High Throughput Analytics for Protein Purification Process Development and Process Control

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.) Sigrid Krogager Hansen
aus Aarhus, Dänemark

Referent: Prof. Dr. Jürgen Hubbuch
Korreferent: Prof. Dr. Matthias Franzreb
Tag der mündlichen Prüfung: 23. September 2013

Abstract

This thesis makes an original contribution to the area of fast protein analytics mainly for application in combination with high throughput process development. A central topic is the more efficient utilization of equipment routinely used for protein analysis and quantification. This was shown to be possible both with purely practical measures as well as by applying multivariate analysis and calibration in areas where traditionally only univariate data is used. With regard to multivariate analysis, it was shown that in combination with spectroscopy, multivariate calibration can exploit the unique characteristics of protein absorption spectra. This enables selective quantification of protein concentrations in protein mixtures which is both label-free and non-invasive. Further, multivariate calibration is used in order to evaluate analytical chromatograms of which the resolution of the components to be quantified was not sufficient to obtain precise results otherwise.

The first part of this thesis is concerned with the development of generic approaches to increase the throughput of protein analysis based on size exclusion chromatography (SEC). By interlacing the sample injections, the lag time before elution of the first sample components can be eliminated. This lag time equals the interstitial volume of the applied SEC column. Importantly, it was possible also to eliminate the lag time post elution of the molecules to be analyzed. This was done by operating two columns in parallel on the same HPLC system. For this configuration, only one additional pump and two valves are necessary. The interlaced injections were performed on both columns at the same time, whereat the samples were injected alternating on the two columns. The exact timing of each injection was adjusted to allow only eluate containing samples components which was to be quantified to pass through the UV detector. The controlling of the chromatography system devices required a special programming to enable correct assignment of a recorded chromatogram to the respective injected sample. The developed approach to increase throughput of SEC was shown to be robust and to deliver results which were identical to SEC performed in conventional mode. A conducted case study showed that for aggregate quantification of a monoclonal antibody, it was possible to obtain a threefold reduction in analysis time. This approach can be applied for any analysis based on SEC. The part of the lag time which can be eliminated will depend on the duration of each lag phase and on the ratio of the overall lag time to the duration of the information phase.

The second approach to increase throughput in SEC was based on dynamic multivariate calibration of the quantitative relationship between a chromatogram and analyte concentrations. The idea was to use multivariate calibration to compensate for low chromatographic resolution in the strive for fast analysis. As an example of use, the quantification of protein-polymer conjugates was chosen. In this particular case, native protein along with mono-, di-, and tri-conjugated protein were present in the samples to be analyzed, all of which were to be quantified. Initially, the choice of a short SEC column in combination with a high flow rate enabled a sufficiently short analysis time. The tradeoff for fast analysis was a very low chromatographic resolution of the di- and tri-conjugated proteins. To achieve precise quantification despite this low resolution, multivariate calibration was used to establish a quantitative correlation between chromatogram and analyte concentrations. The precision of different multivariate regression methods were compared quantitatively and multilinear regression was shown to result in the highest precision. Importantly, a dynamic approach was included in the multivariate

calibration in order to cope with changes in column performance over time which are often inevitable. This measure contributed significantly to the precision of the analysis. In conclusion, it could be shown that multivariate calibration can compensate for low resolution. Hence, multivariate calibration can contribute to the achievement of protein quantification based on analytical chromatography which is fast enough to fit the concept of high throughput process development. In this case, dynamic multivariate calibration was used in combination with interlaced sample injections and together these measures resulted in an analysis time of two minutes.

A significant part of this thesis consists of an original contribution to the application of protein absorption spectra for selective protein quantification. It was investigated whether the information contained in the unique absorption spectra of proteins can be used for selective quantification in protein mixtures. The accomplished work show that multivariate calibration can indeed be used to establish a precise correlation between absorption spectra and selective protein concentrations. With respective calibrations, it was possible to perform selective protein quantification in the range 0.07 - 1.5 g/L and for up to four different proteins. Mid-UV protein absorption spectra are known to be insensitive to moderate changes of the solvent (e.g. ion strength or pH). Validation experiments confirmed the robustness of the spectra against solvent changes in a range relevant for liquid chromatography. The applicability of this new method was illustrated by a resin screening in packed column high throughput format. Because selective protein quantification based on absorption spectra was applied instead of analytical chromatography for analysis of the samples, both the time needed for analysis and the workload for the experimenter was decreased significantly. Importantly, the achieved increase in analytical throughput was achieved without compromising the quality of the results when compared to results obtained by analytical chromatography. A very important aspect of the developed method for selective protein quantification, is the degree of spectral similarity in combination with the accuracy and precision of the device used to record the spectral data. Hence, further work was dedicated to uncover both possibilities and limitations of selective protein quantification based on protein absorption spectra. Principal component analysis was performed on a broad range of protein absorption spectra to investigate their common characteristics and differences. The principal component analysis was used both for cluster analysis and to define a measure for spectral similarity. For binary protein combinations, the calibration precision was shown to decrease exponentially with the defined spectral similarity factor. This correlation will depend on the applied spectrophotometer and it can be used to determine a priori whether a calibration for certain proteins will be successful or not. To enable fast measurement of many samples, most often single beam plate reader devices are used. It was shown that a plate reader device is not precise enough to enable selective quantification of two monoclonal antibodies of which the spectra differed only by 0.6 %. However, when applying a double beam photometer where the measurements are performed in cuvettes, this calibration was successfull.

In the final part of this thesis, preliminary studies of an in-line application of spectral based selective protein quantification for process control is presented. The previously presented work proved that mid-UV protein absorption spectra are very useful for selective protein quantification. The objective here was to show that this approach can also function in a real time application during the performance of preparative chromatography. Such an application requires the use of a diode array detector in order to acquire full spectra on a real time basis. The resolution of a diode array detector is in general lower compared to monochromator based equip-

ment used in the previous work. However, the application of spectral based selective protein quantification proved to result in very precise signals even for proteins with very similar absorption spectra. Hence, it was concluded that selective protein quantification based on mid-UV absorption spectra are indeed qualified for applications related to process control purposes.

Zusammenfassung

In der vorliegenden Doktorarbeit werden neuartige Methoden zur Proteinanalytik für die nwendung in Kombination mit Hochdurchsatz-Prozessentwicklung präsentiert. Ein zentrales Thema ist die effizientere Nutzung von Geräten, die routinemäßig zur Proteinanalytik eingesetzt werden. Es wurde gezeigt, dass die Effizienz gesteigert werden konnte, indem alternative Wege zur Durchführung der Analysen entwickelt wurden. Weiter konnte gezeigt werden, dass der Einsatz von multivariaten Methoden zur Datenanalyse und Datenauswertung die Qualität der angewendeten Analysen erhalten konnte bei gleichzeitiger Reduzierung der zur Analyse benötigten Zeit. In Verbindung mit Spektroskopie wurden multivariate Methoden benutzt, um die einzigartigen spektralen Charakteristika von Proteinen zu verwerten. Somit wurde selektive Proteinquantifizierung über Absorptionsmessungen in Proteinmischungen ermöglicht. Weiter wurde multivariate Kalibrierung bei der Auswertung von analytischen Chromatogrammen eingesetzt, bei denen die Auflösung zwischen den Komponenten zu gering war, um eine genaue Quantifizierung über Peakflächen zu gewährleisten.

Der erste Teil dieser Dissertation befasst sich mit der Entwicklung eines generischen Vorgehens zur Erhöhung des Durchsatzes von Analysen basierend auf Größenausschlusschromatographie (SEC). Es konnte gezeigt werden, dass eine Verschachtelung der Probeninjektion die Verzögerungszeit vor der Elution der Probenkomponenten einsparen konnte. Diese Verzögerungszeit ist gleich dem interstitiellen Volumen der jeweils genutzten Chromatographiesäule. Zudem war es möglich, auch die Verzögerungszeit nach der Elution der letzten zu quantifizierenden Probenmoleküle nahezu vollständig zu eliminieren, indem zwei Säulen gleichzeitig an einer Chromatographieanlage betrieben wurden. Hierzu wurden lediglich eine zusätzliche Pumpe und zwei Ventile benötigt. Anschließend, wurde die Verschachtelung der Injektionen auf beiden Säulen gleichzeitig durchgeführt, wobei die Proben alternierend auf beiden Säulen injiziert wurden. Die genauen Zeitpunkte der einzelnen Injektionen wurden so auf einander abgestimmt, dass annähernd ausschließlich die Informationsphase im Detektor erfasst wurde. Die Steuerung der Geräteeinheiten erforderte eine besondere Programmierung, um einer detektierten Informationsphase die jeweils injizierte Probe zuzuordnen. Die entwickelte Methode erwies sich als sehr robust und wies die gleiche Präzision und Richtigkeit der ursprünglichen Analysemethode. Auf diese Weise konnte der Probendurchsatz an einer einzelnen Chromatographieanlage für die Bestimmung der Aggregatanteilen von Antikörperlösungen verdreifacht werden. Die entwickelte Methodik kann für jede Analyse, die auf Größenausschlusschromatographie basiert, angewendet werden. Die anteilige Verzögerungszeit, die eingespart werden kann, hängt von der Verteilung vor und nach der Informationsphase, sowie dem Verhältnis von Verzögerungszeit zur Dauer der Informationsphase ab. Des Weiteren wurde multivariate Kalibrierung in dynamischer Form für die Auswertung von chromatographischen Daten entwickelt und eingesetzt. Das Ziel war es, mittels multivariater Kalibrierung trotz unvollständiger Trennung von Probenmolekülen bei SEC-Analysen richtige und präzise Analyseergebnisse in quantitativer Form erzielen zu können, und somit eine schnellere Analytik zu ermöglichen. Als Anwendungsbeispiel wurde die Analyse von Protein-Polymer Konjugaten mittels SEC gewählt. Es sollte sowohl natives Protein als auch mono-konjugiertes, di-konjugiertes und tri-konjugiertes Protein quantifiziert werden. Durch Benutzung einer kurzen Säule bei hoher Flussrate verkürzten sich die Analysezeiten deutlich, doch führte dies gleichzeitig zu einer Verminderung der chromatographischen Auflösung.

Um die erniedrigte Auflösung zu kompensieren, wurde multivariate Kalibrierung benutzt, um einen quantitativen Zusammenhang zwischen den Chromatogrammen bekannter Proben und die in den Proben beinhalteten Analytkonzentrationen herzustellen. Die Genauigkeit verschiedener multivariater Kalibrierungsmethoden wurde quantitativ miteinander verglichen. Mittels multilinearer Regression wurden die höchsten Genauigkeiten für die Kalibrierungen erreicht. Des Weiteren wurde ein dynamischer Kalibrierungsansatz entwickelt, um eine Änderung in der Auftrennungsleistung der Säule über die Zeit korrigieren zu können. Dadurch konnte die Genauigkeit der Methode signifikant gesteigert werden. Die Arbeit zeigte, dass die Anwendung von multivariater Kalibrierung eine niedrige Peak-Auflösung kompensieren kann. Schließlich konnte durch die Kombination aus multivariater Kalibrierung und verschachtelter Probeninjektionen eine genaue und präzise Analytik für Protein-Polymer-Konjugate mit einer Analysezeit von nur 2 Minuten erreicht werden.

Der Hauptteil dieser Arbeit liefert einen neuen technologischen Beitrag in der Nutzung von Proteinabsorptionsmessungen zur selektiven Proteinquantifizierung. In diesem Teil der Arbeit wurde untersucht, ob die Information der charakteristischen Absorptionsspektren von Proteinen auch zur selektiven Quantifizierung genutzt werden kann. Die durchgeführten Entwicklungs- und Validierungsarbeiten auf diesem Gebiet zeigten, dass multivariate Kalibrierung eingesetzt werden kann, um eine Korrelation zwischen Absorptionsspektren und Proteinkonzentrationen in einem Gemisch zu bestimmen. Mit jeweiligen Kalibrierungen war es möglich, selektiv Proteinkonzentrationen im Bereich von 0.07 g/L - 1.5 g/L von bis zu 4 Proteinen zu quantifizieren. Validierungsversuche bestätigten die Robustheit der entwickelten Quantifizierungsmethode gegenüber Änderungen der Lösungs-mitteleigenschaften in einem Bereich, der relevant für präparative Chromatographie ist. Die Anwendbarkeit dieser Methode wurde in Screenings von Adsorbermaterialien im Hochdurchsatzformat gezeigt. Im Vergleich zu einer Analyse mittels analytischer Chromatographie konnten sowohl die Analysezeit als auch der Arbeitsaufwand deutlich verringert werden. Es ist hervorzuheben, dass die Steigerung des Durchsatzes erzielt werden konnte, ohne die Qualität der Ergebnisse zu beeinflussen.

Sehr wichtige Einflussparameter der entwickelten analytischen Methode sind das Maß der Unterschiedlichkeit in den Proteinabsorptionsspektren und die Genauigkeit des Messinstrumentes, mit der die Spektren aufgenommen werden. Daher wurden in einem weiteren Abschnitt dieser Arbeit Untersuchungen durchgeführt, die sowohl die Genauigkeit und Robustheit der Analysemethode, als auch ihre Limitierungen aufzeigen sollte. Die Methode der Hauptkomponentenanalyse wurde eingesetzt, um Ähnlichkeiten und Unterschiede in den Absorptionspektren zu untersuchen. Zunächst wurde eine Clusteranalyse durchgeführt. Zudem wurde die Hauptkomponentenanalyse angewendet, um einen Faktor für spektrale Ähnlichkeit zu definieren. Für binäre Proteingemische zeigte sich eine exponentielle Abnahme der Genauigkeit mit zunehmender Ähnlichkeit der Spektren. Vorherige Kenntnis über diesen Zusammenhang kann dazu verwendet werden, um a priori vorhersagen zu können, ob die Analysemethode sich mit ausreichender Genauigkeit etablieren lässt. Um möglichst schnell Proben zu messen, werden oft Einstrahl-Photometer für Mikrotiterplatten verwendet. Es zeigte sich jedoch, dass ein solches Photometer nicht die Genauigkeit bietet, um zwei monoklonale Antikörper, deren Spektren sich nur um 0,6 % unterscheiden, selektiv zu quantifizieren. Bei Verwendung eines Doppeltstrahlphotometers und Messung in Küvetten war selektive Quantifizierung allerdings mit hoher Genauigkeit möglich.

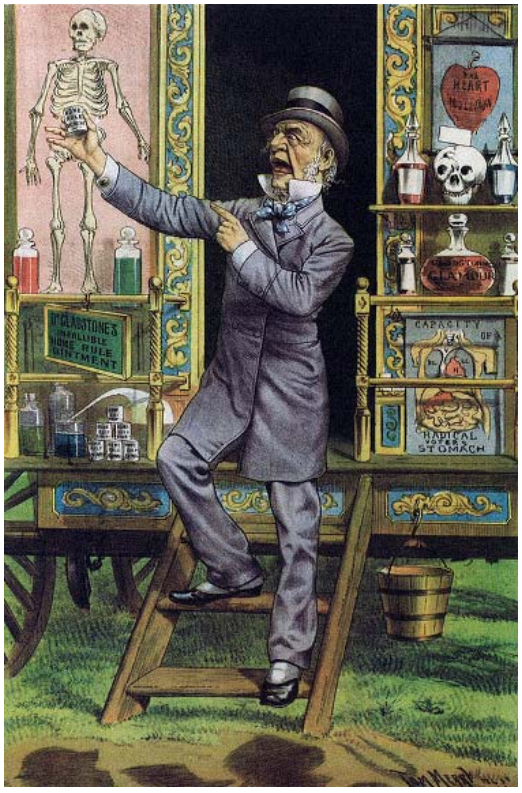
Im letzten Teil der Arbeit wurde eine Machbarkeitsstudie zur einen Echtzeit-Anwendung der Spektralanalytik durchgeführt. Die Möglichkeit selektive Proteinquantifizierung in Echtzeit durchzuführen, könnte zur Prozesskontrolle bei chromatographischen Trennungen von Proteinen im präparativen Maßstab eingesetzt werden. Diese Anwendung benötigt ein Spektrophotometer mit einem Diodenarraydetektor, um ganze Spektren in Echtzeit aufnehmen zu können. Die Auflösung eines Diodenarraydetektors ist im Vergleich zu den in den vorigen Arbeiten benutzen Monochromatorbasierten Geräten niedriger. Dennoch konnte gezeigt werden, dass die Auflösung ausreichend war, um präzise selektive Proteinsignale über den Elutionsbereich zu generieren. Es wird daher als vielversprechend angenommen die Methode auch zur Prozesskontrolle einzusetzen.

Contents

1 Introduction

1.1 Preface - From Quack to Exhaustive Regulation

The business of manufacturing and selling substances with the purpose to bring healing or relief to individuals suffering from illness has undergone vast changes. Until the end of the 19th century, most drugs were prepared in pharmacies and there was no regulation or restriction of claimed efficacy or safety of the drugs. Around this time pharmacies began to develop from small manufacturers into pharmaceutical companies and a new industry was born. Insulin and antibiotics were the first bio-pharmaceutical blockbusters and many of today's large pharmaceutical companies originated around this business. Soon came the first major scandals and as a reaction to each scandal came new laws and regulations. The early legislation mostly aimed to secure a correlation between label claim and actual content. Since then, continuous addition of laws and guidelines have made the pharmaceutical industry subject to very detailed regulation, both with regard to development, testing, production and admission/licensing. Manufacturing is required to be at state of the art and numerous guidelines from different agencies and organizations have been issued describing what the current state of the art is. State of the art protein purification processes are required to be based on identified product quality attributes and their detailed and documented linkage to process parameters. At the stage of manufacturing a detailed process control strategy must be in place. The developed control strategy is aimed at ensuring very high as well as highly reproducible quality based on different control elements and the knowledge of the links between process parameters and product quality. A control strategy consists of numerous elements including *process control elements* such as Procedural Controls and Process Parameter Controls and *testing control elements* such as In-Process Testing, Specifications, and Process Monitoring.

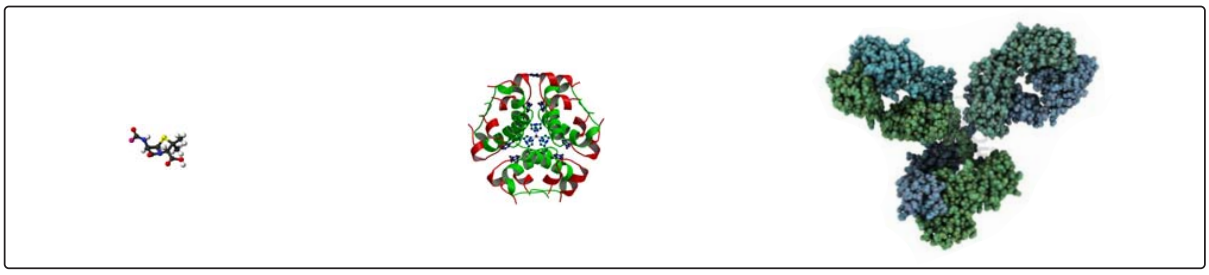


1.2 Therapeutic Proteins

Therapeutic proteins can be divided into three groups: therapeutic proteins with enzymatic or regulatory activity (for replacement therapy e.g. growth hormone, insulin, blood factors), therapeutic proteins with special target activity (e.g. monoclonals antibodies), protein vaccines (against pathogens or treatment of cancer and autoimmune diseases). Together they have an enormous marked value and today it is impossible to imagine a pharmaceutical market without therapeutic proteins.

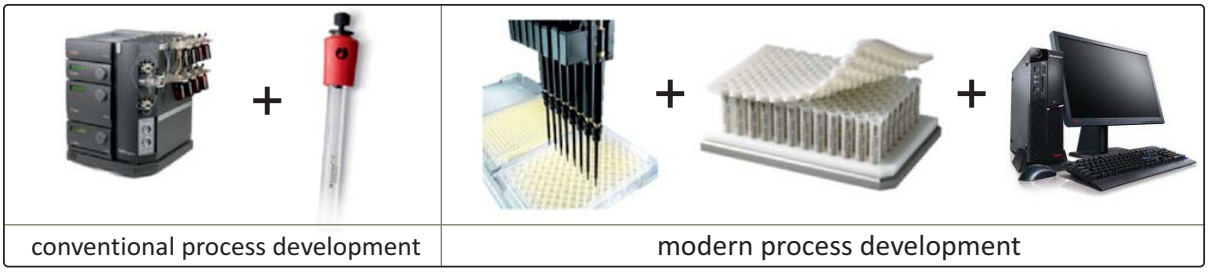
Compared to small molecules therapeutics, protein therapeutics are challenging to manufacture and due to their size and complexity also elaborate to analyze and characterize. Due to the biologic origin and the complexity of the therapeutic proteins, a final drug product administered to a patient will never consist only of numerous identical protein molecules in an appropriate matrix. Much rather it is a well defined population of protein species: Some related to the main product such as charge variants, aggregates, and glycoforms and some which are a reminiscence of the biological host in form of host cell proteins. Purification of therapeutic proteins from their biological source must be performed under gentle conditions in order to preserve the biological function and mostly rely on a set of unit operations including but not limited to filtration, chromatography, centrifugation, precipitation, and crystallization.

The introduction of high throughput techniques has modernized the field of protein purification process development over the past decade. While high throughput techniques such as automation, parallelization, and scale miniaturization has long been applied for screenings in the field of bioprocess development, this was not the case for purification process development at the turn of the millennium. The main driving force in this progress is the need for very short purification process development times and the requirement for a deeper knowledge and understanding of the processes.



1.3 The Dawn of Modern Protein Purification Process Development

As it has been for decades, liquid chromatography is still the central unit operation for biomolecule purification. When developing a purification process step based on liquid chromatography, there is a vast number of solid and liquid phases to choose from and the total number of possible parameters influencing any given separation process exceeds by far what can be tested in a sequential packed column mode. Hence, it is not a surprise that the first reported application of high throughput techniques for protein purification process development was related to chromatography. This first report described a parallel micro scale experimentation in micro titer plate format with automated liquid handling and was published by Thiemann et al. in 2004.



The authors suggested that due to high number of experiments necessary for a systematic variation of the relevant parameters in the optimization of chromatographic process step, automation of the liquid handling would bring substantial benefit with regard to experimental throughput. Where this initial work published by Thiemann et al. was aimed at the optimization of single purification step, Rege et al. in 2006 proposed a micro plate based screening method to identify the best sequence of chromatography steps and hence named their approach *high throughput process development*. Hereby they had named a new field within in protein purification research and became the herald of a new era in protein purification process development.

1.3.1 Literature Summary

A literature search on the subject of high throughput process development[1], of which the quantitative result is displayed in Fig. 1, reveals the novelty of this field. Up to 2007 there were only a few publications dealing explicitly with the theme of high throughput process development. However, in 2010 the interest in the subject had grown to an extend which enabled the first installment of a conference series dedicated to the theme. The increasing interest is also reflected in the number of publications which increased significantly after the year 2010. The literature search was limited to unit operations mainly used for protein purification and hence does not include work on protein refolding or crystallization. In summary, the applications found were liquid chromatography in batch and column mode, precipitation, and extraction in aqueous two phase systems.

Many of the early publications in the field were mainly concerned with the development of tools and approaches for efficient screening procedures (e.g., Thiemann et al., 2004; Rege et al., 2006; Wiendahl et al., 2008). In the publication by Rege et al. (2006), concern was expressed towards the equivalent distribution of the relative small amounts of chromatographic resin used in the conducted experiments. This concern was addressed and a solution proposed the same year by Herrmann et al. who presented a device which could produce equally sized resin plaques in a 96 well format. Most recently, an in-process photometric resin quantification has been presented in order to correct for pipetting error in resin distribution (Kittelmann et al., 2012; Li et al., 2012).

Up to now, successful applications of high throughput process development was mainly demonstrated with monoclonal antibodies (Wierling et al., 2007; Bergander et al., 2008; Coffman et al., 2008; Kelley et al., 2008a,b; Kramarczyk et al., 2008; Wensel et al., 2008; Rodrigo

[1]Search conducted in the on-line collection of scientific publications ScienceDirect and academic search service Web of Knowledge. In both data bases, two searches were conducted. One the therm 'high throughput process development' and one on 'high throughput' together with 'protein purification'.

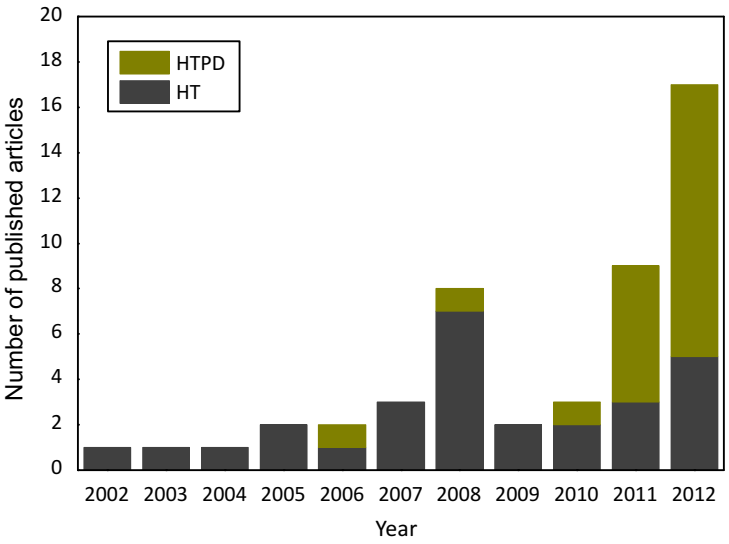


Figure 1: Number of publications on high throughput process development for protein purification. The search was conducted both with the specific term 'high throughput process development' (Legend: HTPD) and with a combination of 'high throughput' and 'protein purification' (Legend: HT). Publication which came up under both searches are included only in the numbers for HTPD publications.

and Nilsson-Välimaa, 2010; Toueille et al., 2011; Oelmeier et al., 2011; Morrison et al., 2011; Gronberg et al., 2011; Treier et al., 2012b; Oelmeier et al., 2012; Konstantinidis et al., 2012b). This is thought to be both a reflection of the vast number of monoclonal antibodies in the biopharmaceutical pipelines as well as the fact that these molecules are subject to platform processing. Platform processes offers a unique opportunity to use high throughput tools *e.g.* for the assessment of platform fit of new molecules (McDonald et al., 2010). A high number of reports was based on work with model proteins in order to bring proof of concept to new approaches within high throughput process development (e.g., Bensch et al., 2007; Susanto et al., 2008; Kumar et al., 2009; Susanto et al., 2009; Treier et al., 2012c,a; Osberghaus et al., 2012b,a; Li et al., 2012). Many of the reports which applied model proteins came from academia and demonstrated approaches to elevate high throughput process development beyond mere screenings. In 2008 Susanto et al. showed how model parameter for mechanistic modeling of large scale chromatography can be determined from high throughput data. Besides mechanistic modeling, also different kinds of search algorithms were reported for use in combination with high throughput experimentation. In 2009 Susanto et al. reported successful application of a genetic search algorithm for process optimization and most recently the application of a hybrid simplex algorithm was reported (Konstantinidis et al., 2012a). Besides search algorithms and mechanistic modeling, the literature also includes applications of statistical approaches such as quantitative structure-efficacy relationship modeling (Mazza et al., 2002), ANOVA modeling (Wensel et al., 2008), and design of experiments (e.g., Toueille et al., 2011; Oelmeier et al., 2011; Treier et al., 2012c; Fogle and Persson, 2012).

1.3.2 Analytics Applied in High Throughput Process Development

With very limited exceptions, the complete work published on high throughput process development uses univariate spectroscopic measurements, either as a preliminary analytics or exclusively. The exclusive use of univariate spectroscopic measurements is only sufficient for complete evaluation of the performed experiments when very pure protein is applied. This approach can be used to determine e.g. binding capacity. However, if more proteins are involved other analytical methods are necessary. Besides spectroscopic measurements, high-performance liquid

chromatography (HPLC) is very often applied as analytical mode of choice when performing high throughput process development. Often, the reported assays are based on size exclusion chromatography (Sunasara et al., 2003; Coffman et al., 2008; Kramarczyk et al., 2008; Rodrigo and Nilsson-Välimaa, 2010; Ma et al., 2010; Toueille et al., 2011; Morrison et al., 2011; Treier et al., 2012b). Also hydrophilic interaction chromatography (Sunasara et al., 2003), ion exchange chromatography (Coffman et al., 2008), and reversed phase chromatography (Mazza et al., 2002; Thiemann et al., 2004) have been reported. Besides HPLC, gel electrophoresis and enzyme-linked immunosorbent assay (ELISA) are the most applied techniques for selective or specific protein analysis and quantification within high throughput process development. Gel electrophoresis was most often reported as analytical method of choice in work that dealt with proteins other than monoclonal antibodies: angiotensin-II-generation enzyme (Thiemann et al., 2004), α-amylase and scFv-β-lactamase (Rege et al., 2006), penicillin G acylase (Zhou et al., 2008), and NN[2] (Sanaie et al., 2012). ELISA was reported for quantification of host cell proteins. The use of mass spectrometry was found in only one application where it was used to quantify host cell proteins (Wierling et al., 2007)

1.4 Analytical Bottleneck

In 2005, Bensch et al. published the first review related to high throughput process development. In this review and in a publication that followed in 2006 by Rege et al., one of the key problems of this new technology which they touched upon, but gave no solution to, was the lack of analytical throughput. Bensch et al. described the problem as an analytical bottleneck which creates a situation with an inherent trade off between analytical information and analysis time.

To be able to gain full advantage of the high experimental speed which has been enabled in protein purification process development, the trade off between analytical information and analytical speed must be minimized. Hence, it is necessary to develop robust analytical methods which can match the experimental throughput which is possible today to bring relief to this analytical bottleneck. As size exclusion chromatography and spectroscopic measurements were found to be the most frequently applied analytical methods within high throughput process development, a more detailed description of the possibilities and draw backs of these methods is presented.

1.4.1 Size Exclusion chromatography

Analysis based on size exclusion chromatography is very straight forward to establish. However, this advantage is combined with a very limited number of measures which can be applied to enable faster analysis. Analysis based on size exclusion chromatography is often time consuming. This is reflected in the reported analysis times which range from 10 min to 40 min per sample (e.g., Thiemann et al., 2004; Coffman et al., 2008). Further, the method can only be parallelized by numbering up the available of HPLC systems.

In general, different measures can be taken in order to enhance the speed of analysis based on chromatography. Steeper gradients, higher flow rates or shorter columns can be applied, however, these measures will all influence resolution in a negative direction and hence decrease

[2]The authors inform that it is a small protein consisting of a highly glycosylated 104-amino acid polypeptide expressed in Chinese hamster ovary cells

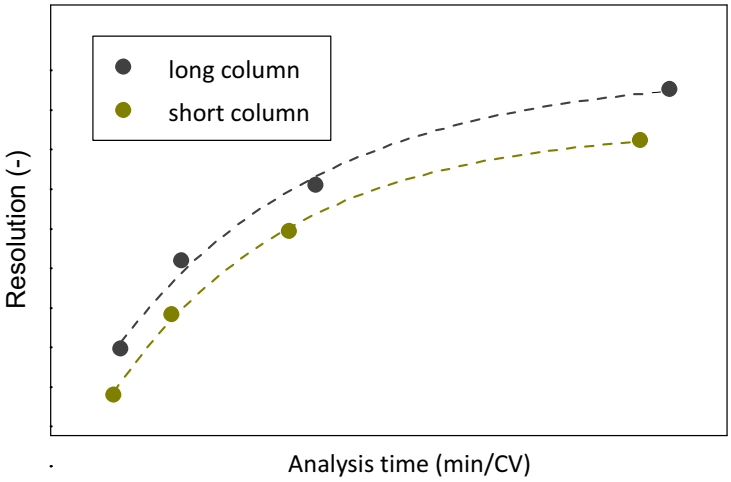


Figure 2: The correlation of analysis time and achieved chromatographic resolution for two different column lengths. The results are based on the separation of the monomer and dimer form af a monoclonal antibody using size exclusion chromatography.

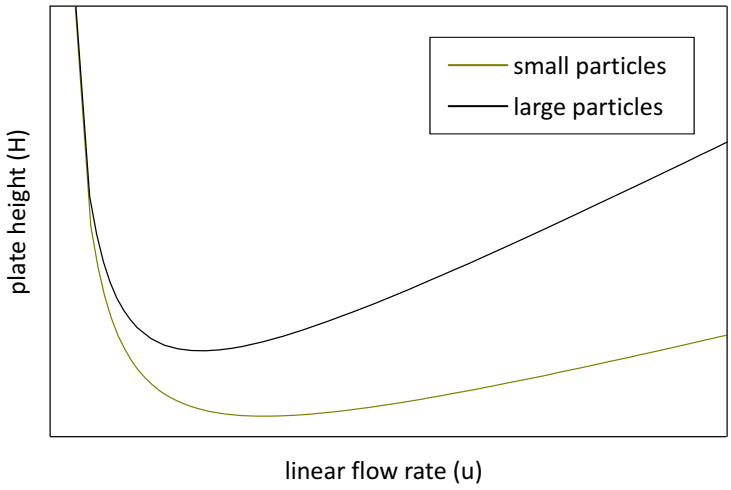


Figure 3: Dependency of theoretical plate height and linear flow rate according to the Van Deemter equation. To illustrate the influence of particle size the dependency is shown for two different particle sizes.

the quality of the obtained analytical information. The van Deemter equation decsribes the resolving power of a chromatographic column as a function of various flow and kinetic parameters. According to the classical Van Deemter equation[3],

$$H = A + \frac{B}{u} + Cu$$

the increase in linear flow rate (u) is directly proportional to an increase in the plate height (H). However in reality the situation is more favorable, in the sense that a two fold increase in linear flow rate will not lead to a two fold increase in plate height. Hence, there is some advantage with respect to analysis time in increasing column length and linear flow rate at the same time (Popovici and Schoenmakers, 2005).

Fig. 2 shows the timely advantage which can be gained by increasing the column length. It can bee seen that with decreasing analysis time (increasing flow rate) the advantage with respect to analysis time decreases. Further, this approach has a limitation in the stability of the resin particles, which is markedly low in the case of size exclusion chromatography. This is due to the fact that resin particles used for size exclusion chromatography must have a very high porosity. As the resistance to mass transfer (C) also has an impact on plate height, the particle diameter

[3]H = plate height, A = Eddy-diffusion, B = diffusion coefficient, C = reisitance to mass transfer, u = linear flow velocity

is important with respect to resolution as it influences the mass transfer. By decreasing the particle size, the plate height is increased and therefore separation efficiency will increase (see Fig 3). If the wish is to increase analysis time, the flow rate can then be increased and resolution maintained when compared to the achievable resolution with larger particles. The limitation to this approach is again the mechanical stability of the resin particles as the back pressure created in the column will increase with decreasing particle size. Hence, all together there is need for alternative ways to increase the throughput of protein analysis based on size exclusion chromatography.

1.4.2 Protein Spectroscopy and Intrinsic Protein mid-UV Absorption

In general, spectroscopic methods for quantitative and qualitative protein analysis are widely applied. Spectroscopic methods are commonly used to asses protein secondary and tertiary structure. Numerous techniques are available such as nuclear magnetic resonance (NMR), X-ray, UV absorption and fluorescence, circular dichroismus (CD), light scatter measurements, and FT-IR spectroscopy. NMR and X-ray are often applied for determination of protein structure. These techniques require very sophisticated instrumentation and are not used for routine measurements. CD spectra are related both to primary sequence and the secondary structure of a protein or peptide. Primarily, CD measurements are used for investigation of changes in conformation e.g. unfolding and refolding and metal-protein interactions. Further, CD is often used to investigate the interaction of chiral drugs with proteins (Bertucci et al., 2010). Of all the spectroscopic methods mentioned, absorption and fluorescence are the most simple both in terms of equipment and interpretation of the results.

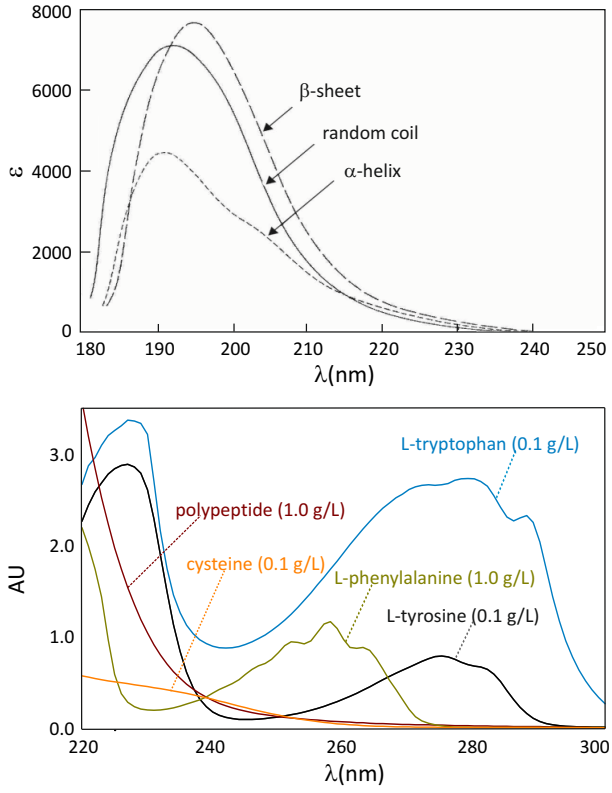


Figure 4: top: Absorbance in the far-UV range of peptide bonds in different conformations (Adapted form Rosenheck and Doty). **bottom:** Spectra of the components mainly contributing to protein mid-UV absorption spectra.

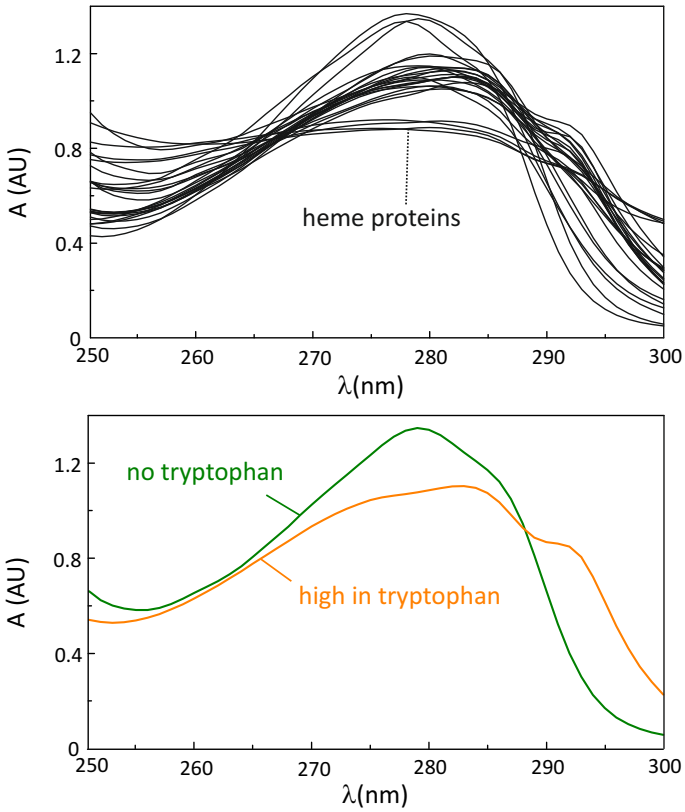


Figure 5: **top:** UV absorption spectra of 27 proteins. **bottom:** Three UV spectra representative of the extremes with in the 27 measured spectra. Proteins containing heme groups were omitted here as they are out of scope in the following work.

The comprehensive advancement of high throughput screening within biological and biochemical research and development has necessitated miniaturized and parallelized methods for protein concentration determination. Hence, many colorimetric and fluorometric protein quantification methods are today adapted to micro plate formats (Stoscheck, 1990; Lorenzen and Kennedy, 1993). Also protein quantification via intrinsic absorption is very well established and widely applied in micro plate formats. Most of the reported applications deal with very pure protein as total protein quantification is then a very straight forward procedure. For absolute protein quantification in protein mixtures via intrinsic absorption, the use of the absorption at 205 nm has been proposed (Kreusch et al., 2003). There has been reported some use of intrinsic absorption spectra measured in plate format for investigation of protein stability (Dasnoy et al., 2011). However, when it comes to selective quantification in mixtures, the general opinion seems to be that absorption and fluorescence measurements in the UV-range lack selectivity, as only the sum signal of the contributing amino acids can be measured (Staub et al., 2011). If a high degree of the spectral data recorded form a protein mixture is caused by the intrinsic absorbance of the proteins, it should be possible to correlate the spectral data to protein concentrations. The robustness of such an approach will depend on the extend of other factors influencing the spectra and to what levels these can be controlled. Further, it will depend on how pronounced the spectral differences are compared to general measurement noise and external influences.

The delocalized electrons of both peptide bonds and aromatic structures in the residues are the originators of the main mid-UV absorption of proteins (see Fig 4). Electromagnetic radiation in the range 200 - 300 nm is defined as mid-UV. The upper part of this range from 250 nm was termed the ‘simple region’ by Wetlaufer due to the small number of contributors to the intrinsic protein absorbance in this region. These are the residues with aromatic structures: tryptophan,

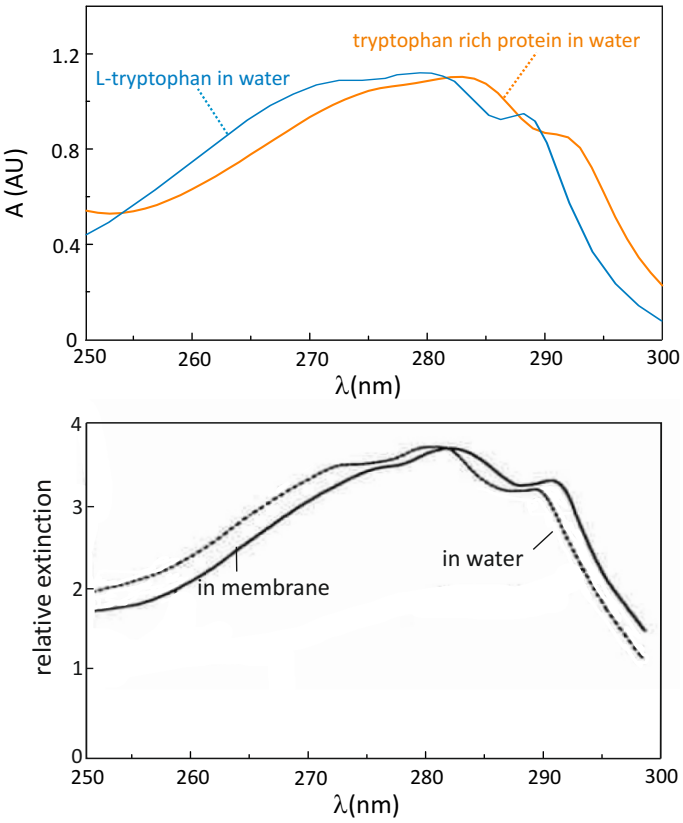


Figure 6: top: UV absorption spectra of L-tryptophan and a tryptophan rich protein, both dissolved in water. **bottom:** UV absorption spectra of the peptide mellitin embedded in a lipid membrane and in water (adapted from Winter and Noll).

tyrosine, and phenylalanine. In the range 200 – 250 nm the absorption of the aromatic residues and the peptide bonds overlap and also the residues containing sulphur absorb in this range. The region below 200 nm is interesting for some applications because the peptide bonds absorb here and particularly because of the relation to conformation.

In Fig. 5 - top, the absorption spectra of a wide range of proteins are displayed. With exception to the proteins containing a heme group, the spectra are rather similar. The main variance is related to the amount of tryptophan present in the proteins. The two spectra in Fig. 5 - bottom, represents main variation which can be expected from protein absorption spectra depending mainly on the tryptophan/tyrosine ratio. Other factors such as the phenyl alanine content and the direct environment of the chromophores will have less significant influence on the spectra in the range 250 – 300 nm. However, the intrinsic absorption of proteins is influenced by the environment of the chromophores. In Fig 6, two examples of a hypsochromic shift caused by changes in the environment of tryptophan are shown. Fig. 6 - bottom compares the absorption of tryptophan in water and the absorption of a protein very rich in tryptophan also dissolved in water. As tryptophan residues are hydrophobic they are very often buried in the inside of the three dimensional protein structure and hence not exposed to the hydrophilic solvent. As can be seen, the water dissolved tryptophan spectra has undergone a significant blue shift which is explained by the change from a predominantly hydrophobic environment to a hydrophilic environment. The same is illustrated in Fig. 6 - bottom comparing the spectra of the peptide mellitin embedded in a lipid membrane and free in aqueous solution. This effect can also be observed in general when comparing the spectra of the three aromatic amino acids with the 27 protein spectra (see Fig. 4 - bottom and 5 - top). The aromatic amino acids exhibit a minimum in their spectra between 230 nm and 240 nm where as the minima of the protein absorption

spectra are shifted to the range 250 – 255 nm.

Extreme pH values will alter the protein absorption spectra due to changes in the ionization stages of the chromophores (Benesch and Benesch, 1955). However, such extreme conditions are not likely to be applied during protein purification. In general, it is expected that the protein absorption spectra will not change significantly under the mild conditions usually present during protein purification processes. This is in contrast to intrinsic protein fluorescence, which is known to be much more environment sensitive than absorptivity.

References

Benesch R, Benesch R. 1955. The acid strength of the-SH group in cysteine and related compounds. Journal of the American Chemical Society 77(22):5877–5881.

Bensch M, Schulze Wierling P, von Lieres E, Hubbuch J. 2005. High Throughput Screening of Chromatographic Phases for Rapid Process Development. Chemical Engineering & Technology 28(11):1274–1284.

Bensch M, Selbach B, Hubbuch J. 2007. High throughput screening techniques in downstream processing: Preparation, characterization and optimization of aqueous two-phase systems. Chemical Engineering Science 62(7):2011–2021.

Bergander T, Nilsson-Välimaa K, Oberg K, Lacki KM. 2008. High-throughput process development: determination of dynamic binding capacity using microtiter filter plates filled with chromatography resin. Biotechnology progress 24(3):632–9.

Bertucci C, Pistolozzi M, De Simone A. 2010. Circular dichroism in drug discovery and development: an abridged review. Analytical and bioanalytical chemistry 398(1):155–66.

Coffman JL, Kramarczyk JF, Kelley BD. 2008. High-throughput screening of chromatographic separations: I. Method development and column modeling. Biotechnology and bioengineering 100(4):605–18.

Dasnoy S, Le Bras V, Préat V, Lemoine D. 2011. High-throughput assessment of antigen conformational stability by ultraviolet absorption spectroscopy and its application to excipient screening. Biotechnology and bioengineering 109(2):502–516.

Fogle J, Persson J. 2012. Effects of resin ligand density on yield and impurity clearance in preparative cation exchange chromatography. II. Process characterization. Journal of Chromatography A 1225:70–78.

Gronberg A, Eriksson M, Ersoy M, Johansson HJ. 2011. A tool for increasing the lifetime of chromatography resins. Mabs 3(2):192–202.

Herrmann T, Schröder M, Hubbuch J. 2006. Generation of equally sized particle plaques using solid-liquid suspensions. Biotechnology progress 22(3):914–8.

Kelley BD, Switzer M, Bastek P, Kramarczyk JF, Molnar K, Yu T, Coffman J. 2008a. High-throughput screening of chromatographic separations: IV. Ion-exchange. Biotechnology and Bioengineering 100(5):950–963.

Kelley BD, Tobler Sa, Brown P, Coffman JL, Godavarti R, Iskra T, Switzer M, Vunnum S. 2008b. Weak partitioning chromatography for anion exchange purification of monoclonal antibodies. Biotechnology and bioengineering 101(3):553–66.

Kittelmann J, Dismer F, Hubbuch J. 2012. 384-well based hts batch chromatography and orientation sensitive qspr modelling of protein adsorption on ion exchange. HTPD, Second International Conference, Avignon, France p. 203.

Konstantinidis S, Chhatre S, Velayudhan A, Heldin E, Titchener-Hooker N. 2012a. The hybrid experimental simplex algorithm - An alternative method for 'sweet spot' identification in early bioprocess development: Case studies in ion exchange chromatography. Analytica chimica acta 743C:19–32.

Konstantinidis S, Kong S, Chhatre S, Velayudhan A, Heldin E, Titchener-Hooker N. 2012b. Strategic Assay Selection for analytics in high-throughput process development: Case studies for downstream processing of monoclonal antibodies. Biotechnology Journal 7(10):1256–1268.

Kramarczyk JF, Kelley BD, Coffman JL. 2008. High-throughput screening of chromatographic separations: II. Hydrophobic interaction. Biotechnology and bioengineering 100(4):707–20.

Kreusch S, Schwedler S, Tautkus B, Cumme Ga, Horn A. 2003. UV measurements in microplates suitable for high-throughput protein determination. Analytical biochemistry 313(2):208–15.

Kumar S, Dalvi DB, Moorthy M, Korde SS, Fondekar KP, Sahasrabudhe SD, Schacht HT, Ekkundi VS, Halik C, Choudhury R, Kumar A, Punekar NS. 2009. Discriminatory protein binding by a library of 96 new affinity resins: A novel dye-affinity chromatography tool-kit. Journal of Chromatography B 877(29):3610–3618.

Li X, de Roo G, Burgers K, Ottens M, Eppink M. 2012. Self-packed filter plates: A good alternative for pre-packed filter plates for developing purification processes for therapeutic proteins. Biotechnology Journal 7(10):1269–1276.

Lorenzen A, Kennedy S. 1993. A fluorescence-based protein assay for use with a microplate reader. Analytical biochemistry 214(1):346–348.

Ma J, Hoang H, Myint T, Peram T, Fahrner R, Chou JH. 2010. Using precipitation by polyamines as an alternative to chromatographic separation in antibody purification processes. Journal of Chromatography B 878(9-10):798–806.

Mazza CB, Rege K, Breneman CM, Sukumar N, Dordick JS, Cramer SM. 2002. High-throughput screening and quantitative structure-efficacy relationship models of potential displacer molecules for ion-exchange systems. Biotechnology and bioengineering 80(1):60–72.

McDonald PJ, Cai N, Cheng J, Lin I, Wong M, Hopp J, Lester P, Kelley B. 2010. Using robotics and high throughput screenings to develop early stage purification processes. HTPD, First International Conference, Cracow, Poland p. 106.

Morrison CJ, Gagnon P, Cramer SM. 2011. Purification of monomeric mAb from associated aggregates using selective desorption chromatography in hydroxyapatite systems. Biotechnology and bioengineering 108(4):813–21.

Oelmeier Sa, Dismer F, Hubbuch J. 2011. Application of an aqueous two-phase systems high-throughput screening method to evaluate mAb HCP separation. Biotechnology and bioengineering 108(1):69–81.

Oelmeier SA, Effio CL, Hubbuch J. 2012. High throughput screening based selection of phases for aqueous two-phase system-centrifugal partitioning chromatography of monoclonal antibodies. Journal of Chromatography A 1252:104–114.

Osberghaus a, Baumann P, Hepbildikler S, Nath S, Haindl M, von Lieres E, Hubbuch J. 2012a. Detection, Quantification, and Propagation of Uncertainty in High-Throughput Experimentation by Monte Carlo Methods. Chemical Engineering & Technology 35(8):1456–1464.

Osberghaus a, Drechsel K, Hansen S, Hepbildikler S, Nath S, Haindl M, von Lieres E, Hubbuch J. 2012b. Model-integrated process development demonstrated on the optimization of a robotic cation exchange step. Chemical Engineering Science 76:129–139.

Popovici ST, Schoenmakers PJ. 2005. Fast size-exclusion chromatography–theoretical and practical considerations. Journal of chromatography A 1099(1-2):92–102.

Rege K, Pepsin M, Falcon B, Steele L, Heng M. 2006. High-throughput process development for recombinant protein purification. Biotechnology and bioengineering 93(4):618–30.

Rodrigo G, Nilsson-Välimaa K. 2010. Developing a MAb Aggregate Removal Step by High Throughput Process Development. Biopharm International 23(4):4–8.

Rosenheck K, Doty P. 1961. The far ultraviolet absorption spectra of polypeptide and protein solutions and their dependence on conformation. Proceedings of the National Academy of Sciences of the United States of America 47:1775–85.

Sanaie N, Cecchini D, Pieracci J. 2012. Applying high-throughput methods to develop a purification process for a highly glycosylated protein. Biotechnology Journal 7(10):1242–1255.

Staub A, Guillarme D, Schappler J, Veuthey JL, Rudaz S. 2011. Intact protein analysis in the biopharmaceutical field. Journal of pharmaceutical and biomedical analysis 55(4):810–22.

Stoscheck CM. 1990. Quantitation of protein. Methods in enzymology 182:50–68.

Sunasara KM, Xia F, Gronke RS, Cramer SM. 2003. Application of hydrophobic interaction displacement chromatography for an industrial protein purification. Biotechnology and Bioengineering 82(3):330–339.

Susanto a, Knieps-Grünhagen E, von Lieres E, Hubbuch J. 2008. High Throughput Screening for the Design and Optimization of Chromatographic Processes: Assessment of Model Parameter Determination from High Throughput Compatible Data. Chemical Engineering & Technology 31(12):1846–1855.

Susanto a, Treier K, Knieps-Grünhagen E, von Lieres E, Hubbuch J. 2009. High Throughput Screening for the Design and Optimization of Chromatographic Processes: Automated Optimization of Chromatographic Phase Systems. Chemical Engineering & Technology 32(1):140–154.

Thiemann J, Jankowski J, Rykl J, Kurzawski S, Pohl T, Wittmannliebold B, Schluter H. 2004. Principle and applications of the protein-purification-parameter screening system. Journal of Chromatography A 1043(1):73–80.

Toueille M, Uzel A, Depoisier JFA, Gantier RA. 2011. Designing new monoclonal antibody purification processes using mixed-mode chromatography sorbents. Journal of Chromatography B 879(13-14):836–843.

Treier K, Berg A, Diederich P, Lang K, Osberghaus A, Dismer F, Hubbuch J. 2012a. Examination of a genetic algorithm for the application in high-throughput downstream process development. Biotechnology Journal 7(10):1203–1215.

Treier K, Hansen S, Richter C, Diederich P, Hubbuch J, Lester P. 2012b. High-throughput methods for miniaturization and automation of monoclonal antibody purification processes. Biotechnology progress 28(3):723–32.

Treier K, Lester P, Hubbuch J. 2012c. Application of genetic algorithms and response surface analysis for the optimization of batch chromatographic systems. Biochemical Engineering Journal 63:66–75.

Wensel DL, Kelley BD, Coffinan JL. 2008. High-throughput screening of chromatographic separations: III. Monoclonal antibodies on ceramic hydroxyapatite. Biotechnology and Bioengineering 100(5):839–854.

Wetlaufer D. 1963. Ultraviolet spectra Of Proteins and Amino Acids. Academic Press, vol. 17 of Advances in Protein Chemistry, pp. 303 – 390.

Wiendahl M, Schulze Wierling P, Nielsen J, Fomsgaard Christensen D, Krarup J, Staby a, Hubbuch J. 2008. High Throughput Screening for the Design and Optimization of Chromatographic Processes Miniaturization, Automation and Parallelization of Breakthrough and Elution Studies. Chemical Engineering & Technology 31(6):893–903.

Wierling PS, Bogumil R, Knieps-gru E. 2007. High-Throughput Screening of Packed-Bed Chromatography Coupled With SELDI-TOF MS Analysis : Monoclonal Antibodies Versus Host Cell Protein. October 98(2):440–450.

Winter R, Noll F. 1998. Methoden der Biophysikalischen Chemie. Stuttgart: B. G. Teubner, 1 st edn.

Zhou J, Wang YH, Chu J, Gou BQ, Zhuang YP, Zhang SL, Yuan ZY. 2008. Penicillin G acylase purification with the aid of high-throughput screening approach. Journal of the Chinese Institute of Chemical Engineers 39(3):195–202.

2 Research Proposal

A process for purification of therapeutic proteins must ensure very high and also very reproducible quality. In order to reach this goal, a detailed control strategy is developed alongside the purification process development. Recently, the interest in applying *in process testing* such as in-line, on-line, and at-line analysis for process control has increased. The motivation behind this is to ensure timely release of processed batches and limit the number of failed batches. The number of failed batches is also closely related to process robustness. Therefore, process robustness and the detailed relation between process parameters and product quality must be addressed thoroughly during development of a protein purification process. As a consequence, the focus of process development has been extended beyond the identification of (optimal) process parameters and there is need for very extensive process characterization during process development. In order to meet this need, the concept of high throughput process development (HTPD) has evolved during the past decade. It is based on miniaturized, parallelized, and automated unit operations for protein purification which enables a high experimental throughput. The concept of HTPD enables an intensified gain of process knowledge during the short time frame available for process development. This is due to the high experimental throughput combined with reduced protein material consumption which together increases the possible number of experiments. In order to get beyond mere screenings, more sophisticated approaches such as factorial designs, intelligent search algorithms, or mechanistic modeling are often includes in the concept of HTPD. The larger amount of process knowledge gained in the miniaturized scale will also reduce the process development work which must be performed at lab-scale. Many of the unit operations applied for protein purification have already been established in a high throughput format. However, to complete the successful establishment of HTPD, efficient analytical tools must be in place as the overall throughput will otherwise decrease significantly. An often applied strategy is to sacrifice analytical information for a higher throughput, however this will limit the possible applications.

This thesis is proposed to address this analytical bottleneck in order to realize a high level of analytical information without compromising throughput. The aim is to create a range of original tools which will increase the efficiency using already applied core technologies for protein analysis. Besides high throughput, the requirements for these analytical tools are robustness, minimal sample preparation, high level of automation and low sample consumption. A further objective of this thesis is to identify or invent analytical methods applicable for in-line, on-line or at-line analysis for use in preparative liquid chromatography.

Analytical chromatography is very often applied for protein analysis during protein purification process development. Hence, a particular focus of this thesis is to enable a more efficient utilization of analytical chromatography systems and to ensure that these approaches will not influence the quality of the obtained analytical information. Further, use of multivariate calibration for analysis based on analytical chromatography will be investigated as an opportunity to enhance speed without compromising the analytical quality.

For in-line process analysis, analytical methods based on spectroscopy are of particular advantage as they are non-invasive and can be performed in flow cells or with fiber optic probes. Also for application within the concept of HTPD, analytical methods based on spectroscopy are of great advantage. Modern spectrophotometers have been developed which can measure samples in micro well plates. Hence, these photometers have a very high sample throughput.

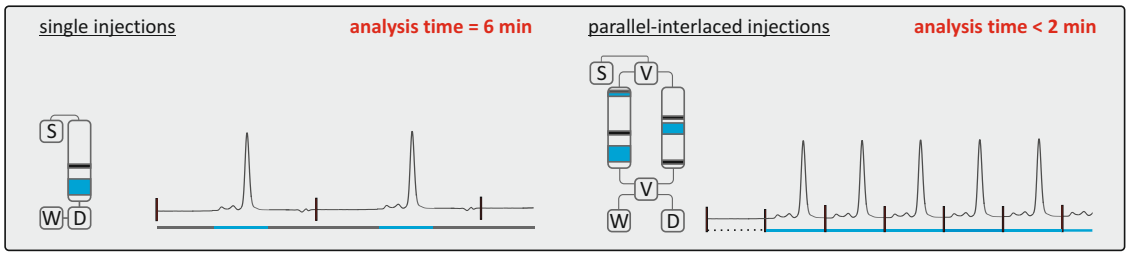
Therefore, another particular focus of this thesis will be on spectroscopic methods both for in-line process analysis and high throughput protein analysis. The aim is to establish quantitative and robust relationships between spectral informations and protein concentration or other quality related information. As protein absorption spectra in the mid-UV range are particular robust, these are thought to be of particular interest. A large investigation into the differences and similarities of mid-UV protein spectra is of particular interest in order to challenge the prevalent opinion that mid-UV absorption spectra lack specificity.

3 Publications & Manuscripts

1. ***A sub-two minutes method for monoclonal antibody-aggregate quantification using parallel interlaced size exclusion high performance liquid chromatography***

   This paper describes a method which enables maximum analytical throughput in size exclusion chromatography. This is achieved by eliminating the lag time inherent in analysis based on size exclusion chromatography. Special focus is on protein aggregate quantification and it is demonstrated that a three fold reduction in analysis time is possible in the particular case described.

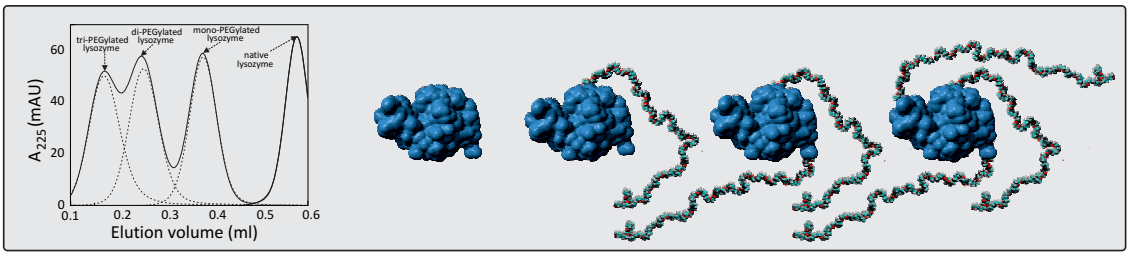
   Journal of Chromatography A, 1218 (2011) 9010-1018



2. ***Rapid quantification of protein-polyethylene glycol conjugates by multivariate evaluation of chromatographic data***

   This paper describes the use of multivariate calibration in the establishment of chromatographic analysis where the achieved resolution is not sufficient for precise quantification of all species based on traditional peak are integration. In the particular case, the samples to be analyzed contained proteins with different numbers of poly ethylene glycol (PEG) chains attached. The capability of performing precise quantification based on a low resolution analytical chromatography, enables faster analysis without loosing analytical quality of the analytical data.

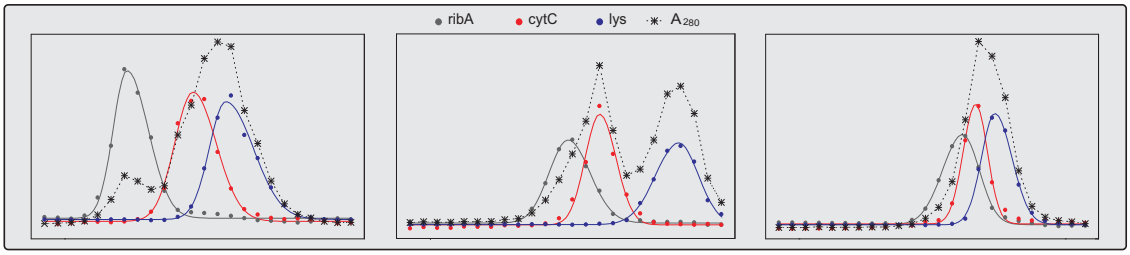
   Journal of Chromatography A, 1257 (2012) 41-47



3. ***A label-free methodology for selective protein quantification by means of absorption measurements***

   Biotechnology & Bioengineering, 108 (2011) 2661-2669

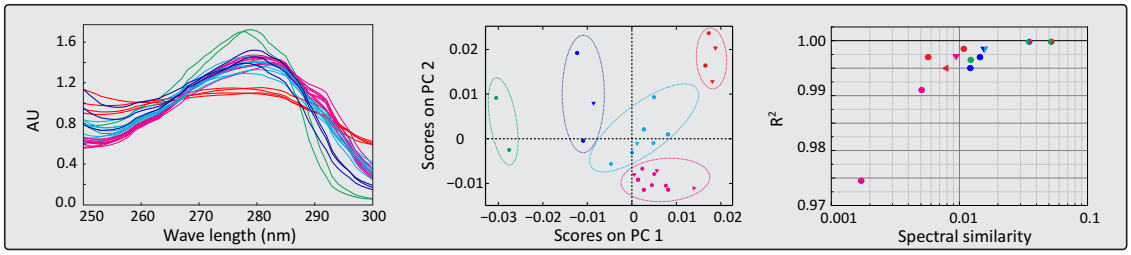
   This paper describes the use of UV absorption spectra of proteins for selective protein quantification in mixtures. A precise quantification method was successfully established for a ternary protein mixture and shown to be valuable in a high throughput process development context. The quantification method based on spectral data was much faster compared to the quantification method based on analytical chromatogrphy otherwise used. Further, it was shown that the quality of the analytical data was the same.



4. ***Selective high throughput protein quantification based on UV absorption spectra***

   Biotechnology & Bioengineering, doi: 10.1002/bit.24712

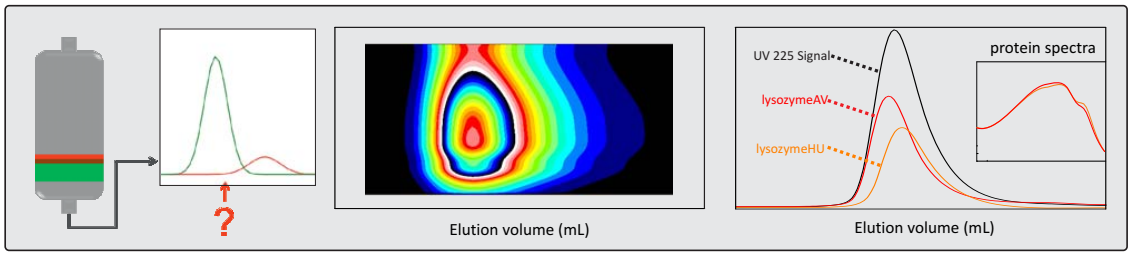
   This paper presents a thorough investigation into protein absorption spectra. The work includes 27 different proteins and their differences and similarities were uncovered and quantified. A special focus was on the definition of a measure for spectral similarity and its correlation with calibration precision.



5. ***In-line analytics for preparative chromatographic separations based on multivariate calibration of protein spectra***

   Presented at the European Congress of Applied Biotechnology 2011, Berlin

   This paper describes the possibility of applying in-line analytics for preparative chromatographic separations based on multivariate calibration of protein spectra. A proof of concept study for separations of two binary protein mixtures is presented.



A Sub-Two Minutes Method for MAb-Aggregate Quantification using Parallel Interlaced Size Exclusion HPLC

Patrick Diederich[+], Sigrid K. Hansen[+], Stefan A. Oelmeier[+], Bianca Stolzenberger, Jürgen Hubbuch*

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

[+] *These authors contributed equally to this publication and appear in alphabetical order*
* *Corresponding author. Tel.: +49 721 608-2557; fax: +49 721 608-6240. E-mail address: juergen.hubbuch@kit.edu*

published the ***Journal of Chromatography A***

Abstract

In process development and during commercial production of monoclonal antibodies (mAb) the monitoring of aggregate levels is obligatory. The standard assay for mAb aggregate quantification is based on size exclusion chromatography (SEC) performed on a HPLC system. Advantages hereof are high precision and simplicity, however, standard SEC methodology is very time consuming. With an average throughput of usually two samples per hour, it neither fits to high throughput process development (HTPD), nor is it applicable for purification process monitoring. We present a comparison of three different SEC columns for mAb-aggregate quantifi cation addressing throughput, resolution, and reproducibility. A short column (150 mm) with sub-two micron particles was shown to generate high resolution (~ 1.5) and precision (coefficient of variation (cv) < 1) with an assay time below six minutes. This column type was then used to combine interlaced sample injections with parallelization of two columns aiming for an absolute minimal assay time. By doing so, both lag times before and after the peaks of interest were successfully eliminated resulting in an assay time below two minutes. It was demonstrated that determined aggregate levels and precision of the throughput optimized SEC assay were equal to those of a single injection based assay. Hence, the presented methodology of paralell interlaced SEC (PI-SEC) represents a valuable tool addressing HTPD and process monitoring.

Keywords: Monoclonal antibody, aggregates, Size Exclusion Chromatography, high throughput analytics, interlaced injection, PI-SEC

1 Introduction

Aggregate levels in monoclonal antibody drugs are a critical quality attribute due to their potential immunogenicity [1, 2]. Aggregates of monoclonal antibodies are often the most abundant product related impurity. The purification process needs to ensure that aggregate levels are reduced to an acceptable level in the final drug product. While the first two steps in a standard mAb downstream process are readily capable of depleting three highly abundant process related impurities, host cell protein, DNA, and water, the reduction of aggregate levels to acceptable levels is often challenging. Thus, monitoring aggregate levels is critical in process development.

One way to reduce process development costs is to increase development throughput. Various process steps have been scaled down to fit into a high throughput process development (HTPD) scheme [3–6]. Additionally, platform processes have been implemented for monoclonal antibody based products, further reducing the efforts needed from process development down to process verification [7]. These improvements have created an analytical bottleneck in process development. To match throughput of the experimentation, reasonably short analysis times need to be achieved.

Size exclusion chromatography (SEC) is the standard method for mAb-aggregate analysis. The standard SEC assay with a throughput of two samples per hour [8, 9] does however not suit a HTPD approach. Several measures are thus in the spotlight to increase throughput in HPLC without changing the analytical technique as such: parallelization and interlacing sample injection. While parallelization using multiple HPLC stations is currently the most often used approach, it is for obvious reasons also the most expensive. Parallelization of multiple columns on a single detector via column switching valves is a way to reduce parallelization cost and has been successfully demonstrated [10]. Most often in this approach, the elution and the regeneration of a chromatographic analysis are separated such that one column regenerates while the other column performs an analysis [11]. In contrast to gradient elution, column regeneration is however not necessary in SEC. Another approach to improve throughput is to run a single column in an interlaced mode. In interlaced chromatography a sample is injected onto the column before the preceding analysis has been completed. This approach requires isocratic conditions. Farnan *et. al* [12] successfully demonstrated its use for aggregate analysis of mAbs and were able to reduce assay time per sample by more than a factor of two from 30 minutes to 14 minutes.

Finally, HPLC equipment capable of higher back pressures has been implemented (most often termed UHPLC) [13]. Shorter columns with smaller column volume and smaller particle sizes can be used with this equipment, thus reducing assay time without sacrificing resolution. While one of the most often used columns for mAb-aggregate analysis has a pressure limit of 7.2 MPa (Tosoh TSKgel 3000 SWxl), two new SEC columns suitable for higher back pressures of 24.1 MPa (Zenix SEC-250 (Sepax Technologies)) to 41.4 MPa (ACQUITY UPLC BEH200 SEC (Waters Corporation)) recently became commercially available.

In this paper, we compare mAb-aggregate analysis performed on these three SEC columns. The columns are compared in terms of assay throughput, resolution, and precision. We demonstrate the application of ACQUITY UPLC BEH200 SEC columns (Waters Corporation) in an interlaced mode as well as by interlaced injections on two columns run in parallel. We demonstrate how throughput can be increased by a factor of 10-15 compared to a standard analysis using a TSKgel 3000 SWxl column. Advantages and disadvantages of the methodology are discussed.

1.1 Theory - Increasing Throughput by Interlacing and Parallelization

While the presented methodology can be applied universally to any type of SEC-column, differences arise in the use of (U)HPLC equipment and the actual pressure rating of the respective SEC-columns and adsorbents. To implement the method developed in this study to its full potential, a prerequisite lies in the use of an (U)HPLC system which is equipped with two independent flow switching valves. An *inlet valve* directs the flow to the columns and autosampler and an *outlet valve* directs the flow from the column outlets to the detector and waste. For maximum throughput two SEC columns can thus be run in parallel applying interlaced injections on each of the two identical columns. The idea of parallel interlaced (PI-) SEC methodology is to eliminate every region of a chromatogram which is not providing any relevant data (e.g. antibody aggregate and monomer). In a first step, data of a single chromatographic SEC analysis therefore serve as a benchmark for the estimation of analysis time and method development as described in the following:

Single Injection

In figure 1 A and 2 A typical chromatograms of common mAb SEC analysis are displayed. The chromatograms can be divided into four main phases. The first phase after sample injection is the initial lag phase (t_{lag}). The time span in which aggregate species and monomer elute is referred to as information phase (t_{inf}). In this work, protein fragments are not considered as species of interest and are not included in t_{inf}. The third phase between monomer peak and the eluting salt fraction is referred to as hold phase (t_{hold}). It is assumed that no protein elute later than the salt fraction of the injected sample. The elution region of salt species is refered to as t_{salt}. A single chromatogram of the sample material provides the user with the retention times of every elution phase for the column used at the specific flow rate. The total time required for the analysis of n samples can be stated as:

$$t_{total} = n \cdot (t_{lag} + t_{inf} + t_{hold} + t_{salt}) \tag{1}$$

Given these retention times, the first step to increase analysis throughput is to eliminate t_{lag} from the resulting chromatograms as explained below.

Interlaced Injection

Farnan *et al.* [12] has described the methodology of interlaced SEC in detail. In a brief, the methodology is based on injecting a subsequent sample before the ongoing analysis of a sample has completed. The subsequent information phase begins immediately after the salt fraction of the preceding sample has eluted. Figure 3 A and B show the transition from a mode of single injection to interlaced injection. By the use of a second timebase (see section 2.2), a separate control program for data aquisition ("program DAD") facilitates distinct chromatograms for each injection and corresponding sample. In figure 3 B it is demonstrated that the lag phase can thus be eliminated from analysis. The total time required for the analysis of n samples can be stated as:

$$t_{total} = t_{lag} + n \cdot (t_{inf} + t_{hold} + t_{salt}) \tag{2}$$

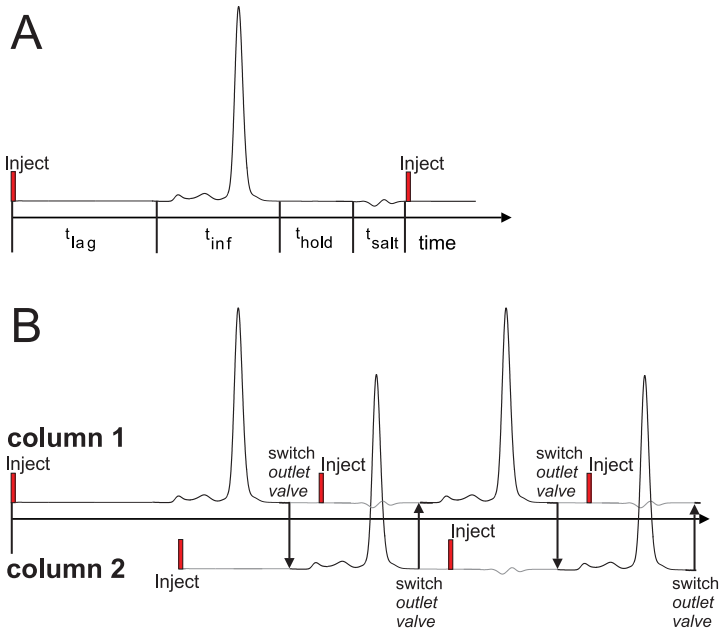


Figure 1: Schematic of PI-SEC methodology applicable in the case of $t_{inf} > t_{hold}$. A: chromatogram of a mAb sample analyzed in single injection mode. Using the elution phases $t_{lag}, t_{inf}, t_{hold}$, and t_{salt}, a PI-SEC program can be set up (B). In this case, samples are injected alternately on two columns, while the *outlet valve* directs the flow from the column outlet to the detector.

Parallel Interlaced Injection

A further increase in throughput can be achieved when applying interlaced injections on two columns which are operated in parallel. Starting from interlaced chromatogaphy, in parallel interlaced SEC the assay time is further reduced by t_{hold}, as is demonstrated in figure 3 B and C. Two switching valves are used to direct the flow alternately between autosampler, two columns and the detector, thus enabling the elimination of t_{lag}, t_{hold} and t_{salt}. In figure 3 D a scheme of the valve switching is displayed. The use of two columns and switching valves require two distinct programs assigned to *timebase 1*, on which pumps, autosampler and column compartment including the switching valves are controlled. The programs contain the same commands, but differ in the direction of both valves switching. As for interlaced chromatography, data aquisition is performed separately by using a second timebase (*timebase 2*) for the detector, now only recording phase t_{inf} of each injected sample.

For programming PI-SEC, three possible cases need to be considered, since elution profiles of a single injection analysis differ in t_{lag}, t_{hold} and t_{salt} depending on column type and sample material. For reason of simplicity, it is assumed that $t_{hold} > t_{salt}$, which is the common case in SEC analysis of antibody samples.

Case-1. $t_{inf} > t_{hold}$:

The first sample is injected on column 1 at:

$$t_1 = 0 \tag{3}$$

The second sample is injected on column 2 at:

$$t_2 = t_1 + t_{inf} \tag{4}$$

The subsequent samples are alternately injected on column 1 and column 2 at times:

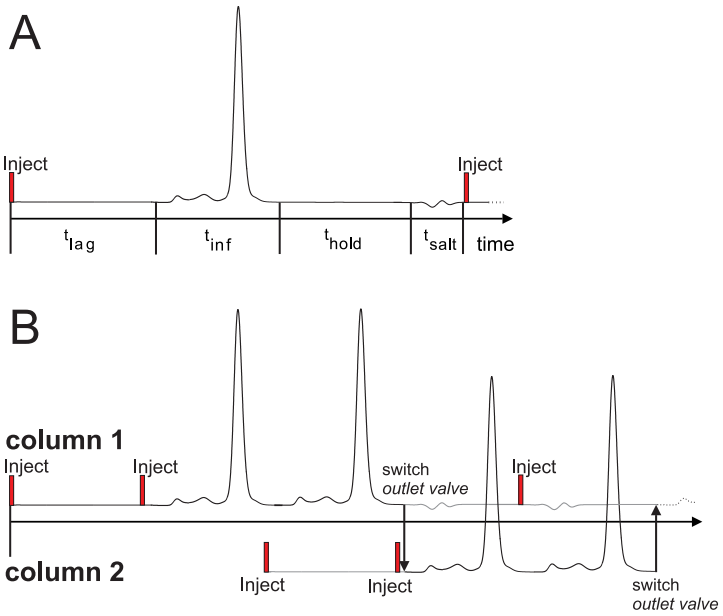


Figure 2: Schematic of PI-SEC methodology applicable in the case of $t_{inf} < t_{hold}$. A: chromatogram of a mAb sample analyzed in single injection mode. Using the elution phases $t_{lag}, t_{inf}, t_{hold}$, and t_{salt}, a PI-SEC program can be set up (B). In this case, two samples are subsequently injected per column before switching to the second column.

$$t_{n,inj} = t_{n,inj-1} + t_{inf} \tag{5}$$

The total assay time for the analysis of n samples can hence be calculated by equation 6. This equation gives the theoretically possible increase in throughput which can be gained via PI-SEC using one single detector.

$$t_{total} = t_{lag} + n \cdot (t_{inf}) + t_{hold} + t_{salt} \tag{6}$$

The *outlet valve* is switched as soon as the information phase of a sample from one column has passed the detector. At that time, the salt peak has completely eluted from the other column. Samples are alternately injected on the two columns and analyzed without any interference of eluting salt fractions. As an example, figure 1 shows a schematic drawing of PI-SEC methodology for the case of $t_{hold} < t_{inf}$.

Case-2. $k \cdot t_{inf} < t_{hold}$:

If $k \geq 1$, one or more informational phases fit into t_{hold} and k additional injections (rounding down of k to whole numbers) on one column become feasible before switching to the second column. The injection times and the time needed for the analysis of n samples can be estimated using the same equations 3 - 6 as given in case one. Figure 2 shows a schematic drawing of the PI-SEC methodology applied for a case 2 elution profile where $1 < k < 2$. Now, two salt peaks elute from one column within the time two information phases elute from the other column.

Although time benefit is the same as in case one, it should be noted that in this mode proteins of multiple, subsequently injected samples pass the salt fraction of the preceding injected samples, whereas for case one the salt fraction of each sample always elute earlier from the column than does the information phase. Multiple injections on

one column is further only applicable, if no species of lower molecular weight than the monomer species is present in the sample material. Otherwise the species of lower molecular weight will elute within the information phase of the subsequent sample injected on the same column.

In the case that $k < 1$ and the *outlet valve* is switched instantly after the information phase of a sample from one column has passed the detector, the salt fraction of the preceding sample has not eluted yet from the second column. Therefore, some additional time (t_{add}) must be added before switching the *outlet valve*. The sum of $t_{add} + t_{inf}$ needs to be greater than $t_{hold} + t_{salt}$. The time needed for the analysis of n samples can be estimated using equation 6, while including t_{add} (9). This delay needs also to be factored in the injection times of the interlaced mode of each column. When the first injection at t_1 is performed, the second injection takes place at:

$$t_2 = t_1 + t_{inf} + t_{add} \tag{7}$$

The injection time of sample n can be hence given by:

$$t_{n,inj} = t_{n,inj-1} + t_{inf} + t_{add} \tag{8}$$

The total assay time for n samples can be calculated using:

$$t_{total} = t_{lag} + n \cdot (t_{inf} + t_{add}) + t_{hold} \tag{9}$$

From a practical aspect it should be mentioned that, if t_{inf} is slightly smaller or exactly equals the sum of $t_{hold} + t_{salt}$, the *outlet valve* is switched just when salt is detected or just arrives at the detector. The baseline determination and an autozero processing of the absorbance signal is hence affected and might lead to unprecise peak integration.

Regarding all described scenarios case one marks the optimal condition for PI-SEC since information phases of samples injected alternately on two columns neither interfere with eluting salt fractions nor are additional times required. With an increasing ratio of t_{inf}/t_{hold}, the benefit of using two columns in parallel over interlaced injection decreases. For the purpose of method robustness, in any of the above described cases additional time for switching the inlet and outlet valves should be implemented: Switching the *inlet valve* should occur a few seconds before the injection takes place and switching of the *outlet valve* should occur a few seconds before the high molecular weight species elute. Thus, baseline determination and peak integration become more precise. To set up the control program, sampling and washing times need to be taken into account. The duration of sampling and washing depends strongly on the used (U)HPLC equipment and might significantly slow down the assay if it exceeds the duration of the information phase. Furthermore, differences in column packing and hence retention times need to be considered.

2 Materials & Methods

2.1 SEC columns

SEC columns from three vendors were used in this work: 1. TSKgel 3000 SWxl (Tosoh Corporation, Tokio, Japan) 2. ACQUITY UPLC BEH200 SEC (Waters Corporation, Milford, MA,

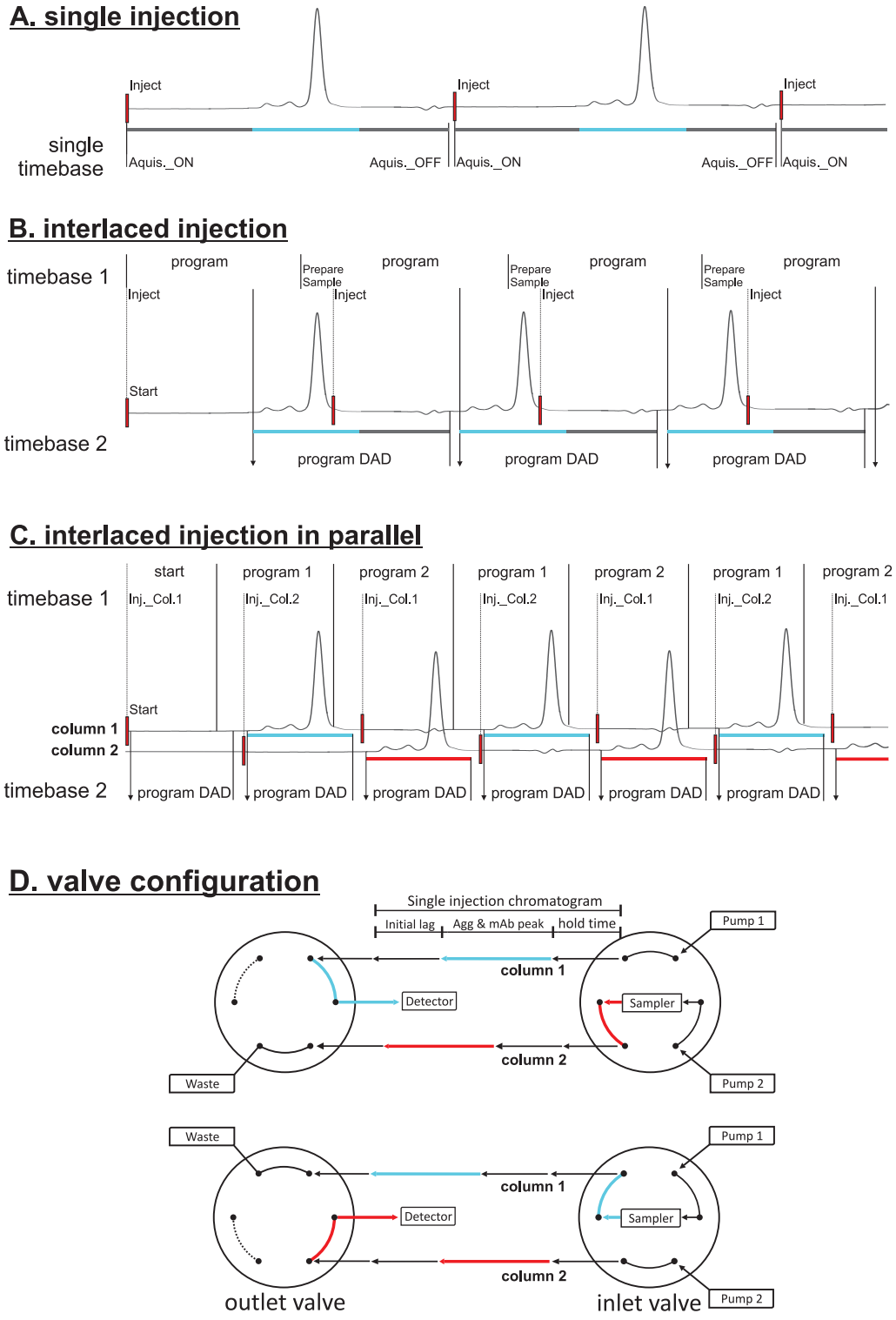


Figure 3: Three modes of operating SEC analysis are displayed. Based on single run chromatograms **(A)** throughput can be improved by interlacing sample injections **(B)** on one SEC column. By using a second timebase (*timebase 2*) for data acquisition, a dedicated chromatogram is generated for every sample injection. Each timebase is controlled by separate programs. Using a second column run in parallel and two timebases **(C)**, throughput can pushed to its theoretical maximum by performing interlaced injections on both columns. Hereby, two programs on *timebase 1* are implemented differing only in the switching direction of the switching valves. A schematic of the configuration of two six-port-valves **(D)** demonstrates the switching procedure which has to be implemented in the control programs 1 and 2.

USA) 3. Zenix SEC-250 (Sepax Technologies, Newark, DE, USA). Columns were fitted with 0.2 μm inlet filter (Opti-Solv EXP , Optimize Technologis, Oregon City, OR, USA). In table I the column properties are listed. The columns differ in macroscopic as well as microscopic dimensions.

Table I Specifications of the HPLC SEC columns used in this study.

vendor description	column dimension	pore size	particle size	maximum pressure	volume column	volume void
ACQUITY UPLC BEH200 SEC	4.6x150 mm	200 Å	1.7 μm	41.5 mPa	2.5 mL	1.97 mL
Zenix SEC-250	4.6x250 mm	300 Å	3.0 μm	24.1 mPa	4.2 mL	3.45 mL
TSKgel 3000 SWxl	7.8x30 0mm	250 Å	5.0 μm	7.8 mPa	14.3 mL	12.23 mL

2.2 UHPLC setup

An UltiMate3000 RSLC x2 Dual system from Dionex (Sunnyvale, CA, USA) was used for UHPLC analysis. The system was composed of two HPG-3400RS pumps, a WPS-3000TFC-analytical autosampler and a DAD3000RS detector. The autosampler was equipped with a sample loop of 5 μl or 20 μl, respectively. The volume of the injection needle was 15 μl, the syringe size was 250 μl. In all experiments, full loop injections were performed. The system included a TCC-3000RS column thermostat to enclose two columns, which were connected to two six-port column switching valves. The *inlet valve* directs the flow between autosampler outlet and column inlets, hence controlling to which column a sample is injected. The *outlet valve* directs the flow between column outlets and UV-detector, hence controlling from which column outlet the UV signal is measured. All column experiments were conducted at 25 °C. For SEC analysis performed in interlaced and parallel-interlaced mode, the system was split in two virtual parts by using two separate timebases. *Timebase 1* controlled pumps, autosampler, valves and column compartment and *Timebase 2* controlled the UV detector. The two timebases were physically linked by connecting a relay assigned to *timebase 1* with an input assigned to *timebase 2*. Switching of the relay in *timebase 1* triggered an input signal in *timebase 2*. This input signal was then used to trigger the UV signal acquisition. By this setup, it was possible to record the information phase of each sample separately.

2.3 Software

Matlab2010a (The Mathworks Natick, ME, USA) was used for data analysis. Chromeleon (6.80 SR10) was used to control the UHPLC equipment and to integrate the elution peaks in the chromatograms. The Chromeleon software was extended to include two timebases.

2.4 Buffer and Sample

SEC analysis were performed using a 0.2 M potassium phosphate buffer at pH 6.2 containing 0.25 M potassium chloride. Buffers were filtered through 0.2 μm filters (Sartorius, Germany) prior to use. When two pumps were used simultaneously (parallel-interlaced protocol), the same buffer preparation was apportioned in two bottles. A proteinA pool of a CHO expressed IgG was used as mAb sample. The concentration was set to a concentration of 1 g/L by dilution with dH2O.

2.5 Aggregate Level and Chromatographic Resolution

For each single injection run, the aggregate level and the resolution was determined. For all interlaced and parallel-interlaced runs only the aggregate level was determined. The aggregate level was defined as the percentage of the species in the mAb sample eluting prior to the monomer. The achieved chromatographic resolution of the mAb monomer and the smallest aggregate (dimer) was calculated based on the EP norm:

$$R = 1.18 \cdot \frac{t_{monomer} - t_{dimer}}{W_{50\%,monomer} + W_{50\%,dimer}} \tag{10}$$

2.6 Single Injection SEC Protocols

The TSKgel column was loaded with 20 μL of sample and the analysis was run at flow rates between 0.235 mL/min and 1.5 mL/min (30 - 188 cm/h). The ACQUITY column was loaded with 5 μL of sample and run at flow rates between 0.05 mL/min and 0.5 mL/min (18 - 181 cm/h). The Zenix column was loaded with 5 μL of sample and run at flow rates between 0.05-0.96 mL/min respectively (18 - 347 cm/h). The exact flow rates are listed in table II.

2.7 Interlaced SEC protocol

For interlaced SEC experiments the chromatography system was split in two virtual parts as described in section 2.2. It should be noted, that this is not a necessary prerequisite in interlaced chromatography, but rather a convenience for the experimenter. By splitting the instrument and running dedicated programs for UV signal acquisition, the relation of chromatogram and injected sample is facilitated. The methodology described in section 1.1 was applied to the use of ACQUITY columns. A single chromatographic run at a flow rate of 0.4 ml/min was used to determine the initial lag phase (t_{lag}) (see fig. 1 A).

In the adapted method, the data acquisition program on *timebase 2* was triggered by switching a relay on *timebase 1* at $t = t_{lag}$ after injection. The withdrawal of the subsequent sample (pulled-loop mechanism) was triggered 51 seconds prior to injection by using the "PrepareNextSample"- command. This avoided additional hold phases between subsequent control programs.

2.8 Parallel-Interlaced SEC Protocol

To improve throughput further, a second column was run in parallel to the first column using two switching valves directing the flow to the columns and to the detector, respectively. The eluate of one column was directed to the waste right after the monomer peak has passed the detector. The eluate of the second column was then directed to the detector, while the salt peak eluted from the first column into the waste. By running both columns simultaneously in an interlaced mode, the maximum possible throughput of the system was realized (section 1.1). In this work, two ACQUITY columns were used at a flow rate of 0.4 mL/min. The time for sample withdrawal was adjusted to 27 seconds (pulled-loop mechanism). Thoroughly washing of the sample loop and the injection needle was set to be performed within 90 seconds.

2.9 Aggregate Spiking Studies

Aggregate spiking studies were conducted in order to evaluate the linearity of aggregate determination of the presented parallel-interlaced methodology. Two solutions containing different levels of aggregate were mixed to control the level of aggregate in the samples. In order to obtain a solution with a high aggregate content, aggregate was isolated from the proteinA pool. This was done by loading the mAb sample onto a Poros 50 HS (GE Healthcare, Germany) column. Before loading the column, the mAb sample had been adjusted to a conductivity of 15 mS/cm and a pH of 5.5. These conditions had been found to provide high selectivity for mAb aggregates compared to mAb monomer. The elution was performed with a sodium chloride gradient from 10-150 mM in 20 mM MES buffer at pH 5.5. The eluate was collected in fractions, analyzed by SEC and merged to create an aggregate pool with approximately 50 % aggregate. Seventeen aggregate levels were tested ranging from 2.1 to 48.7 %. The samples were first analyzed on two different ACQUITY columns in single injection mode, where each sample was measured sixfold. Subsequently, the presented parallel-interlaced assay was applied, using the same two columns and the same samples which were measured sixfold each. The results were compared in terms of coincidence of the linear regression between expected aggregate level and aggregate level determined via the different approaches.

3 Results & Discussion

SEC columns from three different vendors with different particle size, pore size, and length were applied for mAb aggregate quantification. In contrast to the TSKgel column, the ACQUITY and the Zenix columns have entered the market recently. The TSKgel column has been on the market for almost 25 years and a literature survey revealed a marked preference for this particular column in relation with mAb analysis (data not shown). The chosen columns were compared in terms of generated chromatographic resolution, throughput and precision of aggregate quantification. Based on the results, the best suited column and flow rate was chosen and used to establish a in throughput optimized assay by combining interlaced injections with parallel operation of two SEC columns.

3.1 Single Injections

Three different columns were used to analyze identical mAb samples. Figure 4a shows all three resulting chromatograms. The applied flow rates were 108 cm/h for the ACQUITY, 116 cm/h for the Zenix column, and 126 cm/h for the TSKgel column. For comparability, the chromatograms were normalized with respect to void volume of the respective column (figure 4b). The void volume of each column was defined as the elution volume of the sample buffer. These are listed in table I.

The normalized chromatograms revealed similar elution patters for all columns in which the mAb species eluted over a range from approximately 0.45 to 0.85 void volumes. The elution order, based on normalized elution volume of the monomer species from the three different columns ($V_{ACQUITY} < V_{TSKgel} < V_{Zenix}$) correlated with the decreasing pore size of the column matarial (ACQUITY: 200 μm, TSKgel: 250 μm, Zenix: 300 μm). The elution profiles generated by the Zenix and the TSKgel column exhibited a more widely stretched elution of the aggregate species. At very low flow rates, these two columns also revealed a third aggregate species

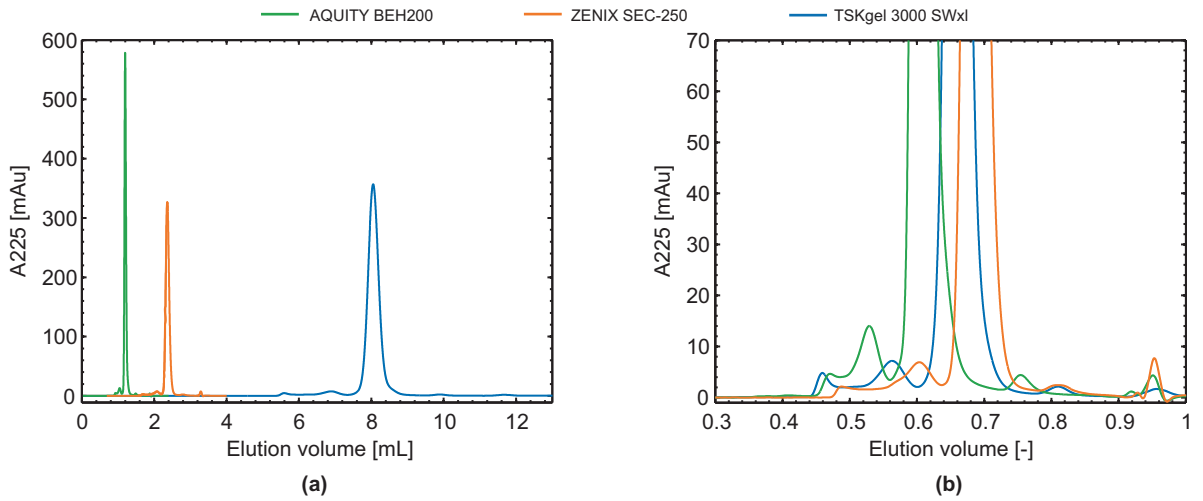


Figure 4: a: Overlay of single injection chromatograms of the mAb sample (1.0 g/L) analyzed on three different SEC columns. b: For comparability, elution volumes were normalized to column void volumes.

in the mAb sample which eluted in between the two main aggregate species (data not shown). However, if an analytical assay aims for the total aggregate level, a resolution of single aggregate species is not necessary. In such a case, the most important parameter is the resolution of the smallest mAb aggregate species (dimer) and the mAb monomer. Hence, in the following the term resolution will refer only to the resolution of mAb monomer and dimer species.

3.1.1 Aggregate Levels and Precision

The determined resolution, aggregate level, and coefficient of variation (cv) for each applied flow rate and column are listed in table II. The columns were shown to generate different results regarding aggregate level, even though the same mAb sample was analyzed. Using the TSK column, the highest and most stable aggregate level (4.80 % ± 0.08) over the tested range of flow rates was determined. Using the ACQUITY column, a lower mean aggregate level was determined (4.17 % ± 0.44) and further the determined aggregate levels exhibited an increase with increasing flow rate (3.79 % - 5.02 %). However, the precision resulting from each tested flow rate was comparable to the accuracy obtained with the TSKgel column ($cv_{mean,TSKgel}$ = 0.91, $cv_{mean,ACQUITY}$ = 0.87). The overall aggregate level determined using the Zenix column (4.20 % ± 0.35) was similar to the one obtained with the ACQUITY column, however the accuracy of the results was lower compared to both other columns ($cv_{mean,Zenix}$ = 1.38). As for the ACQUITY column, the aggregate level determined with the Zenix column exhibited an increase with increasing flow rate (3.69 % - 4.54 %). For all columns, a tendency of higher precision at medium flow rates was observed.

3.1.2 Resolution vs. Analysis Time

The main objective of the presented work, was to establish an ultra-rapid SEC assay for mAb aggregate quantification. Due to the different column dimensions, the correlation between resolution and flow rate does not transmit directly to analysis time. To give an overview of the direct relation between analysis time and chromatographic resolution, the resolution generated for each flow rate and column was plotted as a function of the required time per analysis (figure

Table II: Aggregate levels determined for a mAb sample using three different columns. Each column was operated at several different flow rates. All displayed results are based on six replicates.

TSKgel 3000 SWxl				
Flow rate		Aggregate	cv	resolution
(cm/h)	(mL/min)	(%)	(%)	(-)
30	0.235	4.74	1.91	1.85
44	0.352	4.87	0.85	1.77
63	0.50	4.87	0.60	1.71
94	0.75	4.84	0.27	1.59
126	1.00	4.83	0.52	1.50
157	1.25	4.79	0.48	1.41
188	1.50	4.64	1.75	1.34
ACQUITY UPLC BEH200 SEC				
Flow rate		Aggregate	cv	resolution
(cm/h)	(mL/min)	%	%	(-)
18	0.05	3.79	1.94	1.66
27	0.075	3.90	1.00	1.60
36	0.10	3.90	0.99	1.61
72	0.20	4.00	0.48	1.56
108	0.30	4.16	0.27	1.52
144	0.40	4.36	0.94	1.47
181	0.50	5.07	1.52	1.45
Zenix SEC-250				
Flow rate		Aggregate	cv	resolution
(cm/h)	(mL/min)	%	%	(-)
18	0.05	3.69	2.33	1.35
27	0.075	3.96	0.97	1.33
36	0.10	4.11	0.58	1.30
116	0.32	4.28	0.97	1.14
231	0.64	4.62	1.53	1.01
347	0.96	4.54	1.91	0.92

5). The evaluation was performed in sequential mode, thus time per analysis equals time needed for processing a single column volume (CV). In generel, the decrease in resolution correlated with the particle size of the column material. We found that at assay times above 20 min, the TSKgel column achieved the highest resolution of the columns tested. The resolution achieved under these conditions ranged from 1.59 to 1.85. However, in most cases, a resolution of 1.5 is sufficient for precise quantification. Hence, the high resolution achieved by the TSKgel column at the lower end of the tested flow rates will in some cases be disadvantageous as an unnecessary low throughput is the consequence of the achieved yet dispensable resolution. At lower assay times (increased flow rates) the resolution achieved with the TSK column was shown to decrease faster compared to the ACQUITY column. Of all columns, the ACQUITY column was shown to generate the highest resolution at assay times below 20 min. This finding correlates with the smaller particle size of the ACQUITY column. The tested Zenix column was outperformed by the TSKgel and ACQUITY columns with respect to resolution at all tested assay times. One advantage of the Zenix column was the potentially lower assay time, but the low resolution under these conditions were shown to generate imprecise results (see table II) . However, assay times down to 13 min generated adequate precision (cv < 1) despite the low resolution. Hence, taking the relative low cost for the Zenix column compared to the TKSgel and the ACQUITY column into consideration (which exhibits a factor of 1:1.5:2), this column could pose a favorable alternative to the otherwise comprehensive use of the TSKgel column.

Sufficient resolution (~ 1.5) and precision (cv < 1) was shown feasible with the ACQUITY column even at very low analysis times. This clearly favours the ACQUITY column for development of a high throughput parallel-interlaced SEC assay. A flow rate of 0.4 ml/min was chosen, both to guarantee sufficient accuracy and also not to operate the column close to maximal flow

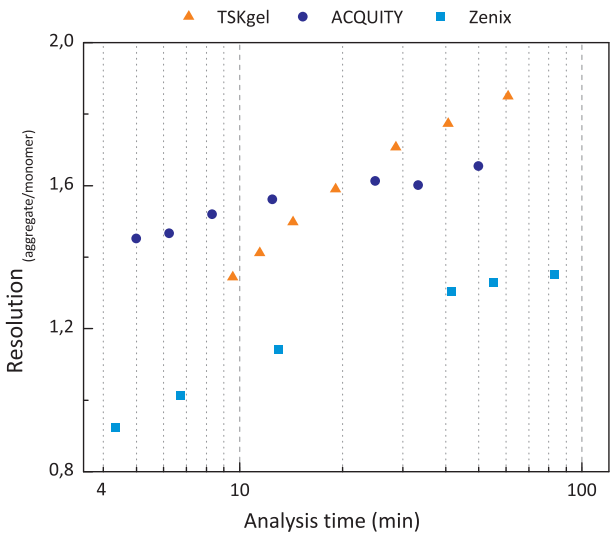


Figure 5: Achieved chromatographic resolution for each tested column displayed as a function of the analysis time. Each data point represents the mean value of six measurements.

rate.

The findings presented above are based on measurements performed with only one column per column type. Hence, the conclusions do not take batch and packing variability into consideration. This influence is shown in the studies below. Further, a buffer optimization was not in the scope of this work and changes in performance under other buffer conditions can not be ruled out.

3.2 Interlaced SEC

Twenty five injections of the same load material were performed on three ACQUITY columns in interlaced mode. Average analysis time per sample was 3:27 minutes at a flow rate of 0.4 ml/min. Figure 6A shows the resulting A225 trace from the detector. It can be seen that the initial lag time was successfully cut from the analysis time. In this mode of operation 1.43 samples were analyzed per column volume. While aggregate levels resulting from all three column were in the same range and normally distributed around their mean, pairwise t-tests ($\alpha = 0.01$) showed that all results differed statistically significantly from one another. The first column resulted in a mean aggregate level of 5.08 % with a standard deviation of 0.04. The second column yielded mean 5.02 % with a standard deviation of 0.05. The third column yielded mean 4.91 % with a standard deviation of 0.04.

By interlacing injections and switching to a column of smaller volume and particle size, the assay time was reduced from 14 minutes reported by Farnan [12] to 3:27 minutes. The obvious advantage of using interlaced injections lies in the improved throughput. However, special care has to be taken in order to correctly relate sample and chromatogram. By splitting the instrument into two virtual parts (timebases) a comfortable solution to this problem can be achieved. While throughput was increased, there was still room for optimization. First, column utilization is not optimal as only the initial lag phase is eliminated by interlaced injections. Second, the next sample was not injected until 15 seconds after the salt fraction of the preceding sample had eluted.

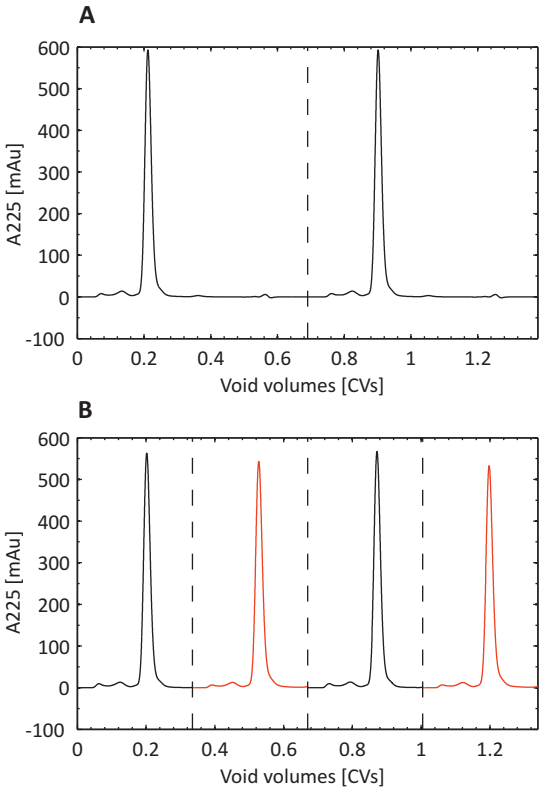


Figure 6: A: A225 absorption data of two injections run in interlaced mode on the ACQUITY column. The dashed line represents the limit between the two samples. 1.43 samples could be analyzed per CV in this mode of operation. B: A225 absorption data of four injections run in parallel-interlaced mode on two ACQUITY columns. The dashed lines represent the moments of switching the column outlet valve to the detector for the subsequent sample. Separate result files are generated for each sample as delimited by the dashed lines. Equally colored lines represent samples analyzed over the same column. Three samples were analyzed per CV in this mode of operation

3.3 Parallel-Interlaced SEC

Program Parameters

By parallelization of two ACQUITY columns operated with interlaced sample injections, chromatograms containing only the aggregate and monomer areas could be generated. As described in chapter 1.1, the control program was set up based on a single run at a flow rate of 0.4 ml/min. The operation commands of the Chromeleon software and the corresponding times in the control programs of *timebase 1* and *timebase 2* are summarized in table III. t_{lag} was set to 2:00 minutes. t_{inf}^{min}, the minimal possible analysis time was 1:12 minutes. Twenty-four seconds were added to t_{inf}^{min} to make the method more robust against changes in sample composition. t_{inf} used for programming the method was thus 1:36 minutes. The determined t_{hold} was 1:18 minutes. A sequence of samples was first started with a dummy run in which the first sample is injected but no protein elutes. DAD data acquisition thus generated a blank sample. Immediately after DAD data acquisition has ended, the *outlet valve* was switched. Fifteen seconds were added to the method to ensure a stable baseline after switching the *outlet valve* (t_{add}^{1} = 0:15). Next, the *inlet valve* was switched. Three seconds were added to the method to flush the autosampler prior to injection (t_{add}^{2} = 0:03). Triggering the data acquisition was performed three seconds after the sample injection by using the following commands: after the `Inject` command triggered sample injection in *timebase 1*, a `Relay.State = ON` command switched a relay which was connected to an input via cable. A `wait Input.State = ON` as first command in the control method for *timebase 2* triggered the start of this control method and thus of DAD data acquisition as soon as relay 3 was switched. 1:27 minutes later the next sample withdrawal

was started using the `PrepareThisSample` command. 0:09 minutes afterwards, DAD data acquisition was stopped thus closing one cycle of sample injection and detection. The process of sample withdrawal took 27 seconds and was performed during the last nine seconds of t_{inf} of the preceding sample and the t^1_{add} and t^2_{add} after switching the *outlet valve* and *inlet valve*.

In general, the operating speed of the autosampler was found to be an important factor when programming the control method. Slower autosampling equipment might hinder the implementation of the method. Compared to the data presented, faster autosampling procedures, for example by using a inline split-loop autosampler instead of the used pulled-loop would take the method closer to its theoretical minimum of 1:12 minutes.

To analyze a batch of samples, two batch files were created, one for each timebase. The batch file for *timebase 1* contained two different control programs with each used for every other sample. The two control programs were equal but for the valve switching commands. The batch file for *timebase 2* consisted of a sequence of the DAD control program. The two batch files were started simultaneously.

Method Performance

Fifty injections (25 on each column) of the same mAb load material were performed in parallel-interlaced mode. The analysis time for this batch was 1:57 minutes per sample. Figure 6B shows the resulting detector signal at a wavelength of 225 nm of four consecutive samples. Compared to the standard analytic (single injections, TSKgel column), throughput was improved by 10-15x. Compared to single injections on the same column type, throughput was increased approximately 3x. In accordance to equation 6 the analysis time per sample for n samples can be calculated as follows:

$$t_{analysis} = \frac{t_{lag}}{n} + (t_{inf} + \sum t_{add}) + \frac{t_{hold}}{n} \tag{11}$$

which in our case amounts to

$$1{:}57 = \frac{2{:}00}{50} + (1{:}36 + (0{:}15 + 0{:}03)) + \frac{1{:}18}{50} \tag{12}$$

It is obvious that t_{lag} and t_{hold} do not contribute substantially to the overall analysis time when running the columns in parallel interlaced mode.

A statistical analysis of the results was performed and two data points differing more than 3 standard deviations from the mean value were excluded from further analysis. Average aggregate content detected was 5.03% with a standard deviation of 0.26 This rather large standard deviation was due to differing results from the two separate columns used. Mean aggregate level determined on the first column was 5.27% with a standard deviation of 0.06. Mean aggregate level determined on the second column was 4.78% with a standard deviation of 0.05. While both columns yielded aggregate levels normally distributed around their mean value, results from both column differed statistically significantly as determined by a t-test ($p < 0.1\%$).

The presented method was shown to achieve large improvements of throughput for the particular analysis investigated. Certain prerequisites for achieving these improvements for any given chromatographic assay should be noted. First, the method works for isocratic elutions only, which is the case for SEC and some IEC/HIC analytics. Second, the improvement in assay throughput is related to the ratio of the information to the non-information phases of the chromatogram as only those parts containing no valuable information can be eliminated from the chromatogram. In the case described here, the information phase was approximately 24%

Table III Control parameters used to control *timebase 1* (TB1; autosampler, pumps, column compartment including switching valves) and *timebase 2* (TB2; DAD). The commands for injecting five samples are shown. The initial flow path was: sampler → column 1 → DAD. Column "Time" shows the actual time during the analysis. Columns "TB 1" and "TB 2" show the time points programmed into the control programs for timebase 1 and timebase 2. The "action" columns adjacent to the "TB 1" and "TB 2" columns contain the commands used at the corresponding time point. Column "Sample" shows the time during which a sample is on a specific column. The first data acquisition on *timebase 2* generates a chromatogram ('dummy #') that only contains the t_{lag} of the first sample. (The two control programs of *timebase 1* differ only in switching valve commands. The data acquisition program on *timebase 2* is started by switching a relay ON.)

Time	Sample	TB 1	Action	Flow path	Action	TB 2
00:00		-0:27	Prepare sample #1			
00:27	**sample #1 / column 1**	0:00	Inject + Start Wash		Wait Input.state = ON	00:00
00:30		0:03	Relay.State = ON		Data Acquisition On	
01:54		1:27	Pump Acquisition OFF			
01:57		1:30	method end		**dummy #**	
01:57		-0:27	Prepare sample #2			
02:06		-0:18	switch outlet valve	column 1 → DAD	Data Acquisition Off	1:36
02:12					method end	1:42
02:21		-0:03	switch inlet valve	sampler → column 2		
02:24	**sample #2 / column 2**	0:00	Inject + Start Wash		Wait Input.state = ON	00:00
02:27		0:03	Relay.State = ON		Data Acquisition On	
03:51		1:27	Pump Acquisition OFF			
03:54		1:30	method end		**sample #1**	
03:54		-0:27	Prepare sample #3			
04:03		-0:18	switch outlet valve	column 2 → DAD	Data Acquisition Off	01:36
04:09					method end	01:42
04:18		-0:03	switch inlet valve	sampler → column 1		
04:21	**sample #3 / column 1**	0:00	Inject + Start Wash		Wait Input.state = ON	00:00
04:24		0:03	Relay ON		Data Acquisition On	
05:48		1:27	Pump Acquisition OFF			
05:51		1:30	method end		**sample #2**	
05:51		-0:27	Prepare sample #4			
06:00		-0:18	switch outlet valve	column 1 → DAD	Data Acquisition Off	01:36
06:06					method end	1:42
06:15		-0:03	switch inlet valve	sampler → column2		
06:18	**sample #4 / column 2**	0:00	Inject + Start Wash		Wait Input.state = ON	00:00
06:21		0:03	Relay.State = ON		Data Acquisition On	
07:45		1:27	Pump Acquisition OFF			
07:48		1:30	method end		**sample #3**	
07:48		-0:27	Prepare sample #5			
07:57		-0:18	switch outlet valve	column 2 → DAD	Data Acquisition Off	01:36
08:03					method end	01:42
08:12		-0:03	switch inlet valve	sampler → column 1		
08:15	**sample #5 / column 1**	0:00	Inject + Start Wash		Wait Input.state = ON	00:00
08:18		0:03	Relay.State = ON		Data Acquisition On	
09:42		1:27	Pump Acquisition OFF			
09:45		1:30	method end		**sample #4**	

of the entire chromatogram. Samples and analysis tasks making use of a larger portion of the chromatogram are amenable to the methodology as described in section 1.1 but might not yield throughput improvements as high as those reported here.

Reliability, robustness, and quantitativeness are the hallmarks of analytical SEC chromatography for mAb-aggregate quantification. Thus, it is preferred over other, even faster analytical methods such as capillary gel electrophoresis. The presented methodology increased sample throughput to an extend that it matches the speed of high throughput experimentation with-

out changing the robust, underlying analytical principle. More detailed studies of aggregation and aggregate depletion during process development and production of mAb based pharmaceuticals can thus be performed.

3.4 Aggregate spiking studies

Aggregate spiking studies resulted in a linear response of the detected aggregate level to the expected aggregate level in the sample throughout the entire range tested (2.1% to 48.7%). The linear regression of measured aggregate level versus expected aggregate level were compared for the two separate columns used and two modes of operation (single and parallel-interlaced injection mode). The linear regression results were found to coincide, slope and intercepts were found to be statistically not different. The overall regression of expected versus measured value was resulted in a R^2 value of 0.9993 with an intercept fixed at 0 and a resulting slope of 1.01. This underlines our conclusion that the method presented herein can replace the standard method of running SEC columns for mAb-aggregate analysis and that the column used is well suited for the analysis task investigated. In theory, increasing aggregate levels could have increased the aggregate peak area to an extend where either monomer-aggregate peak resolution would decrease or where column valve switching times might have had to be adjusted. However, neither was found leading to the conclusion that the presented method is robust regarding aggregate levels of up to 48.7%. Aggregate levels below 2.1% were not investigated owing to the sample material at hand. However, the authors find no reason to believe that lower aggregate levels would pose a problem to the method.

4 Conclusion

In case of total mAb aggregate quantification, we find the ACQUITY column to be the best suited choice of the tested columns, as it enables more than a two fold improvement in throughput when compared to the TSKgel column (assay time comparison at a resolution of 1.5, see figure 5 and table II). Further, due to the relatively low influence of flow rate on the separation which was found for the ACQUITY column, assay throughput can be increased further without compromising resolution significantly. The ACQUITY column also offers the benefits of lower buffer consumption and lower sample volume, latter being of great importance when performing HTPD.

A new methodology to improve throughput for SEC mAb analysis applied in biopharmaceutical science was demonstrated in this paper. By combining interlaced injections with parallel operation of two columns, near optimum utilization of SEC columns for the quantification of monomer and aggregate of a monoclonal antibody solution was achieved. Assay time was reduced to 1:57 minutes per sample as compared to 20-30 minutes using standard analytical protocols. Resulting aggregate levels were found to be comparable between different columns and different modes of operation. As an added benefit, heterogeneity between separate columns is factored into the results by using this method. With analysis times in the range of 2 minutes per sample the method presented in this paper is well suited for current high throughput pharmaceutical process development and process monitoring.

References

[1] M. Vzquez-Rey, D. a. Lang, Aggregates in monoclonal antibody manufacturing processes., Biotechnology and bioengineering 108 (7) (2011) 1494–508.

[2] A. S. Rosenberg, Effects of protein aggregates: an immunologic perspective., The AAPS journal 8 (3) (2006) E501–7.

[3] J. L. Coffman, J. F. Kramarczyk, B. D. Kelley, High-Throughput Screening of Chromatographic Separations: I. Method Development and Column Modeling, Biotechnology and Bioengineering 100 (4) (2008) 605–618.

[4] P. S. Wierling, R. Bogumil, E. Knieps-Grünhagen, J. Hubbuch, High-Throughput Screening of Packed-Bed Chromatography Coupled With SELDI-TOF MS Analysis : Monoclonal Antibodies Versus Host Cell Protein, Biotechnology and Bioengineering 98 (2) (2007) 440–450.

[5] M. Wiendahl, P. S. Wierling, J. Nielsen, D. F. Nielsen, J. Krarup, A. Staby, J. Hubbuch, High Throughput Screening for the Design and Optimization of Chromatographic Processes - Miniturization, Automation, and Parallelization of Breakthrough and Elution Studies, Chemical Engineering & Technology 6 (31) (2008) 893–903.

[6] S. Oelmeier, F. Dismer, J. Hubbuch, Application of an Aqueous Two-Phase Systems High-Throughput Screening Method to Evaluate mAb HCP Separation, Biotechnology and Bioengineering 108 (1) (2011) 69–81.

[7] B. Kelley, Industrialization of mab production technology: the bioprocessing industry at a crossroads., mAbs 1 (5) (2009) 443–52.

[8] J. F. Kramarczyk, B. D. Kelley, J. L. Coffman, High-throughput screening of chromatographic separations: Ii. hydrophobic interaction., Biotechnology and bioengineering 100 (4) (2008) 707–20.

[9] Tosoh Bioscience LLC, Analysis of human immunoglobulins by size exclusion chromatography, application note G00951.

[10] R. King, C. Miller-Stein, D. Magiera, J. Brann, Description and validation of a staggered parallel high performance liquid chromatography system for good laboratory practice level quantitative analysis by liquid chromatography/tandem mass spectrometry., Rapid communications in mass spectrometry : RCM 16 (1) (2002) 43–52.

[11] M. Nelson, J. Dolan, Parallel chromatography - double your money, LC GC North America 22 (4) (2004) 338–343.

[12] D. Farnan, G. Moreno, J. Stults, A. Becker, G. Tremintin, M. van Gils, Interlaced size exclusion liquid chromatography of monoclonal antibodies., Journal of chromatography. A 1216 (51) (2009) 8904–9.

[13] M. Swartz, Uplc: An introduction and review, Journal of Liquid Chromatography & Related Technologies 28 (7) (2005) 1253–1263.

Rapid Quantification of Protein-PEG Conjugates by Multivariate Evaluation of Chromatographic Data

Sigrid K. Hansen, Benjamin Maiser, and Jürgen Hubbuch

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

Published in the ***Journal of Chromatography A***

Abstract

Size exclusion chromatography (SEC) is often applied for characterization protein-polyethylene glycol (PEG) conjugates regarding the number of attached PEG chains (PEGamers). SEC analysis is advantageous as it is precise, robust, and straightforward to establish. However, most SEC based assays have a maximal throughput of a few samples per hour. We present a strategy to increase analytical throughput based on combining a short column with a fast flow rate, and finally multivariate calibration in order to compensate for the resolution lost in the trade off for speed. Different multivariate approaches were compared and multilinear regression was shown to result in the most precise calibrations. Further, a dynamic calibration approach was developed in order to account for changes in column performance over time. In this way, it was possible to establish a highly precise assay for protein PEGamer quantification with a throughput of 30 samples per hour.

Keywords: multivariate calibration; high throughput analytics; high throughput experimentation; PEGamer quantification; protein PEGylation; high throughput process development

1 Introduction

Attaching polyethylene glycol (PEG) polymer chains to proteins (PEGylation) has been shown to improve their pharmacokinetics. The increase in size reduces renal clearance of the therapeutic and there is proof of immunogenicity and antigenicity being reduced by PEGylation [1, 2]. Further, solubility of hydrophobic proteins can be increased by PEGylation [3]. The PEGylation process determines the PEGamer distribution and positions of the PEG molecules attached to the protein. If random PEGylation is performed, the product of the reaction will be very heterogeneous. This poses a problem in a regulative environment which demands defined products of extremely high homogeneity. In general, there are two ways to reach a homogeneously PEGylated product. Either the product mixture of the random PEGylation can be purified to contain only the desired PEGylated species or site specific PEGylation can be performed.

In our lab, small scale PEGylation experiments performed automated on liquid handling stations have been established. This enables high throughput screenings of PEGylation process parameters and is a part of a general trend where automated high throughput experimentation (HTE) is applied for high throughput process development (HTPD) of biologicals [4]. For the evaluation of the performed experiments to stand in relation to experimental speed when performing HTE, the analytical throughput must match the experimental throughput. In some cases this comes easily, for instance if the evaluation of screenings performed in HTE mode can be based merely on univariate spectroscopic measurements such as total protein quantification via UV absorption measurements [5–7]. If selective or specific quantification is necessary, other methods have to be considered. In the case of quantitative separation of protein PEGamers, size exclusion chromatography (SEC) is widely applied. However, for SEC based assays to match the speed of HTE, the analysis time per sample must be reduced to merely a few minutes.

SEC is a standard method for selective and specific quantification of proteins. If exact quantification is the primary objective, most analysts will seek to achieve high resolution ($R > 1.5$) of the components. If a short analysis time is the primary objective, faster flow rates and/or shorter columns can be applied, however, at the expense of resolution. One approach to increase analytical speed without losing resolution is the interlacing of injections and/or parallel operation of two columns on a chromatographic system [8, 9]. When interlaced injections are performed in SEC analysis, the result is elimination of initial lag time between sample injection and start of elution. If interlaced injections are combined with parallel operation of two columns, the waiting time post elution of the smallest sample molecule of interest can be reduced or even eliminated. Thus, in favorable situations, analysis time per sample can be reduced to the time span in which the molecules of interest elute.

The separation of protein PEGamers using SEC poses a very challenging task, as the relative increase in molecular size decreases for each additional PEG molecule attached to the protein [10]. A size based chromatographic separation of native protein and mono-PEGylated species might be achieved effortlessly, however a resolution of the higher protein PEGamers will become increasingly difficult.

Our proposition is that for analytical purposes the degree of separation necessary for correct quantification can be decreased significantly by applying alternative methods for the evaluation of chromatographic data. This idea is based on the assumption that overlapping elution of different species will result in a chromatogram which is a linear superposition of the signals of each single analyte. Therefore, a linear multivariate correlation between elution profile (chromatogram) and amounts of the different components in the respective samples can be expected. This, of course, presumes that the amount of the analytes does not influence the elution profile.

In the linear range of adsorption, this assumption would apply as peak shape is a function of the chromatography system (dead volumn, column packing ect.) and protein characteristics (adsorption and diffusion) but not of protein concentration. Further, the separation mechanism of SEC does not depend on direct interaction with the column material and thus it is not expected that different load concentrations should cause a non-linear change of the elution profile. If the linear correlation between analyte concentration and elution profile is given, it should be possible to calibrate a multivariate regression model based on one chromatogram of each pure component and one defined mixture of all components. Such a multivariate regression model could then give precise determinations of sample composition despite low chromatographic resolution of the sample components.

The presented work describes how multivariate calibration can be used to gain quantitative results of high quality from low resolution chromatograms. By doing so, faster analysis times are achievable through application of short columns and high flow rates that would otherwise be avoided due to inferior separation performance.

In general, a balanced ratio of analytical and experimental time allows a more efficient use of the HTE platform. The goal was to achieve an assay time of maximum two minutes. This would facilitate a 24 hour experiment-analysis cycle for PEGylation screenings in 96 well HTE format, where screening experiments which are performed during day time can be analyzed over night. Further, fast assays can be useful for PEGylation reaction monitoring and subsequent purification process monitoring.

2 Materials & Methods

2.1 Chemicals

Potassium phosphate, potassium chloride and analytical grade ethanol for SEC buffer preparation were purchased from Merck (Darmstadt, Germany). Lysozyme was purchased from Sigma-Aldrich (St. Louis, MO, USA). Methoxy-PEG aldehyde with an average molecular weight of 5 kDa was provided from NOF Cooperation (Tokyo, Japan). Sodium phosphate, sodium chloride and sodium cyanoborohydride ($NaCNBH_3$) for PEGylation buffer and IEC buffer preparation were purchased from Merck (Darmstadt, Germany).

2.2 PEGylated lysozyme

2.2.1 PEGylation Reaction

Lysozyme (5 g/L) and PEG were dissolved in a 25 mM sodium phosphate buffer pH 7.2, containing 20 mM sodium cyanoborohydride. The molar polymer to protein ratio was set to 6:1. The reaction was carried out in a continuously shaken falcon tube at room temperature, for 10 h.

2.2.2 Preparative separation of lysozyme PEGamers

Single lysozyme PEGamers were purified using cation exchange chromatography. Toyopearl GigaCap S-650M resin (Tosoh Biosience GmbH, Stuttgart, Germany) was packed in an Omnifit glass column (25 mm x 400 mm, Diba Industries Ltd., Cambrigde, UK) according to the manufacturer's protocol. The resulting bed volume was 13.4 mL. A gradient elution was performed at a flow rate of 0.7 mL/min with 25 mM sodium phosphate buffer pH 7.2 as mobile phase.

After column equilibration, 50 mL sample mixture was loaded onto the column. Elution was performed with a gradient from 0 to 200 mM sodium chloride over 21 column volumes. The fractionation volume was set to 5 mL.

2.2.3 Molecular weight determination

Peak fractions resulting from the CEX separation were analyzed with respect to lysozyme PEGamer sizes using combined size exclusion chromatography (SEC) and light scattering (LS). This analysis was conducted using an ÄKTA Ettan system from GE Healthcare (Uppsala, Sweden) in combination with a Dawn Heleos 8+ multi-angle light scattering detector and an Optilap rEX refractive index (RI) detector, both from Wyatt Technology (Santa Barbara, USA). For SEC a Superdex 200 GL10/300 (GE Healthcare, Uppsala, Sweden) with a mobile phase of 25 mM sodium phosphate, pH 7.0, containing 150 mM NaCl was used. The flow rate was set to 0.8 mL/min. Injection volumes between 50 and 100 μL were chosen. After UV absorbance monitoring at 280 nm, each sample was measured by LS and RI detection. Molecular weight (M_w) and hydrodynamic radius (R_h) calculation were calculated using the ASTRA software (v.5.3.4.18).

2.3 Chromatography system setup

An UltiMate3000 RSLC x2 Dual system from Dionex (Sunnyvale, CA, USA) was used for UHPLC analysis. The system was composed of two HPG-3400RS pumps, a WPS-3000TFC-analytical autosampler and a DAD3000RS detector. The autosampler was equipped with a 5 μL sample loop. The volume of the injection needle was 15 μL and the syringe size was 250 μL. For control of the UHPLC equipment and for data evaluation the Chromeleon software (6.80 SR10) was used. The software was extended by an additional time base. This enables a virtual separation of the LC system in two parts which can then be controlled separately. Such a setup is necessary in order to facilitate separate data recording of each analyzed sample when performing SEC in interlaced mode. A thorough description of this setup and the performance of interlaced chromatography has been described by Farnen et al. [9] and Diederich et al. [8].

2.4 Size exclusion chromatography

SEC columns (Zenix SEC-300) were purchased from Sepax Technologies (Newark, DE, USA). The Zenix SEC-300 phase is a silica based material with a hydrophilic coating. The 3 μm sized particles have a nominal pore size of 300 Å. Columns of the dimensions 4.6x150 mm and 4.6x300 mm were used. The short column was operated with a flow rate of 0.6 mL/min and the long column with a flow rate of 0.4 mL/min. The columns were mounted with 0.2 μm Opti-Solv®EXP™inlet filters (Optimize Technologies, Oregon City, OR, USA). For analysis, 5 μl sample was injected via full loop injections and a 250 mM potassium phosphate buffer at pH 6.8 with 200 mM potassium chloride was used as running buffer. To prevent fouling of the columns due to PEG binding, 10 % (V/V) analytical grade ethanol was added to the buffer. Interlaced injection mode was performed in order to eliminate lag time between injection and elution of the first sample components.

2.5 Multivariate calibration

Where nothing else is stated, the multivariate regression models were calibrated with six chromatograms: single component chromatograms of each purified lysozyme PEGamer (mono-, di-,

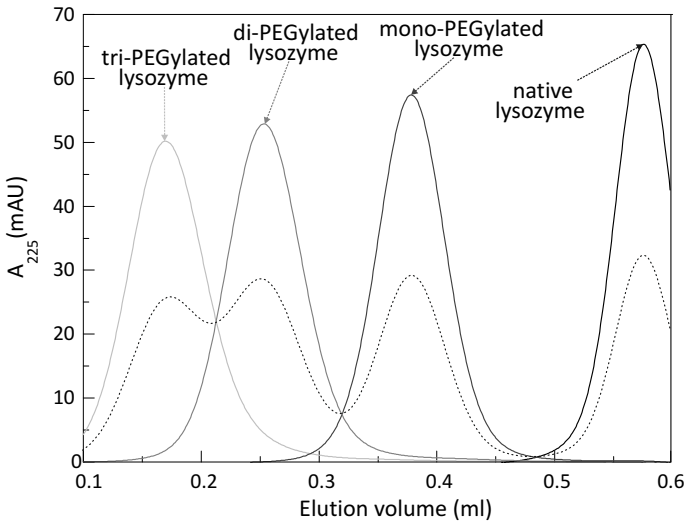


Figure 1: Chromatograms used for calibration of the multivariate regression models. One chromatoram of each pure component with a concentration of 1 g/L with respect to lysozyme and one chromatogram of a mixture of native lysozyme and the three lysozyme PEGamers, all with a concentration of 0.5 g/L with respect to lysozyme.

and triPEG) and native lysozyme, one chromatogram of all components in mass equivalent ratio with respect to lysozyme, and one chromatogram of a blank injection. Chromatograms of the pure components and the mixture of all four components are shown in Fig. 1A. The samples of pure components all had a concentration of 1 g/L. In the mixed sample each component had a concentration of 0.5 g/L. All mentioned concentration are with respect to lysozyme.

Different multivariate regression types were applied: multilinear regression (MLR), partial least squares (PLS) regression, and multivariate curve resolution (MCR). All data processing for multivariate calibration was performed with MATLAB. For multivariate regression with MLR and PLS, the chromatographic data was preprocessed by mean centering. All PLS based regressions were based on 4 latent variables. MCR was performed with the MCR-ALS algorithm. After the MCR model was calibrated with the above mentioned six chromatograms, it was able to deconvolute overlapping chromatograms unknown to the model. This resulted in values corresponding to the integrated area of each pure component present in each of the validation samples. From the initial MCR model calibration, the relation between area and concentration was known and based hereon, the concentration of each component in each validation sample was calculated. Hence, the MCR procedure is a combination of a multivariate approach for chromatogram deconvolution combined with an univariate calibration of the determined area and related concentration.

For multivariate calibrations based on integrated peak areas, calibration samples were generated according to a three layer onion design generated with MODDE (Umetrics, Sweden) which consisted of 32 combinations of the four components.

2.5.1 Chromatographic data

The multivariate calibrations were based either on full chromatograms or four defined points within the chromatogram. The four points were either defined as the UV signal at the elution volume corresponding to the peak maximum of each pure component or as the UV signal at actual peak maxima of the recorded chromatogram (see Fig. 2, left). For the purpose of comparison, calibrations based on integrated peak areas were also performed. Here, vertical peak limits as well as exponential rider skimming were applied to determine peak areas (see Fig. 2, right). Data from full chromatograms consisted of the UV the signal recorded with a frequency of 10 Hz for the short column and 5 Hz for the long column in a defined elution

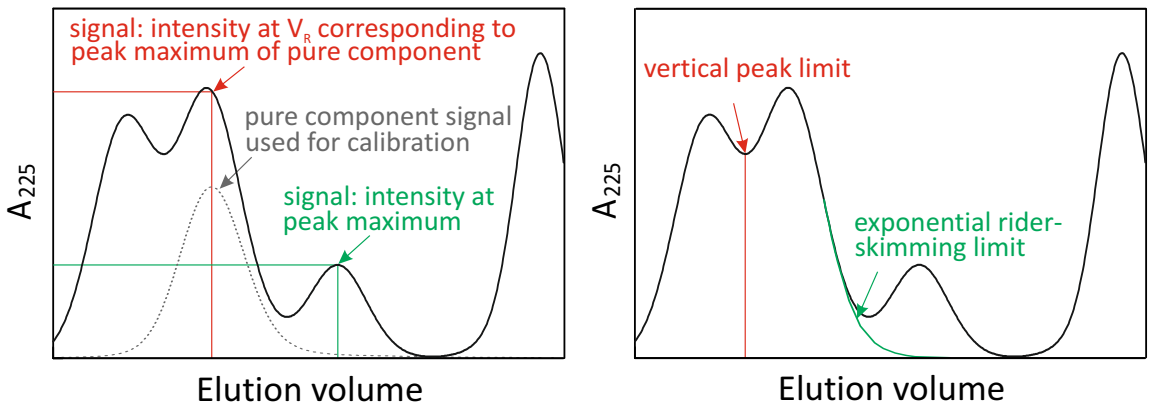


Figure 2: Schematic illustration of the different chromatographic data used for calibration. Left figure: definition of peak maxima for calibration based on single chromatogram data points. Right figure: definition of peak limits for calibration based on area integration.

range. This range was defined as 1.4 – 1.9 mL for the short column and 2.8 – 3.75 mL for the long column with respect to chromatograms based on conventional injection mode as shown in Fig. 3. A full chromatogram consisted of 712 data points when generated with the long column and 500 data points when generated with the short column.

2.5.2 Determination of calibration precision

The calibrated concentration range for native lysozyme and each PEGamer was 0 – 1 g/L with respect to lysozyme. To determine the precision of this suggested approach for evaluation of chromatograms featuring low resolution, validation samples of defined concentrations were analyzed. The validation samples were prepared according to a three layer onion design (designed by MODDE) which resulted in 32 samples with seven different concentration levels (0, 0.25, 0.33, 0.5, 0.67, 0.75, and 1.0 g/L). From each sample three chromatographic runs were made and before each injection of a new sample, a blank run was performed. This procedure was performed using both the long and the short column. Based on the obtained chromatographic data, the calibrated multivariate models were used to predict the content of each lysozyme PEGamer and native lysozyme in the validation samples. For each model and each component the 95 % confidence intervals were calculated based on the total of 96 analyzed samples (threefold injections of 32 validation samples). First the MATLAB 'poly1' fit function was used to fit a linear function to the relation between the concentration predicted by the multivariate model and the nominal concentraion in the validation samples. After the fitting procedure, MATLAB was programmed to return the upper and lower confidence bounds for each linear fit. Finally, the confidence interval for each calibration and component was calculated by adding the upper and lower confidence bound. It should be noticed that the confidence interval does not to give excact information on the precision for different concentration levels within the calibration. Hence, the confidence intervals were only used for comparison of the different calibrations. To obtain more detailed information on presicion, the relative standard devaition (RSD%) was calculated for each concentration interval and presented for a selection of the most precise calibrations.

2.5.3 Dynamic and static calibration

Two modes of multivariate calibration were used. The simplest was a static mode where one set of calibration sample chromatograms recorded immediately before analysis of the validation samples was used for calibration. However, to be able to account for systematic changes in column performance over time, a dynamic approach was developed. In order to do so, calibration sample chromatograms were recorded both before and after the analysis of the validation samples. The regression model was then recalibrated for each validation sample based on a linear interpolation between the chromatograms recorded before and after analysis of the validation samples. The interpolations were calculated using Eq. 1:

$$v_{i,j} = \frac{n-j}{n} \cdot v_{i,pre} + \frac{j}{n} \cdot v_{i,post} \tag{1}$$

where v_i is a vector containing the recorded chromatogram of calibration sample i, n is the total number of measured samples, and j is the number of the sample in the sample sequence for which the model is recalibrated. The indices *pre* and *post* indicates whether a calibration sample was measured before or after the validation samples.

3 Results & Discussion

The aim was to establish a fast SEC assay ($\sim$ two minutes) for the quantification of lysozyme PEGamers and native lysozyme. Hence, a type of SEC columns which allows the use of rather high flow rates up to 3 cm/min was applied. In order to eliminate the lag time inherent in SEC analysis, injections were performed in interlaced mode. Finally, the narrow diameter of the chosen column type allowed for a sample size of 5 μL which minimizes the time necessary for sample injection preparation by the autosampler. All these measures resulted in an assay time of five and two minutes for the long and the short column, respectively. These measures, however, also caused insufficient resolution for precise quantification of all the lysozyme PEGamers when based on determined peak areas. Therefore, multivariate calibration was applied to achieve accurate and precise quantification despite the low resolution.

3.1 Assay time and chromatographic resolution

The resulting separation of lysozyme PEGamers using two different column lengths (150 and 300 mm) is shown in Fig. 3. The presented chromatograms result from samples injected in traditional sequential mode and, as can be seen, less than 50 % of the recorded signal contains relevant information. Hence, interlaced injection mode was applied to reduce the analysis time significantly without decreasing the resolution further. The final assay for the long column was performed with an elution volume of 2.0 mL per injected sample at a flowrate of 0.4 mL/min. This resulted in an assay time of five minutes. The final assay for the short column was performed with an elution volume of 1.2 mL per injected sample at a flowrate of 0.6 mL/min. This resulted in an assay time of two minutes.

column length	Chromatographic resolution of lysozyme PEGamers: *native/mono-PEG*	*mono-/di-PEG*	*di-/tri-PEG*
300 mm	2.72	1.47	0.79
150 mm	1.66	0.98	0.35

Table I: Chromatographic resolution listed for each column length. Each resolution was determined by injecting equal amounts of the two components in question.

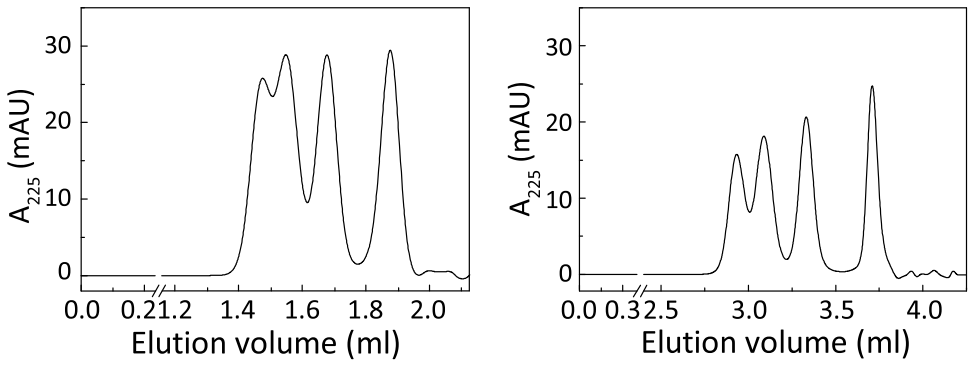


Figure 3: Separation of native lysozyme, mono-, di- and, tri-PEGylated lysozyme with two columns of different lengths. The samples injected on the two columns were identical and contained equal amounts of each component with respect to lysozyme. Left: separation performed on a 150 mm column. Right: separation performed on a 300 mm column.

The determined chromatographic resolutions generated by both columns of lysozyme PEGamers and native lysozyme are listed in table I. These results are based on injection of equal amounts of the two components in question. A resolution of $R \gtrsim 1.5$ is in general sufficient for correct quantification based on integration of peak areas. The short column only generated sufficient resolution for native and mono-PEGylated lysozyme whereas the long column generated sufficient resolution for all components except the tri-PEGylated lysozyme. These results accentuate the challenging task of separating higher PEGylation forms using SEC to a degree sufficient for quantification based on area integration. If sufficient resolution is to be achieved for the higher PEGylation forms, a great amount of dispensable resolution will be generated for native and lower PEGylation forms. Therefore, in order to save time, multivariate calibration was applied in the attempt to achieve precise and accurate quantification despite low resolution.

3.2 Calibration precision

Multivariate models were calibrated based on multilinear regression (MLR), partial least squares (PLS) regression, and multivariate curve resolution (MCR). Further, traditional quantification based on univariate calibration of determined peak areas was performed in order to compare multivariate calibration with the traditional approach. Finally, MLR calibration based on peak areas and peak heights was applied to investigate whether low peak resolution could be compensated by multivariate calibration of these parameters.

Confidence intervals were used to compare the precision of the different calibration approaches, both with regard to regression type and the data used for calibration (i.e. full chromatograms, peak maxima, peak areas). For this purpose, the mean confidence interval of the two components with the lowest chromatographic resolution was determined, in this case the di- and tri-PEGamers. The results are listed in table II and are presented and discussed in the following sections.

3.2.1 Static calibration

In general, a clear increase in precision was achieved by applying static multivariate calibration of chromatographic data. For the long column, the precision increased from 0.0476 g/L to 0.0155 g/L by applying MLR to full chromatogram data instead of univariate calibration of

regression type	model mode	data range	confidence interval
300 mm column			
MLR	static	full chromatogram	0.0155
MLR	static	peak maxima [1]	0.0164
MCR	static	full chromatogram	0.0167
PLS	static	full chromatogram	0.0186
MLR	static	peak maxima [2]	0.0206
MLR	static	peak area [3]	0.0394
Univ.	static	peak area [3]	0.0476
Univ.	static	peak area [4]	0.0768
150 mm column			
MLR	static	peak maxima [1]	0.0226
MCR	static	full chromatogram	0.0226
MLR	static	full chromatogram	0.0247
PLS	static	full chromatogram	0.0347
MLR	static	peak area [3]	0.0453
MLR	static	peak maxima [2]	0.0491
Univ.	static	peak area [3]	0.0580
MLR	dynamic	peak maxima [1]	0.0164
MLR	dynamic	full chromatogram	0.0167
MCR	dynamic	full chromatogram	0.0173
PLS	dynamic	full chromatogram	0.0191

[1)] signal height to base line at retention volumes corresponding to peak maxima of pure components
[2)] signal height to baseline at peak maximum
[3)] vertical peak limits
[4)] exponential rider skimming

Table II: Calibration results presented in order of precision. Different regression types were used to calibrate multivariate models based on different chromatogram data for evaluation of analytical chromatograms. The resulting confidence intervals are listed for each of the different regression types along with model mode and the applied chromatogram data. Further, results based on traditional area integration and univariate calibration are listed. See Fig. 2 for a schematic illustration of the different applied chromatogram data.

peak areas. Based on equivalent calibration for the short column, an increase in precision from 0.0580 g/L to 0.0247 g/L was achieved. For the stronger overlapping peaks generated by the short column, MLR calibration based on only four points in the chromatogram (obtained at V_R of the pure components) resulted in a more precise calibration with a confidence interval of 0.0226 g/L. MLR gave more precise results than PLS and MCR both for the long and the short column. Further, MLR calibration based on peak areas was more precise than univariate calibration of peak areas, however much less precise than calibration based on full chromatogram data. MLR calibration based on peak heights at peak maxima was more precise than MLR calibrations based on peak areas for the long column, however the opposite was the result for the short column. The elution profiles of the pure components (see Fig. 4) shows that severe peak overlapping was limited to approximately 50 % for the long column. Therefore the position of peak maxima was less influenced by the presence of other components when comparing the long column with the short column.

3.2.2 Dynamic calibration

The precision of the univariate peak area based calibration decreased by ~ 25 % when comparing the results obtained by the long column to those obtained by the short column. In comparison, the precision of the multivariate calibration based on MLR and full chromatograms

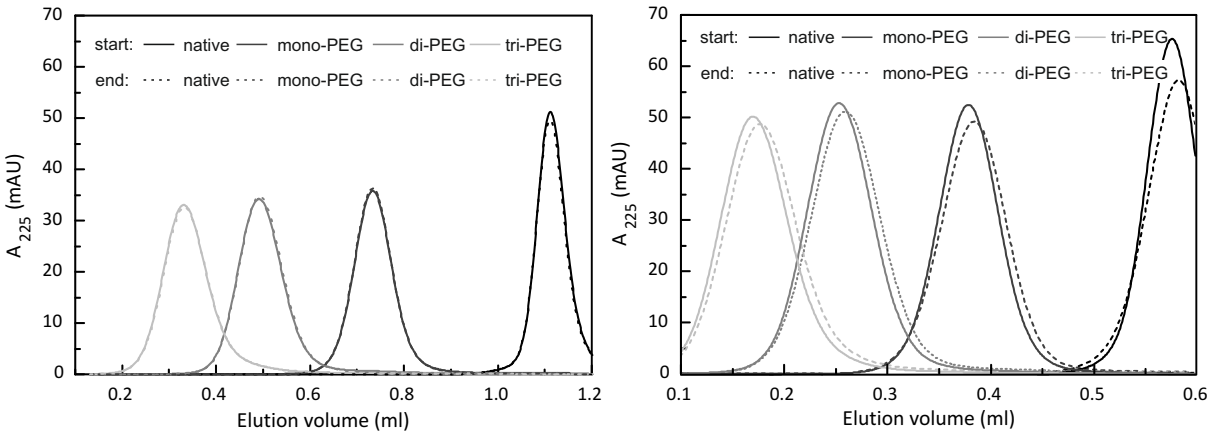


Figure 4: Overlay of single component chromatograms before and after 140 sample injections. Left: 300 mm column. Right: 150 mm column.

decreased by $\sim 70\,\%$. Even though the absolute value of the determined confidence interval was still several times lower for the multivariate calibration compared to univariate calibration (0.0247 *vs.* 0.0580 g/L) when using the short column, the relatively high decrease in precision of the multivariate calibration when switching from the longer to the shorter column (*e.g.* 0.0155 *vs.* 0.0247 g/L) was not expected. In Fig. 4 - left an overlay of pure component chromatograms recorded before and after 140 sample injections on the long column is shown. The data shows stable column performance without a systematic change in peak height or retention volume over time. Fig. 4 - right displays the equivalent data for the short column and here a shift in retention time for all components along with peak broadening can be observed. To overcome the inaccuracy caused by the change in column performance, a dynamic calibration approach was used. This calibration was based on linear interpolation of calibration chromatograms recorded immediately before and after the analysis of the validation samples. It was chosen to apply linear interpolation as the observed change in retention time was linear. The use of dynamic calibration for the short column resulted in an assay of similar precision compared to the static calibration of the long column (0.0167 *vs.* 0.0155 g/L – MLR, full chromatogram).

In Fig. 5 the chromatogram residuals of the validation sample chromatograms are displayed. The shown residuals are derived from both the dynamic and the static PLS calibration for the short column based on the full chromatograms. The variation of the chromatogram residuals remains constant over the period of analysis for the dynamic calibration, where as the variation increases throughout the period of analysis for the static calibration. This supports the decision to use linear interpolation for the dynamic calibration.

In order to visualize the effect of dynamic calibration, three artificial chromatograms were created and displayed along with a true chromatogram of a validation sample (Fig. 6). The artificial chromatograms were created by a linear combination of pure sample chromatograms in a ratio corresponding to the true chromatogram. Two of the artificial chromatograms were based on the calibration chromatograms recorded either before or after the analysis of the validation samples. Neither of these chromatograms were of good resemblance to the validation sample chromatogram. The third artificial chromatogram was based on the linear interpolated chromatograms of the calibration samples which were created for the dynamic calibrations. This chromatogram was built to match the point in time where the displayed true chromatogram was recorded and exhibited a high resemblance to the true validation sample chromatogram. For

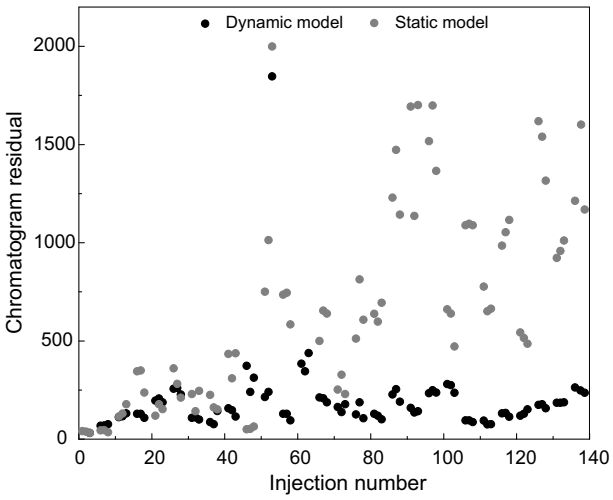


Figure 5: Residuals of the 96 validation sample chromatograms as a function of injection order. Each grey data point represents one sample chromatogram evaluated the static PLS calibration. Each black data point represents one sample chromatogram evaulated with the dynamic PLS calibration. The residuals for the 32 blank samples recorded between the validation samples are not displayed.

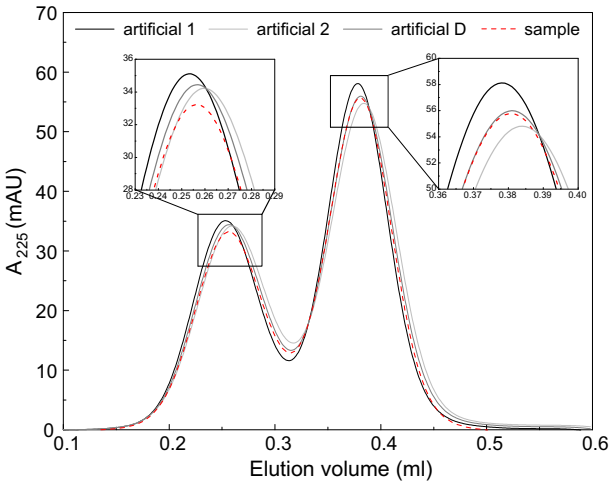


Figure 6: Visual comparison of the effect of dynamic calibration. One true chromatogram and three artificial chromatogram based on linear combinations of pure component chromatograms are displayed. Two of the artificial chromatograms were built from the pure component chromatograms recorded either before (*artificial 1*) or after (*artificial 2*) analysis of the validation samples. The third artificial chromatogram (*artificial D*) was built from the linear interpolations used in the dynamic calibrations. The linear interpolation was built to correspond to the point in time where the true samples was analyzed.

the right peak (monoPEG-protein conjugate), the resemblance was high both for peak intensity and peak shift. The fact that the left peak of the sample chromatogram deviated from the dynamic calibration chromatogram in peak intensity is most likely due to a pipetting error during preparation of the validation samples.

3.2.3 Concentration related precision and sensitivity

The concentration related precision was determined for the dynamic and static MLR calibrations based on the short column and the MLR calibration based on the long column. The relative standard deviation (RSD%) for each contration is displayed in Fig. 7. The displayed RSD% values are mean values of the RSD% for di- and tri-PEGylated lysozyme. For all three calibrations, the precision increased from the first to the second concentration level (0.25 g/L – 0.33 g/L). For all concentration levels, the precision of the static MLR calibration based on the 150 mm column was lowest (RSD $\sim$ 3 – 1.5 %). The precision for all concentration levels was similar for the dynamic MLR calibration based on the 150 mm column and the static MLR calibration based on the 300 mm column. This is in accordance with the finding that the overall precision of these calibrations were similar (95 % confidence interval: 0.0167 g/L and 0.0155 g/L - see table II). This means that both the overall precision and the specific precicion achieved

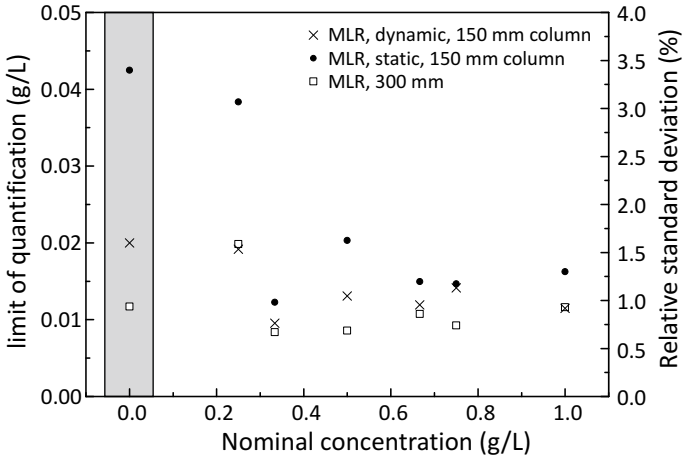


Figure 7: RSD % values for the concentration levels included in the validation sample design. Displayed are mean values of the determined RDS % for di- and tri-PEGylated lysozyme. The value dispayed for the concentration zero (grey bar) are LOQ values. Also the LOQ values are mean values of the determined LOQ for di- and tri-PEGylated lysozyme.

for the long column can be maintained for the short column by applying multivariate calibration. For further comparison and characterisation of the calibrations, the limit of quantification (LOQ) was determined. LOQ was determined as six times the standard deviation of all 'zero' predictions in the calibration samples for each component. Again, the displayed LOQ values are mean values of the determined LOQ value for di- and tri-PEGylated lysozyme. With regard to the LOQ value, the dynamic approach does increase the sensitivity of the method, however not to the level of the MLR calbration based on the 300 mm column.

4 Conclusion and Outlook

The aim of establishing an SEC assay for the quantification of lysozyme PEGamers with an analysis time of two minutes was achieved. The presented work clearly demonstrates that by applying multivariate calibration for the quantitative evaluation of low resolution chromatograms the precision can be enhanced significantly when compared to traditional univariate calibration. Hence, fast chromatographic assays can easily be achieved by applying short columns and fast flow rates. The tested chromatographic assays included resolutions down to 0.35 demonstrating how little resolution is actually sufficient to achieve a highly precise chromatographic assay. The lower limit of resolution necessary for precise calibration is still to be determined. Further, the results demonstrated that a change in column performance over time can be handled without difficulty by using a dynamic calibration approach.

In the presented work, pure samples were used for calibration. If pure material is at hand or easily achievable, this is the most straightforward approach. If pure samples are difficult or impossible to obtain, mixed samples of defined composition can also be used. This then requires an alternative assay or analytics to define the mixed samples and further these samples must contain sufficient variation.

The use of multivariate calibration is of course not limited to SEC. Any robust chromatographic assay established for defined samples can be evaluated by the presented approach. The approach cannot be applied to complex samples in which unknown peaks occur in the area used for calibration and also chromatographic assays based on non-linear chromatography will demand more sophisticated multivariate calibration of non-linear nature. This might increase the calibration complexity to a level which is no longer leveraged by the gained increase in assay speed. However, this is still to be investigated.

In working environments with limited time for assay development and data evaluation, it might be an unmanageable task to perform data evaluation not inherent in the chromatography system software. Hence, we propose the integration of multivariate calibration directly in the commercial software supplied with the chromatography systems.

References

[1] M. J. Knauf, D. P. Bell, P. Hirtzer, Z. P. Luo, J. D. Young, N. V. Katre, Relationship of effective molecular size to systemic clearance in rats of recombinant interleukin-2 chemically modified with water-soluble polymers, J. Biol. Chem. 263 (29) (1988) 15064–70.

[2] F. Fuertges, A. Abuchowski, The clinical efficacy of poly(ethylene glycol)-modified proteins, J. Controlled Release 11 (1990) 139–148.

[3] N. V. Katre, M. J. Knauf, W. J. Laird, Chemical modification of recombinant interleukin 2 by polyethylene glycol increases its potency in the murine Meth A sarcoma model, Proc. Natl. Acad. Sci. U.S.A. 84 (6) (1987) 1487–91.

[4] R. Bhambure, K. Kumar, A. S. Rathore, High-throughput process development for biopharmaceutical drug substances, Trends Biotechnol. 29 (3) (2011) 127–35.

[5] S. Oelmeier, F. Dismer, J. Hubbuch, Application of an Aqueous Two-Phase Systems High-Throughput Screening Method to Evaluate mAb HCP Separation, Biotechnol. Bioeng. 108 (1) (2011) 69–81.

[6] S. Chhatre, R. Francis, D. G. Bracewell, N. J. Titchener-Hooker, An automated packed protein G micro-pipette tip assay for rapid quantification of polyclonal antibodies in ovine serum, J. Chromatogr. B 878 (30) (2010) 3067–75.

[7] M. Bensch, P. S. Wierling, E. von Lieres, J. Hubbuch, High Throughput Screening of Chromatographic Phases for Rapid Process Development, Chem. Eng. Technol. 28 (11) (2005) 1274–1284.

[8] P. Diederich, S. K. Hansen, S. a. Oelmeier, B. Stolzenberger, J. Hubbuch, A sub-two minutes method for monoclonal antibody-aggregate quantification using parallel interlaced size exclusion high performance liquid chromatography, J. Chromatogr. A 1218 (50) (2011) 9010–9018.

[9] D. Farnan, G. T. Moreno, J. Stults, A. Becker, G. Tremintin, M. van Gils, Interlaced size exclusion liquid chromatography of monoclonal antibodies., J. Chromatogr. A 1216 (51) (2009) 8904–9.

[10] C. J. Fee, J. M. Van Alstine, PEG-proteins: Reaction engineering and separation issues, Chem. Eng. Sci. 61 (3) (2006) 924–939.

A Label-Free Methodology for Selective Protein Quantification by Means of Absorption Measurements

Sigrid K. Hansen[1], Erik Skibsted[2], Arne Staby[2] and Jürgen Hubbuch[1]*

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

[2] *CMC Project Planning & Management, Novo Nordisk A/S, Denmark*

Manuscript published in ***Biotechnology and Bioengineering***

Abstract

The application of high throughput experimentation (HTE) in protein purification process development has created an analytical bottleneck. Using a new label-free and non-invasive methodology for analyzing multicomponent protein mixtures by means of spectral measurements, we show that the analytical throughput for selective protein quantification can be increased significantly. An analytical assay based on this new methodology was shown to generate very precise results. Further, the assay was successfully applied as analytics for a resin screening performed in HTE mode. The increase in analytical throughput was obtained without decreasing the level of information when compared to analytical chromatography. This proves its potential as a valuable analytical tool in conjugation with high throughput process development (HTPD). Further, fast selective protein quantification can enhance process control in a commercial production environment and, hence, minimize the need for off-line release analysis.

Keywords: protein analytics, specific protein quantification, label-free, non-invasive, partial least squares, high throughput process development

1 Introduction

Over the past few years, the use of high throughput techniques for development of protein purification processes has increased strongly. The main drivers of this shift towards high throughput process development (HTPD) have been (and still are) the need to speed up both time to clinic and time to market. Automated performance of unit operations in a miniaturized and parallel mode has delivered the high experimental throughput which is the basis of HTPD. The automation has been realized using robotic workstations with liquid handling systems where sampling is commonly performed in microtiter plates.

Alongside the development of high throughput experimentation (HTE) techniques, a trend of replacing the traditional 'one factor at the time' screenings with more sophisticated approaches such as factorial designs, intelligent search algorithms, and mechanistic modeling has evolved [1–3]. These approaches make process development more efficient and can result in a deeper understanding of the developed processes. However, regardless of how sophisticated a search algorithm or design is used to improve the experiment to information ratio, a high number of samples have to be analyzed when performing state of the art HTPD. In order not to compromise the throughput and level of automation, univariate spectroscopic measurements performed in plate readers is the most comfortable analytical solution. If the nature of the experiments is more complex and requires selective or specific protein quantification, analytical chromatography, mass spectrometry, electrophoresis and/or enzyme linked immunosorption assay can be applied. However, all methods have drawbacks related to throughput, sample preparation, and/or automated integration. Hence, the enhanced experimental throughput often creates an analytical bottleneck where the experimenter is forced to make decisions on the basis of a trade off between analysis time and analytical information.

Liquid chromatography is the main workhorse of modern protein purification. Accordingly, liquid chromatography, in batch as well as column mode, is among the unit operations readily established and applied in HTPD [4–6]. Batch experiments can give information on selectivity, uptake kinetics and static binding capacities in studies of different phase combinations. From column experiments, information on dynamic binding capacities and chromatographic resolution can be derived. Both batch and column chromatography in HTE mode can also be used to determine parameters for mechanistic modeling of chromatographic processes [7].

Column chromatography in HTE mode generates hundreds of samples per hour and hence represents a major analytical challenge if the high experimental throughput is to be maintained. The objective of the work presented here, was to develop and establish an assay which would limit the analytical trade off without compromising the throughput of the experimental performance. Such an assay would in general be useful to unit operations relevant to protein purification where selective protein quantification is necessary: *e.g.* membrane chromatography, batch chromatography, selective protein precipitation, protein crystallization and aqueous two-phase systems.

The standard univariate spectroscopic assay used to quantify proteins in aqueous solutions is based on the application of the Lambert-Beer law to a measured UV absorption caused by solute protein molecules. In most cases, the absorption ascribable to the aromatic structures in the residues tryptophan, tyrosine, and phenylalanine is used. Of these amino acids, tryptophan and tyrosine have far the strongest absorption. Fig. 1 displays the absorption spectra in the range from 220 nm to 300 nm of pure tryptophan, tyrosine, and phenylalanine in aqueous solutions. Their spectra are characterized by two maxima in the mid-UV range: the most intense around 230 nm and one of lower intensity and more wide stretching around 280 nm.

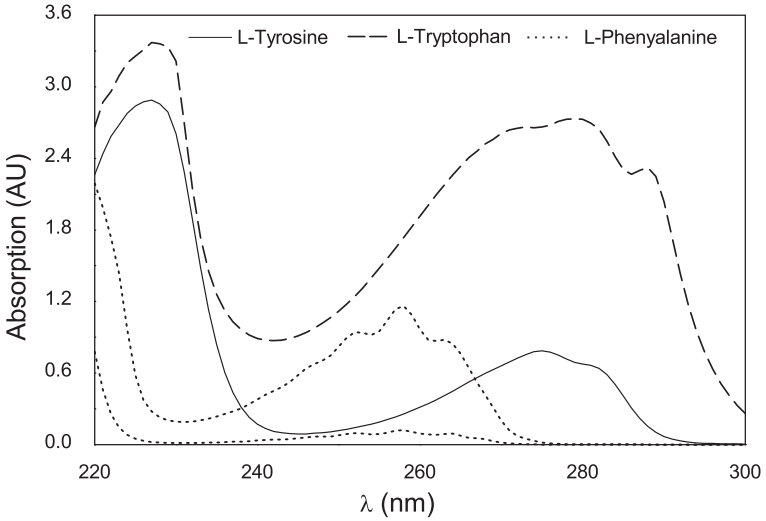


Figure 1: Mid-UV absorption spectra of tyrosine, tryptophan, and phenylalanine. The concentration of tyrosine and tryptophan was 0.1 g/L. Due to the relative low absorptivity of phenylalanine, absorption spectra of both the concentrations 1.0 g/L and 0.1 g/L are displayed.

The absorptivity of phenylalanine is approximately one magnitude lower and both maxima are shifted 20 nm towards lower wavelengths. Due to material and apparatus issues, it is often preferred to measure the absorption at 280 nm. However, as the extinction coefficient (ϵ) differs depending on the amount and kind of aromatic residues present in a protein, the method is only applicable to pure protein solutions. For absolute quantification of proteins with unknown extinction coefficients or protein mixtures, methods combining the absorption at 280 with the absorption at 205 nm where also the peptide bonds absorb can be found in the literature [8, 9].

Mid-UV spectral data have been applied for identification and selective quantification of many different compounds other than proteins: e.g. [10–12]. However, until now mid-UV absorption spectra of proteins have mainly been applied to monitor structural and binding aspects of proteins: e.g. [13–15]. One exception is the application of UV absorption spectra to determine protein and nucleic acid content of viruses [16]. As can be seen in Fig. 1, the absorption spectra of the aromatic amino acids differs significantly, not only in intensity but also in shape. This fact will cause the absorption spectra of different proteins to change depending on the ratios of the aromatic residues present in the proteins. Awareness of this fact combined with the spreading application of chemometric approaches (e.g. multivariate calibration) in biomedical analysis [17] has been the inspiration to a new methodology for analyzing multi-component protein mixtures in a label-free and non-invasive mode. According to the definition by Lavine [18], chemometrics is an approach to analytical and measurement science that uses mathematical, statistical, and other methods of logical science to determine properties of substances that would otherwise be difficult to determine directly. In this case the substance property is the selective protein concentration in a mixture of several proteins and the applied 'logical science' is *partial least squares - projection to latent structures* (PLS) regression. This was chosen as PLS is known to be particular good in coping with expected collinearity of the spectral data [19].

2 Materials & Methods

2.1 Disposables

Sample preparation: 96 well square deep well plates (DWP), volumetric capacity 2 ml, polypropylene (VWR, Germany). **Mid-UV spectra measurement**: 96 well half area UV microtiter plates (MTP), volumetric capacity 180 μL, polystyrene, flat bottom and 384 square well MTP, volumetric capacity 110 μL, polystyrene, flat bottom (Greiner, Germany). **Chromatography**: RoboColumns purchased prepacked with 200 μL of the following resins: CM Ceramic HyperD F, Toyopearl SP 650 M, SP Sepharose FF, and Fractogel EMD SO_3 (Atoll, Germany).

2.2 Model Proteins

The following model proteins were used: lysozyme (lys) from hen egg white, cytochrome C (cytC) from horse heart and ribonuclease A (ribA) from bovine pancreas (Sigma-Aldrich, Germany). These proteins compose a convenient model system for method development of protein purification due to the distribution of their isoelectric points. At neutral pH, separation can be achieved using cation exchange chromatography.

2.3 Liquid Handling Station

The generation of protein samples for model calibration, model validation, and elution buffers as well as the chromatographic experiments were performed automated on a Tecan Freedom Evo 200 (Tecan, Germany) liquid handling station (LHS). To enable performance of column chromatography on the LHS, a custom made column carrier was used. A similar carrier is available under the name Te-Chrom (Tecan, Germany). A thorough description of parallelized chromatography performed automated on a LHS has been published by [6] and [7]. Further, the LHS was equipped with an Infinite 200 plate reader (Tecan, Germany) to measure the mid-UV spectra of the calibration and validation samples as well as the chromatographic fractions. The liquid handling station was operated with EVOware 2.0. Import of values for pipetting volumes and export of spectral data was handled via Excel (Microsoft, Redmond, USA).

2.4 Experimental Setup

2.4.1 Model Calibration and Model Validation

The calibration and validation samples were based on a four level D-optimal onion design generated with MODDE (Umetrics, Sweden). In total, the four level onion design with three center points consisted of 29 samples. The design covered 19 mixing ratios and nine concentration levels from 0-0.7 g/L. In fig. 2A the mixing ratios of all three proteins and the absolute concentrations of ribA and cytC are displayed. An additional concentration level at 0.05 g/L was added to the four layer onion design, however only as pure samples. In total a sample set consisted of 32 samples. Two mL of each sample were prepared from stock solutions, containing 2.1 g/L protein in a sodium phosphate buffer at pH 7. Preparing 2 mL of each sample, enabled testing of different plate types and sample volumes based on the same sample stock. For the measurement of the absorption spectra, a defined volume of the prepared samples were transferred to UV transparent MTPs. The absorption spectra were measured with 1 nm resolution in the range 240-300 nm. For model calibration based on the obtained spectral data, the MATLAB

based PLS toolbox (EigenVector, USA) was used. The spectral data was preprocessed by mean centering which is done by subtracting the mean spectrum of all calibration samples from each protein spectrum. Method validation was performed according to the flow scheme displayed in fig. 2. The setup was designed to obtain a reliable measure of the possible obtainable precision taking into account pipetting errors, measurement error, and errors caused by plate deviations. The setup consisted of six independently pipetted sample sets (fig. 2C,E). Five sample sets were used to calibrate five models (fig. 2D). All five calibrated models were then used to experimentally determine the concentrations of the sixth sample set (fig. 2E). From these values, absolute errors, relative errors, and 95 % confidence intervals were calculated (fig. 2F-H).

2.4.2 Resin Screening

All columns were equilibrated with 5 CV (1 CV = 200 μL) loading buffer (20 mM sodium phosphate, pH = 7) before loading them with 0.25 mg of each protein per column. This protein load equals 3.75 g/L resin of total protein. Before starting the actual column experiments, 24 different buffer solutions with increasing salt concentrations were prepared and stored in DWPs until use. The prepared buffer solutions composed a linear step gradient from 0-500 mM NaCl and was used to elute the proteins in 24 CV. Each gradient step was collected as one fraction. After completed elution, 100 μL of each elution fraction were transferred into UV grade MTPs and the absorption spectra were measured. Based on these spectra, the protein content in the samples was determined.

2.4.3 Peak Fitting and Calculation of Chromatographic Resolution

To evaluate the chromatographic separations, an asymmetric Gaussian function was fitted to the data points. The fitting was performed using the graphing and data analyzing software OriginPro8.5 (OriginLab Corporation, USA) which has an asymmetric Gaussian fit function under the name *bigauss*. The *bigauss* fitfunction is a two-step function which separately describes the leading and the trailing edges of an asymmetric peak using two different gauss curves. The bigauss fitfunction returns the values w_1 (FWHM/2 - left side fit) , w_2 (FWHM/2 - right side fit), x_c (retention at peak maximum), H (maximal peak height), and y_0 (base line) for each peak. Based on the parameters w_1, w_2, and x_c the resolution of two adjacent peaks p1 and p2 was calculated according to:

$$R_{p1,p2} = \frac{2.345(x_{c,p2} - xc_{c,p1})}{4(w_{2,p1} + w_{1,p2})} \tag{1}$$

The above described equation is a derivative of the standard equation for chromatographic resolution based on peak basis width at 4σ (w_b) and the distance between peak maxima (t_{R2}-t_{R1}):

$$R_s = \frac{2(t_{R2} - t_{R1})}{w_{b1} + w_{b2}} \tag{2}$$

2.4.4 Analytical Chromatography

For comparison, the protein concentrations in the samples generated in the resin screening were quantified using a Mono S 4.6/100 PE column (GE Healthcare, Germany). The column was loaded with 20 μL sample and eluted with a multi linear sodium chloride gradient. Buffer A: 20 mM sodium phosphate, pH 7, buffer B: 20 mM sodium phosphate, 500 mM sodium chloride,

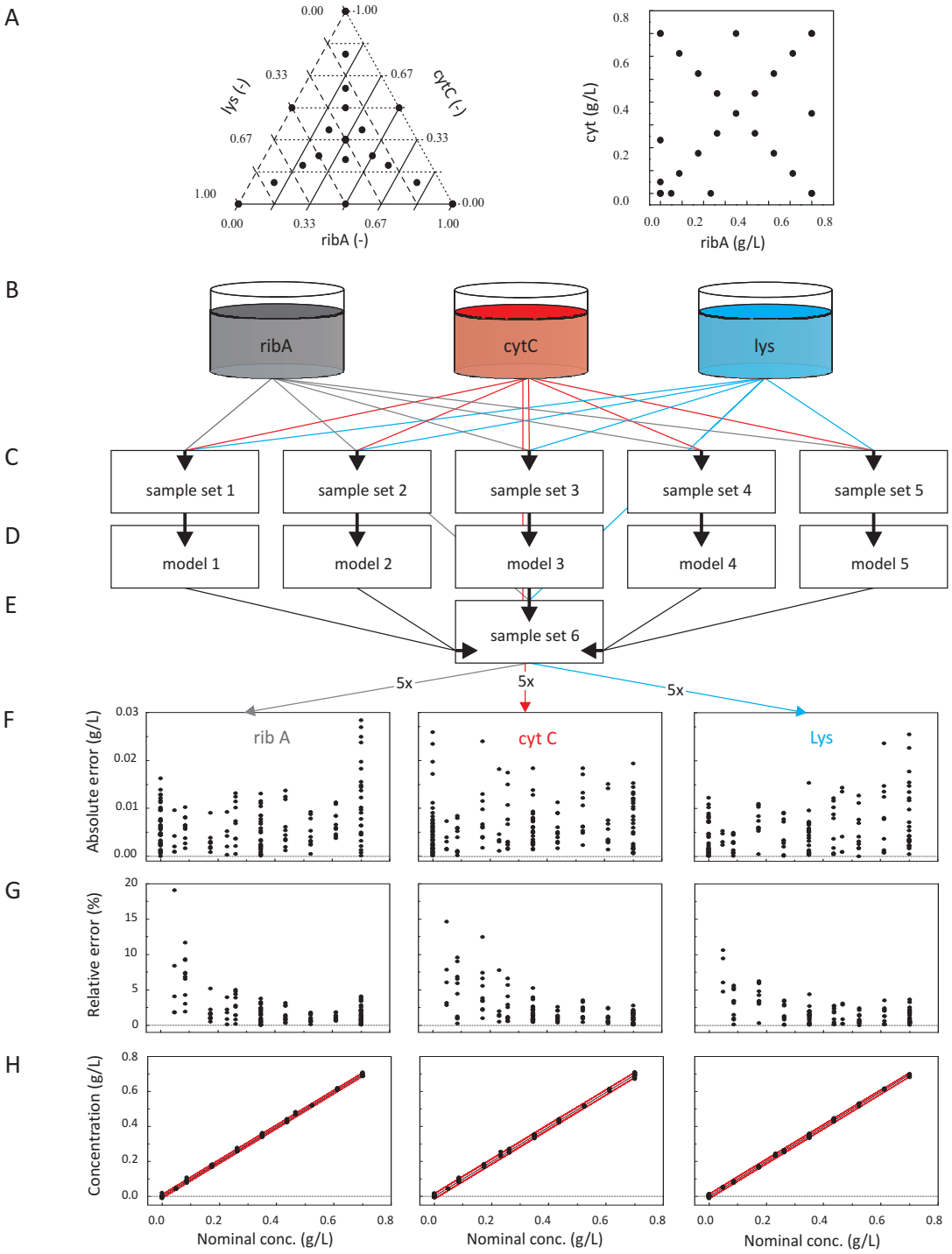


Figure 2: Design of the model calibration and validation samples followed by a flowchart describing the validation procedure. **A:** Mixing ratios of all three proteins and absolute concentrations of cytC and ribA. **B:** Protein stock solutions **C:** Five identical sample sets prepared individually from the protein stock solutions. **D:** Five calibrated models based on the five sample sets. **E:** Sixth sample set of which the concentrations were experimentally determined using the five calibrated models. **F:** Absolute errors of the experimentally determined concentrations displayed as a function of the nominal concentration. **G:** Relative error of each experimentally determined protein concentration as a function of the nominal concentration. **H:** Experimentally determined protein concentration plotted as a function of the nominal concentration. The red lines display 95 % confidence interval for each protein species calculated on basis of the experimentally determined values.

pH 7. Gradient: buffer B 0-20 % from 0-8 min, buffer B 20 % from 8-12 min, buffer B 20-40 % from 12-18 min, buffer B 100 % from 18-20 min, buffer B 0 % from 20-25 min. The flow rate was 1 mL/min.

3 Results & Discussion

In the presented work, we describe a fast and automated assay for selective protein quantification. The label-free and non-invasive methodology on which the assay is based, uses PLS to correlate absorption spectra to selective protein concentrations. The developed assay has been subjected to a thorough validation and the results hereof will be presented and discussed. Further, an application relevant to HTPD is presented and the results are compared to results generated using standard analytical chromatography.

3.1 Protein Mid-UV Absorption Spectra

In general, proteins absorb UV light in the mid-UV range. This absorption is mainly caused by the chromophoric properties of the three aromatic residues; tryptophan, tyrosine, and phenylalanine. Further, some proteins exhibit additional absorption in the visible range due to other chromophoric properties; *e.g.* green fluorescence protein ($\lambda_{max} \sim 488$ nm) and cytochrome C ($\lambda_{max} \sim 526$ nm). Present in a mixture, these proteins can easily be quantified separately by measuring the absorption at wave lengths specific for each protein. In contrast, specific quantification of proteins with absorption characteristics only ascribable to their aromatic residues and peptide bonds, can not be performed by such a straight forward method. However, the ratio and number of aromatic residues will vary among different proteins, and therefore the intensity and shape of the related UV absorption spectrum will vary.

Fig. 3A displays the absorption spectra of the three model proteins ribA, cytC and lys in the range 240-300 nm at a concentration of 0.5 g/L. To depicture the extend to which the spectral shapes vary independently of spectral intensity, difference spectra based on normalized spectra were calculated (Fig. 3B). The normalization was based on the maximal absorption value in the range 270-290 nm which was set to 1.0 AU. Fig. 3C displays the difference spectra build by subtracting the mean of all three normalized spectra from each individual normalized spectrum. Lys and ribA have similar spectral shapes in the range 250-280 nm and deviates in the ranges 240-250 nm and 280-300 nm. The spectrum of cytC is different from both other spectra with the exception of the lys spectrum in the range 280-290 nm.

3.2 Method Validation

A multivariate calibration model links a relation between several input variables to one or more properties of interest. In this case the input variables are the spectral data and the properties of interest are the selective concentrations of the different proteins contained in a sample. To determine the possible precision which can be obtained for the correlation of spectral data and protein concentrations using PLS, a thorough validation was undertaken according to the scheme presented in fig. 2. As a measure for precision, the 95 % confidence interval was used. The confidence interval was based on the five fold experimental determination of the protein concentrations in the samples of *sample set 6* (see fig. 2).

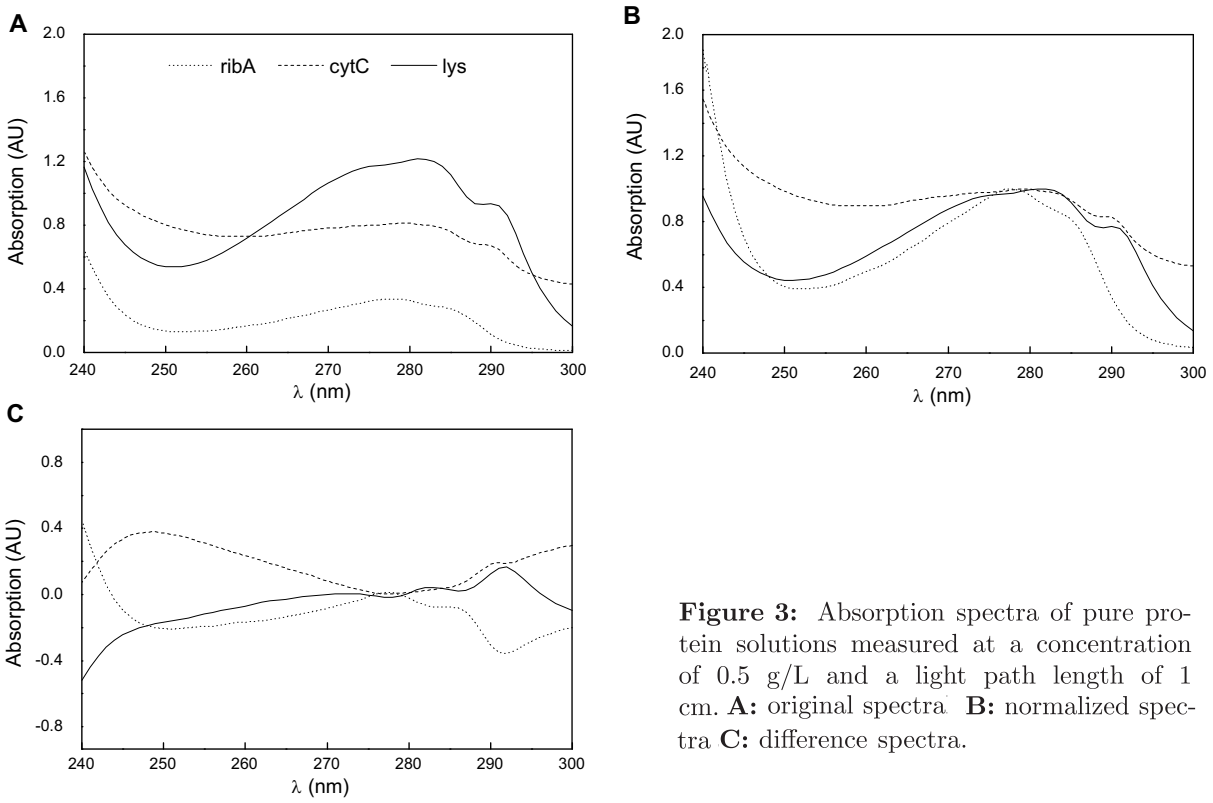


Figure 3: Absorption spectra of pure protein solutions measured at a concentration of 0.5 g/L and a light path length of 1 cm. **A:** original spectra **B:** normalized spectra **C:** difference spectra.

In the following the validation results are presented. This includes the influence of plate type and sample volume, which in MTPs determines the optical path length, method robustness regarding salt concentration and temperature, *in silico* parameters such as choice of preprocessing and number of latent variables used for calibration, and finally PLS was compared to multiple linear regression (MLR).

3.2.1 Number of Latent Variables

PLS calibration is based on a data reduction with the objective to capture only the significant variation needed for calibration. Initially, it was tested how the level of reduction would influence precision of the resulting PLS model. Here, the level of reduction is expressed by the number of latent variables included in the calibration of the PLS model where a low number of latent variables corresponds to a high data reduction. For each tested number of latent variables and for each protein species, the 95 % confidence interval was calculated. Further, the confidence intervals for all three proteins were summarized for each number of latent variables to give a measure for the overall precision of the model. In fig. 4A the confidence intervals are displayed as a function of the number of latent variables in the tested range from 3-11. The curves fitted to the data points serves the visualization. Whereas cytC describes a low but continuous increase over the tested range, both lys and ribA exhibit a minimum at six latent variables. This is consistent with the fact that cytC differs the most from the mean spectrum of all three proteins (see fig. 3). Hence, the most significant variation in the data is correlated to the cytC concentration and this variation will then be described more accurately by the first latent

variables compared to ribA and lys. The sum of the confidence intervals is lowest for six latent variables. However, using six latent variables is a compromise, as the precision of cytC is not optimal at this point. Alternatively, a separately calibrated model for cytC could be used. Due to the very small difference in precision, it was chosen to base further method validation on models calibrated for all proteins based on six latent variables.

3.2.2 Plate Type and Sample Volume

The influence of plate type and sample volume was tested by comparing the resulting confidence intervals. Included were two MTP types: 384 well MTP (square wells) and 96 well MTP (half area wells). In the 384 well MTP the tested volumes were 40, 60, and 80 μL corresponding to path lengths of 3.3, 5.5, and 6.7 mm. In the 96 well MTP the tested volumes were 40, 75, 100, and 150 μL corresponding to path lengths of 2.6, 4.7, 6.2, and 9.0 mm. The influence of sample volume and plate type is displayed in fig. 4B. To simplify the visualization, only mean values and standard deviations of the confidence intervals of all three proteins are displayed. In the range 2.6-5.0 mm, the optical path length correlated with a linear decrease of the confidence interval, independently of plate type. Above 5.0 mm the confidence interval decreased further when using the 96 well MTP, however the variance (expressed by the standard deviation) of the confidence intervals increased for the highest measured path length of 9.0 mm. Hence, it was chosen to perform all other experiments with 100 μL sample in a 96 well MTP where the variance of the single confidence intervals was low (ribA: 0.031 g/L, cytC: 0.032 g/L, lys: 0.028 g/L) and the mean value was not significantly higher compared to what was achieved with 150 μL (0.031 vs. 0.027 g/L).

3.2.3 Protein Concentration Range

The precision over the calibrated concentration range is displayed in fig. 4C (100 μL samples measured in 96 well MPTs and PLS models calibrated with 6 latent variables). The curves fitted to the data points serve to visualize the observed trends. They do not include the lowest concentration of 0.05 g/L, as this concentration level was only included in the form of pure samples and therefore not representative for complex samples. The precision increased with increasing protein concentrations up to 0.525 g/L, which is expressed by the decreasing relative standard deviation. For ribA and cytC the relative standard deviation decreased from 7 % at 0.05 g/L to 1 % at 0.525 g/L. Over the same range, the relative standard deviation for lys decreased from 3 % to 1 %. Above a concentration of 0.525 g/L, the relative standard deviation for all three protein species increased. The descending order of precision within the protein species (lys>ribA>cytC) correlated with the previously determined confidence intervals based on the same data (lys: 0.028 g/L, ribA: 0.031 g/L, cytC: 0.032 g/L). The initial decrease of the relative standard deviation is explained by pipetting errors and measurement noise having the highest impact at low concentrations. The increase of the relative standard deviation above 0.525 g/L is also expectable, since only a very low amount of light passes the sample at high protein concentrations. The lower limits of quantification (determined as 6σ) were 0.04, 0.06, and 0.03 g/L for ribA, cytC, and lys, respectively. This is not immediately consistent with the relative standard deviations at the lowest concentration level presented in fig. 4C. However, as already mentioned, the concentration level at 0.05 g/L was only included in the form of pure protein samples is therefore not representative for complex samples.

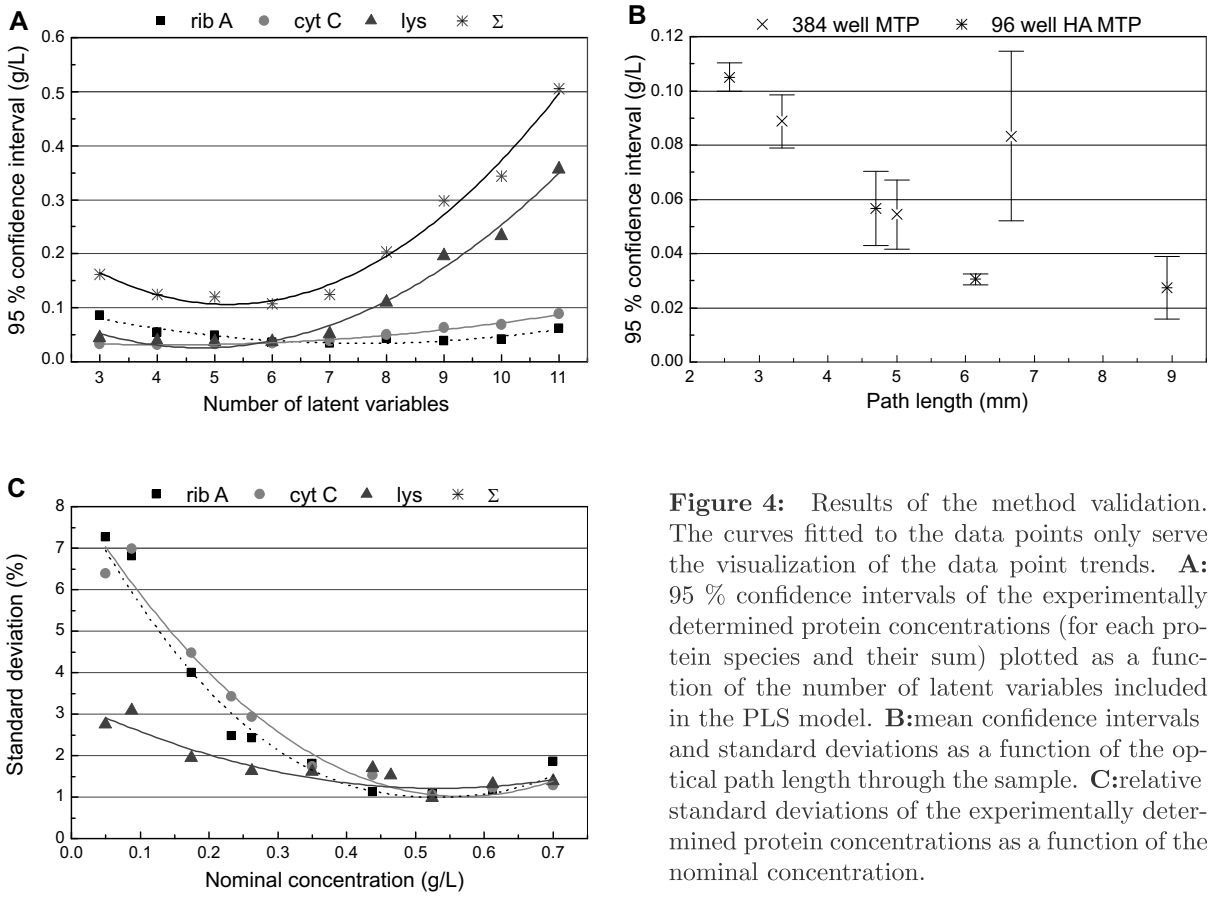


Figure 4: Results of the method validation. The curves fitted to the data points only serve the visualization of the data point trends. **A:** 95 % confidence intervals of the experimentally determined protein concentrations (for each protein species and their sum) plotted as a function of the number of latent variables included in the PLS model. **B:**mean confidence intervals and standard deviations as a function of the optical path length through the sample. **C:**relative standard deviations of the experimentally determined protein concentrations as a function of the nominal concentration.

3.2.4 Calibration and Preprocessing

MLR is a classical method for multivariate calibration. MLR is known to be sensitive to collinearity which is inherent in spectral data, as these do not consist of independent variables. To assert that PLS actually generates more precisely calibrated models, both approaches were tested based on identical data. The measurements were performed in 96 well MTP with 100 μL sample and PLS models were based on six latent variables. Further, it was tested whether mean centering of the spectral data has a positive effect on the calibration. For comparison, the resulting 95 % confidence intervals were used (listed in table 1). The MLR model calibrated with mean centered data was shown to have a precision approximately three times lower compared to a PLS model calibrated with mean centered data. Further, mean centering of the spectral data showed to be essential in obtaining a precise model, as the confidence intervals increased by one order of magnitude when the models were calibrated based on raw data compared with the achieved precision of models calibrated based on mean centered data. Hence, the best result was achieved using mean centered data in combination with PLS.

calibration method	applied preprocessing	confidence interval (g/L) ribA	cytC	lys
PLS	mean centering	0.031	0.032	0.028
PLS	none	0.227	0.349	0.275
MLR	mean centering	0.097	0.089	0.102
MLR	none	0.350	0.590	0.414

Table I: confidence intervals for each protein species based on models calibrated with PLS or MLR. Each model type was based on raw or mean centered data.

3.2.5 Method Robustness

All factors influencing the absorption in the applied range from 240-300 nm will have an effect on precision and accuracy of the method. Hence, it was tested whether the addition of sodium chloride has an effect on precision and accuracy of the calibration. In the tested range from 0-250 mM sodium chloride a significant change in precision or accuracy was not observed. Further, it was tested whether changing temperature of the sample has an influence. The tested range was $23 - 35\,^{\circ}\mathrm{C}$. In a sample containing equal amounts of all three proteins, the determined lys concentration exhibited a decreasing tendency and the determined ribA and cytC concentrations exhibited a increasing tendency, both with increasing temperature. However, the change was not significant.

3.3 Application for Process Development

The applicability of the established analytical assay for selective protein quantification was tested on an adsorber screening performed in HTE mode. The screening including four different cation exchanger resins (CM Ceramic HyperD F, Toyopearl SP 650 M, SP Sepharose FF, and Fractogel EMD SO_3) and generated a total of 96 samples. The objective of the screening was to determine which adsorber would yield the best separation of ribA, cytC, and lysozyme. After determining the protein concentrations in the fractions, an asymmetric Gaussian function was fitted to the data points. In fig. 5A the resulting chromatograms are displayed. Mass balances and chromatographic resolutions were calculated based on the fitted curves. For comparison, the protein concentrations in the fractions were also determined using analytical chromatography. Based on both analytical approaches, the chromatographic resolutions and the mass balances were calculated. The results hereof are displayed in fig. 5B-C. A tendency of marginally higher resolutions was observed when using analytical chromatography for protein quantification. This can be explained by the higher precision of the analytical chromatographic assay at low concentrations (RSD $\sim$ 1 % at 0.05-0.3 g/L). RibA and cytC were separated with similar resolution (R $\sim$ 2) on column 1 and 2. CytC and lys were separated with similar resolution (R $\sim$ 2.5) on column 2 and 4. Only column 2 was able to separate all three proteins and column 3 was not able to separate any of three proteins under the applied conditions. Despite the mentioned differences in the determined resolutions, both analytical approaches delivered the same overall result of resin 2 (Toyopearl SP 650 M) being the best suited and resin 3 (CM Ceramic HyperD F) being the least suited resin for the separation of the three proteins. The mass balances for each of the protein species were calculated based on both quantification methods (fig. 5C). PLS based multivariate calibration resulted in mass balances in the range 83%-109 % and analytical chromatography resulted in mass balances 86%-104 %. This result corresponds with the higher accuracy of the chromatographic assay. Further, a tendency of a lower overall recovery was observed when using the chromatographic assay. In general, the mass balances are very acceptable

for this type of screening. The advantages of the presented fast multivariate assay including the high throughput (> 3 samples per min), the non-invasive and label-free methodology which is important in case the material/proteins are needed for further experiments, and the easy automation far outweighs the slightly lower precision compared to the assay based on analytical chromatography.

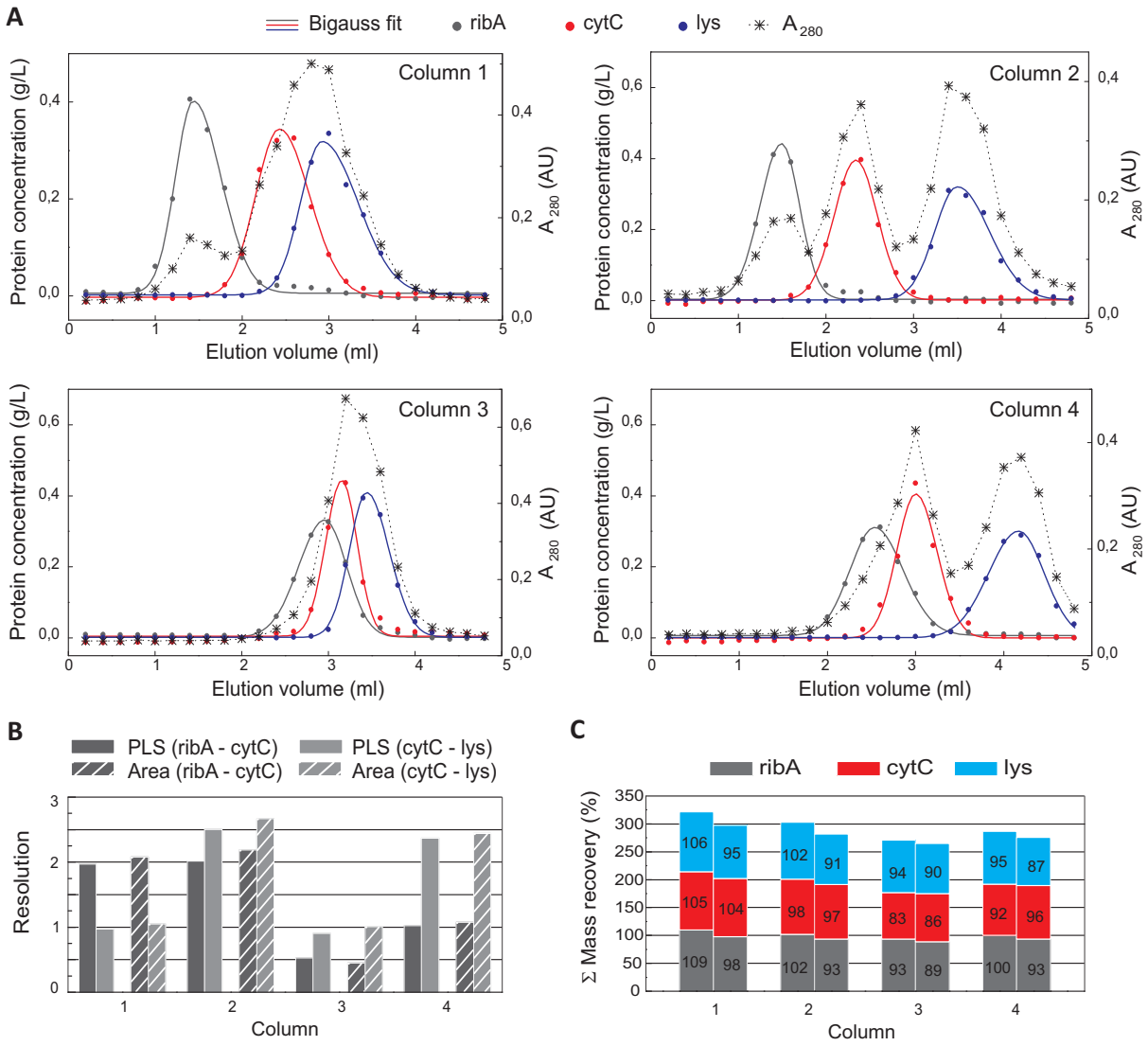


Figure 5: Results of the resin screening. Column 1: SP Sepharose FF, column 2: Toyopearl SP 650 M, column 3: CM Ceramic HyperD F, column 4: Fractogel EMD SO_3 **A:** Resulting chromatogram for each applied resin type. For each fraction, the concentration of the three protein species and the total absorption signal is displayed. The concentrations were determined using the fast absorption spectra based assay. Each data point represent the determined concentration using the fast absorption spectra based assay. An assymetrical Gauss function was fitted to the data points. **B:** Based on the fitted curves the chromatographic resolution of ribA-cytC and cytC-lys was calculated for both analytical approaches. **C:** Based on the fitted curves the mass balances for each column were calculated for both analytical approaches.

4 Conclusion

An assay capable of non-invasive and label-free selective protein quantification in protein mixtures was successfully established. The thorough assay validation showed that it generates very precise results. PLS was shown to generate more precise calibrations when compared to MLR and further mean centering of the spectral data was shown to be essential for precision. To demonstrate the applicability of the assay, it was used as analytics for an adsorber screening performed in HTE mode. The yielded results were highly similar to those achieved by analytical chromatography, hence it can be concluded that a high reduction in analysis time without loss of analytical information was possible. The presented methodology for selective protein quantification therefore holds potential to help eliminate the analytical bottleneck which has been created by the enhanced experimental throughput in downstream process development. In future work, the impact of the number of proteins and their spectral characteristics on the precision of the method will be investigated. Further, possibilities in the area of in-line application as process analytics for preparative chromatography will be pursued.

References

[1] A. Susanto, K. Treier, E. Knieps-Grünhagen, E. von Lieres, J. Hubbuch, High Throughput Screening for The Design and Optimization of Chromatographic Processes: Automated Optimization of Chromatographic Phase Systems, Chemical Engineering & Technology 32 (1) (2009) 140–154.

[2] M. Degerman, K. Westerberg, B. Nilsson, A Model-Based Approach to Determine the Design Space of Preparative Chromatography, Chemical Engineering & Technology 32 (8) (2009) 1195–1202.

[3] S. Oelmeier, F. Dismer, J. Hubbuch, Application of an Aqueous Two-Phase Systems High-Throughput Screening Method to Evaluate mAb HCP Separation, Biotechnology and Bioengineering 108 (1) (2011) 69–81.

[4] J. L. Coffman, J. F. Kramarczyk, B. D. Kelley, High-Throughput Screening of Chromatographic Separations: I. Method Development and Column Modeling, Biotechnology and Bioengineering 100 (4) (2008) 605–618.

[5] P. S. Wierling, R. Bogumil, E. Knieps-Grünhagen, J. Hubbuch, High-Throughput Screening of Packed-Bed Chromatography Coupled With SELDI-TOF MS Analysis : Monoclonal Antibodies Versus Host Cell Protein, Biotechnology and Bioengineering 98 (2) (2007) 440–450.

[6] M. Wiendahl, P. S. Wierling, J. Nielsen, D. F. Nielsen, J. Krarup, A. Staby, J. Hubbuch, High Throughput Screening for the Design and Optimization of Chromatographic Processes - Miniturization, Automation, and Parallelization of Breakthrough and Elution Studies, Chemical Engineering & Technology 6 (31) (2008) 893–903.

[7] A. Susanto, E. Knieps-Grünhagen, E. von Lieres, J. Hubbuch, High Throughput Screening for the Design and Optimization of Chromatographic Processes: Assessment of Model Parameter Determination from High Throughput Compatible Data, Chemical Engineering & Technology 31 (12) (2008) 1846–1855.

[8] J. R. Whitaker, P. E. Granum, An Absolute Method for Protein Determination Based on Difference in Absorbance at 235 and 280 nm, Analytical Biochemistry 109 (1) (1980) 156–9.

[9] R. K. Scopes, Measurement of Protein by Spectrophotometry at 205 nm, Analytical Biochemistry 59 (1) (1974) 277–82.

[10] S. Y. Xie, R. B. Huang, L. S. Zheng, Separation and identification of perchlorinated polycyclic aromatic hydrocarbons by high-performance liquid chromatography and ultraviolet absorption spectroscopy, Journal of Chromatography A 864 (1) (1999) 173–7.

[11] K. Thoma, I. Ziegler, Simultaneous quantification of released succinic acid and a weakly basic drug compound in dissolution media, European Journal of Pharmaceutics and Biopharmaceutics 46 (2) (1998) 183–90.

[12] A. C. Muñoz-Muñoz, J. L. Pichardo-Molina, G. Ramos-Ortíz, O. Barbosa-García, J. L. Maldonado, M. A. Meneses-Nava, N. E. Ornelas-Soto, A. Escobedo, P. L. López-de Alba,

Identification and Quantification of Furanic Compounds in Tequila and Mezcal using Spectroscopy and Chemometric Methods, Journal of the Brazilian Chemical Society 21 (6) (2010) 1077–1087.

[13] G. M. Anantharamaiah, V. K. Mishra, G. Datta, H. C. Cheung, Peptide and Protein Drug Analysis, Informa Healthcare, 1999, Ch. 26, pp. 277–282.

[14] D. Krzyzanowska, M. Lisowski, M. Kochman, UV-difference and CD spectroscopy studies on juvenile hormone binding to its carrier protein, Journal of Peptide Research 51 (2) (1998) 96–102.

[15] C. O. Rock, Environment of the Aromatic Chromophores of Acyl Carrier Protein, Archives of Biochemistry and Biophysics 225 (1) (1983) 122–9.

[16] J. Z. Porterfield, A. Zlotnick, A simple and general method for determining the protein and nucleic acid content of viruses by UV absorbance, Virology 407 (2) (2010) 281–8.

[17] G. Escandar, P. Damiani, H. Goicoechea, A. Olivieri, A review of multivariate calibration methods applied to biomedical analysis, Microchemical Journal 82 (1) (2006) 29–42.

[18] B. K. Lavine, Chemometrics, Analytical Chemistry 72 (12) (2000) 91–98.

[19] T. Næs, B.-H. Mevik, Understanding the collinearity problem in regression and discriminant analysis, Journal of Chemometrics 15 (4) (2001) 413–426.

Selective High Throughput Protein Quantification Based on UV Absorption Spectra

Sigrid K. Hansen[1], Babak Jamali[2], and Jürgen Hubbuch[1]

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

[2]*CMC Project Planning & Management, Novo Nordisk A/S, Denmark*

Manuscript published in ***Biotechnology and Bioengineering***

Abstract

The application of high throughput experimentation (HTE) in protein purification process development has created an analytical bottleneck. Recently, a new label-free and non-invasive methodology for analyzing multicomponent protein mixtures by means of spectral measurements was presented. Analytics based on the methodology was shown to increase analytical throughput for selective protein quantification significantly, however this was only demonstrated for one particular protein combination. In this work the possibilities and limitations of the analytical method are investigated further. Principal component analysis (PCA) was performed on a broad range of absorption spectra to investigate their common characteristics and differences. The PCA was used both for cluster analysis and to define a measure for spectral similarity. For binary protein combinations, the calibration precision was shown to decrease exponentially with the defined spectral similarity factor. Knowledge of this correlation can be used to determine a priori whether a calibration will be successful or not. Calibration robustness was investigated by applying the analytics to liquid chromatography performed in HTE mode. Further it was shown, that a spectral difference of 0.6 % was sufficient to sucessfully preform a spectral based calibration of two IgG1 monoclonals.

Keywords: protein analytics; analytical bottleneck; selective protein quantification; protein UV absorption spectra; high throughput experimentation; high throughput process development

1 Introduction

The paradigm shift towards quality by design (QbD) as standard in the pharmaceutical industry is progressing [1]. As QbD implies identification and evaluation of critical process parameters and detailed documentation of their linkage to product quality, the focus of process development has been extended beyond the sole identification of (optimal) process parameters. Hence, the amount of sound process knowledge obtainable during the short time frame available for protein purification process development is of increasing importance.

High throughput process development (HTPD) is a concept which can intensify the gain of process knowledge during process development [2]. It is based on miniaturized, parallelized, and automated unit operations for protein purification which enables a high experimental throughput, also referred to as high throughput experimentation (HTE). In order to take HTPD beyond traditional 'one factor at the time' screening, HTE is often supported by more sophisticated approaches such as factorial designs, intelligent search algorithms or mechanistic modeling [3–5]. Today, most of the common unit operations applied for protein purification have already been established in a HTE mode and integrated in the HTPD concept [6–10].

Univariate spectroscopic measurements performed in plate readers is the preferred analytics used in combination with HTE as it enables maintenance of the experimental throughput and automation. We recently reported a method capable of selective protein quantification in protein mixtures based on UV absorption spectra [11]. This method combines the advantages of spectroscopy such as easy automation, non-invasiveness, and high throughput with the capability of selective protein quantification for which analytical chromatography or electrophoresis would traditionally be applied.

Spectroscopy is of fundamental importance in protein science. Fields of applications range from determination of protein conformation and reaction channels to simpler purposes such as detection, quantification and visualization in various forms. In the particular case of protein mid-UV (200 - 300 nm) absorption spectroscopy, physical properties of peptide bonds and a few amino acid residues are responsible for the absorption. In Figure 1 the spectra of these absorbing components are shown. Mainly the aromatic structures of phenyl alanine, tyrosine, and phenyl alanine contributes to the mid-UV absorption of proteins. Each absorption spectra of these residues exhibits two peaks, one in the area 250 – 280 nm and another at lower wave lengths around 215 – 230 nm. Cysteine residues and peptide bonds also absorb mid-UV light, mainly below 260 nm. The absorption of cysteine remains rather low, where as the absorption of the peptide bonds increase strongly toward lower wave lengths. These characteristics are

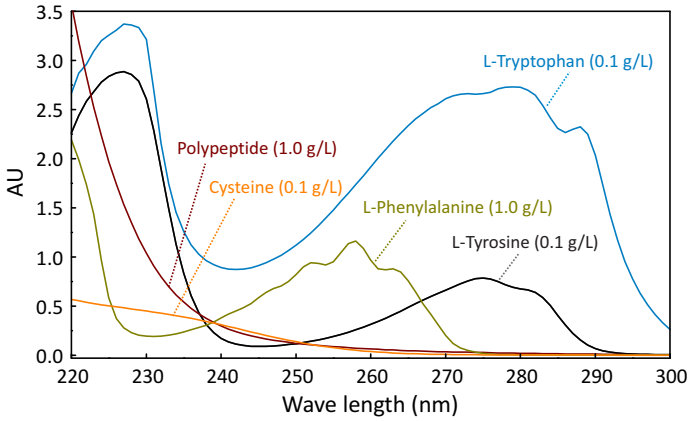


Figure 1: UV absorption spectra of amino acids. Spectra of the amino acids which absorb in the Mid-UV range are displayed together with the spectrum of a leucine polypeptide to also demonstrate the absorption caused by the peptide bonds.

text book knowledge, however the application of protein mid-UV spectra and their unique relation to protein identity has not yet led to a wide application for selective quantification in protein mixtures. At lower wave lengths around 190 – 210 nm mainly the peptide bonds absorb. Hence, it is also possible to obtain information on the composition of a protein mixture via the coefficient of e.g. the absorption at 280 nm and 205 nm. However, due to material and apparatus issues of absorption measurements at such low wave lengths, these are often avoided.

The recently reported method for selective protein quantification based on the use of UV absorption spectra was demonstrated to work for one particular combination of proteins. In this work, the application is extended to a wider number of proteins to prove its general applicability. In total, 26 proteins were included in a spectral cluster analysis to generate an overview of spectral characteristics and variation. Further, it was attempted to define a measure for spectral similarity which a relation to precision of the selective quantification method based on protein spectra.

2 Materials & Methods

2.1 Disposables

PD-10 desalting columns (GE Healthcare, Buckinghamshire, UK) for buffer exchange of protein solutions. Syringe filters with 0.22 μm pore size (Millipore, Billerica, MA, USA) for filtration of non dissolved protein. UV-grade polystyrene micro titer plates (MTP) in round 96 well half area format and a volumetric capacity of 180 μL (Greiner Bio-One, Kremsmnster, Austria) for measurement of UV absorption spectra. Half micro sized UV-grade cuvettes with a working volume $\geq$ 1.5 mL (BRAND GMBH + CO KG, Wertheim, Germany) for spectral measurements in a double beam photometer. Square 96 well deep well plates (DWP) for preparation of elution buffer for HTE chromatography. RoboColumns for performance of HTE chromatography prepacked with 200 μL Macro-Prep Ceramic Hydroxyapatite TYPE I, 80 μm (Atoll, Germany).

2.2 Apparatus

The generation of protein samples for PLS regression model calibration and validation was performed automated on a Tecan Freedom Evo 100 (Tecan, Crailsheim, Germany) liquid handling station (LHS) equipped with 250 μL dilutors. Column chromatography in HTE mode using RoboColumns was performed on a Tecan Freedom Evo 200 LHS equipped with 1 mL dilutors. A thorough description of parallelized chromatography performed automated on an LHS has been published by [12]. The used setup included a custom made column carrier. A similar carrier is commercially available under the name Te-Chrom (Tecan, Crailsheim, Germany). Further, the LHS was equipped with an Infinite 200 plate reader (Tecan, Crailsheim, Germany). The LHS was operated with EVOware 2.0. Import of values for pipetting volumes and export of spectral data was handled via Excel (Microsoft, Redmond, USA). For spectral measurements in cuvettes, the double beam photometer Lambda 35 from Perkin Elmer (Waltham, MA, USA) was used.

2.3 Proteins

Human insulin, hen egg white avidin and immunoglobulinG (recombinant IgG1 monoclonal) were provided by an industrial project partner. Human insulin and hen egg white avidin were provided in solid form and human IgG in aqueous solution at a concentration of 7 g/L. All other

Table I: Alphabetical listing of all proteins used in the work. The factor listed for each protein is a spectral intensity factor, used to normalize protein concentrations in g/L with respect with their absorption spectra in the range 250 - 300 nm.

N°	Name	Origin	Factor
1	avidin	hen egg white	0.80
2	carbonic anhydrase	bovine	0.93
3	catalase	bovine	1.02
4	α-chymotrypsin	bovine	0.65
5	α-chymotrypsinogen	bovine	1.06
6	conalbumin	hen egg white	1.07
7	concanavalin	*C. ensiformis*	1.13
8	cytochromeC	equine	0.47
9	β-glucoronidase	bovine	0.76
10	glucose oxidase	*Aspergillus*	0.84
11	hemoglobinBO	bovine	0.54
12	hemoglobinHU	human	0.85
13	immunoglobulinG	human*[1]	0.72
14	insulin	human*[2]	1.25
15	α-lactalbumin	bovine	0.64
16	β-lactoglobulin	bovine	1.29
17	lysozymeAV	hen egg white	0.49
18	lysozymeHU	human*[3]	0.58
19	myoglobin	equine	0.58
20	ovalbumin	hen egg white	1.51
21	ovomucoid	hen egg white	1.83
22	ribonucleaseA	bovine	1.70
23	serum albuminBO	bovine	1.68
24	serum albuminHU	human	1.72
25	thaumatin	*T. daniellii*	1.06
26	transferrin	bovine	0.82

*derived from trangenetic source:
[1]chinese hamster ovary cells [2]yeast [3]rice

proteins were commercially purchased in solid form (crystallized, precipitated or freeze-dried) from Sigma-Aldrich (St. Louis, MO, USA). The proteins, their origin (for transgenic sources, the sequence origin), and spectral intensity factor are listed in Table I.

Protein solutions for spectral measurements only, were prepared in 20 mM phosphate buffer at pH 7 with a nominal mass concentration of 1.0 g/L. This protein concentration was based merely on weight of the protein powders and did not account for salt or other buffer components present in the purchased solid protein preparations. Some proteins (βglucoronidase, thaumatin, concanavalin) exhibited a solubility below 1 g/L in the used buffer. These solutions were filtered with 0.22 μm syringe filters prior to spectral measurement. Due to their low solubility, these proteins were only included in the cluster analysis. A buffer exchange of all protein solutions was performed to remove buffer salts or other components present in the purchased protein preparations. The spectral intensity factors (listed in Table I) were based on the absorption spectra in the range 250 – 300 nm of mass equivalent protein concentrations. The factor for each protein was determined as the quotient of the mean spectral intensity of all proteins and the spectral intensity of each single protein. Spectral intensity was defined as the sum of the 51 absorbance values contained in a spectrum.

2.4 Protein Spectra

Where nothing else is stated, protein spectra were measured in UV-grade MTPs. Each sample had a volume of 150 μL and the spectra were recorded in steps of 1 nm in the range 250 – 300 nm.

According to manufacture's data the bandwidth of the plate reader is 5 nm in the range used to record the spectra. For spectral measurements used for cluster analysis, spectra of the 1 g/L protein solutions were measured and subsequently the concentrations of these solutions were adjusted for all spectra to give the same sum of measured absorption values. Measurements with the Lambda 35 double beam spectrophotometer were performed with a band width of 1 nm and recording steps of 1 nm.

2.5 Protein Difference Spectra

Difference spectra were uesd for visual comparison of protein spectra. The mean spectrum of the spectra to be compared was calculated and then subsequently substracted from each of the spectra to be compared.

2.6 Cluster Analysis

Prior to cluster analysis, the spectral data was preprocessed by normalization and mean centering. Hierarchical cluster analysis was performed using the MATLAB based PLS toolbox (Eigenvector, Wenatchee, WA, USA). The hierarchical clustering was performed with *Ward's method* which is a minimum variance method and was based on principal component analysis (PCA) of the spectral data. The hierarchical cluster was visualized as a dendrogram. Non hierarchical cluster analysis was also based on PCA. In this case, the cluster analysis was visualized as a scatter plot where the score of each spectrum on the 2^{nd} PC was plotted as a function of the score on the 1^{st} PC.

2.7 Spectral Similarity

The determination of spectral similarity was based on PCA. By using the scores on the first three PCs as determined by the PCA, all spectra were represented by a point in a three dimensional space. For binary systems the spectral similarity was defined as the distance between two protein spectra, *i. e.* the distance between two points representing the respective spectra in a three dimensional space. Hence, a short distance would indicate high similarity and *vice versa*. For ternary systems, the similarity was defined as the area of the triangle spanned by three proteins, *i. e.* the area spanned by three points in the three dimensional space. In this case, a small area would indicate high similarity and *vice versa*.

2.8 Selective Protein Concentration Calibration

Partial least squares (PLS) regression was used to calibrate multivariate models for determination of selective protein concentrations. Each calibrated PLS model was validated by an independent sample set of defined concentrations to determine its precision. Precision was defined as the coefficient of determination (R^2) based on the independent validation samples. For the binary protein combinations, the composition of calibration and validation samples was based on a six layer onion design which consisted of 16 samples. For ternary and quaternary proteins combinations, the composition of calibration and validation samples was based on a four layer onion design which consisted of 32 samples for ternary systems and 40 samples for quaternary systems. The designs were generated with MODDE (Umetrics, Umeå, Sweden). The binary and ternary protein calibrations were performed in a concentration range from

0 – 0.4 g/L and the quaternary calibration was performed in the concentration range from 0 – 0.2 g/L. For each protein combination, two sample sets were prepared of which the first was used for calibration and the second for validation. For calibration and validation, the MATLAB based PLS toolbox (Eigenvector, Wenatchee, WA, USA) was used. Calibration and validation samples were prepared on the LHS from stock solutions with a concentration of 0.8 g/L for binary and quaternary systems and 1.2 g/L for ternary systems. All mentioned concentrations are spectral intensity corrected concentrations. The true concentrations of the stock solutions were determined according to the spectral intensity factor of each protein as listed in Table I. In order to prepare a spectral intensity corrected protein solution with a nominal concentration of 0.8 g/L of *e.g.* lysozymeHU, the intended nominal concentration was multiplied with lysozymeHU spectral intensity factor to give the actual concentration of the solution.

2.9 Column Chromatography

All columns were equilibrated with five column volumes (CV) loading buffer (1 mM sodium phosphate, pH = 6.8) before loading protein on to the columns. The amount of protein loaded to the columns was adjusted according to spectral absorption intensity of each protein. A nominal concentration of 5 g/L was multiplied with the spectral intensity factor to produce the actual concentration of protein in the load solution. Of the protein load solution, 50 μL were loaded to the column. Before starting the actual column experiments, 18 different buffer solutions with increasing sodium phosphate concentrations were prepared and stored in DWPs until use. The prepared buffer solutions composed a linear step gradient from 1 – 300 mM sodium phosphate at pH 6.8 over 18 CV. Each gradient step was collected as one elution fraction. After completed elution, 150 μL of each elution fraction were transferred into UV grade MTPs and the absorption spectra were measured. Based on these spectra, the protein content in the samples was determined using previously calibrated PLS regression models as described in section 2.8.

3 Results & Discussion

It has recently been shown that mid-UV absorption spectra can be applied to perform selective protein quantification in protein mixtures. To demonstrate a broad application of the method, many different proteins have been included in this work. An overview of spectral variety among the proteins was generated and followed by an investigation of the relation between spectral similarity and calibration precision.

3.1 Clustering

3.1.1 Dendrogram

order to obtain a preliminary overview of the spectral data, a hierarchical cluster analysis was performed and visualized in the form of a dendrogram (Figure 2A). Within the 26 protein spectra five clusters were identified and labeled 1 – 5. Cluster identification was based on the increase in variance within each cluster with decreasing number of clusters as plotted in Figure 2B. Below five clusters the maximal variance within the clusters increases significantly and hence this point was chosen as threshold for cluster identification. Cluster 4 had the highest variance of ~ 0.5 and cluster 1 the lowest variance of ~ 0.3. The clusters 2, 3, and 4 had a variance of ~ 0.4. When looking at the spectra in Figure 2C on which the dendrogram is based, the clusters 1 and 2 (green and red) can easily be identified from the complete set of curves by visual inspection,

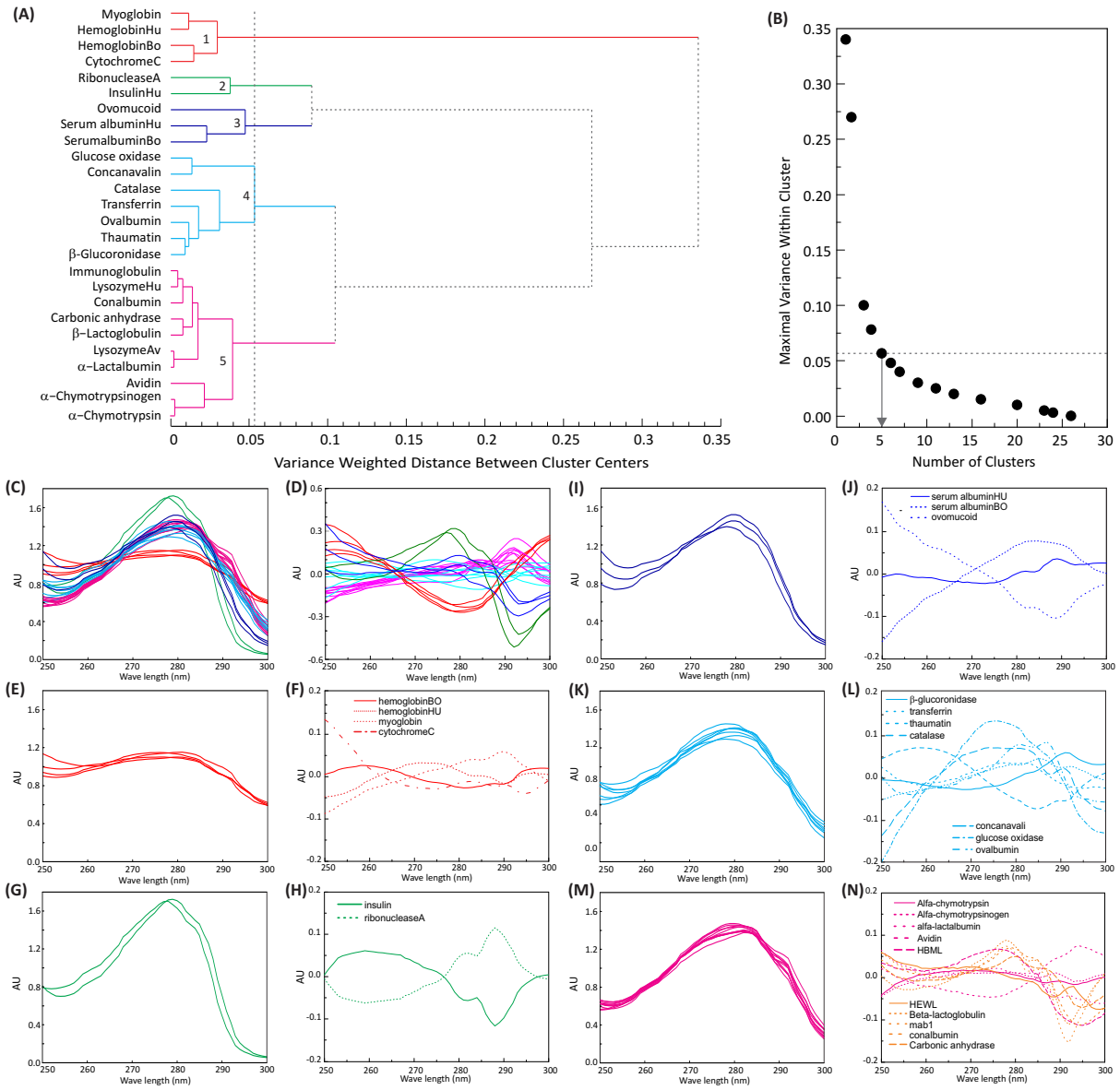


Figure 2: PCA based cluster identification based on spectral data. **A:** From 26 protein spectra a PCA based dendrogram was computed to give an overview of spectral similarity. In the dendrogram, five clusters were identified (labeled $1, 2, 3, 4$, and 5) and given different main colors. **B:** Decrease in the number of clusters as a function of the variance within each cluster. **C:** All Spectra in the color of each respective cluster. **D:** Difference spectra of all measured spectra. **E-N:** Spectra of single clusters colored as the respective clusters in the dendrogram and to the right of each cluster the respective difference spectra.

however the clusters $3, 4$, and 5 are more difficult to distinguish. In Figure 2 E, G, I, K, and M, the spectra of each cluster are plotted separately. Here, the significant traits of all clusters become visible. Cluster 1 has the least internal variance and is also the most isolated cluster. All of the proteins in this cluster are heme proteins, and the heme group is assumed to cause the characteristic flat shape of these spectra. Cluster 2 consists of only two spectra with a very narrow shape in the range $260 - 300$ nm. These are spectra of human insulin and ribonucleaseA

both of which contain no tryptophan. This explains their narrow shape, as tryptophan has a broader spectrum compared tyrosine and above 290 nm tryptophan is the sole contributor to absorption (see Figure 1). The spectra of cluster 3 are broader in the range 260 – 300 nm compared to the spectra of cluster 2 and the spectra of cluster 4 and 5 follows the pattern of the spectra becoming broader. As the spectra becomes broader, the shoulder at 290 nm becomes more distinct which specifically indicates an increase in tryptophan content. Hence, the hierarchical cluster analysis generated clusters mainly based on the relative tryptophan content in the proteins with an increase in tryptophan content through cluster 2 to 5. This correlates with the fact that tryptophan is the strongest contributor to the UV absorption spectra with a seven times stronger absorption per mole in the 250 – 300 nm when compared to tyrosine and 70 times stronger when compared to phenyl alanine. The observation that many of the spectra exhibits an increase in absorption below 255 nm does not correlate with the pure spectra of the aromatic amino acids and can also not be explained by the cysteine residues or peptide bonds alone. The observation is most likely due to a solvatochromic effect, as the aromatic amino acids are mostly situated in the hydrophobic core of a native protein and not in hydrophilic surroundings as the aqueous buffer in which the were in fact measured.

3.1.2 PCA Scatter

Hierarchical clustering provides a helpful overview of the spectral data, however the quantitative measure obtained is only the distance between cluster centers and when ascending in the hierarchy only distance between merged cluster centers. Therefore, it can not be used to give a measure on the spectral similarity of every random combination of proteins. In order to obtain a quantitative measure and visualization of spectral similarity of all spectra, PCA was performed and visualized as a scatter plot. PCA is a projection method, which is able to condensate significant information from larger data sets. Figure 3 shows the resulting scatter plot of the scores on two first PCs for each protein spectrum included in the PCA.

The PCA showed that two PCs described 96.6 % of the variation in the spectral data with 55.52 % described by the first PC and 40.48 % described by the second PC. The third PC described 2.2 % and is also considered relevant, however the plot was limited to the display of two dimensions in order to simplify the visual interpretation. The data points are colored with respect to cluster belonging as determined by the hierarchical chluster analysis. The clusters identified by the hierarchial cluster analysis corresponded well to the PC scatter plot, *i. e.* each color symbolizing cluster belonging as determined by the hierarchial cluster analysis also forms a cluster in the PC scatter plot. In the scatter plots, the clusters of narrow spectra are positioned to the left and the clusters move to the right and downwards as the spectra gets broader (see spectra of single clusters in Figure 2).

Seven of the protein spectra were used to validate the PCA model. These are displayed as triangles in the scatter plot. Their positions correlate with the main clusters determined by the hierarchical cluster analysis, *i. e.*, avidin is closest to α-chymotrypsinogen and α-chymotrypsin, α-lactalbumin is closest to lysozymeAV, and so forth. This indicates that the PCA model has predictive power and can be used to determine the similarity of protein spectra not used to build the initial PCA model. This is important as the idea was to establish a correlation between spectral similarity and possible precision obtainable for a calibration of selective protein concentrations based on UV spectra. If the PCA model has predictive power and a correlation is established between spectral similarity and calibration precision, the PCA model can be used to give an a priori estimate on the precision of a calibration.

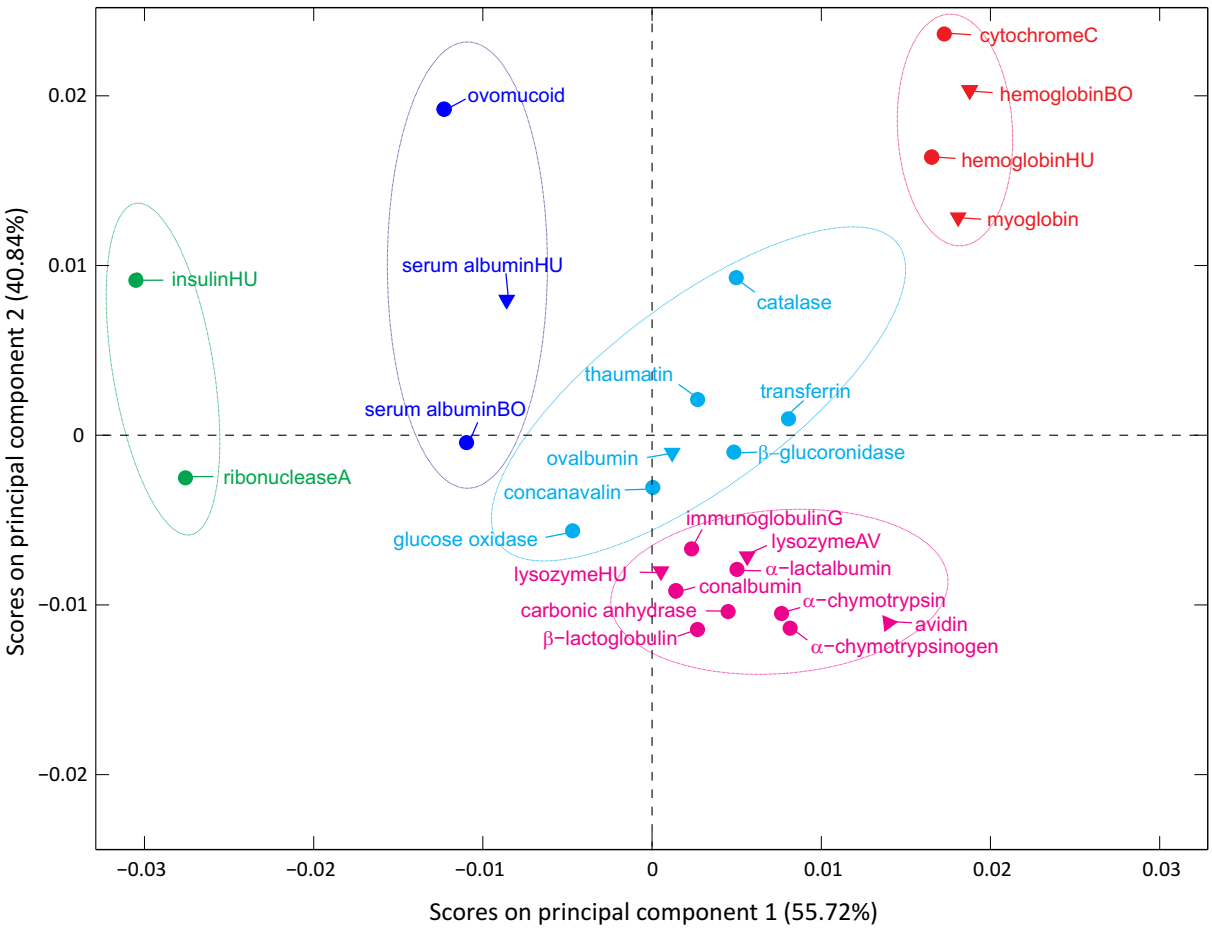


Figure 3: PCA based scatter plot of spectral data. Principal component plot of the protein spectra in which each data point represent the scores on the 1^{st} and 2^{nd} principal component as calculated from one protein spectrum. The data points are colored with respect to the clusters defined by the dendrogram cluster analysis (see fig. 2A). Eight spectra were chosen to validate the PCA model and their the corresponding data points are represented by triangles.

As in the dendrogram, the scatter plot shows that cluster 1 is the most isolated cluster. However, the fact that cluster 2 is the second most isolated cluster is an information which could not have be derived from the dendrogram. Further, it would not have been possible to identify whether *e. g.* cluster 1 is more or less similar to any of the other clusters without looking at the actual spectra from the clusters. From the scatter plot this can be determined by looking at the distance between each cluster center. As can be seen, the distances from cluster 1 to the centers of cluster 3 and 5 are similar (~ 0.3) and the distance to cluster 3 is a bit shorter (~ 0.25). Accordingly, the similarity between any two spectra can be determined from the scatter plot.

3.2 Calibrations

In order to get a measure for the similarity of the measured protein spectra, each spectra was defined by its principal component scores on the first three principal components. For a binary systems, the similarity was then defined as the distance and for ternary systems the similarity was defined by the spanned area. The determined similarities were set in relation

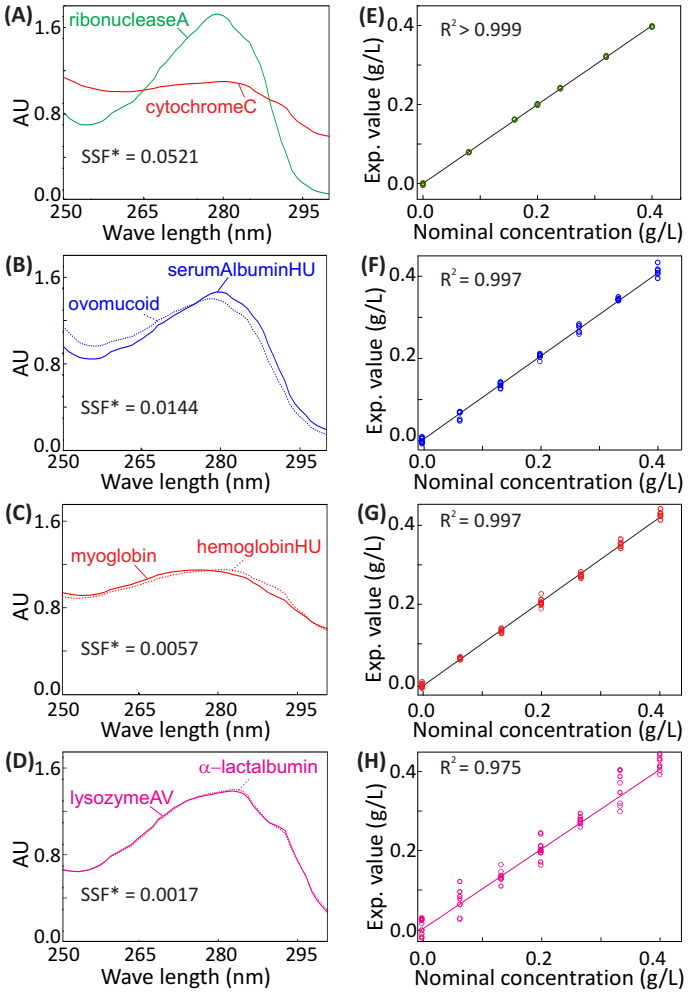


Figure 4: Protein spectra and calibration validation of four binary protein combinations. **A-D:** Protein spectra of four of the eleven protein combinations chosen for calibration. The colors of the spectra represent their respective cluster belonging as determined by the dendrogram cluster analysis (see fig. 2A). **E-H:** Visual display of the calibration result for the respective protein combination shown to the left. *spectral similarity factor

to the calibration precision achieved for each of the respective systems in order to investigate the correlation between these two parameters. A low spectral similarity value indicates a high spectral similarity and *vise versa*.

3.2.1 Binary systems

PLS regression models were calibrated for eleven different protein combinations. All models were validated with independent samples and the precision of each calibrated PLS regression model was defined as the coefficient of determination based on the respective validation samples. In Figure 4 the spectra of four protein combinations with increasing spectral similarity are shown along with validation results. There is a clear correlation between spectral similarity and precision at the extremes of high or low spectral similarity (Figure 4A/E and D/H), however when comparing the systems in between the picture becomes ambiguous. In Figure 5B and C the spectra and calibration curves of the systems serum albuminHU/ovomucoid and myoglobin/hemoglobinHU are displayed. Both systems achieved the same calibration precision of 0.997 even though the spectral similarity factors were very different, *i. e.* 0.0057 for myoglobin/hemoglobinHU and 0.0145 for the serum albuminHU/ovomucoid system. In Figure 5A the determined calibration precision for all binary systems is shown as a function of determined spectral similarity. The validation of the calibrated PLS regression models revealed coefficients

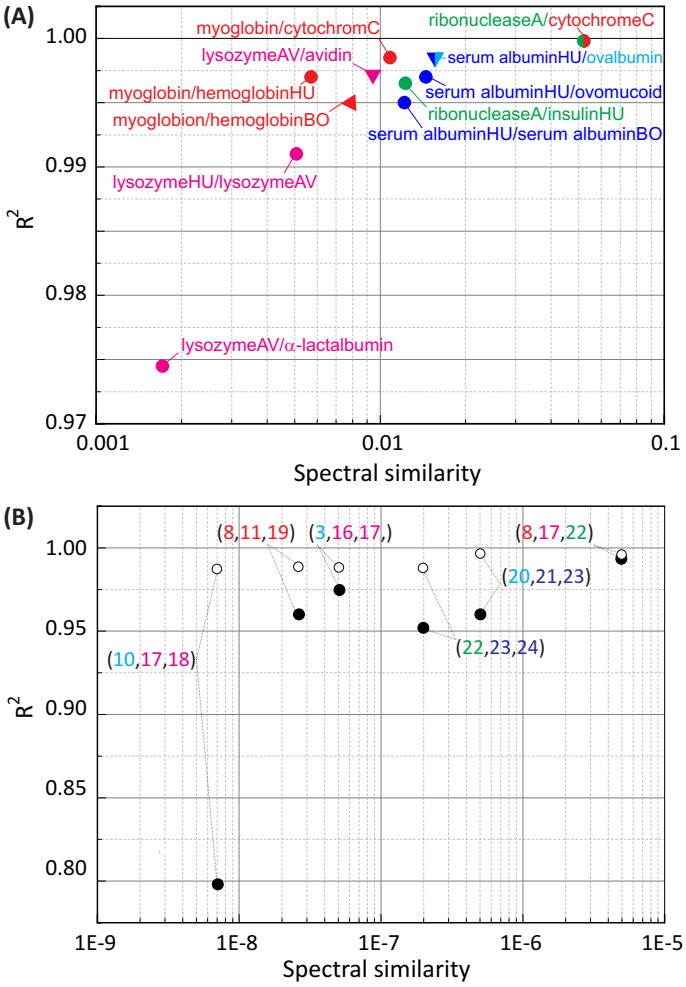


Figure 5: Calibration of selective protein concentrations based on recorded UV spectra including eleven binary and six ternary protein combination. **A**: Coefficient of determination for each binary system displayed as a function of the spectral similarity coefficient (distance between the first three principal components of the two spectra). The colors refer to cluster belonging. **B:** Coefficient of determination for each ternary system displayed as a function of the spectral similarity coefficient (area of triangle spanned by the first three principal components of the three spectra). Closed circles represent calibrations based on 3 latent variables and open circles represent calibrations based on 4 latent variables. The numbers refer to the proteins in Table I and the colors refer to cluster belonging.

of determination ranging from 0.975 to >0.999. The most precise calibration was achieved for the ribonucleaseA/cytochromeC system which is also the least similar system with a spectral similarity of ~ 0.052 and these proteins also had the two most different spectra of all measure spectra. The least precise calibration was achieved for the lysozymeAV/α-lactoglobulin system which correlated with the fact that these proteins exhibited the highest spectral similarity with a similarity factor of 0.0016.

The before mentioned example of protein combinations with rather different spectral similarity factors resulting in identical precision, represents an extreme value in presented data, however it does mark a general trend that calibrations of proteins from cluster 1 and 5 are more precise than calibrations from cluster 2, 3, and 4. The PCA analysis on which the determined spectral similarities were based, was evaluated by an external data set consisting of seven protein spectra. Many of these spectra were of proteins used in the calibration experiments. Data points represented by triangles are systems where both proteins were used for PCA validation. In all other systems one of the proteins was used for validation, with exception ribonucleaseA/cytochromeC and ribonucleaseA/insulin all of which were not used for validation. Systems consisting of two proteins of which the spectra were used for validation, does neither create nor disturb the general trend of the cluster belonging having an influence on calibration precision and can therefore not explain the observation. As reflected by the spectral intensity

factors listed in Table I, a lower absorbance per mass unit was characteristic for the proteins from the clusters 2, 3, and 4. Because the protein solutions were prepared to have the same absorption intensity in the recorded spectral range, the applied protein solutions had different mass concentrations. The rationale for the spectral intensity corrections of the used protein concentrations, was based on the assumption that proteins with a low spectral intensity would otherwise be more influenced by noise. To exclude this effect, mass concentrations corrected for spectra intensity were used. The stock solutions used to prepare the calibration samples were in the range 0.47 – 0.84 g/L for proteins from cluster 1 and 5 and 1.25 – 1.83 g/L for proteins from cluster 2, 3, and 4 (see Table I). It has been described that increasing protein concentration decreases the surface tension of aqueous protein solutions [13]. Liquids with a relative lower surface tension tends to be more difficult to pipet and this could explain the differences in calibration precision which can not be explained by the difference in spectral similarity alone. Further, the surface tension also influences the surface geometry of a liquid in a MTP well. Pure buffer has a straight surface and with decreasing surface tension (increasing protein concentration) the surface warps into meniscal shape [14]. As the light beam passes vertically through the samples in a plate reader, the shape of the horizontal surface influences the light path length and amount of scattered light and, hence, the measurement precision.

The precision of PLS regression models based on protein UV spectra is only comparable with analytical chromatography in the most favorably cases, as generally a coefficient of determination above 0.999 can be expected of an HPLC based protein quantification assay. Only the system with the two proteins having most different spectra achieved a calibration with a comparable precision. The necessary precision of a given calibration of course depends on the application and therefore no absolute measure for the necessary spectral difference. Nine out of eleven protein combinations resulted in a coefficient of determination ≥ 0.995. For all combinations with a spectral similarity factor ≥ 0.01, the coefficient of determination was ≥ 0.995. For many HTE applications, this is thought to be sufficient. Further, the advantages of the method would in many cases compensate the trade off in precision.

3.2.2 Ternary Systems

PLS regression models were calibrated for six ternary protein combinations and validated with independent samples sets. Based on the validation samples, the precision of each calibration was determined. For ternary systems, the spectral similarity factor was defined as the area of the triangle spanned by the PC scores of the three spectra. Figure 5B displays each determined coefficient of determination as a function of the respective spectral similarity factor. Each system had been calibrated twice; once where the PLS regression model was based on three latent variables and once based on four latent variables. The precision of the PLS regression models based on three latent variables decreased from 0.993 to 0.798 with decreasing spectral similarity factor. However, only the endpoints exhibited a clear correlation. Of the four combinations in between, the two ternary combinations with proteins from the clusters 2, 3, and 4 exhibit a local decrease in precision with a decrease in spectral similarity factor and the same observation was made for the two combinations with proteins mainly from the clusters 1 and 5. As the two last mentioned protein combination exhibited the highest precision of these four combinations, the previously observed trend that binary protein combinations from cluster 1 and 5 result in more precisely calibrated PLS regression models recurs for ternary protein mixtures where the calibration was based on three latent variables. In all cases where the calibration had been based on four latent variables the precision was significantly higher compared to those based

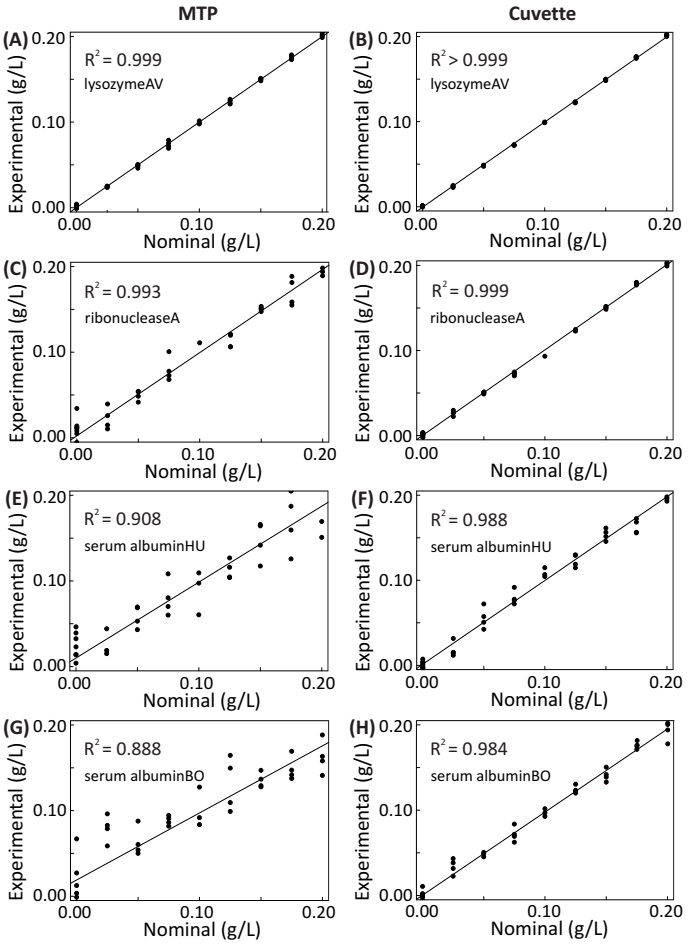


Figure 6: Validation result of a four component calibration. Displayed are the results obtained by two different photometers: left - single beam plate reader, right: double beam cuvette photometer. For each component the data points represent the determined protein concentration *vs.* nominal concentration.

on three latent variables. Further, the overall decrease in precision between the endpoints is less pronounced. For the two protein combinations with the lowest spectral similarity the determined coefficient of determination was 0.997 and for the four protein combinations with the highest spectral similarity 0.988 ± 0.06 %.

3.2.3 Quaternary Systems

The possibility to calibrate protein concentration with absorption spectra in a quaternary protein mixture consisting of lysozymeAV (lys), ribonucleaseA (rib), serum albuminHU (HSA), and serum albuminBO (BSA) was investigated. The spectra of the calibration and validation samples were measured with two different photometers: the single beam plate reader used to measure all other spectra and a double beam photometer which required the use of cuvettes as sample container. The calibration results based on both photometers are shown in Figure 6 (left column: plate reader, right column: double beam photometer). The calibration precision achieved with the double beam photometer was higher for all four proteins. As can be seen, the effect is considerable and is explained by the higher resolution and precision of the double beam photometer as well as the changed sample container geometry used herein. The effect of the sample container is due to the fact that protein solutions in MTP wells exhibit a convex surface which causes both path length inaccuracy and a higher amount of light scattering which both affect the quality of the recorded spectra.

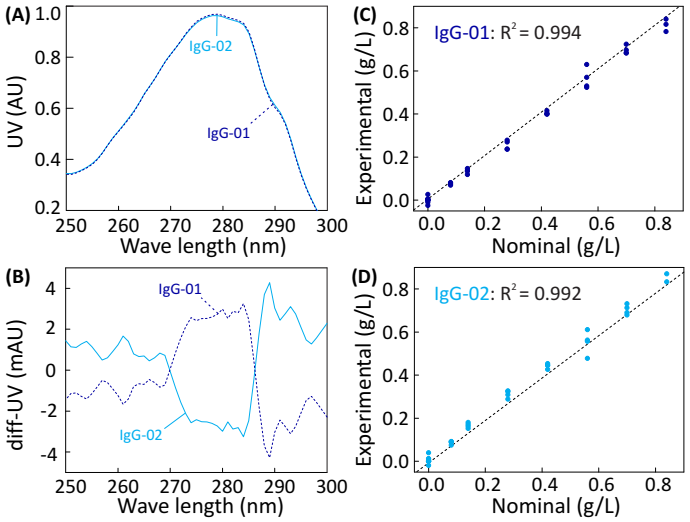


Figure 7: Validation result of a two component IgG calibration. **A-B:** Spectra and difference spectra of the two IgGs. **C-D:** For each IgG, the data points represent the experimentally determined IgG concentration *vs.* nominal concentration.

The achieved precision for each of the four components was rater different, however with the same ascending order lys > rib > HSA > BSA for both photometer types. Lys achieved a coefficient of determination ≥ 0.999 using both photometers. Also rib achieved a coefficient of determination = 0.999, however only using the double beam photometer. The coefficients of determination for HSA and BSA were ~ 0.99 for the double beam photometer and ~ 0.9 for the single beam plate reader. The two of the four proteins with the highest similarity are BSA and HSA which are both from cluster 3 (spectral similarity factor: 0.012). This correlates with the lower precision of the calibration for those two proteins. As can be seen in the PCA scatter plot (Figure 3), the BSA spectrum has a higher similarity to the spectra of rib and lys than the spectrum of HSA. This correlates with the finding that the precision for BSA was the lowest of all four components. However, the fact that the calibration was less precise for rib than for lysA can not be explained by differences in spectral similarity, as both components have very similar spectral similarity factors when set in relation to BSA and HSA (rib/HSA = 0.0179, lys/HSA = 0.0183, rib/BSA = 0.0230, lys/BSA = 0.0216). The finding does however correlate with the previous finding that proteins of which the spectra were assigned to cluster 1 and 5 generated calibrations of higher precision compared to proteins of which the spectra were assigned to the clusters 2, 3, and 4. Thus, the very different degrees of precision seems to be caused by a combination the cluster related spectral intensity factor (listed in Table I) and the different spectral similarities of the four proteins.

3.3 Case Study I: Calibration of two Monoclonal IgGs

In Figure 7A, the spectra of two different monoclonal IgGs (In the following termed IgG-01 and IgG-02) are shown. As can be seen, the spectral similarity for these two proteins is very high and by visual inspection of the spectra, almost no difference can be determined. The difference spectra (Figure7B), however, reveals spectral differences in the range $\sim 2-8$ mAU which is approximately 0.5 % of the average spectral signal (average spectral difference 3.57 mAU and average spectral signal 0.62 AU). The spectral similarity factor for the two IgGs calculated on basis of the PCA analysis was 0.0009. According to the investigated binary protein combinations, a calibration based on the spectra of these two IgGs should result in a coefficient of determination well below 0.975 (see Figure 5B) which was the coefficeint of determination measured for the lysozymeAV/α-lactalbumin combination. Hence, based on the initial studies of

the IgG spectra, a calibration was predicted to be of low quality. However, the study including four proteins measured in cuvettes in the double beam photometer, had demonstrated how it was possible to increase precision by changing measurement instrument and sample container. Hence, a calibration of the two IgGs was attempted using cuvetts and the double beam photometer. In the calibrated range 0 – 0.8 g/L, this resulted in a coefficient of determination of 0.994 for IgG-01 and 0.992 for IgG-02. To emphasize the increase in precision, the samples were also measured in MTP and as expected this resulted in a significantly lower calibration precision (IgG-01: $R^2 = 0.944$ and IgG-02: $R^2 = 0.932$).

3.4 Case Study II: HTE Chromatography

The presented analytical method enables non-invasive and label-free selective protein quantification performed automated on an LHS. Hence, the method is a valuable addition to univariate spectroscopic analytics or even a replacement of other selective quantification methods such as analytical chromatography or electrophoresis. HTE chromatography is a setup of miniaturized and parallel chromatography performed automated on an LHS. In order to demonstrate the presented methods applicability, HTE chromatography was performed with two binary protein combinations of different spectral similarity. Different degrees of spectral similarity were chosen to check for agreement with the previously demonstrated correlation between spectral similarity and calibration precision. The two protein combinations were serum albuminBO/serum albuminHU (BSA/HSA) from cluster 3 and and lysozymeAV/lysozymeHU (lysA/lysH) from cluster 5 (see Figure 2). Of the two combinations, lysA/lysH had the highest spectral similarity with a spectral similarity factor of 0.005 where as the combination BSA/HSA had a lower spectral similarity with a spectral similarity factor of 0.012 (see Figure 4). The precision of the calibrated regression models was highest for the BSA/HSA combination ($R^2 = 0.995$) and somewhat lower for lysA/lysH combination ($R^2 = 0.991$).

3.4.1 HTE Chromatogram Evaluation

Four column runs were performed per protein combination: one blank, one with both proteins, and two with pure proteins. The Figures 8A-D display the UV_{280} traces and the spectral fit error of the four column runs related to the BSA/HSA combination. Below those, in Figure 8E-G, the respective selective protein concentrations are shown. In Figure 8I-P the equivalent data for the lysA/lysH runs is displayed. The spectral fit error is a measure for the part of a spectrum which cannot be fitted to the PLS regression model calibrated for the protein combination in question. Causes therefore can be regular noise, influences of salt in the elution buffer, or separation of impurities present in the purchased protein preparations and hence calibrated into the PLS regression model with the main protein components. If the latter would be the case, the spectra of all fractions in which the co calibrated components are no longer present in the original ratio, will be partly unknown to the calibration model. The severity of the impact on the determination of selective protein concentrations will depend on the spectral differences. If the noise level is constant and no 'impurity separation' occurs, the spectral error is expected to be proportional to the UV signal intensity. This was the case for the lysA/lysH combination, however, not for the BSA/HSA combination. This could indicate that the purchased lysozyme preparations were of higher purity compared to the serum albumins. However, both lysA and lysH eluted from the RoboColumns in only 2-3 fractions and almost no separation of the two proteins was achieved. Hence, potential impurities would most likely coelute with the lysozyme species and a possibly impurity ratio would therefore remain identical to that of the

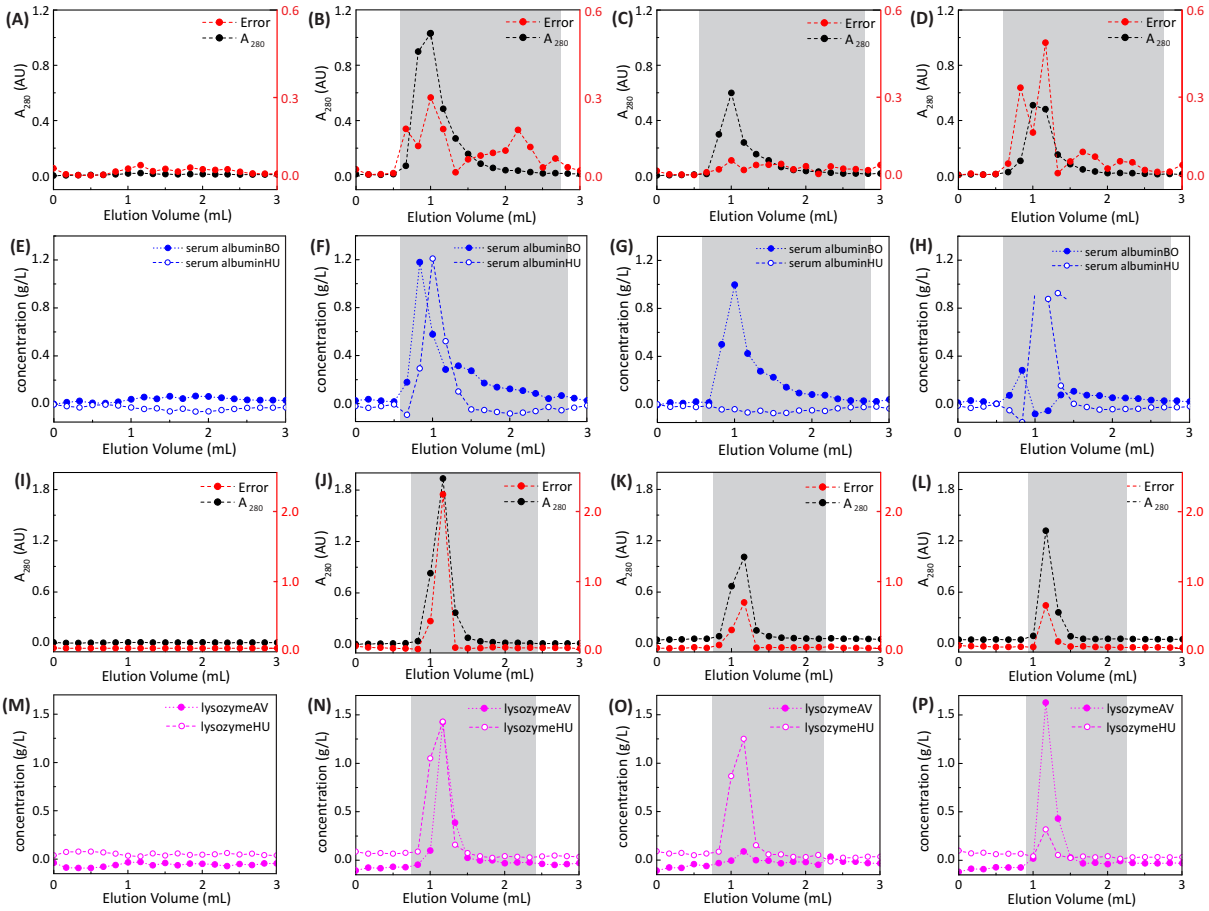


Figure 8: HTE chromatography performed with RoboColumns. In the HTE chromatograms the UV_{280} value, the spectral error value, and the selective protein concentrations as determined in each collected fraction are displayed. The HTE chromatograms displaying protein concentrations are all located below the respective chromatogram showing the UV_{280} values. The first column (A, E, I, M) shows the HTE chromatograms of the blank runs where only buffer was loaded to the RoboColumns. The second column (B, F, J, N) displays the HTE chromatograms of both protein combinations (serum albuminHU/serum albuminBO and lysozymeHU/lysozymeHU). The two last columns (C, D, G, H, K, L, O, P) displays the HTE chromatograms of each single component.

calibration. The fact that the spectral error was higher in the lysA/lysH run when compared to the single protein runs might caused by to the high protein concentrations in the elution fractions which required a significant extrapolation of the original lysA/lysH calibration. The protein concentrations of the blank run exhibited a baseline drift for both lys signals in opposite directions. The baseline drift of both lys protein signals also occurred in all three runs with protein. The runs with only lysA or lysH, revealed a positive interaction of the protein signals. This is most likely due to an inaccuracy in the calibration of the very similar spectra. Apart from the interaction, the selective concentration profiles of the single component runs matched the UV_{280} profiles of the respective runs very well. Further, they also matched the selective protein signals of the lysA/lysH run.

The spectral fit error for the BSA/HSA combination was approximately one to two orders of magnitude lower when compared to the lysozymeAV/lysozymeHU combination. This correlated with the lower similarity of the serum albumin spectra, however it is not understood whether the

	Mass Balances (%)			
Proteins	**serum albuminBO**		**serum albuminHU**	
loaded:	$UV_{250-300}$	UV_{280}	$UV_{250-300}$	UV_{280}
serum albuminHU/BO	114	101*	85	101*
serum albuminBO	95	102	-7**	-
serum albuminHU	8**	-	82	98
Proteins	**lysozymeHU**		**lysozymeAV**	
loaded:	$UV_{250-300}$	UV_{280}	$UV_{250-300}$	UV_{280}
lysozymeAV/HU	102	92*	84	92*
lysozymeHU	87	102	6**	-
lysozymeAV	11**	-	91	93

*Identical mass balance values for both proteins, as UV_{280} based quantification only enables quantification of total protein content

**hypothetical mass balance based on the nominal mass of the protein loaded to the column

Table II: Listing of protein mass balances for the HTE chromatography runs. Mass balances were calculated using both traditional univariate total protein quantification (UV_{280}) and selective quantification based on absorption spectra ($UV_{250-300}$). In Figure 6, which displays the HTE chromatograms, the gray areas indicate the fractions on which the presented mass balances are based.

correlation is causal. The selective protein concentrations of the blank run exhibited a baseline drift for both proteins in opposite directions of each other. The baseline drift of both serum albumin protein signals also occurred in all chromatograms with protein. In the BSA run, the error signal at first correlated with the intensity of the UV_{280} signal and then increased again through the 'shoulder' of the BSA peak. However, the noise level was only slightly higher than for the blank run. The HSA run exhibited a significantly higher spectral error when compared to the BSA run and also did not follow the course of the UV_{280} signal. This indicated the presence of one or more impurities in the purchased HSA, which had been at least partly separated on the column. This was reflected in the protein signals of HSA in the HTE chromatograms of both runs with HSA. In the HSA run, the first two fractions (fraction 5 and 6) of which the UV_{280} signal indicated eluting protein, the HSA signal exhibited negative values. At the same time, the BSA signal exhibited positive values even though this protein was not present. By subtracting the BSA signal of the HSA run from the BSA signal of the BSA/HSA run, the BSA signal of the BSA/HSA run became very similar to the BSA signal of the BSA run (data not shown). This supports the assumption of an impurity in the purchased HSA to be the cause of the differences in the elution profiles of BSA in the two runs with BSA. Apart from the interactions caused by the impurity present in HSA and the baseline drift, the selective protein signals of the single component runs correlated with the UV_{280} signals.

For both protein combinations, the baseline of the blank runs deviated from zero and in both cases the calculated protein concentration of one component mirrored the other around the x-axis. Due to the stable baseline of the UV_{280} signal at zero and also the mirrored curse of the selective concentrations, it is unlikely that these protein signals were caused by elution of protein remains from earlier runs. More likely, it is an calibration artifact as the baseline course of the blank runs also repeat in the runs with protein. It might be caused by a combined effect of increasing phosphate concentration in the buffer and difficulties of the calibrated model to cope with spectra of low intensity.

3.4.2 Mass Balance Evaluation

For the calculation of the mass balances, it was assumed that the baseline drifts were static and not dependent on protein concentration. Hence, for the mass balance calculations, a baseline

correction was undertaken by subtracting the baseline signals of the blank runs from the protein signals in the runs with protein. The mass balances were calculated based on the gray marked area in the HTE chromatograms. For comparison, also mass balances based on the UV_{280} signal were determined. All mass balances are listed in Table II.

The mass balances of the BSA/HSA runs exhibited a higher variability (82 - 114 % - mean: 94 %) in comparison to the lysA/lysH runs (84 - 102 % - mean: 89 %) when based on selective protein concentrations. For the BSA/HSA runs the UV_{280} based mass balances indicated correct mass recovery (98 - 102 % - mean: 101 %). Hence, deviations in the mass balances based on the selective protein signals cannot be ascribed to general inaccuracies or errors in the experimental setup. In the BSA run, the mass balance was 95 % for BSA and -7 % for HSA when based on the selective protein concentrations and 102 % when based on UV_{280}. This indicates an interaction between the two components in the applied calibration, which could not be eliminated by the performed baseline correction alone. The low recovery for HSA in the HSA run was most likely cuased by partly separated impurites present in the purchased HSA. This is caused by the fact that the previously calibrated model cannot know the spectra of partly separated impurities. Therefore, the concentrations calculated after the inpurities were separated from the main protein component would be faulty (see Figure 6 F,H). The 8 % BSA identified in this run was most likely also due to the faulty concentration determination due toimpurity separation, possibly in combination with onter interactions not eliminated completely by the baseline correction. In the BSA/HSA run, the mass balances based on the selective protein concentrations reflected the behaviour of the single component runs. The mass balance for BSA was > 100 % as HSA alone creates a netto positive signal for BSA and vice versa.

The mass balance variability of the lysA/lysH runs was lower compared to the HSA/BSA runs when based on selective concentration (84 - 102 % - mean: 89 %). However, the UV_{280} based mass balances for the lysA/lysH runs exhibited a higher variablity and a lower mean value when compared to the BSA runs. This indicated an incomplete mass recovery (92 - 102 % - mean: 94 %) in the lys runs. In the run with both lysA and lysH, the mass balances based on selective concentrations showed a higher recovery for lysH (102 %) than for lysA (84 %). This difference correlated with the interaction determined in the single component runs; in the lysA run 11 % lysH was detected and in the lysH run 6 % lysA was detected. The mean value of the mass balance based on selective concentrations (93 %) were very similar to the mass balance based on UV_{280}. As lysA and and lysH eluted from the column in a small volume, the concentration in the fractions was generally higher compared to the fractions with BSA and HSA. Hence, the high protein concentration in the run with both lysA and lysH possibly exeeded the range of correct extrapolation for both the univariate UV_{280} based calibrations and the spectral based calibration of the respective PLS regression model. This would then explain the lower recovery. The fact that the UV_{280} based mass balance for the run with only lysA was only 93 % might indicate some experimentral error as the absorption intensity was not considerably higher than the run with HSA/BSA or lysH.

Even though the combination lysA/lysH had a higher spectral similarity compared to BSA/HSA, the HTE chromatograms of lysA/lysH did not exhibit a lower quality. For both protein combinations, the elution profiles based on selective concentration matched the UV_{280} elution profiles very well. For the BSA/HSA combination, an even better result could have been achieved if the applied HSA had been of a higher purity. If so, the mass balances for BSA and HSA based on selective protein concentrations would be expected to have a higher accuracy. This might also show a more significant effect of the lower spectral similarity of BSA/HSA compared to lysA/lysH which in theory should generate chromatograms of higher accuracy for

HSA/BSA. However, it was clearly demonstrated that it is possible to generate selective concentration elution profiles which contained much more information than a respective univariate UV trace.

4 Conclusion and Outlook

The investigation into spectral similarities of a broad range of proteins based on hierarchical and non hierarchical cluster analysis determined five groups of proteins with similar spectral characteristics. One group was comprised of heme proteins whereas the other four groups were defined primarily by tryptophan content. By principal component analysis, a measure for spectral similarity was defined for bi- and ternary protein combinations. As previously demonstrated, selective high throughput protein quantification can be performed based on protein absorption spectra. For binary protein combinations, the precision of the quantification was shown to decrease exponentially with spectral similarity. The PCA model used as basis for spectral similarity determination was shown to have predictive power. Hence, it can be used to predict the possible precision of the method when used for quantification of proteins unknown to the PCA model. This can help estimate a priori whether a certain calibration will be successful for the purpose intended, i.e., whether the achieved precision will be adequate.

The finding that instrument device and sample container both play a crucial role for the achievable calibration precision, was tested for a combination of two monoclonal IgGs for which the PCA analysis had indicated that calibration would not be successful using MTP and UV plate reader. It was demonstrated, that even for monoclonal IgGs with a spectral difference below 1 % it was possible to achieve a coefficient of determination > 0.99 when using a double beam photometer and cuvettes as sample container.

For ternary combinations, no clear correlation was established between spectral similarity and precision, suggesting that the defined measure for spectral similarity of ternary protein combinations did not stand in direct relation to method precision. Future work will be aimed at finding a measure for spectral similarity with a clear relation to method precision also for ternary protein combinations. Further, as the ternary protein combinations also resulted in precise calibrations, further efforts will be dedicated to demonstrate their general robustness when used as analytics for HTE.

On the basis of liquid chromatography, which is a widely used unit operation for protein purification, a HTE application of spectral based selective protein quantification was successfully demonstrated. The application is of course not restricted to HTE chromatography and can be applied as high throughput analytics for practically any unit operation where selective protein quantification is required. The complexity of the quantification task in question and the required detection limits and precision will determine whether an application will be successful. However, as the demonstrated application was perpharmaceutiske fed with proteins of very high spectral similarity we expect that this will often be the case.

References

[1] A. S. Rathore, H. Winkle, Quality by design for biopharmaceuticals, Nature Biotechnology 27 (1) (2009) 24–36.

[2] R. Bhambure, K. Kumar, A. S. Rathore, High-throughput process development for biopharmaceutical drug substances., Trends in biotechnology 29 (3) (2011) 127–35.

[3] S. Oelmeier, F. Dismer, J. Hubbuch, Application of an Aqueous Two-Phase Systems High-Throughput Screening Method to Evaluate mAb HCP Separation, Biotechnology and Bioengineering 108 (1) (2011) 69–81.

[4] A. Susanto, K. Treier, E. Knieps-Grünhagen, E. von Lieres, J. Hubbuch, High Throughput Screening for The Design and Optimization of Chromatographic Processes: Automated Optimization of Chromatographic Phase Systems, Chemical Engineering & Technology 32 (1) (2009) 140–154.

[5] M. Degerman, K. Westerberg, B. Nilsson, A Model-Based Approach to Determine the Design Space of Preparative Chromatography, Chemical Engineering & Technology 32 (8) (2009) 1195–1202.

[6] M. Bensch, P. S. Wierling, E. von Lieres, J. Hubbuch, High Throughput Screening of Chromatographic Phases for Rapid Process Development, Chemical Engineering & Technology 28 (11) (2005) 1274–1284.

[7] C. Dinter, a. Schuetz, T. Blume, H. Weinmann, M. Harre, H. Neh, Automated Solubility Determination Using a Customized Robotic System and a Turbidity Probe, Journal of the Association for Laboratory Automation 10 (6) (2005) 408–411.

[8] M. Bensch, B. Selbach, J. Hubbuch, High throughput screening techniques in downstream processing: Preparation, characterization and optimization of aqueous two-phase systems, Chemical Engineering Science 62 (2007) 2011 – 2021.

[9] P. S. Wierling, R. Bogumil, E. Knieps-Grünhagen, High-Throughput Screening of Packed-Bed Chromatography Coupled With SELDI-TOF MS Analysis : Monoclonal Antibodies Versus Host Cell Protein, October 98 (2) (2007) 440–450.

[10] J. L. Coffman, J. F. Kramarczyk, B. D. Kelley, High-Throughput Screening of Chromatographic Separations: I. Method Development and Column Modeling, Biotechnology and Bioengineering 100 (4) (2008) 605–618.

[11] S. K. Hansen, E. Skibsted, A. Staby, J. Hubbuch, A label-free methodology for selective protein quantification by means of absorption measurements., Biotechnology and bioengineering 108 (11) (2011) 2661–2669.

[12] M. Wiendahl, P. Schulze Wierling, J. Nielsen, D. Fomsgaard Christensen, J. Krarup, A. Staby, J. Hubbuch, High Throughput Screening for the Design and Optimization of Chromatographic Processes Miniaturization, Automation and Parallelization of Breakthrough and Elution Studies, Chemical Engineering & Technology 31 (6) (2008) 893–903.

[13] T. Okubo, K. Kobayashi, Surface Tension of Biological Polyelectrolyte Solutions., Journal of colloid and interface science 205 (2) (1998) 433–442.

[14] M. G. Cottingham, C. D. Bain, D. J. T. Vaux, Rapid method for measurement of surface tension in multiwell plates., Laboratory investigation; a journal of technical methods and pathology 84 (4) (2004) 523–9.

In-line Analytics for Preparative Chromatographic Separations Based on Multivariate Calibration of Protein Spectra

Sigrid K. Hansen[1], Arne Staby[2], Erik Skibsted[2], Jürgen Hubbuch[1]*

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

[2] *CMC Project Planning & Management, Novo Nordisk A/S, Denmark*
** Corresponding author. Tel.: +49 721 608-2557; fax: +49 721 608-6240. E-mail address: juergen.hubbuch@kit.edu*

Presented at the ***European Congress of Applied Biotechnology 2011, Berlin***

Abstract

Online analysis of quality attributes (CQA) during biopharmaceutical processing is an essential part of the process analytical technology (PAT) concept (PAT Guidance for Industry – A Framework for Innovative Pharmaceutical Development, Manufacturing and Quality Assurance, FDA 2004 [1]). The goal is to enable management of batch variability and deviations of critical process parameters by using in-line and on-line analysis. In this way, consistent quality of the processed product can be assured and batch rejections can be avoided. Liquid chromatography is one of the most applied unit operations in purification of biopharmaceuticals. In most cases, the peak cutting criteria is the most important process parameter to be controlled during product elution. Commonly, the peak cutting decision is based on elution volume or univariate absorption measurement. Recently, a new method for selective protein quantification based on mid-UV protein absorption spectra was presented. The method applied multivariate calibration to link the information in the absorption spectra to the selective protein concentrations present in a sample. In the presented form, a monochromator-based photometer was used for acquisition of the spectral data. When, instead of a monochromator-based, a diode array-based photometer is applied, the spectral data can be acquired on a real-time scale. In theory, this enables in-line monitoring of preparative chromatography processes. Based on a model system, we investigated the possibility to perform in-line monitoring of chromatographic separations based on absorption spectra measured in the column eluate.

1 Introduction

Product purity is the most important objective in the manufacturing of protein therapeutics. Of course, process parameters such as yield and throughput are also important, however without purity there will be no product. State of the art protein purification processes are required to be based on identified product quality attributes and their detailed and documented linkage to process parameters. At the stage of manufacturing a detailed process control strategy must be in place. The developed control strategy is aimed at ensuring very high as well as highly reproducible quality based on different control elements and the knowledge of the links between process parameters and product quality. A control strategy consists of numerous elements including *process control elements* such as Procedural Controls and Process Parameter Controls and testing control elements such as In-Process Testing, Specifications, and Process Monitoring. In-process testing is defined as measurements performed during processing. If the analytical procedure on which the test is based is performed in the process stream, the term in-line analytics is used.

A common strategy to achieve high purity is the use of orthogonal purification steps, each of which is designed to robustly remove specific impurities. The robustness of a purification step is often achieved by using well defined pooling criteria such as complete 'load flow through' or complete 'eluate' collection. The draw back of such a process design is the limited application of high resolution steps as these would require more sophisticated pooling criteria.

An established purification process must capable to cope with variabilities in the material to be processed. However, this can lead to very conservative process parameters settings, in order to ensure quality. By having real time in-process testing in place, the process can react to variability such as e.g changes in column performance, buffer variability, or the amount of impurities. Recently, Kaltenbrunner et al. [6] suggested the application of a chromatographic process analyzer and based the pooling criteria on the data generated by the process analyzer. A schematic of the set up is displayed in Fig.1.

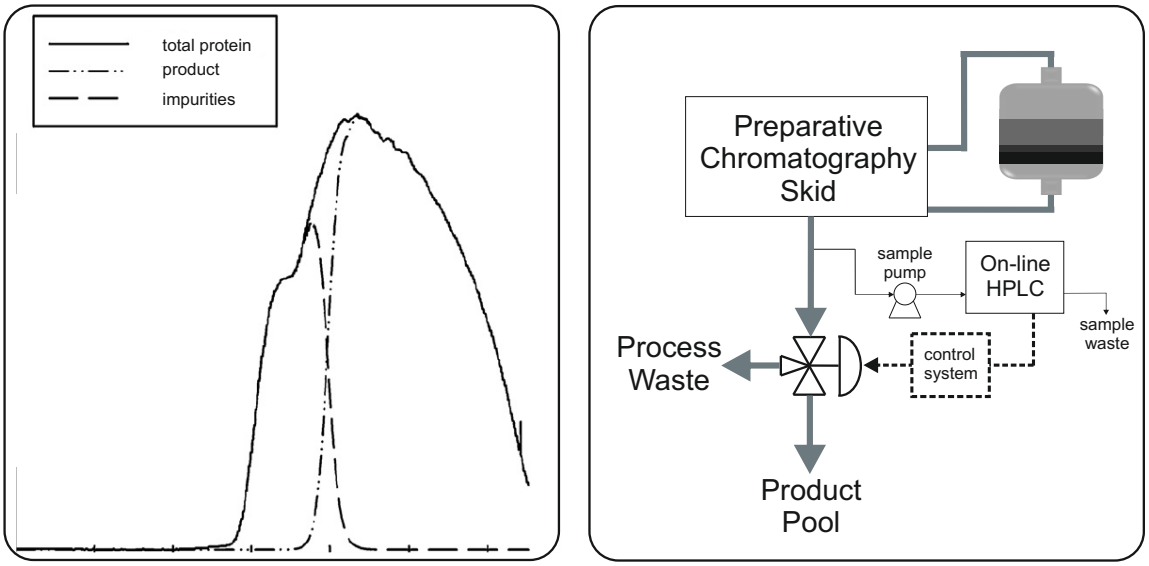


Figure 1: Fast on-line HPLC analytic for in-process control. The chromatogram shows a purification step after a refolding process where misfolded protein is separated from the target protein. The figure is adapted form the original work by Kaltenbrunner et al. [6].

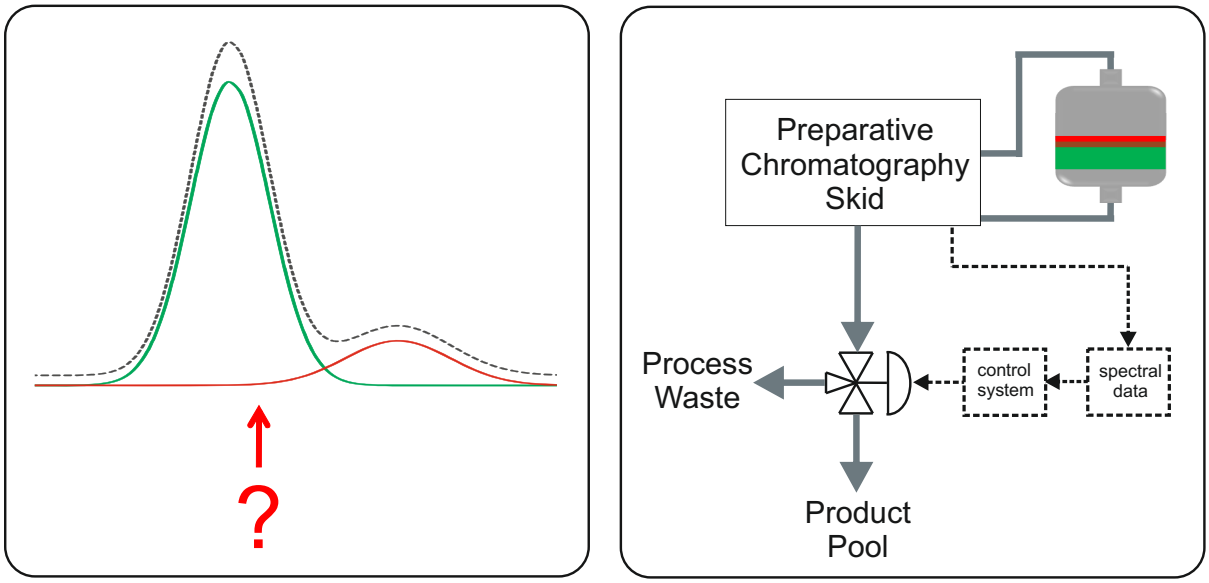


Figure 2: Proposed process control scheme based on analysis of spectral data. If the spectral of the target product and the impurity are different, process control can be based on the spectral data recorded during separation.

We have recently developed strategies to perform very fast HPLC based protein analytics [2, 4]. These methodologies have potential to be used to establish very fast analytics as needed when performing fast on-line HPLC analytic for process control. However, it is still very challenging to get a precise chromatographic analysis fast enough for real time in-process control. As a consequence, some kind of delay compensation must be build in the control system. We have recently shown that protein mid-UV absorption spectra contain information which can be robustly correlated to selective protein concentration in a protein mixture [5, 3]. The used of a diode array detector (DAD) makes it possible to perform complete wave length scans with a high frequency during a preparative chromatography step. We hence propose the that in cases where the spectra of the target protein and the impurity which is to be removes are different, the spectral data can be used as in-line process analytic, by generating selective protein signals during processing. In Fig.2 a schematic of the control system based on spectral data is displayed.

2 Methods and Materials

2.1 Chromatography System

An UltiMate3000 RSLC system from Dionex (Sunnyvale, CA, USA) was used. The system was composed of one HPG-3400RS pump, a WPS-3000TFC-analytical auto sampler and a DAD3000RS detector. The auto sampler was equipped with a 100 μL sample loop. The volume of the injection needle was 15 μL and the syringe size was 1000 μL. For control of the HPLC equipment and for data evaluation the Chromeleon software (6.80 SR10) was used.

2.2 Selective Protein Concentration Calibration

100 μL of each calibration sample, in this case pure protein, was injected directly into the diode array detector via the auto sampler. The calibration was based on spectra from each pure

protein, one spectrum created from a linear combination of both pure spectra, and one buffer spectrum. The calibration was based on PLS regression with two latent variables and mean centering was used as preprocessing.

2.3 Proteins

All proteins were purchased in solid form from Sigma-Aldrich (St. Louis, MO, USA). The applied proteins were: hen egg white lysozyme (HEWL), human breast milk lysozyme (HBML) expressed in rice, bovine α-lactalbumin (α-lac), and bovine ribonucleaseA (ribA).

2.4 SEC Runs

The SEC runs were performed using a 0.2 M potassium phosphate buffer at pH 6.2 containing 0.25 M potassium chloride. Buffers were filtered through 0.2 μm filters (Sartorius, Germany) prior to use. All protein solutions were prepared in this buffer. The chromatography runs were performed with a 15 cm ZENIX SEC-250 column (Sepax Technologies, Newark, DE, USA). The column was fitted with a 0.2 μm inlet filter (Opti-Solv, Optimize Technologis, Oregon City, OR, USA). The flow rate was 0.4 mL/min. The data collection settings were as follows: data collection rate = 1 Hz, bunch width = 1 nm, slit width = narrow.

3 Results and Discussion

As the setup of the used system did not yet allow for real time access to the spectral data, the data processing was performed after completion of each chromatographic run. The data was automatically saved in a folder. The numerical computing environment MATLAB was used to process the data. First, the calibration data were loaded and the regression model was build. The model was based on each of the pure spectra, one computed mixture spectrum, and one

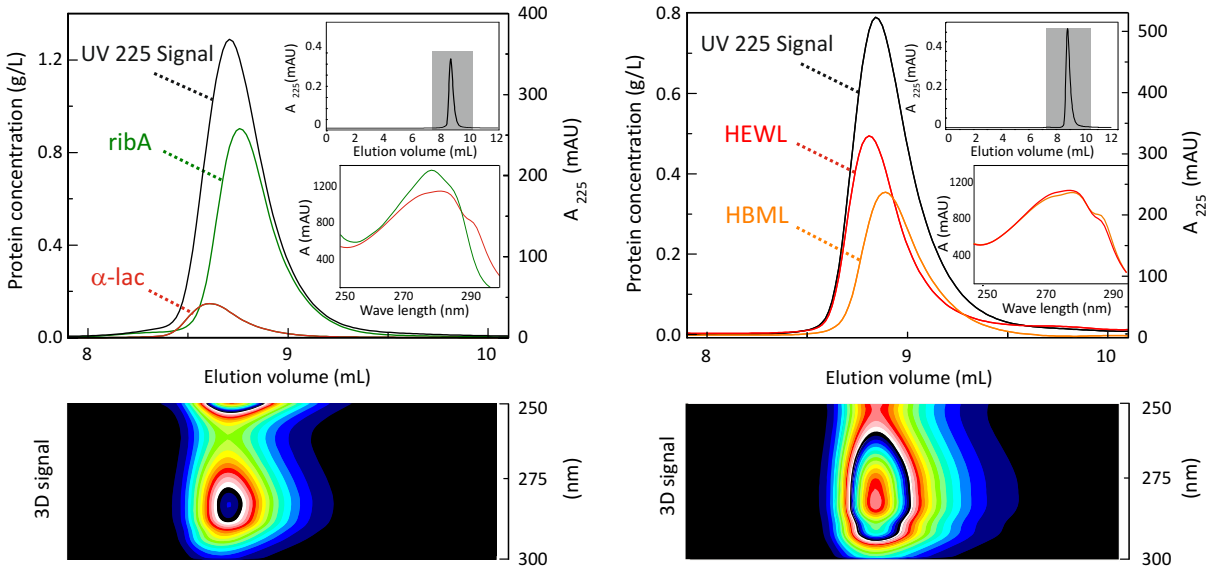


Figure 3: Selective protein signals generated on the basis of the mid-UV protein spectra. The concept was shown for two binary protein mixtures of which the spectra were similar to a different degree.

buffer spectrum. After the model was calibrated, the data form the chromatographic run was loaded and the selective protein concentration for each point in time where a spectrum was measured was calculated. This process only lasts a few seconds which makes it superior to any process analyzer based on HPLC analytics. In Fig. 3, the result of two chromatographic runs are shown. Left is the separation of ribA and α-lac which have rather different spectra. Right is the separation of HEWL and HBML which have very similar spectra. Even for the two proteins with very similar spectra, the quality of spectral data was high enough to give stable protein signals. The mass balances were in the range from 80 – 110 %. This will of cours not be acceptable if the procedure is to be used as an in-line process control tool. The reasons for the deviating mass balances is thought to be due to the pipetting of the calibration samples and protein mixtures. Further, the used model proteins does come with some impurities. If these impurities are separated form the main component during the SEC run, the calibrated model will calculate falsely concentrations.

4 Conclusion and Outlook

The presented work show that selective protein signals can be generated based spectra data recorded with a DAD. This is advantageous as it opens up for the possibility to used this for in.line process control. As it was not possible to access the spectral data in real time with the present setup of the chromatography system, this work was limited to a proof of principle based on the off-line data after a chromatographic run was performed. This work serves as the starting point for a PhD project which is planned establish the necessary requisites for real time data evaluation. Further, it will be attempted to combine the generated selective protein signals with mechanistic modeling aiming at a very high level of process control.

References

[1] PAT Guidance for Industry – A Framework for Innovative Pharmaceutical Development, Manufacturing and Quality Assurance. US Department of Health and Human Services, Food and Drug Administration (FDA), Center for Drug Evaluation and Research (CDER), Center for Veterinary Medicine (CVM), Office of Regulatory Affairs (ORA), 2004 .

[2] 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. Journal of chromatography A 1218(50):9010–9018.

[3] Hansen SK, Jamali B, Hubbuch J. 2012. Selective high throughput protein quantification based on UV absorption spectra. Biotechnology and bioengineering xxx(xxx):1–13.

[4] Hansen SK, Maiser B, Hubbuch J. 2012. Rapid quantification of protein polyethylene glycol conjugates by multivariate evaluation of chromatographic data. Journal of Chromatography A 1257:41–47.

[5] Hansen SK, Skibsted E, Staby A, Hubbuch J. 2011. A label-free methodology for selective protein quantification by means of absorption measurements. Biotechnology and bioengineering 108(11):2661–2669.

[6] Kaltenbrunner O, Lu Y, Sharma A, Lawson K, Tressel T. 2012. Risk-benefit evaluation of on-line HPLC analysis for pooling decisions in large-scale chromatography. Journal of Chromatography A 1241:37–45.

4 Conclusion & Outlook

Original methodologies to increase throughput in protein analysis were developed. In this thesis, a more efficient use of chromatography systems was shown to result in a three fold decrease in analysis time for aggregate quantification based on SEC analysis. The decrease in analysis time came from the elimination of the lag time otherwise inherent in SEC analysis. Hence, the actual information produced by the analysis was not altered. The methodology can be realized for any chromatographic analysis which is performed exclusively in isocratic mode. Further, multivariate calibration was proven capable of generating precise quantification despite low chromatographic resolution. In this context was shown that a dynamic calibration approach was able to cope with changes in analytical column performance over time which poses a problem when using a static multivariate calibration. Each of these methodologies can be used alone or in combination depending on the analytical task. In order to make these approaches readily available in settings with limited time for establishment of analytical methods or where the expertise is not at hand, it would be highly advantageous to make the presented methodologies available in the commercial chromatography software.

Original work on the use of spectral based selective protein quantification proved that mid-UV protein absorption spectra are very useful for this purpose. The analytical methods which were developed based on the mid-UV absorption spectra can be performed in micro titer plates which enables a very high throughput for selective protein quantification. The spectra of a great range of proteins were investigated, and it was shown that the approach is applicable even for proteins with very similar spectra. One of the limitations was shown to be inaccuracies related to the geometry of the liquid surface of the protein samples in the wells of a micro titer plate. Hence, future work is aimed at developing chip based UV measurement devise, where the protein samples can be measured with a defined path length. The challenge will be to maintain throughput and sensitivity of the current plated based spectrophotometers.

The potential of the use of mid-UV absorption spectra for in-line process control of preparative chromatography was shown. However, difficulties in accessing the data in real time were experienced. Future work will aim intensively at the establishment of an in-line process control method based on spectral data. It would be very interesting to combine generated selective protein signals with a mechanistic chromatography model. Such a combination might enable a accurate prediction of process yield and remaining process time. This information can be used to intensify process control and process monitoring. Further, the information could support the preparation of subsequent process steps and hence increase overall process efficiency.

The use of mid-UV spectral data is currently finding use in several new research projects. One application which has already been published is a method for determination of accurate retention time of co-eluting proteins without prior knowledge on their spectra. This method can for example enable parameter determination from protein mixtures for mechanistic modeling of chromatography. Further, the method can be used to replace electrophoresis as the second dimension in feedstock characterization.

Sigrid Krogager Hansen

Wilhelmstraße 68
76137 Karlsruhe
Deutschland

Geburtsdatum/Ort:
10/07/1978, Aarhus, Dänemark

Lebenslauf

06/2008 – jetzt	Wissentschaftlicher Mitarbeiterin am Lehrstuhl für Molekulare Aufarbeitung für Bioprodukten, Karlsruhe Institut für Technologie
	Hochdurchsatz-Analytiken für Prozessentwicklung und Prozessüberwachung im Bereich der Proteinaufreinigung
	Betreuung von Doktoranden und Studenten
05/2008	Diplom in Biotechnologie von der TU Berlin
06/2007 – 02/2008	Diplomarbeit am Forschungszentrum Jülich, Institut für Biotechnologie 2
	Purification of Monoclonal Antibodies – Development of High Throughput Methods for Process Optimization
05/2005 – 12/2006	Wissentschaftliche Hilfskraft Institut für Technische Biochemie TU Berlin
	Hauptaufgabe: Stoffflussanalyse der Sterolmetabolismus von Saccharomyces serevisiae
06/2004 – 12/2004	Studienarbeit am Institut für technische Biochemie TU Berlin
	Wachstum von* Dehalococcoides *Stamm CBDB1 mit 2,3-dichlorophenol als Elektronenakzeptor
01/2004 – 04/2004	Praktikum im Deutsches Rheuma-Forschungszentrum Berlin in AG Andreas Radbruch
	Hauptaufgaben: Aufreinigung und Kultivierung von primären CD4+ T Lymphozyten, Transfektion und Durchflusszytometrie
10/2001 – 01/2008	Studium der Biotechnologie and der TU Berlin
	Hauptfächer: chemisch-technisch Analyse, Trennverfahren, Bioprozess-Technik, bakterielle Genetik
01/1999	Umzug nach Deutschland
08/1995 – 06/1998	Gymnasium, *Aarhus Statsgymnasium*, Aarhus, Dänemark
	Hochschulreife

Sigrid Krogager Hansen

Wilhelmstraße 68
76137 Karlsruhe
Germany

Date and place of birth:
10/07/1978, Aarhus, Denmark

Curriculum Vitae

06/2008 – current	PhD student at the Chair of Biomolecular Separation Engineering, Karlsruhe Institute of Technology (KIT)
	High throughput analytics for protein purification process development and process monitoring
	Supervision of PhD and master students
05/2008	Degree in Biotechnology from TU Berlin
06/2007 – 02/2008	Diploma thesis at Research Center Jülich, Institute for Biotechnology 2
	Purification of Monoclonal Antibodies – Development of High Throughput Methods for Process Optimization
05/2005 – 12/2006	Student worker at the Institute of Bioprocess Engineering, TU Berlin
	Main work: Flux analysis of the sterol metabolism in *Saccharomyces serevisiae*
06/2004 – 12/2004	Minor thesis at the Institute of Technical Biochemistry, TU Berlin
	Growth of* Dehalococcoides *Strain CBDB1 with 2,3-dichlorophenol as Electron Acceptors
01/2004 – 04/2004	Internship at Deutsches Rheuma-Forschungszentrum Berlin in the group of Andreas Radbruch
	Main Work: Purification and cultivation of primary CD4+ T lymphycytes, cell transfection, and flow cytometry
10/2001 – 01/2008	Studies of Biotechnology
	Main subjects: chemical-technical analysis, protein separation and purification, microbial genetic engineering, microbial bioprocessing
01/1999	Relocation to Germany
08/1995 – 06/1998	Secondary school, Aarhus Statsgymnasium, Aarhus, Denmark
	University entry qualification

Scientific Publications

Oral Presentations:

S. Hansen, E. Skibsted, A. Staby, J. Hubbuch
Selective protein quantification using multivariate calibration - An analytical tool for high-throughput experimenting and inline monitoring of chromatographic separations
High-Throughput Process Development (HTPD) 2010, Krakau

S. Hansen, E. Skibsted, A. Staby, J. Hubbuch
Fast Analytical Tools Applicable for Protein Purification Process Development and Monitoring
ACS National Meeting 2012, San Diego

S. Hansen, F. Dismer, S. Oelmeier, P. Diederich, E. Skibsted, A. Staby, J. Hubbuch
Fast Protein Analytics – Multivariate Calibration Minimizes the Tradeoff between Speed and Information
Recovery of Biological Products 2012, Stowe

Posters:

S. Hansen, Jürgen Hubbuch
Evaluation of chromatographic separations for selective protein quantification based on multivariate calibration
ISPPP 2010, Bologna

S. Hansen, E. Skibsted, A. Staby, J. Hubbuch
Exploring the practical application of protein spectra for selective protein quantification
GVC/Dechema-Tagung „Bioverfahrenstechnik an Grenzflächen“ 2011, Potsdam

S. Hansen, E. Skibsted, A. Staby, J. Hubbuch
Inline Analytics for Preparative Chromatographic Separations Based on Multivariate Calibration of Protein Spectra
European Congress of Applied Biotechnology 2011, Berlin (Poster Award)

Articles:

L. Adrian, S. Hansen, J. M. Fung, H. Gorisch, S. H. Zinder
Growth of* Dehalococcoides *Strains with Chlorophenols as Electron Acceptors
Environmental Science & Technology, 41 (2007) 2318-2323

S. Hansen, E. Skibsted, A. Staby, J. Hubbuch
Label-free methodology for selective protein quantification by means of absorption measurements
Biotechnology and Bioengineering, 108 (2011) 2661-2669

P.Diederich*, S. Hansen*, S. Oelmeier*, B. Stolzenberger, J. Hubbuch
*Authors contributed equally to this work
Sub-two minutes method for monoclonal antibody-aggregate quantification using parallel-interlaced size exclusion high performance liquid chromatography.
Journal of Chromatography A, 1218 (2011) 9010-9018

A. Osberghaus, K. Drechsel, S. Hansen, S.K. Hepbildikler, S. Nath, M. Haindl, E. von Lieres, J. Hubbuch
Model-integrated process development demonstrated on the optimization of a robotic cation exchange step
Chemical Engineering Science, 76 (2012) 129-139

K. Treier, S. Hansen, C. Richter, P. Diederich, J. Hubbuch, P. Lester
High-throughput methods for miniaturization and automation of monoclonal antibody purification processes
BIOTECHNOLOGY PROGRESS, 28 (2012) 723-732

S. Hansen, B. Maiser, J. Hubbuch
Rapid Quantification of Protein-PEG Conjugates by Multivariate Evaluation of Chromatographic Data
Journal of Chromatography A, 1257 (2012) 41-47

S. Hansen, E. Skibsted, J. Hubbuch
Selective High Throughput Protein Quantification Based on UV Absorption Spectra
Biotechnology and Bioengineering, 2012, accepted

F. Dismer, S. Hansen, S. Oelmeier, J Hubbuch
Accurate retention time determination of co-eluting proteins in analytical chromatography by means of spectral data
Biotechnology and Bioengineering, 2012, accepted