variability.

Biological processes are very complex and exhibit a high degree of variability compared to their chemical equivalents. Focusing at only one chromatographic step of a whole purification chain, a certain degree of variability enters already the unit operation, as illustrated by Figure 1.9. This variability is combined with uncontrollable influences such as column aging or lot-to-lot variability of resins. Additionally, there is a certain degree of freedom for the controllable operation conditions to be adapted by the manufacturer such as elution conditions or residence times. This all results in an output, the chromatogram of the separation. As a consequence of all variability, this output is variable as well and is forwarded to the next separation step. Advanced monitoring and control strategies for chromatographic separations could however manage process variability. The result of this approach would be an increased reproducibility and thus product quality as well as more efficient processes. Additionally, the realization of PAT is a prequisite for continuous processing and real-time release.

The major focus of this thesis was hence the development of technologies, which cannot only monitor quality and performance attributes, but also manipulate the process in a way such that these attributes are kept in a certain predefined range. The corresponding technologies should preferably operate in real-time and should be a applicable for real-life separation issues. Requirements for real-life separation issues include the calibration of the analytical technologies using feedstocks or in-process materials as well as the operation under conditions where the columns are completely loaded. The most interesting phases to be controlled during a chromatographic separation comprise the load phase and the elution phase. An interesting attribute to be controlled during the column loading is the termination of this phase as soon as the resin is saturated with the product and

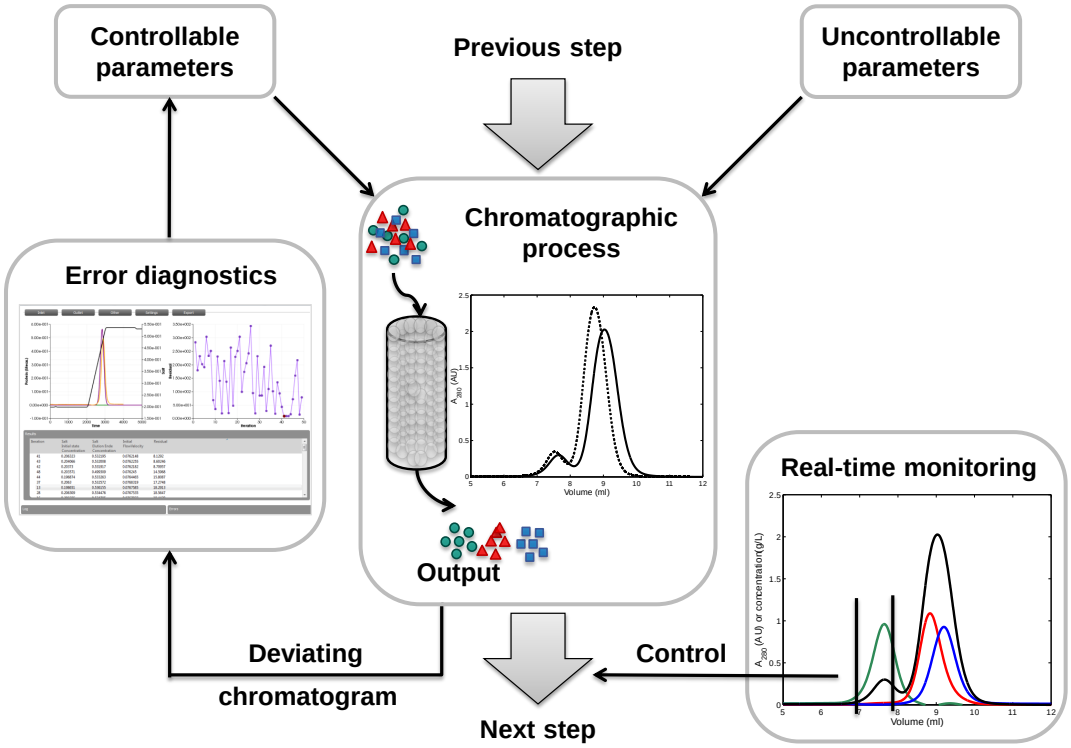


Figure 1.9: Research objective for this thesis: A major challenge in protein manufacturing is process variability, resulting in a variable output of a chromatographic separation. The variability can be compensated by applying technologies that enable real-time monitoring and control. The reason for an observed deviation in a chromatogram can be investigated using mathematical models in order to fix the cause of a deviation in the next cycle or batch.

the product starts to accumulate in the column effluent. During elution, the control of the peak-cutting is desirable. If the elution profiles of the contributing species can be obtained in real-time, as displayed in Figure 1.9, product purity or yield based pooling can be performed. Both the control of the load and the elution phase could compensate process variability and achieve the previously mentioned improvements. However, this approach cannot explain the cause for an observed deviation in a chromatogram. Another goal of this thesis was hence to apply mathematical models for root cause investigations as illustrated by Figure 1.9. If the cause for a deviating chromatogram is identified, an adaption could be made in the next cycle or batch to correct a detected problem.

1.4 Outline

The following chapter 2 describes the development of a PAT tool for the real-time monitoring of selective protein concentrations and the control of the elution phase in chromatography. Chapter 3 and 4 characterize further developments, which enabled the application of the PAT tool for real-life separation issues. While chapter 3 addresses the issue of calibrating the analytical tool using feedstock or in-process materials, chapter 4 deals with preparative conditions, where the columns are completely loaded. In the following chapter 5, the control of the load phase using the further developed PAT tool is discussed. Applying the obtained technologies for the control of the load and the elution phase, process variability can be compensated. Eventually, chapter 6 covers the identification of causes for deviating chromatograms. Inverse mechanistic chromatography modelling in combination with a selective concentration determination by the PAT tool was applied to perform root cause investigations. An overview of all publications and manuscripts of the thesis is given in the following listing:

Chapter 2: A Tool for Selective Inline Quantification of Co-Eluting Proteins in Chromatography Using Spectral Analysis and Partial Least Squares Regression

N.Brestrich, T. Briskot, A. Osberghaus, and J. Hubbuch

Manuscript published in Biotechnology & Bioengineering 111 (2014), p. 1365-1373.

This article presents the development of a tool for selective inline quantification of co-eluting proteins in chromatography. The tool was successfully applied for an inline peak deconvolution of a co-eluting ternary protein mixture consisting of lysozyme, ribonuclease A, and cytochrome c on SP Sepharose FF. Using the gained peak deconvolution by the inline quantification tool, precise purity-based real-time pooling decisions were possible.

Chapter 3: Advances in Inline Quantification of Co-Eluting Proteins in Chromatography: Process-Data-Based Model Calibration and Application Towards Real-Life Separation Issues

N. Brestrich, A. Sanden, A. Kraft, K. McCann, J. Bertolini, and J. Hubbuch

Manuscript published in Biotechnology & Bioengineering 112 (2015), p. 1406-1416.

This article describes the establishment of a process-data-based calibration method for the application of Partial Least Squares regression (PLS) models in protein chromatography. The method uses recorded mid-UV absorption spectra from inline measurements that are correlated with offline fraction analytics to calibrate PLS models. The calibrated models allowed a successful inline quantification of the co-eluting species in a cation-exchange-based aggregate and fraction removal during monoclonal antibody purification. Additionally, the inline quantification of co-eluting serum proteins in an anion-exchange-based purification of Cohn supernatant I was demonstrated.

Chapter 4: Selective Protein Quantification for Preparative Chromatography using Variable Pathlength UV/Vis Spectroscopy

N. Brestrich*, M. Rüdt*, D. Büchler, and J. Hubbuch (*contributed equally)

Manuscript in preparation, to be submitted to Biotechnology & Bioengineering.

This manuscript explores the application of variable pathlength UV/Vis absorption spectroscopy for a selective protein quantification in preparative chromatography. Absorption measurements at variable pathlenghts allowed for an exact detection of both the highly concentrated product peak and lower concentrated contaminants in the peak flanks. In combination with Partial Least Squares regression modelling, variable pathlength UV/Vis spectroscopy was successfully used for a selective quantification of co-eluting lysozyme (fictitious product) and cytochrome c (fictitious contaminant) as well as co-eluting antibody monomer and aggregates.

Chapter 5: Real-time Monitoring and Control of the Load Phase in a Protein A Capture Step

M. Rüdt*, N. Brestrich*, L. Rolinger, and J. Hubbuch (*contributed equally)

Manuscript submitted to Biotechnology & Bioengineering.

This manuscript implements a real-time detection of a monoclonal antibody during the load phase of a Protein A capture step. The method relies on UV/Vis spectroscopy and Partial Least Squares regression to monitor the monoclonal antibody content in the effluent. Besides from the antibody, the effluent consisted of many protein and non protein-based contaminants. Based on the estimated antibody signal, an automated process control for terminating the load phase was established.

Chapter 6: Application of Spectral Deconvolution and Inverse Mechanistic Modelling as a Tool for Root Cause Investigation in Protein Chromatography

N. Brestrich, T. Hahn, J. Hubbuch

Manuscript published in Journal of Chromatography A 1437 (2016), p. 158-167.

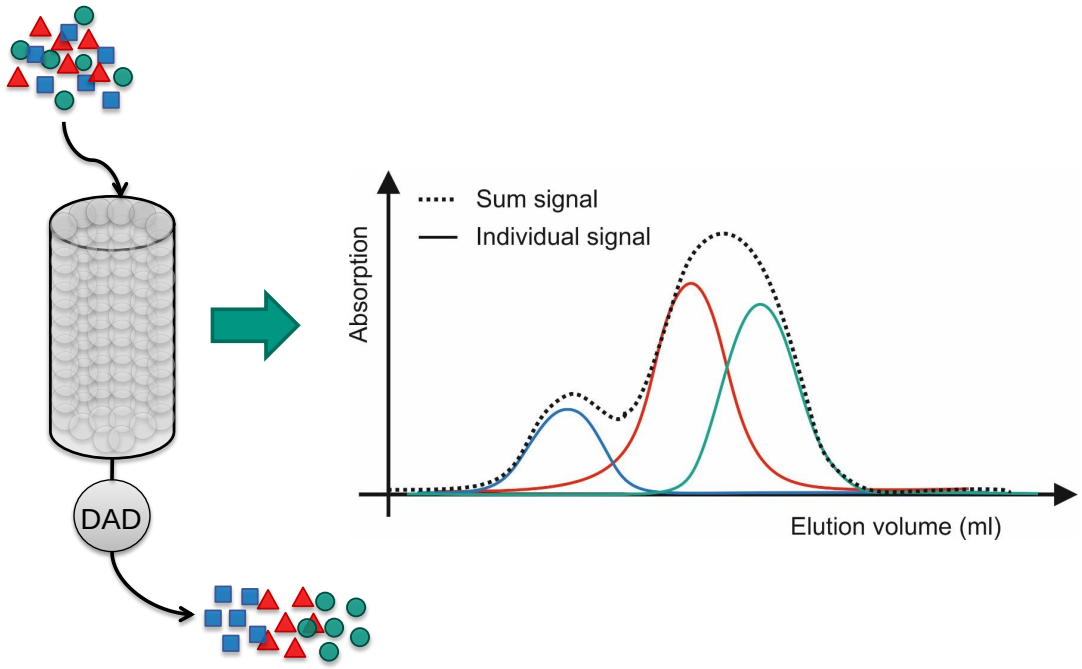
This article describes the development of a tool for rapid root cause investigation in protein chromatography. The tool is based on a combination of spectral deconvolution and inverse mechanistic chromatography modelling. It was capable of identifying the causes for deliberately caused deviations in chromatograms. This allows for a correction of a detected problem in the next cycle or batch.

2 A Tool for Selective Inline Quantification of Co-Eluting Proteins in Chromatography using Spectral Analysis and Partial Least Squares Regresion

N. Brestrich[1], T. Briskot[1], A. Osberghaus[1], J. Hubbuch[1,2]

[1] : Institute of Process Engineering in Life Sciences, Section IV: Biomolecular Separation Engineering, Karlsruhe Institute of Technology, Engler-Bunte-Ring 3, 76131 Karlsruhe, Germany

[2] : Corresponding author. *Email*: juergen.hubbuch@kit.edu



Abstract

Selective quantification of co-eluting proteins in chromatography is usually performed by off-line analytics. This is time-consuming and can lead to late detection of irregularities in chromatography processes. To overcome this analytical bottleneck, a methodology for selective protein quantification in multicomponent mixtures by means of spectral data and Partial Least Squares Regression (PLS) was presented in two previous studies. In this paper, a powerful integration of software and chromatography hardware will be introduced that enables the applicability of this methodology for a selective inline quantification of co-eluting proteins in chromatography. A specific setup consisting of a conventional liquid chromatography system, a diode array detector, and a software interface to Matlab® was developed. The established tool for selective inline quantification was successfully applied for a peak deconvolution of a co-eluting ternary protein mixture consisting of lysozyme, ribonuclease A, and cytochrome c on SP Sepharose FF. Compared to common off-line analytics based on collected fractions, no loss of information regarding the retention volumes and peak flanks was observed. A comparison between the mass balances of both analytical methods showed, that the inline quantification tool can be applied for a rapid determination of pool yields. Finally, the achieved inline peak deconvolution was successfully applied to make product purity-based real-time pooling decisions. This makes the established tool for selective inline quantification a valuable approach for inline monitoring and control of chromatographic purification steps and just in time reaction on process irregularities.

Keywords: Process Analytical Technology, inline monitoring, Chemometrics, Partial Least Squares Regression, Selective protein quantification, Protein analytics, Bioprocess monitoring

2.1 Introduction

Liquid chromatography continues to be the main workhorse of pharmaceutical protein purification, delivering the required product purity and high yields. However, irregularities in chromatography processes due to aging of columns and lot-to-lot fluctuations of feed concentrations, resins, or buffers can lead to reduced pool purities or yields. The real-time monitoring and control of critical process parameters in chromatography can help to manage these irregularities and ensure a stable product quality, an approach in accordance with the US Food and Drug Administration's (FDA) Process Analytical Technology (PAT) concept (FDA, 2004).

The selective titers of the target protein and co-eluting contaminants in the column flow-through are, among others, critical process parameters in protein chromatography and crucial for making pooling decisions or controlling the protein load onto the column. Selective quantification of proteins in the flow-through is commonly realized by fractionation of the elution peak or breakthrough curve of interest and examination of the fractions using off-line analytical chromatography. This is not only elaborate and conflicting the PAT concept, but can also lead to late reaction on process irregularities and thus discarding of batches.

To overcome this analytical bottleneck, various real-time or near real-time monitoring tools for the selective titers of co-eluting proteins in the column flow-through have been developed (Bracewell et al., 1998; Fahrner et al., 1998; Fahrner and Blank, 1999a; Kaltenbrunner et al., 2012; Nilsson et al., 1992; Proll et al., 2004; Rathore et al., 2008a,b, 2009, 2010b). Near-real time monitoring can be realized by atline analysis, the performance of assays in close proximity to the process stream (FDA, 2004). Atline analysis has been applied to monitor the titers of co-eluting misfolded and correctly folded proteins by fluorescence spectroscopy in order to make pooling decisions (Rathore et al., 2009). However, atline monitoring is time-consuming, can lead to infrequent or delayed data and thus late detection of production irregularities, and is consequently rather inefficient for process control (van de Merbel et al., 1996).

A less laborious method for process monitoring, allowing frequent data collection, is the application of online monitoring, the automated diversion and measurement of samples from the production stream (FDA, 2004). Online monitoring allows the use of additional chemicals and the application of established assays. Flow-injection immunoassays (Nilsson et al., 1992) and biosensors (Bracewell et al., 1998; Proll et al., 2004) have been proposed for online monitoring of chromatography processes. However, these monitoring tools only allow the titer determination of a single protein in a multicomponent mixture. Further, online HPLC and UHPLC have been applied for selective protein quantification in multicomponent mixtures (Fahrner et al., 1998; Fahrner and Blank, 1999a; Kaltenbrunner et al., 2012; Rathore et al., 2008a,b, 2010b) and have successfully been used to make real-time pooling decisions (Fahrner et al., 1998; Kaltenbrunner et al., 2012; Rathore et al., 2008a,b, 2010b) and to control antibody loading in protein A chromatography (Fahrner and Blank, 1999a). In addition to the expensive equipment, the major disadvantage of online HPLC or UHPLC is the rather long response time (usually several minutes)(van de Merbel et al., 1996). Although efforts have been made to decrease assay times (Fahrner et al., 1998; Fahrner and Blank, 1999a; Kaltenbrunner et al., 2012; Rathore et al., 2008a,b, 2010b), the response time may require a slowdown of the flow rate (Kaltenbrunner et al., 2012; Rathore et al., 2008a,b, 2010b) and/or the application of mathematical models to predict the protein titers in the upcoming fraction (Rathore et al., 2008a,b, 2010b).

An alternative methodology for selective quantification of substances in multicomponent mixtures is spectral analysis in combination with chemometrics-based, multivariate regression techniques such as Principal Component Regression (PCR) or Partial Least Squares Regression (PLS). This methodology has been promoted by the FDA's PAT guidance for process monitoring and control (FDA, 2004). It is rapid, non-invasive, cost-effective and therefore very suitable for online and inline monitoring, the continuous measurement using sensors directly inserted in the production stream (FDA, 2004). Among the available multivariate regression techniques, PLS has become the most popular due to the quality of the regression models and the simplicity of its implementation (Lavine, 2000). Especially for selective quantification using spectral data, PLS is very suitable, as it can handle the collinearity in the data (Næs and Mevik, 2001).

In the past decade, several tools using spectral analysis in combination with PLS for online or inline monitoring in biotechnology have been reported (Arnold et al., 2002; Boehl, 2003; Haack et al., 2004; Jung et al., 2002; Kara et al., 2010; Navrátil et al., 2005;

Tamburini et al., 2003). In upstream-processing, fluorescence, near-infrared, and mid-infrared spectra in combination with PLS have been applied for online/inline monitoring of biomass (Arnold et al., 2002; Boehl, 2003; Haack et al., 2004; Navrátil et al., 2005; Tamburini et al., 2003), substrate (Jung et al., 2002; Kara et al., 2010; Mazarevica et al., 2004; Navrátil et al., 2005; Tamburini et al., 2003), product (Boehl, 2003; Kara et al., 2010; Mazarevica et al., 2004), and metabolite concentrations (Navrátil et al., 2005; Tamburini et al., 2003). In contrast to that, the use of spectral analysis in combination with chemometrics-based, multivariate regression techniques for downstream process monitoring is rather uncommon (Rathore et al., 2011). For instance, atline monitoring of a flocculation process for removing contaminants such as cell debris, host cell proteins, and nucleic acid could be achieved using near infrared spectroscopy and PLS (Yeung et al., 2000). During an ultrafiltration-based separation of a ternary protein mixture, atline monitoring of the selective protein concentrations in the retentate and permeate was performed by fluorescence spectra in combination with PLS (Elshereef et al., 2010).

Further, several studies in downstream processing showed that spectral analysis in combination with PLS can be used for selective off-line quantification of biomolecules and declared this methodology as promising for process monitoring (Capito et al., 2013b; Hansen et al., 2011, 2013; Kamga et al., 2013; Zang et al., 2013). Fourier transform mid-infrared spectroscopy in combination with PLS has been used for selective off-line quantification of host cell proteins in harvested cell culture fluid (Capito et al., 2013b). In the area of chromatographic purification of glycosaminoglycans, near-infrared spectroscopy combined with PLS has successfully been applied to perform selective off-line quantification of the product in collected fractions (Zang et al., 2013).

In two previously published studies (Hansen et al., 2011, 2013), a methodology for selective off-line quantification of proteins in multicomponent mixtures using UV-visible spectra and PLS was presented. The used spectral data was in the band of 240-300 nm, where mainly the aromatic structures of phenylalanine, tyrosine, and tryptophan contribute to UV-absorption. Due to unequal number of these amino acid residues in different protein species, protein mid-UV spectra vary and allow selective protein quantification in multicomponent mixtures using PLS. The method has been shown to work for various protein mixtures and has successfully been used as analytics for high throughput chromatography experiments using binary and ternary protein mixtures. Recently, other researchers have also confirmed the applicability of mid-UV spectra in combination with PLS for selective off-line protein quantification in multicomponent mixtures (Kamga et al., 2013).

Off-line analytics using mid-UV spectra in combination with PLS can increase the analytical throughput of collected fractions significantly. However, the methodology does not allow a real-time detection of process irregularities in chromatography. Consequently, no product-purity based pooling decisions are possible. In this study, a powerful integration of software and chromatography hardware was established. This integration enables the applicability of mid-UV spectra in combination with PLS for a selective inline quantification of co-eluting proteins in chromatography. A specific setup consisting of a conventional liquid chromatography system, a diode array detector, and a software interface to Matlab® was established. As model system, a ternary protein mixture consisting of lysozyme, ribonuclease A, and cytochrome c was selected. A PLS model for

the ternary protein mixture was calibrated and validated using a D-optimal onion design. The calibrated PLS model and the established specific setup were afterwards applied for a selective inline quantification and peak deconvolution of the co-eluting ternary protein mixture on SP-Sepharose FF. To validate the established inline quantification tool, the elution peak was furthermore analyzed by fractionation and off-line analytical chromatography. Finally, the established tool for selective inline protein quantification was adapted to make product purity-based real-time pooling decisions.

2.2 Materials and Methods

2.2.1 Model Proteins and Buffers

As model system, a ternary protein mixture consisting of lysozyme from hen egg white (lys), ribonuclease A from bovine pancreas (rib A), and cytochrome c from equine heart (cyt c) was selected. This system was chosen due to the stability of the proteins in aqueous solutions regarding the formation of multimers and due to their advantageous isoelectric point distribution, allowing separation at neutral pH by cation exchange chromatography. All proteins were purchased from Sigma-Aldrich, St. Louis, USA.

The buffer for the preparation of calibration and validation protein samples for the PLS model and for loading in chromatography was 20 mM sodium phosphate (pH 7). Elution was performed using a high-salt buffer consisting of 20 mM sodium phosphate as well as 500 mM sodium chloride (pH 7). All buffer components were purchased from VWR, West Chester, USA. Buffers were 0.22 μm filtrated and degassed by sonification before usage.

2.2.2 Liquid Handling Station

The generation of protein solutions for the calibration and validation of the PLS model were carried out using a Tecan Freedom Evo® 75 liquid handling station (LHS), equipped with a Liquid Handling Arm and a Robotic Moving Arm (Tecan, Männedorf, Swiss). The LHS was operated with EVOware® 2.3. The import of values for pipetting volumes was handled using Matlab® (MathWorks, Natick, USA).

2.2.3 Chromatographic Instrumentation

The protein spectra measurement of the calibration and validation samples as well as the chromatographic separation were performed using a specific experimental setup consisting of a conventional liquid chromatography system, a diode array detector, and a software interface to Matlab®. The liquid chromatography system was an Akta® purifier 10 equipped with pump P-900, UV monitor UV-900 (10 mm optical path length), conductivity monitor C-900, pH monitor pH-900, autosampler A-905, and fraction collector Frac-950 (all GE Healthcare, Chalfont St Giles, UK). The instrumentation was controlled with Unicorn® 5.31 software. In order to perform inline protein spectra measurements, a Dionex UltiMate® 3000 Diode Array Detector (DAD) equipped with a semi-preparative flow cell (0.4 mm optical path length) and operated with Chromeleon® 6.80 software

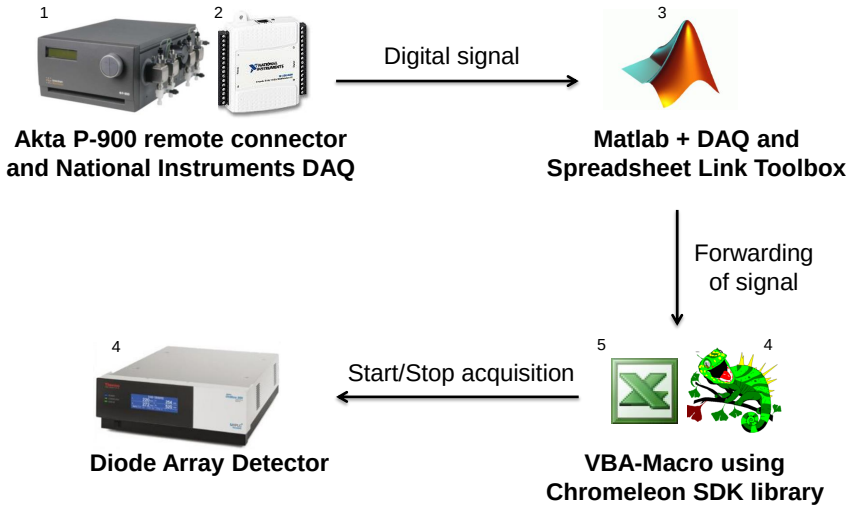


Figure 2.1: Trigger for the data acquisition of the DAD: The fact of sample injection or the end of an experiment was first submitted to Matlab® by sending a digital signal to the Matlab® DAQ toolbox using the REMOTE connector of the Akta® P-900 pump in combination with an USB-6008 DAQ device. Afterwards, the signal was forwarded to an VBA Macro using the Matlab® Spreadsheet Link EX toolbox. The VBA Macro then instructed the Chromeleon® software to start or stop data acquisition using the Chromeleon® Software Developer Kit. Graphics taken from / copyright by 1: GE Healthcare, 2: National Instruments, 3: Mathworks, 4: Dionex, 5: Microsoft.

(Thermo Fisher Scientific, Waltham, USA) was connected to the Akta® Purifier 10. The data acquisition of the DAD was triggered after a sample injection as follows: The fact of sample injection was first submitted to Matlab® by sending a digital signal to the Matlab® Data Acquisition (DAQ) toolbox using the remote connector of the Akta® P-900 pump in combination with an USB-6008 DAQ device (National Instruments, Austin, USA) (cf. Fig. 2.1). Afterwards, the signal was forwarded to an Excel Visual Basic for Application (VBA) Macro (Microsoft, Redmond, USA) using the Matlab® Spreadsheet Link EX toolbox. The VBA Macro then instructed the Chromeleon® software to start raw data acquisition using the Chromeleon® Software Developer Kit (Thermo Fisher Scientific, Waltham, USA). At the end of an experiment, the DAD was triggered to stop data acquisition using the same interface.

2.2.4 PLS Model Calibration and Validation

The calibration of a PLS model was based on a four level D-optimal onion design, generated with MODDE (Umetrics, Umeå, Sweden). The design consisted of 29 samples, including three center points, and covered 19 mixing ratios and 9 concentration levels from 0 to 0.7 g/L. An additional concentration level at 0.05 g/L was added to the design as pure samples to support a good model quality at low protein concentrations. The validation of the model was performed using a three level D-optimal onion design, generated with MODDE. The validation design consisted of 25 samples, including three center points, and was covering 16 mixing ratios and 8 concentration levels from 0 to 0.7 g/L. As for the calibration design, an additional concentration level with low protein concentra-

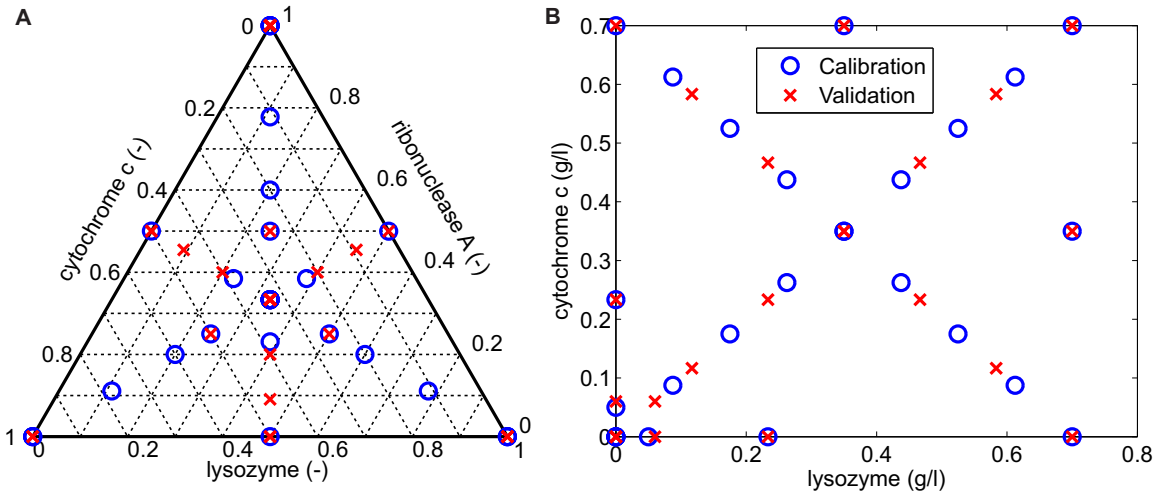


Figure 2.2: Mixing ratios of all three proteins (A) and absolute concentrations of cytochrome c and lysozyme (B) applied in the calibration and validation sample set. The calibration of the PLS model was based on a four level D-optimal onion design, while the validation of the model was based on a three level D-optimal onion design.

tions of 0.06 g/L was added as pure samples. The mixing ratios of all three proteins for both the calibration and the validation sample set are displayed in Figure 2.2 A. Further, the absolute concentrations of lys and cyt c for both the calibration and the validation are displayed in Figure 2.2 B. The mixing ratios were prepared in ABgene polypropylene 96 deep well plates (volumetric capacity of 1 ml, Thermo Fisher Scientific, Waltham, USA) using the LHS. All mixing ratios were prepared from stock solutions, containing 3.5 g/L protein in 20 mM sodium phosphate buffer.

In order to avoid systematic errors, the pipetting order of the calibration and validation samples was randomized using Matlab®. The spectral data measurement of the samples was realized by injecting them into the described chromatographic setup with DAD until a plateau of the UV absorption at 280 nm was reached. For each injected calibration solution, the absorption spectrum in the band of 240-300 nm with 1 nm resolution was measured. The PLS model calibration based on the obtained spectral data was finally performed with Matlab® using the *plsregress* function. The *plsregress* function preprocesses the spectral data by mean centering and uses the SIMPLS-algorithm for the PLS model generation.

PLS aims to reduce data by only applying significant variance in data for correlating several input variables with variables of interest. To achieve a data reduction, variables containing similar information are summarized in so-called latent variables (LV's). Thus, the number of LV's expresses a measure of data reduction, where a low number of LV's corresponds to a high reduction. The information content included in each LV is high for the first LV and decreases for the following LV's. This means that the variance in the data can be explained by only a few LV's, leaving out higher LV's that mainly contain detector noise. The higher the number of LV's used in a PLS model, the more precise is the established model for the calibration sample set. However, the use of too many latent variables can result in overfitting. In this case, the prediction of the variables of interest for other samples, such as the validation samples, decreases. Therefore, the

thorough examination of the number of latent variables used in a PLS model is crucial for its prediction precision (Eriksson et al., 2006). The determination of a reasonable number of LV's was realized by calculating the root mean square error (RMSE) for both calibration and validation in dependence on the numbers of LV's used in the PLS model. The final PLS model was applied to calculate the selective protein concentrations of the validation samples from their measured protein spectra. The residuals between predicted and actual protein concentrations were calculated to validate the PLS model. Normally distributed residuals below 0.05 g/l were declared as acceptable.

2.2.5 Inline Quantification of Co-eluting Proteins

A HiTrap® 1 ml column prepacked with SP Sepharose FF (GE Healthcare, Chalfont St Giles, UK) was first equilibrated with 6 column volumes (CV) of loading buffer and then loaded with 0.2 mg of each protein. After the protein load, the DAD was triggered to start data acquisition and recorded protein spectra in the band of 240-300 nm with 1 nm resolution. As displayed in Figure 2.3, the spectra were transferred to Matlab® for selective protein quantification using the calibrated PLS model. The interface between Chromeleon® and Matlab® was realized with a VBA Macro in combination with the Chromeleon® Software Developer Kit. After protein loading, a wash of 1 CV loading buffer was performed and the proteins were then eluted by a linear gradient from 0 to 500 mM sodium chloride in 3 CV's. The flow rate was 0.3 mL/min for all steps. The elution peaks were collected in a microtiter plate (Greiner BioOne, Kremsmünster, Austria) in 150 μl fractions for analytical chromatography. Based on the resulting chromatogram of the inline measurement, mass balances for all three proteins were calculated using the *trapz* function of Matlab®.

2.2.6 Analytical Chromatography

In addition to selective inline protein quantification by spectral analysis, off-line analytical chromatography was performed with the collected fractions, using a Dionex UltiMate® 3000 liquid chromatography system (Thermo Fisher Scientific, Waltham, USA). The system was composed of a HPG-3400RS pump, a WPS-3000TFC-analytical autosampler, a TCC-3000RS column thermostat, and a DAD3000RS detector. For each fraction, a Proteomix SC X-NP1.7 2.5 ml column (Sepax Technologies, Delaware, USA) was first equilibrated with 2 CV's of loading buffer, loaded with 20 μl sample, and eluted by a linear gradient from 0 to 500 mM sodium chloride in 4 CV's. The flow rate was 0.5 mL/min. As for the inline quantification, mass balances were calculated by Matlab® based on the determined concentrations in the fractions.

2.2.7 Real-time Pooling Decisions

The established tool for selective inline quantification of co-eluting proteins was applied for product purity-based real-time pooling decisions. Therefore, rib A was declared as target protein for a first experiment and was separated from the two other proteins on the HiTrap® column prepacked with SP Sepharose FF with a gradient length of 3 CV's. In a second experiment, cyt c was selected as target and separated from the two other proteins

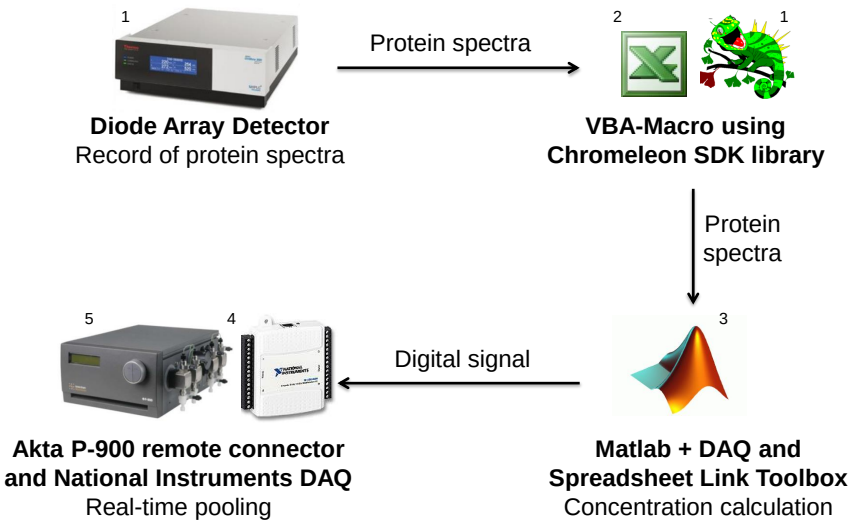


Figure 2.3: Selective inline protein quantification and real-time pooling decision: The DAD recorded protein spectra in the band of 240-300 nm with 1 nm resolution. The spectra were transferred to Matlab® for selective protein quantification using the calibrated PLS model. The interface between Chromeleon® and Matlab® was realized with a VBA Macro in combination with the Chromeleon® Software Developer Kit. When the criterion to start or stop pooling was detected by the Matlab®, a digital signal was sent to the REMOTE connector of the Akta® P-900 pump via the Matlab® Data Acquisition toolbox and the USB-6008 data acquisition device. Graphics taken from / copyright by 1: Dionex, 2: Microsoft , 3: Mathworks, 4: National Instruments, 5: GE Healthcare.

with a gradient length of 5 CV's. The flow rate was 0.3 mL/min for both experiments. The criterion to start pooling was a target protein concentration above 0.002 g/L as well as a mass fraction of the target above 0.8 in the column flow-through. As soon as this criterion was detected by the Matlab® function that was calculating the selective protein concentrations, a digital signal was sent to the REMOTE connector of the Akta® P-900 pump via the Matlab® Data Acquisition toolbox and the USB-6008 data acquisition device (cf. Fig. 2.3). This digital signal was then used in Unicorn® as criterion to start pooling. The criterion to stop pooling was a mass fraction of the target protein below 0.8. The end of pooling was managed using the interface between Matlab® and Unicorn® that was described above. The time span between the detection of the criterion to start or to stop pooling by Matlab® and the reaction on the signal by Unicorn® was below 1 s. Based on the selective inline quantification, the pool purities and yields were calculated.

2.3 Results and Discussion

Product purity-based real-time pooling decisions in liquid chromatography require a fast and reliable analysis of the column flow-through. This work demonstrates the applicability of a methodology for selective protein quantification (Hansen et al., 2011, 2013) as an inline monitoring tool in chromatography. The methodology uses a PLS model to correlate protein UV-visible spectra with selective protein concentrations. The inline application of the methodology for selective protein quantification requires a specific experimental setup consisting of a liquid chromatography system, a DAD, and a software

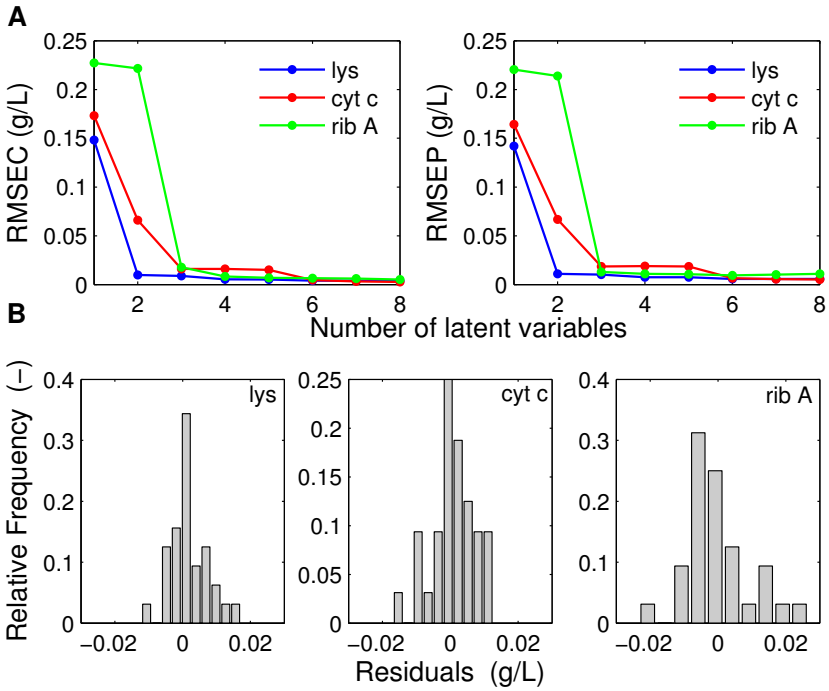


Figure 2.4: **A**: Calculated root mean square error for the calibration (RMSEC) and validation (RMSEP) with respect to the number of latent variables used for the PLS Model calibration. **B**: Relative frequency distribution of the residuals of lysozyme, cytochrome c, and ribonuclease A in the model validation for 6 latent variables.

interface to Matlab®. The applied detector type (DAD) was different from the detector used in (Hansen et al., 2011, 2013). Consequently, a new PLS model for the used ternary protein mixture needed to be calibrated and validated. The calibrated model and the established experimental setup were afterwards applied for selective inline protein quantification in the column flow-through. The selective inline protein quantification was thoroughly validated by comparing it with off-line analytical chromatography. In order to prove the practical relevance of this new monitoring tool, it was shown to be applicable for real-time pooling decisions.

2.3.1 PLS Model Calibration and Validation

PLS aims to reduce data by only applying significant variance in data for correlating several input variables (here protein spectra) with variables of interest (here selective protein concentrations). To achieve a data reduction, variables containing similar information are summarized in latent variables (LV's). A thorough examination of the number of latent variables used in a PLS model is crucial for its prediction precision (Eriksson et al., 2006). In order to determine the optimal number of latent variables, the root mean square error for the calibration samples (RMSEC) and the root mean square error for the prediction (RMSEP) of the selective concentrations in the validation sample set were calculated for eight different PLS models with one to eight LV's used in the calibration.

Figure 2.4 A displays the calculated RMSEC's and RMSEP's of the eight PLS models in dependence on the number of latent variables used in the model calibration. For both

lys and cyt c, a significant decrease in the RMSEC and RMSEP was observed when two instead of one latent variable was used, which shows that the second latent variable contained a major part of the information about the concentration of these two proteins. The first latent variable mainly contained information about the total protein concentration (data not shown). The major part of the information about the rib A concentration was in the third latent variable. While the RMSEC was found to decrease for all proteins with increasing number of LV's, the RMSEP decreased continuously for lys and cyt c and exhibited a minimum for ribonuclease A at six latent variables. Additionally, a significant decrease of the RMSEC and RMSEP of cyt c was observed, when the number of latent variables was increased from five to six. Therefore, a PLS model with six LV's was selected. The six LV's explained 99.96 percent of the variance in the protein concentrations of the calibration sample set and 99.92 percent of the variance in the protein concentrations of the validation sample set.

For the PLS model with six LV's, the relative frequency distribution of the residuals between predicted and actual protein concentrations of the validation sample set was calculated for all three proteins. The results are displayed in Figure 2.4 B. The residuals of all three proteins were nearly normally distributed, which indicates that systematic errors occurred during the PLS model calibration are rather unlikely. The residuals lay between -0.0121 and 0.0171 g/L for lys, -0.0168 and 0.0126 g/L for cyt c, and between -0.0232 and 0.0265 g/L for rib A, implying a high prediction precision of the calibrated PLS model with six LV's.

2.3.2 Inline Quantification of Co-eluting Proteins

The calibrated PLS model was applied for an inline quantification of co-eluting lys, cyt c, and rib A during a chromatographic separation on SP sepharose FF with a gradient length of 3 CV's. The established experimental setup consisting of an Akta® liquid chromatography system, a diode array detector, and a software interface to Matlab®, allowed a selective protein quantification in real-time. Figure 2.5 displays the resulting chromatograms of the inline measurement. The predicted, selective protein concentrations by the PLS model (cf. Fig. 2.5 B, solid lines) are compared with both the absorption at 280 nm (A_{280}) and at 527 nm (A_{527}) (cf. Fig. 2.5 A) and with the determined protein concentrations in the collected fractions using off-line analytical chromatography (cf. Fig. 2.5 B, dashed lines).

Good agreement was found between the retention volume of the first peak in the A_{280} signal and the rib A concentration profile from the inline measurement. The retention volume of cyt c from the inline measurement corresponded to the peak maximum of the A_{527} signal, which is characteristic for cyt c. Further, both the selective inline quantification and the off-line analytical chromatography resulted in the same retention volume for all three model proteins and exhibited good agreement in the peak flanks. However, the protein peaks determined from the off-line analytical chromatography exhibited negligibly more tailing to some degree, which might be explained by additional dispersion in the capillary between DAD and fraction collector. The good agreement with both the absorption measurement and the results from the analytical chromatography prove the precision and applicability of the established inline quantification tool.

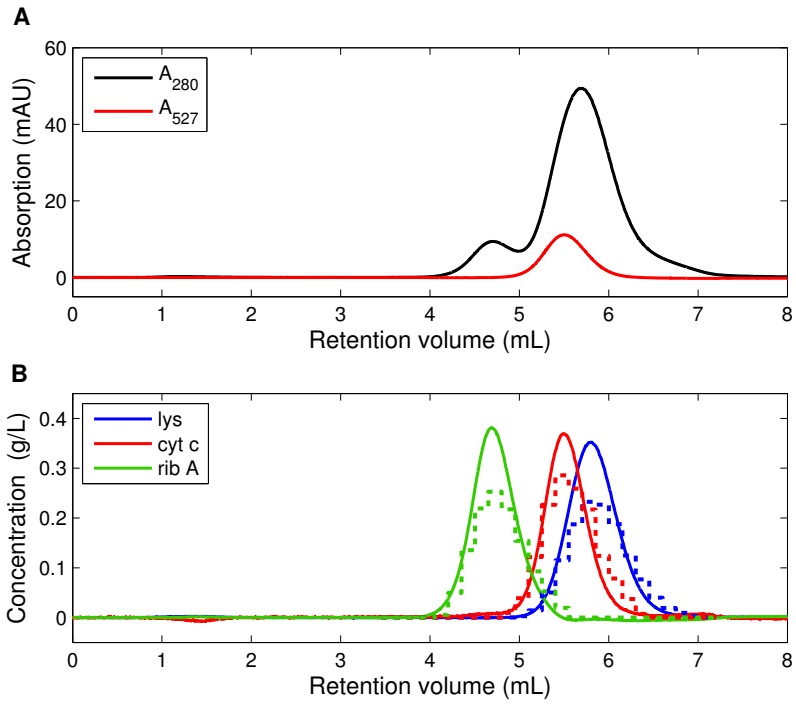


Figure 2.5: Resulting chromatograms of the inline measurement and off-line analytical chromatography. **A**: Absorption at 280 and 527 nm plotted against the retention volume. **B**: Predicted protein concentrations by the PLS model (solid lines) and determined protein concentrations in collected fractions using off-line analytical chromatography (dashed lines) plotted against the retention volume.

Regarding the protein concentration at the peak maximums of all three proteins, small deviations were found between both analytical methods (cf. Fig. 2.5 A). A slightly better agreement in the protein concentrations at the peak maximums might have been achieved, if the resolution of the collected fractions had been increased. However, the inline protein quantification tended to show slightly higher protein concentrations in the peak maximums compared to analytical chromatography.

Table 2.1 shows the calculated mass balances for both analytical methods. The mass balances resulted in a range between 108-117% for the inline quantification and in a range of 89-108% for analytical chromatography. While the inline quantification was found to slightly overestimate the protein masses in the column flow-through, analytical chromatography resulted in mass balances below 100% for lys and rib A. The slight overestimation of protein masses by the inline quantification mainly resulted from an

Table 2.1: Mass balances (in %) for lys, cyt c, and rib A for both selective inline protein quantification and off-line analytical chromatography.

	Analytical chromatography	inline quantification
Lys	95.80	116.60
Cyt c	108.68	113.45
Rib A	88.78	107.85

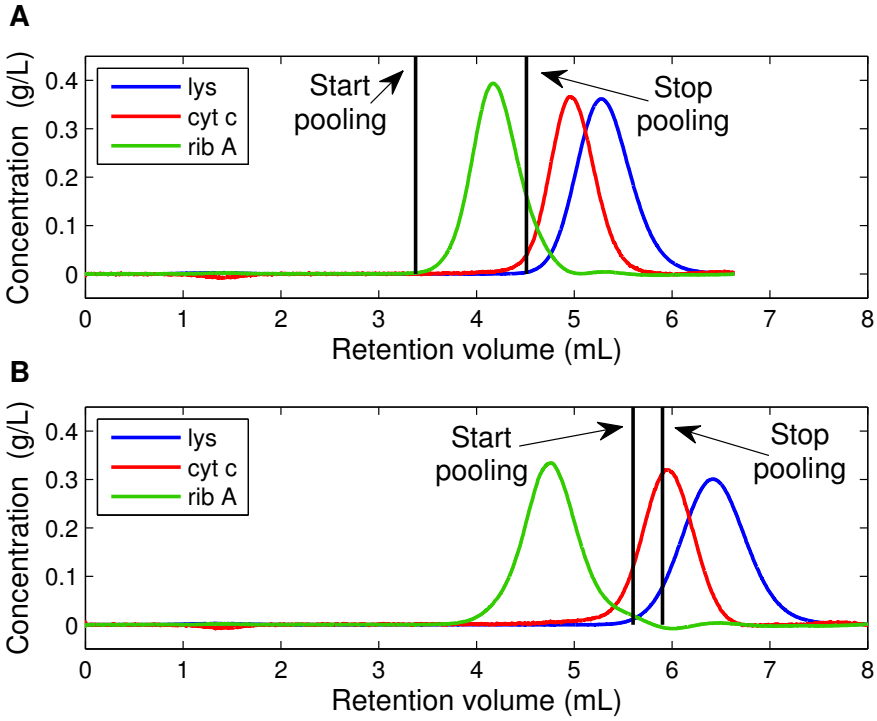


Figure 2.6: Resulting chromatograms of the product purity-based real-time pooling during a chromatographic separation of lys, cyt c, and rib a on SP Sepharose FF. The black lines in the chromatograms visualize the the pool borders that were automatically detected by Matlab® and forwarded to the software Unicorn® . **A**: Rib A was declared as target protein and separated with a gradient length of 3 CV's. **B** Cyt c was declared as target protein and separated with a gradient length of 5 CV's.

overestimation of the protein concentrations at the peak maximums. However in the validation of the PLS model, an overestimation for samples with similar concentrations as found in the peak maximums was not observed. The residuals between actual and predicted protein concentrations in the validation were all below 0.03 g/L (cf. Fig. 2.4 B). Consequently, an error in the PLS model calibration for this concentration area can be excluded. A substantiated explanation for the observed overestimation is however currently not at hand. But although the mass recovery of the analytical chromatography was marginally more accurate, the predicted selective masses by the PLS model are acceptable regarding the fact that they can be determined inline. Consequently, the established tool can also be applied for a rapid determination of pool yields.

2.3.3 Real-time Pooling Decisions

The established tool for selective inline quantification of co-eluting proteins was finally applied for real-time pooling decisions and a rapid determination of pool purities and yields. In a first exemplary experiment, rib A was declared as target protein and separated from the "contaminants" lys and cyt c on SP sepharose FF with a gradient length of 3 CV's. In a second experiment, cyt c was the target and attempted to be separated using a gradient length of 5 CV's. For both experiments, the criterion to start pooling was a target protein concentration above 0.002 g/L and a mass fraction of the target above 0.8. The criterion to stop pooling was a mass fraction of the target below 0.8. The

established, selective inline quantification tool as well as the established interface between Matlab® and Unicorn® allowed an automated, product purity-based pooling in real-time. The resulting chromatograms of both experiments are displayed in Figure 2.6. The black lines in the chromatograms visualize the the pool borders that were automatically detected by Matlab® and forwarded to the software Unicorn® which was operating the liquid chromatography system. Finally, the mass balances gained by the selective inline quantification and the pool borders were used to calculate the product yield and purity in the collected pools. The calculated yields were 88.8% for rib A and 36.3% for cyt c respectively. The determination of the pool purity resulted in 96.9% for the rib A pool and in 83.5% for the cyt c pool.

2.4 Conclusion and Outlook

A tool for selective inline quantification of co-eluting proteins in chromatography was successfully established. A specific setup consisting of a conventional liquid chromatography system and a DAD was developed to perform inline measurements of protein spectra. To correlate the protein spectra with selective protein concentrations, a PLS model was calibrated. An established software interface allowed a selective inline quantification of the co-eluting proteins by Matlab® using the calibrated PLS model and the protein spectra. The tool for selective inline quantification was rigorously compared with off-line analytical chromatography using collected fractions and found to deliver precise results for both the peak flanks and the retention volumes of co-eluting proteins. Further, a comparison between the mass balances based on analytical chromatography and inline quantification showed, that the established tool can be applied for a rapid determination of pool yields. Using the gained peak deconvolution by the inline quantification tool, precise purity-based real-time pooling decisions were possible. Consequently, a reliable inline monitoring tool was established that can be used to detect process irregularities. The presented tool has a great potential to manage irregularities in preparative chromatography processes by real-time pooling. Future work will focus on more complex applications with pharmaceutically important proteins and separation challenges. One first step in this context will be the elimination of the necessity of having each protein as a pure component for the PLS model calibration. This could be achieved by performing the calibration with samples from the chromatographic separation process to be monitored. As each component in a process needs to be accounted for in the calibration in order to use the inline quantification tool, a possible application could be the inline monitoring of co-eluting product and product-related impurities during an intermediate purification or polishing step.

Acknowledgment

This work was supported by a grant from the Ministry of Science, Research, and the Arts of Baden-Württemberg, Germany (Az.: 33-7533-7-11.6-2).

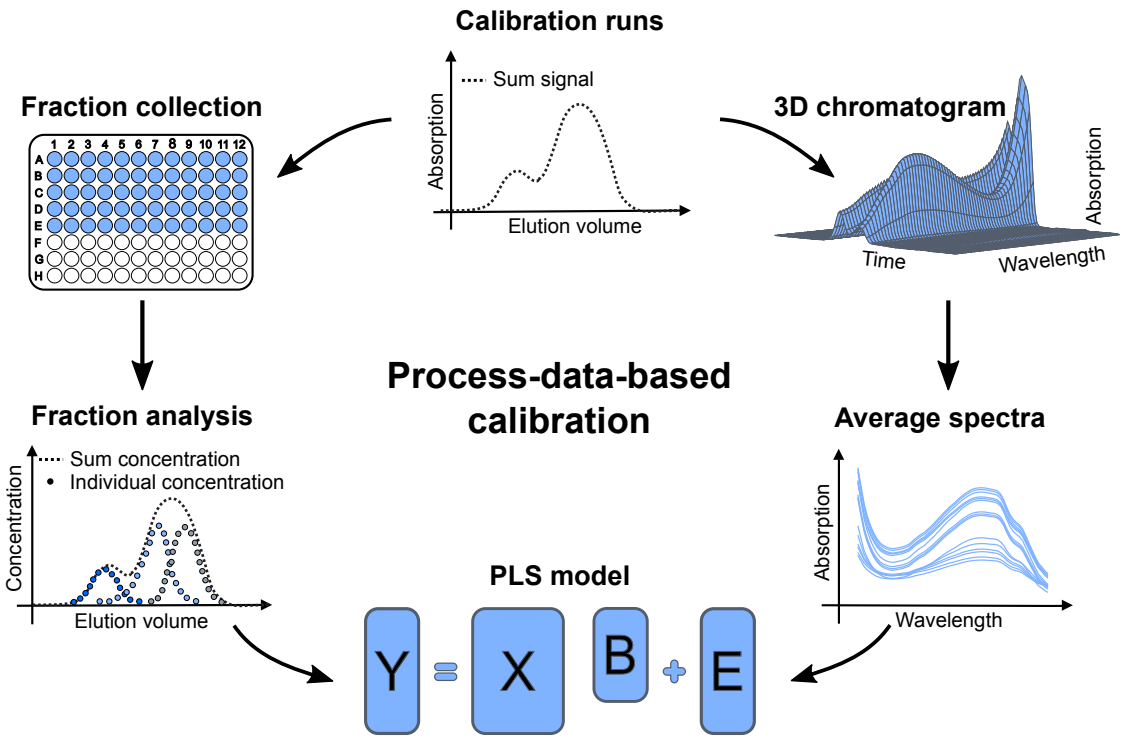
3 Advances in Inline Quantification of Co-Eluting Proteins in Chromatography: Process-Data-Based Model Calibration and Application Towards Real-Life Separation Issues

N. Brestrich[1], A. Sanden[1], A. Kraft[1], K. McCann[2], J. Bertolini[2], J. Hubbuch[1,3]

[1] : Institute of Process Engineering in Life Sciences, Section IV: Biomolecular Separation Engineering, Karlsruhe Institute of Technology, Engler-Bunte-Ring 3, 76131 Karlsruhe, Germany

[2] : CSL Behring Australia, 189-209 Camp Rd, Broadmeadows, VIC 3047

[3] : Corresponding author. *Email*: juergen.hubbuch@kit.edu



Abstract

Pooling decisions in preparative liquid chromatography for protein purification are usually based on univariate UV absorption measurements that are not able to differentiate between product and co-eluting contaminants. This can result in inconsistent pool purities or yields, if there is a batch-to-batch variability of the feedstock. To overcome this analytical bottleneck, a tool for selective inline quantification of co-eluting model proteins using mid-UV spectra and Partial Least Squares Regression (PLS) was presented in a previous study and applied for real-time pooling decisions. In this paper, a process-data-based method for the PLS model calibration will be introduced that allows the application of the tool towards chromatography steps of real-life processes. The process-data-based calibration method uses recorded inline spectra that are correlated with offline fraction analytics to calibrate PLS models. In order to generate average spectra from the inline data, a Visual Basic for Application macro was successfully developed. The process-data-based model calibration was established using a ternary model protein system. Afterwards it was successfully demonstrated in two case studies that the calibration method is applicable towards real-life separation issues. The calibrated PLS models allowed a successful quantification of the co-eluting species in a cation-exchange-based aggregate and fraction removal during the purification of monoclonal antibodies and of co-eluting serum proteins in an anion-exchange-based purification of Cohn supernatant I. Consequently, the presented process-data-based PLS model calibration in combination with the tool for selective inline quantification has a great potential for the monitoring of future chromatography steps and may contribute to manage batch-to-batch variability by real-time pooling decisions.

Keywords: Process Analytical Technology, Inline monitoring, Chemometrics, Partial Least Squares Regression, Selective protein quantification, Protein analytics, Bioprocess monitoring

3.1 Introduction

In the production of biopharmaceuticals, a consistent product quality is of high importance. A key issue compromising a consistent quality is batch-to-batch variability. In order to manage variability and to ensure a consistent product quality, the US Food and Drug Administration's (FDA's) Process Analytical Technology (PAT) initiative promotes the monitoring and control of critical quality attributes and process parameters as well as the use of modern, multivariate data acquisition tools (FDA, 2004). Especially in liquid chromatography, the main workhorse in biopharmaceutical protein purification, batch-to-batch variability represents a major issue. Typical reasons are aging of columns, variations in the composition of buffers and feed-concentrations, or lot-to-lot variability of resins, which may result in different retention times or peak shapes of the eluting proteins. This makes the decision on when to start or to stop the pooling of the target fraction a great challenge, especially in the case of co-eluting proteins.

Pooling decisions in preparative chromatography are usually based on univariate UV absorption measurements, as this approach is easy to implement (Follman and Fahrner,

2004; Zhou et al., 2007). However, univariate UV absorption measurements do not allow a differentiation between co-eluting proteins. This is especially disadvantageous during polishing steps that aim to remove product-related impurities. These impurities exhibit similar physicochemical properties as the product, which usually results in a poor resolution. In these cases, pooling based on univariate UV absorption results in inconsistent pool purities or yields, if there is a batch-to-batch variability of the feedstock.

In order to tackle this issue, various PAT tools for chromatography have been developed in the last years that enable the differentiation between product and co-eluting contaminants (Fahrner et al., 1998; Kaltenbrunner et al., 2012; Mendhe et al., 2015; Rathore et al., 2008a,b, 2009, 2010b; Westerberg et al., 2010). For instance, near real-time atline fluorescence spectroscopy has been applied to differentiate between misfolded and correctly folded proteins in order to make pooling decisions (Rathore et al., 2009). However, atline monitoring is time-consuming, can lead to infrequent or delayed data and thus late detection of production irregularities, and is consequently rather inefficient for process control (van de Merbel et al., 1996).

An approach allowing frequent data collection is online HPLC and UHPLC. PAT tools based on this approach have been applied for a selective quantification of the product and co-eluting contaminants and have successfully been used to make real-time pooling decisions (Fahrner et al., 1998; Kaltenbrunner et al., 2012; Mendhe et al., 2015; Rathore et al., 2008a,b, 2010b). In addition to the expensive equipment, the major disadvantage of online HPLC or UHPLC is the rather long response time (usually several minutes)(van de Merbel et al., 1996). Although efforts have been made to decrease assay times (Fahrner et al., 1998; Kaltenbrunner et al., 2012; Mendhe et al., 2015; Rathore et al., 2008a,b, 2010b), the response time may require a slowdown of the process flow rate (Kaltenbrunner et al., 2012; Mendhe et al., 2015; Rathore et al., 2008a,b, 2010b) and/or the application of mathematical models to predict the protein titers in the upcoming fraction (Rathore et al., 2008a,b, 2010b).

Next to online HPLC and UHPLC, PAT tools based on mathematical models in combination with univariate UV absorption or conductivity measurements have been proposed for inline monitoring and control of chromatography processes (Mendhe et al., 2015; Westerberg et al., 2010). Mendhe et al. could successfully establish a model based on the retention time of a characteristic peak eluting prior to the product peak. This model allowed product purity-based, real-time pooling decisions using univariate UV absorption measurements. The proposed PAT tool is easy to implement and does not require a modification of the chromatography system, but is not applicable for all separation issues as it requires a pre-peak as reporter. The PAT tool proposed by Westerberg et al. uses direct (conductivity of the loading buffer) and indirect (univariate UV absorption) measurements of critical process parameters that were identified during a chromatography model-based sensitivity analysis. However, the two presented case studies were based on simple separation issues. For real processes with many components, a chromatography model-based sensitivity analysis may not be easy to implement as it requires a prior model parameter determination for each component.

An alternative approach to differentiate between product and co-eluting contaminants that also allows a selective quantification of the species is spectral analysis in combination with Partial Least Squares Regression (PLS). Due to unequal number of the aromatic

amino acid residues in different protein species, protein mid-UV spectra vary and allow selective protein quantification in multicomponent mixtures using PLS. PLS has been successfully applied to calibrate models that correlate UV spectra with selective protein concentrations. The calibrated PLS models were successfully applied for a selective quantification by spectral analysis in various multicomponent protein mixtures. The approach has also been successfully used as analytics for high throughput chromatography experiments (Hansen et al., 2011, 2013). The applicability of UV spectra in combination with PLS for selective protein quantification in multicomponent mixtures has also been confirmed by other researchers (Kamga et al., 2013).

Recently, a powerful integration of software and chromatography hardware was introduced that enables the applicability of this approach for a selective inline quantification of co-eluting proteins in chromatography (Brestrich et al., 2014). This tool for selective inline quantification consists of a conventional Akta purifier liquid chromatography system, a diode array detector for the spectral measurement, a National Instruments data acquisitions device, and a software interface to Matlab®. The tool was rigorously compared with common offline analytics and found to deliver precise results for the retention times and peak flankes of a ternary model protein system. Further, the tool for selective inline quantification could be successfully applied for product purity-based, real-time pooling decisions.

So far, however, the tool was only applied for model proteins, as a Design of Experiments (DoE)-based PLS model calibration, requiring pure protein components, was used (Hansen et al., 2011, 2013). In complex applications with pharmaceutically relevant proteins, this would require an isolation of the target protein and all impurities. This is not only elaborate, but also impracticable regarding the material limitations during biopharmaceutical process development. Additionally, when used for inline applications, the DoE-based calibration method requires a time-consuming injection of each sample into the applied inline detector (Brestrich et al., 2014).

To overcome these limitations, a method was developed in this study that enables the PLS model calibration using recorded process data. The process data was generated by performing various chromatography runs with batch-to-batch variability and by recording the corresponding inline spectra. The selective protein concentrations of the elution peaks were determined using offline reference analytics based on collected fractions. In order to generate average spectra from the inline data and to correlate them with the reference analytics in a PLS model, a Visual Basic for Application (VBA) macro was developed.

This process-data-based calibration method was established using a ternary model protein system consisting of lysozyme (lys), cytochrome c (cyt c), and ribonuclease A (rib A). Afterwards, the method was applied to calibrate PLS models for two different separation issues to demonstrate that the proposed tool for selective inline quantification (Brestrich et al., 2014) can be applied for the monitoring and control of real-life processes. In a first case study, models for a selective quantification of monoclonal antibody (mAb), its aggregates (HMWs - High Molecular Weights) and fragments (LMWs - Low Molecular Weights) were calibrated. The quantified species were co-eluting during a cation exchange (CEX) step in mAb purification. The second examined separation issues was the purification of transferrin (trf) and IgG from Cohn supernatant I using flow-through mode anion-exchange (AEX) chromatography. A detailed description of the process can

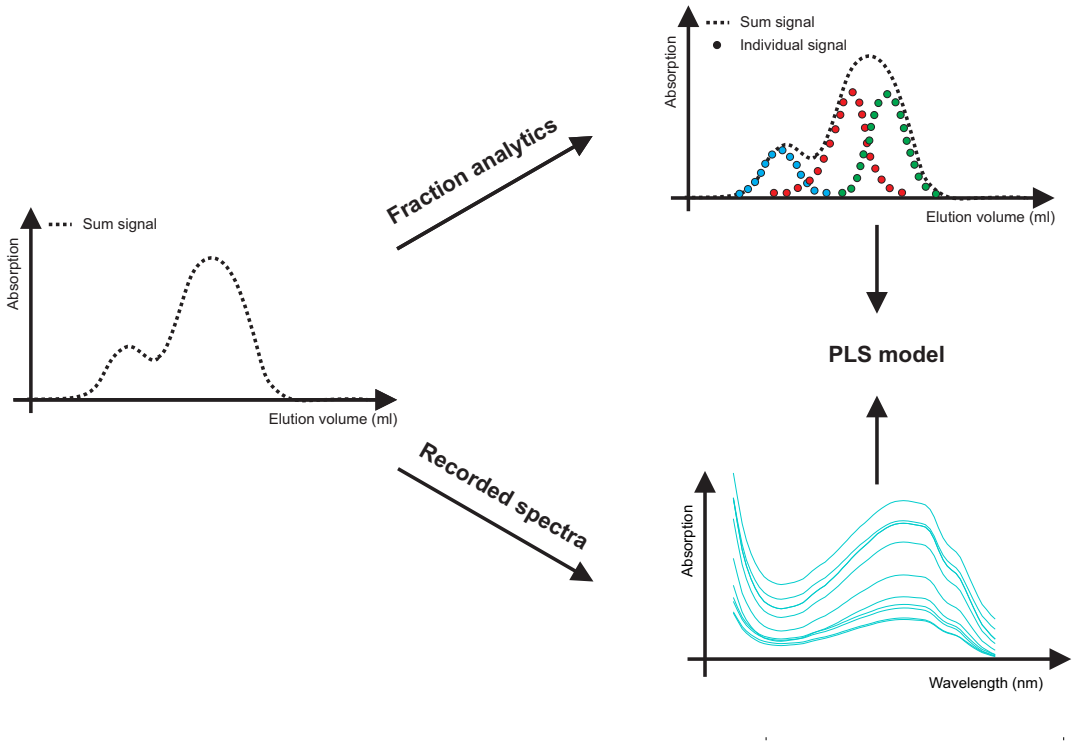


Figure 3.1: Process-data-based model calibration: Various chromatography runs with batch-to-batch variability were performed. The elution peaks were afterwards deconvoluted by reference analytics based on collected fractions. Additionally, the UV spectra of the runs or of collected fractions were recorded. For the PLS model calibration, the results of the reference analytics were finally correlated with the corresponding spectra.

be found in (McCann et al., 2005). Besides from the co-eluting target molecules trf and IgG, the AEX flow-through was containing other, lower concentrated UV-absorbing species.

3.2 Materials and Methods

For both the method establishment and the two case studies, the general experimental procedure for the process-data-based model calibration was the same and is displayed in Fig. 3.1: In a first step, various chromatography runs with batch-to-batch variability were performed. The elution peaks were afterwards deconvoluted by reference analytics based on collected fractions. Additionally, the UV spectra of the runs were recorded. For the method establishment and the first case study, this was realized using an inline diode array detector. The average spectra of each fraction were afterwards determined using a customized VBA macro. For the second case study, only the UV spectra of the fractions could be recorded as the study was performed at CSL Behring, Broadmeadows. The PLS models were finally generated by correlating the results of the fraction analytics with the corresponding spectra.

3.2.1 Method Establishment Using Model Proteins

Proteins and Buffers

Lys from hen egg white, rib A from bovine pancreas, and cyt c from equine heart were purchased from Sigma-Aldrich, St. Louis, USA. This system was chosen due to the stability of the proteins in aqueous solutions regarding the formation of multimers and due to their advantageous isoelectric point distribution, allowing separation at neutral pH by cation exchange chromatography. Cyt c was declared as "product" and the two other proteins as "contaminants". For the chromatography runs, the loading buffer was 20 mM sodium phosphate (pH 7). Elution was performed using a high-salt buffer consisting of 20 mM sodium phosphate as well as 500 mM sodium chloride (pH 7). For the reference analytics (analytical CEX chromatography), the loading buffer was 25 mM Tris (pH 8) and elution was performed using 25 mM Tris with 600 mM sodium chloride (pH 8). All buffer components were purchased from VWR, West Chester, USA. Buffers were 0.22 μm filtered and degassed by sonification before usage.

Chromatographic Instrumentation

The chromatography runs were performed using a specific experimental setup consisting of a conventional liquid chromatography system, a diode array detector, and a software interface to Matlab®. The liquid chromatography system was an Akta® purifier 10 equipped with pump P-900, sample pump P-960, UV monitor UV-900 (10 mm optical path length), conductivity monitor C-900, pH monitor pH-900, autosampler A-905, and fraction collector Frac-950 (all GE Healthcare, Chalfont St Giles, UK). The instrumentation was controlled with Unicorn® 5.31 software. In order to perform inline protein spectra measurements, a Dionex UltiMate® 3000 Diode Array Detector (DAD) equipped with a semi-preparative flow cell (0.4 mm optical path length) and operated with Chromeleon® 6.80 software (Thermo Fisher Scientific, Waltham, USA) was connected to the Akta® Purifier 10. The data acquisition of the DAD was triggered using a customized software interface. A detailed description can be found in (Brestrich et al., 2014).

Chromatography Runs

A HiTrap® 1 mL column prepacked with SP Sepharose FF (GE Healthcare) was first equilibrated with 5 column volumes (CVs) of loading buffer and then loaded with 0.4 mg of the "product" and 0.2 mg of each "contaminant". Afterwards, a wash of 5 CVs loading buffer was performed and the proteins were then eluted by various linear gradients from 0 to 500 mM sodium chloride. The gradient length was 1, 3, 5, and 7 CVs. The variable gradient length was performed to imitate batch-to-batch variability for a model protein study. Variable gradients lead to different resolutions and thus mixing ratios of the proteins as well as to different peak heights and thus maximal concentrations of the species. At the start of the gradient, the DAD was triggered to start data acquisition and recorded protein spectra in the band of 240-300 nm with 1 nm resolution. After the gradient, the column was regenerated for 5 CVs using elution buffer. The flow rate was 0.2 mL/min for all steps and experiments. The elution peaks were collected in deep well

plates (VWR) in 150 μl fractions. Afterwards, they were transferred to half area plates (Greiner BioOne, Kremsmünster, Austria) for the reference analytics. This procedure allowed an appropriate fraction collection as well as the required sample height in the wells for the reference analytics. As the reference analytics was based on collected fractions, protein concentrations that are averaged over the fraction size were determined. In contrast to that, the DAD recorded a continuous inline signal (3d field of the DAD). Therefore, spectra averaged over the fraction size were calculated from the recorded data before correlating them with the reference analytics by PLS. In order to generate these average spectra, a VBA macro was developed. The macro uses classes of the Chromeleon Software Developer Kit (Thermo Fisher Scientific) to export the recorded spectra of each fraction to Matlab®. The average spectrum of each fraction was then determined by Matlab®.

Reference Analytics

As reference analytics, analytical CEX chromatography was performed with the collected fractions, using a Dionex UltiMate® 3000 liquid chromatography system (Thermo Fisher Scientific). The system was composed of a HPG-3400RS pump, a WPS-3000TFC-analytical autosampler, a TCC-3000RS column thermostat, and a DAD3000RS detector. For each fraction, a Proswift SCX-1S 4.6 x 50 mm column (Thermo Fisher Scientific) was first equilibrated for 2.7 min with loading buffer, loaded with 20 μl sample, washed for 1 min with loading buffer and then eluted by multiple salt steps: 20 % elution buffer for 0.9 min, 75 % elution buffer for 0.7 min, and 100 % elution buffer for 6.4 min. The flow rate was 0.5 mL/min for all steps.

3.2.2 First Case Study: CEX Step for mAb Purification

Process Material and Buffers

mAb Protein A eluate, donated by Rentschler (Laupheim, Germany), was 0.22 μm filtered. The mAb concentration based on a provided extinction coefficient was determined using a NanoDrop 2000 UV-Vis Spectrophotometer (Thermo Fisher Scientific). For the chromatography runs, the loading buffer was 10 mM sodium citrate (pH 4.1). Elution and regeneration was performed using 10 mM sodium citrate with 1000 mM sodium chloride (pH 4.1). For the reference analytics (size exclusion chromatography), the buffer was 200 mM potassium phosphate with 250 mM potassium chloride (pH 7). All buffer components were purchased from VWR, West Chester, USA. Buffers were 0.22 μm filtered and degassed by sonification before usage.

Chromatographic Instrumentation

For the chromatographic instrumentation used, cf. section 3.2.1.

Chromatography Runs

A MediaScout®MiniChrom 1 mL column prepacked with POROS® HS 50 (Carlsbad, CA, USA) was purchased from Atoll, Weingarten, Germany. The column was equilibrated

with 5 CVs of loading buffer and then loaded with filtered protein A pool. Afterwards, a wash of 5 CVs loading buffer was performed and the elution was carried out using various linear gradients from 0 to 70 % elution buffer. Gradient length, imitating the batch-to-batch variability, was 10, 20, and 30 CVs. At the start of the gradient, the DAD was triggered to start data acquisition and recorded protein spectra in the band of 240-410 nm with 1 nm resolution. Finally, the column was regenerated for 5 CVs using 100 % elution buffer and was then stripped with 0.5 M sodium hydroxide over night. The flow rate was 0.2 mL/min for all steps and experiments. The elution peaks were collected in 150 μl fractions as described in section 3.2.1. Further, the average spectra were determined as explained in 3.2.1.

Reference Analytics

As reference analytics, the total mAb concentration of each fraction was first of all determined using the NanoDrop 2000. Afterwards, size exclusion chromatography (SEC) was performed with the collected fractions to determine the mass fractions of mAb, HMWs, and LMW using the same Dionex UltiMate® 3000 liquid chromatography system as described in section 3.2.1. For each fraction, 20 μl sample were loaded on a 4.6 x 150 mm TSKgel SuperSW mAb HTP column (Tosoh, Tokio, Japan). In order to reduce the assay time, interlaced injections were used (Diederich et al., 2011).

3.2.3 Second Case Study: AEX Step in Protein Purification from Cohn Supernatant I

Process Material and Buffers

Lipid- and euglobulin-depleted Cohn supernatant I was taken directly from the full scale process as an intermediate sample and was 0.22 μm filtered. The major components were albumin (approx. 73 %), IgG (approx. 19 %), and trf (approx. 5 %). Other minor components were IgA and IgM. For the chromatography runs, the loading buffer used was sodium acetate at pH 5.2. Elution was performed using sodium acetate at pH 4.5. All buffers were 0.22 μm filtered before usage.

Chromatographic Instrumentation

The chromatography runs were performed using an Akta explorer 100 system equipped with pump P-900, sample pump P-960, UV monitor UV-900 (10 mm optical path length), conductivity monitor C-900, pH monitor pH-900, and fraction collector Frac-950 (all GE Healthcare) . The instrumentation was controlled with Unicorn® 5.31 software.

Chromatography Runs

An XK16/40 column (GE Healthcare) was packed to a height of 17.5 cm using DEAE Sepharose FF resin by GE Healthcare, and the quality of the packed column was assessed by measuring HETP. A packed column was rejected if the HETP came out to be > 0.06 cm. The column was equilibrated using loading buffer. A sample of lipid- and euglobulin-depleted supernatant I was loaded onto the column at 65 g/L. Loosely bound

protein was flushed off the column using loading buffer. The flow-through was collected in fractions of 5 mL. The fractions were then analyzed by measuring their individual absorbance spectra between 240 nm and 300 nm using a Cary 50 Spectrophotometer by Agilent Technologies (Santa Clara, California, USA) and a quartz glass cuvette. Finally, bound protein was eluted using the corresponding buffer. The flow rate was 3.33 mL/min for all steps and experiments.

Reference Analytics

The levels of plasma proteins were measured by nephelometry using an Immage 800 Immunochemistry system by Beckman-Coulter (Pasadena, California, USA). The Immage system was calibrated using standard samples available from Beckman-Coulter.

3.2.4 Partial Least Squares Regression

In order to correlate the average UV spectra with the selective protein concentrations Partial Least Squares Regression (PLS) technique was applied (Eriksson et al., 2006; Höskuldsson, 1988; Martens and Næs, 1989). PLS aims to reduce data by only applying significant variance in data for correlating several input variables with variables of interest. To achieve a data reduction, variables containing similar information are summarized in so-called latent variables (LVs). Thus, the number of LVs expresses a measure of data reduction, where a low number of LVs corresponds to a high reduction. The information content included in each LV is high for the first LV and decreases for the following LVs.

This means that the variance in the data can be explained by only a few LVs, leaving out higher LVs that mainly contain detector noise. The higher the number of LVs used in a PLS model, the more precise is the established model for chromatography runs used in the calibration. However, the use of too many latent variables can result in overfitting. Therefore, the thorough examination of the number of LVs used in a PLS model is crucial for its prediction precision. In order to determine a reasonable number of LVs, the root mean square error for the prediction (RMSEP) in dependence on the number of LVs used in the calibration was calculated for a validation run (Eriksson et al., 2006).

The PLS model calibration based on the obtained average spectra and concentrations from the corresponding reference analytics was performed using SIMCA (Umetrics, Umeå, Sweden). The software was used because of the easiness of data preprocessing and visualization. SIMCA applies the NIPALS-algorithm for PLS. Before performing PLS, spectra were preprocessed by mean centering using SIMCA. However, it is also possible to perform the PLS regression using Matlab® or other available software. After their calibration, the PLS models were applied to calculate protein concentrations from inline spectra.

3.3 Results and Discussion

An application of the tool for selective inline quantification (Brestrich et al., 2014) towards real-life separation issues demands for a fast and reliable calibration of PLS models. The calibration method should not require pure protein components and thus an elaborate

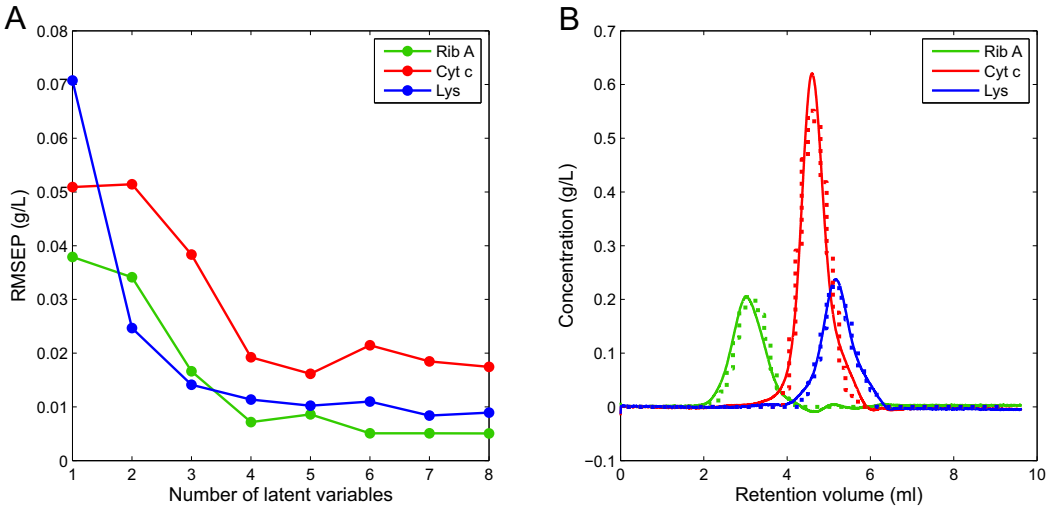


Figure 3.2: A: Calculated root mean square error for the prediction (RMSEP) of rib A, cyt c, and lys in the validation run with respect to the number of latent variables used for the PLS model calibration. B: Resulting chromatogram of the inline peak deconvolution for the validation run (solid lines) and offline reference analytics displayed as stairs (dashed lines).

isolation of the target protein and all impurities. This work demonstrates that process data can be successfully applied to calibrate PLS models for the inline monitoring of co-eluting proteins in chromatography. In order to generate average spectra of the recorded process data, a VBA macro was successfully developed. The methodology was established using model proteins. Afterwards, it was demonstrated in two case studies that the process-data-based calibration is applicable towards real-life separation issues.

3.3.1 Method Establishment Using Model Proteins

In order to imitate batch-to-batch variability for a model protein study, the gradient length was varied. This resulted in different resolutions and thus mixing ratios of the model proteins as well as different peak heights and thus maximal concentrations of the species. For the PLS model calibration, the results of the fraction analytics by analytical CEX chromatography for the 1, 3, and 7 CV gradient run were correlated with the corresponding average spectra using PLS. Eight PLS models with one to eight LVs were calibrated. The PLS models were afterwards applied to predict the selective protein concentrations of the run with a 5 CV gradient length from the average spectra.

The RMSEP between the predicted concentrations and the results from the reference analytics with respect to the number of LVs used in the calibration is displayed in Figure 3.2 A. As to be expected, the RMSEP was first decreasing with increasing number of LVs for all proteins and then slightly increasing at certain number of LVs when overfitting was reached. The best combination of low RMSEP and low number of LVs was given at 6 LVs for rib A, at 5 LVs cyt c, and at 7 LVs for lys. As this combination was different for each protein, an individual PLS model for each protein using this determined optimal number of LVs was calibrated using the 1, 3, and 7 CV gradient runs. The RMSEP of the final model was 0.008 g/L for rib A, 0.016 g/L for cyt c, and 0.005 g/L for lys.

The calibrated PLS models were afterwards applied for a peak deconvolution of the

validation run with a 5 CV gradient using inline spectra. This means that in contrast to the PLS model calibration, where average spectra of each fraction were applied, the continuous inline signal (3d field of the DAD) was used. The application of PLS models generated by spectra averaged over fraction size to evaluate inline spectra was possible due to high resolution of the collected fractions. The resulting chromatogram of the inline peak deconvolution is displayed in Figure 3.2 B (solid lines) and compared with the offline fraction analytics (dashed lines). A good agreement between the prediction by the PLS model and the reference analytics for both the peak maxima and peak flanks was observed. This clearly shows the the precision of the models generated by the process-data-based calibration and the applicability of this calibration method towards inline monitoring of real-life separation issues.

Compared to the prediction of the PLS model calibrated by the DoE-method (Brestrich et al., 2014), where calibration samples from pure protein stock solutions are mixed according to a DoE, a higher precision at the peak maxima was achieved. This can be explained by the fact that using a DoE-based calibration, all possible mixing rations and concentration levels of the model proteins are accounted for using as few calibration samples as possible. This results in PLS models that are valid for a large concentration level and mixing ratio space, but which are not specialized for the mixing ratios and concentration levels occurring in the chromatography run to be monitored. In contrast to that, PLS models calibrated with process data exhibit a high precision within the process window they were calibrated for but decrease in their prediction precision outside of this window.

One possibility to display this process window is plotting the scores of the calibration samples on all LVs used in a model. Exemplarily, Figure 3.3 A displays the scores of the calibration samples (1, 3, and 7 CV gradient) for the cyt c model on the first and second LV, which together explained a major part (88 percent) of the variance in the cyt c concentration. Additionally, the scores of the 5 CV gradient validation run are displayed. As they are located within the process window, the prediction precision was high for cyt c in the validation run (cf. Fig. 3.2 B). Obviously, this also has to be examined for the higher LVs (data not shown). Furthermore, it can be seen in Figure 3.3 A that a model calibrated using the 1, 5, and 7 CV gradient run can precisely predict the 3 CV gradient run (cf. Fig. 3.3 B). In contrast to that, it is neither possible to apply the 1 CV nor the 7 CV gradient run as a validation run, as they the would then lie outside of the process window spread out by the calibration samples (cf. Fig. 3.3 A). The prediction of one of these two runs by using the other ones for the model calibration equals an extrapolation of the corresponding PLS model for some concentrations and mixing ratios. This leads to a lower prediction precision (cf. Fig. 3.3 C-D). Consequently, all batch-to-batch variability that might occur in a process should be accounted for in the corresponding models. The required data for this purpose might be obtained from process development and robustness studies in order to minimize the experimental effort.

3.3.2 First Case Study: CEX Step for mAb Purification

In mAb platform processes, the protein A capture and low pH virus inactivation is usually followed by two polishing chromatography steps aiming to reduce remaining host cell

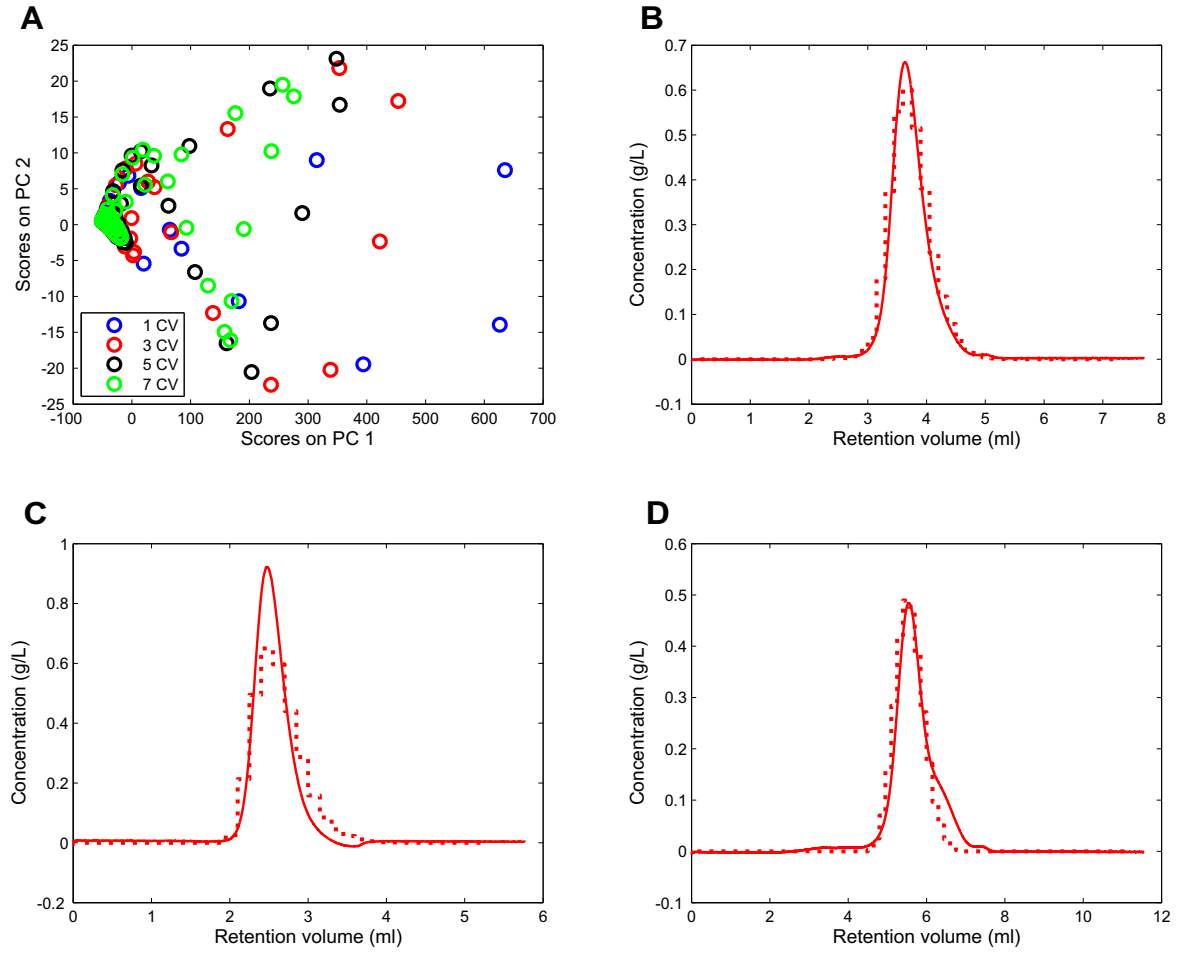


Figure 3.3: A: Scores on the first and second LV for the cyt c model. B-D: Resulting chromatogram of the inline peak deconvolution of cyt c (solid lines) for a 3 CV (B), 1 CV (C), or 7 CV (D) gradient length validation run and offline reference analytics displayed as stairs (dashed lines). The remaining samples were applied as calibration samples.

proteins, HMWs, LMWs, DNA and leached Protein A (Shukla et al., 2007). The CEX step typically aims to reduce HMWs and LMWs. These two species are exhibiting similar physico-chemical properties as the product, which often results in a poor resolution. As the pooling in this purification step is typically based on univariate UV absorption, batch-to-batch variability results in inconsistent pool purities. The goal of this study was to imitate batch-to-batch variability by different gradient lengths and to calibrate PLS models for the inline quantification of mAb, HMWs, and LMWs using this process data. Based on the inline quantification by the PLS models, product-purity based, real-time pooling decisions would be possible.

For the model calibration, the results of the fraction analytics by SEC for the 10 and 30 CV gradient run were correlated with the corresponding average spectra using PLS. Ten PLS models with one to ten LVs were calibrated. The PLS models were afterwards applied to predict the selective concentrations of monomer, HMW1, HMW2, and LMW in the run with a 20 CV gradient length from the average spectra. The RMSEP between prediction and reference analytics with respect to the number of LVs used in the calibration is displayed in Figure 3.4. The best combination of low RMSEP and low

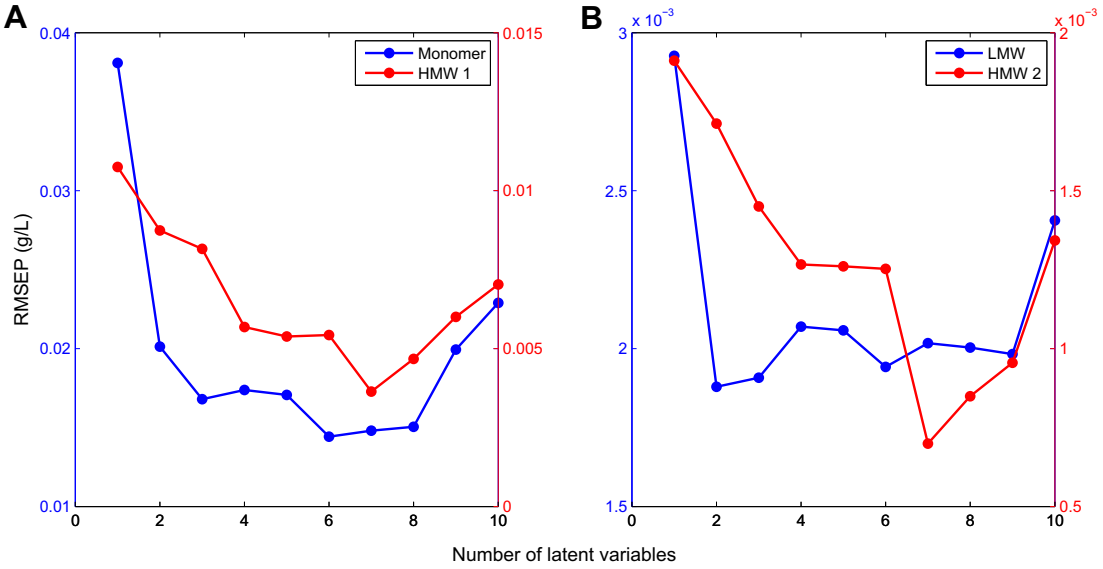


Figure 3.4: A: Calculated root mean square error for the prediction (RMSEP) of monomer and HMW 1 in the validation run with respect to the number of latent variables used for the PLS model calibration. B: RMSEP of LMW and HMW 2 in the validation run with respect to the number of latent variables used for the PLS model calibration.

number of LVs was given at 6 LVs for monomer, at 7 LVs for HMW 1, at 2 LVs for LMW, and at 7 LVs for HMW 2. As for the model proteins, individual PLS models using this determined optimal number of LVs were calibrated using the 10 and 30 CV gradient run. The RMSEP of the final model was 0.014 g/L for monomer, 0.0036 g/L for HMW 1, 0.0019 g/L for LMW, and 0.0007 g/L for HMW 2.

The calibrated PLS models were applied for an inline peak deconvolution of the validation run with a 20 CV gradient. The resulting chromatograms are displayed in Figure 3.5 (blue lines) and compared with the offline fraction analytics (red stair functions). A good agreement between the prediction of the PLS models and the reference analytics was observed for all four monitored species. The agreement between prediction and reference for the aggregate species (HMW 1 and 2) was found to be slightly lower. This might be explained with a lower precision of the reference analytics for theses species, as no complete baseline separation could be achieved during the SEC. But the results are acceptable regarding the fact that the overall trend, especially the peak flanks, could be predicted.

Based on these findings, the question on the origin of spectral differences between monomers and aggregates arises. An effect based on molecular structure might be conformational changes during the aggregation process which lead to changes in the microenvironment around the aromatic amino acid residues (Donovan, 1969; Ragone et al., 1984; Kueltzo et al., 2003). More recently, Thakkar et al. (Thakkar et al., 2012) found out that even protein-protein interactions in highly concentrated solutions can lead to changes in the microenvironment around the aromatic amino acid residues and thus to changes in UV absorbance. With larger aggregates these structural effects might be overruled by scattering effects. Scattering of incident light on solute protein multimers prevents the light from reaching the detector and results in an artificial increment of the absorbance values (Jiskoot and Crommelin, 2005). Usually, the scattering artefact is corrected. How-

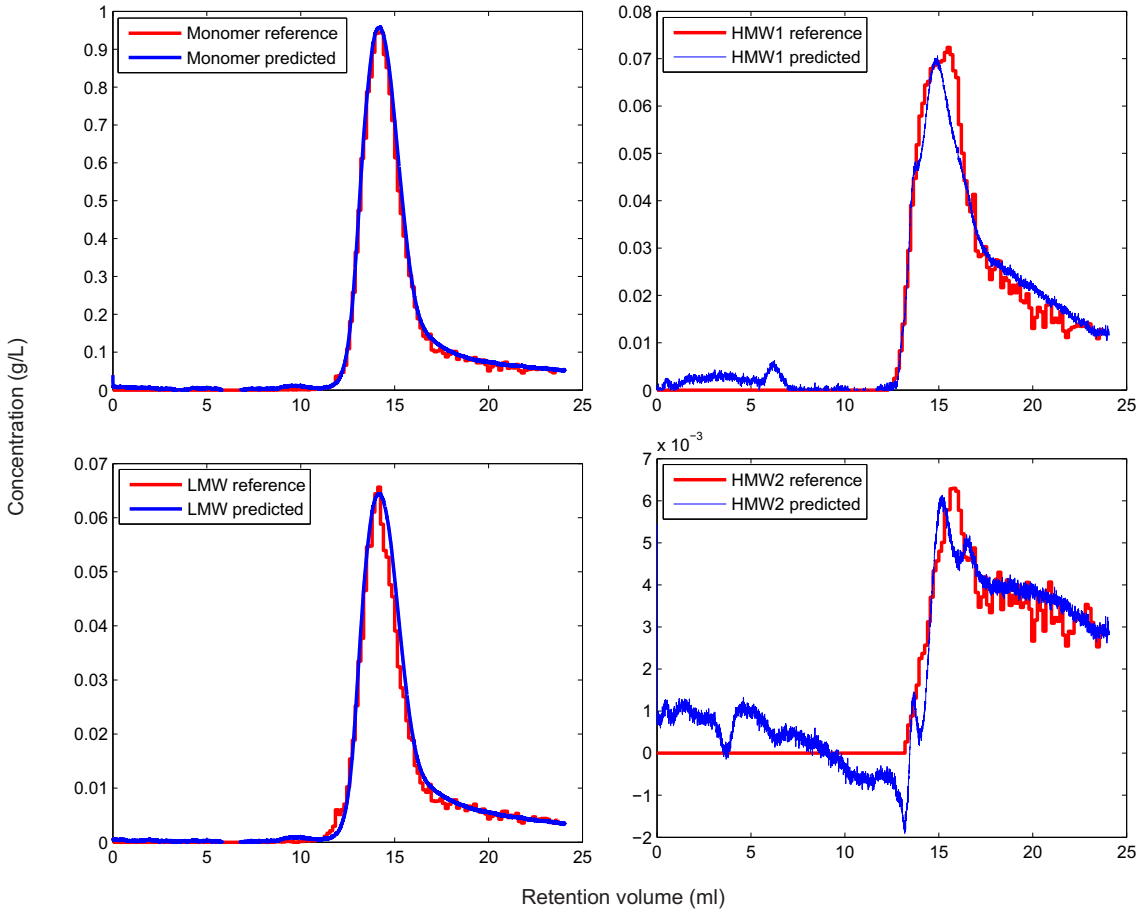


Figure 3.5: Resulting chromatograms of the inline peak deconvolution for monomer, LMW, HMW1, and HMW2 in the validation run as well as the resulting offline reference analytics. The results of the reference analytics (SEC chromatography) based on collected fractions are displayed as stairs.

ever in this study it was desired. The spectral differences between LMWs and the other species can be ascribed to differences in the number and ratio of the aromatic amino acid residues.

3.3.3 Second Case study: AEX Step in Protein Purification from Cohn Supernatant I

In contrast to the first case study and the method establishment using model proteins, the spectra of the second case study were obtained from an offline spectrophotometer as the study was performed at CSL Behring, Broadmeadows. For this study, the mid-UV spectra of collected fractions were recorded. Consequently, the calibrated model cannot be used for an inline monitoring as it is specific for the applied detector. However, process data was used for the PLS model calibration and the developed procedure to generate average spectra from inline data could be easily transferred to this study.

At CSL Behring, DEAE Seaphorse FF chromatography is used to purify albumin, IgG, and trf from Cohn supernatant I. Therefore, supernatant I is loaded at pH 5.2, allowing albumin (pI 4.8) to bind to the resin while IgG (pI 4.35-9.95) passes the column

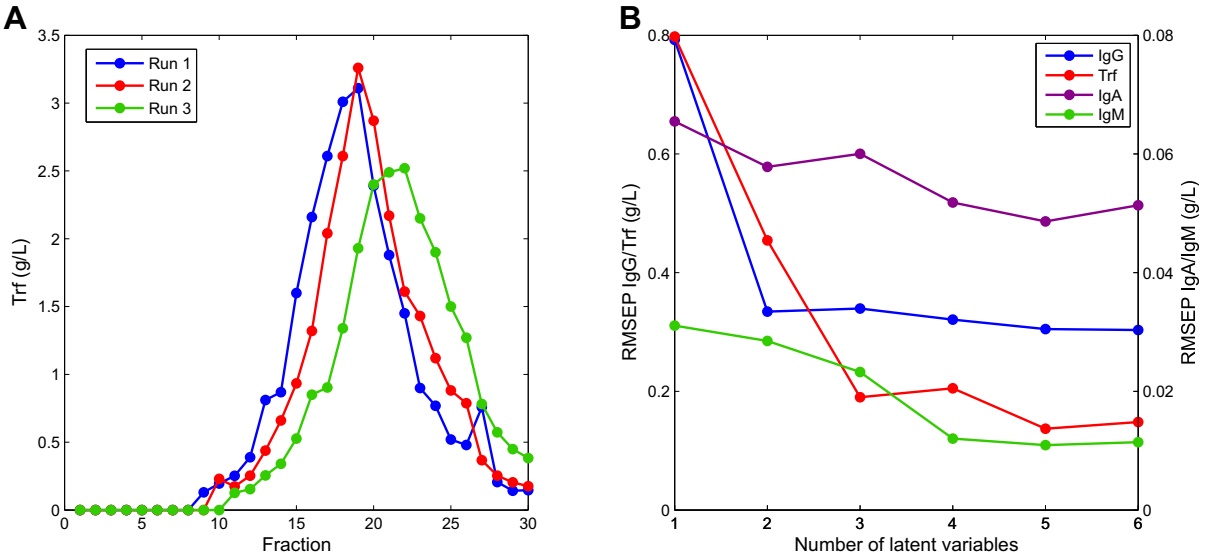


Figure 3.6: A: Batch-to-batch variability in the trf elution profiles of the three runs with a standard resin load of 65 g/L. B: Calculated root mean square error for the prediction (RMSEP) of IgG, trf, IgA, and IgM in the validation run with a resin load of 65 g/L.

unretained (Prin et al., 1995). Other proteins in Cohn supernatant I such as trf (pI 5.0-6.0), IgA (pI 4.0-7.1) and IgM (pI 4.0-9.1) are partially retained and therefore present in both the IgG- and the albumin-rich fraction (Prin et al., 1995; Morgan, 1981). The DEAE flow-through is pooled in two fractions in order to gain both trf and IgG. However, the resolution of both proteins is strongly dependent on the sample loading, protein concentration in supernatant I, as well as on the conductivity of supernatant I and of the equilibration buffer (McCann et al., 2005). Consequently, the main goal for the PLS model calibration was the deconvolution of co-eluting trf and IgG in the flow-through in order to enable real-time pooling of these two target proteins. Nevertheless, IgM and IgA, which are separated at a later point of the process, were included in the model as well.

The standard process exhibited a batch-to-batch variability as displayed by the trf elution profiles in Figure 3.6 A. Therefore, two of these runs were applied for the model calibration and a third run was used as validation run. Six PLS models with one to six LVs were calibrated and then afterwards applied to predict the selective protein concentrations of IgG, trf, IgM, and IgA of the validation run. Figure 3.6 B shows the determined RMSEP between prediction and reference analytics with respect to the number of LVs used in the calibration. The best combination of low RMSEP and low number of LVs was given at 5 LVs for all proteins examined. Consequently only one PLS model with 5 LVs was applied. The RMSEP of this model was 0.305 g/L for IgG, 0.137 g/L for trf, 0.011 g/L for IgM, and 0.049 g/L for IgA.

The calibrated PLS model was used to predict the protein concentrations of the validation run from the average spectra of the collected fractions. The results for all four proteins examined are displayed in Figure 3.7 and compared with the corresponding reference analytics. A good prediction of the elution profiles for the higher concentrated target proteins IgG and trf could be achieved. Additionally, a good agreement between prediction and reference of the lower concentrated IgM was observed. Solely for IgA differences

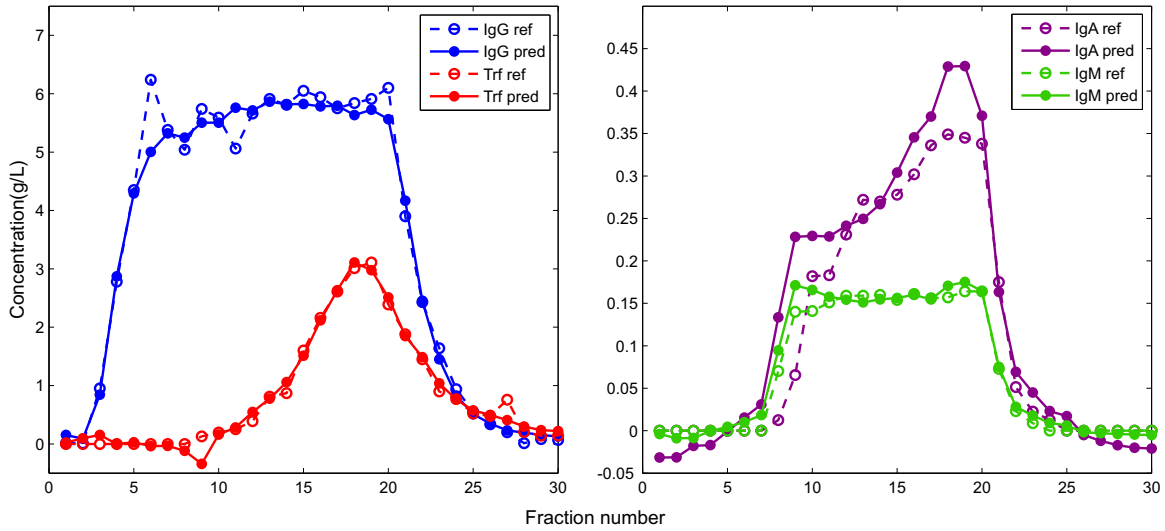


Figure 3.7: PLS model prediction based on collected fractions and reference analytics for the validation run with a sample loading of 65 g/L.

in the total concentrations were found indicating a lower precision of the model for this protein. But the overall trend of the elution profile could be predicted. The cause was however not further pursued because the goal of the study was the accurate prediction of the IgG and trf as this has the greatest impact on process efficiency and IgG recovery. IgA is a minor constituent of the crude IgG and does not have a significant impact on the downstream processing of the IgG fraction. However, if the exact determination of IgA was important for the process, more calibration runs would be needed in order to calibrate a model that can predict IgA more precisely.

3.4 Conclusion and Outlook

The use of inline spectra in combination with PLS for a real-time monitoring of co-eluting proteins in chromatography demands for a fast and reliable calibration method of PLS models, which does not require pure proteins. In this paper, a process-data-based PLS model calibration was successfully developed. The process data was generated by performing various chromatography runs with batch-to-batch variability and recording the corresponding inline spectra. The elution peaks were deconvoluted by offline reference analytics based on collected fractions. In order to generate average spectra from the recorded inline data, a VBA Macro was developed. The average spectra and the reference analytics were correlated in various PLS models and successfully applied for inline peak deconvolutions of co-eluting proteins in validation runs. The process-data-based calibration method was found to deliver precise results within the calibrated process window. As to be expected, the prediction precision outside this window decreased. Consequently, the final variability of a process should be accounted for in a corresponding PLS model. Suitable data for this could be obtained from process robustness studies. The process-data-based calibration method was established using a ternary model protein system and could successfully be transferred on real-life separation issues. Consequently, the presented calibration method together with the proposed tool for selective inline quan-

tification (Brestrich et al., 2014) has a great potential for the real-time monitoring of future chromatography steps. This approach offers a tool to manage batch-to-batch variability by real-time pooling and to enable consistent pool purities, which is one of the key requirements within the FDA's PAT concept. Future work will focus on monitoring of HCPs and on a qualitative method for inline peak deconvolution, not requiring a calibration.

Acknowledgment

This work was supported by a grant from the Ministry of Science, Research, and the Arts of Baden-Württemberg, Germany (Az.: 33-7533-7-11.6-2). We kindly thank Rentschler (Laupheim, Germany) for donating Protein A pool.

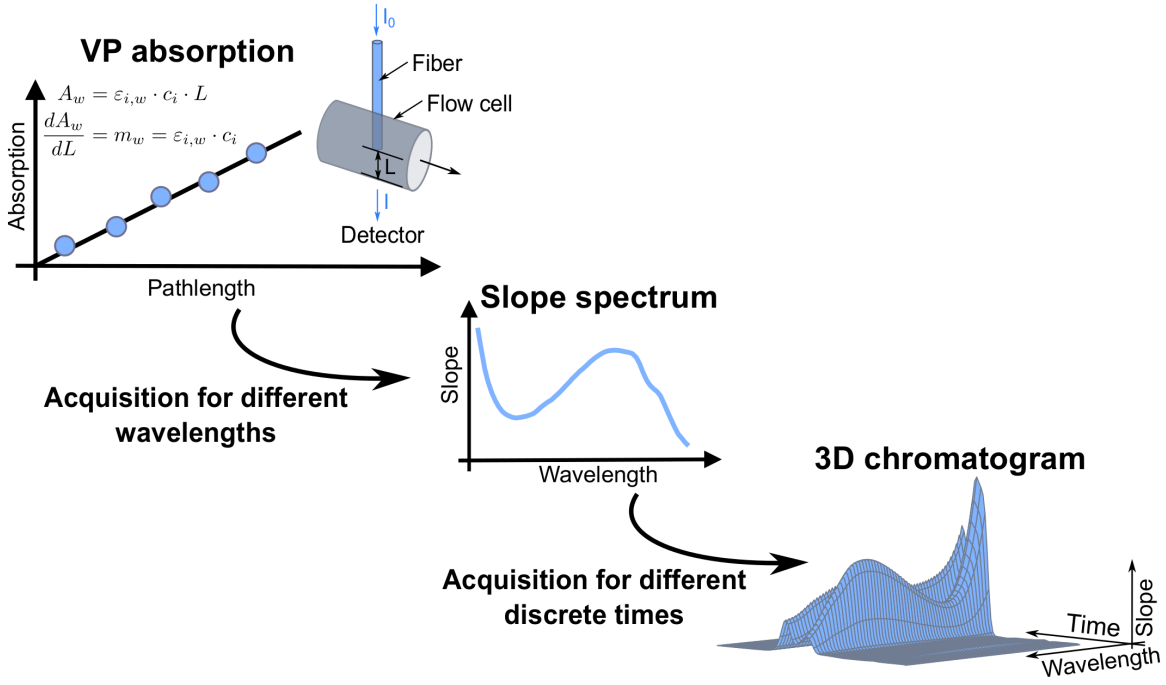
4 Selective Protein Quantification for Preparative Chromatography using Variable Pathlength UV/Vis Spectroscopy

N. Brestrich[1,*], M. Rüdt[1,*], D. Büchler[1], J. Hubbuch[1,2]

[1] : Institute of Process Engineering in Life Sciences, Section IV: Biomolecular Separation Engineering, Karlsruhe Institute of Technology, Engler-Bunte-Ring 3, 76131 Karlsruhe, Germany

[*] : contributed equally

[2] : Corresponding author. *Email*: juergen.hubbuch@kit.edu



Abstract

In preparative chromatography, high dynamic ranges of protein concentrations as well as co-elution of product and impurities is common. Despite being the standard in biopharmaceutical production, monitoring of preparative chromatography is generally limited to surrogate signals, e.g. UV absorption at 280 nm. In this paper, variable pathlength (VP) spectroscopy in conjunction with Partial Least Squares regression (PLS) was used to monitor preparative chromatography. While VP spectroscopy enabled the acquisition of spectral data for a broad concentration range, PLS modelling allowed for the differentiation between the protein species. The approach was first implemented for monitoring the separation of lysozyme from cytochrome c at a loading density of 92 g/l. The same method was then applied to the polishing step of a monoclonal antibody (mAb) at 40 g/l loading density. For the PLS model prediction of the mAb monomer and the high molecular weight variants (HMWs), the root mean square error (RMSE) was 1.15 g/l and 0.48 g/l respectively. Consequently, the presented VP spectroscopy in conjunction with PLS modelling has a great potential for the monitoring of future chromatography steps at large scale.

Keywords: Preparative Chromatography, Process Analytical Technology, Inline Monitoring, Variable Pathlength Spectroscopy, Selective Protein Quantification, Partial Least Squares Regression

4.1 Introduction

In the current purification process of biopharmaceuticals, preparative liquid chromatography is key for separating the target product from media components, DNA, host cell proteins, and product related impurities (Carta and Jungbauer, 2010). The method is used because of its high separative power while minimizing product loss. Despite being the standard, monitoring of preparative chromatography is generally limited to surrogate signals, e.g. UV absorption at 280 nm. The Process Analytical Technology (PAT) initiative of the US Food and Drug Adminstration (FDA) however promotes the acquisition of critical quality attributes in real-time (FDA, 2004). Especially for chromatography, PAT is still an active field of research (Rathore et al., 2010a; Rathore and Kapoor, 2015).

Preparative chromatographic processes are generally run at high loading densities to realize efficient processes (Carta and Jungbauer, 2010). High loading densities subsequently lead to a broad range of protein concentrations eluting from the column. Thus, monitoring techniques must feature a broad dynamic range to capture the peak concentration as well as lower concentrated impurities in the peak flanks. Furthermore, due to the limited resolution of preparative chromatographic columns, baseline separation between product and impurities is rarely achieved. PAT techniques should therefore selectively quantify the product and the main impurities. Due to the high non-linearity of chromatographic processes, the typical decision time within chromatography is limited (Rathore et al., 2010b).

In literature, a number of different approaches have been proposed for real-time monitoring of chromatography. On-line monitoring and control of preparative chromatography

has been realized by automated sampling and subsequent analysis by high performance liquid chromatography (Fahrner et al., 1998; Kaltenbrunner et al., 2012; Rathore et al., 2010b). The broad concentration ranges in preparative liquid chromatography can be managed easily by varying injection volumes. Disadvantages comprise time delay due to sample handling and assay times as well as the risk of contamination.

Fourier Transform Infrared (FTIR) spectroscopy has been applied previously for at-line monitoring of downstream processes (Capito et al., 2015b). Samples were taken at various stages of downstream processing, optionally dried and subsequently measured by FTIR. The method allowed to quantify product, high molecular weight variants (HMWs), and host cell proteins over a broad range of concentrations. Despite showing promising results, at-line measurements bear the risk of introducing operator errors and may increase the risk of contamination. Furthermore, the delay due to sample handling and measuring times may be too long for the typical decision times in chromatography.

Previously, UV/Vis spectroscopy has been proposed as a method for real-time monitoring of chromatographic processes (Brestrich et al., 2014, 2015). Its usefulness was shown for multicomponent mixtures of model proteins and for real-life separation problems. Based on UV/Vis spectral data and Partial Least Squares regression (PLS) modelling, real-time process decisions were taken such as the beginning and end of product pooling. Other applications include high throughput process development (Hansen et al., 2011; Baumann et al., 2016) and coupling with chromatography modelling (Brestrich et al., 2016). While being very fast and accurate, previous applications of real-time monitoring using UV/Vis spectroscopy only took process decisions for separation problems in diluted conditions. The problem of high dynamic concentration ranges occurring in preparative liquid chromatography was not addressed.

To increase the dynamic range of spectroscopic acquisitions, measurement cells have been designed which allow to continuously change the optical pathlength to achieve ideal sensitivity for virtually any analyte concentration (Flowers and Callender, 1996). The methodology has been further developed by a commercial vendor for protein related applications. With the commercialized product, spectra of proteins were acquired at a broad range of concentrations (Thakkar et al., 2012). Samples were studied for spectral effects of protein-protein interactions of UV/Vis with protein concentrations up to 250 g/l. Recently, an additional product line has been launched for inline variable pathlength (VP) measurements.

In this publication, we demonstrate real-time monitoring of preparative chromatography runs with UV/Vis VP spectroscopy in conjunction with PLS modelling. While VP spectroscopy allowed to monitor chromatographic processes at almost arbitrary protein concentrations, PLS models selectively quantified multiple co-eluting species from spectral data. The approach was first implemented for the separation of a mixture of the model proteins lysozyme (lys) and cytochrome c (cyt c) at high loading densities. Former was considered the product, latter the contaminant with high respectively low concentration in the feed. Subsequently, a preparative polishing step of a monoclonal antibody (mAb) and its HMWs was monitored.

4.2 Materials and Methods

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

4.2.1 Chromatography Instrumentation

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

In order to perform VP spectroscopy, different pathlengths L were set by moving the fiber of the FlowVPE within the flow path as illustrated by Figure 4.1. That way, VP absorption measurements were realized. For each measurement, the linear range of the

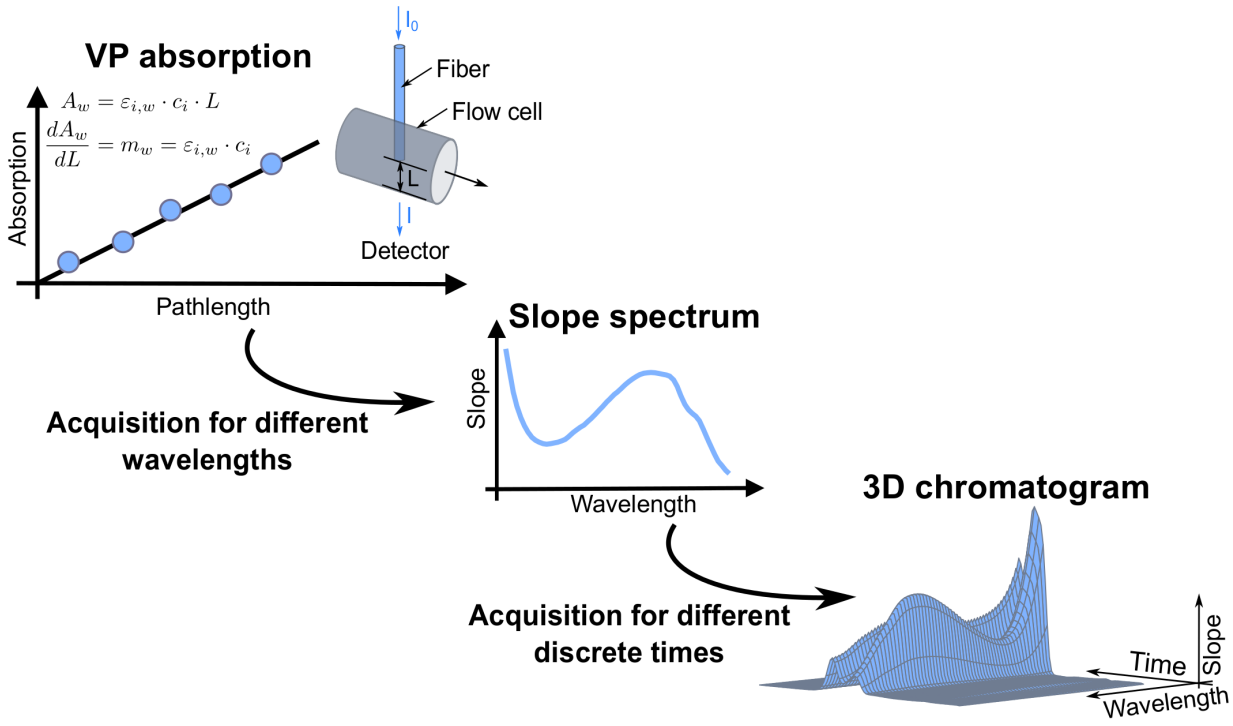


Figure 4.1: Slope spectroscopy in chromatography: Variable pathlengths L were set by moving the fiber of the Flow VPE. After the identification of the linear range of the dependency of the absorption A from L, UV absorption spectra at five different pathlengths were acquired. The slope m of A as a function of L was calculated for every wavelength to obtain a slope spectrum for one discrete time. For a whole chromatography run, this results in a time/wavelength/slope 3D chromatogram.

dependency of the absorption A_w from L was first determined for one wavelength w (280 nm). Within that range, an UV absorption spectrum was obtained at five different pathlengths. The slopes m_w of A_w as a function of L were subsequently calculated for every wavelength w, which resulted in a slope spectrum (cf. Fig. 4.1). The total duration of one measurement (including the screening of the linear range) was about 30 s. As slope spectra were obtained during a whole chromatography run, the result of one experiment was a time/wavelength/slope 3D chromatogram. The obtained slope spectra are correlated with the protein concentration c_i of a species i with proportionality factors $\varepsilon_{i,w}$.

The reference analytics of collected fractions was performed using a Dionex UltiMate 3000 rapid separation liquid chromatography system (Thermo Fisher Scientific, Waltham, USA). The system was composed of a HPG-3400RS pump, a WPS-3000 analytical autosampler, a TCC-3000RS column thermostat, and a DAD3000RS detector.

4.2.2 Case Study I: Separation of Cyt c from Lys

Proteins and Buffers

As model system, a protein mixture consisting of lysozyme from hen egg white (lys) and cytochrome c from equine heart (cyt c) was applied (both Sigma-Aldrich, St. Louis, USA). For the preparative runs, the equilibration buffer was 20 mM sodium phosphate (pH 7). Elution was carried out with 20 mM sodium phosphate and 500 mM sodium chloride (pH 7). For the reference analytics (analytical cation exchange chromatography), equilibration was performed with 20 mM Tris (pH 8) and elution was carried out with 20 mM Tris and 700 mM sodium chloride (pH 8). All buffer components were purchased from VWR, West Chester, USA. The buffers were prepared with Ultrapure Water (PURELAB Ultra, ELGA LabWater, Viola Water Technologies, Saint-Maurice, France), filtrated with a cellulose acetate filter with a pore size of 0.22 μm (Pall, Port Washington, USA), and degassed by sonification before usage.

Chromatography Runs

A HiTrap 16x25 mm column prepacked with SP Sepharose FF (GE Healthcare) was first equilibrated for 5 column volumes (CVs) and then loaded with 418 mg lys and 41.8 mg cyt c. After a wash of 1 CV with equilibration buffer, the proteins were eluted by performing a linear gradient from 0 to 100 % elution buffer. The gradient length was 2, 4, 6, and 8 CVs. While the results of the runs with a length of 2, 4, and 8 CVs were applied for the PLS model calibration, the run with a gradient length of 6 CVs was used to confirm the model. At the beginning of the linear gradient, the data acquisition of the Flow VPE was started and slope spectra from 240-300 nm with 2 nm resolution were obtained. After the linear gradient elution, the column was regenerated for 3 CVs with the elution buffer. The flow rate was 0.5 ml/min for all steps and experiments. The elution peaks were collected in deep well plates (VWR) in 1000 μl fractions.

Reference Analytics

The collected fractions were analyzed by analytical cation exchange chromatography. For each fraction, a Proswift SCX-1S 4.6 x 50mm column (Thermo Fisher Scientific) was first equilibrated for 2 min, loaded with 20 μl sample, washed for 0.5 min with equilibration buffer, and eluted with a linear salt gradient from 10-100 % elution buffer in 2 min. The column was subsequently regenerated with 100 % elution buffer for 1 min. The flow rate was 1.5 mL/min for all steps.

4.2.3 Case Study II: Separation of HMWs from mAb Monomer

Proteins and Buffers

Protein A pool was obtained from Lek Pharmaceuticals d.d. (Mengeš, Slovenia) and stored at −80°C before experimentation. Because of the low aggregate level of the applied mAb, a part of the Protein A pool was pH stressed and applied to spike the rest of the Protein A pool to an aggregate level of 10 %. For the preparative runs, the equilibration buffer was 20 mM sodium citrate (pH 6) and elution was performed with 20 mM sodium citrate and 500 mM sodium chloride (pH 6). For the reference analytics (analytical size exclusion chromatography), a buffer with 200 mM potassium phosphate and 250 mM potassium chloride (pH 7) was used. All buffer components were purchased from VWR. All buffers were prepared with Ultrapure Water (PURELAB Ultra). Prior to the experiments, the buffers and the feed were filtrated with a cellulose acetate filter with a pore size of 0.22 μm (Pall). The buffers were also degassed by sonification before usage.

Chromatography Runs

The HiTrap 16x25 mm SP Sepharose FF column was first equilibrated for 5 CVs and then loaded with 200 mg mAb (monomer and HMWs). After a wash of 3 CV with equilibration buffer, variable linear gradients from 0 to 100 % elution buffer were performed. The gradient length was 4, 5, 6, and 7 CVs. The results of the runs with a length of 4, 5, and 7 CVs were used to calibrate the PLS model, while the run with a length of 6 CVs was applied to confirm the model. As for the model protein study, the data acquisition of the Flow VPE was started at the beginning of the gradient and slope spectra from 240-340 nm with 2 nm resolution were recorded. After the linear gradient elution, the column was regenerated for 3 CVs with the elution buffer. The flow rate was 0.5 ml/min for all steps and experiments. The elution peaks were collected in deep well plates (VWR) in 1000 μl fractions.

Reference Analytics

As reference analytics, size exclusion chromatography was performed with the collected fractions to determine the concentration of mAb monomer and HMWs. For each fraction, 20 μl sample were loaded on a 4.6 x 150 mm TSKgel SuperSW mAb HTP column (Tosoh, Tokio, Japan). The flow rate was 0.3 ml/min for all examined samples.

4.2.4 Data Analysis

As the total duration of a slope spectrum acquisition (including the screening for the linear range) was slightly variable, the results of the offline reference analytics were linearly interpolated such that they matched the slope spectra. The slope spectra were then preprocessed by mean centering and correlated with the results of the offline analytics using PLS technique (Eriksson et al., 2006; Höskuldsson, 1988; Martens and Næs, 1989). The number of latent variables in the corresponding PLS model was based on the root mean square error (RMSE) of the model prediction in a cross validation.

For both case study, the concentrations of the different species were subsequently smoothed over time by a Savitzky-Golay filter (11 support points) in Matlab® (MathWorks, Inc., Natrick, USA) (Savitzky and Golay, 1964). The Savitzky-Golay filter was used in a symmetric form allowing to smooth the central data point in a given window.

4.3 Results and Discussion

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

4.3.1 Application of VP Spectroscopy for Chromatography

In VP spectroscopy, UV absorption measurements at different pathlengths are obtained. Figure 4.2 A illustrates a typical chromatogram obtained by VP spectroscopy at one wavelength (here 280 nm). The green lines represent the obtained absorption values at the different pathlengths applied in the run, which are illustrated by the grey lines. At each time point, the slope (orange line) was determined by regression analysis between five absorption values and pathlengths. The slope is proportional to the total protein concentration. It is worth noting that the absorption values stay approximately constant during protein elution. This can be explained by the fact that the FlowVPE determines the linear range of the dependency of the absorption from the pathlength prior to data acquisition. For the VP absorption measurement, the longest pathlength is then selected such that the corresponding absorption is about 1 AU. As a consequence, the set pathlengths are inversely proportional to the slope and thus to the protein concentration (cf. Fig. 4.2 A).

While Figure 4.2 A displays a typical chromatogram for one wavelength, a chromatogram for multiple wavelength measurements is presented in Figure 4.2 B. Instead of one slope, a slope spectrum was acquired. As the slopes were recorded during a whole chromatography run, this resulted in a time/wavelength/slope 3D chromatogram. The obtained 3D chromatograms of the linear gradient chromatography experiments were the

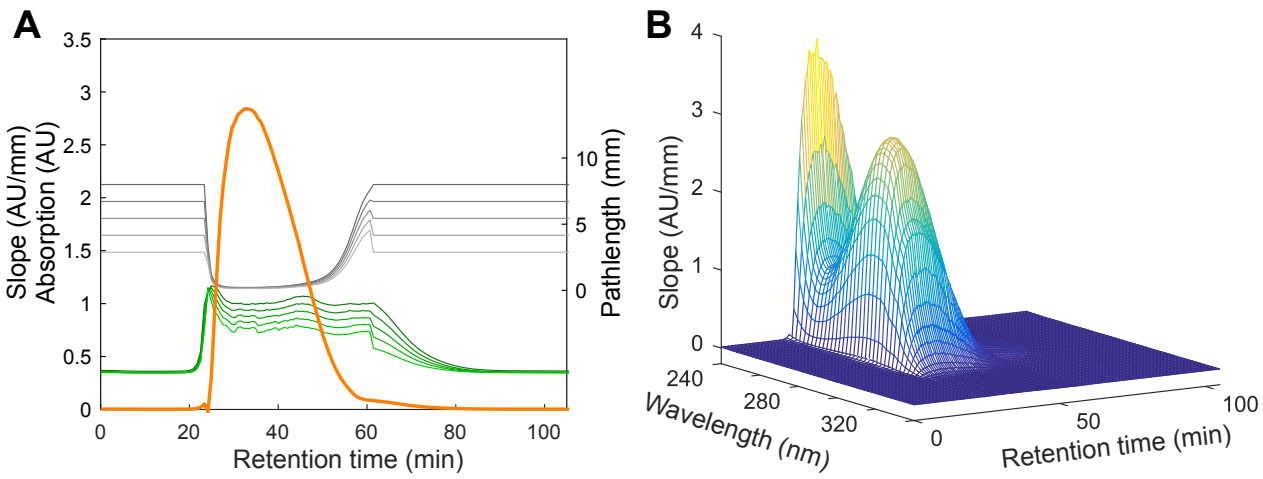


Figure 4.2: Typical chromatograms obtained by VP spectroscopy for one wavelength (A) and multiple wavelengths (B). A: The green lines represent the obtained absorption values at the different pathlengths applied in the run (grey lines). At each point in time, the slope (orange line) was determined by regression analysis between five absorption values and pathlengths. B: If slope spectra are recorded during a chromatography run, a 3D chromatogram is obtained.

starting point for the PLS model calibration and confirmation, which is discussed for both case studies in the following sections.

4.3.2 Case study I: Separation of Cyt c from Lys

Based on a cross validation, a PLS2 model with three latent variables was selected for the model protein case study. The resulting PLS model predictions for the three calibration runs (gradient length of 2, 4, and 8 CVs) are displayed in Figure 4.3 A, B, and D. The figures compare the PLS model prediction for lys (solid blue lines) and cyt c (solid red lines) with the results of the corresponding reference analytics (blue bars for lys and red bars for cyt c). A good agreement was observed between PLS model prediction and reference for all calibration runs. It should be however noted, that the PLS model was based on the slope spectra of these runs.

To confirm the model, predictions for a 6 CV gradient run were made, which closely follow the corresponding reference analytics (cf. Fig. 4.3 C). The RMSE of this run was 0.52 g/l for cyt c and 1.84 g/l for lys. The combination of VP spectroscopy and PLS modelling allowed for a selective quantification of lys during chromatography runs with highly loaded columns and maximal lys peak concentrations of up to 80 g/l. As the pathlength was adapted dynamically during the runs, the method also enabled a precise detection of the lower concentrated cyt c.

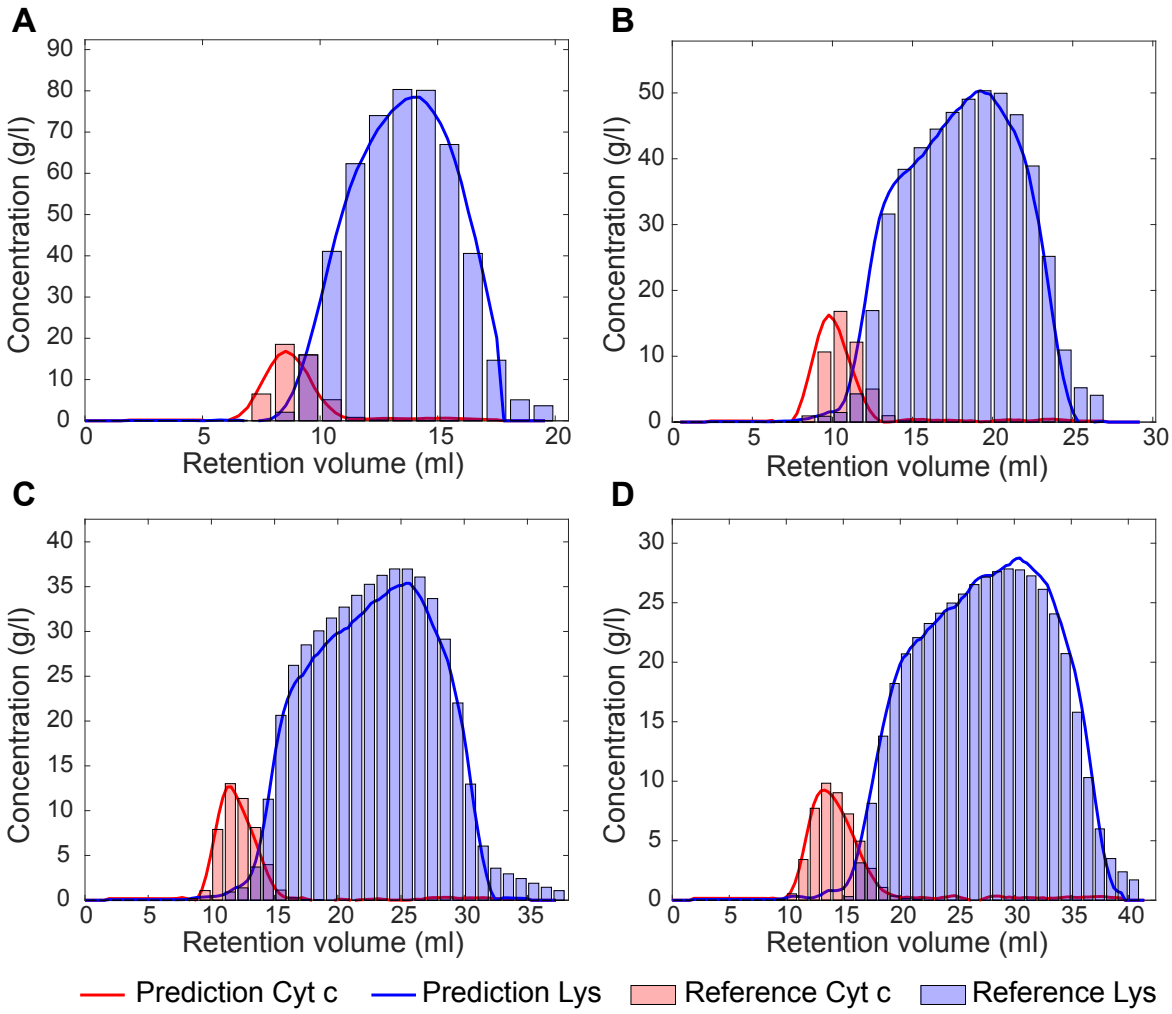


Figure 4.3: Comparison of the PLS model prediction for lys and cyt c with the results of the offline reference analytics for a gradient length of A: 2 CV, B: 4 CV, C: 6 CV (confirmation run), and D: 8 CV.

4.3.3 Case Study II: Separation of HMWs from mAb Monomer

Based on the cross validation, a PLS1 model with 4 latent variables was selected for the mAb monomer, while a PLS1 model with 9 latent variables was used for the HMWs. Figure 4.4 A, B, and D illustrate the PLS model prediction for the three calibration runs (gradient length of 4, 5, and 7 CVs), while Figure 4.4 C displays the results of the confirmation run (6 CV gradient). The figures show a comparison between the PLS model prediction for the mAb monomer (solid blue lines) and the HMWs (solid red lines) with the results of the reference analytics (blue bars for mAb monomer and red bars for HMWs). In all four runs, the model prediction complied with the reference analytics. The RMSE of prediction in the confirmation run was 1.15 g/l for the mAb monomer and 0.48 g/l for the HMWs respectively. This demonstrates the applicability of the method for proteins, with only slight differences in the corresponding absorption spectra.

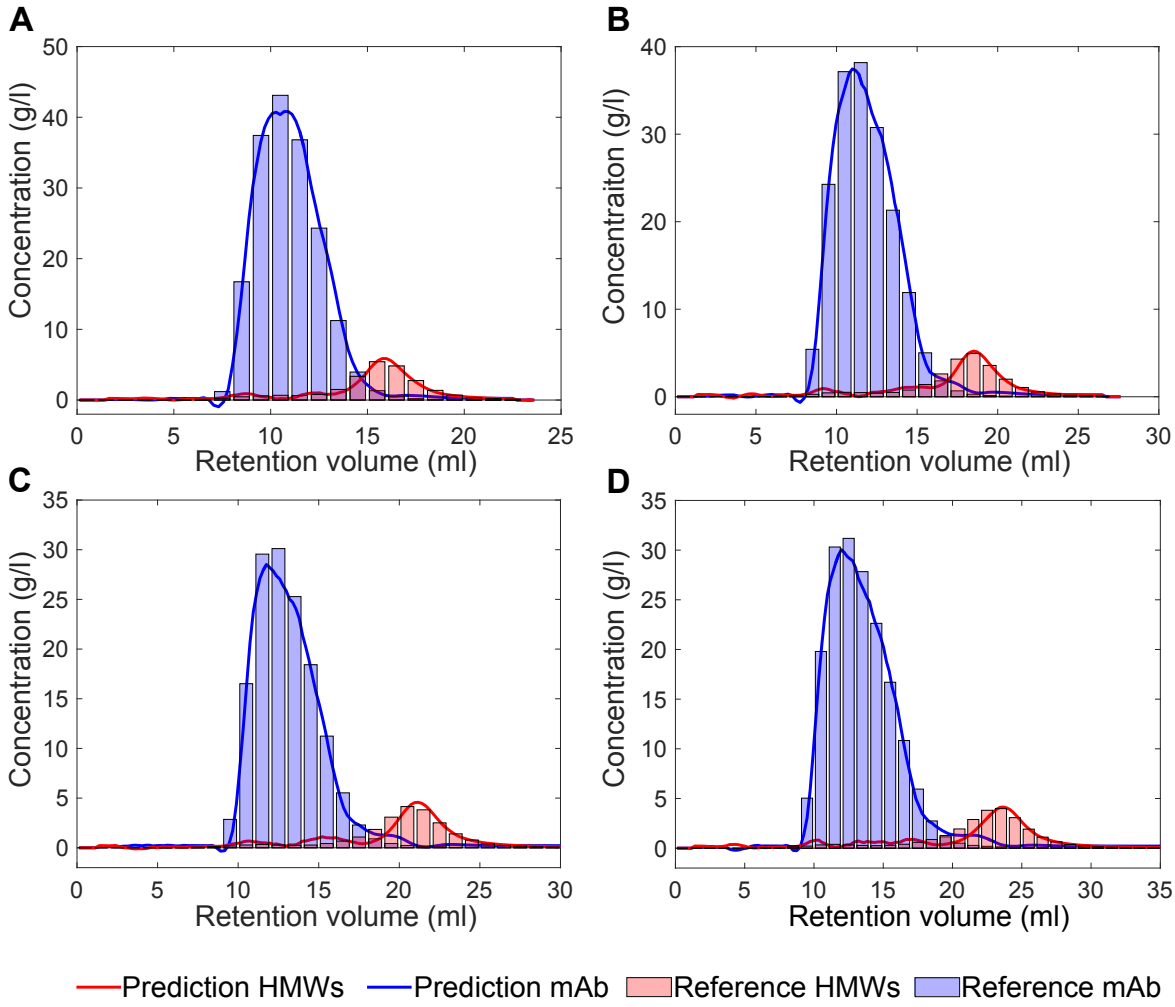


Figure 4.4: Comparison of the PLS model prediction for mAb and HMWs with the results of the offline reference analytics for a gradient length of A: 4 CV, B: 5 CV, C: 6 CV (confirmation run), and D: 7 CV.

4.4 Conclusion and Outlook

The inline monitoring of preparative chromatography was successfully realized in this study. It was demonstrated, that the conjunction of variable pathlength UV/Vis spectroscopy and PLS modelling allows for a selective inline protein quantification in a broad dynamic range of concentrations. The method enabled the monitoring of chromatography runs with highly loaded columns. The product peak concentration varied between 30-80 g/l, while the contaminant peak concentration was only 4-20 g/l. Consequently, the proposed method has potential for the real-time monitoring and control of preparative chromatography. It might also be applicable for controlling switching times in continuous chromatography. Future challenges are especially related to the scale up and robustness of the method as well as to the optimization of the measurement time in VP spectroscopy.

Acknowledgment

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 635557. We kindly thank Lek Pharmaceuticals d.d. (Mengeš, Slovenia) for providing the Protein A pool.

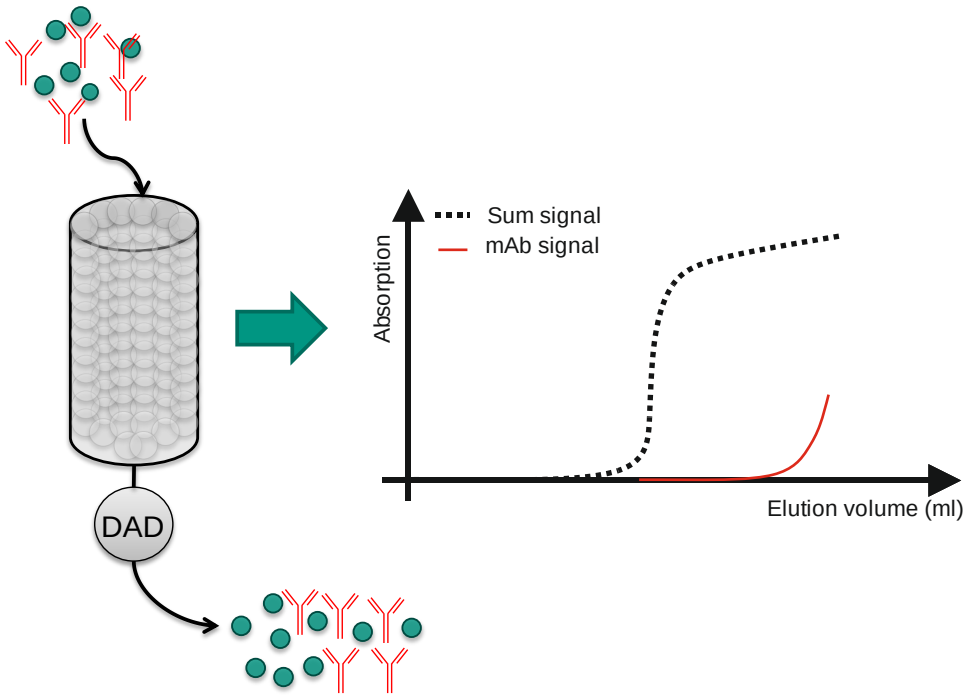
5 Real-time Monitoring and Control of the Load Phase in a Protein A Capture Step

M. Rüdt[1,*], N. Brestrich[1,*], L. Rolinger[1], J. Hubbuch[1,2]

[1] : Institute of Process Engineering in Life Sciences, Section IV: Biomolecular Separation Engineering, Karlsruhe Institute of Technology, Engler-Bunte-Ring 3, 76131 Karlsruhe, Germany

[*] : contributed equally

[2] : Corresponding author. *Email*: juergen.hubbuch@kit.edu



Abstract

The load phase in preparative Protein A capture steps is commonly not controlled in real-time. The load volume is generally based on an offline quantification of the monoclonal antibody (mAb) prior to loading and on a conservative column capacity determined by resin-life time studies. While this results in a reduced productivity in batch mode, the bottleneck of suitable real-time analytics has to be overcome in order to enable continuous mAb purification. In this study, Partial Least Squares Regression (PLS) modelling on UV/Vis absorption spectra was applied to quantify mAb in the effluent of a protein A capture step during the load phase. A PLS model based on several breakthrough curves with variable mAb titers in the harvested cell culture fluid was successfully calibrated. The PLS model predicted the mAb concentrations in the effluent of a validation experiment with a root mean square error (RMSE) of 0.06 mg/ml. The information was applied to automatically terminate the load phase, when a product breakthrough of 1.5 mg/ml was reached. In a second part of the study, the sensitivity of the method was further increased by only considering small mAb concentrations in the calibration and by subtracting an impurity background signal. The resulting PLS model exhibited a RMSE of prediction of 0.01 mg/ml and was successfully applied to terminate the load phase, when a product breakthrough of 0.15 mg/ml was achieved. The proposed method has hence potential for the real-time monitoring and control of capture steps at large scale production. This might enhance the resin capacity utilization, eliminate time-consuming offline analytics, and contribute to the realization of continuous processing.

Keywords: Process Analytical Technology, Capture Step, Protein A Chromatography, Selective Antibody Quantification, Partial Least Squares Regression

5.1 Introduction

A capture step is the first unit operation in the protein purification process which is used to bind the target protein from crude harvested cell culture fluid (HCCF). It increases product concentration as well as purity and prevents proteolytic degradation. Due to its high selectivity, Protein A capture is widely used in current monoclonal antibody (mAb) purification platform processes (Hahn et al., 2006; Shukla et al., 2007; Shukla and Thömmes, 2010; Tarrant et al., 2012; Tsukamoto et al., 2014).

A difficulty in Protein A capture is a lack of real-time analytics for mAb quantification in the HCCF and in the column effluent during loading. As both the mAb and impurities contribute to the absorption at 280 nm (A_{280}), single wavelength measurements are not suitable as selective analytics (Gupta, 2013). To determine the mAb titer in the HCCF, elaborate offline analytics is commonly performed (Fahrner and Blank, 1999b). As mAb titers are influenced by variability in the cell culture, this offline analytics has to be repeated for every lot in order to adapt the load volume onto the column (Fahrner and Blank, 1999a). While this results in a reduced productivity in batch mode, the bottleneck of suitable real-time analytics has to be overcome to enable continuous mAb purification.

In addition to the mAb titer in the HCCF, the optimal load volume onto the column is also influenced by the resin capacity. Due to leaching and degradation of the Protein A

ligands as well as pore and ligand blocking by leftover impurities or product, the capacity of the resin decreases over cylce time (Jiang et al., 2009). In batch mode, a conservative loading is commonly applied to avoid breakthrough of the expensive product at the cost of productivity. In contrast to that, columns are overloaded in continuous mode to maximize productivity (Angarita et al., 2015). In this case, the determination of the the percentual product breakthrough is necessary for process control (Warikoo et al., 2012).

To perform (near) real-time process monitoring and control, several process analytical technology (PAT) tools have been developed to enable fast mAb quantification in the cell culture fluid and in the column effluent during loading. For instance, atline mid-IR spectroscopy in combination with multivariate data analysis has been applied for secreted mAb quantification during a Chinese Hamster Ovary (CHO) cell culture process (Capito et al., 2015b). Selective mAb quantification in upstream processing was also successfully realized by atline matrix-assisted laser desorption/ionization mass spectrometry (Steinhoff et al., 2015). For the control of the load phase of a two column continuous protein A chromatography process, which was connected to a CHO perfusion culture, atline analytical chromatography was applied (Karst et al., 2015). Atline monitoring however bears the risk of human errors resulting in contamination, time-delays, or missing data.

In order to minimize human impact, automated sampling can be applied. Automated analytical chromatography has been used in upstream processing to monitor the mAb titers (Chase, 1986; Ozturk et al., 1995; Paliwal et al., 1993). In downstream processing, this technique was successfully used for mAb quantification in the column effluent during the load phase of Protein A chromatography. As soon as 1 % mAb breakthrough was detected, the load phase was automatically terminated (Fahrner and Blank, 1999a). Automated analytical chromatography is relatively easy to develop and equipment is commercially available. However, the equipment is expensive and the technique error-prone. Besides from the risk of contamination, the time delay between sampling and analytical results bears the risk of late reaction or requires a slow-down of the process.

PAT tools that operate in real-time, such as UV-based methods, overcome these limitations. In a patent application, a UV-based control method for determining binding capacities in Protein A capture was disclosed (Bängtsson et al., 2012). The method is based on the calculation of a difference signal between two detectors situated at the column in- and outlet. During the load phase, the post column signal is supposed to stabilize and is referred to as impurity baseline. As soon as the mAb breaks through, there is an increase in the post-column UV signal above the impurity baseline which corresponds to a breakthrough level of the product. Consequently, the method is very suitable for determining column switching times in continuous Protein A capture. It allows for an equal loading in terms of percentual breakthrough regardless of the mAb titer variability in the feed or decreasing column capacities. However, it requires two detectors posing a risk of unequal detector drifts. A further limitation might be displacement effects of contaminants that prevent a stabilized impurity baseline. The technique might also be limited to the equipment of the future patent holder.

Another recently published UV/Vis-based method for monitoring and control in protein chromatography applies UV/Vis absorption spectra instead of single wavelength measurements (Brestrich et al., 2014, 2015). Different protein species exhibit distinct

variations in their UV absorption spectra. Consequently, Partial Least Squares Regression (PLS) technique has been used to correlate absorption spectra with selective protein concentrations. The method was successfully applied for a selective inline protein quantification and for product purity-based pooling decisions in real-time. However, no load control in Protein A chromatography has been performed so far using this technique.

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

5.2 Materials and Methods

5.2.1 Cell Culture Fluid and Buffers

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

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

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

5.2.2 Chromatographic Instrumentation

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

The communication between Unicorn and Chromeleon was implemented analogous to the protocol published in (Brestrich et al., 2014). Shortly, Unicorn triggers the DAD data acquisition by sending a digital signal to a Matlab script (MathWorks, Natick, USA), which communicates with Chromeleon via a Visual Basics for Application Macro (Microsoft, Redmond, USA). If a certain condition such as a defined mAb concentration is fulfilled, the Matlab script sends a signal back to Unicorn to terminate a phase in the chromatographic method.

Reference analysis of collected fractions was performed using a Dionex UltiMate 3000 rapid separation liquid chromatography system (Thermo Fisher Scientific). The system was composed of a HPG-3400RS pump, a WPS-3000 analytical autosampler, a TCC-3000RS column thermostat, and a DAD3000RS detector.

5.2.3 Chromatography Runs

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

5.2.4 Analytical Chromatography

As displayed in Figure 5.1, the collected fractions of all runs were examined by analytical Protein A chromatography to obtain the mAb concentrations. For each sample, a 2.1x30 mm POROS prepacked Protein A column (Applied Biosystems, Foster City, USA) was equilibrated with 2.6 column volumes (CVs) of equilibration buffer, flowed by an injection of 20 μl sample. The column was then equilibrated with 0.8 CVs of equilibration buffer and eluted with 1.4 CVs of elution buffer. The flow rate was 2 ml/min for all phases and experiments.

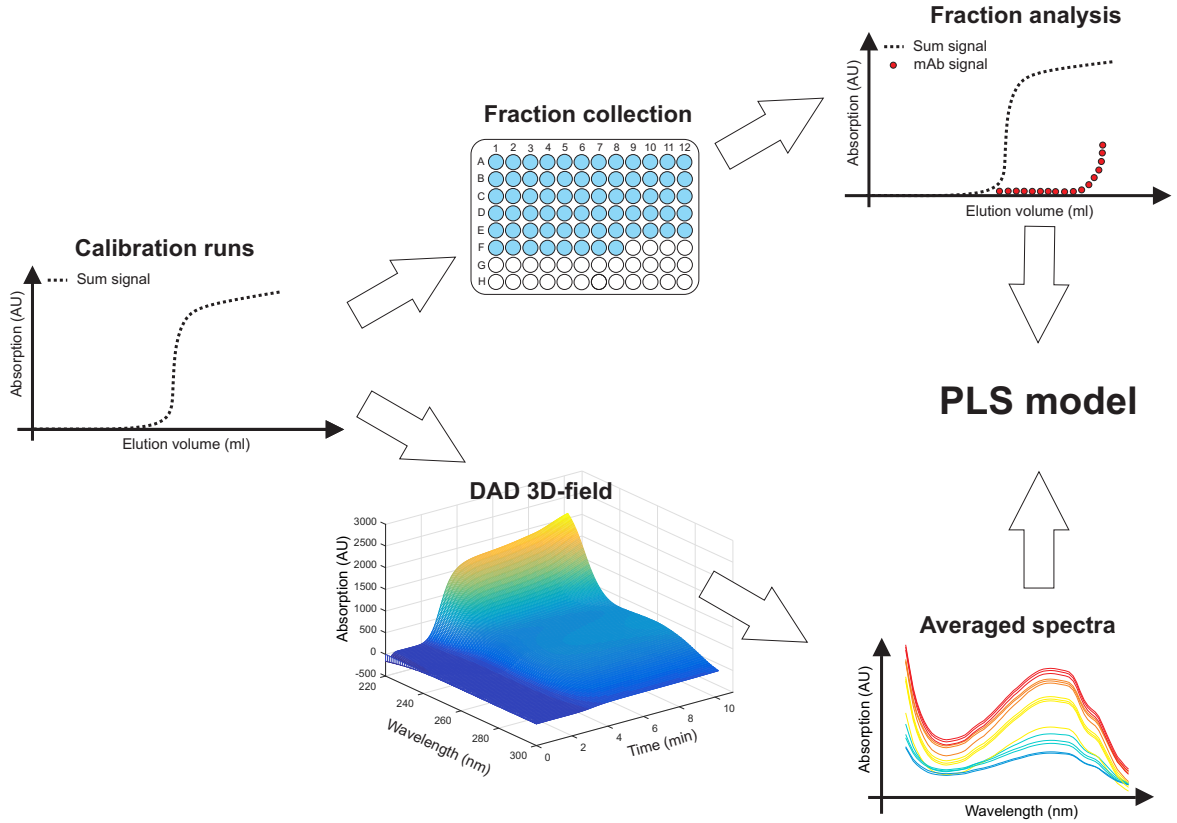


Figure 5.1: Experimental procedure for the PLS model calibration: For each calibration run, 200 μl fractions were collected and analyzed by analytical Protein A chromatography to obtain the mAb breakthrough curves. In addition, averaged spectra corresponding to the fraction size were calculated from the time, wavelength, and absorption 3D-field. Averaged spectra and mAb concentrations were eventually correlated using PLS technique.

5.2.5 Data Analysis

For the correlation of the absorption spectra with the mAb concentrations, PLS technique was applied using SIMCA (MKS Data Analytics Solutions, Umeå, Sweden). SIMCA applies the NIPALS-algorithm for PLS. Before performing PLS, all spectra were preprocessed by mean centering using SIMCA. PLS finds variation in the spectral data matrix, which is relevant for the correlation with the mAb concentrations and thereby separates information in the matrix from detector noise (Eriksson et al., 2006; Höskuldsson, 1988; Martens and Næs, 1989). In order to achieve this separation, collinearity in the data is reduced by summarizing variables (here wavelengths) with similar information in latent variables (LVs). This is done in a way such that the content of relevant information for the correlation included in each LV is highest for the first LV and decreases for the following ones. The number of applied LVs in a PLS model is hence a measure of data reduction and only a few LVs are required to obtain the correlation between absorption spectra and mAb concentrations.

The number of applied LVs has to be evaluated thoroughly to avoid under- or overfitting of a model. In order to determine a reasonable number of LVs, the root mean square error (RMSE) for the prediction of validation samples is usually determined in dependence on the number of LVs applied in a PLS model. The minimum corresponds to

the optimal number of LVs. In this study, cross validation was performed to determine an optimal number of LVs. Therefore, the calibration data was separated into seven groups. One group was then excluded during model calibration and the RMSE for theses samples was calculated subsequently. For every number of LVs, this procedure was performed until each group was excluded. Based on the so obtained number of LVs, completely independent runs were predicted to evaluate the final models.

The PLS model calibration was based on the results of the runs with the following mAb titers in the feed: 2.7, 2.85, 3.15, and 3.3 mg/ml. The results of the corresponding spectral acquisitions are time, wavelength and absorption 3D-fields. The 3D-fields were averaged in time according to the fraction duration as displayed in Figure 5.1. The results of theses calculations were stored in an absorption matrix. Afterwards, PLS was carried out to correlate the mAb concentrations of the collected fractions with the the corresponding absorption matrix.

5.2.6 Real-Time Monitoring and Control

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

In a second run, the stop criterion to terminate the load phase was set to a target concentration of 0.15 mg/ml (5 % product breakthrough). To enhance the sensitivity of the method for this purpose, a second PLS model was calibrated. Only samples with mAb concentrations below 0.5 mg/ml were considered in the model calibration. Additionally, a background subtraction was performed. As soon as the change in absorption signal fell under a predefined threshold, an average absorption was calculated for every wavelength. This impurity background was subtracted from the absorption of all following data points.

5.3 Results and Discussion

As described above, the breakthrough of mAb was monitored in real-time by UV/Vis spectroscopy in combination with a PLS model. To calibrate the PLS model, 4 chromatographic runs at mAb concentrations of 2.7, 2.85, 3.15 and 3.3 mg/ml in the feed were performed and analyzed by offline analytics. The model was eventually confirmed by performing a real-time control of two runs with a mAb titer of 3 mg/ml. The difference in the mAb titers in the feed ensured variable mixing ratios between product and contaminants. This was done to imitate variability in upstream processing and to span a calibrated design space for the PLS model.

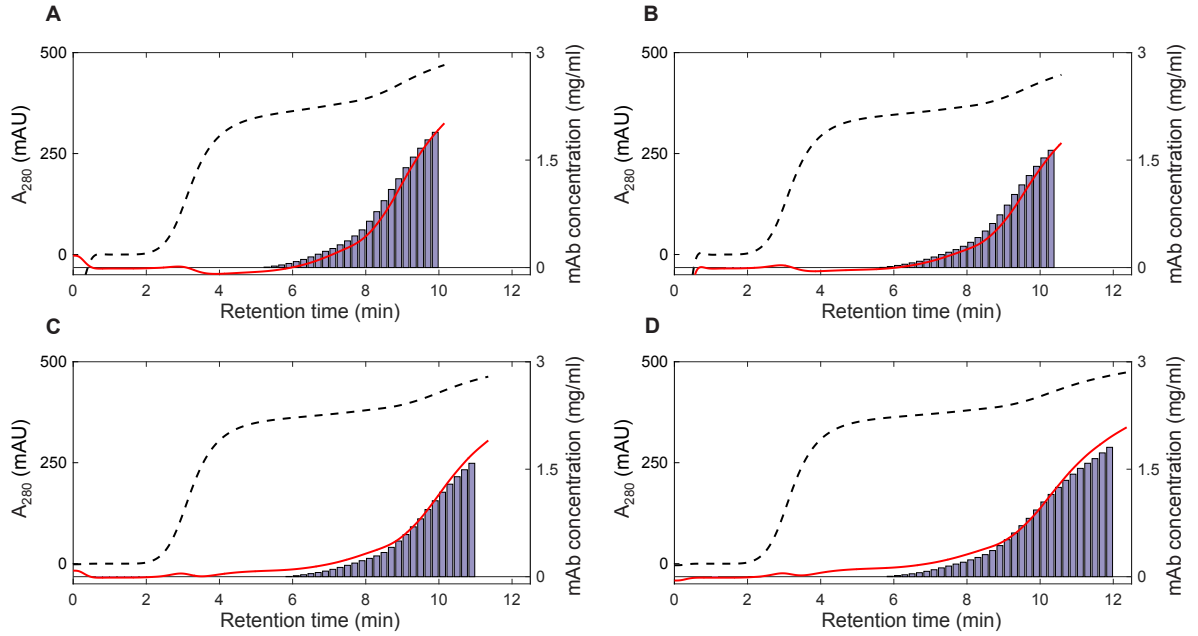


Figure 5.2: Results of the PLS model calibration. The A_{280} (measured at a pathlength of 0.4 mm and displayed as dashed black line) is compared with the results of the offline analytics for mAb quantification (blue bars). The PLS model prediction is illustrated as red lines. The four runs exhibited variable mAb titers in the feed A: 3.3 mg/ml, B: 3.15 mg/ml, C: 2.85 mg/ml, D: 2.7 mg/ml.

5.3.1 PLS Model Calibration

The results of the model calibration are illustrated by Figure 5.2. It compares the A_{280} (recorded at a pathlength of 0.4 mm and displayed as dashed black line) to the concentrations measured by offline analytics (blue bars) and the signal calculated by the calibrated PLS model (solid red lines). The number of LVs was set to 4 based on a minimal RMSE of 0.08 mg/ml in the cross validation. The calibrated PLS model was applied to evaluate all 3D-fields. In contrast to model calibration, where averaged spectra were used, the spectral raw data at each time point was translated into concentrations. The estimated concentrations by the PLS model closely follow the measured values by offline analytics. It is worth noting that no clear plateau of the A_{280} is reached after the breakthrough of media components. Instead, the A_{280} continuous to increase. This may be caused by different impurities being retained differently on the membrane. Indeed, it has previously been shown, that major interactions between HCPs, the stationary phase and mAbs may occure (Shukla and Hinckley, 2008; Aboulaich et al., 2014). The advent of mAb breakthrough cannot be clearly distinguished from A_{280} alone. Based on the multivariate spectral data, the PLS model is able to predict protein concentrations, which allows for real-time monitoring and control.

5.3.2 Real-Time Monitoring and Control

For the confirmation of the obtained results, the calibrated PLS model was used to control the load phase of a Protein A capture step in real-time. In a first run, a target breakthrough concentration of 1.5 mg/ml was set, which corresponds to 50 % product breakthrough. Figure 5.3 A shows the A_{280} (dashed black line), the real-time prediction

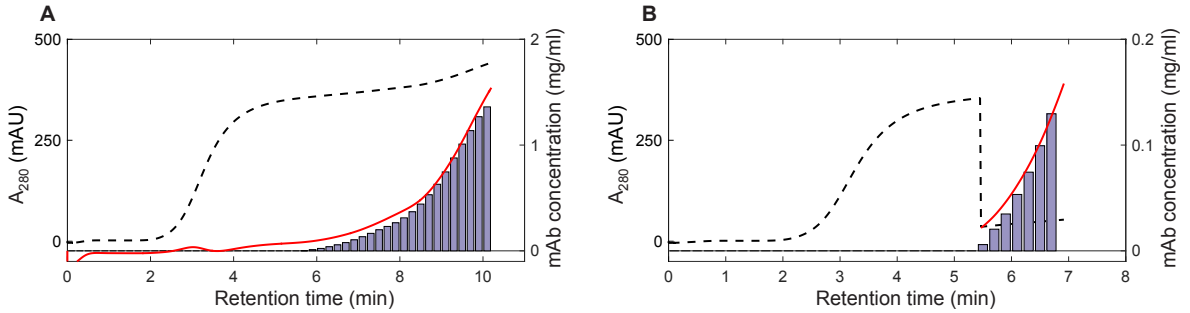


Figure 5.3: Results of the model evaluation by performing a real-time control of the load phase using a mAb titer of 3 mg/ml in the feed. The PLS model prediction (red lines) is compared with the results of the offline analytics (blue bars) as well as the A_{280} (measured at a pathlength of 0.4 mm and displayed as dashed black line). The load phase was automatically terminated, when a mAb concentration in the effluent of A: 1.5 mg/ml or B: 0.15 mg/ml was reached. The sudden decrease in the A_{280} arises from the background subtraction.

of mAb concentrations (solid red line) and the corresponding offline analytics (blue bars). The model reached an RMSE for prediction of 0.06 mg/ml compared to the offline analytics.

To further improve model predictions at low mAb concentrations, a second PLS model was calibrated based on the calibration data set as described in the method section. The model was used predict and stop a load phase in a second run at 0.15 mg/ml, which corresponds to 5 % product breakthrough. The results of this second run are displayed in Figure 5.3 B. As an impurity background was subtracted to increase the sensitivity of the method, the A_{280} suddenly decreases. The second PLS model reached an RMSE for prediction of 0.01 mg/ml.

During both runs, the respective load phases were successfully terminated close to the intended breakpoints. In Table 5.1, a summary of intended and measured mAb concentrations in the last fraction of both confirmation runs is shown. The Matlab script sent a digital signal to Unicorn and terminated the load phase, when the targeted breakthrough concentration was reached. As the targeted breakthrough set points were concentrations at discrete time points, they are expected to be slightly higher than the concentrations of the last fraction determined by offline analytics. This was observed for both confirmation runs (cf. Table 5.1). For an easier comparison between model and offline analytics, a concentration based on an averaged absorption spectrum was calculated for the last fractions of both runs and compared with the corresponding offline analytics. For the first run, the deviation between prediction and reference was 8.0 %, while for the second

Table 5.1: Results of both confirmation runs: The targeted concentration to terminate loading is compared with the mAb concentration in the last fraction determined by offline analytics. In addition, a PLS model prediction for the last fraction based on an averaged absorption spectrum is shown for comparison.

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

run a deviation of 2.3 % was found. This demonstrates that the described method can be successfully used to control the load phase in a Protein A capture step.

5.4 Conclusion and Outlook

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

Acknowledgment

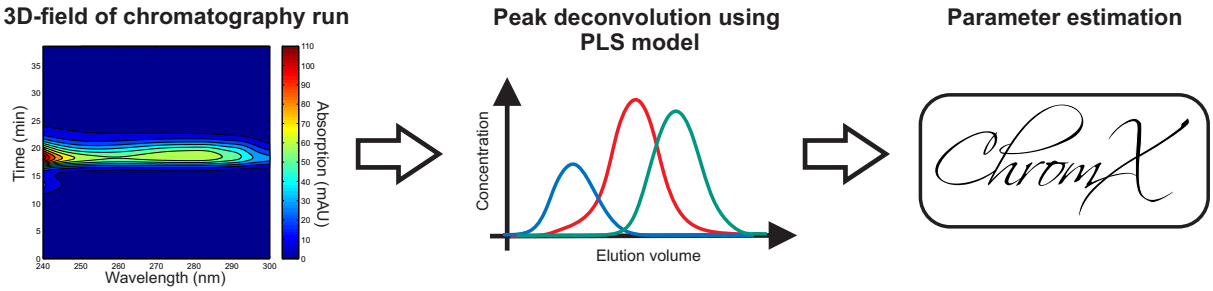
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 635557. We kindly thank Lek Pharmaceuticals d.d. (Mengeš, Slovenia) for providing the HCCF and mock.

6 Application of Spectral Deconvolution and Inverse Mechanistic Modelling as a Tool for Root Cause Investigation in Protein Chromatography

N. Brestrich[1], T. Hahn[1], J. Hubbuch[1,*]

[1] : Institute of Process Engineering in Life Sciences, Section IV: Biomolecular Separation Engineering, Karlsruhe Institute of Technology, Engler-Bunte-Ring 3, 76131 Karlsruhe, Germany

[*] : Corresponding author. *Email*: juergen.hubbuch@kit.edu



Abstract

In chromatographic protein purification, process variations, aging of columns, or processing errors can lead to deviations of the expected elution behavior of product and contaminants and can result in a decreased pool purity or yield. A different elution behavior of all or several involved species leads to a deviating chromatogram. The causes for deviations are however hard to identify by visual inspection and complicate the correction of a problem in the next cycle or batch. To overcome this issue, a tool for root cause investigation in protein chromatography was developed. The tool combines a spectral deconvolution with inverse mechanistic modelling. Mid-UV spectral data and Partial Least Squares Regression were first applied to deconvolute peaks to obtain the individual elution profiles of co-eluting proteins. The individual elution profiles were subsequently used to identify errors in process parameters by curve fitting to a mechanistic chromatography model. The functionality of the tool for root cause investigation was successfully demonstrated in a model protein study with lysozyme, cytochrome c, and ribonuclease A. Deviating chromatograms were generated by deliberately caused errors in the process parameters flow rate and sodium-ion concentration in loading and elution buffer according to a design of experiments. The actual values of the three process parameters and, thus, the causes of the deviations were estimated with errors of less than 4.4 %. Consequently, the established tool for root cause investigation is a valuable approach to rapidly identify process variations, aging of columns, or processing errors. This might help to minimize batch rejections or contribute to an increased productivity.

Keywords: Root Cause Investigation, Process Analytical Technology, Inline Monitoring, Mechanistic Modelling, Partial Least Squares Regression, Selective Protein Quantification

6.1 Introduction

The application of advanced Process Analytical Technology (PAT) concepts in biopharmaceutical production has been discussed increasingly in the recent years (Challa and Potumarthi, 2013; Glassey et al., 2011; Mercier et al., 2014; Rathore and Kapoor, 2015; Read et al., 2010). Especially the evaluation of continuous processing has given rise to an increasing interest in advanced PAT tools, as the implementation of monitoring and control strategies is a critical aspect for the success of continuous processing (Konstantinov and Cooney, 2015; Rathore et al., 2015). The successful application of PAT tools to minimize process variability and to ensure a consistent product quality is also promoted by the US Food and Drug Administration's (FDA's) PAT initiative (FDA, 2004).

In chromatographic protein purification, process variability, errors, or aging of columns can lead to changes in retention volumes or peak shapes of the eluting species and can consequently cause deviating chromatograms. This can result in inconsistent pool purities or yields, if the chromatographic process lacks feed-forward and/or feed-back control. In the case of co-eluting proteins, a fast peak deconvolution is required to enable feed-forward control in the context of real-time pooling. For feed-back control, the causes for

a deviating chromatogram have to be identified before the problem can be corrected in the next cycle or batch.

PAT tools enabling fast peak deconvolution of co-eluting proteins have been examined in various studies (Brestrich et al., 2014, 2015; Fahrner et al., 1998; Kaltenbrunner et al., 2012; Mendhe et al., 2015; Rathore et al., 2008a,b, 2010b). Especially PAT tools based on online-(U)HPLC have been applied to obtain the individual elution profiles of the co-eluting species and to perform real-time pooling decisions (Fahrner et al., 1998; Kaltenbrunner et al., 2012; Mendhe et al., 2015; Rathore et al., 2008a,b, 2010b).

Aside from online-(U)HPLC, a PAT tool for inline peak deconvolution based on mid-UV absorption spectra has been published recently (Brestrich et al., 2014). As the number and ratio of aromatic amino acids varies in different protein species, protein mid-UV absorption spectra exhibit distinct differences. The variance in mid-UV absorption spectra can be utilized to calibrate Partial Least Squares Regression (PLS) models, correlating the information in the spectra with selective protein concentrations. After calibration, PLS models can be applied to calculate selective protein concentrations from inline absorption spectra. The obtained peak deconvolution can be used to perform real-time pooling decisions. The PAT tool has been successfully applied to deconvolute the peaks of co-eluting serum proteins and of a co-eluting monoclonal antibody, its fragments, and aggregates (Brestrich et al., 2015). However, both online-(U)HPLC and spectral deconvolution only allow for the detection of changes in the elution profiles of co-eluting proteins. Hence, they are limited to feed-forward control. For a feed-back control, a mechanistic understanding is required of how changes in the process influence the resulting chromatogram.

A mechanistic understanding of the chromatographic process can be achieved using mechanistic chromatography models. Mechanistic chromatography modelling has been widely used in academia for process optimization (Helling et al., 2012; Huuk et al., 2014; Jakobsson et al., 2005) and robustness analysis (Degerman et al., 2009a,b; Jakobsson et al., 2005, 2007; Westerberg et al., 2012), even for crude feedstocks (Baumann et al., 2015). Despite its broad application, mechanistic modelling has not been explored yet for process analysis in biopharmaceutical protein purification.

In this study, inverse mechanistic modelling was used to identify causes for deviations in chromatograms. In inverse modelling, parameters influencing a chromatogram are altered systematically until the modelled chromatogram matches the observed one. However, this can be challenging if the proteins are co-eluting as the commonly applied mechanistic models are based on molar concentrations. While molar concentrations can be calculated for well-separated proteins after sensor calibration, offline analysis is required for co-eluting species. Offline analysis is however time-consuming and not applicable for process control. To overcome these limitations, inverse mechanistic modelling was combined with spectral deconvolution in this study. This tool for rapid root cause investigation first determines the individual elution profiles. For this purpose, a PLS model is applied to calculate selective protein concentrations from mid-UV absorption spectra. The individual elution profiles are subsequently used to identify deviations between actual value and setpoint of process parameters by inverse modelling.

The functionality of the tool was demonstrated in a case study with lysozyme (lys), cytochrome c (cyt c), and ribonuclease A (rib A). The three proteins were co-eluting during

their chromatographic separation on a SP sepharose FF column. After calibration of PLS and mechanistic model, deviating chromatograms were generated by deliberately caused errors in the process parameters flow rate and Na^+ concentration in loading and elution buffer. The actual values of the three parameters and thus the causes for deviations were subsequently estimated by inverse modelling with the deconvoluted chromatograms. The three parameters were selected as they are easy to manipulate in the laboratory and to demonstrate exemplary the functionality of the tool for error diagnostics. The actual values of the flow rate or the Na^+ concentration in loading and elution buffer can, of course, be measured easily using flow rate or conductivity sensors. However, solely measuring theses parameters does not allow inferences with regards to deviating chromatograms.

6.2 Theory

6.2.1 Partial Least Squares Regression

For spectral deconvolution, PLS was carried out using NIPALS algorithm (Eriksson et al., 2006; Höskuldsson, 1988; Martens and Næs, 1989). The PLS technique was applied to calibrate models correlating mid-UV absorption spectra with selective protein concentrations. PLS reduces the dimension of the spectral data matrix by only applying significant variance for the correlation with the protein concentrations. The data reduction is achieved by summarizing variables (wavelengths) of the spectral matrix containing similar information in so-called latent variables (LVs). The information content included in each LV is high for the first LV and decreases for the following LVs. Consequently, an optimal number of applied LVs in a PLS model has to be evaluated thoroughly to avoid overfitting of the model to the training data set.

6.2.2 Mechanistic Model

In this paper, ion exchange chromatography with low concentrated samples is used. A transport dispersive model (Guiochon et al., 1994; Schmidt-Traub, 2005) in connection with a Stoichiometric Displacement isotherm (Velayudhan and Horvath, 1988) is a reasonable choice for this setting. Ion exchange interaction are fast and comparably long ranging, lumping film transfer and pore diffusion is hence legitimate. In addition, no strong tailing was observed (cf. section 6.4). For the protein mass loaded onto the column in this study (cf. section 6.3), it is not to be expected that we encounter steric shielding. The models are explained in detail in the following.

Transport Dispersive Model

The transport dispersive model is a macroscopic description of the convective and diffusive concentration transport processes within a packed bed of porous particles. The mobile phase is divided into the interstitial volume with concentration c_i of species i and the pore volume with concentration $c_{p,i}$ (Guiochon et al., 1994; Schmidt-Traub, 2005). The concentration in the interstitial volume is transported with interstitial velocity u_{int} along the column and undergoes Fickian diffusion with an axial dispersion coefficient D_{ax}. The

concentration exchange with the particles considers the differences in volumes expressed with the bed porosity ε_b and introduces a lumped film transfer/pore diffusion coefficient k_{eff}. Complemented with Danckwerts boundary conditions, the model equations for the interstitial concentration of a column with length L are:

$$\begin{aligned}
\frac{\partial c_i}{\partial t}(x,t) &= -u_{int}(t)\frac{\partial c_i}{\partial x}(x,t) \\
&\quad + D_{ax}\frac{\partial^2 c_i}{\partial x^2}(x,t) - \frac{1-\varepsilon_b}{\varepsilon_b}\frac{3}{r_p}k_{eff,i} \\
&\quad \cdot (c_i(x,t) - c_{p,i}(x,t)) \quad \forall_i, && (6.1)\\
\frac{\partial c_i}{\partial x}(0,t) &= \frac{u_{int}(t)}{D_{ax}}(c_i(0,t) - c_{in,i}(t)) \quad \forall_i, && (6.2)\\
\frac{\partial c_i}{\partial x}(L,t) &= 0 \quad \forall_i, && (6.3)
\end{aligned}$$

where $3/r_p$ stems from the surface to volume ratio of spheres and is usually factored out of k_{eff} for a packing of spherical beads with radius r_p.

Depending on the particles' porosity ε_p, the concentration exchange is performed with the mobile phase in the pore volume or directly with the stationary phase concentration q_i:

$$\begin{aligned}
&\frac{3}{r_p}k_{eff,i}(c_i(x,t) - c_{p,i}(x,t)) \\
&= \varepsilon_p\frac{\partial c_{p,i}}{\partial t}(x,t) + (1-\varepsilon_p)\frac{\partial q_i}{\partial t}(x,t) \quad \forall_i. && (6.4)
\end{aligned}$$

Stoichiometric Displacement Isotherm

The Stoichiometric displacement isotherm (SDM - Stoichiometric displacement model) is able to model the multi-point binding of proteins in ion-exchange chromatography using a characteristic charge ν_i. In a stoichiometric exchange, ν_i counter ions are freed when a protein adsorbs. From the law of mass action Velayudhan and Horvath (Velayudhan and Horvath, 1988) derived the equation

$$k_{eq,i} = \frac{c_{p,salt}^{\nu_i}(x,t)\, q_i(x,t)}{q_{salt}^{\nu_i}(x,t)\, c_{p,i}(x,t)} \quad \forall_{i\neq salt}, \tag{6.5}$$

where the index *salt* depicts the respective counter-ion concentration. In the case considered here, we do not expect steric shielding such that we can express the salt concentration in the stationary phase as

$$q_{salt}(x,t) = \Lambda - \sum_i \nu_i q_i(x,t) \tag{6.6}$$

with total ionic capacity of the stationary phase Λ.

Numerical Solution

To solve the resulting equation system, the in-house software ChromX was employed (Hahn et al., 2015). ChromX uses finite elements for space discretization, here linear basis and test functions. For discretization in time, a fractional step θ-scheme was used because of its superior stability. The remaining non-linear equation systems were solved with fixed point iteration. As linear solver, a direct LU factorization method was chosen.

For isotherm parameter estimation as well as error diagnostics, a heuristic method was combined with a deterministic one. As demonstrated in (Hahn et al., 2014), deterministic solvers need a suitable starting point which was identified using Adaptive Simulated Annealing (Oliveira et al., 2012). Afterwards, the solution was refined using an implementation of the Levenberg-Marquardt algorithm (Lourakis, 2004).

Estimation of Parameter Certainty

As a measure of parameter certainty, an estimate of the covariance matrix is used. The diagonal elements of the covariance matrix represent the variance of the estimates. Obtaining the exact covariance matrix is a complex mathematical problem that leads most studies to use a Jacobian matrix (Bauer et al., 2000) or a Fisher information matrix (FIM) instead (Walter and Pronzato, 1997; Ucinski, 2004). The asymptotic variance of the maximum likelihood estimator is given by the Cramer-Rao lower bound (Cramer, 1957; Goodwin, Graham C, Payne, 1977), the reciprocal of the Fisher information:

$$Cov(\theta) \geq FIM(\theta)^{-1} = \left(\sum_{k=1}^{N_{exp}} FIM_k(\theta) \right)^{-1} \tag{6.7}$$

with the Fisher information matrix of each experiment k being defined with the parameter sensitivities of a Function $F_k(t)$ (Qureshi et al., 1980) :

$$FIM_k(\theta) = \frac{1}{{\sigma_k}^2} \int_T \frac{\partial F_k}{\partial \theta_j} \frac{\partial F_k}{\partial \theta_j}^T dt. \tag{6.8}$$

The confidence intervals contain the true parameter of interest with an a prioi defined probability. Given an estimate of the covariance matrix created from $F_k = (c_{meas,k}(t) - c_{sim,k}(t))^2$, a confidence interval for θ_j with a probability of 95 % is given by

$$\theta_j \pm c\sqrt{diag\left(Cov\left(\theta\right)\right)} \tag{6.9}$$

where c is the respective quantile of the Student's t-distribution.

6.3 Materials and Methods

6.3.1 Proteins, Buffers, and Columns

Lysozyme (lys) from hen egg white, ribonuclease A (rib A) from bovine pancreas, and cytochrome c (cyt c) from equine heart were purchased from Sigma-Aldrich, St. Louis,

USA. This system was chosen because of the stability of the proteins in aqueous solutions regarding the formation of multimers and their advantageous isoelectric point distribution, allowing separation at neutral pH by cation exchange chromatography. For all preparative chromatography runs, the loading buffer was 10 mM tribasic sodium citrate with sodium chloride concentrations ranging from 150 to 250 mM (180 to 280 mM total Na^+ concentration), depending on the experiment (pH 5). Elution was carried out using a high-salt buffer consisting of 10 mM tribasic sodium citrate with sodium chloride concentrations ranging from 400 to 600 mM (430 to 630 mM total Na^+ concentration), depending on the experiment (pH 5). Reference analysis (analytical cation exchange chromatography) was carried out using a loading buffer with 20 mM Tris (pH 8) and an elution buffer with 20 mM Tris and 700 mM sodium chloride (pH 8). All buffer components were purchased from VWR, West Chester, USA. Buffers were 0.22 μm filtered and degassed by sonification before usage. All preparative chromatography runs were performed with a HiTrap 7 x 25 mm column prepacked with SP Sepharose FF (average particle diameter of 90 μm, GE Healthcare, Chalfont St Giles, UK). Reference analysis was carried out using a Proswift SCX-1S 4.6 x 50 mm column (Thermo Fisher Scientific, Waltham, USA).

6.3.2 Chromatographic Instrumentation

All preparative chromatography runs were performed using a setup consisting of a conventional liquid chromatography system and a diode array detector (DAD). The liquid chromatography system was an Akta purifier 10 equipped with pump P-900, sample pump P-960, UV monitor UV-900 (10 mm optical pathlength), conductivity monitor C-900, pH monitor pH-900, autosampler A-905, and fraction collector Frac-950 (all GE Healthcare). The instrumentation was controlled with Unicorn 5.31 software. To measure protein absorption spectra inline, a Dionex UltiMate 3000 DAD equipped with a semi-preparative flow cell (0.4 mm optical pathlength) and operated with Chromeleon 6.80 software (Thermo Fisher Scientific) was connected to the Akta Purifier 10. The data acquisition of the DAD was triggered from Unicorn using the following interface: The fact of a sample injection was first submitted to Matlab (MathWorks, Natick, MA) by sending a digital signal to the Matlab Data Acquisition (DAQ) toolbox using the remote connector of the Akta P-900 pump in combination with an USB-6008 DAQ device (National Instruments, Austin, TX). The signal was subsequently forwarded to an Excel Visual Basic for Application (VBA) Macro (Microsoft, Redmond, WA) using the Matlab Spreadsheet Link EX toolbox. The VBA Macro then instructed the Chromeleon software to start raw data acquisition using the Chromeleon Software Developer Kit (Thermo Fisher Scientific). At the end of each experiment, the DAD was triggered to stop data acquisition using the same interface.

Reference analysis of collected fractions was performed using a Dionex UltiMate® 3000 liquid chromatography system (Thermo Fisher Scientific). The system was composed of a HPG-3400RS pump, a WPS-3000TFC-analytical autosampler, a TCC-3000RS column thermostat, and a DAD3000RS detector.

6.3.3 System and Column Characterization

The determination of extra column effects, porosities, and axial dispersion was carried out according to (Altenhöner et al., 1997; Schmidt-Traub, 2005). All dead volumes of the chromatographic instrumentation upstream of conductivity cell and DAD were characterized using pulse injections of 1 %(v/v) acetone solution (Merck, Darmstadt, Germany) and elution buffer. The determined dead volumes calculated from the 280 nm and conductivity signals were used for the correction of all measured raw data. The dead volume between DAD and fraction collector was determined by weighing the liquid inside the capillary and was accounted for as fractionation delay using Unicorn.

Bed and bead porosity were calculated from the retention volumes of 20 μl tracer injections at a flow rate of 0.2 ml/min. Sodium chloride was applied as pore-penetrating, non-interacting tracer, while filtrated 10 g/L dextran 2000 kDa solution was used as non-pore-penetrating, non-interacting tracer (Sigma, St. Louis, MO, USA). All experiments were performed in triplicate. The axial dispersion coefficient D_{ax} was determined by curve fitting of three sodium chloride tracer experiments. Film and pore diffusion effects of sodium chloride were uncoupled by setting the effective film diffusion coefficient k_{eff} to $r_p/3$. For minimization of the sum of squared errors, Levenberg Marquardt algorithm was used.

For the determination of the total ionic capacity Λ, acid-base titration was performed. The SP sepharose FF column was first flushed with 0.1 M hydrochloric acid and subsequently washed using ultra pure water. Both steps were performed until constant UV and conductivity signals were achieved. The column was then titrated with 0.1 M sodium hydroxide until an increase in the conductivity signal was detected. Λ was eventually calculated from the Na^+ concentration of the applied titrant and its volume. Flow rate was 0.2 ml/min for all performed steps. The titration experiment was performed in triplicate.

6.3.4 Linear Gradient Elutions

For each run, the SP Sepharose FF column was first equilibrated with 5 column volumes (CVs) of loading buffer and then loaded with 0.4 mg of each protein. Simultaneously, the DAD was triggered to start data acquisition and recorded protein spectra in the band of 240-300 nm with 1 nm resolution. A wash of 5 CVs loading buffer was performed subsequently and the proteins were eluted by a linear gradient of 3 CVs with elution buffer. The gradient steepness was varied using different sodium chloride concentrations in loading and elution buffer: 250 to 400 mM, 150 to 600 mM, 230 to 460 mM, 170 to 540mM, and 200 to 500 mM. Run 1 to 4 were used for calibration of both PLS and mechanistic model. Run 5 was performed for model validation. After linear gradient elution, the column was regenerated for 5 CVs using the respective elution buffer. The flow rate was 0.2 mL/min for all steps and experiments. The elution peaks were collected in microtiter plates (Greiner BioOne, Kremsmünster, Austria) in 200 μl fractions.

6.3.5 Reference Analysis

As reference analysis, analytical cation exchange chromatography was performed with the collected fractions. For each fraction, the Proswift SCX-1S 4.6 x 50 mm column was first

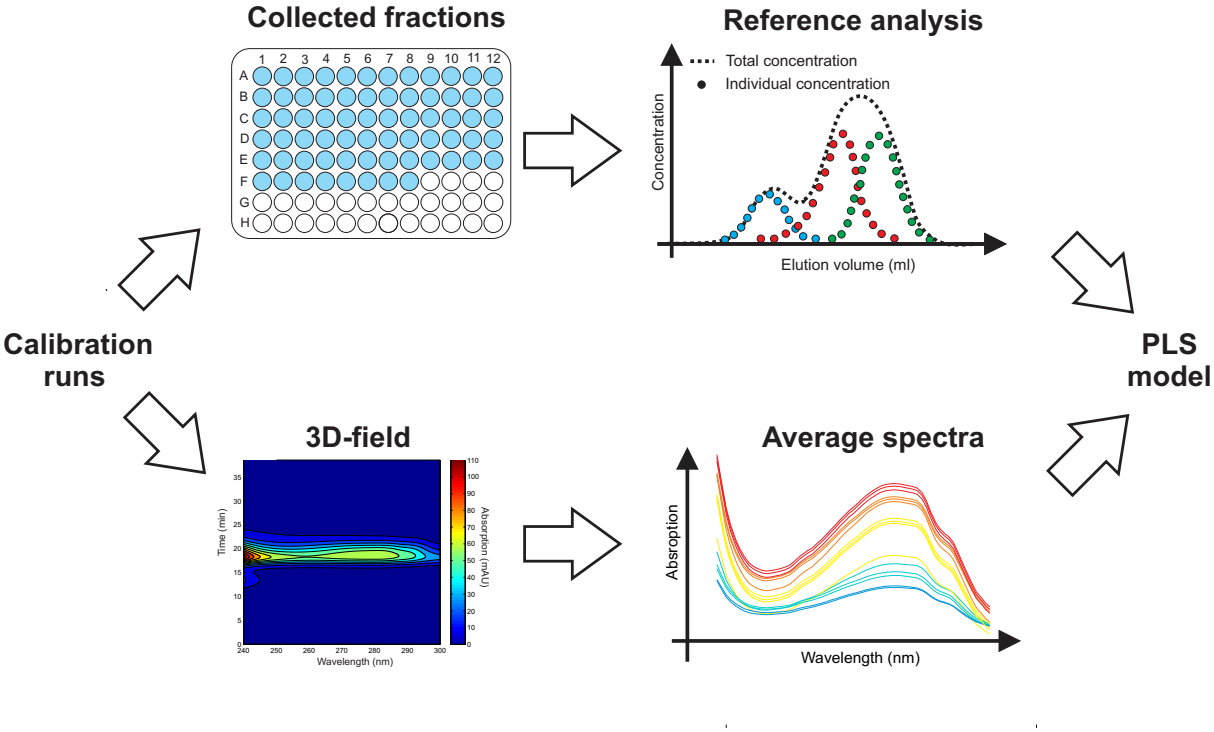


Figure 6.1: PLS model calibration: For each chromatography run employed for PLS model calibration, 200 μl fractions were collected and analyzed by reference analysis. In addition, average absorption spectra corresponding to the collected fractions were calculated from the 3D-field of the DAD. The average spectra and the results of reference analysis were eventually correlated using PLS technique.

equilibrated for 2 min with loading buffer, loaded with 20 μl sample, washed for 0.5 min with loading buffer and then eluted with a piecewise linear salt gradient: 0-10 % elution buffer in 2 min, 10-100 % elution buffer in 2 min, and regeneration with 100 % elution buffer for 1 min. Flow rate was 1.5 mL/min for all steps.

6.3.6 PLS Model Calibration and Validation

For the calibration of a PLS model, average spectra corresponding to each collected fraction were calculated from the recorded time, wavelength, and absorption 3D-field of the DAD as displayed in Figure 6.1. The calculated average spectra were correlated with the determined concentrations from reference analysis using PLS technique. PLS modelling was performed using Simca (Umetrics, Umeå, Sweden). Simca applies the NIPALS-algorithm for PLS. As training data set, the spectra and concentrations of all fractions from run 1-4 were applied. Before performing PLS, all spectra were preprocessed by mean centering. The optimal number of LVs for the PLS model was determined via cross validation (leave one out)(Eriksson et al., 2006). The PLS model was finally validated by application to an inline peak deconvolution of run 5. This run exhibited different mixing ratios and concentrations levels of the proteins than the runs applied for model calibration. In order to perform an inline peak deconvolution, the PLS model was applied to evaluate the 3D-field of the DAD instead of using average spectra. Consequently, the spectrum at each time point was applied to predict selective protein concentrations using the PLS model. Inline peak deconvolution was performed in Matlab.

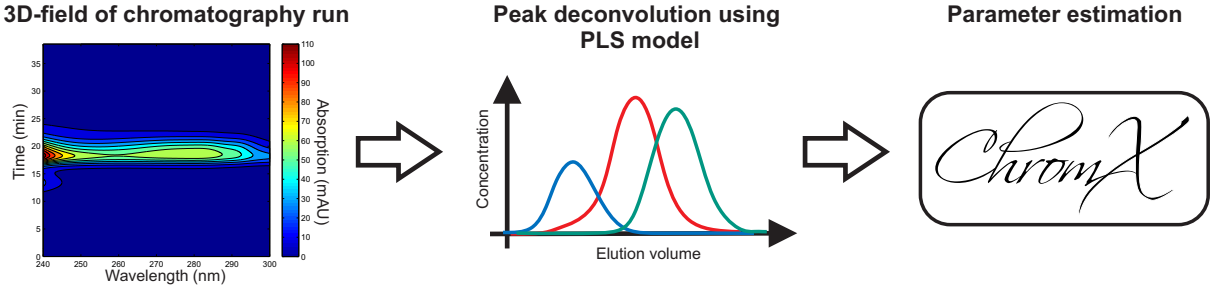


Figure 6.2: Parameter estimation for mechanistic modelling and for model-based root cause investigation: For each chromatography run, the 3D-field of the DAD was evaluated by the PLS model to obtain a peak deconvolution. The selective protein concentration profiles were subsequently imported into ChromX for parameter estimation.

6.3.7 Parameter Estimation for Mechanistic Modelling

Parameter estimation for mechanistic modelling was carried out using the in-house developed software ChromX. The effective mass transfer coefficients k_{eff} of rib A and cyt c were determined by curve fitting of three protein pulse injections (20 μl sample) under non-binding conditions (600 mM sodium chloride) at 0.2 ml/min. For minimization of the sum of square errors, Levenberg Marquardt algorithm was used. As lys exhibited interactions with the adsorber at 600 mM sodium chloride, the average of k_{eff} determined for rib A and cyt c was applied for this protein. The SDM isotherm parameters k_{eq} and ν were estimated from the deconvoluted chromatograms of run 1-4. As displayed in Figure 6.2, the PLS model was first applied to the 3D-field of each run to determine the individual elution profile of each protein. The deconvoluted chromatograms were subsequently imported into ChromX and the isotherm parameters for each protein were estimated by curve fitting. For the minimization of the sum of square errors, Adaptive Simulated Annealing algorithm was used followed by Levenberg Marquardt algorithm for refinement.

6.3.8 Root Cause Investigation

To imitate process variation or experimental errors, the process parameters flow rate and Na^+ concentration in loading and elution buffer were manipulated. Flow rate was varied between 0.18 and 0.22 ml/min, the Na^+ concentration in the loading buffer between 0.21 and 0.25 M, and the Na^+ concentration in the elution buffer between 0.48 and 0.58 M. The combinations of all errors were examined systematically by applying a Central Composite Design (cf. Figure 6.3). Except for the manipulated process parameters, the method described in section 6.3.4 was applied. For each combination of experimental errors, the PLS model was used to obtain selective protein concentrations as displayed in Figure 6.2. The selective protein concentrations were subsequently imported into ChromX and the process parameters (flow rate, Na^+ concentration in loading and elution buffer) were estimated by curve fitting. For the minimization of the sum of square errors, Adaptive Simulated Annealing algorithm followed by Levenberg Marquardt algorithm for refinement was used.

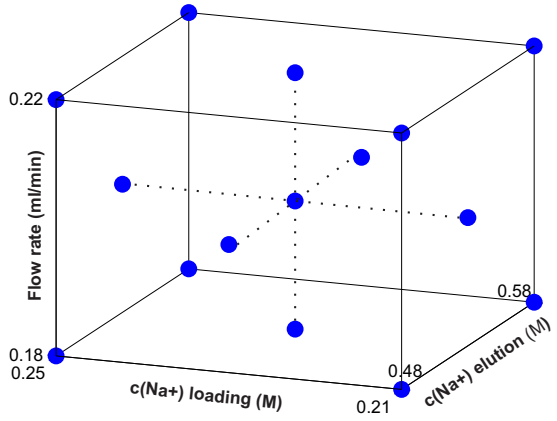


Figure 6.3: Central Composite Design for the systematical examination of experimental errors in the flow rate as well as in the Na^+ concentration in loading and elution buffer.

6.4 Results and Discussion

6.4.1 System and Column Characterization

The column porosity ε_b and the bead porosity ε_p were calculated from the retention volumes of sodium chloride and 2000 kDa dextran tracer injections. For ε_b, a value of 0.274 was determined which is very close to the value for hexagonal close packing (0.26). The determined value for ε_p of 0.854 is in well in agreement with observations of other researchers for SP Sepharose FF (DePhillips and Lenhoff, 2000; Nash and Chase, 1998). The axial dispersion coefficient D_{ax} was determined by curve fitting of three sodium chloride tracer injections assuming no mass transfer hindrance. The total ionic capacity Λ was measured by acid-base titration. All raw data was corrected by the determined system dead volumes. The results of column characterization are summarized in Table 6.1.

The result of curve fitting for the determination of D_{ax} is displayed in Figure 6.4. Good agreement was found between the simulated conductivity with the estimated D_{ax} and the measured conductivity traces of sodium chloride injections. Solely the slight fronting in the experimental data was not described by the simulation. This is to be expected as the slight fronting resulted from an unequally packed bed which is not accounted for in the simulation.

Table 6.1: Determined column parameters ε_b, ε_p, D_{ax}, and Λ as well as method of parameter determination. Λ is specified as mol sodium ions per L of resin backbone. All column parameters were determined in triplicate with no measurable difference between experiments. For curve fitting, estimated confidence intervals according to 6.2.2 are presented.

	Method of determination	Result
ε_b	Tracer retention volumes	0.274
ε_p	Tracer retention volumes	0.854
D_{ax}	Curve fitting	0.022 +/- 0.00024 $\frac{mm^2}{s}$
Λ	Acid base titration	2.932 M

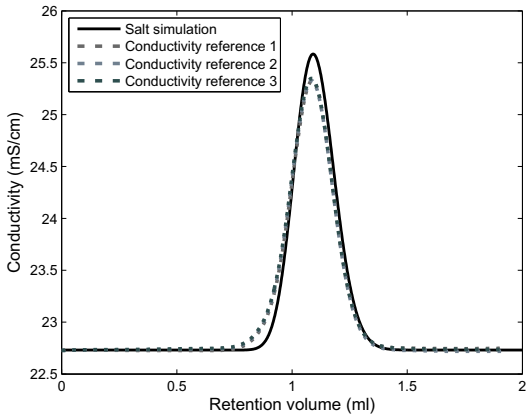


Figure 6.4: Result of D_{ax} estimation by curve fitting: Simulated conductivity in comparison with results of three sodium chloride tracer injections.

6.4.2 PLS Model Calibration and Validation

For PLS model calibration, the results of fraction analysis by analytical CEX chromatography of all calibration runs were correlated with the corresponding average mid-UV absorption spectra using PLS. The optimal number of latent variables was determined by cross validation (leave one out). This means that one sample is excluded during model calibration and the prediction error for this sample is calculated subsequently. For different numbers of LVs, this procedure is performed until each sample was excluded. For three latent variables, a minimal root mean square error of 0.018 g/L for rib A, 0.014 g/L for cyt c, and 0.015 g/L for lys was found.

The calibrated PLS model with three latent variables was subsequently applied to predict the selective protein concentrations for all calibration runs and for the validation run. For model prediction, the PLS model was applied to evaluate the 3D-field of the DAD instead of using average spectra. Figure 6.5 displays the comparison between this inline peak deconvolution (solid lines) and the results of offline reference analysis (dashed stair functions) for all calibration runs. Figure 6.6 A shows the same comparison for the validation run. A good agreement between the prediction by the PLS model and the reference analysis for all calibration runs and the validation run was observed. This proves a high precision of the calibrated PLS model. For both shallower gradients (cf. Fig. 6.5 A and C), the PLS model predicted a small rib A peak in the peak flank of cyt c. This peak was not detected in offline analysis and is most likely an artifact of the PLS model. A possible explanation for this peak is an isoform of cyt c with a spectrum similar to rib A. For shallower gradients, this isoform might be more resolved and might cause the small rib A peak in the peak flank of cyt c. However, the maximum of this peak was only 0.02 g/L. The rib A artifact peak might be avoided by considering all cyt c isoforms individually during PLS model calibration. In addition to cyt c, rib A exhibited several isoforms. The rib A isoforms were resolved during isocratic elution in the experiments with higher sodium chloride content in the loading buffer (cf. Fig. 6.5 A and C). The spectra of these isoforms are however very similar and allowed for PLS modelling of Rib A as one species.

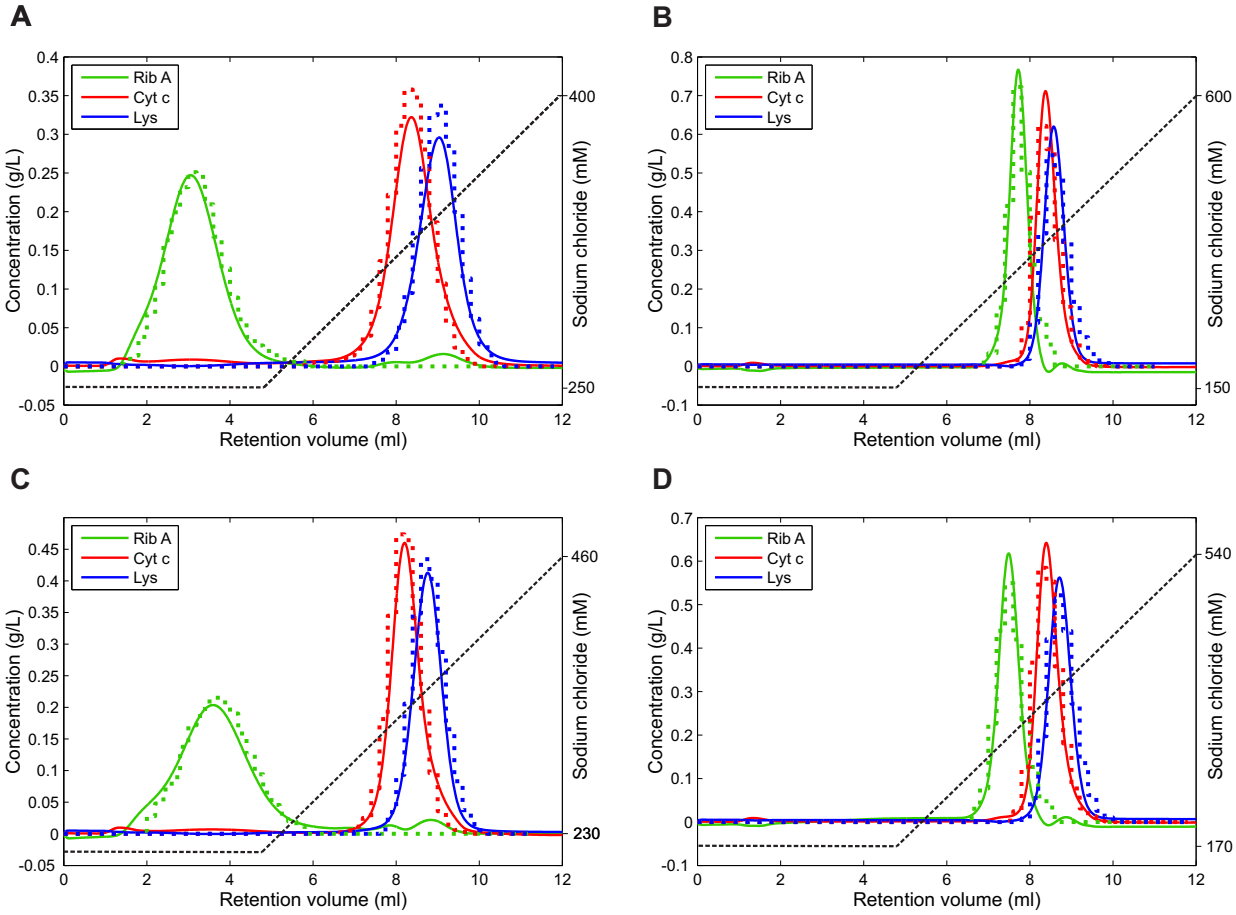


Figure 6.5: PLS model prediction for all calibration runs (solid lines) in comparison with reference analysis (dashed stair functions). Gradient steepness in calibration runs was varied by applying different sodium chloride concentrations in loading and elution buffer. A: 250 to 400 mM, B: 150 to 600 mM, C: 230 to 460 mM, D: 170 to 540 mM.

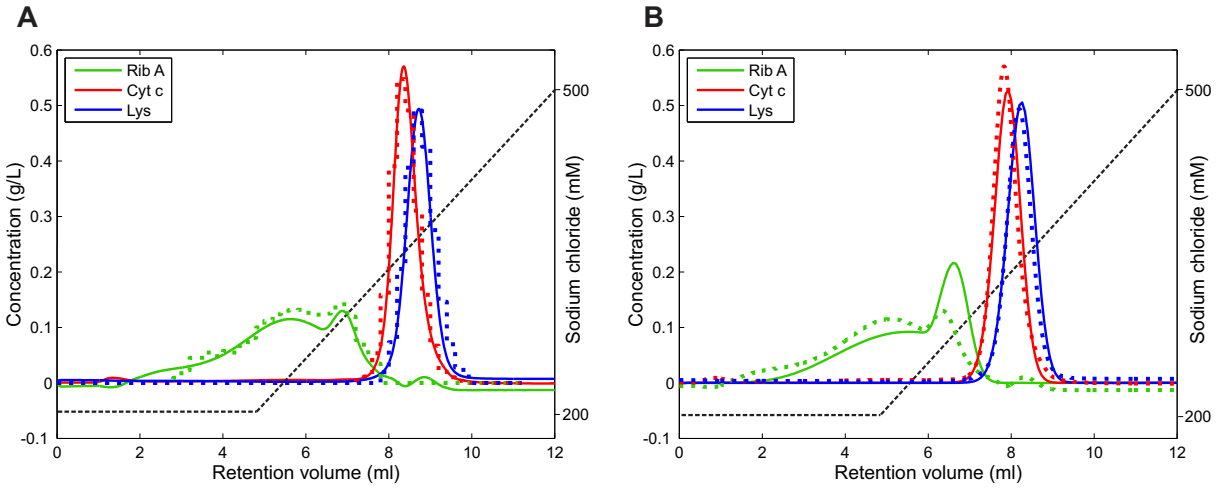


Figure 6.6: Valdiation run (Gradient from 200 mM sodium chloride in loading to 500 mM in elution buffer). A: PLS model prediction (solid lines) in comparison with reference analysis (dashed stair functions). B: Mechanistic model prediction (solid line) in comparison with PLS model inline peak deconvolution (dashed line).

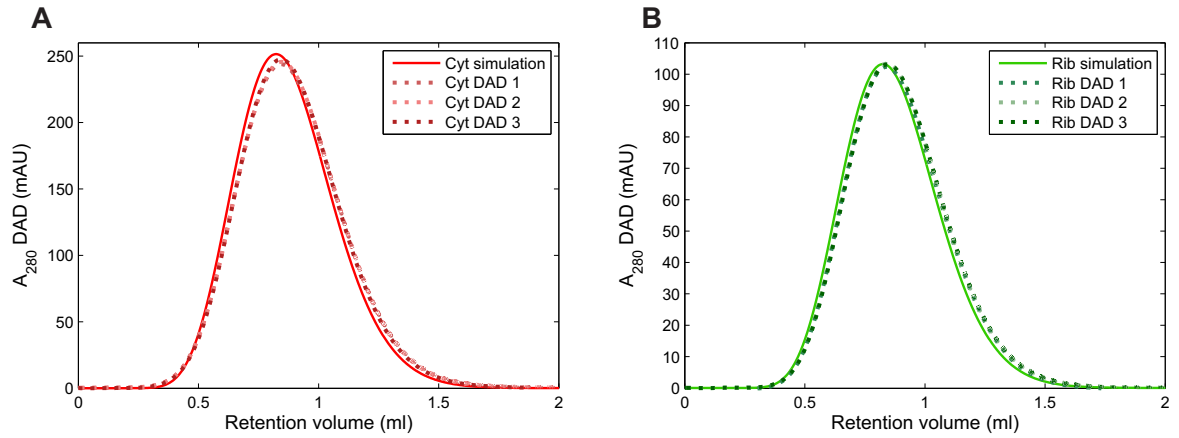


Figure 6.7: Result of k_{eff} estimation by curve fitting: A: Simulated A_{280} signal of cyt c and measured A_{280} signals at the DAD, B: Simulated A_{280} signal of rib A and measured A_{280} signals at the DAD. All absorption values are specified for a pathlength of 10 mm.

6.4.3 Parameter Estimation for Mechanistic Modelling

The effective mass transfer coefficients k_{eff} of rib A and cyt c were determined by curve fitting of three protein pulse injections under non-binding conditions (600 mM sodium chloride). As lys exhibited interactions with the adsorber at 600 mM sodium chloride, the average of k_{eff} determined for rib A and cyt c was applied for this protein. The resulting fits are presented in Figure 6.7 A for cyt c and 6.7 B for rib A. For both proteins, a comparison between simulated A_{280} and measured signal (in triplicate) is shown. The estimated k_{eff} was 0.0015 mm/s for both rib A and cyt c. As the proteins have similar molecular weights (cyt c: 12.384 kDa and rib A: 13.700 kDa) (nearly) equal mass transfer resistance is to be expected. This also justifies the assumption to use a k_{eff} of 0.0015 mm/s for lys with a molecular weight of 14.300 kDa. The upper limit of k_{eff} is given by $3/r_p$ and is 0.015 mm/s for SP Sepharose FF. The SDM isotherm parameters k_{eq} and ν were estimated from the deconvoluted chromatograms of the calibration runs by curve fitting. All estimated parameters as well as an estimate for their confidence intervals according to section 6.2.2 are summarized in Table 6.2.

While for all three proteins similar charges between 6.1 and 6.5 were estimated, k_{eq} was varying significantly with $8.8 \cdot 10^{-6}$ for rib A, $2.1 \cdot 10^{-5}$ for cyt c, and $5.3 \cdot 10^{-5}$ for lys. Variance in k_{eq} was correlated with the proteins' elution order. This is to be expected as k_{eq} corresponds to the slope of the isotherm and thus correlates with the proteins' affinity to the resin. A possible explanation for similar characteristic charges ν_i might be a large

Table 6.2: Estimated effective mass transfer coefficients k_{eff} and SDM parameters k_{eq} and ν for all proteins. Parameters were determined by curve fitting of protein tracer injections and linear gradient elutions. Confidence intervals for all parameters were estimated according to 6.2.2.

	k_{eff} $[mm/s]$		k_{eq} [-]		ν [-]	
Rib A	0.0015	$+/-0.00016$	$8.8 \cdot 10^{-6}$	$+/- 8.0 \cdot 10^{-8}$	6.1	$+/-0.002$
Cyt c	0.0015	$+/-0.00012$	$2.1 \cdot 10^{-5}$	$+/- 1.6 \cdot 10^{-7}$	6.5	$+/-0.002$
Lys	0.0015	not available	$5.3 \cdot 10^{-5}$	$+/- 2.2 \cdot 10^{-7}$	6.3	$+/-0.0008$

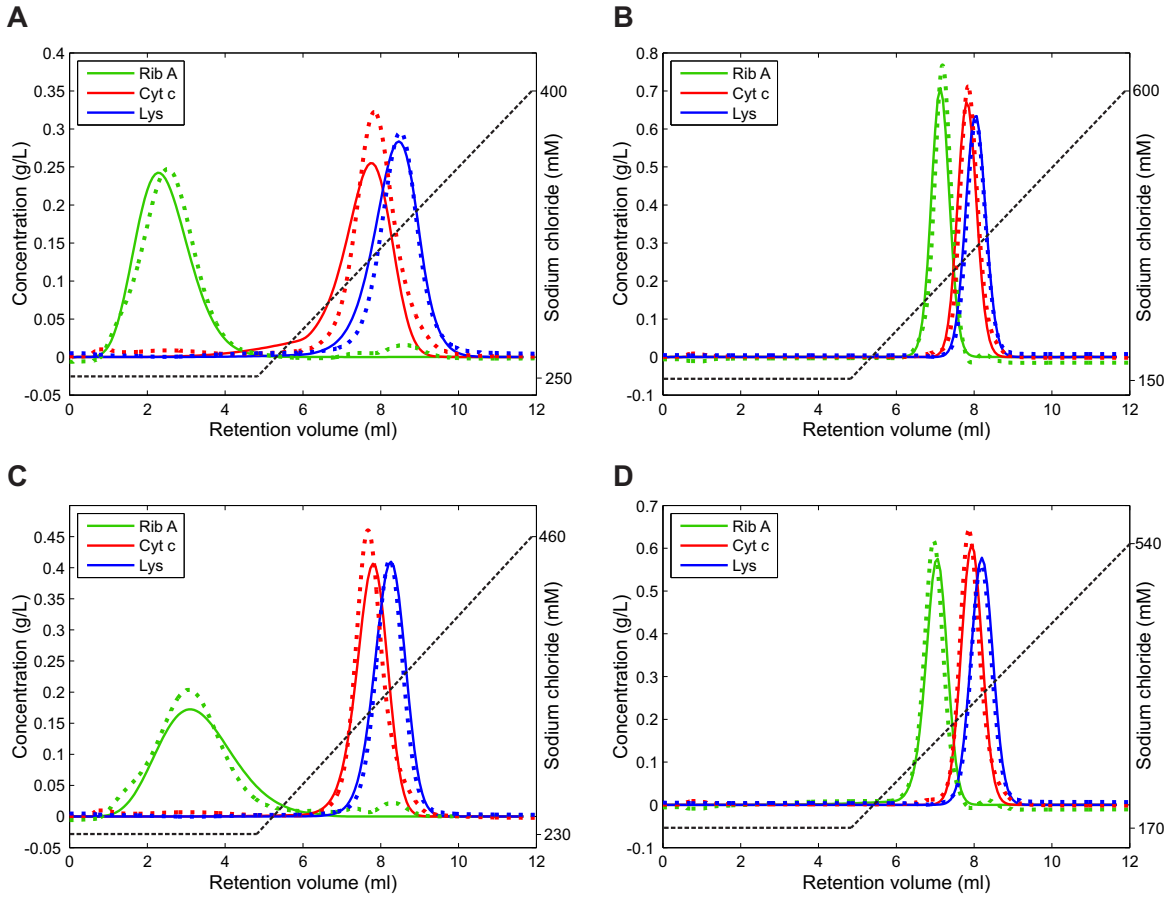


Figure 6.8: Results of SDM parameter determination by curve fitting of calibration runs: Simulations using mechanistic model (solid lines) are displayed in comparison with the PLS model inline peak deconvolution (dashed lines). Gradient steepness in calibration runs was varied by applying different sodium chloride concentrations in loading and elution buffer. A: 250 to 400 mM, B: 150 to 600 mM, C: 230 to 460 mM, D: 170 to 540 mM.

distance of the operating pH (5) to the pI of the three proteins applied (rib A, 9.6; cyt c, 10.5, lys: 11.4). However, more data is needed to actually prove this behavior.

The resulting fits of the SDM parameter estimation are displayed in Figure 6.8. All simulations using the mechanistic model (solid lines) are shown in comparison with the corresponding PLS model inline peak deconvolutions (dashed lines). The precision of the mechanistic model was demonstrated by simulating the protein elution profiles in the validation run. Results are presented in Figure 6.6 B. The determined SDM parameters (cf. Table 6.2) lead to a good agreement between the simulation by mechanistic model and the inline peak deconvolution reference. The highest conformity between simulation and reference for calibrations runs and the validation was observed for lys. Lys was also the purest protein without any isoforms. For cyt c, deviation between simulation and reference was larger for the two shallow gradients (cf. Fig. 6.8 A and C). This might be associated with the fact that cyt c exhibited several isoforms which were more resolved in these experiments. Similar behavior was found for rib A, where isoforms where resolved during isocratic or partially isocratic elution (cf. Fig. 6.8 A and C and Fig. 6.6 B). Mechanistic model precision might be further improved by considering all isoforms of cyt c and

rib A separately. However in general, prediction precision of the mechanistic model was high as demonstrated by the validation run and model precision was sufficient to perform a model-based root cause investigation.

6.4.4 Model-based Root Cause Investigation

For the model-based root cause investigation, experimental errors in the process parameters flow rate and Na^+ concentration in loading and elution buffer were caused deliberately. The combinations of all errors were examined systematically by applying a Central Composite Design. For each experiment, peak deconvolution was performed by applying the PLS model to evaluate the 3D-field of the DAD. The actual values of the three process parameters were subsequently estimated by curve fitting of each deconvoluted chromatogram using the mechanistic model. The results for the Na^+ concentration in the loading buffer are presented in Figure 6.9 A, the results for the Na^+ concentration in the elution buffer in 6.9 B and the results for the flow rate in 6.9 C.

For each experiment in the Central Composite Design, the absolute percental error between the estimated value of a process parameter and its actual value in the experiment is shown. The estimation of the Na^+ concentration in the loading buffer was mainly determined by the curve fitting of rib A. Here, the highest deviations of 2.6-3.5 % were observed for 0.23 M Na^+ in the loading buffer. This is in compliance with the observed deviation of the mechanistic model prediction for rib A in the validation run. In addition, two experiments with 0.21 M Na^+ in the loading buffer and 0.48 M Na^+ in the elution buffer exhibited higher deviations of 3.4 %, indicating a lower precision of the mechanistic and/or the PLS model for this parameter combination.

The estimation of the flow rate by curve fitting was determined by the retention times of the proteins and exhibited the lowest errors of 2 % or less. This is to be expected as the retention time is not influenced by the precision of the PLS model (prediction of exact concentrations). The estimation of the Na^+ concentration in the elution buffer was dominated by both retention times and peak heights of the proteins during gradient elution. This leads to larger deviations between error diagnostics and the actual values for both the flow rate and the Na^+ concentration in the elution buffer, if a peak height is not correctly predicted by either PLS or mechanistic model. This effect was observed for the experiment with a process parameter combination 0.21 M, 0.48 M, 0.22 ml/min. Here, the deviation between error diagnostics and actual value was 4.4 % for the Na^+ concentration in elution buffer and 2 % for the flow rate. However regarding standard deviations of other analytical technologies, the observed deviations are acceptable for all three process parameters examined.

6.5 Conclusion and Outlook

This study demonstrates that a combination of spectral deconvolution and inverse mechanistic modelling can be applied as a tool for root cause investigation in protein chromatography. The tool was capable of identifying the causes for a deviating chromatogram. This was achieved by inverse modelling with a mechanistic chromatography model. Parameters which can cause deviations were altered systematically until the simulated chro-

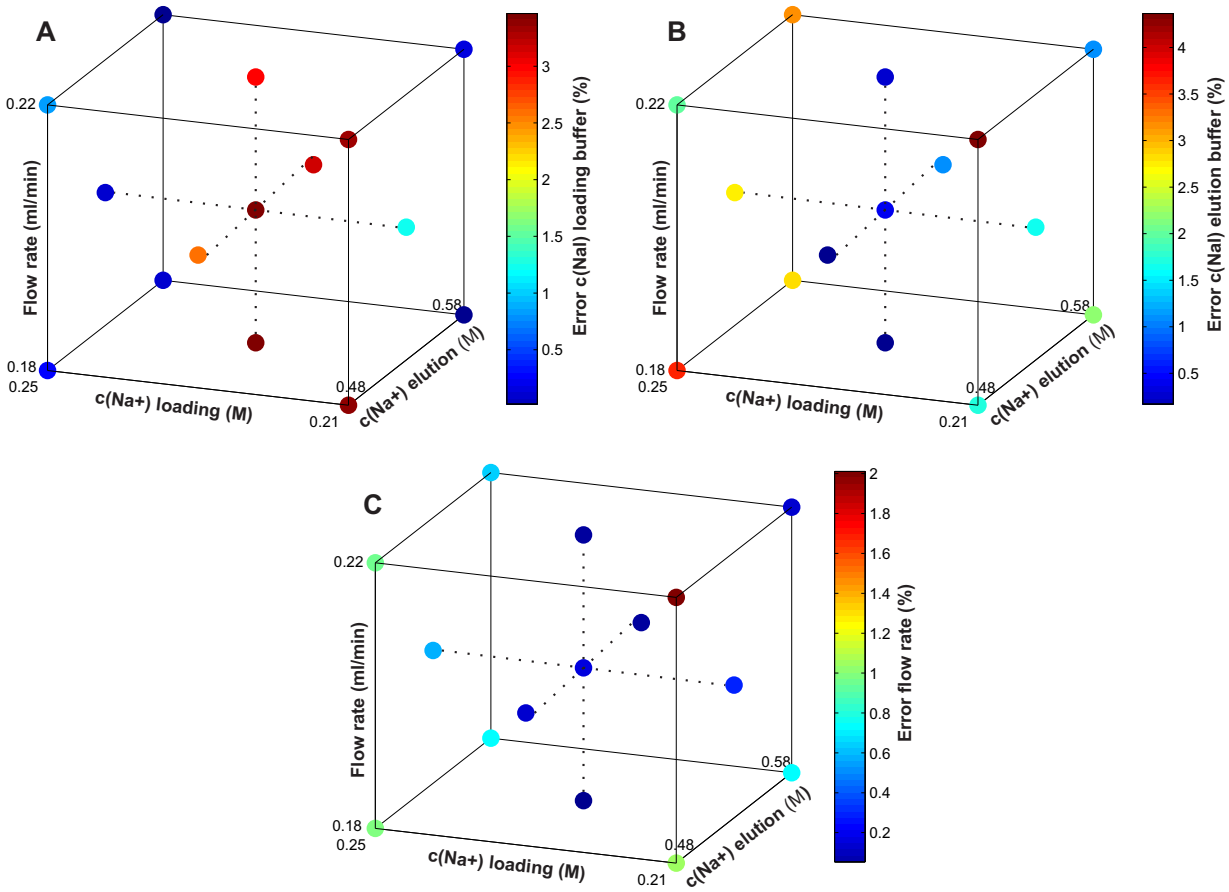


Figure 6.9: Results of the model based root cause investigation for the Na^+ concentration in the loading buffer (A), for the Na^+ concentration in the elution buffer (B), and for the flow rate (C). For each experiment in the Central Composite Design, the absolute percental error between the estimated value of a process parameter and its actual value in the experiment is shown.

matogram matched the observed one. As this method required molar concentrations of all involved proteins, PLS modelling with mid-UV absorption spectra was applied to obtain the individual elution profiles. The functionality of the tool for root cause investigation was successfully demonstrated in a case study with lysozyme, cytochrome c, and ribonuclease A which were co-eluting on a SP-Sepharose FF column. Deviations in chromatograms were generated by deliberately caused errors in the parameters flow rate and sodium-ion concentration in loading and elution buffer according to a Central Composite Design. The combination of spectral deconvolution and inverse mechanistic modelling allowed for a correct estimation of the actual values of all parameters. The largest deviation between an estimation and the applied value in an experiment was 4.4 %. Thus, the established tool for root cause investigation can reliably identify causes for deviating chromatograms. This allows for the correction of a detected problem in the next cycle or batch. Consequently, the tool has a great potential for a feed-back control of chromatography processes. Before an application in process control, the sensitivity of the tool for error diagnostics, which is influenced by both PLS and mechanistic model precision, needs to be evaluated carefully. In this study, even small deviations in the three examined process parameters were detectable. For future work, the tool should be tested for the identification of changes in column capacity or porosity caused by column aging and

should be developed for other modes of chromatography than ion exchange.

Acknowledgment

This work was supported by a grant from the Ministry of Science, Research, and the Arts of Baden-Württemberg, Germany (Az.: 33-7533-7-11.6-2).

7 | Conclusion and Outlook

This thesis addressed the lack of suitable analytical technologies for the real-time process monitoring and control in the downstream processing of therapeutic proteins. As chromatography is the widest applied unit operation in downstream processing, the objective of this thesis was to develop PAT for chromatography based protein purification. The main focus was to solve the issue of process variability in protein manufacturing by developing technologies that can both monitor critical quality and performance attributes and keep these attributes in a certain predefined range. In this context, a PAT tool for real-time monitoring and control of the load and elution phase in chromatography was developed. Another challenge addressed in this thesis was the investigation of causes for variability. As chromatographic processes are very complex and react highly non-linear to changes in the input, this challenge was dealt with mechanistic chromatography modelling.

In the first part of the thesis, a PAT tool for selective inline quantification of proteins was developed. It consists of a DAD, a PLS model, and customized interfaces between different types of software to enable real-time data evaluation and process control. Instead of performing univariate absorption measurements at 280 nm, which is the industrial standard, the DAD acquires UV absorption spectra. These spectra were correlated with selective protein concentrations in a PLS model. After a calibration and validation, the PLS model was applied to monitor selective protein concentrations in a chromatographic separation in real-time. By this way, the peaks of co-eluting proteins were deconvoluted into the elution profiles of all contributing species. Based on the achieved peak deconvolution, a control of the elution phase was performed. The peaks were cut in real-time using product-purity criterions.

In contrast to the commonly used process control based on the absorption at 280 nm, the newly developed PAT tool also allows a process control in cases, where the peaks of product and contaminants are not well separated. As the process control is fully automated, human errors as well as the personnel deployment might be minimized compared to the atline control strategies proposed in literature. The flow-cell for the spectroscopic measurement could even be made in a disposable fashion and no additional risk of contamination is to be expected for this method. This is an advantage compared to automated sampling strategies such as online chromatography. Additionally, the acquisition of UV absorption spectra is fast and hence very suitable for the control of chromatographic processes, where the concentration changes of the contributing species occur rather fast.

A further goal of this thesis was to demonstrate applicability of the PAT tool for real-life separation issues. This required a further development of the PAT tool. A first issue to be addressed was the calibration of the corresponding PLS models using feedstocks

or in-process materials. Hence, a calibration method which is based on process-data was established. This method applied absorption spectra which were acquired during the chromatographic step as well results of an offline analysis based on collected fractions. The absorption spectra were then averaged in time according to the fraction size and correlated with the results of the offline analytics using the PLS technique. This newly developed method allowed the application of the PAT tool for a selective inline quantification of co-eluting mAb monomer, fragments, and aggregates and of co-eluting serum proteins in a chromatographic step of the Cohn process.

Another prerequisite for the applicability for large scale preparative chromatography was the usability of the PAT tool for conditions were the columns are completely loaded. An issue that had to be dealt with in this context was the large concentration difference between product and contaminants. Especially in polishing steps, the product peak is highly concentrated, while co-eluting contaminants in the peak flanks exhibit rather low concentrations. This issue was solved by applying VP spectroscopy. Instead of measuring the UV absorption spectra at one fixed pathlength with a DAD, a novel device for VP measurements was integrated into the PAT tool. The device acquires the slope in the linear range of the dependency of the absorption from the pathlength. Hence, PLS models based on the slope at every wavelength were calibrated and validated. These models allowed a peak deconvolution for separations, where the elution peak consisted of high and low concentrated species.

Besides from monitoring and control of the elution phase, the PAT tool was also successfully applied during the load phase of a capture step. In this context, it was shown that PLS modelling with UV absorption spectra is also applicable for the quantification of mAb in the background of many protein and non protein-based impurities. The method was applied to monitor the mAb breakthrough in the column effluent during the load phase of a protein A capture step. This allowed the termination of the load phase, when a defined mAb concentration in the column effluent was reached.

By applying the PAT tool for the control of the load phase in future capture steps, a better resin capacity utilization might be realized. If the column capacity is variable because of aging effects, the load volume could automatically be adapted such that the actual resin capacity is used. In addition, time-consuming offline determination of the variable product titer previous to capture steps, which are performed to determine the load volume onto the column, could be eliminated. Thus, more efficient processes could be realized. In contrast to the alternative commercially available method, which uses difference signals between two detectors, no stabilized impurity baseline during the load phase is required for the control strategy developed in this thesis. Benefits compared to automated sampling methods comprise a reduced risk of contamination as well as the speed of data acquisition and evaluation.

In a last part of the thesis, the investigation of causes for variability was addressed. Inverse mechanistic chromatography modelling was applied to identify the causes for deviations in chromatograms. Parameters which can cause deviations were altered systematically until the simulated chromatogram matched the experimental one. As inverse modelling requires the concentrations of all contributing species, peak deconvolutions based on PLS modelling with UV absorption spectra were applied prior to mechanistic

modelling. The approach allowed the correct identification of deliberately caused errors in the parameters flow rate and sodium-ion concentration in the loading and elution buffer.

In summary, this thesis not only contributes to solve the issues of process variability in protein chromatography, but also shows a possibility how causes for variability could be investigated. The application of the developed PAT tool in future chromatographic separations could increase the reproducibility and product quality in protein manufacturing. This complies with the requirements of the FDAs PAT concept. Additionally, the control of the load and elution phase of future chromatography steps might contribute to an increased process efficiency and to the realization of continuous processing as well as real-time release. Eventually, the application of inverse mechanistic modelling might be a valuable approach for root cause investigation, if a batch fails. This might enable the correction of the detected problem in the next chromatographic cycle or in the next batch.

For future work, the root cause investigation could be used for a feed-back control in chromatography. If a cause for a deviation in a chromatogram is identified, the controllable parameters of a chromatographic separation could be adapted such that a certain objective for the separation is achieved. Exemplarily, this objective could be the resolution between the product and a contaminant. This is however currently limited by the computational time. The identification of the causes for deviations in a chromatogram requires several minutes. In addition to that, the optimum of an objective function would have have to be found to obtain the optimal parameters for the next cycle, which requires computational time as well.

While this thesis was only based on UV absorption spectroscopy, the examination of other spectroscopic methods as well as data fusion of several techniques might further improve the PAT tool. For instance, fluorescence spectroscopy might be more sensitive. A limitation for this method might however be the sensitivity to changes in the solvent. Infrared spectroscopy detects different chemical groups than UV absorption and fluorescence and is hence an orthogonal method. A benefit of infrared spectroscopy might be an improved selectivity (in the mid-infrared range). Limitations of mid-infrared spectroscopy comprise the strong absorption of water and the required measurement times of several minutes to obtain good signal to noise ratios. While this might be disadvantageous for chromatography, an application for ultra-/diafiltration, where the changes over time are slower, might be thinkable.

References

Aboulaich N, Chung WK, Thompson JH, Larkin C, Robbins D, Zhu M. 2014. A novel approach to monitor clearance of host cell proteins associated with monoclonal antibodies. Biotechnol Prog 30:1114–1124.

Altenhöner U, Meurer M, Strube J, Schmidt-Traub H. 1997. Parameter estimation for the simulation of liquid chromatography. J Chromatogr A 769:59–69.

Angarita M, Müller-Späth T, Baur D, Lievrouw R, Lissens G, Morbidelli M. 2015. Twin-column capture SMB: A novel cyclic process for protein A affinity chromatography. J Chromatogr A 1389:85–95.

Arnold SA, Gaensakoo R, Harvey LM, McNeil B. 2002. Use of at-line and in-situ near-infrared spectroscopy to monitor biomass in an industrial fed-batch escherichia coli process. Biotechnol Bioeng 80:405–413.

Balestrieri C, Colonna G, Giovane A, Irace G, Servillo L. 1978. Second-derivative spectroscopy of proteins. A method for the quantitative determination of aromatic amino acids in proteins. Eur J Biochem 90:433–440.

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

Bauer I, Bock HG, Körkel S, Schlöder JP. 2000. Numerical methods for optimum experimental design in DAE systems. Comput Appl Math 120:1–25.

Baumann P, Hahn T, Hubbuch J. 2015. High-throughput micro-scale cultivations and chromatography modeling: Powerful tools for integrated process development. Biotechnol Bioeng 112:2123–2133.

Baumann P, Huuk T, Hahn T, Osberghaus A, Hubbuch J. 2016. Deconvolution of high-throughput multicomponent isotherms using multivariate data analysis of protein spectra. Eng Life Sci 16:194–201.

Beaven G, Holiday E. 1951. Ultraviolet absorption spectra of proteins and amino acids. Adv Prot Chem 7:319–386.

Binns C. 2010. Introduction to nanoscience and nanotechnology. Wiley Survival Guides in Engineering and Science, Wiley.

Boehl D. 2003. Chemometric modelling with two-dimensional fluorescence data for claviceps purpurea bioprocess characterization. J Biotechnol 105:179–188.

Bracewell DG, Gill A, Hoare M, Lowe PA, Maule CH. 1998. An optical biosensor for real-time chromatography monitoring: Breakthrough determination. Biosens Bioelectron 13:847–853.

Brestrich N, Briskot T, Osberghaus A, Hubbuch J. 2014. A tool for selective inline quantification of co-eluting proteins in chromatography using spectral analysis and partial least squares regression. Biotechnol Bioeng 111:1365–1373.

Brestrich N, Hahn T, Hubbuch. 2016. Application of spectral deconvolution and inverse mechanistic modelling as a tool for root cause investigation in protein chromatography. J Chromatogr A 1437:158–167.

Brestrich N, Sanden A, Kraft A, McCann K, Bertolini J, Hubbuch J. 2015. Advances in inline quantification of co-eluting proteins in chromatography: Process-data-based model calibration and application towards real-life separation issues. Biotechnol Bioeng 112:1406–1416.

Broly H, Mitchell-Logean C, Costioli MD, Guillemot-Potelle C. 2010. Cost of goods modeling and quality by design for developing cost-effective processes. Biopharm Int 23:26–35.

Brooks CA, Cramer SM. 1992. Steric mass-action ion exchange: Displacement profiles and induced salt gradients. AIChE J 38:1969–1978.

Browne S, Al-Rubeai M. 2007. Selection methods for high-producing mammalian cell lines. Trends Biotechnol 25:425–432.

Capito F, Skudas R, Kolmar H, Hunzinger C. 2013a. Mid-infrared spectroscopy-based antibody aggregate quantification in cell culture fluids. Biotechnol J 8:912–917.

Capito F, Skudas R, Kolmar H, Hunzinger C. 2015a. At-line mid infrared spectroscopy for monitoring downstream processing unit operations. Process Biochem 50:997–1005.

Capito F, Skudas R, Kolmar H, Stanislawski B. 2013b. Host cell protein quantification by fourier transform mid infrared spectroscopy (FT-MIR). Biotechnol Bioeng 110:252–259.

Capito F, Skudas R, Stanislawski B, Kolmar H. 2013c. Matrix effects during monitoring of antibody and host cell proteins using attenuated total reflection spectroscopy. Biotechnol Prog 29:265–274.

Capito F, Zimmer A, Skudas R. 2015b. Mid-infrared spectroscopy-based analysis of mammalian cell culture parameters. Biotechnol Prog 31:578–584.

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

Challa S, Potumarthi R. 2013. Chemometrics-based process analytical technology (PAT) tools: Applications and adaptation in pharmaceutical and biopharmaceutical industries. Appl Biochem Biotechnol 169:66–76.

Chase HA. 1986. Rapid chromatographic monitoring of bioprocesses. Biosensors 2:269–286.

Chi EY, Krishnan S, Randolph TW, Carpenter JF. 2003. Physical stability of proteins in aqueous solution: Mechanism and driving forces in nonnative protein aggregation. Pharmaceut Res 20:1325–1336.

Chirino AJ, Mire-Sluis A. 2004. Characterizing biological products and assessing comparability following manufacturing changes. Nat Biotechnol 22:1383–1391.

Cramer H. 1957. Mathematical methods of statistics. Princeton University Press.

Cramer SM, Holstein MA. 2011. Downstream bioprocessing: Recent advances and future promise. Curr Opin Chem Eng 1:27–37.

Crommelin DJ, Storm G, Verrijk R, de Leede L, Jiskoot W, Hennink WE. 2003. Shifting paradigms: Biopharmaceuticals versus low molecular weight drugs. Int J Pharm 266:3–16.

Danzon PM, Furukawa MF. 2006. Prices and availability of biopharmaceuticals: An international comparison. Health Aff 25:1353–1362.

Degerman M, Westerberg K, Nilsson B. 2009a. Determining critical process parameters and process robustness in preparative chromatography – A model-based approach. Chem Eng Technol 32:903–911.

Degerman M, Westerberg K, Nilsson B. 2009b. A model-based approach to determine the design space of preparative chromatography. Chem Eng Technol 32:1195–1202.

DePhillips P, Lenhoff AM. 2000. Pore size distributions of cation-exchange adsorbents determined by inverse size-exclusion chromatography. J Chromatogr A 883:39–54.

DeVries JH, Gough SCL, Kiljanski J, Heinemann L. 2015. Biosimilar insulins: A european perspective. Diabetes Obes Metab 17:445–451.

Diederich P, Hansen SK, Oelmeier SA, Stolzenberger B, Hubbuch J. 2011. A sub-two minutes method for monoclonal antibody-aggregate quantification using parallel interlaced size exclusion high performance liquid chromatography. J Chromatogr A 1218:9010–9018.

Donovan JW. 1969. Chemistry and metabolism of macromolecules: Changes in ultraviolet absorption produced by alteration of protein conformation. J Biol Chem 244:1961–1967.

Elshereef R, Budman H, Moresoli C, Legge RL. 2010. Monitoring the fractionation of a whey protein isolate during dead-end membrane filtration using fluorescence and chemometric methods. Biotechnol Prog 26:168–178.

Emerton DA. 2013. Profitability in the biosimilars market. Bioprocess Int 11:6–23.

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

European Commission. 2008. DGC. Pharmaceutical Sector Inquiry Preliminar Report. Technical report.

Fahrner RL, Blank GS. 1999a. Real-time control of antibody loading during protein A affinity chromatography using an on-line assay. J Chromatogr A 849:191–196.

Fahrner RL, Blank GS. 1999b. Real-time monitoring of recombinant antibody breakthrough during protein A affinity chromatography. Biotechnol Appl Biochem 29:109–112.

Fahrner RL, Lester PM, Blank GS, Reifsnyder DH. 1998. 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 827:37–43.

FDA. 2004. Guidance for industry. PAT – A framework for innovative pharmaceutical development, manufacturing, and quality assurance.

Fernández BC, Martínez-Hurtado JL. 2012. Biosimilars: Company strategies to capture value from the biologics market. Pharmaceuticals 5:1393–1408.

Flatman S, Alam I, Gerard J, Mussa N. 2007. Process analytics for purification of monoclonal antibodies. J Chromatogr B 848:79 –87.

Flowers P, Callender S. 1996. Variable path length transmittance cell for ultraviolet, visible, and infrared spectroscopy and spectroelectrochemistry. Anal Chem 68:199–202.

Follman DK, Fahrner RL. 2004. Factorial screening of antibody purification processes using three chromatography steps without protein A. J Chromatogr A 1024:79–85.

Genetic Engineering and Biotechnology News. 2013. http://www.genengnews.com/insight-and-intelligence/top-20-best-selling-drugs-of-2012/77899775/?page=2. Status as of February 2016.

Glassey J, Gernaey KV, Clemens C, Schulz TW, Oliveira R, Striedner G, Mandenius CF. 2011. Process analytical technology (PAT) for biopharmaceuticals. Biotechnol J 6:369–377.

Godawat R, Konstantinov K, Rohani M, Warikoo V. 2015. End-to-end integrated fully continuous production of recombinant monoclonal antibodies. J Biotechnol 213:13–19.

Goodwin, Graham C, Payne RL. 1977. Dynamic system identification: Experiment design and data analysis. Academic Press.

Guiochon G, Shirazi S, Katti A. 1994. Fundamentals of preparative and nonlinear chromatography. Academic Press.

Gupta M (Ed.). 2013. Methods for affinity-based separations of enzymes and proteins. Birkhäuser Basel.

Haack MB, Eliasson A, Olsson L. 2004. On-line cell mass monitoring of saccharomyces cerevisiae cultivations by multi-wavelength fluorescence. J Biotechnol 114:199–208.

Hahn R, Shimahara K, Steindl F, Jungbauer A. 2006. Comparison of protein A affinity sorbents III. Life time study. J Chromatogr A 1102:224–231.

Hahn T, Huuk T, Heuveline V, Hubbuch J. 2015. Simulating and optimizing preparative protein chromatography with ChromX. J Chem Educ 92:1497–1502.

Hahn T, Sommer A, Osberghaus A, Heuveline V, Hubbuch J. 2014. Adjoint-based estimation and optimization for column liquid chromatography models. Comput Chem Eng 64:41–54.

Hansen S, Brestrich N, Hubbuch J. 2016. Mid-UV protein absorption spectra and partial least squares regression as screening and PAT tool. Wiley, in press.

Hansen SK, Jamali B, Hubbuch J. 2013. Selective high throughput protein quantification based on UV absorption spectra. Biotechnol Bioeng 110:448–460.

Hansen SK, Skibsted E, Staby A, Hubbuch J. 2011. A label-free methodology for selective protein quantification by means of absorption measurements. Biotechnol Bioeng 108:2661–2669.

Harbour PJ. 2007. The competitive implications of generic biologics, remarks of commissioner Harbour, ABA Sections of Antitrust and Intellectual Property Law, Intellectual Property Antitrust: Strategic Choices, Evolving Standards, and Practical Solutions, Federal Trade Commission, San Francisco.

Helling C, Borrmann C, Strube J. 2012. Optimal integration of directly combined hydrophobic interaction and ion exchange chromatography purification processes. Chem Eng Technol 35:1786–1796.

Höskuldsson A. 1988. PLS regression methods. J Chemom 2:211–228.

Hotelling H. 1933. Analysis of a complex of statistical variables into principal components. J Educ Psychol 24:417–441.

Huuk TC, Hahn T, Osberghaus A, Hubbuch J. 2014. Model-based integrated optimization and evaluation of a multi-step ion exchange chromatography. Sep Purif Technol 136:207–222.

ICH. 2008. Pharmaceutical development. Harmonised tripartite guideline: Q8(R1).

Ito M. 1960. The effect of temperature on ultraviolet absorption spectra and its relation to hydrogen bonding. J Mol Spectrosc 4:106–124.

Jakobsson N, Degerman M, Nilsson B. 2005. Optimisation and robustness analysis of a hydrophobic interaction chromatography step. J Chromatogr A 1099:157–166.

Jakobsson N, Degerman M, Stenborg E, Nilsson B. 2007. Model based robustness analysis of an ion-exchange chromatography step. J Chromatogr A 1138:109–119.

Jiang C, Liu J, Rubacha M, Shukla AA. 2009. A mechanistic study of protein A chromatography resin lifetime. J Chromatogr A 1216:5849–5855.

Jiskoot W, Crommelin D. 2005. Methods for structural analysis of protein pharmaceuticals. AAPS Press.

Jung B, Lee S, Yang InH, Good T, Cote GL. 2002. Automated on-line noninvasive optical glucose monitoring in a cell culture system. Appl Spectrosc 56:51–57.

Jungbauer A. 2013. Continuous downstream processing of biopharmaceuticals. Trends Biotechnol 31:479–492.

Kaltenbrunner O, Lu Y, Sharma A, Lawson K, Tressel T. 2012. Risk-benefit evaluation of on-line high-performance liquid chromatography analysis for pooling decisions in large-scale chromatography. J Chromatogr A 1241:37–45.

Kamga MH, Woo Lee H, Liu J, Yoon S. 2013. Rapid quantification of protein mixture in chromatographic separation using multi-wavelength UV spectra. Biotechnol Prog 29:664–671.

Kara S, Anton F, Solle D, Neumann M, Hitzmann B, Scheper T, Liese A. 2010. Fluorescence spectroscopy as a novel method for on-line analysis of biocatalytic CC bond formations. J Mol Catal B: Enzym 66:124–129.

Karst D, Steinebach F, Morbidelli M. 2015. Integrated continuous processing for the manufacture of monoclonal antibodies, in: Integrated Continuous Biomanufacturing II, Chetan Goudar, Amgen Inc. Suzanne Farid, University College London Christopher Hwang, Genzyme-Sanofi Karol Lacki, Novo Nordisk Eds, ECI Symposium Series.

Kessler R. 2006a. Prozessanalytik: Strategien und Fallbeispiele aus der industriellen Praxis. Wiley-VCH.

Kessler W. 2006b. Multivariate Datenanalyse: für die Pharma, Bio- und Prozessanalytik. Wiley-VCH.

Klutz S, Magnus J, Lobedann M, Schwan P, Maiser B, Niklas J, Temming M, Schembecker G. 2015. Developing the biofacility of the future based on continuous processing and single-use technology. J Biotechnol 213:120–130.

Knäblein (Ed.). 2008. Modern biopharmaceuticals. Wiley-VCH.

Knezevic I, Griffiths E. 2011. Biosimilars – Global issues, national solutions. Biologicals 39:252–255.

Konstantinov KB, Cooney CL. 2015. White paper on continuous bioprocessing. May 20-21, 2014 Continuous Manufacturing Symposium. J Pharm Sci 104:813–820.

Kowalski B, Sharaf M, Illman D. 1986. Chemometrics. Wiley.

Krättli M, Steinebach F, Morbidelli M. 2013. Online control of the twin-column countercurrent solvent gradient process for biochromatography. J Chromatogr A 1293:51–59.

Kuczewski M, Schirmer E, Lain B, G. ZP. 2011. A single-use purification process for the production of a monoclonal antibody produced in a PER.C6 human cell line. Biotechnol J 6:56–65.

Kueltzo LA, Ersoy B, Ralston JP, Middaugh CR. 2003. Derivative absorbance spectroscopy and protein phase diagrams as tools for comprehensive protein characterization: A bGCSF case study. J Pharm Sci 92:1805–1820.

Lavine BK. 2000. Chemometrics. Anal Chem 72:91–97.

Liu HF, Ma J, Winter C, Bayer R. 2010. Recovery and purification process development for monoclonal antibody production. mAbs 2:480–499.

Lourakis M. 2004. Levmar: Levenberg-marquardt nonlinear least squares algorithms in C/C++. http://users.ics.forth.gr/lourakis/levmar/. Status as of February 2016.

Malinowski ER. 2002. Factor analysis in chemistry. Wiley-VCH, 3rd edition.

Martens H, Næs T. 1989. Multivariate calibration. Wiley.

Mazarevica G, Diewok J, Baena JR, Rosenberg E, Lendl B. 2004. On-line fermentation monitoring by mid-infrared spectroscopy. Appl Spectrosc 58:804–810.

McCamish M, Woollett G. 2011. Worldwide experience with biosimilar development. mAbs 3:209–217.

McCann KB, Hughes B, Wu J, Bertolini J, Gomme PT. 2005. Purification of transferrin from Cohn supernatant I using ion-exchange chromatography. Biotechnol Appl Biochem 42:211–217.

McGlaughlin M. 2010. An emerging answer to the downstream bottleneck. BioProcess Int 10:58–61.

McKeage K. 2014. A review of ct-p13: An infliximab biosimilar. BioDrugs 28:313–321.

Mendhe R, Muralikrishnan T, Patil N, Rathore AS. 2015. Comparison of PAT based approaches for making real-time pooling decisions for process chromatography – use of feed forward control. J Chem Technol Biotechnol 90:341–348.

van de Merbel N, Lingeman H, Brinkman U. 1996. Sampling and analytical strategies in on-line bioprocess monitoring and control. J Chromatogr A 725:13–27.

Mercier SM, Diepenbroek B, Wijffels RH, Streefland M. 2014. Multivariate PAT solutions for biopharmaceutical cultivation: Current progress and limitations. Trends Biotechnol 32:329–336.

Morgan EH. 1981. Transferrin, biochemistry, physiology and clinical significance. Mol Aspects Med 4:1–123.

Næs T, Mevik BH. 2001. Understanding the collinearity problem in regression and discriminant analysis. J Chemom 15:413–426.

Nash DC, Chase HA. 1998. Comparison of diffusion and diffusion-convection matrices for use in ion-exchange separations of proteins. J Chromatogr A 807:185–207.

Navrátil M, Norberg A, Lembrén L, Mandenius CF. 2005. On-line multi-analyzer monitoring of biomass, glucose and acetate for growth rate control of a vibrio cholerae fed-batch cultivation. J Biotechnol 115:67–79.

Nilsson M, Håkanson H, Mattiasson B. 1992. Process monitoring by flow-injection immunoassay. Evaluation of a sequential competitive binding assay. J Chromatogr 597:383–389.

Novonordisk.com. 2014. Annual report. http://www.novonordisk.com/content/dam/denmark/hq/commons/documents/novo-nordisk-annual-report-2014.pdf. Status as of February 2016.

Oelmeier SA, Ladd-Effio C, Hubbuch J. 2013. Alternative separation steps for monoclonal antibody purification: Combination of centrifugal partitioning chromatography and precipitation. J Chromatogr A 1319:118–126.

Oliveira AE, Junior H, Ingber L, Petraglia A, Rembold Petraglia M, Augusta Soares Machado M. 2012. Adaptive simulated annealing, in: Stochastic global optimization and its applications with fuzzy adaptive simulated annealing. Springer Berlin Heidelberg. volume 35, pp. 33–62.

Otto R, Santagostino A, Schrader U (Eds.). 2014. From science to operations: Questions, choices and strategies for success in biopharma. Mc Kinsey.

Ozturk SS, Thrift JC, Blackie JD, Naveh D. 1995. Real-time monitoring of protein secretion in mammalian cell fermentation: Measurement of monoclonal antibodies using a computer-controlled HPLC system (biocad/rpm). Biotechnol Bioeng 48:201–206.

Paliwal SK, Nadler TK, Wang DIC, Regnier FE. 1993. Automated process monitoring of monoclonal antibody production. Anal Chem 65:3363–3367.

Prin C, Bene MC, Gobert B, Montagne P, Faure GC. 1995. Isoelectric restriction of human immunoglobulin isotypes. Biochim Biophys Acta 1243:298–290.

Proll G, Kumpf M, Mehlmann M, Tschmelak J, Griffith H, Abuknesha R, Gauglitz G. 2004. Monitoring an antibody affinity chromatography with a label-free optical biosensor technique. J Immunol Methods 292:35–42.

Przybycien TM, Pujar NS, Steele LM. 2004. Alternative bioseparation operations: Life beyond packed-bed chromatography. Curr Opin Biotech 15:469 –478.

Qureshi ZH, Ng TS, Goodwin GC. 1980. Optimum experimental design for identification of distributed parameter systems. Int J Control 31:21–29.

Ragone R, Colonna G, Balestrieri C, Servillo L, Irace G. 1984. Determination of tyrosine exposure in proteins by second-derivative spectroscopy. Biochem 23:1871–1875.

Rathore AS, Agarwal H, Sharma AK, Pathak M, Muthukumar S. 2015. Continuous processing for production of biopharmaceuticals. Prep Biochem Biotechnol 45:836–849.

Rathore AS, Bhambure R, Ghare V. 2010a. Process analytical technology (PAT) for biopharmaceutical products. Anal Bioanal Chem 398:137–154.

Rathore AS, Bhushan N, Hadpe S. 2011. Chemometrics applications in biotech processes: A review. Biotechnol Bioeng 27:307–313.

Rathore AS, Kapoor G. 2015. Application of process analytical technology for downstream purification of biotherapeutics. J Chem Technol Biotechnol 90:228–236.

Rathore AS, Li X, Bartkowski W, Sharma A, Lu Y. 2009. Case study and application of process analytical technology (PAT) towards bioprocessing: Use of tryptophan fluorescence as at-line tool for making pooling decisions for process chromatography. Biotechnol Prog 25:1433–1439.

Rathore AS, Parr L, Dermawan S, Lawson K, Lu Y. 2010b. Large scale demonstration of a process analytical technology application in bioprocessing: Use of on-line high performance liquid chromatography for making real time pooling decisions for process chromatography. Biotechnol Prog 26:448–457.

Rathore AS, Wood R, Sharma A, Dermawan S. 2008a. Case study and application of process analytical technology (PAT) towards bioprocessing II: Use of ultra-performance liquid chromatography (UPLC) for making real-time pooling decisions for process chromatography. Biotechnol Bioeng 101:1366–1374.

Rathore AS, Yu M, Yeboah S, Sharma A. 2008b. 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 100:306–316.

Read EK, Park JT, Shah RB, Riley BS, Brorson KA, Rathore AS. 2010. Process analytical technology (PAT) for biopharmaceutical products: Part I. Concepts and applications. Biotechnol Bioeng 105:276–284.

Rege K, Pepsin M, Falcon B, Steele L, Heng M. 2006. High-throughput process development for recombinant protein purification. Biotechnol Bioeng 93:618–630.

Roche.com. 2014. Financial report. http://www.roche.com/fb14e.pdf. Status as of February 2016.

Rosa P, Azevedo A, Ferreira I, de Vries J, Korporaal R, Verhoef H, Visser T, Aires-Barros M. 2007. Affinity partitioning of human antibodies in aqueous two-phase systems. J Chromatogr A 1162:103 –113.

Rosenberg N. 1974. Science, invention and economic growth. The Econ J 84:90–108.

Rosenheck K, Doty P. 1961. The far ultraviolet absorption spectra of polypeptide and protein solutions and their dependence on conformation. P Natl Acad Sci USA 47:1775–1785.

Savitzky A, Golay. 1964. Smoothing and differentiation of data by simplified least squares procedures. Anal Chem 36:1627–1639.

Schiestl M, Stangler T, Torella C, Čepeljnik T, Toll H, Grau R. 2011. Acceptable changes in quality attributes of glycosylated biopharmaceuticals. Nat Biotechnol 29:310–312.

Schmidt-Traub H. 2005. Preparative chromatography of fine chemicals and pharmaceutical agents. Wiley.

Sekhon BS, Saluja V. 2011. Biosimilars: An overview. Biosimilars 1:1–11.

Shukla AA, Hinckley P. 2008. Host cell protein clearance during protein A chromatography: Development of an improved column wash step. Biotechnol Prog 24:1115–1121.

Shukla AA, Hubbard B, Tressel T, Guhan S, Low D. 2007. Downstream processing of monoclonal antibodies – Application of platform approaches. J Chromatogr B 848:28–39.

Shukla AA, Thömmes J. 2010. Recent advances in large-scale production of monoclonal antibodies and related proteins. Trends Biotechnol 28:253–261.

Sim SL, He T, Tscheliessnig A, Mueller M, Tan RB, Jungbauer A. 2012. Branched polyethylene glycol for protein precipitation. Biotechnol Bioeng 109:736–746.

Statistica.com. 2015. http://de.statista.com/statistik/daten/studie/36502/umfrage/weltweit-umsatzstaerkste-biotech-medikamente/, http://de.statista.com/statistik/daten/studie/36502/umfrage/weltweit-umsatzstaerkste-biotech-medikamente/. Status as of February 2016.

Steinhoff RFK, J. D, Steinebach F, Kopp MR, Schmidt GW, Stettler A, Krismer J, Soos M, Pabst M, Hierlemann A, Morbidelli M, Zenobi R. 2015. Microarray-based maldi-tof mass spectrometry enables monitoring of monoclonal antibody production in batch and perfusion cell cultures. Methods, in press .

Tamburini E, Vaccari G, Tosi S, Trilli A. 2003. Near-infrared spectroscopy: A tool for monitoring submerged fermentation processes using an immersion optical-fiber probe. Appl Spectrosc 57:132–138.

Tarrant RDR, Velez-Suberbie ML, Tait AS, Smales CM, Bracewell DG. 2012. Host cell protein adsorption characteristics during protein A chromatography. Biotechnol Prog 28:1037–1044.

Thakkar SV, Allegre KM, Joshi SB, Volkin DB, Middaugh CR. 2012. An application of ultraviolet spectroscopy to study interactions in proteins solutions at high concentrations. J Pharm Sci 101:3051–3061.

Thömmes J, Etzel M. 2007. Alternatives to chromatographic separations. Biotechnol Prog 23:42–45.

Trusheim MR, Aitken ML, Berndt ER. 2010. Economic aspects of small and large molecule pharmaceutical technologies. Forum Health Econ Policy 13:1–43.

Tsukamoto M, Watanabe H, Ooishi A, Honda S. 2014. Engineered protein A ligands, derived from a histidine-scanning library, facilitate the affinity purification of IgG under mild acidic conditions. J Biol Eng 8:1–9.

Tsuruta LR, Lopes dos Santos M, Moro AM. 2015. Biosimilars advancements: Moving on to the future. Biotechnol Prog 31:1139–1149.

Ucinski D. 2004. Optimal measurement methods for distributed parameter system identification. CRC Press LLC.

University of Copenhagen. 2015. http://www.models.life.ku.dk/teaching. Status as of October 2015.

Velayudhan A, Horvath C. 1988. Preparative chromatography of proteins – Anaysis of the multivalent ion-exchange formalism. J Chromatogr 443:13–29.

Walsh G. 2010. Biopharmaceutical benchmarks 2010. Nat Biotechnol 28:917–924.

Walter E, Pronzato L. 1997. Identification of parametric models: From experimental data. Springer-Verlag.

Warikoo V, Godawat R, Brower K, Jain S, Cummings D, Simons E, Johnson T, Walther J, Yu M, Wright B, McLarty J, Karey KP, Hwang C, Zhou W, Riske F, Konstantinov K. 2012. Integrated continuous production of recombinant therapeutic proteins. Biotechnol Bioeng 109:3018–3029.

Westerberg K, Broberg Hansen E, Degerman M, Budde Hansen T, Nilsson B. 2012. Model-based process challenge of an industrial ion-exchange chromatography step. Chem Eng Technol 35:183–190.

Westerberg K, Degerman M, Nilsson B. 2010. Pooling control in variable preparative chromatography processes. Bioprocess Biosyst Eng 33:375–382.

Wetlaufer D. 1962. Ultraviolet spectra of proteins and amino acids. Adv Prot Chem 17:303–390.

Wheatley PJ. 1964. The crystal and molecular structure of aspirin. J Chem Soc :6036–6048.

Xenopoulos A. 2015. A new, integrated, continuous purification process template for monoclonal antibodies: Process modeling and cost of goods studies. J Biotechnol 213:42 –53.

Yeung KSY, Hoare M, Thornhill NF, Williams T, Vaghjiani JD. 2000. Near-infrared spectroscopy for bioprocess monitoring and control. Biotechnol Bioeng 63:684–693.

Zang H, Wang J, Li L, Zhang H, Jiang W, Wang F. 2013. Application of near-infrared spectroscopy combined with multivariate analysis in monitoring of crude heparin purification process. Spectrochim Acta, Part A 109:8–13.

Zheng R. 2010. The game changer. BioProcess Int 8:4–9.

Zhou JX, Dermawan S, Solamo F, Flynn G, Stenson R, Tressel T, Guhan S. 2007. pH-conductivity hybrid gradient cation-exchange chromatography for process-scale monoclonal antibody purification. J Chromatogr A 1175:69–80.

List of Abbreviations and Symbols

Abbreviations

AEX	Anion Exchange
CEX	Cation Exchange
CHO	Chinese Hamster Ovary
CPP	Critical Process Parameters
CQA	Critical Quality Attribute
CV	Column Volume
Cyt c	Cytochrome c
DAD	Diode Array Detector
DAQ	Data Aquisition Toolbox
DoE	Design of Experiments
ELISA	Enzyme-Linked Immunosorbent Assay
FDA	Food and Drug Administration
FIM	Fisher Information Matrix
FTIR	Fourier transform infrared spectroscopy
HCCF	Harvested Cell Culture Fluid
HCP	Host Cell Protein
HMWs	High Molecular Weight Species
HPLC	High Performance Liquid Chromatography
IgG	Immunoglobulin G
LHS	Liquid Handling Station
LMWs	Low Molecular Weight Species
LV	Latent Variable
Lys	Lysozyme
MCSGP	Multicolumn Countercurrent Solvent Gradient Purification
MS	Mass Spectrometry
MV	Membrane Volume
MVDA	Multivariate Data Analysis

P	Index for Protein
mAb	Monoclonal Antibody
PAT	Process Analytical Technologies
PC	Principal Component
PCA	Principal Component Analysis
PCR	Principal Component Regression
PLS	Partial Least Squares Regression
QbD	Quality by Design
qPCR	Quantitative Real Time Polymerase Chain Reaction
Rib A	Ribonuclease A
RMSE	Root Mean Square Error
RMSEC	Root Mean Square Error of Calibration
RMSECV	Root Mean Square Error of Cross Validation
RMSEP	Root Mean Square Error of Prediction
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
SDM	Stoichiometric Displacement Model
SEC	Size Exclusion Chromatography
Trf	Transferrin
UHPLC	Ultra High Performance Liquid Chromatography
VBA	Visual Basic for Application
VP	Variable Pathlength

Symbols

a	Number of selective protein concentrations	-
A_{280}	Absorption at 280 nm	AU
A_{527}	Absorption at 527 nm	AU
$\mathbf{B} \in \mathbb{R}^{mxa}$	Matrix of regression coefficients/PLS model	$\frac{g}{l \cdot AU}$
c_i	Concentration of species i in the interstitial volume	$\frac{mol}{l}$
$c_{p,i}$	Concentration of species i in the pore volume	$\frac{mol}{l}$
D_{ax}	Axial dispersion coefficient	$\frac{mm^2}{s}$
$\mathbf{E} \in \mathbb{R}^{nxm}$	Residual matrix of the $\mathbf{X}$ factorization	AU
$\mathbf{F} \in \mathbb{R}^{nxa}$	Residual matrix of the $\mathbf{Y}$ factorization	$\frac{g}{l}$
$\mathbf{G} \in \mathbb{R}^{nxa}$	Residual matrix of PLS	$\frac{g}{l}$
$\mathbf{I} \in \mathbb{R}^{kxk}$	Inner relation between the score matrices $\mathbf{T}$ and $\mathbf{U}$	-
ε_b	Bed porosity	-
ε_p	Particles' porosity	-

k	Number of applied PCs or LVs	-
k_{eff}	Lumped film transfer/ pore diffusion coefficient	$\frac{\text{mm}}{\text{s}}$
k_{eq}	Adsorption equilibrium coefficient	-
L	Column length	mm
Λ	Total ionic capacity of the stationary phase	$\frac{\text{mol}}{\text{l}}$
ν_i	Characteristic charge of species i	-
$\mathbf{P} \in \mathbb{R}^{mxk}$	Loading matrix of the $\mathbf{X}$ factorization	-
$\mathbf{Q} \in \mathbb{R}^{axk}$	Loading matrix of the $\mathbf{Y}$ factorization	-
m	Number of measured variables	-
n	Number of measured objects	-
q_i	Stationary phase concentration of species i	$\frac{\text{mol}}{\text{l}}$
r_p	Particle radius of adsorbent	mm
t	Time dimension	s
$\mathbf{T} \in \mathbb{R}^{nxk}$	Score matrix of the $\mathbf{X}$ factorization	AU
$\mathbf{t} \in \mathbb{R}^{nx1}$	Vector of the score matrix of the $\mathbf{X}$ factorization	AU
$\bar{t}$	Mean of $\mathbf{t}$	AU
t	Score value in $\mathbf{T}$	AU
u_{int}	Interstitial mobile phase velocity	$\frac{\text{mm}}{\text{s}}$
$\mathbf{U} \in \mathbb{R}^{nxk}$	Score matrix of the $\mathbf{Y}$ factorization	AU
$\mathbf{u} \in \mathbb{R}^{nx1}$	Vector of the score matrix of the $\mathbf{Y}$ factorization	AU
$\bar{u}$	Mean of $\mathbf{u}$	AU
u	Score value in $\mathbf{U}$	AU
x	Space dimension	mm
$\mathbf{X} \in \mathbb{R}^{nxm}$	Matrix of UV absorption spectra	AU
$\mathbf{Y} \in \mathbb{R}^{nxa}$	Matrix of selective protein concentrations	$\frac{\text{g}}{\text{l}}$
$\mathbf{Y_{pred}} \in \mathbb{R}^{nxa}$	Predicted matrix of selective protein concentrations	$\frac{\text{g}}{\text{l}}$
$\mathbf{Y_{ref}} \in \mathbb{R}^{nxa}$	Reference matrix of measured protein concentrations	$\frac{\text{g}}{\text{l}}$
$\mathbf{y} \in \mathbb{R}^{nx1}$	Vector of one protein concentration	$\frac{\text{g}}{\text{l}}$
y_{pred}	Predicted selective protein concentration	$\frac{\text{g}}{\text{l}}$
y_{ref}	Measured reference protein concentration	$\frac{\text{g}}{\text{l}}$