

OPTIMIZATION OF CHROMATOGRAPHIC MULTI-COMPONENT SEPARATIONS IN SILICO USING HTS-DATA

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

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

genehmigte
DISSERTATION
(Tag der mündlichen Prüfung: 28.09.2012)

Dipl.-Math. Anna Zofia Osberghaus
aus Bocholt

Referent: Prof. Dr. Jürgen Hubbuch
Korreferent: Prof. Dr. Matthias Franzreb

Acknowledgments

I would like to express my deep and sincere gratitude to my supervisor, Professor Dr. Jürgen Hubbuch, Head of the Biomolecular Separation Engineering Group at the Institute of Process Engineering in Life Sciences, Karlsruhe Institute of Technology. His scientific knowledge and forthright opinions have been of great value for me. His encouraging and personal guidance have provided a good basis for the present thesis.

I am also deeply grateful to my supervisor, Dr. Eric von Lieres, Head of the Modeling and Simulation Group at the Institute of Bio- and Geosciences, Research Centre Juelich. He provided me not only with very well commented M-Files for simulation but beyond that with profound modeling knowledge, deep interest and constructive comments supporting my work.

I wish to express my warm and sincere thanks to Prof. Dr. Matthias Franzreb for being not only the second reviewer of this thesis but a helpful advisor from the beginning on. In him I had a close-by and very enthusiastic expert for chromatography modeling.

I warmly thank Dr. Susanne Nath, Dr. Hans Rogl, Dr. Markus Haindl, and Dr. Stefan Hepbildikler at Roche Diagnostics in Penzberg, who supported my work not only in financial respect but first of all with constructive comments and inquiry on relevancy of methods, data quality and optimization tools. A special thanks to Katharina Doninger and Alexander Kurtenbach working in the laboratories at Roche for being not only project partners, but colleagues and experimentators full of interest, friendliness and fantasy.

During this work I have collaborated with many colleagues for whom I have great regard, and I wish to extend my warmest thanks to all those who have helped me with my work in the Biomolecular Separation Engineering Group at the Institute of Process Engineering in Life Sciences, Karlsruhe Institute of Technology. It was not only the hours in disputation on our research topics and the role of modeling in science, that led me to climb the next step in my own research, but also the time spent together in the cinema, at a party or at the barbecue, that created an atmosphere of trust and friendship, in which I could follow my research ideas.

I owe special thanks to Dipl.-Ing. Jörg Kittelmann, who introduced Roxanne to me and supported and escorted the first contacts between a mathematician and a robotic platform with a lot of patience. It has always been pleasant and fruitful to share office and thoughts with you.

I warmly thank as well Dr.-Ing. Katrin Treier, Dr.-Ing. Florian Dimer, Dipl.-Ing. Patrick Diederich and Dr.-Ing. Stefan Oelmeier, with whom I had genial discussions on modeling and statistics. You also helped me to get a clue about 'engineer thinking' and 'communication on lab data'.

I also owe gratitude to Dr.-Ing. Arthur Susanto, who encouraged my first individual steps in Matlab when I was a young and unexperienced student apprentice. Arthur had

always a comforting and supporting answer to the oddest questions and problems.

I also wish to thank Prof. Dr. Christine Müller and Dr. Katharina Nöh, whose personalities and opinions, probably without intention and without knowing it by themselves, had a positive influence on some important decisions in the workflow of my thesis and my scientific life.

I owe my loving thanks to my husband Daniel and my daughter Katharina. They may have lost some of my time or thoughtfulness due to my research. Without their encouragement and understanding it would have been impossible for me to finish this work. My special gratitude is also due to my parents for their loving support.

Es muss das Herz bei jedem Lebensrufe
Bereit zum Abschied sein und Neubeginne,
Um sich in Tapferkeit und ohne Trauern
In andre, neue Bindungen zu geben.
Und jedem Anfang wohnt ein Zauber inne,
Der uns beschützt und der uns hilft zu leben.
(H. Hesse)

Abstract

In this thesis critical issues concerning the application of mechanistic models for the development and optimization of chromatography processes are addressed and improved. A central topic of the thesis is the evaluation of chances and challenges for simulation and optimization when data from standard high-throughput experimentation (HTE) is included into model calibration. To summarize, this approach showed to be successful and exceptionally efficient when supported by selected mathematical applications such as Design of Experiments (DoE), Monte Carlo sampling and partial least squares (PLS) regression. The thesis can be divided into two parts:

1. Examination, comparison and improvement of methods for model calibration, especially focussing on modeling protein sorption to adsorbent surfaces (Langmuir model, steric mass action (SMA) model)
2. Application of mechanistic models to in silico optimization of chromatography processes; comparison of this mechanistic approach to established empiric methods for process optimization; further improvement of the model-based methodology for process development applications

For the examination and improvement of model calibration based on static HTE data, typical high-throughput batch binding studies with lysozyme and the strong cation exchanger SP Sepharose FF were implemented in 96 well format on a robotic platform. SP Sepharose FF is commonly used in biopharmaceutic industry for applications on all scales. Based on experimental data and the Langmuir model for protein sorption, a mechanistic in silico model for the binding studies was implemented. By Monte Carlo examinations on this in silico model the impact of uncertainties in single process steps on the results of binding studies was evaluated. In addition, various case scenarios were analysed considering a broad range of uncertainties in single process steps. This approach allowed for a visualization and quantification of uncertainty propagation. To sum up, the manual production of adsorbent plaques with a scheduled volume of 20.8 μl proved to be the main source of uncertainty in high-throughput batch binding studies. Furthermore, Langmuir model calibration based on HTE batch binding studies showed to be extremely sensitive to the investigated uncertainties. Especially higher salt concentrations in the binding buffer cause very flat isotherms, hindering a precise parameter determination. Thus, further sorption models and calibration approaches had to be investigated.

In contrast to the Langmuir model, the SMA model considers salt concentrations of buffers in the calculation of sorption equilibria; a very important quality considering the simulation of salt gradients in ion exchange chromatography. Consequently, various approaches for SMA parameter determination were evaluated on three standard proteins, ribonuclease A, cytochrome *c* and lysozyme at pH 5 and pH 7. These proteins have the advantage of being accessible in high purities; in addition, on- and offline monitoring of mixtures is facilitated due to the option to monitor cytochrome *c* by complementary absorption measurements at 528 nm. Two very promising methods for SMA parameter determination were applied and compared in detail - an experimental method according to instructions in Shukla *et al.* (1998) and an approach based on an inverse method and

a mechanistic lumped-rate model for chromatography. Both methods were based on the same experimental data, five gradient elutions with gradient volumes of 5 cv (column volumes), 10 cv, 30 cv, 60 cv and 120 cv and an additional breakthrough experiment for each of the protein components at pH 5 respectively pH 7. All required experiments for SMA parameter determination were performed on prepacked 1 ml columns on a standard Äkta chromatography system. The results for SMA parameter determination and model predictivity with respect to an accurate prediction of retention times were compared for both methods and their performance and challenges in application discussed. It was shown that both methods lead to reasonable parameter sets for sorption modeling and allow for a highly precise prediction of retention times. While the model equations according to Shukla *et al.* (1998) neglect time-dependent processes and are therefore of lower complexity, the calibrated lumped-rate model has the advantage of being fit for the prediction of complete chromatograms, peak shapes and further process steps.

This led to the decision, to further use lumped-rate models for chromatography modeling and to intensify and optimize the use of inverse methods for model calibration. Thus, all relevant experiments for model calibration were repeated in miniaturized and parallelized mode on a robotic platform with prepacked RoboColumns[®] (0.2 ml bed volume) in order to examine, how HTE data might support mechanistic modeling from the outset of process development. The reduced data quality (lower density of measurement points and higher background noise) posed an important challenge in this issue. Nevertheless, by introducing elaborate methods for datapoint density enhancement and PLS-based high-throughput protein quantification on the robotic system, it was possible to calibrate the mechanistic model mainly with fractionation data from HTE chromatography. Additionally, gradient elution experiments were planned by Monte Carlo sampling in order to let them contain maximal information for the estimation of SMA parameters by inverse methods. Thus, a reduced number of four miniaturized elution experiments proved to be sufficient to determine a complete set of SMA parameters for a specific combination of protein, pH condition and adsorbent. This approach was verified with lysozyme, cytochrome *c* and ribonuclease A on four cation exchange adsorbents (SP Sepharose FF, Capto S, Toyopearl CM 650M and Toyopearl GigaCap S 650M) at three pH-conditions (pH 5, pH 6 and pH 7).

In the second part of the thesis, calibrated chromatography models were employed for the numeric optimization of a multicomponent separation step of ribonuclease A, cytochrome *c* and lysozyme on SP Sepharose FF on various scales. This separation problem is challenging and interesting due to very close isoelectric points of cytochrome *c* and lysozyme leading to peak overlaps in non-optimized separation steps. Furthermore, this system serves as a model to many separation problems, where the component of interest has to be separated from a previously and an afterwards eluting contaminant.

The established lumped-rate model showed to be highly predictive and even allowed for upscale predictions with an excellent quality. It was further demonstrated that no recalibration was necessary when optimizing separations considering changing objectives, as for example:

- objective **A**: minimal overall peak overlap
- objective **B**: mean retention time for lysozyme is 800 seconds

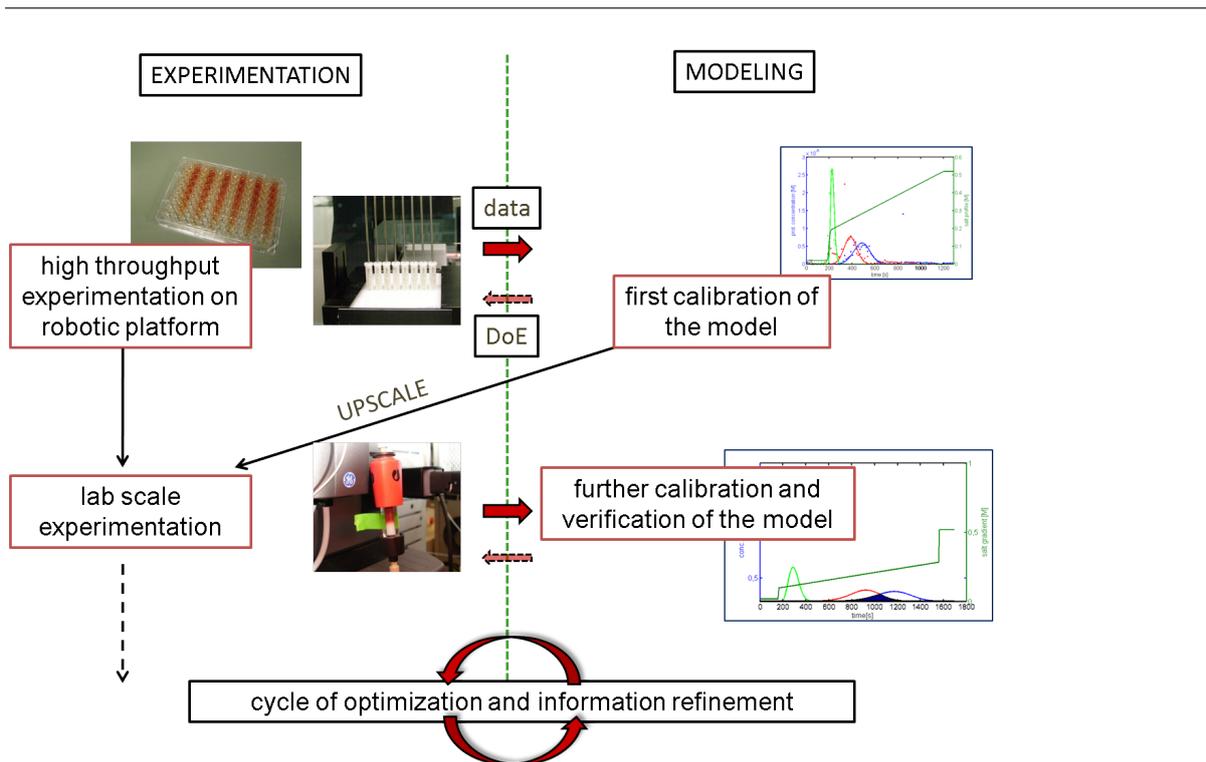


Figure 1: Concept scheme for model-integrated process development

- objective C: optimal separation between cytochrome *c* and lysozyme

The performance of the mechanistic modeling approach was further compared to the widely established approach of response surface modeling (RSM). The nonlinear elution gradient that was to be optimized was characterized by three factors describing the initial proportion of elution buffer in the running buffer, the slope of the elution gradient and the gradient length. The comparison of mechanistic and empiric approach revealed clear advantages of mechanistic modeling for process optimization. While both approaches allowed for the calculation of process robustness, the RSM method, however, failed to predict optimal separation performance within and outside the design space; only rough tendencies were predicted successfully. On the contrary, the predictions based on the established lumped-rate model were in every case precise; the optimal nonlinear gradients for the separation were determined and experimentally verified.

These findings led to the establishment of a novel concept for model-integrated process development (illustrated in figure 1). In this concept, the mechanistic model accompanies process development from the outset with initial model calibration based on HTE data. The upscale is then supported by model-based DoE allowing for experiments containing optimal information for model recalibration and refinement. This concept demonstrates a very efficient cooperation of HTE in process development and mechanistic modeling and was successfully applied to the optimization of an multicomponent separations.

Zusammenfassung

Die vorliegende Arbeit befasst sich mit wesentlichen Aspekten der Verwendung mechanistischer Modelle zum Zweck der Entwicklung und Optimierung chromatographischer Prozesse. Ein zentraler Bestandteil dieser Arbeit ist dabei die Bewertung der Chancen und Herausforderungen, die sich für die Kalibrierung solcher Modelle und den Optimierungsvorgang durch die Berücksichtigung von Daten aus Hochdurchsatz-Experimenten (HTE) ergeben. Diese Herangehensweise erwies sich vor allem dann als erfolgreich und überaus effizient, wenn sie durch mathematische Methoden wie Design of Experiments (DoE), Monte Carlo sampling und Partial Least Squares (PLS)-Regression gestützt wurde. Die Arbeit besteht aus zwei aufeinander aufbauenden Abschnitten:

1. Untersuchungen, Vergleichsstudien und Methodenentwicklung im Bereich der Modellkalibrierung mit einem besonderen Augenmerk auf die Modellierung der Sorptionsvorgänge an der Adsorberoberfläche (Langmuir Modell und Steric Mass Action (SMA) Modell)
2. Anwendung mechanistischer Modelle für die In-silico-Optimierung von Chromatographieprozessen, Vergleich dieser mechanistischen Herangehensweise mit gängigen empirischen Methoden für Prozessoptimierung, Unterstützung der modell-basierten Methoden mit HTE-Daten in Hinsicht auf den Einsatz von Modellen in der Prozessentwicklung

Für die Untersuchung und Optimierung einer Modellkalibrierung, die auf statischen HTE-Daten basieren sollte, wurden typische Hochdurchsatz-Bindungsstudien mit Lysozym und dem starken Kationentauscher SP Sepharose FF im 96-Well-Format auf einer Roboterplattform implementiert (Bindepuffer: Natriumphosphatpuffer bei pH 7). SP Sepharose FF ist ein gebräuchlicher Adsorber, der in der biopharmazeutischen Industrie in allen Prozessmaßstäben verwendet wird. Unter Berücksichtigung des Langmuir-Modells für Proteinsorption konnte ein mechanistisches 'In-silico'-Modell für die Bindungsstudien etabliert werden. Mit Hilfe von Monte Carlo sampling konnte daraufhin die Fortpflanzung von Fehlern in einzelnen Prozessschritten auf das Resultat von Bindungsstudien, die Isotherme, eingehend untersucht werden. Unter einzelnen Prozessschritten sind beispielsweise sämtliche manuelle oder robotergestützte Pipettierschritte zu verstehen aber auch die parallelisierte Herstellung von Adsorberplaques und Stammlösungen. Durch Monte Carlo sampling etlicher Fallbeispiele mit ausgewählten Fehlerverteilungen wurde eine Visualisierung und Quantifizierung der Fehlerfortpflanzung ermöglicht. So konnte die Fertigung der Adsorberplaques als hauptsächliche Fehlerquelle in Hochdurchsatz-Bindungsstudien identifiziert werden. Die HTE-basierte Kalibrierung des Langmuir-Modells erwies sich als äußerst sensitiv gegenüber den untersuchten Messpunktfehlern in den Isothermen. Vor allem Bindungsstudien mit höheren Salzkonzentrationen im Puffer ergeben sehr flache Isothermen, bei denen eine Bestimmung der Langmuir-Parameter nur eingeschränkt möglich ist. Deshalb mussten weitere Sorptionsmodelle sowie Kalibrieransätze verfolgt werden.

Im Gegensatz zum Langmuir-Modell berücksichtigt das SMA-Modell die Rolle der Salzionen bei der Berechnung von Gleichgewichtskonzentrationen auf der Adsorberoberfläche.

Dies ist eine bedeutsame Eigenschaft des Modells vor allem im Hinblick auf die Simulation von Elutionen mit Salzgradienten in der Ionenaustauschchromatographie. Deshalb wurden verschiedene Ansätze zur Bestimmung von SMA-Parametern für die drei Modellproteine Ribonuklease A, Cytochrom *c* und Lysozym bei pH 5 und pH 7 verfolgt. Diese Proteine haben den Vorteil, in größeren Mengen und hochrein zur Verfügung zu stehen; zudem ist die Konzentrationsbestimmung der einzelnen Proteine in Mischungen dadurch erleichtert, dass Cytochrom *c* neben dem für Proteine typischen Absorptionsmaximum bei 280 nm ein weiteres bei 528 nm aufweist. Zwei besonders vielversprechende Ansätze zur Bestimmung von SMA-Parametern wurden sehr detailliert untersucht und verglichen - eine Methode, die auf eine Publikation von Shukla *et al.* (1998) zurückgeht, sowie ein inverser Ansatz in Kombination mit einem mechanistischen Lumped-Rate-Modell. Beide Ansätze basierten hauptsächlich auf Chromatogrammen zu fünf Gradientenelutionen mit Gradientenvolumina von 5 cv (column volumes), 10 cv, 30 cv, 60 cv und 120 cv sowie einem zusätzlichen Durchbruchsexperiment für jedes Protein jeweils bei pH 5 und pH 7. Die benötigten Experimente zur Bestimmung der SMA-Parameter wurden mit industriell gepackten 1 ml-Säulen an einem standardisierten Äkta-LC-System durchgeführt. Die Ergebnisse der Parameterbestimmung und der Untersuchungen zur Voraussagekraft der beiden Ansätze in Hinsicht auf die präzise Voraussage von Retentionszeiten wurden verglichen und auch im Hinblick auf ihre anwenderfreundlichkeit ausgiebig diskutiert. Es konnte gezeigt werden, dass beide Ansätze eine sinnvolle Parameterbestimmung zuließen und sehr präzise Voraussagen betreffs der Retentionszeit getroffen werden konnten. Während die Methode nach Shukla *et al.* (1998) etwas anwenderfreundlicher war, da durch Vernachlässigung zeitabhängiger Prozesse nur einfache Gleichungssysteme gelöst werden mussten, hatte das kalibrierte Lumped-Rate-Modell, ein System partieller Differentialgleichungen unter anderem den Vorteil, dass komplette Chromatogramme, also neben Retentionszeiten auch Peakformen und Überschneidungen für jegliche Gradientenlänge und -steigung vorausgesagt werden konnten.

Daher wurde eine eingehendere Beschäftigung mit Lumped-Rate-Modellen für Chromatographiemodellierung unternommen, um die Anwendung inverser Methoden zur Modellkalibrierung zu untersuchen und zu optimieren. Alle relevanten Experimente zur Modellkalibrierung wurden miniaturisiert und parallelisiert auf einer Roboterplattform mit vorgepackten RoboColumns[®] (0.2 ml Bettvolumen) wiederholt, um herauszufinden, in welchem Maße HTE-Daten die mechanistische Modellierung von Beginn der Prozessentwicklung an unterstützen können. Eine wesentliche Herausforderung stellte hierbei die reduzierte Qualität der Daten dar, die durch eine geringere Datenpunktdichte und höheres Hintergrundrauschen als an der Äkta gegeben war. Durch die Einführung raffinierter Experimente zur Erhöhung der Datenpunktdichte und PLS-basierte hochdurchsatzfähige Proteinquantifizierungsmethoden, war es trotzdem möglich, das mechanistische Modell auf Grundlage von HTE-Chromatogrammen zu kalibrieren. Zudem wurden mittels statistischer Versuchsplanung und Monte Carlo sampling Gradientenelutionsexperimente bestimmt, die besonders viel Information bezüglich der Schätzung von SMA-Parametern enthielten. So war es möglich, mit der reduzierten Anzahl von vier miniaturisierten Gradientenelutionen einen kompletten Satz an SMA-Parametern für eine Protein-Adsorber-pH-Kombination zu bestimmen. Dieser Ansatz konnte für die drei Modellproteine Ribonuklease A, Cytochrom *c* und Lysozym mit vier Adsorbentien (SP Sepharose FF, Capto S, Toyopearl CM 650M und Toyopearl GigaCap S 650M) bei drei pH-Bedingungen (pH 5,

pH 6 und pH 7) verifiziert werden.

Der zweite Teil der vorliegenden Arbeit beschäftigt sich mit der Anwendung bereits kalibrierter Chromatographiemodelle für die numerische Optimierung von Dreikomponenten-Trennprozessen mit Ribonuklease A, Cytochrom *c* und Lysozym auf SP Sepharose FF in verschiedenen Prozessmaßstäben. Diese Trennung ist besonders herausfordernd und interessant, da die Ähnlichkeit der isoelektrischen Punkte von Cytochrom *c* und Lysozym bei nicht-optimalen Bedingungen zur Überlappung der Peaks führt. Zudem dient dieses System als Modellsystem für verschiedene Trennprobleme, bei denen die Zielkomponente von jeweils einer davor und einer danach eluierenden Kontaminante sauber getrennt werden soll.

Es zeigte sich, dass Lumped-Rate-Modelle eine hohe Voraussagekraft besaßen und sogar Voraussagen für den Upscale mit einer exzellenten Qualität erlaubten. Zudem konnte gezeigt werden, dass sie erfolgreich und ohne erneute Kalibrierung für die Optimierung von Trennprozessen unter verschiedenen Zielsetzungen eingesetzt werden können, wie zum Beispiel:

- Zielsetzung **A**: minimale Peaküberschneidung
- Zielsetzung **B**: die Retentionszeit für Lysozym soll im Mittel 800 Sekunden betragen
- Zielsetzung **C**: optimale Trennung zwischen Cytochrom *c* und Lysozym

Die Leistung mechanistischer Modelle konnte zudem durch einen Vergleich mit einer gängigen empirischen Methode, dem Response Surface Modeling (RSM) anhand der Optimierung des Trennprozesses über einen nichtlinearen Gradienten, der durch drei Faktoren gegeben war, überprüft werden. Die drei zu optimierenden Faktoren waren: der Anteil an Elutionspuffer im Laufpuffer zu Beginn des Gradienten, die Steigung und die Länge des Elutionsgradienten. Der Vergleich des mechanistischen mit dem empirischen Modell zeigte klare Vorteile des mechanistischen Ansatzes auf. Obwohl beide Ansätze geeignet waren, um die Robustheit des Systems im Faktorraum zu berechnen, versagte die empirische Methode völlig bei der Berechnung des optimalen Gradienten, sowohl für Faktoren innerhalb als auch außerhalb des Raumes; nur grobe Tendenzen der Trennleistung konnten vorausgesagt werden. Voraussagen hingegen, die auf einem Lumped-Rate-Modell beruhten, waren in jedem Fall präzise und der optimale Gradient konnte berechnet und experimentell verifiziert werden.

Diese Ergebnisse führten zur Etablierung eines gänzlich neuen Konzeptes für die modellintegrierte Prozessentwicklung, das in Abbildung 1 dargestellt ist. In diesem Konzept begleitet ein geeignetes mechanistisches Modell für Chromatographie die Prozessentwicklung von Beginn an und wird bereits mit den ersten Datensätzen der HTE-Chromatographie kalibriert. Diese Kalibrierung wird durch statistische Versuchsplanung (DoE) unterstützt, indem experimentelle Setups berechnet werden, die besonders viel Information im Hinblick auf die Modellkalibrierung und -verbesserung enthalten. Das vorgestellte Konzept zeigt, wie HTE-Daten besonders effektiv in die modellbasierte Prozessentwicklung eingebunden werden können und konnte in dieser Arbeit erfolgreich an der Optimierung von Multikomponenten-Trennprozessen durchgeführt werden.

Contents

1	Introduction	13
1.1	Physicochemical properties of proteins	14
1.2	Sorption processes	16
1.3	Mechanistic models in chromatography	18
1.4	Model calibration - experimental and inverse approaches	21
1.4.1	Parameter determination based on conventional methods	21
1.4.2	Parameter estimation based on an inverse method	23
1.5	Data quality in chromatography	23
1.6	Optimization of chromatographic separations	24
1.6.1	Search algorithms and empiric modeling	24
1.6.2	Model-based optimization in silico	25
1.6.3	Model-based Design of Experiments	26
1.6.4	Model-integrated process development	26
2	Research Proposal	28
3	Publications & Manuscripts	30
	Detection, quantification and propagation of uncertainty in high throughput experimentation by Monte-Carlo methods	31
	Determination of parameters for the steric mass action model - A comparison between experimental and modeling approaches	59
	Optimizing a chromatographic three component separation: A comparison of mechanistic and empiric modeling approaches	83
	Model-integrated process development demonstrated on the optimization of a robotic cation exchange step	103
4	Conclusion & Outlook	123
5	Abbreviations and Symbols	125
6	References	127
7	Akademischer Lebenslauf/Curriculum Vitae	130

1 Introduction

Chromatography (gr.: $\chi\rho\tilde{\omega}\mu\alpha$ [chroma]: colour and $\gamma\rho\tilde{\alpha}\varphi\epsilon\iota\nu$ [graphein]: to write) has now been applied for separation purposes for over 100 years. In 1903, M.S. Tswett, a russian botanist, discovered and screened adsorption-based methods for the separation of chlorophylls and carotenoids. In the publication on his findings (published in Russian and translated into English in 1989 (Berezkin, 1989)), the fascination for 'chromatography', like he called the application, is still present and reminds me of my own very first experiences and experiments with chromatography, tracking the red colour from cytochrome *c* with anticipation on its course through the chromatography column. Full of presentiment, Tswett's paper closes with the sentence : 'The empirical determination of the adsorption properties of different substances found in living organisms [...] would lead to the elaboration of certain adsorption-analytical approaches for various problems.' In fact, chromatography has been extensively developed and upgraded to important applications in high-resolution analytics as well as in preparative industrial production since Tswett's publication. However, the main qualities of the technique that Tswett examined and observed in 1903 remained unaffected in this course of development: The separation process is based on mass transfer between two phases and interactions between the components of interest and the phases. These interactions are dependent on the physical properties from both, the components and the phases, and decide on the quality of the separation.

First research on the application of chromatography for separating protein mixtures began with the development of size-exclusion and ion-exchange techniques in the 1950's, marked by the invention of absorbent phases that are based on cellulose matrices in 1954 (Sober and Peterson, 1954). Soon, further matrices were developed, the most important being the cross-linked dextrans (Sephadex) (Porath and Flodin, 1959) and matrices based on agarose (Hjertén, 1964). The first application of protein chromatography in a large scale was insuline production in the 1970's. Since then, protein chromatography has emerged step by step to a workhorse in food and biopharmaceutical industry, like it is for example illustrated in a publication from Curling on the history of chromatography (Curling, 2007). The contemporary full-scale biopharmaceutical industry based on blood plasma fractionation releases a current annual production in the 100-tons scale with a sales volume of millions of US \$ per annum (exemplary price for immunoglobulins: 60-70 \$/g, (Buchacher and Iberer, 2006, Kelley, 2007)). In the European Union, 88 recombinant biomolecules and mabs (monoclonal antibodies) have been approved by 2002, representing 36 % of all new approvals since 1995 under the centralized European drug approval system (Walsh, 2003).

While most biological products can be applied or consumed as crude extracts, biopharmaceuticals typically require exceptional purity. Thus, chromatographic purification is of singular importance, being mostly the only separation process delivering products of satisfactory purity. However, the separation process can cover up to 70 % of the total production cost. This is the main reason for intense attempts to simplify and optimize chromatography processes, mainly industrial preparative processes, aiming at the same time for higher yields, purities and productivity of the component of interest. Successful approaches to the optimization of chromatographic separations always include a

detailed consideration of the physicochemical properties of involved components as well as thorough examinations of the separation process itself and all interactions between the proteins and the adsorbent phase. Thus, necessary details on protein properties and sorption processes will be given in the next two sections. Based on these details, the application of mathematical optimization tools, particularly modeling tools, will be intensely discussed in the following two sections, focussing on the model equations and efficient model calibration. After a brief section on data quality, that is highly dependent on the employed chromatographic system, a concept for effective and reasonable application of modeling methods, considering their advantages and disadvantages as well as a general concept for 'in silico optimization' will be introduced in the last section of this introduction.

1.1 Physicochemical properties of proteins

Proteins are biomolecules composed of long chains of amino acids. Dependent on the specific sequence of amino acids, the primary structure of a protein, a secondary structure (α -helix or β -sheet) is determined by hydrogen bonds and sulfur linkings between some amino acids, like it is depicted in figure 2.

The tertiary structure of a protein is given by the steric configuration of sequences of α -helices and β -sheets; optional conglomerations of protein subunits provide quaternary structures. The high variability in the amino acid sequence in combination with protein folding and posttranslational modifications, results in multitudinous variations of protein conformations and properties. Therefore, proteins are fit to fulfill a broad spectrum of important tasks and operations in living nature - like for example building elastic fibres in muscles, serving as transmembrane transport vehicles, catalyzing chemical reactions (enzymes) or providing main functionalities in the immune system. Based on protein variability, the specific properties and operation possibilities, the production of proteins and their application in food industry, agriculture, materials science and medicine is potent and increasing.

Of course, the physicochemical properties of proteins show a wide spread due to their structural variability. These properties mainly decide on the interactions with the adsorbent and hence on the chromatography modus optimally applied for separation of a specific protein from a mixture. Table 1 sums up most of the established chromatography modi and the protein properties they are based on.

A common characteristic of all proteins is absorbance of ultraviolet light at 280 nm due to the strong absorbance of the amino acid tryptophan at this wavelength. This characteristic is most relevant for protein chromatography as it allows for a quantitative determination of protein in batch and flowthrough mode by absorption measurements. In addition, some proteins, like for example the heme protein cytochrome *c* or GFP (green fluorescent protein), absorb at wavelengths in visible range. This characteristic is an advantage in the determination of protein concentrations in multicomponent systems/mixtures, hence, these proteins are favoured model and test proteins.

The **size** of folded proteins lies in ranges between 3 to 10 nm in diameter. Regarding to the weight, most proteins lie in the range of 15 to 200 kDa (1 Da = $1/12$ m(^{12}C)) with one of the smallest proteins being insulin (5.8 kDa) up to multimeric glycoproteins (20,000 kDa). Thus, proteins are rather large biomolecules and can clearly be separated

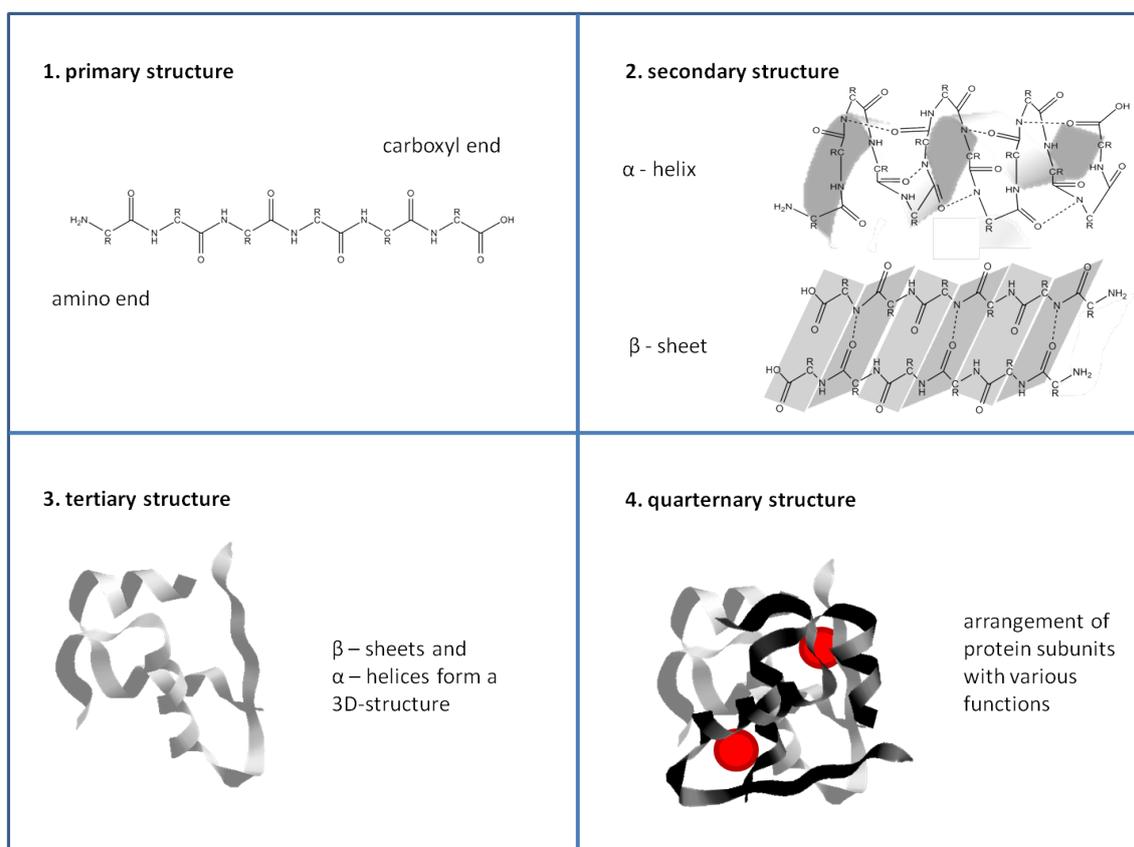


Figure 2: The reasons for wide ranges in physical properties of proteins can be traced back on different levels of structural complexity - here introduced as primary structure (sequence of amino acids), secondary structure (determined by hydrogen bonds and sulfur linkings between amino acids leading to α -helix and β -sheet conformation), tertiary structure (steric configuration of sequences of α -helices and β -sheets) and quaternary structure (arrangement of protein subunits)

from viruses and smaller molecules like sugars or amino acids by their size.

Another important physicochemical property for protein chromatography is given by the amphoteric nature of proteins. The net **charge** of a protein depends on the number of ionizable amino acid residues and their pKa-values as well as on (buffer) substances in the environment. At a particular pH, known as the isoelectric point (pI), positively and negatively charged residues are balanced and protein net charge becomes zero. Knowledge on the isoelectric points of the protein of interest and of the contaminants for example mainly decides on the optimal pH conditions for ion exchange chromatography. Despite of the fact that protein net charge in neutral surroundings can be approximately calculated by adding the pKa-values of the residues, the pH- and salt-dependent changes in protein charge are still difficult to determine, especially for the fact of possible conformation changes in the protein structure. Only recently some promising approaches have been developed based on molecular dynamic simulations, for example in (Dismer and Hubbuch, 2010).

Other properties of proteins, also critical in downstream processing, are **hydrophobicity** and **solubility**. Hydrophobicity is determined by the residues of the non-polar amino acids. The density and distribution of these residues at the surface constitute the

chromatography modus	main utilized property
size-exclusion chromatography (SEC)	protein size
affinity chromatography (AC)	conformation specificity
ion exchange chromatography (IEC)	charge
hydrophobic interaction chromatography (HIC)	hydrophobicity
reversed-phase chromatography (RPC)	polarity
countercurrent chromatography (CCC)	solubility and distribution coefficients

Table 1: Chromatography modi and the corresponding protein properties they are based on

basis for hydrophobic interaction chromatography (HIC). Thus, HIC can for example be applied for the separation of native proteins from misfolded isoforms.

The solubility of proteins can vary dramatically with pH, ionic strength and salt type. In general, it is lowest at the isoelectric point; salts at low concentrations increase the solubility (salting-in-effect) and, conversely, at high concentrations salts reduce solubility (salting-out-effect). Solubility considerations are important for separation techniques based on two liquid phases (CCC) or in crystallization, but they also limit design spaces in process design for other chromatography modi.

1.2 Sorption processes

Due to the various physicochemical properties of proteins the accumulation of protein molecules in the adsorbent phase is strongly dependent of the conditions in the mobile phase (like pH condition, ionic strength and salt type). An additional requirement for chromatography in biopharmaceutical production is mostly given by the need for complete reversibility of the sorption processes, so there always have to exist conditions leading to complete protein removal from adsorbent surface.

Considering protein sorption to the adsorbent phase, a natural analogy to other uptake kinetics is obvious. In fact, the most simple equations for sorption processes were published by I. Langmuir (Langmuir, 1916) only three years after L. Michaelis and M.L. Menten communicated their famous kinetic uptake equations for the enzyme invertin (Michaelis and Menten, 1913). Both publications describe equilibrium processes being only dependent on uptake/adsorption, release/desorption rates and overall capacity.

Provided constant pH conditions, a fixed salt concentration in the mobile phase and monolayer building on the adsorbent surface, the Langmuirian kinetic equation for the time-dependent sorption of a single protein component q is given by:

$$\frac{\partial q}{\partial t} = k_{ads} (q_{max} - q) c - k_{des} q \quad (1)$$

with q denoting the protein concentration on the adsorbent and c the concentration in the mobile phase. k_{ads} and k_{des} are the ad- and desorption coefficients and q_{max} denotes the maximal capacity of the adsorbent phase. Under the assumption of rapid equilibrium ($\frac{\partial q}{\partial t} \approx 0$), equation (1) can be reduced to the Langmuir isotherm equation:

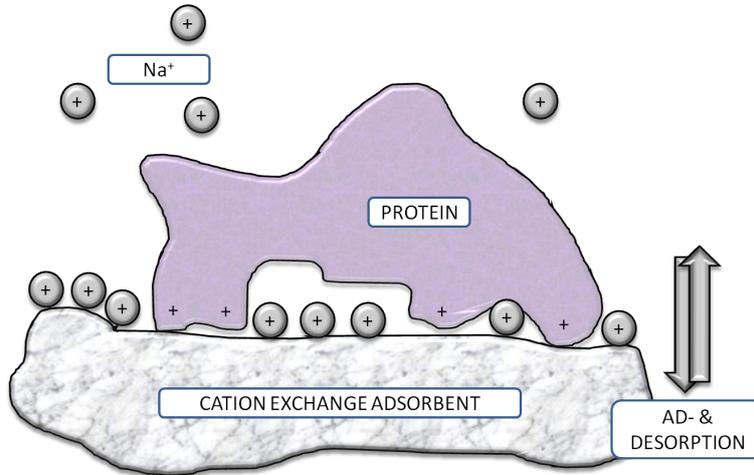


Figure 3: The basic idea of the steric mass action model by (Brooks and Cramer, 1992) is illustrated in this figure: the tridimensionality of the protein and its rivalry to the salt ions (here Na^+) in the mobile and stationary phase defines the equations of the SMA-model. In addition to ad- and desorption rates steric shielding is considered as well as the characteristic charge of the protein component.

$$q = q_{max} \frac{k_{eq}c}{1 + k_{eq}c} \quad (2)$$

with the equilibrium coefficient k_{eq} being the ratio of the ad- and desorption coefficient. Analytical chromatography processes with low protein concentrations can be located in the linear part of the curve described by equation (2) but for preparative processes the nonlinearity in the upper part of this curve has to be considered when calculating equilibria.

However, again comparable to the field of enzyme kinetics, experimentally derived isotherms show often more complex shapes that can not accurately be modeled by equation (2) (Giles *et al.*, 1974). Therefore, since Langmuir's publication in 1916, a considerable number of alternative sorption equations have been suggested in order to account for the extremely varied sorption behavior of adsorbents and proteins. A collection of alternative equations can be for example be found in (Guiochon *et al.*, 2006).

A very important example for alternative sorption kinetics equations in ion exchange chromatography is given by the steric mass action model (SMA model), derived by Brooks and Cramer in 1992 (Brooks and Cramer, 1992).

This model explicitly accounts for the fact that in IEC the equilibrium state is determined based on a competition between protein components and salt ions as well as steric effects caused by tertiary and quaternary protein structures. The concept of the SMA model is illustrated in figure 3. Summing up the competition depicted in this figure, every component interacting with the particle surface owns four characterizing parameters:

- ν (characteristic charge): mean number of binding sites of the component
- σ (steric factor): mean number of shielded/covered binding sites on adsorbent surface (due to the tertiary and quaternary structure of the protein)

- k_{ads} and k_{des} (ad- and desorption coefficient): ad- and desorption rates

Thus, the proportionality in stoichiometric exchange of protein component i and exchangeable salt counterions can be represented by:

$$c_i + \nu_i \bar{q}_1 \propto q_i + \nu_i c_1 \quad (3)$$

with c_i being the concentration of the unbound and q_i the concentration of the bound protein for $i \neq 1$. \bar{q}_1 denotes the exchangeable salt counterions and c_1 the unbound counterions. Based on this, the kinetic equations of the SMA-model for n components ($n = \text{salt} + \text{number of protein components}$) are given by:

$$\frac{\partial q_i}{\partial t} = k_{ads,i} c_i \bar{q}_1^{\nu_i} - k_{des,i} c_1^{\nu_i} q_i \quad i > 1 \quad (4)$$

$$\Lambda = q_1 + \sum_{i=2}^n \nu_i q_i \quad (5)$$

$$\bar{q}_1 = q_1 - \sum_{i=2}^n \sigma_i q_i \quad (6)$$

Equation (4) expresses the time dependent change of the concentration of surface bound component i ($\frac{\partial q_i}{\partial t}$). $k_{ads,i}$ denotes the adsorption rate and $k_{des,i}$ the desorption rate. The parameter Λ (ionic capacity of the adsorbent) limits the available binding places and displays the rivalry between salt concentration q_1 and the other bound components q_i , $2 \leq i \leq n$ with their specific characteristic charges ν_i . \bar{q}_1 , the concentration of bound salt ions available for exchange with the protein, is given by the total salt ion concentration q_1 less the shielded ions determined by the protein specific steric factors (σ_i) in equation (6). If the assumption of rapid equilibrium is valid ($\frac{\partial q_i}{\partial t} \approx 0$), equations (4), (5) and (6) can be linked to the SMA isotherm:

$$c_i = \left(\frac{q_i}{k_{eq,i}} \right) \left(\frac{c_1}{\Lambda - \sum_{i=2}^n (\nu_i + \sigma_i) q_i} \right)^{\nu_i} \quad i > 1 \quad (7)$$

with the parameters $k_{eq,i}$ denoting the ratio of ad- and desorption coefficients.

The SMA model has been successfully applied for mechanistic modeling of protein elution behaviour, as was demonstrated in (Gallant *et al.*, 1995) for step gradients, for linear gradients in (Gallant *et al.*, 1996) and for displacement systems in (Natarajan *et al.*, 2000). However, further model upgrade and the development of new concepts are urgently required, as for example pH-dependencies are relevant in IEC, but up to now, no convincing equations exist for modeling of these effects.

1.3 Mechanistic models in chromatography

The SMA model for sorption processes is in general considered as a *mechanistic model*, as physical processes are described by time-dependent equations with parameters of physical

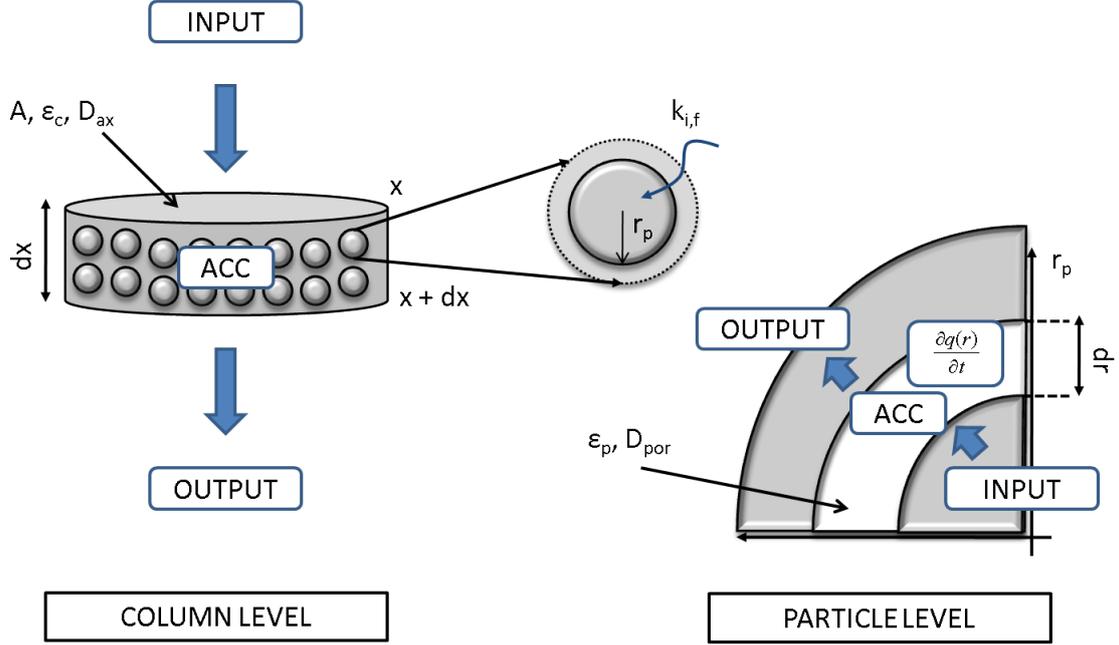


Figure 4: The basic idea of mechanistic modeling in chromatography is illustrated. Transport and mass transfer equations are derived based on physical processes on column level (left-hand side of the figure) and particle level (right-hand side of the figure) [reproduced from Schmidt-Traub (2006)]

relevance, at least to a certain degree (cmp. figure 3). However, a chromatographic separation relies not only on intra-particle sorption processes but also on physical processes on column level. The construction of a model for chromatography, appropriate for the simulation of physical processes on column and particle level and therefore predictive for the whole process, is a natural consequence of the previous considerations.

Rate equations on column level can be set up, considering the chromatography column to be a plug flow reactor, homogeneously filled with porous particles. For reasons of simplification, in most models constant temperature and pressure as well as a constant flow velocity in the column are assumed.

Consequently, the following rate equations on column level can be set up for a component i with respect to a column slice with the width dx and flowthrough area A (cmp. to the left hand side of figure 4):

$$\begin{aligned} \varepsilon_c \cdot A \left(u_{int} c_i - D_{ax,i} \frac{\partial c_i}{\partial x} \right)_x - \varepsilon_c \cdot A \left(u_{int} c_i - D_{ax,i} \frac{\partial c_i}{\partial x} \right)_{x+dx} \\ = A \cdot dx \left(\varepsilon_c \frac{\partial c_i}{\partial t} + (1 - \varepsilon_c) \frac{3}{r_p} k_{f,i} (c_i - c_{p,i})|_{r=r_p} \right) \end{aligned} \quad (8)$$

INPUT - OUTPUT = ACCUMULATION

with c_i being the concentration of component i in the mobile phase and $c_{p,i}$ the concentration in particle pores. ε_c displays the column porosity, u_{int} the convective interstitial flow velocity of the mobile phase and $D_{ax,i}$ the axial dispersion of component i , a coefficient

lumping wall effects, eddy-diffusions and other dispersive processes in the column. Mass transfer to the adsorbent particle with radius r_p is modeled by a linear passage through a film on the accumulation side of equation (8) with the film coefficient $k_{f,i}$ (compare figure 4). Dividing equation (8) by $\varepsilon_c \cdot A \cdot dx$ and letting dx approach 0, the differential equation for time-dependent concentration change of component i on column level (c_i) is given by:

$$\frac{\partial c_i}{\partial t} = -u_{int} \frac{\partial c_i}{\partial x} + D_{ax,i} \frac{\partial^2 c_i}{\partial x^2} - \frac{1 - \varepsilon_c}{\varepsilon_c} \cdot \frac{3}{r_p} k_{f,i} (c_i - c_{p,i})|_{r=r_p} \quad (9)$$

Assuming the following boundary condition at particle border

$$D_{por,i} \frac{\partial c_{p,i}(r)}{\partial t} = k_{f,i} (c_i - c_{p,i})|_{r=r_p}$$

with $D_{por,i}$ being intra-particle pore diffusion (analogously to $D_{ax,i}$ in equation (8)), the differential rate equation for the adsorbent phase is then analogously to equation (9) given by (cmp. also to the right hand side of figure 4):

$$\frac{\partial c_{p,i}(r)}{\partial t} = D_{por,i} \left(\frac{\partial^2 c_{p,i}}{\partial r^2} + \frac{2}{r} \cdot \frac{\partial c_{p,i}}{\partial r} \right) - \frac{1 - \varepsilon_p}{\varepsilon_p} \frac{\partial q_i(r)}{\partial t} \quad (10)$$

with $\frac{\partial q_i(r)}{\partial t}$ displaying the interaction of protein and particle surface - normally a sorption equation (see previous section). This system of partial differential (and algebraic, if isothermal behaviour of protein sorption is assumed) equations (PD[A]E-system) is a typical 'General Rate Model' (GRM) for chromatography modeling.

As intraparticle mass transfer is difficult to measure (details are for example given in Susanto *et al.* (2007) and Gallant (2004)) and often neglectable with respect to mass transfer resistances in the column-to-particle-transition, simplified 'lumped rate models' have been developed (for example by Bak *et al.* (Bak *et al.*, 2007) and Staby *et al.* (Staby *et al.*, 2007)) and successfully applied to simulation of protein chromatography with sufficient accuracy.

In the chromatography models, this thesis is based on, intra-particle pore diffusion and mass resistances are lumped to the parameter of *effective film-diffusion* $k_{eff,i}$. Thus, the model consists of the following equation

$$\frac{\partial c_i}{\partial t} = -u_{int} \frac{\partial c_i}{\partial x} + D_{ax} \frac{\partial^2 c_i}{\partial x^2} - \frac{1 - \varepsilon_c}{\varepsilon_c} \cdot \frac{3}{r_p} k_{eff,i} [c_i - c_{p,i}] \quad (11)$$

on column level and

$$\frac{\partial c_{p,i}}{\partial t} = \frac{3}{\varepsilon_p r_p} k_{eff,i} [c_i - c_{p,i}] - \frac{1 - \varepsilon_p}{\varepsilon_p} \frac{\partial q_i}{\partial t} \quad (12)$$

on particle level. The parameter determination for lumped rate models (model calibration) can be handled with conventional methods like batch uptake and column experiments and will be described in the next section.

system & method	bed & particle	mass transfer & sorption
system dead volume	bed length	film transfer coefficient
mobile phase flow	particle diameter	pore diffusion
number of components	axial dispersion	sorption parameters (SMA, Langmuir,..)
initial concentrations	bed porosity	
bind and elution mode	particle porosity	

Table 2: Table of parameter subgroups in mechanistic models for chromatography

1.4 Model calibration - experimental and inverse approaches

A lot of parameters determining the physical processes in chromatography have to be considered in the mechanistic modeling equations that have been derived in the previous section. They can be divided into three subgroups of parameters that are listed up in table 2.

The determination of these parameters will be discussed in this section. Typical experimental approaches and an alternative approach based on inverse modeling will be introduced, the latter being exceptionally qualified for the determination of model parameters on particle level. Further advantages and disadvantages of experimental and inverse approaches with respect to the determination of sorption parameters are intensely discussed in (Osberghaus *et al.*, 2012b).

1.4.1 Parameter determination based on conventional methods

The determination of system and method parameters (cmp. table 2) is mostly very simple and self-explanatory. **Dead volumes** of the chromatography system can be calculated by comparison of pulse experiments with and without connection to the chromatographic column. The velocity of **mobile phase flow** is normally controlled and defined by the experimenter. More appealing is the determination of the **number of components** and their **initial concentrations**.

This is obviously simple in well defined (academic) separation problems, but in industrial processes, mixtures are composed of the component of interest and an additional unknown number of components, summarized for example under the concept of *host cell protein*. Then it is convenient to sum components with similar biochemical properties up to pseudocomponents, like for example demonstrated in (Chan *et al.*, 2008). Still the analytics and determination of initial concentrations are most challenging.

The **bind and elution mode** parameters include the volumes/duration of typical process steps like injection, washes and elution as well as the elution mode itself (typical elution modes are illustrated in figure 5 on the right-hand side). Normally, these parameters are fixed for a specific process by the applied chromatography method but in screenings, for example with respect to optimal gradient settings, the parameters change from run to run. It is crucial for excellent simulation results that the complete chromatographic method is always considered and calculated in correct timelines.

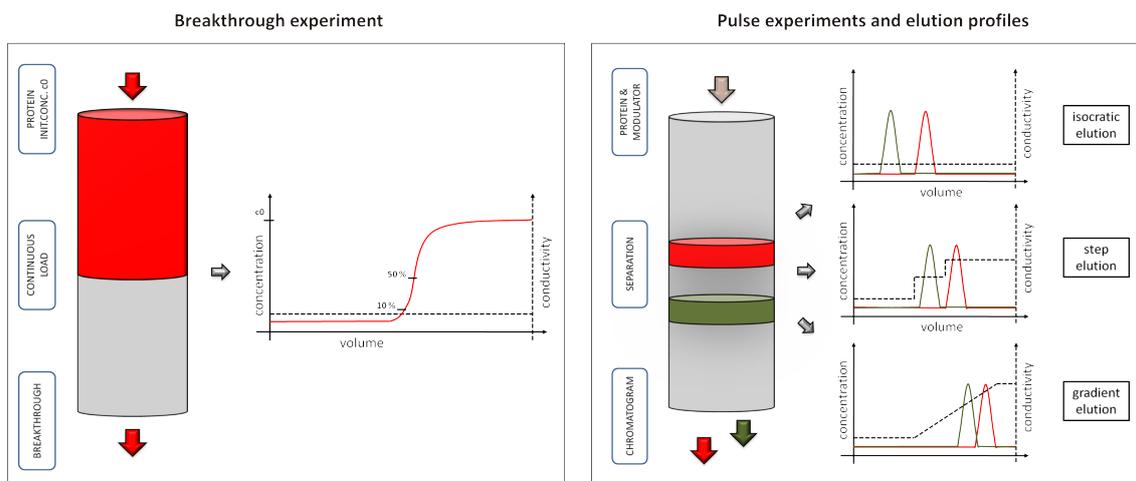


Figure 5: Common experiments for the characterization of chromatographic processes and parameter determination for mechanistic modeling of chromatography: breakthrough experiments (left-hand side) and various modi of elution experiments (right-hand side)

While **bed length** and **particle diameter** are often given by the manufacturer of a chromatography column, the other bed and particle characterizing parameters (cmp. 2nd column in table 2), **axial dispersion** and **porosities**, can be determined by breakthrough and pulse experiments with non-binding tracer substances. Typical tracer substances in ion exchange chromatography are acetone (pore-penetrating) and blue dextran (non-penetrating). Detailed instructions can be found in (Altenhoener *et al.*, 1997) and (Schmidt-Traub, 2006). The determination of **film transfer** coefficients and **pore diffusion** is very laborious, but can be performed for example by correlation equations (Schmidt-Traub, 2006), or directly based on confocal laser microscopy (Susanto *et al.*, 2007) or a recently developed method of 'peak parking' (Gritti and Guiochon, 2011).

The direct determination of **sorption model parameters** (3rd column in table 2) based on confocal laser scanning microscopy is rather seldom (an example can be found in (Teske *et al.*, 2009)); for feasibility reasons, sorption parameters are mostly determined indirectly with breakthrough and pulse experiments. There are various instructions for the determination of stationary Langmuir- and SMA-parameters in batch mode (Schmidt-Traub, 2006, Langmuir), (Barz *et al.*, 2009, SMA) or based on column experiments (Andrzejewska *et al.*, 2009, Langmuir), (Gadam *et al.*, 1993, Shukla *et al.*, 1998, SMA). For both modes, the experimental determination of sorption parameters is generally laborious and expensive due to a high protein consumption and large elution volumes, that are time- and material-consuming. Miniaturized and parallelized experimentation on robotic platforms reduces some of these disadvantages. An alternative method for the direct experimental determination of these parameters state *inverse methods*, that are introduced in the next section. The efficiency of this method and its performance is discussed in detail in the second manuscript enclosed in this thesis (Osberghaus *et al.*, 2012b).

1.4.2 Parameter estimation based on an inverse method

The parameter estimation based on an inverse method follows the concept of finding modeling parameters providing the best fit between model and experimental chromatography data. Let $c(t_j)$ be the chromatogram monitored at column outlet at the points in time $t_j = t_0 \cdots t_{end}$, preprocessed to a concentration time series. Let $\hat{c}(t_j)$ be the solution of a mechanistic model for chromatography at the same location and points in time; $\hat{c}(t_j)$ can then be compared to the chromatograms. Let now θ_{fix} be the set of all model input parameters that are fixed on a constant value and θ_{est} the set of model input parameters that can be manipulated by the algorithm solving the inverse problem (e.g. for estimating the SMA parameters $\theta_{est} = \{\nu, k_{eq}, \sigma\}$). Then the inverse problem can be stated as an minimization of a least squares residual given by:

$$res(\theta_{est}) = \sum_{j=0}^{end} (\hat{c}(t_j, \theta_{fix}; \theta_{est}) - c(t_j))^2 \quad (13)$$

The minimization of equation (13) can for example be performed with the Matlab[®] procedure `lsqnonlin`.

1.5 Data quality in chromatography

The determination of model parameters in both approaches, the conventional and the inverse method, is decisively dependent on data quality, which is in turn dependent on the available analytics and monitors. For example, flowthrough data from established LC-systems with bed volumes larger than 0.5 ml are commonly measured online and continuously. On the contrary, chromatograms from miniaturized and parallelized chromatography on robotic systems are given by offline measured discrete data from collected fractions. These conditions influence data density and information content.

A possibly most exact determination of process parameters is crucial for model quality and is the first step to a highly predictive model that can be employed for optimization and experimental planning purposes. However, the impact of data quality on model parameter determination or estimation showed to be high. Thus, special data densifying methods or methods for high throughput protein quantification had to be developed for peak retention time determination and resolution calculation in data sets from robotic platforms. Data qualities and their impact on parameter determination will be discussed mainly in the fourth enclosed manuscript (Osberghaus *et al.*, 2012a).

Figure 6 illustrates three possible data qualities that are discussed in this thesis with respect to their impact on model calibration and *in silico* optimization of separations. The left subfigure shows continuous data from a classical LC-system [Äkta]. A three-component mixture including cytochrome *c* is separated in IEC by a specific elution gradient. The absorption at 280 nm (overall protein concentration) and 528 nm (concentration of cytochrome *c*) is measured **online** and **continuously**. The center subfigure shows corresponding fractionated data from a miniaturized and parallelized mode on robotic platforms. As online monitoring is not possible in this mode, absorption at 280 nm and 528 nm is here measured **offline** and in discrete fractions. The same fractionated data evaluated with a recently developed spectral method for high-throughput protein quantification by (Hansen *et al.*, 2011) is shown in the right-hand side subfigure.

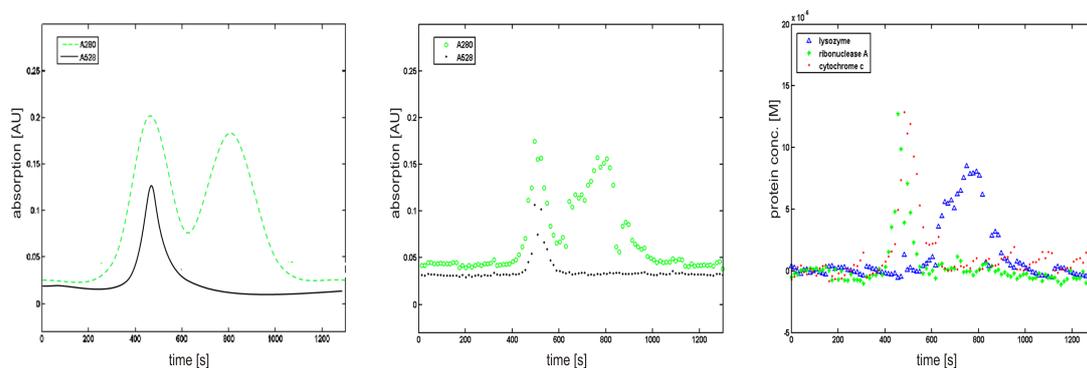


Figure 6: Comparison of chromatograms from a standard experiment monitored with different data analytics and evaluation methods. The left subfigure shows continuous data from a classical LC-system [Äkta], where absorption at 280 nm (overall protein concentration) and 528 nm (concentration of cytochrome *c*) is measured **online** and **continuously**. The centered subfigure shows a corresponding chromatogram from a miniaturized and parallelized elution on a robotic platform. Here, absorption at 280 nm and 528 nm is measured **offline** and in discrete fractions. Compared to these measurements, protein quantification in high-throughput mode by a recently developed spectral method is shown in the right-hand side subfigure.

1.6 Optimization of chromatographic separations

1.6.1 Search algorithms and empiric modeling

In most chromatographic processes, the main focus lies on an optimal resolution between the eluting peaks (analytic purposes) or the separation of a target component from a protein mixture (preparative purposes). However, the retention times and peak widths, i.e. the quality of a separation, is dependent on many process factors, among others the employed buffers and pH conditions and the time dependent change of salt concentration in the elution step. Moreover, the objective of a separation step is mostly not only defined by a high resolution, but also by yield and product purity and additional constraints, concerning process robustness, financial or ecologic issues. Thus, the optimization of a separation step is a multiparametric and multiobjective problem.

The approaches to tackle such problems are various and can be roughly divided into search algorithms and modeling methods. Successful application of search algorithms for the optimization of chromatographic processes started about 30 years ago, for example, by (Fast *et al.*, 1982), who employed a simplex algorithm for separation optimization in reversed-phase chromatography. Still search algorithms enjoy great popularity, what is demonstrated exemplarily in (Susanto *et al.*, 2009) where a genetic algorithm is applied for the optimization of bilinear elution gradients in chromatography on robotic platforms. While the low mathematical effort of search methods and their sufficient performance even in noisy systems was demonstrated in many research publications, a critical drawback of these methods is given by the enormous experimental effort and the low knowledge gain about the examined system, particularly about sensitivity and robustness aspects.

However, the importance of process understanding, as well as robustness and sensitivity analyses was only recently emphasized in the Process Analytical Technology (PAT) guidelines, published by the US Food and Drug Administration (F.D.A.U.S. Department of Health and Human Services, 2004). Consequently, the application of multivariate opti-

mization approaches based on design of experiments (DoE) and response surface modeling (RSM) increased in bioseparation process development, because these methods allow for the characterization of design factor spaces and for the calculation of optimal system settings as well as robustness analyses. Similar to the application of search algorithms, first publications on the DoE-RSM approach in the field of chromatography have been released in the eighties, for example (Pullan, 1988) who optimized a separation step in reversed-phase chromatography based on a full-factorial design. Since then, DoE-RSM techniques were applied in chromatography studies with a rising level of complexity in design and modeling, for example in (Bachman and Stewart, 1989) (full factorial design), (Bergqvist and Kaufmann, 1993) (block design and partial least squares-regression) and (Nguyen *et al.*, 2010) (partial factorial design, application of modeling software). Comparison studies of regression algorithms (Bylund *et al.*, 1997), reviews on DoE-RSM methods, like in (Ferreira *et al.*, 2007) or the formulation of very specific regression functions like in (Lebrun *et al.*, 2008) demonstrate that this method is well established in chromatography separation optimization.

1.6.2 Model-based optimization in silico

As the calculation power of computers is increasing, the application of mechanistic modeling for the optimization of chromatography processes as an alternative to RSM is recently on the rise (cmp. argumentation line in (Beckley *et al.*, 2009) and (Lieres and Andersson, 2010)). This method has important advantages compared to RSM that lead to a broad field of applications. A mechanistic model for chromatography is comprehensive and displays the physical background of complex chromatography processes (cmp. to section 1.3). Thus, it surpasses the options of an empiric model by far and owns a mechanistically founded predictivity. Once the model is calibrated for a specific system of adsorbent and components, it can be employed for the prediction of chromatograms for all kinds of bind and elution steps. Furthermore it allows for the optimization of a specific chromatography step in silico.

Let for example the objective be to optimize a specific gradient shape with regard to a minimal peak-overlap in a threecomponent separation. Then, for the in silico optimization of the separation process, the unknown parameters $\theta_{optgrad}$ describing the gradient of least overlap, are the solution of following minimization problem:

$$res_{12}(\theta_{grad}) + res_{23}(\theta_{grad}) + res_{13}(\theta_{grad}) \longrightarrow min! \quad (14)$$

with

$$res_{k,l} = \sum_{j=0}^{end} (\min(\hat{c}_k(t_j, \theta_{fix}, \theta_{grad}), \hat{c}_l(t_j, \theta_{fix}, \theta_{grad}))) \quad (15)$$

$\hat{c}_i(t_j, \theta)$ being the concentration profile/chromatogram for component i calculated by the mechanistic model, θ_{fix} are the model parameters that are fixed in the optimization (SMA parameters inclusive), θ_{grad} denotes the optimizable parameters and consists for example of the three factors **Start**, **Length** and **Slope** defining the elution gradient. A more detailed discussion and comparison of the DoE-RSM-approach and the approach of optimization in silico with mechanistic modeling is enclosed in the third paper of this thesis (Osberghaus *et al.*, 2012c).

1.6.3 Model-based Design of Experiments

Alternative to commonly used design of experiments like full-factorial designs or partial-factorial designs, which deliver regular screening patterns over a factor space and provide the information for multilinear response models, experiments with a high information content can also be planned based on the mechanistic model.

Let all controllable and uncontrollable factors that influence an experimental result be called *design factors*. For example, the three factors **Start**, **Length** and **Slope** defining the elution gradient (see above) are design factors in a chromatographic separation. Experimental planning allows for the determination of design factor sets, *experimental designs*, that lead to experiments containing a maximum of information with respect to a specific objective. Let this objective be, for example, the D-optimal determination of SMA-parameters based on an inverse method with a small estimation variance, D-optimality defined by the following definition:

Definition 1 Let $FI(\zeta)$ be the Fisher Information matrix based on the design factors $\zeta_1 \dots \zeta_n$. A defined experimental design $\zeta^* = \zeta_1^* \dots \zeta_n^*$ is called D-optimal if and only if

$$\det FI(\zeta^*) \in \max_{\zeta \in V^+} \det FI(\zeta) \quad (16)$$

with V^+ being the space of all possible experimental designs (Bandemer and Bellmann, 1994).

As the inverse of the Fisher Information matrix is equal to the covariance matrix under specific assumptions (cmp. Mardia *et al.* (1979), Fahrmeir and Hamerle (1995)), experimental designs minimizing the determinant of the covariance matrix are also called *D-optimal experimental designs*. A popular way for the calculation of covariance matrices without a demand for linearization assumptions are bootstrap methods based on Monte-Carlo sampling. Based on the mechanistic model, a huge number of in silico experiments are performed corresponding to a specific experimental design and afflicted with noise, that is characteristic for the examined chromatography process. Then, for example, the SMA-parameters are estimated based on these in silico chromatograms (cmp. equation (13)). The deviations and the covariance matrix of these estimations can be calculated and the information contents of the specific experimental designs can be compared according to equation (16). Further details on this approach are given in Efron and Tibshirani (1993) and Joshi *et al.* (2006).

1.6.4 Model-integrated process development

Based on the intensive discussion of strategies for model calibration and in silico optimization, as well as specific and target-orientated design of experiments, a new concept for model integrated process development was developed in this thesis. The concept is established by a tight cooperation between experimentation, modeling and model-based experimental design and allows for the optimization of chromatographic processes, robustness analyses and upscale predictions. It has already been successfully applied to different qualities of chromatographic data (cmp. section 1.5) in systems with model proteins and to an industrial application (separation of a monomer from aggregates and HCP). The

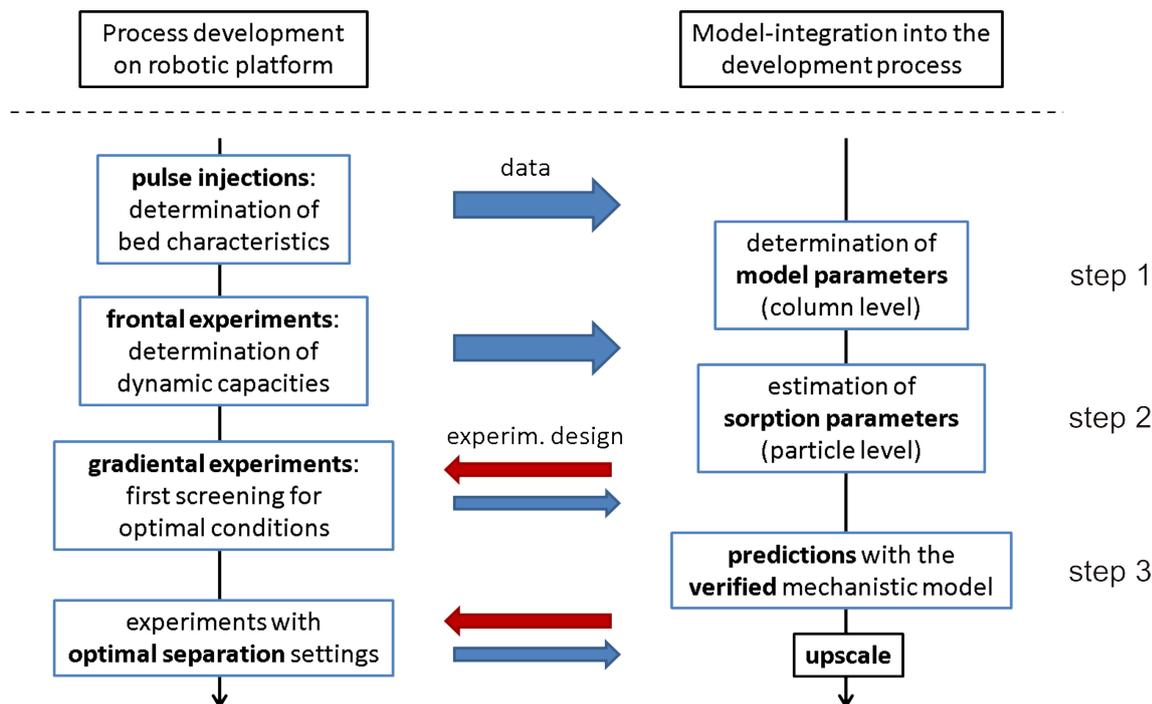


Figure 7

three step scheme for the concept of model integration in process development is shown in figure 7. In the first step of process development, pulse and frontal experiments are performed for a characterization of the column and the packed bed (porosity, dispersion, capacity). The results can be simultaneously used for the calibration of the mechanistic model on column level. In the second step of process development, frontal and gradient elution experiments support the decision on optimal elution conditions with respect to pH, adsorbent, salt type etc. These gradient elution experiments will be designed beforehand based on mechanistic modeling in order to contain increased information with respect to model calibration on particle level/determination of SMA parameters. The completely calibrated model allows for the *in silico* prediction of optimal separation gradients in the third step. Convenient objectives, like for example the resolution between the peaks or a certain retention behaviour, can be chosen as performance variables for optimization. The planned experiments are performed and model predictivity has to be verified on the experimental results by comparing peak shapes, resolutions and retention times. The verified model is then fit for upscale predictions.

The efficiency of this concept was demonstrated in (Osberghaus *et al.*, 2012a) based on data from miniaturized and parallelized chromatography. The chromatograms, planned by model-based design of experiments, contributed valuable information for process development, the model was easily calibrated and predictions showed a high precision. Thus, it was shown, that mechanistic modeling allows for *in silico* optimization based on high-throughput data, i.e. from miniaturized scale on.

2 Research Proposal

Due to its high selectivity, chromatography is a main technique in protein separation sciences. Most production processes for biopharmaceutical proteins contain at least one chromatography step, but often more. Like in most other separation techniques, a thorough selection of conditions (pH, buffers, chromatography mode) is necessary to get efficient processes. Optimal choices for these conditions can often be derived based on the properties of the component of interest (size, charge, hydrophobicity), but still the optimization of a chromatographic process poses a multivariate and multiobjective challenge. Time-pressure in industrial applications does often not allow for a careful multilevel full-factorial examination of the design space and the establishment of optimal processes. Three techniques have been developed, to improve this situation:

- high throughput screenings (HTS) in miniaturized and parallelized scale
- Design of Experiments and response surface modeling for the characterization of the design space
- mechanistic modeling in chromatography

Mostly, one of these techniques is exclusively applied to the examination and/or optimization of a specific separation problem. While high-throughput screenings and response surface modeling are quite established in chromatography process optimization as applied in industrial research groups, mechanistic modeling has still the reputation of being a complicated, expensive and laborious tool. Reasons for this reputation can probably be seen in the imperative to handle and calibrate a system of partial differential equations and furthermore in the large pool of mathematical/statistical instruments that has to be applied for model validation and optimization. In addition, the performance of mechanistic modeling is critically dependent of efficient algorithms, software and calculation power. The idea of mechanistic modeling got recently new impetus with two developments outside from chromatography science: rising calculation power in combination with algorithm efficiency and increased pressure of administrative institutions that are responsible for the approval of biopharmaceutical ingredients. These institutions recently initiated the establishment of statistical approaches considering process robustness, characterizations of design spaces and process control; modeling seems to be the best answer to these requirements. The application of empiric response surface modeling in this area soon revealed shortcomings and a limited predictivity, especially with respect to complex chromatography processes. These facts increased the interest for mechanistic modeling for the characterization and optimization of chromatography processes.

This thesis was set up to consider several issues connected with the practical application of mechanistic modeling in industrial processes with a focus on qualified model calibration and the handling of data from high-throughput screenings. The aim was to develop strategies for an efficient combination of the three previously mentioned techniques for the optimization of chromatography steps: high-throughput screening, Design of Experiments and modeling. This aim included analyses on appropriate model complexity for different chromatography scales (miniaturized scale, 1 ml lab scale and 15 ml process development scale) as well as the highly predictive up- and downscaling by employment of the mechanistic model.

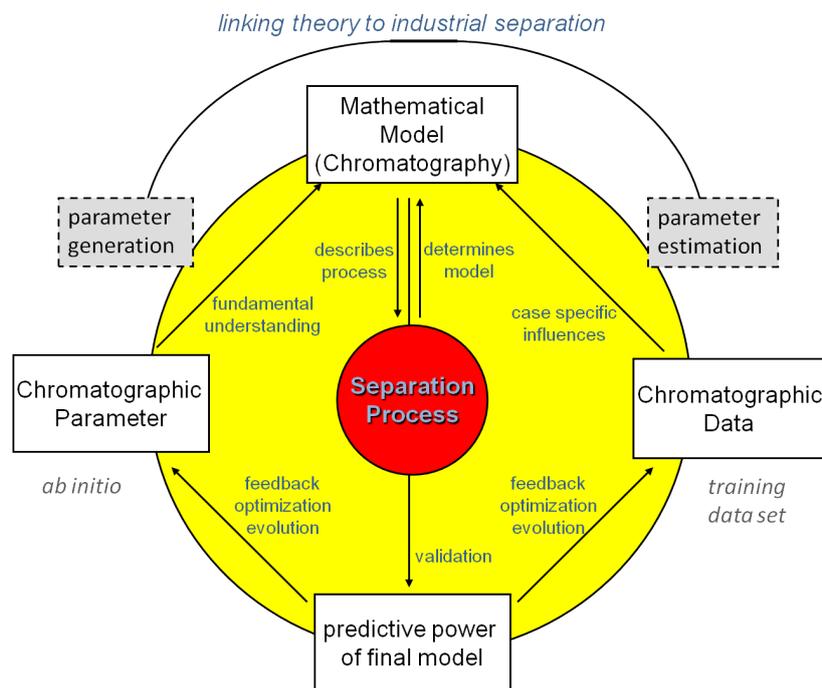


Figure 8: Concept circle illustrating the research proposal for this thesis

Experimental results from high-throughput screenings showed to be useful on the one hand for model calibration (isothermal and kinetic measurements) and on the other hand, miniaturized column chromatography provided necessary data in sufficient quality for model-based *in silico* optimization in this very early process development step. Another issue was the application of model-based Design of Experiments in order to let all experiments contain a maximum of information with respect to the specific objective (model calibration, etc). This led to the employment of modern statistical concepts like Monte-Carlo sampling methods and a model-based development of D-optimal experimental setups. A system of model proteins (ribonuclease A, cytochrome *c* and lysozyme) together with a set of most common cation exchange adsorbents served for the examination and evaluation of the developed concepts. For optimization examinations, the cation exchanger SP Sepharose FF was especially qualified, as this adsorbent material is often used in industrial purposes for low costs, however, the derivation of optimal separation conditions is quite difficult. Succeeding, the developed applications were applied to a specific industrial problem from the project partner, leading to further knowledge on systems and refining the model as well as the applications themselves.

The research proposal is illustrated in figure 8. The combination of fundamental understanding of the underlying physical processes together with experimental planning and model calibration based on designed training data leads obviously to a cycle of model optimization and refinement. Most important in the concept of the proposal is the continuous verification of the predictive power of the chromatography model on experimental results. Closing this cycle contributes new insights into the chances and challenges of the application of mechanistic modeling in chromatography.

3 Publications & Manuscripts

1. *Detection, quantification and propagation of uncertainty in high throughput experimentation by Monte-Carlo methods*

Anna Osberghaus, Pascal Baumann, Stefan Hepbildikler, Susanne Nath, Markus Haindl, Eric von Lieres, Jürgen Hubbuch

accepted by Chemical Engineering Technology,

This paper discusses the chances and challenges for the application of Monte-Carlo methods for the calculation of uncertainty propagation. A special focus is the evaluation of high-throughput processes. Thus, the methods are demonstrated based on a case study including the determination of Langmuir parameters for chromatography modeling with miniaturized batch experiments.

2. *Determination of parameters for the steric mass action model - a comparison between experimental and modeling approaches*

Anna Osberghaus, Stefan Hepbildikler, Susanne Nath, Markus Haindl, Eric von Lieres, Jürgen Hubbuch

Journal of Chromatography A, 1233 (2012) 54-65

This paper describes two methods, an experimental and an inverse approach for SMA parameter determination. Both methods are applied, compared and evaluated based on a case study with three proteins and two pH conditions.

3. *Optimizing a chromatographic three component separation: A comparison of mechanistic and empiric modeling approaches*

Anna Osberghaus, Stefan Hepbildikler, Susanne Nath, Markus Haindl, Eric von Lieres, Jürgen Hubbuch

Journal of Chromatography A (2012), doi:10.1016/j.chroma.2012.03.029

This paper investigates and compares two methods for the optimization of chromatography processes - the empiric response surface modeling and optimization based on a mechanistic model. Advantages and disadvantages of both methods are revealed and discussed in order to develop new strategies for chromatography optimization.

4. *Model-integrated process development demonstrated on the optimization of a robotic cation exchange step*

Anna Osberghaus, Katharina Drechsel, Sigrid Hansen, Stefan Hepbildikler, Susanne Nath, Markus Haindl, Eric von Lieres, Jürgen Hubbuch

accepted by Chemical Engineering Science,

This paper discusses the challenges arising with the use of HTS-data for chromatography modeling. Based on a case study including a miniaturized and parallelized

3 PUBLICATIONS & MANUSCRIPTS

three component separation on a cation exchanger, a new concept for model-based process development is introduced and demonstrated.

Detection, quantification and propagation of uncertainty in high-throughput experimentation by Monte Carlo methods

A. Osberghaus¹, P. Baumann¹, S. Hepbildikler², S. Nath², M.Haindl², E. von Lieres³
and J. Hubbuch^{1,*}

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

² : Pharmaceutical Biotech Production, Roche Diagnostics GmbH, Nonnenwald 2, 82377 Penzberg, Germany

³ : Institute of Bio- and Geosciences 1, Research Center Jülich, 52425 Jülich, Germany

* : Corresponding author. *E-mail-address*:juergen.hubbuch@kit.edu (J. Hubbuch)

ABSTRACT

Diverse bioprocess applications are established on robotic platforms in screening protocols and high-throughput experimentation. These miniaturized and parallelized applications pose new challenges in detecting and quantifying uncertainties in standard bioprocesses and in calculating uncertainty propagation influencing the process results. Since the efficiency and speed of computing has increased significantly in the last decades, 'in-silico'-approaches, for example quasi-experimental analyses based on mechanistic simulations combined with Monte Carlo methods, are on the rise for uncertainty analyses and estimation of uncertainty propagation. The power and convenience of these approaches for high-throughput processes will be demonstrated with a case study including miniaturized screenings on robotic platforms: a binding study for lysozyme on the adsorbent SP Sepharose FF in 96 well format. In this case study, all relevant uncertainties during the experimental preparations and the automated high-throughput experimentation were identified, quantified and then embedded in a simulation algorithm for the calculation of uncertainty propagation based on Monte Carlo sampling. A proof of concept for this approach is given, followed by the simulation-based analysis of various case scenarios. The high flexibility and simple handling of the Monte Carlo-based approach is convincing and can easily be transferred to uncertainty analyses in other high-throughput processes, potentially revealing bottlenecks in high-throughput analytics and laying solid foundations for process optimization.

Keywords: simulation, technology assessment, chromatography, Monte Carlo method, uncertainty propagation

1. INTRODUCTION

In experimentation for food and drug processes as well as in clinical testing, high reproducibilities and precision are particularly required. In addition, regulation agencies, for example the Food and Drug Administration in the US, expect the detection and quantification of uncertainty and analyses on robustness prior to the approval of production processes [1]. Consequently, robustness

and uncertainty analyses are not only essential for currently implemented processes, but furthermore a certain quality has to be guaranteed throughout process lifetime. In this context, knowledge of the contributions of single process steps to the overall uncertainty of the result leads to a better understanding of weaknesses in the routine and action points for uncertainty minimization.

However, new challenges arise with the revolutionary progress of high-throughput process development on robotic platforms. High-throughput screenings and experimentation are very efficient due to automation, parallelization and miniaturization techniques (cmp. [2], [3], [4] and [5]). The employed robotic platforms/workstations mostly offer an integrated concept of liquid handling including pipetting steps, transport and high-throughput analytics in 96 well format. For a comprehensive uncertainty analysis it is thus of high importance to determine uncertainty contributions from all relevant steps in the robotic workflow. Every step in the miniaturized process, such as liquid handling/pipetting, shaking, heating or centrifuging of samples, is subject to external influences (e.g. temperature or pressure variation), that afflict the scheduled volumes and concentrations with stochastic variations (errors). In addition to the robotic workflow, beforehand manual preparations of the samples and additives are often necessary and have also to be examined in the uncertainty propagation calculation as they might lead to additional uncertainty on the process results. In summary, all single steps, robotic and manual handling of the samples, have to be analysed with respect to magnitude and origin of uncertainties on the scheduled volumes or concentrations.

Even though most single process steps can be considered separately with respect to uncertainty on the step's results and their distribution, the calculation of the contribution to the total uncertainty on the final result can be still highly demanding. Classical methods of uncertainty propagation have been applied to the examination of experimental results in life sciences for a long time. Reviews on these methods and their application can for example be found in [6] and [7]. However, established equations for uncertainty propagation analysis are often not feasible, as demonstrated in [8], describing how classical equations and approaches have to be adapted and expanded with rising complexity of problems. For some problems, including for example discretely distributed uncertainties or non-analytical differential equations, the classical equations are awkward or even unusable [9].

Monte Carlo sampling, a concept of numerical in-silico experimentation based on the law of large numbers, came up in the 1970s parallelly to rising computational power and is frequently applied to problems of uncertainty propagation, e.g. in physics and chemistry (cmp. [10]). Applications of Monte Carlo sampling based uncertainty propagation in the field of bioseparation technologies can for example be found in [11], applied to Simulated Moving Bed separation or in [12], [13] and [14] or separations with High-Performance Liquid Chromatography. However, binding and separation studies or similar processes on a robotic station have to the authors knowledge never been subject to a thorough uncertainty analysis. Though interesting attempts of including deviations of process

steps to a parameter estimation based on a binding study have been published by [15], their approach was neither based on high-throughput experimentation nor on practical uncertainty analysis of the single process steps; they considered uncertainties in the estimation procedure by adding several 'uncertainty parameters'.

In addition, there is only little research on Monte Carlo based case scenarios or sensitivity analyses in the field of high-throughput processes, although for example even small deviations in the pipetted volumes of the robotic platform can have significant effects on the quality of the results when a sample has undergone multiple pipetting steps. Thus, a quantitative determination of such effects is of high interest. This paper illustrates the previously mentioned Monte Carlo based uncertainty analyses and broadens the application field by contributing new aspects of uncertainty analysis for complete processes, where a robotic platform is integrated and the workflow deviations account for the total uncertainty. The methods will be explained and demonstrated with a case study from the field of bioseparation - a binding study with lysozyme on the strong cation exchange adsorbent SP Sepharose FF. Though the proof of principle and uncertainty calculations for an system with a rather flat isotherm, might have been more beneficial from an academical point of view and perhaps less complicated in the analysis, the chosen system is a typical example for a system of industrial relevance and routine. The resulting curve, characterizing the binding behaviour of lysozyme on SP Sepharose FF, has a very steep beginning, indicating a high affinity of the protein to the adsorbent. The so-called 'isotherm' is a curve that characterizes capacity and selectivity of the examined adsorbent with respect to the applied components/proteins and is thus of high relevance in the development and optimization of chromatography separation processes.

An example for an isothermal curve is given in figure 1. In the sorption equilibrium, the concentration of protein bound to the adsorbent (q) is displayed as a function of protein concentration in the supernatant (c). In this case study, all relevant uncertainties during the experimental process, consisting of manual preparations and the high-throughput process on the robotic platform, were identified, quantified and then applied for the Monte Carlo based calculation of uncertainty propagation in a simulation algorithm. Thus, the effects of deviations of single process steps on the total result could be examined and the simulation algorithm varified. In addition, Monte Carlo sampling was employed for the examination of three case scenarios. In these scenarios, the influences of pipetting uncertainties as well as deviations in the adsorbent plaque volume and concentration of the stock solution were quantified and their effect on the total result analysed and identified.

2. THEORY

2.1. Binding studies.

2.1.1. *Equations and parameters for the characterization of protein binding to adsorbent phase.* Binding processes between a protein and the adsorbent surface can be examined by exposing various

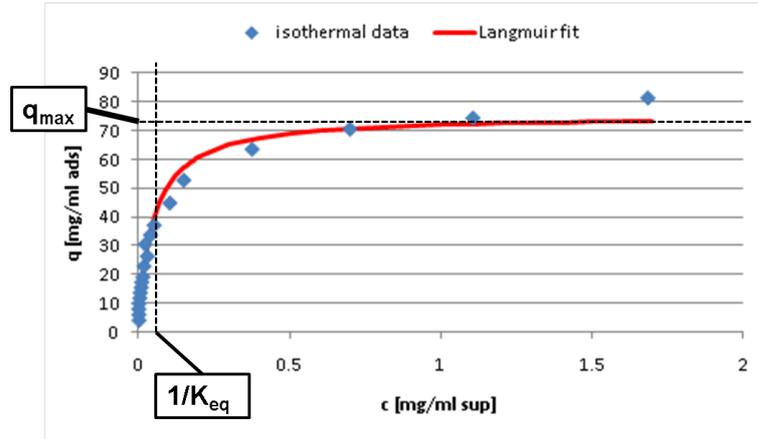


FIGURE 1. An isotherm describes the equilibrium concentrations of protein bound to the chromatographic adsorbent (q) relative to protein concentration in the supernatant (c). The curve characterizes the adsorbent selectivity (slope) and capacity (saturation point of the curve). The continuous line indicates a fit of the Langmuir isotherm equation to the measurement points.

protein quantities to a well defined volume of adsorbent followed by measurements of the equilibrium concentrations on isothermal conditions. The total mass balance for an isotherm is then given by

$$c_0 \cdot V_{sup} = c \cdot V_{sup} + q \cdot V_{ads} \quad (1)$$

V_{sup} is the supernatant volume in the experiment and V_{ads} refers here to the volume of the adsorbent particle inclusive the liquid inside the adsorbent particle, the 'adsorbent phase'. Consequently, c and q are the equilibrium protein concentrations in supernatant, respectively adsorbent phase, dependent from the adsorbent properties. For simplicity, q will be referred to as 'concentration of the bound protein' further on. Mostly, the correlation of protein concentrations at equilibrium follows a saturation curve (cmp. figure 1) which can be described or modeled for example with the Langmuir isotherm equation:

$$q = q_{max} \frac{K_{eq} \cdot c}{1 + K_{eq} \cdot c} \quad (2)$$

with q_{max} describing the saturation concentration of protein on the specific adsorbent and K_{eq} displaying an equilibrium coefficient similar to the Michaelis-constant in the Michaelis-Menten-model for enzyme catalysis [16] (cmp. to figure 1).

Thus, relevant parameters in a binding study are:

- concentrations and volumes
 - c_0 , the initial concentration of protein solution

-
- V_{sup} , the supernatant volume in the experimental setup
 - V_{ads} , the adsorbent volume in the experimental setup
 - isothermal parameters
 - q_{max} , saturation coefficient
 - K_{eq} , equilibrium coefficient

On the one hand, the two parameters q_{max} and K_{eq} can be determined by an inverse fit of equation (2) to experimental data. On the other hand, if the parameters q_{max} and K_{eq} are given, the equilibrium concentrations can be easily calculated by the solution of an equation system consisting of the mass balance equation (equation (1)) and the Langmuir isotherm equation (equation (2)). The insertion of equation (2) in equation (1) gives:

$$(V_{sup} \cdot K_{eq}) \cdot c^2 + (V_{sup} + K_{eq}q_{max}V_{ads} - K_{eq}c_0V_{sup}) \cdot c - c_0V_{sup} = 0 \quad (3)$$

leading to the unique positive solution:

$$c = \frac{-V_{sup} - K_{eq}(q_{max}V_{ads} + c_0V_{sup})}{2V_{sup}K_{eq}} + \frac{\sqrt{(V_{sup} + K_{eq}(q_{max}V_{ads} - c_0V_{sup}))^2 + 4V_{sup}^2K_{eq}c_0}}{2V_{sup}K_{eq}} \quad (4)$$

q can then be calculated with the insertion of c in equation (1) or (2).

2.1.2. *Robotic workflow.* A typical workflow of binding studies performed in high-throughput mode on robotic platforms can be divided into three main steps, the first step being manual preparations which are followed by pipetting and measurement steps on the robotic platform:

- (1) Preparations [manual]
 - preparation of adsorbent plaques with a specific volume V_{ads} stored in a certain volume of buffer V_{stor}
 - preparation of protein stock solution with the concentration c_{prot}
- (2) Pipetting & shaking
 - pipetting of specific dilutions of the protein stock solution to the adsorbent plaques with buffer volume V_{buff} and protein stock solution volume V_{prot}
 - shaking the experiment until equilibrium is reached
 - centrifuging for settling of the adsorbent in the well
 - pipetting of the supernatant to measurement plates
- (3) UV absorbance measurements at 280 nm for protein quantification in th supernatant

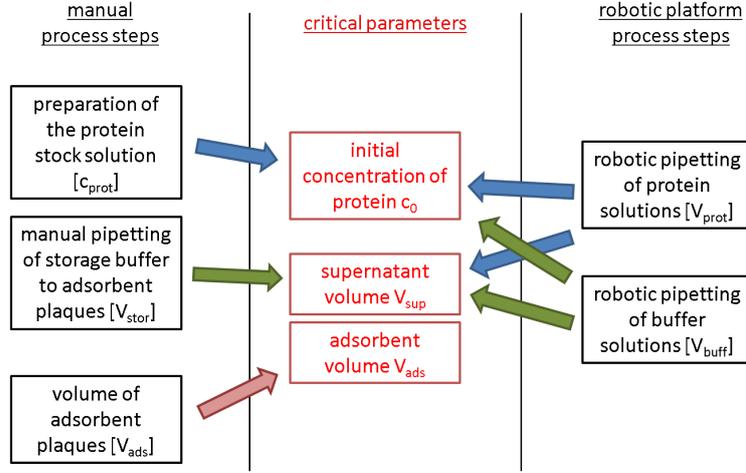


FIGURE 2. The connections between critical volumes and concentrations in the case study and the deviation-afflicted steps influencing the results of binding studies. The deviations of volumes and concentrations from scheduled values were identified and embedded into a simulation algorithm for the quantification of the contribution of single steps to total uncertainty.

Every substep should be considered with respect to its potential contribution to the total uncertainty. In this case study the contribution of shaking, centrifuging and the measurements at the end of the process are neglected, as in these substeps the deviations can be assumed to be negligibly small compared to the remaining substeps. Also the buffer solution is assumed to be standardized and of constant quality.

As the isothermal parameters q_{max} and K_{eq} characterize the specific binding behaviour between protein and adsorbent phase, the critical volumes and concentrations in this process, which are influenced by experimental deviation, are c_0 , V_{ads} and V_{sup} .

V_{sup} and c_0 are in turn given by

$$V_{sup} = V_{stor} + V_{buff} + V_{prot} \quad (5)$$

and

$$c_0 = \frac{c_{prot} V_{prot}}{V_{sup}} \quad (6)$$

Thus we can identify four basic volumes and concentrations affecting results of a binding study:

- V_{ads} , deviations in the production of adsorbent plaques
- V_{stor} , deviations in the manual pipetting of storage buffer to the adsorbent plaques
- c_{prot} , deviations in the production of protein stock solution
- V_{buff} , V_{prot} , deviations in the pipetting on the robotic station

The influences of the different steps in the workflow on the result of a binding study are also illustrated in figure 2 for a better overview.

2.2. Uncertainty propagation analyses based on Monte Carlo sampling. Monte Carlo (MC) sampling allows for in silico experimentation based on mechanistic equations describing the process. Thus, a high number of experiments can be simulated for the determination of the uncertainty contribution of single process steps on the overall result. The performance and mathematical reasoning for MC sampling are explained for example in [9], [17] and [18]. Roughly said, the MC-approach is based on the law of large numbers, high computational power and information on uncertainty distributions. In this case-study only independent stochastic deviations were considered, because systematic errors can seldom be identified in a process.

Stochastic deviations are mainly a consequence of instrument measurement limits and various stochastic influences on the experiment. Statistical location and dispersion parameters to describe these kind of influences on experimental results are the **mean** and the **standard deviation**. Let x_1, x_2, \dots, x_N be identically and independently distributed (iid) results from a stochastically influenced experiment, then the **mean** of this results is given by

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \quad (7)$$

and the standard deviation by

$$sd(x) = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x}) \quad (8)$$

In addition, the standard deviation in correspondence to the mean ($\frac{sd(x)}{\bar{x}}$) is called the **relative standard deviation**. As this parameter is dimensionless, it can easily be applied in uncertainty propagation analysis.

Provided that the mean and the standard deviation of every single process steps is known, deviation-afflicted isothermal results can be simulated based on equations (3), (4), (5) and (6). This would then be the result of a single in-silico experiment performed by the simulation algorithm. The repetition of this simulation leads to a distribution of isothermal results and by changing the deviation of a specific process step, the influence of this step and its deviation on the total result can easily be analyzed and quantified. While in this case-study normally distributed deviations were considered, because the normality could be proven by distribution tests, MC- sampling does not require normality; other specific discrete or continuous distributions of deviations may be as well considered in the simulation.

3. MATERIALS AND METHODS

3.1. Materials. Lysozyme (chicken egg white) was purchased from Sigma (St. Louis, MO, USA). Sodium monobasic phosphate, sodium dibasic phosphate and sodium hydroxide for titration were purchased from Merck KGaA (Darmstadt, Germany). The adsorbent SP Sepharose FF, a strong cation-exchanger with sulfopropyl ligands, was purchased from GE Healthcare (Buckinghamshire, United Kingdom).

3.2. Apparatus & Software. For a triple binding study on the robotic platform 96 adsorbent plaques with a volume of 20.8 μl each were produced parallelly according to the instructions in [19] with a ResiQuot[®] plaques device from Atoll (Weingarten, Germany). Figure 3 A shows the size and appearance of typical adsorbent plaques. As the 36 plaques at the border of the ResiQuot[®] device show slightly higher variances in shape and weight (cmp. [19]), they were completely left out in experimentation, thus only the 60 inner plaques in the DWP were used for the binding studies. The plaques were stored in a sealed DWP in 100 μl of 20 mM sodium phosphate buffer until usage (7 days max.).

For drying purposes a cleanroom drying oven (Mettler Ohaus GmbH + Co.KG, Schwabach, Germany) was used. For all weighing analytics a microbalance with automated readout (Mettler Toledo, Greifensee, Switzerland) was employed. The robotic platform in use was a Tecan Freedom Evo 200 workstation (Tecan, Maennedorf, Switzerland). The station is equipped with one liquid handling arm (LiHA) connected to 1 ml dilutors, a gripper, an integrated Hettich Rotanta 46RSC centrifuge (Andreas Hettich, Tuttlingen, Germany), a Variomag Teleshake horizontal lab shaker with four shaking positions (H+P Labortechnik, Oberschleissheim, Germany) and an infinite M200 Reader (Tecan, Maennedorf, Switzerland).

For the control of the robotic platform the software Evoware[®] 2.3 (Tecan, Maennedorf, Switzerland) was used. All succeeding calculations, data manipulation, simulation and visualization were accomplished in Excel[®] (Microsoft, Redmont, WA, US) and in Matlab[®] (The Mathworks, Natick, ME, USA).

3.3. high-throughput binding studies on a robotic platform. In all experiments the employed buffer was a 20 mM sodium phosphate solution at pH 7 and the scheduled concentration of lysozyme stock solution was 7 mg/ml. The manual preparations for a binding study on the robotic platform have to provide:

- a DWP with 60 adsorbent plaques, each plaque in 100 μl of storage buffer
- 14 ml of protein stock solution per isotherm
- 80 ml of buffer solution

Then, the robotic station pipettes solution series with 20 measurement points into the DWP with the adsorbent plaques according to the scheme given in table 1. Thus, it is possible to place three

buffer [μl]	protein [μl]	c_0 [mg/ml]
684	16	0.140
676	24	0.210
668	32	0.280
660	40	0.350
652	48	0.420
644	56	0.490
636	64	0.560
628	72	0.630
620	80	0.700
604	96	0.840
588	112	0.980
572	128	1.120
556	144	1.260
540	160	1.400
520	180	1.575
500	200	1.750
460	240	2.100
420	280	2.450
380	320	2.800
340	360	3.150

TABLE 1. Pipetting scheme for binding studies on robotic platforms with 20 measurement points. The binding studies are performed with various initial concentrations c_0 and a constant adsorbent plaque volume of 20.8 μl .

binding studies on a single DWP. The arrangement of three binding studies on a DWP is shown on the right hand side of figure 3. The pipetting scheme in table 1 leads, inclusively the storage buffer volume of 100 μl , to a supernatant volume V_{sup} of 800 μl per well (cmp. equation (5)). The DWP is shaken with 1500 rpm for 120 minutes to ensure equilibrium state, then it is centrifuged and the protein concentration in the supernatant is determined by absorbance measurements at 280 nm.

3.4. Detection of deviations for relevant process steps. In each of the experiments for the detection of deviations for relevant process steps, the experimental results were examined visually (histogram) and statistically (ANOVA) for normal distribution.

3.4.1. Determination of the deviation in adsorbent plaques volume. In [19] the deviation in plaque volumes was determined indirectly by the results of binding studies on robotic platforms, assuming the volume uncertainty of adsorbent plaques to be the only, or at least main source for deviations. However, the aim of this case study was an independent examination of uncertainty contributions of all relevant steps in binding studies on a robotic platform, assuming that many process steps

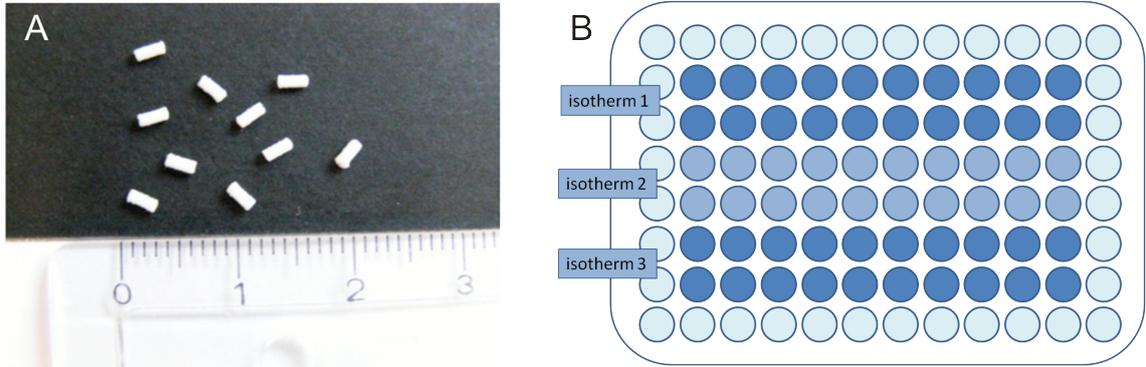


FIGURE 3. In subfigure A typical adsorbent plaques with a volume of $20.8 \mu\text{l}$ are shown to give an impression of size and form. In subfigure B a DWP-scheme for the parallel execution of three binding studies on a robotic platform is illustrated - each study is based on 20 measurement points.

contribute to the total uncertainty on binding study results. Consequently, the detection of deviations in the adsorbent plaque volumes had to be performed by a more direct method. After plaque production (cmp. section 3.2), the 60 inner plaques were released and then desiccated in an oven for 3 days at 40°C . Under the assumption of equal densities of the plaques ρ , the relative standard deviation of the oven-dry mass, determined with an analytical balance, is equal to the relative standard deviation of the wet volume:

$$\frac{\sigma(\text{mass})}{\bar{x}(\text{mass})} = \frac{\sigma(\text{mass}) \cdot \rho}{\bar{x}(\text{mass}) \cdot \rho} = \frac{\sigma(\text{vol})}{\bar{x}(\text{vol})} \quad (9)$$

The dried plaques were weighed separately and the distribution of weights as well as the standard deviation were determined. This procedure was repeated twenty times to get a larger amount of data and information on the plaque volume distribution.

3.4.2. *Determination of deviations on the manual storage buffer pipetting.* A twelve channel pipette was used to pipette a scheduled volume of $100 \mu\text{l}$ of the storage buffer to the ready plaques in the DWP. The standard deviation of manually pipetted volumes was measured by multiple pipetting into preweighed 1 ml Eppendorf cups and weighing them again. With analogous reasoning like in the previous section, the distribution and relative standard deviation of the pipetted volumes could be determined by measuring the distribution and the relative standard deviation of the pipetted masses.

3.4.3. *Determination of the deviations in the concentration of manual prepared protein stock solutions.* The standard procedure for manual stock solution preparation contains a weighing step - the protein sample is weighed in on a microbalance - and a pipetting step - the vial is filled with the adequate volume of buffer. The production of stock solutions with five volumes in the range from 0.357 ml to 50 ml and a concentration of 7 mg/ml protein was executed 20 times for each

scheduled volume, thus different protein masses were weighed in. Neglecting measurement noise from the spectrophotometer as mentioned before, it can be assumed, that the relative standard deviation of absorbance measurements of the stock solutions at 280 nm is equal to the relative standard deviation in solution concentrations.

3.4.4. *Determination of deviations in the pipetting of the robotic LiHA.* As pipetting deviations are the main uncertainty contribution of the robotic system to isothermal results, a volume-dependent relative standard deviation of pipetting steps had to be determined. The robotic tips were previously shown to be comparable and independent. Then, volumes of 10, 20, 40, 100, 200 and 800 μl buffer, respectively protein stock solution, were pipetted with the LiHA under consideration of precalibrated Liquid Classes. The pipetted range of volumes covers the pipetting scheme for binding studies (cmp. to table 1). The microbalance with automated read-out was installed on the robotic platform and the pipetting deviation was determined by pipetting 10 or more times the same volume with each robotic tip.

3.5. **Structure of the simulation algorithm.** The simulation algorithm is programmed for the mechanistic simulation of binding studies on robotic platforms. It includes Matlab[®] command lines describing the initial state, the pipetting and the calculation of the experimental results based on the assumption of a specific equilibrium behaviour described by two previously chosen Langmuir-parameters. These parts of the simulation algorithm will be explained in detail now and are also illustrated in figure 4 (cmp. also to section 2.1.1): The initial parameters for the simulation algorithm are:

- constants
 - scheduled concentration of the protein stock solution c_{prot}
 - scheduled volume of storage buffer V_{stor}
 - scheduled adsorbent volume V_{ads}
 - pipetting scheme for buffer and protein solution [cmp. table 1]
 - Langmuir parameters K_{eq} and q_{max}
- deviation descriptors
 - relative standard deviation of protein stock solution concentration $\sigma_{rel,cprot}$
 - relative standard deviation of manual pipetting of storage buffer $\sigma_{rel,stor}$
 - relative standard deviation of volume of adsorbent plaques $\sigma_{rel,ads}$
 - a functional relation between the LiHA-pipetted volume and its relative standard deviation $\sigma_{rel,pipp} = f(V_{pipp})$, which was found to be applicable for the pipetting of either buffer or protein stock solution

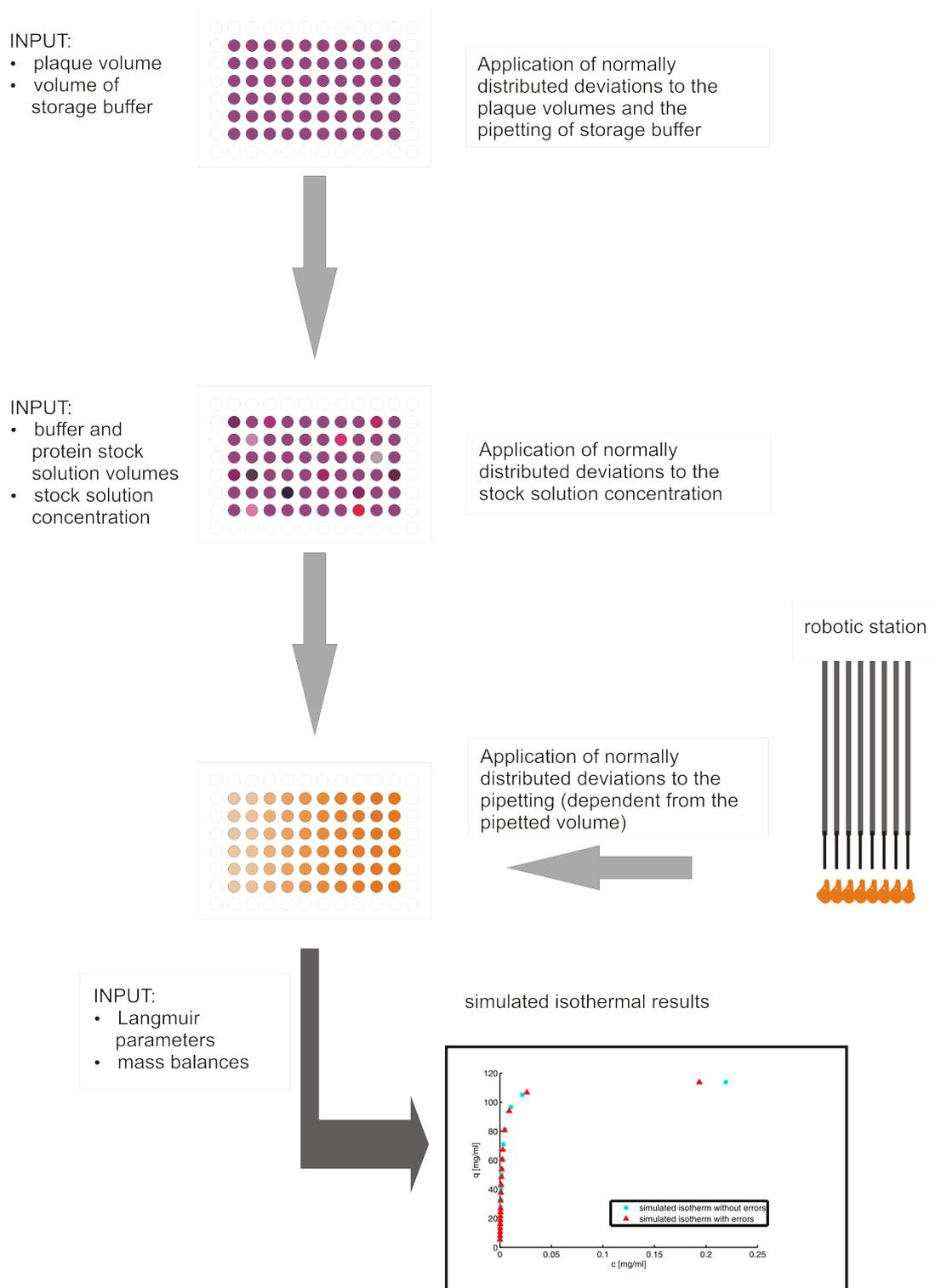


FIGURE 4. This flow scheme shows the structure for the simulation of a binding study including the application of specific relative standard deviations to the plaque volumes, to the robotic pipetting and to the concentration of the stock solution. On the left-hand side of the flow scheme, the input information and their locus apparenti is shown. On the right-hand side the application of deviations is illustrated.

In the next part of the simulation algorithm, all deviations are applied to the given constants. As the results of the experimental determination of standard deviations of the single process steps showed to be independently and normally distributed, the deviation application is given by:

$$\tilde{V}_{ads} = V_{ads} + \varepsilon_{ads} \quad (10)$$

$$\tilde{V}_{stor} = V_{stor} + \varepsilon_{stor} \quad (11)$$

$$\tilde{c}_{prot} = c_{prot} + \varepsilon_{cprot} \quad (12)$$

with ε_x independent normally distributed random errors with mean 0 and relative standard deviation $\sigma_{rel,x}$. To the pipetting scheme for buffer and protein solution, the deviations are added according to a functional relationship between the scheduled LiHA-pipetted volume and its relative standard deviation:

$$\tilde{V}_{pippp} = V_{pippp} + \varepsilon_{pippp} \quad (13)$$

$$(14)$$

V_{pippp} being here V_{buff} respectively V_{prot} . ε_{pippp} are independent and normally distributed errors with mean 0 and a relative standard deviation dependent from the scheduled pipetting volume and calculated by a functional relationship of this volume $f(V_{pippp})$, as will be shown in the results in section 4.2.4. The deviation in robotic pipetting with the LiHA is calculated for every measurement point separately. The simulation results for the binding study are subsequently calculated for each measurement point in the DWP based on the deviation-afflicted values \tilde{V}_{ads} , \tilde{V}_{stor} , \tilde{c}_{prot} which are together with the deviation-afflicted volume

$$\tilde{V}_{tot} = \tilde{V}_{stor} + \tilde{V}_{buff} + \tilde{V}_{prot} \quad (15)$$

inserted into equations (1), (2) and (4). Analogously \tilde{c}_0 is calculated (see equation (6)) and inserted.

Based on this simulation algorithm, the Monte Carlo sampling routine for analysis of deviation propagation consists of a loop, repeating the deviation-afflicted binding study simulations 10000 times and evaluating the mean and standard deviations of the results.

3.6. Proof of concept and case scenarios. For a proof of concept, the standard deviations of the equilibrium concentrations of twelve experimental binding studies were directly compared to the deviations calculated based on Monte Carlo-simulations. For a better comparison, ratios between the deviations were calculated for every relevant measurement point:

$$Q = \frac{\sigma_{rel,sim}}{\sigma_{rel,meas}} \quad (16)$$

For the case scenarios some of the deviations applied to the constants in the simulation algorithm were manipulated (increased or decreased) and the effect on the resulting simulations analysed. Three case scenarios will be introduced and analyzed in this paper:

- **What if** the deviations in protein stock solution concentration are higher or lower than normal?
- **What if** the deviations in adsorbent plaque volume are higher or lower than normal?
- **What if** the deviations in robotic pipetting are higher or lower than normal?

For the first two case scenarios a grid was constructed and Monte Carlo sampling was performed for eleven levels of plaque volume deviations in the range from 0% to 3.7% and eleven levels of deviations on the protein stock solution in a range from 0% to 2.1%. For the third case scenario constant relative standard deviations on the robotic pipetting were assumed to determine the effects of different deviation intensities on the total uncertainty on binding studies. The applied relative standard deviations in robotic pipetting covered the range of 0% to 5%. The ranges for the case scenarios were selected to include at least the experimentally derived mean and two times the standard deviations of the uncertainties from the single process steps.

4. RESULTS

4.1. Binding studies on the robotic platform. A binding study on the robotic platform (lysozyme on SP Sepharose FF) was performed twelve times according to the instructions in section 3.3. Means and standard deviations were calculated for each measurement point. The results are illustrated in figure 5 with error bars that indicate twice the relative standard deviation for the calculated equilibrium protein concentrations in the supernatant (c) and bound protein (q). Please pay attention to the different reference volumes (supernatant/adsorbent) for the concentration calculation. While the means for the concentrations of the unbound protein c_1 to c_{17} were very close to zero, because for low protein concentration all protein will bind to the adsorbent, the means and relative standard deviations on the concentration of unbound protein were rising for measurement points in the saturation part of the isotherm [$sd(c_{18})$: 12.4 %, $sd(c_{19})$: 18.38 %, $sd(c_{20})$: 24.5 %]. On the contrary, the relative standard deviations for the measurement of bound protein q lie mainly between 1.6 % and 2.1 % and show no trends. The Langmuir-parameters for these twelve isotherms were determined by inverse least squares fits with the Matlab[®]-routine `lsqnonlin` and are specified in table 2. The estimation results for Langmuir parameters show, that the estimation for q_{max} has a significantly smaller relative standard deviation [4.6 %] than the estimation for parameter K_{eq} [50.8 %]. That implies, that the standard deviations in the saturation region of the isotherm (last three measurement points in figure 5) have a very low effect on the determination of characterizing parameter q_{max} . The high deviations on K_{eq} are mainly caused by the sensitivity of K_{eq} for small changes in the first part of the isotherm, what was observed previously by [20]. The mean of the

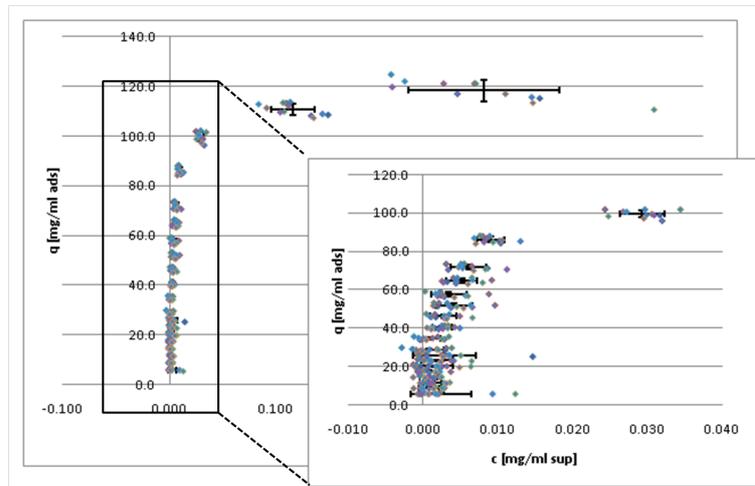


FIGURE 5. This figure shows results from twelve experimental determinations of a lysozyme isotherm on SP Sepharose FF (pH 7, 20 mM sodium phosphate buffer). c [mg/ml supernatant] denotes the concentration of overall unbound protein whereas q [mg/ml adsorbent] denotes the concentration of bound protein. The error bars indicate twice the standard deviation of the data.

experiment	q_{max} [mg/ml]	K_{eq} [ml/mg]
1	114.5	355.8
2	121.6	480.9
3	116.9	343.8
4	114.3	207.1
5	113.7	432.9
6	103.7	355.9
7	113.7	472.6
8	119.7	644.5
9	111.4	1091.3
10	108.6	252.6
11	121.0	475.9
12	119.7	824.8
\bar{x}	114.9	494.9

TABLE 2. Langmuir parameters estimated for twelve results from binding studies with lysozyme on SP Sepharose FF performed on the robotic platform. The estimations are based on an inverse fit of the Langmuir equation (cmp. (2)) to the data shown in figure 5.

parameter estimations was $\overline{K_{eq}} = 494.9$ ml/mg and $\overline{q_{max}} = 114.9$ mg/ml. These means were attributed to be the Langmuir parameters for simulation; consequently, Monte Carlo sampling is based on these parameters.

untrained user		trained user	
$\bar{x}(\text{ads})$ [g]	0.00209	$\bar{x}(\text{Ads})$ [g]	0.00217
$\sigma(\text{ads})$ [g]	0.00011	$\sigma(\text{Ads})$ [g]	0.00004
$\sigma_{rel,ads}$ [%]	5.487	$\sigma_{rel,ads}$ [%]	1.832

TABLE 3. Means and (relative) standard deviations on adsorbent plaques with a scheduled volume of $20.8 \mu\text{l}$. The values for the 'untrained user' apply for the first uses of the ResiQuot[®] device, the values for the 'trained user' apply to expert experimenters with quite constant deviations in adsorbent plaque production.

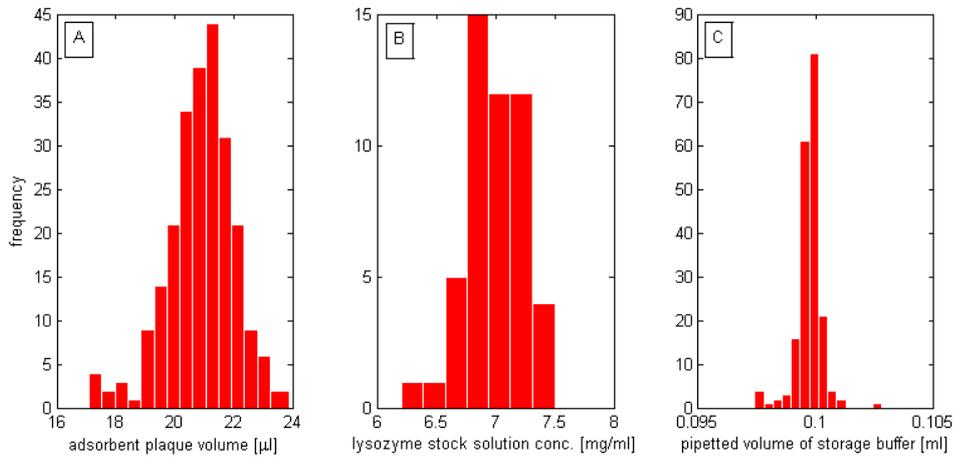


FIGURE 6. Example histograms for deviations on the volume of adsorbent plaques (subfigure A, scheduled volume: $20.8 \mu\text{l}$), deviations in the lysozyme stock solution concentration (subfigure B, scheduled concentration: 7mg/ml) and deviations in the pipetted volume of storage buffer (subfigure C, scheduled volume: 0.1 ml)

4.2. Standard deviations of the process steps.

4.2.1. *Deviations in adsorbent plaque volume.* A decrease in the relative standard deviation from adsorbent plaques volume was observed dependent of the training of the experimenter on the plaque device. Whereas the standard deviations from the very first set of plaques were high (in table 3 referred to as 'untrained user'), the plaque quality was quite constant from a specific time of plaque device use on (in table 3 referred to as 'trained user'). Obviously, by training on the device, the relative standard deviation was reduced for more than 50 %. As the binding study was performed by a 'trained user', a relative standard deviation of around 2 % could be assumed in the simulation algorithm for the deviation of adsorbent plaques volume produced with the ResiQuot[®] device. However, this observation attaches importance to the case scenarios on the effect of deviations on adsorbent plaque volume on the overall results. The distribution of adsorbent plaque volumes

\bar{x} [g]	0.099709
σ [g]	0.00056
$\sigma_{rel,stor}$ [%]	0.562

TABLE 4. Mean and (relative) standard deviations for the manual pipetting of 100 μ l of the storage buffer with a twelve-channel pipette.

weighed in protein [g]	\bar{x} (UV280)	σ (UV280)	$\sigma_{rel,cprot}$ [%]
2.5	0.8567	0.181	21.13
5	0.7373	0.0646	8.76
10	0.8726	0.0305	3.50
100	0.8401	0.0082	0.98
350	0.864	0.0036	0.42

TABLE 5. Mean and (relative) standard deviations for the production of various volumes of protein stock solution.

based on 240 plaques is shown in histogram A in figure 6. By visual analysis of the histogram and statistical analysis of the data (ANOVA) the distribution can be assumed to be normal.

4.2.2. *Deviations in manual pipetting of storage buffer.* The measured deviations in manual pipetting of storage buffer are shown in table 4. The mean (0.0997 g) converted by the buffers density of 0.998 g/ml to a volume of 0.0999 ml is very close to the scheduled volume of 100 μ l and the relative standard deviation in pipetted volumes was around 0.6 %, a rather small deviation compared for example to the deviations in adsorbent plaque volume. The distribution of manual pipetting of storage buffer based on 196 pipetted volumes is shown in histogram B in figure 6. By visual analysis of the histogram and statistical analysis of the data (ANOVA) the distribution can be assumed to be normal.

4.2.3. *Deviations in the production of protein stock solution.* The deviations in the production of stock solutions with respect to different scheduled final volumes is given in table 5 with respect to the necessary protein input for the final concentration of 7 mg/ml.

Obviously, the lower the aspired final volume is, the less protein had to be weighed in and the higher is the relative error on the final protein concentration. With a nominal protein input of 2.5 mg in the stock solution (volume of stock solution: 0.357 ml), the relative standard deviation with 21.1 % is more than twenty times higher than the relative standard deviation in stock solution production with 100 mg protein (volume of stock solution: 14.286 ml). This observation attaches importance to the case scenarios on the effect of deviations in stock solution concentration on the overall results as the amount of weighed in protein could possibly be increased for higher precision in

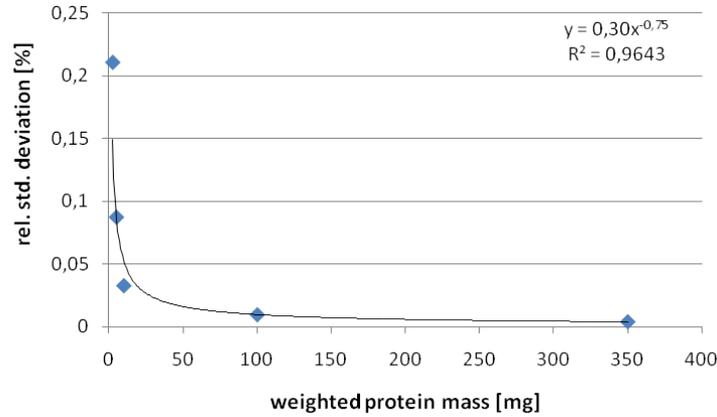


FIGURE 7. The relative standard deviation in the manual production of protein stock solutions correlates to the weighed in protein mass. The data was fitted to the power function shown in the figure.

the experimental results. The example distribution of concentrations in the production of protein stock solution with 10 mg of lysozyme based on 50 stock solutions is shown in histogram C in figure 6. By visual analysis of the histogram and statistical analysis of the data (ANOVA) the distribution could be assumed to be normal.

A power function was fitted to mathematically describe the effect of initial protein input on the relative standard deviation of the final protein concentration. The functional relation is given by

$$\sigma_{rel,cprot} = 0.3 \cdot mass^{-0.75} \quad (17)$$

with a coefficient of determination of 96.43 %. This functional relationship, shown in figure 7, could for example be employed to calculate the standard deviation in stock solution concentration in the simulation algorithm for different experimenters performing the same binding study with individual operating procedures for stock solution production. In the binding studies an overall protein stock solution volume of 50 ml was used (protein input of 350 mg), thus, a relative standard deviation of 0.42 % on the concentration of the stock solution was assumed in the simulation algorithm.

4.2.4. *Deviations in robotic pipetting.* Volumes of 10, 20, 40, 100, 200 and 800 μl were pipetted with the LiHA under consideration of specifically prepared Liquid Classes for buffer respectively protein solution. The results for the robotic pipetting were again proved to be normally distributed; characteristics of the distributions are given in table 6 [buffer solution] and in table 7 [protein solution].

Obviously the relative standard deviation decreases significantly with increasing scheduled pipetting volume in both cases. Pipetting a volume of 10 μl , the relative standard deviation was 6.5 % (buffer) respectively 9.8 % (protein); pipetting higher volumes of around 200 μl , the relative standard deviation was only 1.5 % (buffer) respectively 1.3 % (protein). The increase of relative standard

scheduled pipetted volume [μl]	$\bar{x}(\text{mass})$	$\sigma(\text{mass})$	$\sigma_{rel,pipp}$ [%]
10	8.91	0.57	6.5
20	19.45	0.65	3.4
40	41.44	1.48	3.6
100	99.99	1.84	1.8
200	197.41	2.89	1.5
800	799.30	4.06	0.5

TABLE 6. Mean and (relative) standard deviations for the pipetting of buffer solution with the LiHA

scheduled pipetted volume [μl]	$\bar{x}(\text{mass})$	$\sigma(\text{mass})$	$\sigma_{rel,pipp}$ [%]
10	8.07	0.79	9.8
20	18.80	0.94	5.0
40	40.20	1.12	2.8
100	99.35	1.88	1.9
200	198.01	2.49	1.3
800	799.44	2.78	0.4

TABLE 7. Mean and (relative) standard deviations for the pipetting of protein stock solution with the LiHA

deviations for buffer and protein pipetting and the pipetted means are very similar; therefore, the pipetting data was summarized and analogously to the previous section a functional relationship between the data and their relative standard deviation could be stated to:

$$\sigma_{rel,pipp} = 0.3 \cdot Vol_{pipp}^{-0.6} \quad (18)$$

with a coefficient of determination of 95.83 %. The fit of relative standard deviations from both liquids, buffer and protein pipetting, to this functional relation curve is demonstrated in figure 8.

4.3. Results based on Monte Carlo sampling.

4.3.1. *Proof of concept.* For the proof of concept all previously determined relative standard deviations were applied in the simulation algorithm in order to compare simulated uncertainties to the experimental uncertainties of the binding studies. Figure 9 shows the relevant quotients Q for the comparison of relative standard deviations (cmp. equation (16)). The supernatant concentrations c_1 to c_{17} are left out, because their means were negligibly small in both approaches. Most of the quotients are very close to 1 or slightly larger, what indicates a consistence between $\sigma_{rel,meas}$ and $\sigma_{rel,sim}$. The simulation algorithm is rather overestimating the standard deviations than underes-

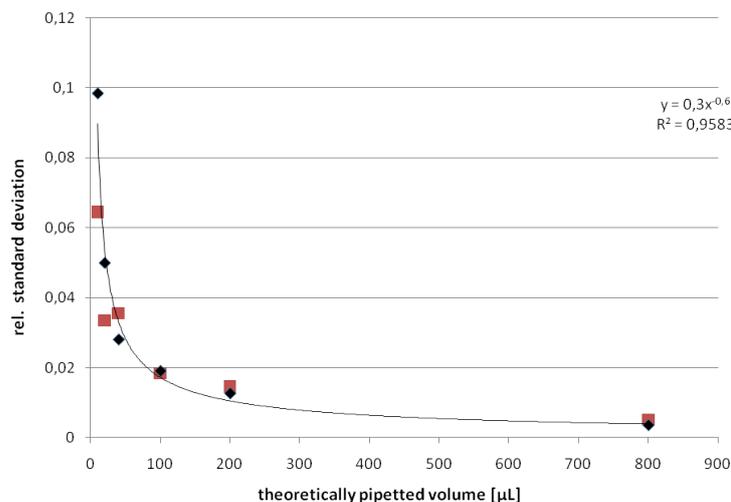


FIGURE 8. This figure shows the relative standard deviation of pipetted volumes on a Tecan Evo robotic station with dilutors holding 1 ml for the pipetting of protein solution (diamonds) respectively buffer (squares). The relative standard deviation rises with pipetting smaller volumes. This correlation was described with the power function, that is shown in the top right-hand corner of the figure.

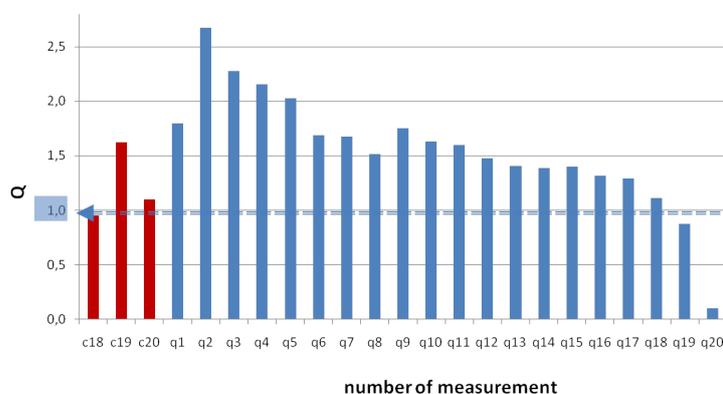


FIGURE 9. Quotients Q are shown, comparing the relative standard deviations for measurement points of experimental isotherms to the deviations calculated by the simulation algorithm including the experimentally derived uncertainties of single process steps (cmp. equation (16)). A quotient Q of 1 would indicate an exact agreement of experimental and simulated deviation. The figure shows, that the simulation algorithm is rather overestimating the standard deviations than underestimating them.

timating them. Only q_{20} is very close to 0, what indicates, that for the last measurement point of the isotherm, the relative standard deviations from the simulations are significantly smaller than the deviation calculated based on experimental results.

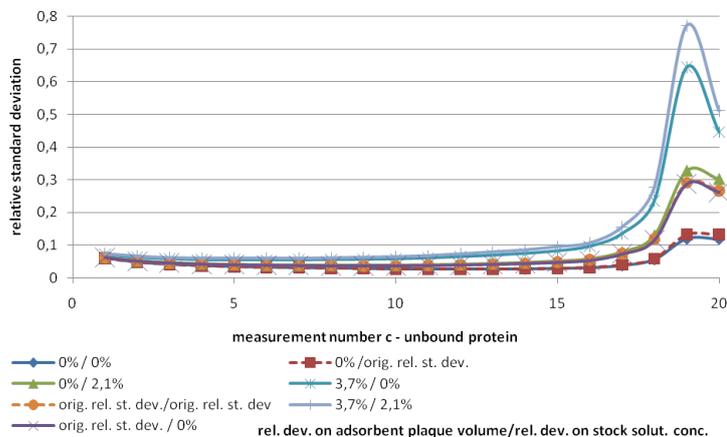


FIGURE 10. The relative standard deviations for twenty simulated measurement points for unbound protein c_{sim} are illustrated for seven selected combinations of uncertainty in adsorbent volume and uncertainty in the concentration of lysozyme stock solution. The original standard deviations account for the values that were determined experimentally: 1.8 % relative standard deviation in the adsorbent volume and 0.42 % relative standard deviation in the concentration of the lysozyme stock solution.

4.3.2. *Case scenarios.* Having shown a good agreement of experimental and simulated isotherms, case scenarios were set up based on the simulation algorithm. For the case scenarios considering the effect of uncertainty in adsorbent plaque volume and stock solution concentration a grid was constructed and 10000 simulations were performed in each case for eleven kinds of plaque volume uncertainties in the range from 0% to 3.7% and eleven uncertainty values on the protein stock solution in a range from 0% to 2.1%. The uncertainties for the pipetting of storage buffer and the LiHA-pipetting were in all simulations fixed to the experimentally derived values (see above). Selected results from the 121 possible combinations of uncertainties are illustrated in figure 10 for the twenty measurement points of c_{sim} , respectively in figure 11 for the simulations of the measurements of the bound protein q_{sim} . The curves for 0 % standard deviation for both process steps, adsorbent plaque production and production of stock solution show the effect of only pipetting uncertainties. A comparison between the curves, where on the one hand only the deviation on adsorbent plaque volume (orig. rel. st. dev./0 %), on the other hand only the stock solution concentration deviation (0 %/orig. rel. st. dev) were set to the original experimentally derived values, shows, that uncertainties on the plaque volume together with the pipetting error are mainly responsible for the overall error on the results from binding studies. Only a very high uncertainty of 2.1 % on the concentration of the stock solution would have a similar effect compared to the 'normal' uncertainty on adsorbent plaque volume. The worst case, represented by the curve with 3.7 % relative standard deviation on the adsorbent plaque volume and of 2.1 % on the concentration

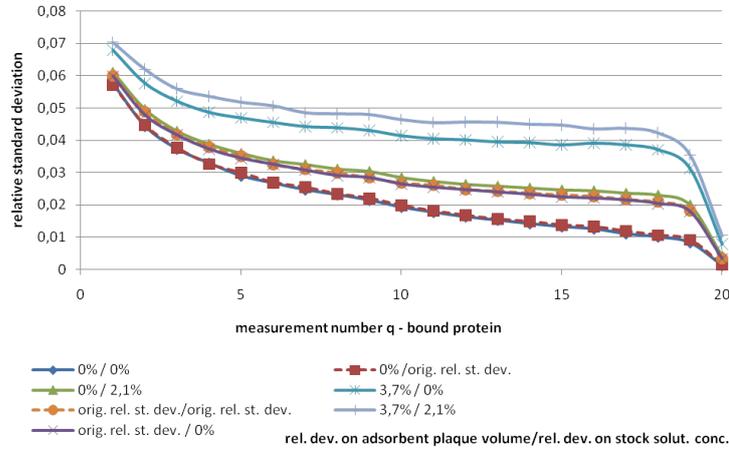


FIGURE 11. The relative standard deviations for twenty simulated measurement points for bound protein q_{sim} are illustrated for seven selected combinations of uncertainty in adsorbent volume and uncertainty in the concentration of lysozyme stock solution. The original standard deviations account for the values that were determined experimentally: 1.8 % relative standard deviation in the adsorbent volume and 0.42 % relative standard deviation in the concentration of the lysozyme stock solution.

of the stock volume shows a relative standard deviation on $c_{sim,19}$ of around 75 % and on $q_{sim,19}$ of 3.5 %. In general the uncertainties on the simulated measurements for the bound protein q_{sim} lie in the range of 0 % to 7 % and are slightly decreasing in direction of the saturation part of the isotherm while the uncertainties on the simulated measurements for the unbound protein c_{sim} lie in the range of 0 % to 80 % and are increasing in direction of the saturation part of the isotherm. This is not correct for the last point of the isotherm $c_{sim,20}$, that generally shows very small simulated uncertainties.

For the case scenarios considering the effect of LiHA pipetting uncertainty 10000 Monte Carlo samples were simulated for absolute pipetting deviations in a range of 0 % to 5 % in 0.1 %-steps. Figures 12 and 13 visualize the effect of LiHA pipetting for the standard deviations on the concentrations of unbound, respectively bound protein. Again the uncertainty in pipetting has a large effect on the measurement points of the unbound protein (range of standard deviations: 0 % to 90 %) and a more than ten times smaller effect on the measurements of bound concentration (range of standard deviations: 0 % to 6 %). In general, the uncertainty in pipetting has a rather constant effect on the deviations in the concentration measurements of bound protein q , that increases with increasing uncertainty. Conversely, the effect on the concentration of the unbound protein increases in direction of the saturation level of the isotherm. This is again not correct for the last point of the isotherm $c_{sim,20}$, that generally shows very small simulated uncertainties.

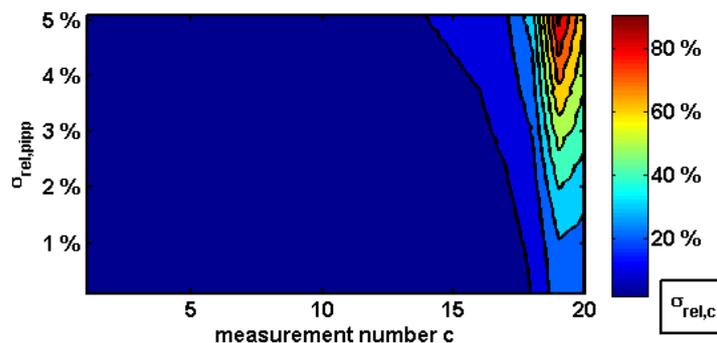


FIGURE 12. A contour plot showing the relative standard deviations on simulated measurements of **unbound** protein c_{sim} in simulated binding studies for the assumption of constant uncertainties on the robotic pipetting in a range of 0 % to 5 %.

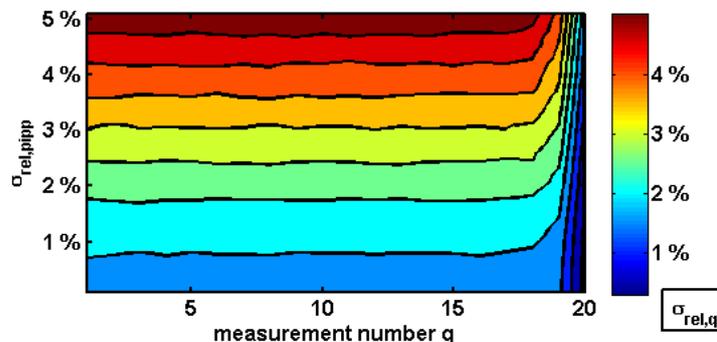


FIGURE 13. A contour plot showing the relative standard deviations on simulated measurements of **bound** protein q_{sim} in simulated binding studies for the assumption of constant uncertainties on the robotic pipetting in a range of 0 % to 5 %.

5. DISCUSSION

The presented results show, that based on a reasonable segmentation of the whole process of interest in 'single process steps' and carefully selected experiments for the determination of step-specific deviations, it is possible to establish simulation algorithms, that reproduce the process in silico and can be used for uncertainty propagation analyses and the simulation of case scenarios. It was illustrated, that the relative standard deviations calculated based on the simulation algorithm are close to the deviations on the experimental results (cmp. figure 9). The slight overestimation of total uncertainty by the simulation algorithm is most likely caused by the rather 'pessimistic' application of deviations to the simulated results, as outliers have not been sorted out in the determination of single step deviations. Secondly, correlations between uncertainties in the single steps have been neglected in this approach. Such correlations would consequently be examined in the next step of uncertainty propagation analysis if desired. The reason for the underestimation of uncertainty in the last measurement point of the isotherm, c_{20} and q_{20} , both, in the 'proof of

concept' and in the following case scenarios, is probably caused by the necessity to fix Langmuir parameters in the simulation algorithm. They were fixed to constant values (cmp. table 2), the means of Langmuir parameters estimated based on the twelve experimental isotherms. Though this parameter fixing showed to be convenient, the parameters are not 'true' and as the last measurement point of the isotherms is located in the saturation part and is therefore strongly correlating with the Langmuir parameter q_{max} , a fixed parameter naturally induces small deviances. Of high importance would now be a following study which concentrates on the influences of uncertainties in the process on the estimation of Langmuir parameters. However, the general quantity of uncertainty for all other measurements could be experimentally observed and simulated in the simulation algorithm including the tendencies, for example the uncertainty on c increasing in direction of the isotherm saturation level (cmp. figure 5).

The simulation algorithm was set up modularly enabling further experimenters to include or change individually the modules for robotic action, specific process steps and the adequate uncertainties. Often, the measurement of uncertainties allows for a deeper insight in corresponding single process steps and for a discovery of previously unidentified dependencies. In this case study, for example, a 'training-dependency' of the relative standard deviation on the adsorbent plaque volume was observed in adsorbent plaque production (cmp. table 3). This 'training-dependency' is significant and can be explained by the handling of the device, which bears some difficulties especially in the tight assembly of component parts and when the plaques have to be outdone from the device into the 96 well plate. However, the values assigned to be the deviations in adsorbent plaque volume by a 'trained experimenter are close to the deviations published by [19]. This result and the results from Monte Carlo sampling on the effect of deviation on adsorbent plaque volume (cmp. figures 10 and 11), confirm quantitatively the assumption from [19], that the quality of the results in binding studies on robotic platforms is mainly dependent from the quality of the adsorbent plaques. The quality of the adsorbent plaques is also responsible for the increase of uncertainty in direction of the saturation part of the isotherm (cmp. again figure 5), as irregularities in the adsorbent plaque volume have here a direct effect on the equilibrium. In the first part of the isotherm, when all protein is adsorbed in the plaque, the protein concentrations in the supernatant are close to zero; therefore, irregularities in adsorbent plaque volume are of negligible consequence in this part of the isotherm.

Compared to these findings, the deviation in manual pipetting of storage buffer was very small (cmp. table 4) and appeared to have the least influence on the overall results. The identification of negligible deviations of process steps is very important, because it allows for simplification in further studies. On the contrary, the dependence of the uncertainty on the results of the binding study from deviations in stock solution concentration (cmp. table 5) was not negligible (see again figures 10 and 11) and could be quantified in the case scenarios. For the standard deviation of stock

solution concentration a dependency from the weighed in protein mass could be derived. It is due to a decrease in weighing accuracy for smaller masses of protein, as small masses of lysozyme can easily attach to the spoon or the outer part of the vial by adhesive or electrostatic forces. This dependency was included into the simulation algorithm by a power function (cmp. figure 7).

The influence of deviation in pipetting with the robotic LiHA (cmp. tables 6 and 7) could also be described by a functional relationship between the relative pipetting deviation and the pipetted volume (cmp. figure 8) and applied in the simulation algorithm based on a power function. This relationship is dependent from the used Liquid Class on the robotic system, but a general dependency of pipetted volume to standard deviations on the pipetting could be observed for many examined Liquid Classes and pipetting methods (data not shown). Pipetting smaller volumes is generally more deviation-prone, as drops can stick to the pipetting tip or adhesive forces hinder pouring out the complete volume in the tip. The power functions, which were applied for the determination of deviation relationships in the case study suggest uncertainty propagation underlying the examined steps. These dependencies could have been disassembled further, but this was not necessary for the simulation. This example shows the importance of a reasonable segmentation of the whole process of interest in 'single process steps'.

With the case scenario simulating different levels of uncertainty for LiHA pipetting, the influences of higher pipetting uncertainties on the total uncertainty of the binding study were demonstrated and quantified (cmp. figures 12 and 13). This study is for example relevant with respect to wear effects in the examined system. A complete process of wearout of instruments could in this way be simulated.

6. CONCLUSION & OUTLOOK

In general, it could be demonstrated on a simple example, that a simulation algorithm in combination with Monte Carlo sampling is applicable for uncertainty propagation calculation in a high-throughput process as long as the main uncertainty sources can be identified and the distribution and deviations be observed and quantified. Functional relationships between scheduled volumes and concentrations and their relative standard deviation have been successfully integrated into the simulation algorithm. Based on the findings of the case scenarios, the process can now be improved, especially the production of adsorbent plaques should be substituted by a more defined and standardized method, that is less user-dependent. In addition, the pipetting uncertainty of the LiHA should be regularly controlled. Of general interest would be also an assessment of the influence of uncertainties in process steps on the estimation of (Langmuir) parameters based on the process results.

The simulation algorithm can now be expanded to more complicated processes on robotic systems, for example experiments with miniaturized chromatography columns or aqueous two phase systems.

The identification and quantification of analytic weaknesses influencing overall results will lead to the improvement of existing processes and the development of more suitable high-throughput analytics and robust processes. Thus, based on this simple example, a general strategy for uncertainty analysis in more complex high-throughput experimentation and screening protocols should be developed and be applied as a standard, including deviations and error distributions from relevant process steps as well as qualified decisions on suitable mechanistic equations modeling the process.

REFERENCES

- [1] F.D.A.U.S. Department of Health and Human Services (Ed.), *Guidance for Industry PAT - A Framework for Innovative Pharmaceutical Manufacturing and Quality Assurance*, 2004.
- [2] H. Wan, M. Rehgren, High-throughput screening of protein binding by equilibrium dialysis combined with liquid chromatography and mass spectrometry, *Journal of chromatography A* 1102 (1-21-2) (2006) 125–34.
- [3] A. Staby, R. Jensen, M. Bensch, J. Krarup, J. Nielsen, M. Lund, S. Kidal, T. Hansen, I. Jensen, Comparison of chromatographic ion-exchange resins vi. weak anion-exchange resins, *Journal of Chromatography A* 1164 (2007) 82–94.
- [4] 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 (7) (2007) 2011–2021.
- [5] 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 (2009) 140–154.
- [6] W. Kessel, *ISO/BIPM Guide: Uncertainty of measurement*, 1999.
- [7] X. Fuentes-Arderiu, Uncertainty of measurement in clinical laboratory sciences, *Clinical chemistry* 46 (9) (2000) 1437–1438.
- [8] C. Dietrich, *Uncertainty, calibration, and probability: The statistics of scientific and industrial measurement*, 2nd Edition, A. Hilger, 1991.
- [9] J. Ogilvie, A Monte-Carlo approach to error propagation, *Computers & Chemistry* 8 (3) (1921) 205–207.
- [10] C. Efstathiou, Monte-Carlo simulation for the study of error propagation in the double known addition method with ion-selective electrodes, *Analytical Chemistry* 54 (9) (1982) 1525–1528.
- [11] A. Kurup, H. Subramani, M. Harris, A monte carlo-based error propagation analysis of simulated moving bed systems, *Separation and Purification Technology* 62 (2008) 582–589.
- [12] D. Hibbert, Propagation of uncertainty in high-performance liquid chromatography with uv, *Analytica Chimica Acta* 443 (2) (2001) 205–214.
- [13] N. Faber, Quantifying the effect of measurement errors on the uncertainty in bilinear model predictions: a small simulation study, *Analytica Chimica Acta* 439 (2) (2001) 193–201.
- [14] E. Hund, D. Massart, J. Smeyers-Verbeke, Comparison of different approaches to estimate the uncertainty of a liquid chromatographic assay, *Analytica Chimica Acta* 480 (2003) 39–52.
- [15] T. Barz, V. Loeffler, H. Arellano-Garcia, G. Wozny, Optimal determination of sma-parameters for the adsorption isotherms of β -lactoglobulin under non-isocratic conditions, *AIChE Journal* 38 (1992) 1969 – 1978.
- [16] L. Michaelis, M. Menten, Die Kinetik der Invertinwirkung, *Biochemische Zeitschrift* 49 (333-369) (1913) 352.

-
- [17] B. Efron, R. Tibshirani, *An Introduction to the Bootstrap*, Chapman and Hall, CRC Press, London, Boca Raton, 1993.
 - [18] J. Tellinghuisen, Statistical error propagation, *The Journal of Physical Chemistry A* 105 (15) (2001) 3917–3921.
 - [19] T. Herrmann, M. Schroeder, J. Hubbuch, Generation of equally sized particle plaques using solid-liquid suspensions., *Biotechnology progress* 22 (3) (2006) 914–8.
 - [20] R. Ritchie, T. Prvan, A simulation study on designing experiments to measure the km of michaelis-menten kinetics curves, *Journal of Theoretical Biology* 178 (3) (1996) 239–254.

Determination of parameters for the steric mass action model - A comparison between two approaches

A. Osberghaus¹, S. Hepbildikler², S. Nath², M. Haindl², E. von Lieres³ and J. Hubbuch^{1,*}

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

² : Pharmaceutical Biotech Production, Roche Diagnostics GmbH, Nonnenwald 2, 82377 Penzberg, Germany

³ : Institute of Bio- and Geosciences 1, Research Center Jülich, 52425 Jülich, Germany

* : Corresponding author. *E-mail-address*:juergen.hubbuch@kit.edu (J. Hubbuch)

ABSTRACT

The application of mechanistic modeling for the optimization of chromatographic steps increased recently due to time efficiency of algorithms and rising calculation power. In the modeling of ion exchange chromatography steps, the sorption processes occurring on adsorbent particle surfaces can be simulated with the steric mass action (SMA) model introduced by (C. Brooks and S. Cramer, 1992). In this paper, two approaches for the determination of SMA parameters will be carried out and discussed concerning their specific experimental effort, quality of results, method differences, reasons for uncertainties and consequences for SMA parameter determination:

Approach I: *estimation of SMA parameters based on gradient and frontal experiments according to instructions in (C. Brooks and S. Cramer, 1992) and (A. Shukla, S. Bae and J. Moore, 1998)*

Approach II: *application of an inverse method for parameter estimation, resulting in SMA parameters that induce a best fit of chromatographic data to a mechanistic model for column chromatography.*

These approaches for SMA parameter determination were carried out for three proteins (ribonuclease A, cytochrome c and lysozyme) at pH 5 and pH 7. The results were comparable and the order of parameter values and their relations to the chromatographic data similar. Nevertheless, differences in the complexity and effort of methods as well as the parameter values themselves were observed. The comparison of methods demonstrated that discrepancies depend mainly on model sensitivities and additional parameters influencing the calculations. However, the discrepancies do not affect predictivity; predictivity is high in both approaches. The approach based on an inverse method and the mechanistic model has the advantage that not only retention times but complete elution profiles can be predicted. Thus, the inverse method based on a mechanistic model for column chromatography is the most comfortable way to establish highly predictive SMA parameters lending themselves for the optimization of chromatography steps and process control.

Keywords: steric mass action, inverse method, lumped rate model, parameter estimation, chromatography modeling

1. INTRODUCTION

Ion Exchange Chromatography (IEC) is one of the key procedures in bioseparation processes. As chromatography steps cover up to 70% of the overall financial effort in biopharmaceutical production, the development of optimal and efficient chromatography processes is a central issue. This issue is even more urgent with respect to the concept of Quality by Design, launched by the US Food and Drug Administration, which requires additional attention to process robustness and reproducibility matters [1]. The search for a favorable and robust operating point of a separation process represents a complex multi-factor optimization problem. One way to tackle this problem is given by screenings for optimal factors in the design space based on design of experiments (DoE). This procedure is often complemented by empiric response surface modeling (RSM). The DoE-RSM-approach is quite established for chromatography optimization; a review on this approach is, for example, given in [2].

However, the application of mechanistic modeling for the optimization of chromatographic steps is on the rise due to increased time efficiency of algorithms and progresses in calculation power (cf. argumentation lines in [3], [4] and [5]). Based on mechanistic modeling, highly precise predictions of chromatograms can be achieved, as was demonstrated in [6] for step gradients, for linear gradients in [7] and for displacement systems in [8]. In addition, mechanistic modeling lends itself for efficient robustness and sensitivity analyses, which was demonstrated in [9] and [10]. Thus, a simulation of chromatographic processes based on mechanistic modeling can, similar to the DoE-RSM approach, support and cheapen the search for optimal conditions and provides additionally troubleshooting and error diagnostic tools for process development.

Despite of some obvious advantages of mechanistic modeling, a main drawback, particularly in comparison with the DoE-RSM-approach, is given by the high effort for preliminary model calibration. A calibrated mechanistic model simulates the flow of the mobile phase through the column and imitates interactions between the mobile phase and the adsorbent surface; a most important but hardly monitorable piece of the whole IEC process. These simulations base on parameters that determine the chromatographic system on column and particle level. Numerous publications with proposals for most effective model calibration have been released, like for example [11], discussing mainly the determination of parameters characterizing the packed bed, [12], dealing with parameters of mass transfer kinetics and [13], where a specific set of experiments is proposed for the determination of all model parameters. These publications reveal that system parameters (dead volume, etc.) and bed characterizing parameters (axial dispersion, porosities, etc.) can be determined with a few basic pulse experiments. However, the determination of parameters for the steric mass action (SMA) model that describes the interaction between the proteins and the adsorbent surface in IEC [14] is based on time-consuming isocratic or gradiental experiments and material-consuming frontal experiments (cf. for example the instructions for SMA parameter

determination in [14], [15], [16] and [17]). In this context, approaches for the determination of SMA parameters based on batch experiments, like in [18] and [19] are interesting but not yet fully established and validated.

A quite established alternative to the experimental determination of SMA parameters is given by model-based inverse methods. In these methods the parameters are calculated by the best fit between data and chromatography model response. Recently, this method was applied more often for sorption parameter determination, for example in [20], [21] and [4], but so far no direct comparison between the results of the suggested methods in [14], [15], [16] and the results of approaches based on an inverse method has been given. It is expected that within the context of high throughput process development which has had an immense impact in the last couple of years within the field of industrial process development, new methods which might be less precise but quicker and more intuitive in their realization will pave the way of model based process development. Inverse modeling might for example be performed with historic data, data already existing from process development, process data etc. Thus it offers a by far more potent application tool than the other approaches. However, a precise determination of sorption parameters allowing a high predictivity is essential for model based process development, as ad- and desorption and adsorbent capacities remarkably decide on retention time and separation quality. Thus, only a direct comparison between different approaches can act as a background for discussion on the optimal determination of modeling parameters and influence of noise in chromatographic data on parameter estimation. Such a comparison should on the one hand discuss experimental effort and parameter qualities, on the other hand it should pay attention to prediction performance, chances, advantages and disadvantages of both approaches. It would thus be of interest, how both approaches differ and what consequences this has for the determination of sorption parameters and which approach is to be favored in future SMA parameter determinations. The latter might lead to a clear distinction if one is interested in the physical meaning of the underlying isothermal concept or simply aims towards a tool for modern process development schemes.

In this paper two approaches in SMA parameter determination will be experimentally executed and discussed concerning specific experimental effort, quality of results, method differences, reasons for uncertainties and consequences for the determination of sorption parameters:

Approach I: estimation of SMA parameters based on gradient and frontal experiments according to instructions in [14] and [16]

Approach II: application of an inverse method for parameter estimation resulting in SMA parameters that induce a best fit of chromatographic data to a mechanistic model for column chromatography

Although the considered methods can as well be applied to protein mixtures with industrial importance, the determination of SMA parameters will be performed based on a case study for three

proteins (lysozyme, ribonuclease A and cytochrome *c*) at pH 5 and pH 7 on a prepacked 1 mL column with the strong cation exchange adsorbent SP Sepharose FF. This case study guarantees a complete comparability of datasets and a comparison of methods based on a well known system of proteins, that highlights advantages and disadvantages of the methods. The discussion will be supported by a comparative literature review and a Monte-Carlo study on the influence of noise in data on the quality of the determined SMA parameters.

2. THEORY

2.1. SMA model for sorption processes in IEC. For IEC processes, a highly regarded characterization of sorption is given by the steric mass action (SMA) model introduced by Brooks and Cramer [14], which accounts for the influence of charged modifiers and their rivalry with proteins for binding sites on the adsorbent surface. In the understanding of [14], in the SMA model every component interacting with the particle surface owns four characterizing parameters (three in case of a rapid equilibrium assumption):

- ν (characteristic charge): average number of binding sites of the component (under the assumption of a single charged counterion)
- σ (steric factor): average number of shielded/covered binding sites on adsorbent surface due to the 3D-structure of the protein components
- k_{ads} and k_{des} (ad- and desorption coefficient): the ratio of the ad- and desorption coefficients is lumped to a single parameter k_{eq} when a rapid equilibrium is assumed (compare with the following equations).

Based on these initial considerations, the time dependent change of the concentration of component i on the adsorbent surface ($\frac{\partial q_i}{\partial t}$) in the SMA model is given by:

$$\frac{\partial q_i}{\partial t} = k_{ads,i} c_i \bar{q}_1^{\nu_i} - k_{des,i} c_1^{\nu_i} q_i \quad i > 1 \quad (1)$$

$$\Lambda = q_1 + \sum_{i=2}^n \nu_i q_i \quad (2)$$

$$\bar{q}_1 = q_1 - \sum_{i=2}^n \sigma_i q_i \quad (3)$$

with respect to n components ($n = 1[\text{salt}] + \text{number of protein components}$). Λ , the parameter describing the ionic capacity of the adsorbent, limits the available binding places and displays the rivalry between salt ion concentration q_1 and the other bound components q_i with their characteristic charges ν_i (cf. equation (2)). \bar{q}_1 , the concentration of bound salt ions available for exchange with the protein, is given by the total salt ion concentration q_1 less the shielded ions determined by the

protein specific steric factors (σ_i) in equation (3). If the assumption of a rapid equilibrium is valid ($\frac{\partial q_i}{\partial t} = 0$), the equation for the SMA-isotherm can be derived from the above equations to be:

$$k_{i,eq} = \left(\frac{q_i}{c_i}\right) \left(\frac{c_1}{\Lambda - \sum_{i=2}^n (\nu_i + \sigma_i) q_i}\right)^{\nu_i} \quad i > 1 \quad (4)$$

The results of SMA parameter determination following instructions in [14] and [16] (**approach I**) will be compared to the estimation of SMA parameters by fitting a mechanistic model for IEC chromatography processes to the monitored and time resolved concentrations of a protein component at column outlet.

2.2. Approach I: Determination of parameters for the SMA model according to instructions in [14] and [16]. The determination of the parameters ν and k_{eq} for the SMA model based on gradiental chromatographic experiments was performed based on an equation of Parente and Wetlauffer [22] modified by [16]:

$$V_R = \left(\left(c_{a,s}^{\nu+1} + \frac{V_d k_{eq} \varepsilon_c \Lambda^\nu (\nu + 1) (c_{e,s} - c_{a,s})}{V_G} \right)^{\frac{1}{\nu+1}} - c_{a,s} \right) \frac{V_G}{c_{e,s} - c_{a,s}} \quad (5)$$

The determination of the parameter σ based on the previously determined parameters ν and k_{eq} and additional frontal chromatographic experiments was carried out based on equation (6) given by [14]:

$$\sigma = \frac{\beta}{c_{prot} \vartheta} \left(\Lambda - c_{salt} \left(\frac{\vartheta}{\beta k_{eq}} \right)^{\frac{1}{\nu}} \right) - \nu \quad (6)$$

with

$$\vartheta = \left(\frac{V_B}{V_0} - 1 \right)$$

Equation (5) provides a correlation between gradient volume and elution volume, where ν and k_{eq} appear implicitly. Equation (6), derived based on [14], poses an explicit expression for the steric factor σ based on the previously estimated parameters ν and k_{eq} . β describes the column phase ratio $\frac{1-\varepsilon_t}{\varepsilon_t}$. The necessary information for the solution of these equations can be divided in 'experiment-enclosed information' - information from the gradient and breakthrough experiments - and supplementary parameters from other sources - 'external information':

experiment-enclosed information

- V_R, V_G - retention time V_R with respect to a specific elution gradient volume V_G

- $c_{a,s}$, $c_{e,s}$ - salt concentration at gradient begin, respectively end
- c_{prot} - protein concentration in the stock solution (breakthrough)
- c_{salt} - salt concentration in the buffer (breakthrough)
- V_B - breakthrough volume at 10% of the complete breakthrough

external information

- Λ - ionic capacity of the adsorbent
- V_d - column dead volume
- ε_c - column porosity
- ε_t - total porosity
- V_0 - breakthrough volume at 10% of a nonretarded tracer

2.3. Approach II: Determination of parameters for the SMA model by an inverse method.

2.3.1. *Introduction to the employed mechanistic model for chromatography.* A mechanistic model for chromatography consists of equations describing convective and dispersive transport, mass transfer resistances and equations describing sorption kinetics, for example the SMA model in IEC. Here, a short overview with respect to the employed model equations, a transport-dispersive model, is given, details on the equations and the implementation of their solution can for example be found in [23], [24] and [5].

On column level, the time- and position-dependent change of concentration for the i -th component, $\partial c_i / \partial t$, is described by:

$$\frac{\partial c_i}{\partial t} = -u_{int} \frac{\partial c_i}{\partial x} + D_{ax} \frac{\partial^2 c_i}{\partial x^2} - \frac{1 - \varepsilon_c}{\varepsilon_c} \cdot \frac{3}{r_p} k_{eff,i} [c_i - c_{p,i}] \quad (7)$$

where the first term on the right hand side describes the convective transport through the column, the second term the dispersive transport and the third term the mass transfer to the particle surface, $k_{eff,i}$ representing the lumped film diffusion coefficient and r_p the particle radius. u_{int} denotes the interstitial velocity, ε_c the column porosity, D_{ax} displays the axial dispersion, more precisely, a combined effect of dispersion and diffusive processes, dispersion being eddies and all effects implied by three dimensionality.

Analogously, the time-dependent change of concentration on particle level for the i -th component, $\partial c_{p,i} / \partial t$, is described by:

$$\frac{\partial c_{p,i}}{\partial t} = \frac{3}{\varepsilon_p r_p} k_{eff,i} [c_i - c_{p,i}] - \frac{1 - \varepsilon_p}{\varepsilon_p} \frac{\partial q_i}{\partial t} \quad (8)$$

with q_i denoting the concentration of particle-bound component i and ε_p the particle porosity. The second term of equation (8) describes ad- and desorption processes on particle level, i.e. the

interaction between mobile and bound phase. Thus, the expression $\partial q_i / \partial t$ is defined in equation (1) respectively, when a rapid equilibrium is assumed, in equation (4).

For the solution of the whole differential-algebraic equation system, Danckwert's boundary conditions were employed [25]. This model was solved in MatLab on a Dual Core Processor with 2.81 GHz in approximately 10 seconds with a density of 200 knots over the whole column length. That is a reasonable time span since the model has to be solved hundreds of times for the inverse method (approach II).

2.3.2. *A model-based inverse method for the determination of SMA parameters.* Let $c(t_j)$ be the chromatogram monitored at column outlet at the points in time $j = t_0 \cdots t_{end}$, preprocessed to a concentration time series. Let $\hat{c}(t_j)$ be the solution of a mechanistic model for chromatography at the same location and points in time; $\hat{c}(t_j)$ can then be compared to the chromatograms. Let now θ_{fix} be the set of all model input parameters that are fixed on a constant value and θ_{est} the set of model input parameters that can be manipulated by the algorithm solving the inverse problem (for estimating the SMA parameters $\theta_{est} = \{\nu, k_{eq}, \sigma\}$). Then the inverse problem can be stated as an minimization of a least squares residual given by:

$$res(\theta_{est}) = \sum_{t=t_0}^{t_{end}} (\hat{c}(t_j, \theta_{fix}; \theta_{est}) - c(t_j))^2 \quad (9)$$

The minimization of equation (9) was in all cases performed with the Matlab procedure `lsqnonlin`.

Analogously to the determination of SMA parameters described in section 2.2, the solution of the inverse method demands for 'experiment-enclosed information' provided by column chromatographic data and 'external information' - additional parameters from other sources:

experiment-enclosed information

- the chromatogram
- $c_{a,s}$, $c_{e,s}$ - salt concentration at gradient begin, respectively end
- c_{prot} - protein concentration in the stock solution (breakthrough)
- c_{salt} - salt concentration in the buffer (breakthrough)

external information

- information characterizing the packed bed
 - Λ - ionic capacity of the adsorbent
 - D_{ax} - axial dispersion
 - u_{int} - interstitial velocity
 - ε_c - column porosity
 - ε_p - particle porosity
 - ε_t - total porosity

-
- k_{eff} - lumped film diffusion coefficient
 - parameters characterizing column geometry
 - L_C - column length
 - r_p - particle radius

The determination of external information as referred to here and at the end of section 2.2, will be briefly described in the next section. Detailed instructions can be found for example in [11].

3. MATERIALS AND METHODS

3.1. Materials. Blue dextran 2,000,000, cytochrome *c* (horse heart), lysozyme (chicken egg white) and ribonuclease A (bovine pancreas) were purchased from Sigma (St. Louis, MO, USA). The buffer substances, acetic acid, sodium monobasic phosphate, sodium dibasic phosphate and sodium hydroxide as well as phenolphthalein for titration were purchased from Merck KGaA (Darmstadt, Germany). Acetone and ethanol (99.8 %) were purchased from Carl Roth GmbH + Co.KG (Karlsruhe, Germany) and sodium chloride from AppliChem (Darmstadt, Germany).

3.2. Apparatus & Software. Prepacked HiTrap SP Sepharose FF 1 mL columns [0.025 m length, 0.7E-2 m ID] from GE Healthcare (Buckinghamshire, United Kingdom) were applied for all column experiments. The experiments were performed on an Ettan LC system from GE Healthcare with a system flow rate of $2.17 \cdot 10^{-4}$ m/s [0.5 mL/min]. Tubing connections and all other LC system parameters were standardized and kept constant. The absorption at 280 nm and 528 nm was measured online in all experiments. Primary analyses and documentation of the chromatograms were performed with the control software Unicorn. All further data analysis as well as the solution of model equations and all applications connected to the model was performed with MatLab (The Mathworks, Natick, ME, USA).

3.3. Gradient elution experiments. The running buffer for all experiments at pH 5 was a 0.02 M acetate buffer. The running buffer for all experiments at pH 7 was a 0.02 M sodium phosphate buffer; the same buffers with additional 0.5 M NaCl served for elution purposes. The proteins were each dissolved in the running buffer to a concentration of 0.2 mM. All gradient elution experiments were performed similar to the instructions in [16]. At first the column was equilibrated with the running buffer for four column volumes (cv), then 20 μ l of the protein solution were injected. This step was followed by another wash step with four cv running buffer. Afterwards, a linear gradient from 0% to 100% high salt buffer was set to elute the protein. This experimental setup was executed for five different gradient lengths: 5 cv, 10 cv, 30 cv, 60 cv and 120 cv.

3.4. Breakthrough experiments. Breakthrough experiments are necessary for the determination of the protein-specific steric factor σ_i (compare equation (6)). Thus, breakthrough experiments for ribonuclease A, cytochrome *c* and lysozyme at pH 5 and pH 7 were performed with the respective

running buffers and stock solutions. After equilibrating the column for 10 cv with running buffer, it was isocratically loaded with a system flow rate of $2.17 \cdot 10^{-4}$ m/s until the breakthrough was complete.

3.5. Determination of parameters for 'external information'. For the determination of external information in approach I and approach II (ε_c , ε_p , ε_t , the column, particle and total porosity, Λ , the ionic capacity of the column, D_{ax} , the axial dispersion coefficient and V_0 , the breakthrough volume of a non-retarded species), pulse and displacement experiments were performed according to instructions in [11] and [26]. All experiments were at least three times repeated for a check of reproducibility and variances. The lumped film diffusion coefficient k_{eff} was estimated to be $1.5 \cdot 10^{-6}$ m/s by the inverse fit of the mechanistic model to the tracer peaks from the pulse experiments described in the next section.

3.5.1. Pulse experiments. Dextran blue was provided as nonbinding and nonpenetrating tracer; acetone (1% in deionized water) was used as nonbinding but penetrating tracer. The absorption at 280 nm was measured online and retention times were corrected with respect to system dead volume. Porosities were calculated with the method of central moments based on several repetitions for all employed columns. Based on the total column volume of $0.96 \cdot 10^{-6}$ m³, the total porosity ε_c was calculated to be 0.92 ± 0.025 , the column porosity to be 0.36 ± 0.0009 and the particle porosity ε_p to be 0.85 ± 0.038 . The axial dispersion coefficient was calculated according to equation (10):

$$D_{ax} = \frac{\sigma_{mom}^2}{\mu_{mom}^2} \frac{L_C u_{int}}{2} \quad (10)$$

with u_{int} being the interstitial velocity, L_C the column length and μ_{mom} and σ_{mom} the first respectively second central moment of the nonbinding and nonpenetrating tracer peaks. The result for axial dispersion was $1.574 \cdot 10^{-10}$ m²s⁻¹ with a relative standard deviation of about 1.75 %.

3.5.2. Displacement experiments. The determination of total ionic capacity Λ was performed according to instructions in [26]. The packed column was equilibrated with deionized water and then isocratically loaded with acetic acid. The system was washed for another ten column volumes with deionized water and then the acetic acid was eluted with 1 M KNO₃-solution. 10 μ L of phenolphthalein solution (10 mg/mL in ethanole) were added to the eluate and the mixture was titrated with 0.01 M NaOH. The ionic capacity of the column was calculated by

$$\Lambda = \frac{c_{NaOH} V_{NaOH}}{V_C (1 - \varepsilon_t)} \quad (11)$$

with the column volume V_C and the concentration and volume of NaOH used for titration (c_{NaOH} , V_{NaOH}). Three determinations of the ionic capacity of the column were averaged resulting in $\Lambda = 800$ mM with a relative standard deviation of about 5 %.

3.6. Monte-Carlo method for the noise sensitivity of model-based SMA parameter determination.

The inverse method described in section 2.3 can as well be employed for sensitivity analyses of parameter determination. By Monte-Carlo simulations it is possible to quantify the effect of noise in chromatograms or of retention time shifts on the parameter estimation (more details on this method can be found in [27] and [28]). In short, 10000 chromatograms were simulated based on the model equations given in section 2.3.1 and then certain effects were attached to every single chromatogram, here, a normally distributed absolute noise or a normally distributed shift in time:

The influence of noise on chromatographic data was applicated with an absolute standard deviation on the data [in mM] in seven levels [$a = 0, 2, 4, 8, 16, 32, 50$] using the following equation:

$$data = data + a \cdot 10^{-7} \cdot randn \quad (12)$$

with `randn` a Matlab function providing normally distributed numbers with mean 0 and standard deviation 1. The influence of shift-noise in time was applicated as a normally distributed time-shift in the simulation in seven levels [$b = 0, 2, 4, 8, 16, 32, 50$, standard deviation of time shift in seconds] using the following equation:

$$time = time + b \cdot randn \quad (13)$$

Both noise applications were performed for three gradient lengths; 5 cv, 20 cv and 60 cv. Then, SMA parameters were re-estimated based on the deviation-afflicted chromatograms and the variances, correlations and distribution of the 10000 estimation results analyzed.

4. RESULTS

4.1. Gradient elution and breakthrough experiments. In both approaches for SMA parameter determination, chromatograms from gradient elution experiments and breakthrough data are the main source of information (cf. sections 3.3 and 3.4). These experiments were performed for the proteins ribonuclease A, cytochrome *c* and lysozyme at pH 5 and pH 7. Example results for gradient elutions with five gradient lengths (5 cv, 10 cv, 30 cv, 60 cv and 120 cv) and breakthrough data at pH 7 are given in figure 1; for better comparability, the experimental results from the single component experiments are superimposed. The first eluting component is ribonuclease A (light grey), followed by cytochrome *c* (dark grey) and lysozyme (black). The elution peaks of cytochrome *c* and lysozyme overlap. Obviously, an increase in elution gradient length increases the gap between

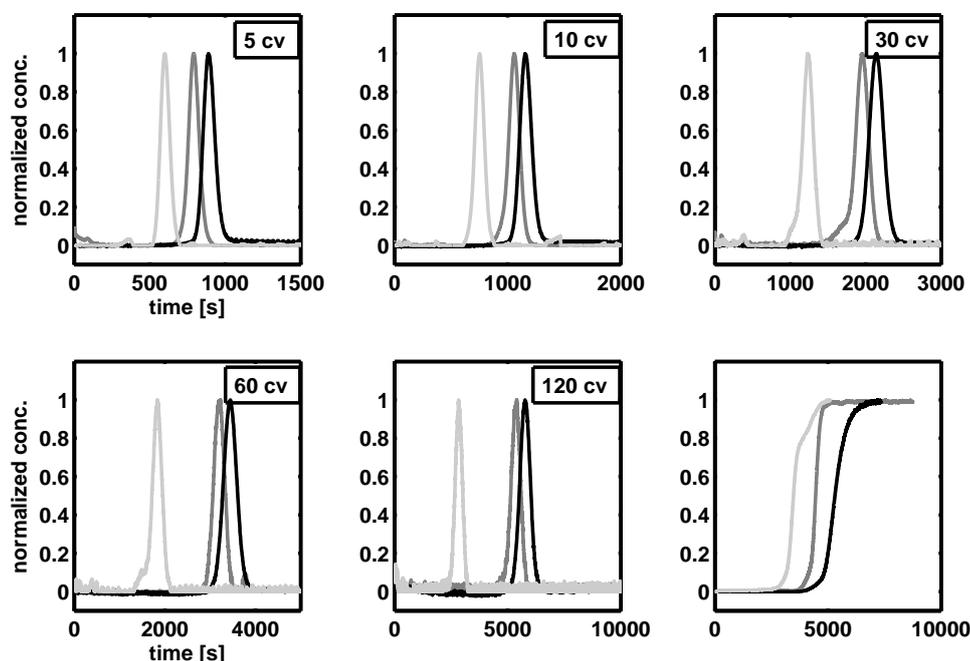


FIGURE 1. Superimposed results of single gradient and frontal experiments at pH 7 (SP Sepharose FF, flow rate: $2.17 \cdot 10^{-4}$ m/s, elution volumes: 5 cv, 10 cv, 30 cv, 60 cv and 120 cv). The gradient elution experiments are displayed with normalized concentration for better comparability [light grey continuous line: ribonuclease A, dark grey: cytochrome *c*, black: lysozyme].

the retention time of ribonuclease A and the other components. The frontal experiments show that the capacity of SP Sepharose FF for ribonuclease A is lower than for cytochrome *c* and lysozyme (highest capacity). Repetitions of gradient and frontal experiments showed a high reproducibility; absolute deviances in retention times determined by gradient experiments were always smaller than 57.725 s (0.5 mL), respectively smaller than 115.45 s (1 mL) for breakthrough volumes.

4.2. Determination of SMA parameters according to approach I. In figure 2 the correlation between retention time and gradient volume, given by equation (5), is exemplarily illustrated for cytochrome *c*. The measurement points for cytochrome *c* at pH 5 are displayed by ∇ -symbols, for pH 7 by Δ -symbols. The least squares fit of equation (5) to the data is displayed by the dotted line (pH 5), respectively the continuous line (pH 7), each fit having a coefficient of determination R^2 of 0.99. The correlation is positive proportional and slightly convex for both pH conditions. Retention times at pH 5 are generally larger and this effect even increases with increasing elution volume. Based on equation (5) and equation (6), on the five chromatograms at 5 cv, 10 cv, 30 cv, 60 cv and 120 cv and the breakthrough curve for every single protein, SMA parameters were determined for pH 5 and pH 7. The results for SMA parameters for ribonuclease A, cytochrome *c* and lysozyme

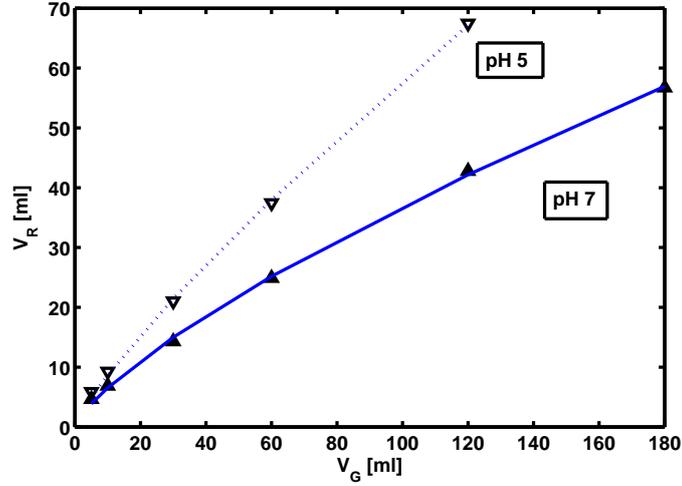


FIGURE 2. The correlation between elution gradient volume V_G and retention time V_R of a protein is described by equation (5). This figure illustrates the correlation for cytochrome c at pH 5 (∇ - measurement points, dotted line - fit) and pH 7 (Δ - measurement points, continuous line - fit)

parameter	ribonuclease A	
pH	5	7
ν	5.11	2.39
k_{eq}	0.148	0.233
σ	28.88	29.34
parameter	cytochrome c	
pH	5	7
ν	5.0	3.31
k_{eq}	0.307	0.356
σ	28.7	40.8
parameter	lysozyme	
pH	5	7
ν	4.72	4.07
k_{eq}	0.441	0.17
σ	36.8	29.74

TABLE 1. SMA parameters determined according to instructions in [14] and [16] [**approach I**].

at pH 5 and pH 7 are given in table 1. The characteristic charges for the proteins at pH 5 are quite close together in the limits from 4.72 to 5.11. The values are decreasing corresponding to the elution order of the proteins. For pH 7 the characteristic charges are more distinct (ribonuclease A: 2.39, cytochrome c : 3.31 and lysozyme: 4.07) and they are increasing corresponding to the elution

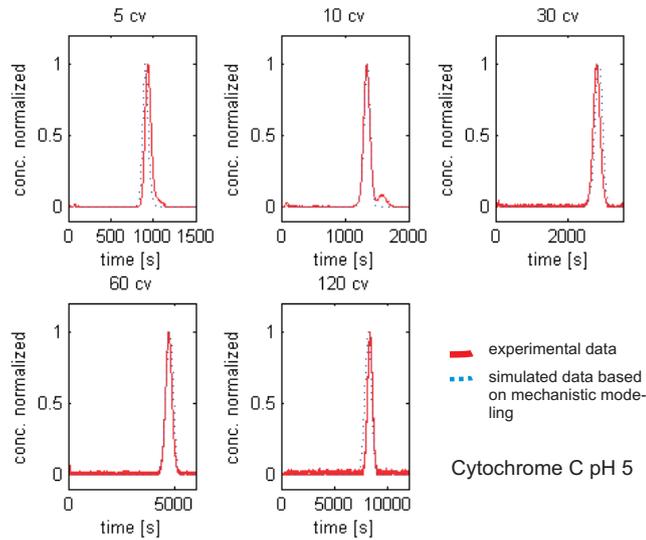


FIGURE 3. Example of inverse method fits based on mechanistic modeling. The experimental data of cytochrome c at pH 5 is displayed by a continuous line and the dotted line shows the inverse model-based fit.

order. The equilibrium coefficients are located between 0.148 (ribonuclease A at pH 5) and 0.441 (lysozyme at pH 5). At pH 5 they show an increasing trend corresponding to the elution order of the proteins. The steric factor of the proteins lies in the ranges between 28.7 (cytochrome c at pH 5) and 40.8 (cytochrome c at pH 7).

4.3. Determination of SMA parameters according to approach II. Example of simultaneous least-squares fits of the mechanistic model to the five chromatograms at 5 cv, 10 cv, 30 cv, 60 cv and 120 cv (cf. section 2.3.2) for every single protein at pH 5 and pH 7 are shown for cytochrome c at pH 5 respectively at pH 7 in figures 3 and 4. The continuous line displays the experimental data, whereas the dotted line shows the model response for simulations with SMA parameter estimations from table 2. The fit of the model response to the gradiental elution data is for most of the datasets at both pH-conditions highly precise and was not corrupted by noise in the data, for example the small side-peak in the subfigure for 10 cv elution in figure 3. Only for short gradients with a length of 5 cv the model response slightly deviates from the data. The SMA parameters that were estimated by the inverse method are given in table 2. Structured multiple start guesses at the beginning of the optimization process (cf. equation (9)) provided the deviances in the optimization results that are given next to the estimated parameters in table 2. These deviances show, that the estimation deviance is dependent on the considered parameter. Thus, the parameters ν and k_{eq} are more determined, when estimated based on gradient elution data and a breakthrough, than the steric factor a fact, that was qualitatively shown earlier, for example in [17].

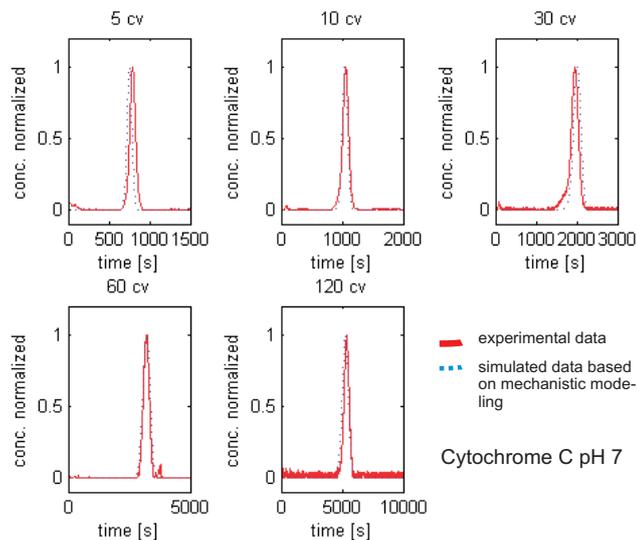


FIGURE 4. Example of inverse method fits based on mechanistic modeling. The experimental data of cytochrome c at pH 7 is displayed by a continuous line and the dotted line shows the inverse model-based fit.

parameter	ribonuclease A	
pH	5	7
ν	5.42 ± 0.09	3.25 ± 0.008
k_{eq}	0.037 ± 0.004	0.013 ± 0.0002
σ	28.5 ± 2.2	57.6 ± 0.24
Parameter	cytochrome c	
pH	5	7
ν	5.3 ± 0.03	4.09 ± 0.009
k_{eq}	0.094 ± 0.006	0.041 ± 0.0002
σ	29.8 ± 0.15	53.8 ± 0.28
Parameter	lysozyme	
pH	5	7
ν	5.07 ± 0.004	4.72 ± 0.01
k_{eq}	0.118 ± 0.0005	0.0372 ± 0.0005
σ	31.2 ± 0.19	38.75 ± 0.74

TABLE 2. SMA parameters determined based on mechanistic modeling (inverse method) [**approach II**]. The specific deviations on the estimations were determined by multiple start guesses.

The characteristic charges for the proteins at pH 5 are again close together, now in the limits from 5.07 to 5.42. They are also decreasing corresponding to the elution order of the proteins. For pH 7 the characteristic charges are again more distinct (ribonuclease A: 3.25, cytochrome c :

4.09, lysozyme: 4.72) and they are increasing corresponding to the elution order. The equilibrium coefficients lie between 0.013 (ribonuclease A at pH 7) and 0.118 (lysozyme at pH 5). At pH 5 and pH 7 they show an increasing trend corresponding to the elution order of the proteins. The steric factor of the proteins lies in the ranges between 28.5 (ribonuclease A at pH 5) and 57.6 (cytochrome *c* at pH 7). Apart from some outliers, the estimations for the steric factor and for the equilibrium coefficients show significantly higher mean relative deviations (about 0.5 %) than the estimations for the characteristic charge (about 0.25 %), but in general, the deviations on the estimated SMA parameters are very small.

4.4. Predictivity examinations on both approaches. For both approaches, retention times for gradient lengths of 25, 80 and 106 cv were predicted and the adequate experiments performed for every protein at pH 5 and pH 7. The predictions for approach I are based on interpolation [IP], the predictions in approach II are based on the solution of a system of differential equations [ODE]. No predictions outside the calibration range were examined, as the considered range with gradient lengths from 5 to 120 cv is very broad. Gradients outside this range might be of little use in practical applications. In table 3, the predictions and a posteriori experimentally determined retention volumes [in mL] for these gradient lengths are listed as well as the experimentally determined retention volumes [E]. The predictions for both approaches are very close to the experimental results; most of the deviances between prediction and validation are smaller than 1 mL, even for long elution gradients with 106 cv. Examples for the prediction of chromatograms based on the SMA parameters estimated by an inverse method are shown in figure 5. For cytochrome *c* at pH 5 the three subfigures on the top of the figure show predictions for gradiental elutions with gradient volumes of 25 cv, 80 cv and 106 cv (dotted line) and the experimental validation results (continuous lines); results for cytochrome *c* at pH 7 are shown in the three subfigures beneath. Obviously, data quality decreased slightly with increasing elution volumes, mainly due to baseline drift and peak broadening. The prediction quality is very high for both, small and large elution volumes, thus, independent of the noise.

4.5. Sensitivities for SMA parameters in approach I. In approach I the parameters ν and k_{eq} are determined simultaneously based on a correlation equation for gradient and retention volume (cf. equation (5)). Though the results are unique, the parameters estimation of k_{eq} is significantly influenced, when ν is fixed to a defined value while the change in the coefficient of determination is negligible. Figure 6 shows the results for a single determination of parameter k_{eq} , when different values for ν in equation (5) are set to be fixed. All remaining parameters in equation (5) were valid for lysozyme at pH 7. The coefficient of determination R^2 was close to 0.99 in every estimation and the residual in every estimation comparably small with differences in the third decimal place. A negative correlation between the parameter estimations could be observed. Small positive changes

	ribonuclease A					
pH	5			7		
	IP	ODE	E	IP	ODE	E
25 cv	16.55	16.59	16.45	7.68	8.27	7.32
80 cv	43.23	44.31	43.05	16.52	18.65	16.03
106 cv	54.51	55.22	55.45	19.8	22.18	19.27
	cytochrome <i>c</i>					
pH	5			7		
	IP	ODE	E	IP	ODE	E
25 cv	18.47	18.41	19	12.95	12.90	12.34
80 cv	48.12	50.83	50.23	31.32	32.73	30.15
106 cv	60.65	63.44	64	38.74	40.06	37.87
	lysozyme					
pH	5			7		
	IP	ODE	E	IP	ODE	E
25 cv	18.85	18.64	18.75	13.75	13.87	14
80 cv	48.63	50.25	49.32	34.33	35.39	34.8
106 cv	61.14	62.6	62.7	42.8	42.94	43.41

TABLE 3. Predictions and validation results of retention volumes [ml] for salt elution gradients with the lengths of 25 cv, 80 cv and 106 cv. The predictions entitled with [IP] are based on the correlation given by equation (5). Predictions from the mechanistic model are entitled with [ODE] and results from experimental validation with [E].

in the value of the characteristic charge ν induce significant negative changes in the estimation value of the equilibrium coefficient k_{eq} . For example, the characteristic charge determined by approach I (4.07) implies an equilibrium coefficient of 0.17 whereas a characteristic charge of 4.72, like it was determined by approach II, would result in an equilibrium coefficient estimation of about 0.06 (compare with the dotted lines in figure 6).

4.6. Sensitivities of SMA parameter estimation with respect to noisy data (approach II). The sensitivities of SMA parameter determination due to absolute noise on chromatograms, respectively, noise shifts in time, were examined by Monte-Carlo simulations. In figure 7 different levels of absolute noise on chromatogram concentration data are correlated with the relative deviation in the estimation of the steric parameter ν . The correlation is linear. It could be shown that even a strong absolute noise on chromatograms (± 0.05 M, original initial concentration: 0.0002 M) leads only to deviances in the second position after decimal point in the estimation of the characteristic charge ν . Similar correlations were observed for all of the SMA parameters. Thus, the application of absolute noise to the chromatograms had nearly no influence on the SMA parameter

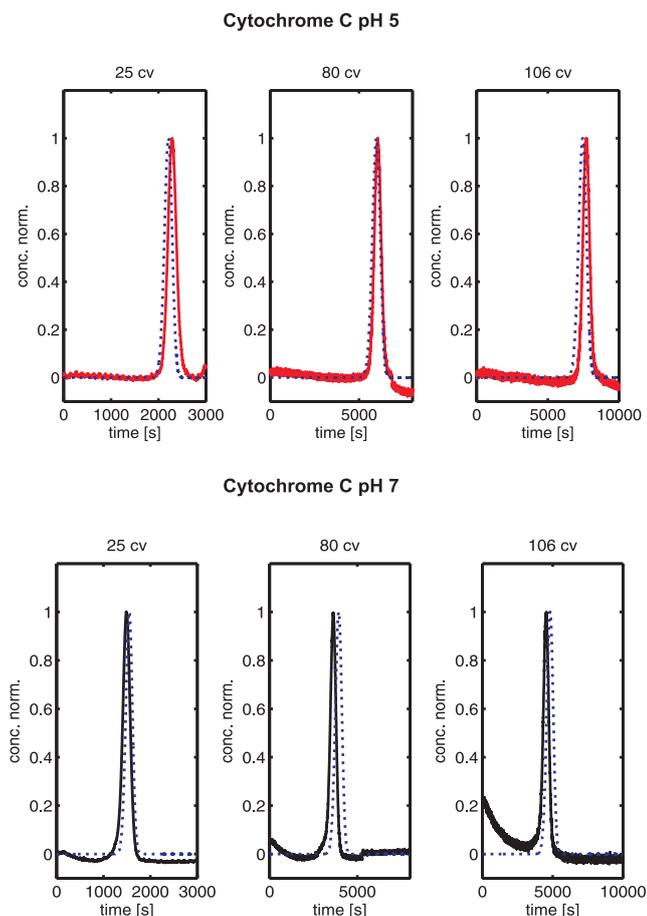


FIGURE 5. Prediction and experimental results of gradient elution data. The model-based prediction for the elution of cytochrome *c* at pH 5 is shown in the three subfigures on the top (dotted line) and the experimental results superimposed the predictions with a continuous line. Results for cytochrome *c* at pH 7 are shown in the subfigures at the bottom.

estimations by the inverse method. System-dependent noise (for example changes in tubing configurations) that leads to shifts in the retention time is more influential on the estimation of SMA parameters. The effects of this kind of noise on the estimation of SMA parameters ν , σ and k_{eq} are illustrated in figure 8. It becomes obvious that all SMA parameters are significantly more sensitive on retention time noise than on absolute noise on data measurements. ν (rel. std. deviation of 0.075 for peak shifts of about 50 s (0.43 mL)) is the least sensitive parameter to time-dependent shifts and σ (rel. std. deviation of almost 1.5 for peak shifts of about 50 s (0.43 mL)) the most sensitive parameter. These observations were valid for all three examined elution gradient lengths of 5 cv, 20 cv and 60 cv.

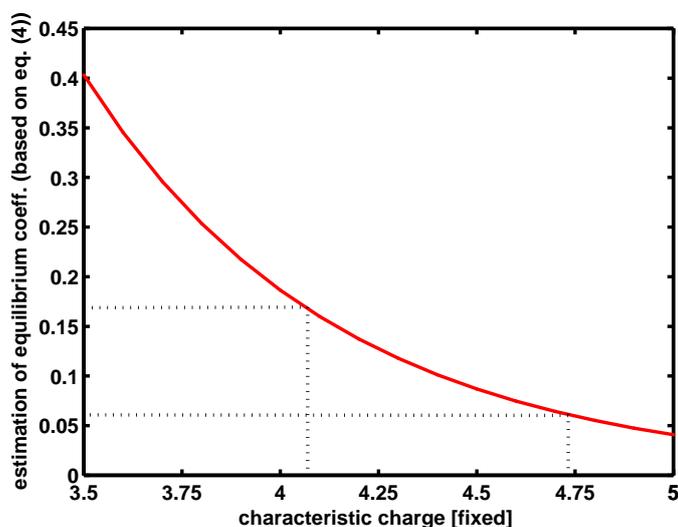


FIGURE 6. Results for the determination of SMA parameter k_{eq} in equation (5) in case of a fixed SMA parameter ν (based on data for lysozyme pH 7). The correlation curve shows the high sensitivity of equation (5) towards changes in ν .

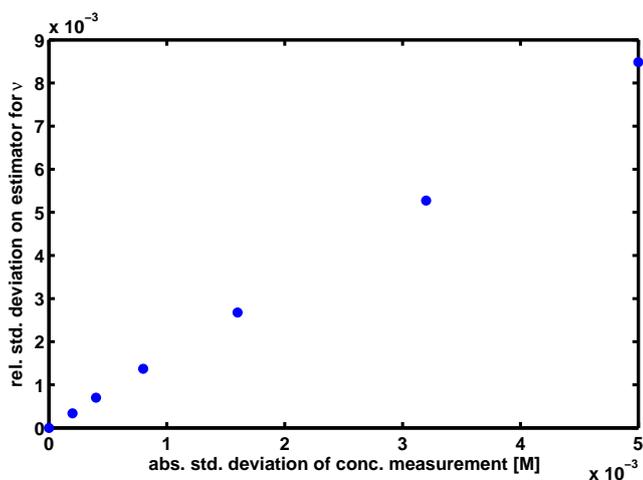


FIGURE 7. Illustration of the effects of **absolute noise on concentration measurement** in chromatographic data on the estimation of the SMA parameter of characteristic charge. This effect was determined for various noise levels (x-axis) by Monte-Carlo simulations.

4.7. Literature review. This literature review is based on [17], [29], [30], [31], [32] and other publications cited in this paper. Table 4 shows specific publications including SMA parameters for the proteins ribonuclease A, cytochrome *c* and lysozyme.

The literature review on the determination of SMA parameters shows that absolute values for SMA parameters are not only dependent on the examined protein and the pH value, but in the same way on the adsorbent and column properties like bed geometry or porosity. This is also obvious

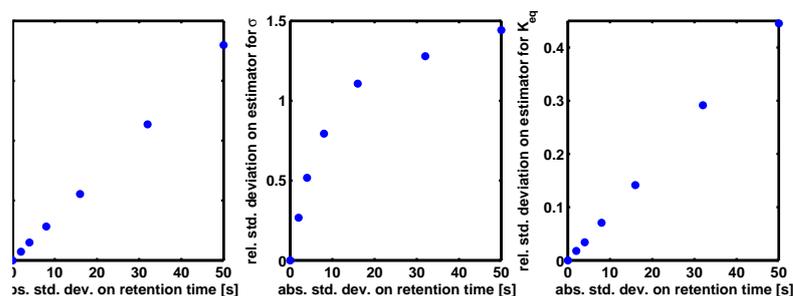


FIGURE 8. Illustration of the effects of **absolute shift noise on retention time** in chromatograms on the estimation of SMA parameters. This effect was determined for various noise levels (x-axis) for the three SMA parameters ν , σ and k_{eq} by Monte-Carlo simulations.

regarding the 'external information' that is necessary for all approaches of determination and can nicely be demonstrated by a comparison of two SMA parameter sets for α -chymotrypsinogen A, both determined at pH 6 on 40- μ m-Waters adsorbent (published by [33] [$\nu = 5.20 \pm 0.07$, $k_{eq} = 0.003$ and $\sigma = 45 \pm 3$] and [32] [$\nu = 4.8 \pm 0.17$, $k_{eq} = 0.0066$ and $\sigma = 52$]). These sets, although determined at the same pH-condition and the same adsorbent, deviate from each other for about 4 % (values for ν) up to about 14 % (values for σ). SMA parameters determined by different experimental approaches, deviate even more: For example in [7] and [16] the characteristic charges ν for lysozyme at pH 6 are 5.95 [isocratic elution experiments] respectively 4.97 ± 0.37 [gradiental elution experiments] and the values for k_{eq} 0.124 respectively 0.91 ± 0.16 .

This shows that even though in [14] the meaning of the SMA parameters is described with respect to protein characteristics like binding sites and shielding. their real virtue comes forth in comparative analysis, what has been indirectly shown before by affinity rankings in [34]. These rankings illustrate a method for the prediction of elution order based on the comparison of SMA parameter relations that seem to be valid for specific pH and similar adsorbent-systems. The intern order in SMA-values of proteins, for example $\nu_{cytc} > \nu_{ribA} > \nu_{achyA}$ at pH 6, coincided in almost every publication given in table 4 and the published SMA parameters for lysozyme, cytochrome *c* and ribonuclease A on cation exchange adsorbents, show mostly the same order relations to the parameters determined in this paper.

5. DISCUSSION

The intention behind the work presented in this study was to simply apply the SMA isotherm relationship as an equation describing ad- and desorption within a mechanistic chromatographic model; the latter finding its way into current high throughput process development strategies. Thus the paper does not aim to mechanistically explain ad- and desorption processes, but to develop a strategy for calibration of a mechanistic model for model based process development in the light of

publication	pH	column	proteins	
Gadam, S. et al. (1993). Characterization of non-linear adsorption properties of dextran-based polyelectrolyte displacers in ion-exchange systems, J.Chrom, (630) 37-52	6	strong cation exchanger with sulfopropyl groups	cytochrome <i>c</i>	$\nu = 6$ $k_{eq} = 0.0106$ $\sigma = 53.6$
			lysozyme	$\nu = 5.3$ $k_{eq} = 0.0148$ $\sigma = 34$
Gallant, S. et al. (1995). Optimization of step gradient separations: Consideration of nonlinear adsorption. Biotechnology and bioengineering, 47(3), 355-72	6	strong cation exchanger with sulfopropyl groups	cytochrome <i>c</i>	$\nu = 6.15$ $k_{eq} = 0.00637$ $\sigma = 53.4$
			lysozyme	$\nu = 5.95$ $k_{eq} = 0.124$ $\sigma = 9.5$
Gallant, S. et al (1997). Productivity and operating regimes in protein chromatography using low-molecular-mass displacers. J.Chrom. A, 771, 9-22	6	Source 15S	cytochrome <i>c</i>	$\nu = 6$ $k_{eq} = 0.12$ $\sigma = 28$
			lysozyme	$\nu = 5.5$ $k_{eq} = 1.1$ $\sigma = 14$
Ghose, S. et al. (2001). Characterization and modeling of monolithic stationary phases: application to preparative chromatography. J.Chrom. A, 928(1), 13-23	6	UNO S6, strong cation exchanger	ribonuclease A	$\nu = 5.69$ $k_{eq} = 0.00335$ $\sigma = 118$
			cytochrome <i>c</i>	$\nu = 6.08$ $k_{eq} = 0.01239$ $\sigma = 125$
Ladiwala, A. et al. (2005). A priori prediction of adsorption isotherm parameters and chromatographic behavior in ion-exchange systems. Proceedings of the National Academy of Sciences of the United States of America, 102(33)	5	SP Sepharose FF	ribonuclease A	$\nu = 5.4$ $k_{eq} = 0.0296$ $\sigma = 17.2$
			cytochrome <i>c</i>	$\nu = 5.9$ $k_{eq} = 0.0295$ $\sigma = 15.8$
			lysozyme	$\nu = 5.6$ $k_{eq} = 0.0763$ $\sigma = 17$

TABLE 4. Literature overview - SMA parameters for the proteins ribonuclease A, cytochrome *c* and lysozyme in publications

industrial needs, habits and data available. The first step on this path is of course a comparison of established and newer methods for parameter estimation.

The gradiental and frontal experiments for ribonuclease A, cytochrome *c* and lysozyme at pH 5 and pH 7 show the typical behaviour and elution order of these proteins on SP Sepharose FF, already

examined for example in [35] (cf. figure 1). The experiments are qualified for the determination of SMA parameters with both approaches. In approach I, the parameters for the SMA-isotherm were determined along the instructions of [14] and [16]. A significant correlation between elution volume and retention volume can be observed ($R^2 = 0.99$, cf. figure 2) and the determined parameters (cf. table 1) have reasonable results compared to literature values published in the papers, listed in table 4. In approach II, the SMA parameters were determined by a model-based inverse method (cf. table 2). The fit of the model response to the gradiental and breakthrough data is very precise (cf. figures 3 and 4). The observed small deviations of the model fit to the data of elution gradients with 5 cv can be attributed to kinetic effects that were neglected in order to empossible the comparison of both approaches (only isothermal SMA parameters can be determined with approach I). The fit is significantly improved by fitting the data of gradients shorter than 30 cv to equation (1) (data not shown).

By the multistart method, deviances in parameter estimations based on the inverse method could be determined. The comparison of these deviances for different SMA parameters shows that the information for the estimation of the steric factor is not as precisely given in gradiental and frontal experiments as for the estimation of the characteristic charge. This observation very likely explains difficulties to determine the steric factor, respectively higher deviances of this parameter, for example reported in [17] and [33]. Two datasets (ribonuclease A at pH 5 and lysozyme at pH 7) show overall larger uncertainties for the parameter estimations. The reason for this are most probably higher deviances in the retention times in gradiental experiments, caused by time-lags and lot-changes between the experiments. By Monte-Carlo-based sensitivity analysis it could be shown that a significant reason for uncertainty in parameter estimations are time-dependent shifts in chromatograms (cf. figures 7 and 8). These shifts may for example appear, when experimental runs are performed with interruptions or other irregularities. However, absolute noise in chromatographic data has only negligible influence on the parameter estimation quality.

The sensitivity analysis and the comparison of SMA parameters from both methods (tables 1 and 2) shows that the influences of external information and the slightly different model structures lead to different parameter values. While the steric factor σ has very similar values in both approaches, the characteristic charge ν determined by the inverse method is always higher than the value determined by approach I. Conversely, the equilibrium coefficients determined by the model-based inverse method are about ten times smaller than the values based on approach I. These deviations are plausably explained by the findings in the sensitivity analysis of equation (5), displayed in figure 6. The high sensitivity of the equilibrium coefficient considering a fixed characteristic charge is most probably due to the parameter's position in the exponent. This position of the parameter might be omitted by using isocratic elutions and a log-log-plot, like it is described in [14]. Still a transformation of the original equation is necessary and sensitivity seems to be influenced by

this transformation: [22] describe, that small errors were amplified by the log-log nature of the ion-exchange isocratic retention model they employed. Thus, a thorough sensitivity assessment for both approaches would be desirable but was out of the focus of this manuscript, which was laid on parameter predictivity.

Despite of the differences in parameter values, the order relations between the parameters are conserved in both approaches, what can be observed by comparison of the results in tables 1 and 2. These differences can not be algorithm-specific as the same algorithm (`lsqnonlin` from MatLab) was used for the solution of implicate equations (equations (5) and (9)). This suggests that parameter **values** are dependent on the column characteristics and the employed model-specific equations, but the **order relations** of SMA parameters rely significantly on adsorbent type and pH-conditions. This assumption is confirmed by the literature review. Furthermore, comparing the two approaches, the quality of fit was very satisfying (cf. figures 2 and 4). Even more important is the fact that for both approaches it could be demonstrated that they provide parameters of comparable high predictivity (cf. table 3). The model-based approach has the important advantage of predicting complete chromatograms (cf. figure 5).

6. CONCLUSION AND OUTLOOK

It could be shown that SMA parameters with comparable internal relations and equal predictivity could be determined based on both methods, the experimental method according to [16] and an model-based inverse method. Considering the physical significance of SMA parameters, the intention for SMA parameter determination is the crucial argument. The authors do not deny the physical significance of the parameters and the good reasoning behind approach I and similar approaches based on [14]. Still, from the view of high qualitative predictions and model calibrations, the physical significance slightly shifts into background leaving place to the very important predictive power of the determined parameters, that was very high for both approaches. In addition, the inverse method has the obvious advantage of predicting complete chromatograms and best fit between model response and data. This leads to the possibility of further usage of the mechanistic model as troubleshooting and error diagnostic tool for the process. Thus, based on the findings in this paper, the inverse method for SMA parameter determination is recommended for fast process development. However, a certain amount and quality of data has to be provided for the inverse method for a precise determination of parameters and reliable predictions. This issue has to be further examined and the design of experiments optimized. With optimal experimental design and an *a priori* analysis of already existent data of the system of interest, it should be possible, to find an efficient way to estimate SMA parameters of equal predictivity directly from process data with no or only few additional experiments. Monte-Carlo simulations might support the analyses, as was shown in this study.

REFERENCES

- [1] F.D.A. U.S. Department of Health and Human Services (Ed.), Guidance for Industry PAT - A Framework for Innovative Pharmaceutical Manufacturing and Quality Assurance, 2004.
- [2] S. Ferreira, R. Bruns, E. da Silva, W. Dos Santos, C. Quintella, J. David, J. de Andrade, M. Breitzkreitz, I. Jardim, B. Neto, Statistical designs and response surface techniques for the optimization of chromatographic systems, *Journal of chromatography A* 1158 (1-2) (2007) 2–14.
- [3] G. Guiochon, Preparative liquid chromatography, *Journal of chromatography A* 965 (1-2) (2002) 129–61.
- [4] K. Beckley, P. Verhaert, L. van der Wielen, J. Hubbuch, M. Ottens, Rational and systematic protein purification process development: the next generation, *Trends in Biotechnology* 27 (12) (2009) 673 – 679.
- [5] E. von Lieres, J. Andersson, A fast and accurate solver for the general rate model of column liquid chromatography, *Computers & Chemical Engineering* 34 (8) (2010) 1180–1191.
- [6] S. Gallant, A. Kundu, S. Cramer, Optimization of step gradient separations: Consideration of nonlinear adsorption, *Biotechnology and bioengineering* 47 (3) (1995) 355–72.
- [7] S. Gallant, S. Vunnum, S. Cramer, Optimization of preparative ion-exchange chromatography of proteins: linear gradient separations, *Journal of Chromatography A* 725 (2) (1996) 295–314.
- [8] V. Natarajan, B. Bequette, S. Cramer, Optimization of ion-exchange displacement separations. i. validation of an iterative scheme and its use as a methods development tool, *Journal of chromatography A* 876 (1-2) (2000) 51–62.
- [9] N. Jakobsson, D. Karlsson, J. Axelsson, G. Zacchi, B. Nilsson, Using computer simulation to assist in the robustness analysis of an ion-exchange chromatography step, *Journal of chromatography A* 1063 (1-21-2) (2005) 99–109.
- [10] N. Jakobsson, M. Degerman, E. Stenborg, B. Nilsson, Model based robustness analysis of an ion-exchange chromatography step, *Journal of Chromatography A* 1138 (1-2) (2007) 109 – 119.
- [11] U. Altenhoener, M. Meurer, J. Strube, H. Schmidt-Traub, Parameter estimation for the simulation of liquid chromatography, *Journal of Chromatography A* 769 (1) (1997) 59 – 69.
- [12] G. Carta, A. Ubiera, T. Pabst, Protein mass transfer kinetics in ion exchange media: Measurements and interpretations, *Chemical Engineering & Technology* 28 (11) (2005) 1252–1264.
- [13] P. Persson, P. Gustavsson, G. Zacchi, B. Nilsson, Aspects of estimating parameter dependencies in a detailed chromatography model based on frontal experiments, *Process Biochemistry* 41 (8) (2006) 1812 – 1821.
- [14] C. Brooks, S. Cramer, Steric mass-action ion exchange: Displacement profiles and induced salt gradients, *AIChE Journal* 38 (1992) 1969 – 1978.
- [15] S. Gallant, A. Kundu, S. Cramer, Modeling non-linear elution of proteins in ion-exchange chromatography, *Journal of Chromatography A* 702 (1-2) (1995) 125 – 142, 1994 International Symposium on Preparative Chromatography.
- [16] A. Shukla, S. Bae, J. Moore, K. Barnhouse, S. Cramer, Synthesis and characterization of high-affinity, low molecular weight displacers for cation-exchange chromatography, *Industrial & Engineering Chemistry Research* 37 (10) (1998) 4090–4098.
- [17] H. Iyer, S. Tapper, P. Lester, B. Wolk, R. van Reis, Use of the steric mass action model in ion-exchange chromatographic process development, *Journal of Chromatography A* 832 (1999) 1–9.
- [18] Q. Shi, Y. Zhou, Y. Sun, Influence of pH and ionic strength on the steric mass-action model parameters around the isoelectric point of protein., *Biotechnology Progress* 21 (2) (2005) 516–23.

-
- [19] T. Barz, V. Loeffler, H. Arellano-Garcia, G. Wozny, Optimal determination of sma-parameters for the adsorption isotherms of β -lactoglobulin under non-isocratic conditions, *AIChE Journal* 38 (2009) 1969 – 1978.
- [20] H. Kempe, A. Axelsson, B. Nilsson, G. Zacchi, Simulation of chromatographic processes applied to separation of proteins, *Journal of Chromatography A* 846 (1-21-2) (1999) 1–12.
- [21] N. Titchener-Hooker, S. Chan, D. Bracewell, A systematic approach for modeling chromatographic processes - application to protein purification, *AIChE Journal* 54 (4) (2008) 965–977.
- [22] E. Parente, D. Wetlaufer, Relationship between isocratic and gradient retention times in the high-performance ion-exchange chromatography of proteins, *Journal of Chromatography* 355 (1984) 29 – 40.
- [23] G. Guichon, A. Felinger, D. Shirazi, A. Katti, *Fundamentals of Preparative and Nonlinear Chromatography*, Elsevier Academic Press, San Diego, USA, 2006.
- [24] H. Schmidt-Traub (Ed.), *Preparative Chromatography*, Wiley-VCH, Weinheim, Germany, 2006.
- [25] P. Danckwerts, Continuous flow systems: Distribution of residence times, *Chemical Engineering Science* 2 (1) (1953) 1–13.
- [26] N. Tugcu, S. Bae, J. Moore, S. Cramer, Stationary phase effects on the dynamic affinity of low-molecular-mass displacers., *Journal of chromatography A* 954 (1-21-2) (2002) 127–35.
- [27] B. Efron, R. Tibshirani, *An Introduction to the Bootstrap*, Chapman and Hall, CRC Press, London, Boca Raton, 1993.
- [28] M. Joshi, A. Seidel-Morgenstern, A. Kremling, Exploiting bootstrap method for quantifying parameter confidence intervals in dynamical systems, *Metabolic Engineering* 8 (2006) 447455.
- [29] A. Ladiwala, K. Rege, C. Breneman, S. Cramer, J. Prausnitz, A priori prediction of adsorption isotherm parameters and chromatographic behavior in ion-exchange systems, *Proceedings of the National Academy of Sciences of the United States of America* 102 (33) (2005) 11710–11715.
- [30] C. Mazza, S. Cramer, Evaluation of lot-to-lot consistency in ion exchange chromatography, *Journal of Liquid Chromatography & Related Technologies* 22 (111) (1999) 1733–1758.
- [31] C. Teske, E. von Lieres, M. Schröder, A. Ladiwala, S. Cramer, J. Hubbuch, High Throughput Screening for the Design and Optimization of Chromatographic Processes: Automated Optimization of Chromatographic Phase Systems, *Chemical Engineering & Technology* 32 (2009) 140–154.
- [32] V. Natarajan, S. Cramer, A methodology for the characterization of ion-exchange resins, *Separation Science and Technology* 35 (111) (2000) 1719–1742.
- [33] V. Natarajan, S. Cramer, Modeling shock layers in ion-exchange displacement chromatography, *AIChE Journal* 45 (1) (1999) 27–37.
- [34] J. Liu, Z. Hilton, S. Cramer, Chemically selective displacers for high-resolution protein separations in ion-exchange systems: effect of displacer-protein interactions, *Analytical chemistry* 80 (9) (2008) 3357–64.
- [35] 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 (2009) 140–154.

Optimizing a chromatographic three component separation: A comparison of mechanistic and empiric modeling approaches

A. Osberghaus¹, S. Hepbildikler², S.Nath², M. Haindl², E. von Lieres³ and
J. Hubbuch^{1,*}

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

² : Pharmaceutical Biotech Production, Roche Diagnostics GmbH, Nonnenwald 2, 82377 Penzberg, Germany

³ : Institute of Bio- and Geosciences 1, Research Center Jülich, 52425 Jülich, Germany

* : Corresponding author. *E-mail-address*:juergen.hubbuch@kit.edu (J. Hubbuch)

ABSTRACT

The search for a favorable and robust operating point of a separation process represents a complex multi-factor optimization problem. This problem is typically tackled by design of experiments (DoE) in the factor space and empiric response surface modeling (RSM); however, separation optimizations based on mechanistic modeling are on the rise. In this paper, a DoE-RSM-approach and a mechanistic modeling approach are compared with respect to their performance and predictive power by means of a case study - the optimization of a multicomponent separation of proteins in an ion exchange chromatography step with a nonlinear gradient (ribonuclease A, cytochrome c and lysozyme on SP Sepharose FF). The results revealed that at least for complex problems with low robustness, the performance of the DoE-approach is significantly inferior to the performance of the mechanistic model. While some influential factors of the system could be detected with the DoE-RSM-approach, predictions concerning the peak resolutions were mostly inaccurate and the optimization failed. The predictions of the mechanistic model for separation results were very accurate. Influences of the experimental factors could be quantified and the separation was optimized with respect to several objectives. However, the discussion of advantages and disadvantages of empiric and mechanistic modeling generates synergies of both methods and leads to a new optimization concept, which is promising with respect to an efficient employment of high throughput screening data.

Keywords: ion exchange chromatography, design of experiments, response surface modeling, mechanistic modeling, steric mass action (SMA), separation optimization

1. INTRODUCTION

Ion exchange chromatography (IEC) is a widely used application in biomolecular downstream processing. In IEC, the main focus is the separation of a target component from a protein mixture - preferably in a step elution, but complex separation problems may require linear or even nonlinear gradient elutions. In addition to the shape of the elution gradient, the quality of a separation depends on several process factors, among others the employed buffers and salts. Furthermore, the objectives of a separation step are not only defined by high yields and product purity, but also

by additional demands, such as process robustness, financial and ecological constraints. Thus, the optimization of a separation step is a multiparametric and multiobjective problem.

The approaches to tackle problems of this kind are various and can be roughly divided into search algorithms and modeling methods. Various successful applications of search algorithms for separation optimization have been published during the last 30 years. The application of simplex algorithms, for example, has been proved successful for example in [1], [2] and [3]. Recently, more robust search algorithms like neural network approaches (see in [4], [5] and [6]), simulated annealing (for example in [7]) and evolutionary algorithms (see [8] and [9]) have been successfully applied to optimization in chromatography. While the low mathematical effort of search methods and their high performance in noisy systems was demonstrated in these research publications, a critical drawback of search methods is given by the tremendous experimental effort and the low knowledge gain about the examined system, particularly about sensitivity and robustness aspects.

However, the importance of process understanding, as well as robustness and sensitivity analyses was only recently emphasized in guidelines, published by the US Food and Drug Administration [10]. Consequently, multivariate optimization approaches based on design of experiments (DoE) and empiric response surface modeling (RSM) are increasingly applied in bioseparation process development, because they allow for the characterization of design factor spaces and for the calculation of optimal system settings and their robustness. Similar to the application of search algorithms, first publications on the application of DoE-RSM in the field of ion exchange chromatography have been published in the eighties, for example by [11] who optimized a separation step in reversed-phase chromatography based on a full-factorial design. DoE-RSM techniques were successfully applied and further developed in chromatography studies, for example in [12] (full factorial design), [13] (block design and partial least squares-regression) and [14] (fractional factorial design, application of modeling software). Reviews on DoE-RSM methods, like in [15], comparison studies of regression algorithms [16] or the formulation of very specific regression functions like in [17] demonstrate that DoE-RSM is well established in the optimization of IEC steps.

Alternatively to this empiric modeling approach, the application of mechanistic modeling for the optimization of IEC steps is on the rise due to time efficiency of algorithms and increased calculation power (see argumentation lines in [18], [19]). Mechanistic modeling means to employ functional relationships between physical parameters in chromatography and retention times or even complete chromatograms. Important reviews on mechanistic modeling are for example [20], [21], [18] or [22]. Successful optimizations based on mechanistic modeling of IEC processes have been demonstrated for step gradients [23], linear gradients [24] and displacement systems [25]. Additionally, a validated model proves to be an accurate prediction tool and lends itself to application in process control, which was demonstrated for example in [26] and [27]. A drawback of separation optimization based on a mechanistic model seemed to be the very time consuming procedure of repeatedly solving the underlying partial differential equation system. However, recently a very efficient and time-optimized solver was introduced by [28] that allows for model-based optimization in a few minutes.

As a result, the optimization based on mechanistic modeling is now competitive to the DoE-RSM approach with respect to time efficiency. This will be the dooropener concerning research on the advantages and benefits of mechanistic modeling, especially compared to the established approaches in chromatography process optimization. As to the authors' knowledge there are no studies, where

both approaches are compared based on the same set of data. Too little information is available on the predictivity quality of mechanistic modeling in comparison to the DoE-RSM approach, inside and beyond the design space. Furthermore, to the best of the authors' knowledge, there is only little research on the performance of DoE in separation problems with low robustness. However, low robustness is very common in separation problems, as slight changes in the level of salt buffers in step or gradient elutions have significant influence on retention times and peak shapes. Considering this, the optimization performance of a model based on mechanistic understanding should exceed the performance of an empiric model. An important aim of the manuscript is to show, if the difference in performance is significant.

Another drawback of the previously cited optimization studies is the fact that they have mostly been limited to a fixed objective for the separation process. However, as shown before, separation issues are normally multiobjective or objectives are changed in the development of a chromatography process. Thus, approaches for optimization should be flexible with respect to changing objectives and should not demand for re-calibration.

In this paper, the DoE-RSM approach and the mechanistic modeling approach are compared with respect to the mentioned issues. After a theoretical comparison of both approaches, they are applied to a case study - the optimization of a multicomponent-separation in an IEC step. As two of the proteins have close isoelectric points (cytochrome *c*: 10.0 - 10.5, lysozyme: 11.35), a bilinear gradient that is a series of two linear salt elution gradients, was chosen for the separation step in analogy to Refs. [26] and [8]. Due to the bilinear gradient, this model system is rather complex and demands for robustness analyses. According to a D-optimal onion design, experimental data for optimization was planned and the chromatography runs randomly executed. Based on this randomly derived DoE-planned data, the RSM-approach as well as the approach of mechanistic modeling in IEC were used for determination of the factor effects on the chromatographic result and for optimization. Further, the additional effort for separation optimization with respect to changing objectives was analysed, as well as model predictiveness regarding factor sets beyond the original design space. The application of both modeling approaches to this case study allowed for an improved comparison of performance and effort with respect to multivariate separation issues.

2. THEORY

2.1. Response surface modeling and Design of Experiments. Response surface modeling (RSM) is a statistical technique for the a posteriori analysis of experimental data; a regression function of whatever nature - the response surface model - is fitted to the experimental results. Common applied response surface models in IEC have linear or quadratic complexity and are empiric (not mechanistic). Popular regression models are, for example, multivariate quadratic functions.

Let x_1, x_2, \dots, x_n be the n selected factors for process description and y_i the response/objective value to a specific factor setting $x_{1i}, x_{2i}, \dots, x_{ni}$. The regression fit of an n -variate quadratic function to the set of m responses y_1, \dots, y_m , is described by:

$$\begin{aligned}
 y_i = & a_1 + b_1x_{1i} + b_2x_{2i} + \dots + b_nx_{ni} \\
 & + c_1x_{1i}^2 + c_2x_{2i}^2 + \dots + c_nx_{ni}^2 \\
 & + d_{1,2}x_{1i}x_{2i} + d_{1,3}x_{1i}x_{3i} + \dots + d_{n-1,n}x_{n-1,i}x_{ni}
 \end{aligned} \tag{1}$$

for all $1 \leq i \leq m$. The parameter a_1 is a constant added to the function (see intercept terms in linear regression); furthermore, the values of the parameters b_k for $1 \leq k \leq n$ display the magnitude of linear influence of the factors x_k . The values of the parameters c_k with $1 \leq k \leq n$ quantify quadratic influences of the factors x_k and the mixed effects/interaction terms of two-components are quantified by the parameters d_{12} to $d_{n-1,n}$. A higher than quadratic complexity in the examined system leads to high prediction errors. However, high prediction errors give no hints as to the reasons in detail and no direction how to correct the model.

RSM is often behold as a DoE-technique, which is not correct. On the contrary, if a quadratic surface has to be fitted to the results, a well-selected DoE provides an adequate planning of the experiments. Thus, the idea of 'DoE' summarizes a diversified collection of statistical approaches for the maximization of specific information in experimental planning. The advantage of the DoE-RSM-approach, compared to simple screenings, is the provision of experimental designs with high information contents, quick information on reasonable factor ranges and first evidence of factor effects and system robustness.

Common and frequently used experimental plans are full-factorial designs or fractional-factorial designs. They deliver regular screening patterns over a factor space and provide the information for multilinear quadratic response models. Other designs meet special experimental constraints or a-priori-information on the system. For example, space filling designs are best when there is little or no information about the underlying effects of factors on responses while D-optimal designs guarantee high information in the single experiment by minimizing the covariance of the parameter estimates.

2.2. Theory on mechanistic modeling. A mechanistic model imitates the physical processes that occur in the observed system and describes them based on a set of mathematical equations. Thus, typical rate models for chromatographic processes contain convective and diffusive flows through a compressed pile of particles on the column level and the imitation of mass transfer resistances and surface interactions on particle level. In IEC modeling the transition of components from column to particle level is commonly modeled assuming a film; the sorption of protein on the particle surface can be imitated by the steric-mass-action-(SMA) model, developed by [29] and commonly used for the modeling of salt gradient elutions in IEC, for example in [30], [31] and [32]. For the solution of the whole differential-algebraic equation system, Danckwerts' boundary conditions were applied [33]. For more details on rate models see [18].

In this case study, the decision concerning the most reasonable model complexity was taken in favor of the lumped transport-dispersive approach. This model, including convective, dispersive processes, mass transfer resistances and the SMA model for sorption kinetics, was solved in MatLab[®] on a Dual Core Processor with 2.81 GHz in approximately 10 seconds with a density of 200 knots over the whole column length. That is a time span of reasonable brevity, since the model has to be solved hundreds of times in model-based optimization.

The time- and position-dependent change of concentration on column level for the i -th component, $\partial c_i / \partial t$, is described by Eq. (2). The first term on the right hand side of Eq. (2) describes the convective transport through the column, the second term the dispersive transport and the third term the transport through a film to the particle surface.

$$\frac{\partial c_i}{\partial t} = -u_{int} \frac{\partial c_i}{\partial x} + D_{ax} \frac{\partial^2 c_i}{\partial x^2} - \frac{1 - \varepsilon_c}{\varepsilon_c} \cdot \frac{3}{r_p} k_{eff,i} [c_i - c_{p,i}] \quad (2)$$

u_{int} denotes the interstitial velocity, ε_c the column porosity, r_p the particle radius and $k_{eff,i}$ the lumped film diffusion coefficient. D_{ax} displays the axial dispersion, more precisely, a combined effect of dispersion and diffusive processes, dispersion being eddies and all effects implied by three-dimensionality.

The time and position-dependent change of concentration on particle level for the i -th component, $\partial c_{p,i}/\partial t$, is analogously described by Eq. (3):

$$\frac{\partial c_{p,i}}{\partial t} = \frac{3}{\varepsilon_p r_p} k_{eff,i} [c_i - c_{p,i}] - \frac{1 - \varepsilon_p}{\varepsilon_p} \frac{\partial q_i}{\partial t} \quad (3)$$

with q_i denoting the concentration of particle-bound component i and ε_p the particle porosity. The first term on the right hand side of Eq. (3) displays the mass transfer to particle surface and the second term describes ad- and desorption processes on particle level, i.e. the interaction between mobile and bound phase.

For the description of ad- and desorption processes, the SMA approach was embedded into the mechanistic model. The model equations for n components ($n = 1[\text{salt}] + \text{number of protein components}$) are given by

$$\frac{\partial q_i}{\partial t} = k_{ads,i} c_i \bar{q}_1^{\nu_i} - k_{des,i} c_1^{\nu_i} q_i \quad i > 1 \quad (4)$$

$$\Lambda = q_1 + \sum_{i=2}^n \nu_i q_i \quad (5)$$

$$\bar{q}_1 = q_1 - \sum_{i=2}^n \sigma_i q_i \quad (6)$$

Eq. (4) expresses the time dependent change of the concentration of surface bound component i . $k_{ads,i}$ denotes the adsorption rate and $k_{des,i}$ the desorption rate. The parameter Λ (ionic capacity of the adsorbent) limits the available binding places and displays the rivalry between salt concentration q_1 and the other bound components q_i , $2 \leq i \leq n$ with their specific characteristic charges ν_i . \bar{q}_1 , the concentration of bound salt ions available for exchange with the protein, is given by the total salt ion concentration q_1 less the shielded ions determined by the protein specific steric factors (σ_i) in Eq. (6). If the assumption of rapid equilibrium is valid ($\frac{\partial q_i}{\partial t} = 0$), Eqs. (4), (5) and (6) can be linked to the SMA isotherm:

$$c_i = \left(\frac{q_i}{k_{eq,i}} \right) \left(\frac{c_1}{\Lambda - \sum_{i=2}^n (\nu_i + \sigma_i) q_i} \right)^{\nu_i} \quad i > 1 \quad (7)$$

where the parameter $k_{eq,i}$ is the ratio of ad- and desorption coefficient.

SMA parameters for the mechanistic model can be determined based on data from gradient and breakthrough experiments (compare [34]). The inverse method states a second, equally predictive approach and is directly based on process data and the mechanistic model (see [35]).

3. MATERIALS AND METHODS

3.1. Apparatus, column and software. The case study aims at an optimal separation of a three component mixture on the adsorbent SP Sepharose FF by bilinear gradients. The running buffer in all experiments was 20 mM sodium phosphate buffer at pH 7. The buffer for elution purposes contained additional 0.5 M NaCl. The three component mixture consisted of lysozyme (from chicken egg white, L651), ribonuclease A (bovine pancreas, R4875) and cytochrome *c* (equine heart, C2506) from Sigma (St. Louis, MO, USA) dissolved in the low salt working buffer to a concentration of $0.2 \cdot 10^{-3}$ M. Salts and 1 M NaOH for pH adjustment were purchased from Merck (Darmstadt, Germany). The chromatographic setup consisted of a prepacked HiTrap SP Sepharose FF 1 mL column and an Ettan LC system, both purchased from GE Healthcare (Buckinghamshire, United Kingdom). The software MODDE (Umetrics, Umeå, Sweden) was used for DoE and RSM handling. The software Matlab (The Mathworks, Natick, ME, USA) was used for the handling of the mechanistic model.

3.2. Gradient elution experiments. In all experimental setups the column was at first equilibrated with running buffer for 10 column volumes (cv). This step was followed by an automated sample load of 20 μ L protein mixture. Then the column was washed for another two cv, before a bilinear elution gradient was initiated. Every elution gradient was applied for exactly 30 cv and quit with 100 % high salt elution buffer, followed by a 5 cv high salt wash step. Conductivity and UV-absorbances at 280 nm and 528 nm were measured online at column outlet. Cytochrome *c* absorbs not only radiation at 280 nm but additionally at 528 nm; this extra information was taken into account when calculating the resolutions between the peaks and for the SMA parameter estimation by the inverse method (see sections 3.3 and 3.4). The flow rate was set constantly to 0.5 mL/min [ca. $0.22 \cdot 10^{-3}$ m/s] in every process step.

The specific shape of a bilinear elution gradient was given by three characteristic factors:

- initial proportion of elution buffer in the running buffer: **Start** [%]
- proportion increment of elution buffer in the running buffer at the end of the first part of the bilinear elution gradient: **Slope** [%]
- length of the first gradient of the bilinear elution gradient: **Length** [cv]

while the overall gradient length was set constant to 30 cv and the final salt concentration of the gradient to 0.5 M NaCl. Figure 1 shows how the shape of a bilinear gradient is defined by the three characteristic factors **Start**, **Length** and **Slope**. A manipulation of these factors influences the axis intercept, length and slope of the first gradient. For the second part of the gradient these characteristics are implicit, due to the fixed end point of the bilinear gradient at 30 cv and 100 % elution buffer (0.5 M NaCl). Figure 2 shows a typical chromatogram resulting from an experiment with a bilinear gradient. The preset shape of the gradient is depicted in the dotted line. In all following chromatograms, always the preset gradient shape will be shown and not the measured conductivity, as these measurements did not go into modeling. However, the comparison of the actual conductivity to the results from mechanistical modeling showed excellent consistency. The grey line depicts the absorption signal at 528 nm, which measures the concentration of cytochrome *c*. The black line depicts the absorption signal at 280 nm. The first peak corresponds to the concentration of ribonuclease A and the second peak is the sum signal for the concentrations of cytochrome *c* and lysozyme.

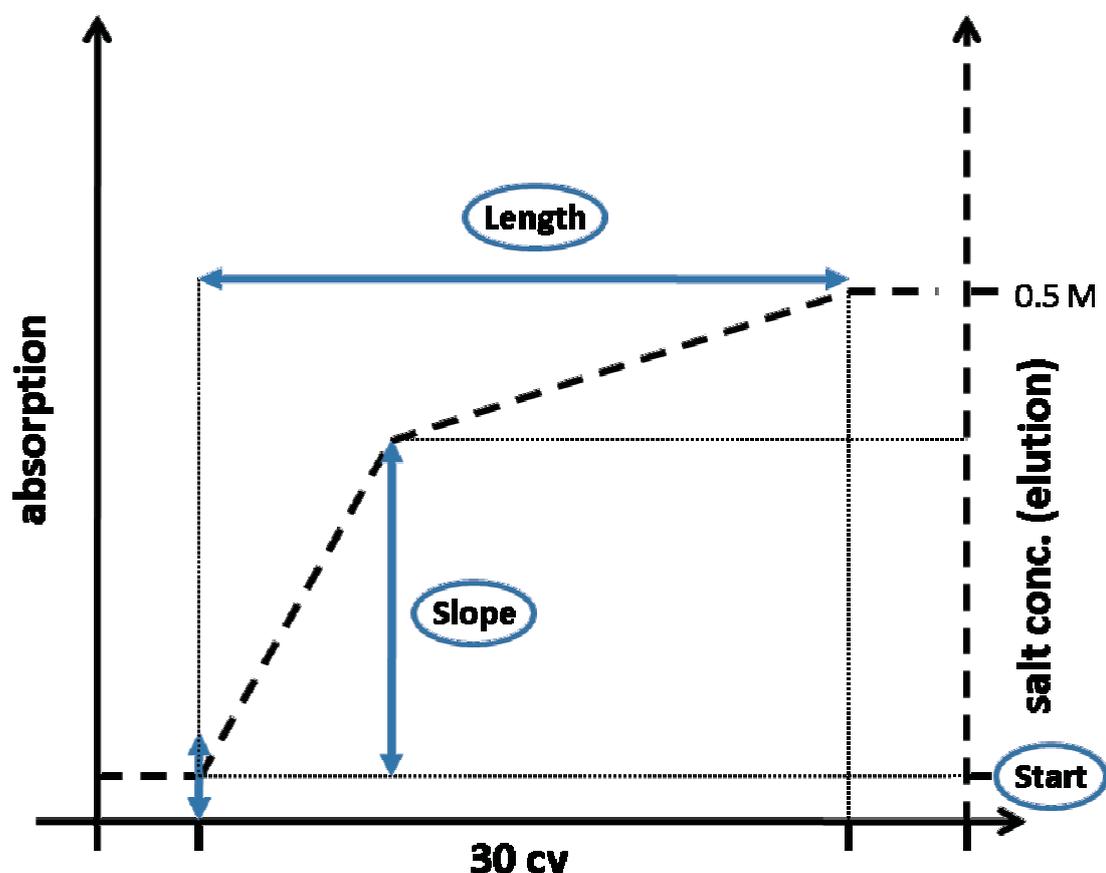


FIGURE 1. Three characteristic factors define the shape of the first gradient. The second gradient is determined by the total elution volume (30 cv) and the final salt concentration of 0.5 M NaCl.

3.3. The response surface modeling approach. The experiments were planned based on two D-optimal onion designs. Onion designs are space-filling designs recommended for situations where the factor correlations are not well known. The factor space is divided into layers around a center point, the number of layers and the layer setup is determined by optimality criteria. Onion designs provide information for nonlinear RSM and have the additional advantage of not having to perform experiments at the edges of the factor space (see [36] and [37] for more details). Although a significant smaller number of experiments would have been possible for a general DoE-RSM approach, a design with 32 measurement points was chosen in this case study. This choice was made in order to prevent a failure of the DoE-RSM-approach due to lack of information. The first design in the presented case study (onion design 1) proposed 29 factor sets in the ranges given in table 1 and a threefold repetition of a central point for check of reproducibility. The number and distribution of experiments in a layer fulfilled the criteria of G-optimality, minimizing the maximum variance of the predicted values. The distribution of measurements can be seen in figure 3. The ranges for the second design that was used in the case study are given in table 2. A single factor set, consisting of three specific values for **Start**, **Length** and **Slope**, described the unique shape of a bilinear gradient. All experiments, including the three center points, were performed in random order. The

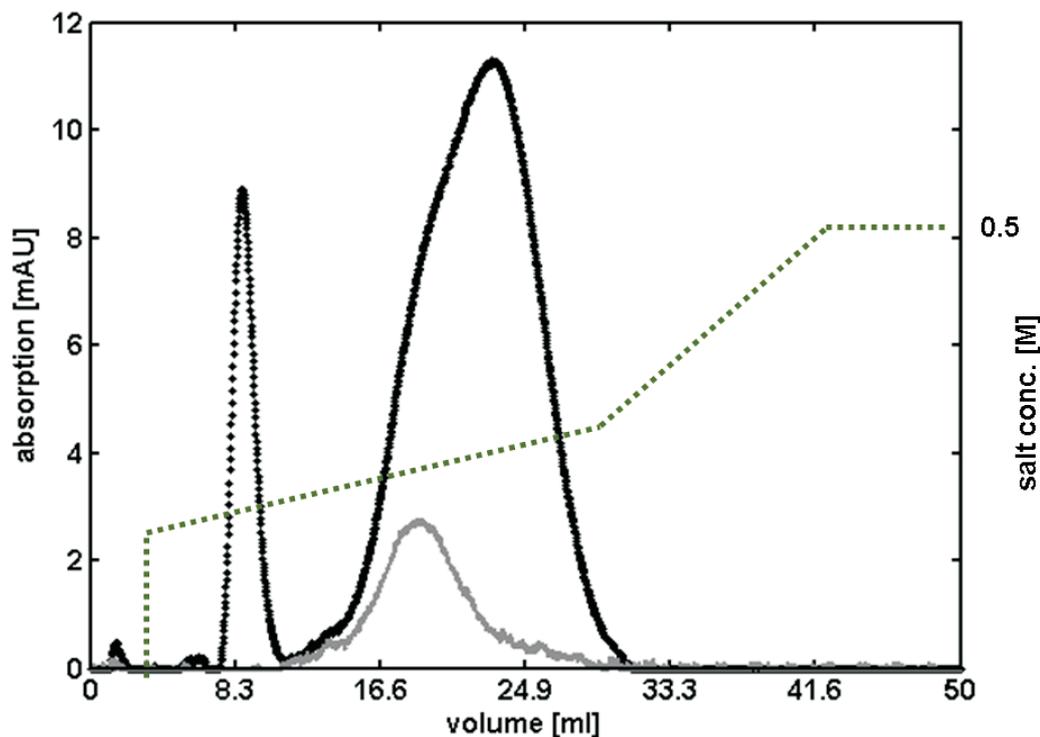


FIGURE 2. Typical chromatogram for the separation of ribonuclease A, cytochrome *c* and lysozyme with a bilinear gradient. The dotted line shows the settings for the elution gradient. Absorption at 528 nm (cytochrome *c*, grey) and 280 nm (all three proteins, black) is measured online and continuously.

factor	unit	range
Start	%	20 - 40
Length	cv	10 - 20
Slope	%	5 - 55

TABLE 1. Factor ranges for onion design 1

center point of onion design 1 was placed in a region, where good separation results were predicted by previous high-throughput screening studies on a robotic platform [8].

The DoE-RSM-approach establishes a functional relationship between the factors **Start**, **Length**, **Slope** on the one hand and the overall peak resolution on the other hand (see section 2.1). Consequently, the sum of the adjacent peak resolutions was selected to be the objective value that was to be maximized.

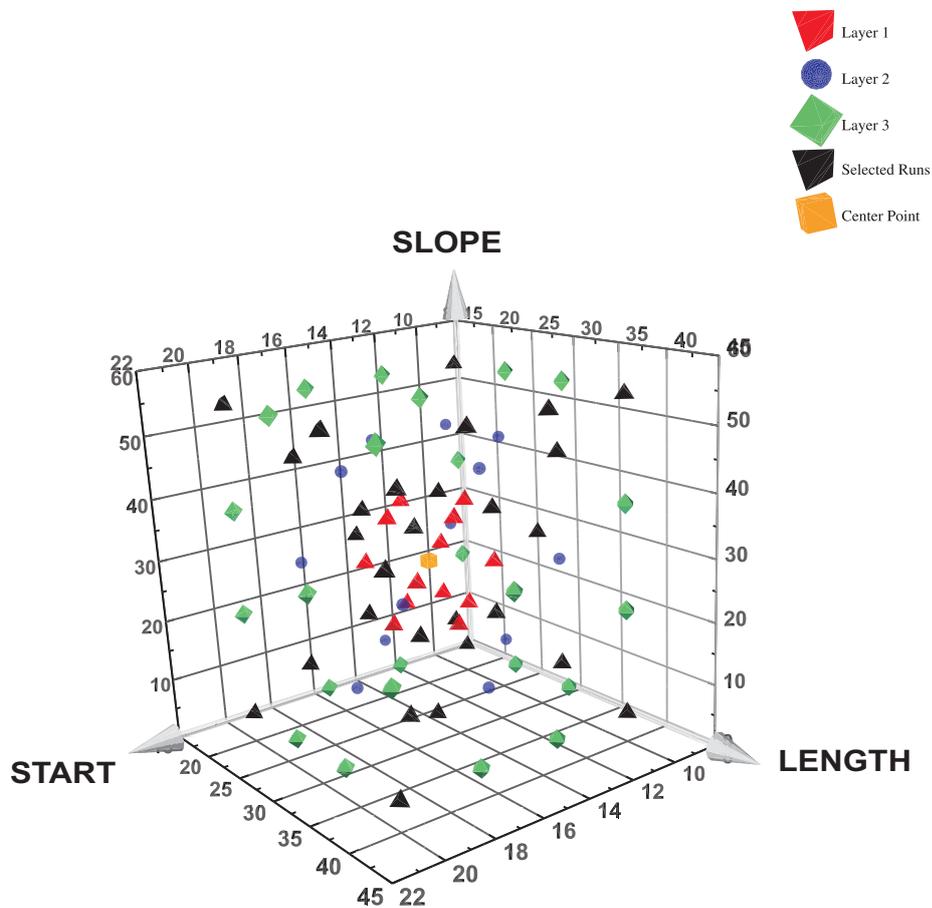


FIGURE 3. 3 D scatter plot of onion design 1

factor	unit	range
Start	%	5 - 25
Length	cv	15 - 30
Slope	%	0 - 15

TABLE 2. Factor ranges for onion design 2

The resolutions were calculated along Eq. (8) (compare to [38]):

$$res_{P1,P2} = \frac{2(\mu_{P1} - \mu_{P2})}{4\sigma_{P1} - 4\sigma_{P2}} \quad (8)$$

μ_{P1} , μ_{P2} and σ_{P1} , σ_{P2} are the characteristic first and second central moments, which describe the location and width of a peak. The peaks were deconvoluted based on the additional chromatogram for cytochrome *c* displaying the absorbance at 528 nm. In Eq. (8) two objectives (small peak width and large distance of retention times) are connected to an objective function. This is a common way to handle multi-objective problems: the objectives are combined and weighted in objective functions that have to be maximized or minimized. Choosing resolution as objective function, both models will optimize the resolution exclusively. In practice, the decision on the objective function is crucial and strongly situation-dependent.

Let f_{RSM} be the empiric model function, a multivariate quadratic function/response surface fitted to the resolution values. All coefficients of f_{RSM} have to be estimated by the inverse method, detecting estimators that induce the best fit of the response surface function to the resolution data. Then, the factor values at the maximum of this function are the characterization of a bilinear gradient that separates the three model proteins best, according to the selected objective, the maximal sum of resolutions.

3.4. The mechanistic model approach. Similar to the coefficients in RSM, the mechanistic model has parameters that have to be determined before model employment. The parameters of mechanistic models for chromatography are characteristic values describing the geometry of the column, the porosities of the packed bed, etc. (compare to section 2.2). In this study only the sorption parameters for the binding of protein to the adsorbent surface had to be established; all other model parameters had been determined beforehand in [35]. The sorption parameters were determined by the inverse method, shortly explained in the next paragraph.

3.4.1. Determination of SMA parameters. Let $c_{280}(t_j)$ be the time series of protein concentrations monitored by absorption at 280 nm at column outlet at the points in time $j = t_0 \dots t_{end}$. Let $c_{528}(t_j)$ analogously be the concentration of cytochrome *c* monitored at 528 nm at column outlet. Let $\hat{c}_{280}(t_j, \theta)$ be the solution of the described mechanistic model for chromatography for the sum of all three components concentrations at the same points in time. This solution is dependent from the modeling parameters θ . Let furthermore $\hat{c}_{528}(t_j, \theta)$ be the solution of the same mechanistic model for the component representing cytochrome *c*. Let θ_{fix} be the previously determined set of model input parameters and θ_{est} the set of model input parameters that are estimated based on all available data sets. Here, θ_{est} are the SMA parameters that have to be estimated for all three components, ribonuclease A, cytochrome *c* and lysozyme, based on 62 chromatograms (two 29 point onion designs with triple center points) corresponding to the factor settings $\theta_{grad,k} = \{\text{Start, Slope, Length}\}$, $1 \leq k \leq 62$. Aiming at a best fit between model response and chromatographic data, the problem of the inverse method can be stated as an optimization of a sum of least squares:

$$res(\theta_{est}) = \sum_{k=1}^{62} \sum_{j=0}^{end} [(\hat{c}_{280}(t_j, \theta_{fix}, \theta_{grad,k}; \theta_{est}) - c_{280}(t_j))^2 + (\hat{c}_{528}(t_j, \theta_{fix}, \theta_{grad,k}; \theta_{est}) - c_{528}(t_j))^2] \quad (9)$$

minimizing $res(\theta_{est})$. The minimization of Eq. (9) was in all cases performed with the Matlab[®]-procedure `lsqnonlin`.

3.4.2. Separation optimization. The case study aimed for an optimization of the separation regarding high resolution between adjacent protein peaks. The resolution between the peaks can be maximized by minimizing the peak overlaps. Let θ_{grad} denote the optimizable parameters, the three factors **Start**, **Length** and **Slope**. For the numerical optimization of the separation process, the unknown parameters $\theta_{optgrad}$ inducing the gradient of least overlap, are the solution of following minimization problem:

$$res_{12}(\theta_{grad}) + res_{23}(\theta_{grad}) + res_{13}(\theta_{grad}) \longrightarrow min! \quad (10)$$

with

$$res_{k,l} = \sum_{j=t_0}^{t_{end}} (\min(\hat{c}_k(t_j, \theta_{fix}, \theta_{grad}), \hat{c}_l(t_j, \theta_{fix}, \theta_{grad}))) \quad (11)$$

$\hat{c}_i(t_j, \theta)$ being the concentration profile/chromatogram for component i calculated by the mechanistic model, θ_{fix} are the mechanistic model parameters that are fixed in the optimization (SMA parameters inclusive).

3.5. Comparison of approaches. The comparison between the described modeling approaches has to be qualitatively as slightly different objective functions have been chosen (compare Eqs. (8) and (10)). This choice has been made in order to keep close to real applications by choosing objective functions corresponding to the typical model response. While the sum of resolutions calculated by Eq. (8) is directly inserted into the DoE-RSM approach for model calibration the mechanistic model is calibrated based on the complete chromatograms that were previously transformed to a time series of concentrations. Thus, the response of the DoE-RSM approach will be a sum of resolutions while the response of the mechanistic model will be a complete noiseless chromatogram with a perfect baseline. To really perform a quantitative comparison with the same objective function on these different model responses, significant data transformations and studies on noise in chromatographic data would have been necessary. This was out of the scope of the manuscript. Although a direct quantitative comparison between the approaches is impossible, the authors are positive, that the shown results allow for a meaningful qualitative comparison and a qualified discussion on advantages and disadvantages.

4. RESULTS

The separation of the three component mixture (ribonuclease A, cytochrome c and lysozyme on SP Sepharose FF) was to be optimized. Two approaches, a response surface modeling approach on the one hand and a mechanistic modeling approach on the other hand, were employed for the separation optimization. Both approaches were based on DoE-planned experiments and were to be compared as to their optimizing and predictivity performance.

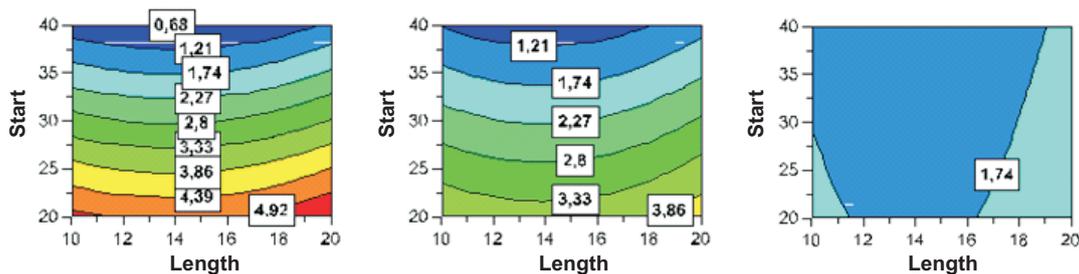


FIGURE 4. Surface contour plots based on 31 experiments from onion design 1. The factors **Start** and **Length** span the space, the factor **Slope** is illustrated in three levels: 5 (left-hand subplot), 25 (center), 55 (right-hand). The contour lines show the predicted values for the overall sum of resolutions in the design space.

4.1. Results of the response surface modeling approach. The plot in figure 4 is based on the experiments from onion design 1 (see table 1) and shows the response surface regression of the three explanatory factors (**Start**, **Length** and **Slope**) to the resulting sum of peak resolutions, denoted on the level curves. The three subplots illustrate three levels of the factor **Slope**. The optimal region of factors in the design space leading to an overall sum of resolution close to 5, is located in the bottom right-hand corner of the left-hand side subplot. Thus, an initial concentration of about 20% salt in elution buffer, a gradient length of the first gradient of 20 cv and a flat slope are predicted to lead to optimal separation results. The supply of predictions over a region that as far as possible surrounds the optimal process conditions is necessary for information on robustness. To keep the approach simple and follow typical procedures, the previously used DoE settings were applied for a second time with slightly shifted and enlarged ranges. Subsequently, the whole set of 62 results was analysed. The experimental reproducibility rep was 0.97, calculated by the variation at the center points compared to the total variation of the responses:

$$rep = 1 - \frac{\frac{1}{2} \sum_{i=1}^3 (cp_i - \overline{cp})^2}{\frac{1}{62-1} \sum_{i=1}^6 2(x_i - \bar{x})^2} \quad (12)$$

where cp_i denote the objective values of the threefold repeated center points, x_i the responses of 62 data points (including the center points) and the bar over a variable implies its mean value. To these data the RSM-method was applied. A quadratic model as initial modeling guess is the most common approach in RSM. In a data-based model discrimination, the model with highest coefficient of determination and with no non-significant parameters (p -value ≤ 0.05) was chosen. The best-fitting response surface was a quadratic model function with interaction terms. The regression fit itself had an adjusted coefficient of determination R^2 of about 0.78, what already indicates a limited predictivity. The coefficient plot in figure 5 shows the scaled and centered coefficients for the factors having most effect on the separation result and the factor interactions. The analysis shows that a long first part of the elution gradient induces positive effects on the resolution between the three protein components. This effect increases considerably with rising gradient length, as the coefficient for '**Length*Length**' is positive. The factor **Slope** has a slightly positive effect on the resolution of the peaks. Conversely, an increasing salt concentration at gradient begin (factor **Start**), has a negative effect on the overall sum of peak resolutions. These results suggest that gradients with a

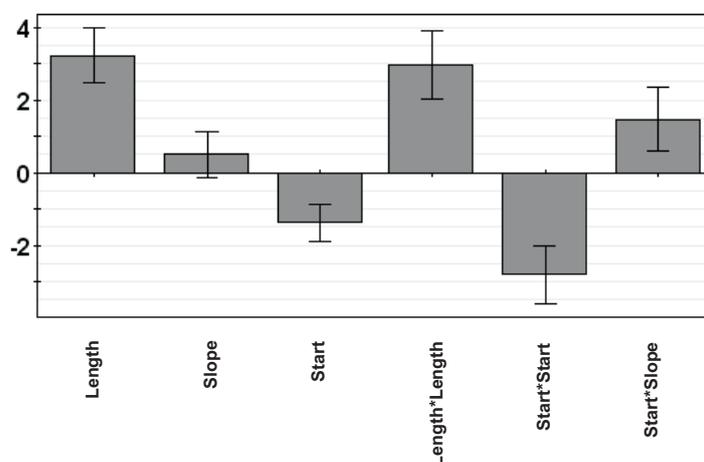


FIGURE 5. Coefficient plot for the response surface regression of the three gradient defining factors **Start**, **Length** and **Slope** to the sum of resolutions between the peaks.

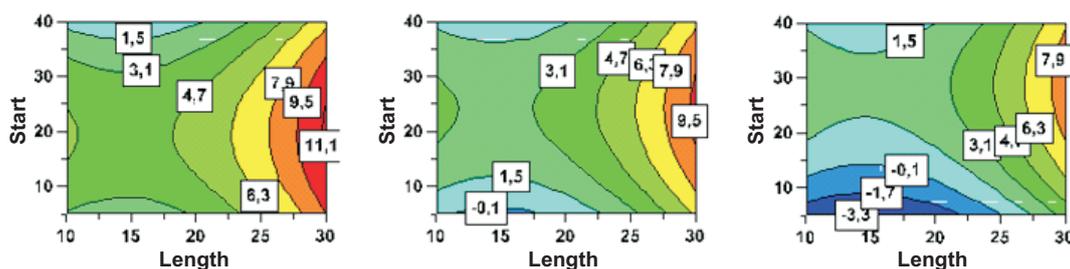


FIGURE 6. Surface contour plots based on all 62 experiments. The factors **Start** and **Length** span the space, the factor **Slope** is illustrated in three levels: 0 (left-hand subplot), 27.5 (center), 55 (right-hand). The contour lines show the predicted values for the overall sum of resolutions in the design space.

gentle slope and a low salt concentration at gradient beginning were most successful with respect to the separation problem. In addition, a positive interaction between **Start** and **Slope** with regard to high resolutions is predicted by the model. This interaction effect can be explained by the fact that the combination of both parameters mainly decides on the slope of the first gradient.

The contour plots in figure 6 are based on all results from onion design 1 and onion design 2. The three subplots illustrate three levels of the factor **Slope**. The optimal set of factors with respect to a high resolution is located at the right-hand side of the left-hand side subplot. Thus, an initial concentration of about 20% salt in elution buffer and a gradient length of 30 cv together with a very even slope showed the best results. Though placed at the border of the design space, the optimal gradient length of 30 cv could not be further optimized, since the maximum possible gradient length was fixed to 30 cv. The optimal region is very small and the gradient within the contour

factor	maximal resolution	medium resolution	minimal resolution
Start	10	19.26	40
Length	30	10	26
Slope	0	2.08	55
predicted resolution	9.46	4.8	-3.3
exp. determined resolution	0.1	4.2	2.8

TABLE 3. Predictions for factor sets inducing maximal, medium and minimal overall resolution in the three component system. These predictions are based on the **DoE-RSM**-approach.

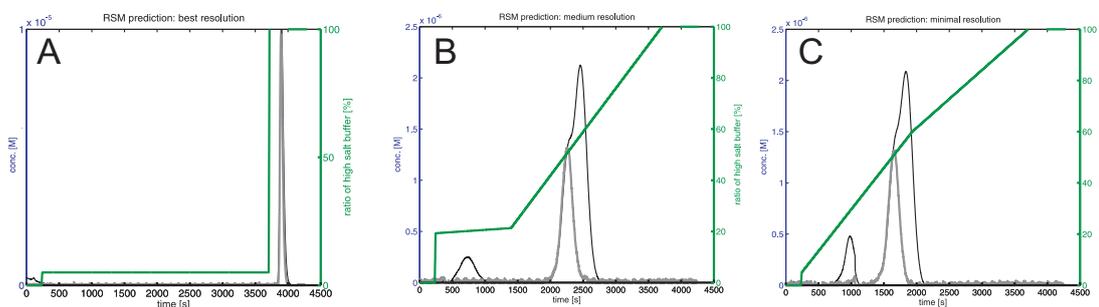


FIGURE 7. Experimental results for the RSM-based predictions for a maximal (subfigure A), a medium (subfigure B) and a minimal (subfigure C) resolution in the multicomponent separation of ribonuclease A, cytochrome *c* and lysozyme on SP Sepharose FF (see table 3). The salt gradient is displayed as ratio of elution buffer. In the chromatograms the black continuous line displays the total protein concentration and the grey line the concentration of cytochrome *c*.

plot is step. This indicates a low robustness of the examined system. Small changes in the elution gradient will have significant effects on the separation quality.

Three quantitative RSM-based predictions for the optimal set of factors with respect to a maximal, a medium and a minimal overall sum of resolutions are given in table 3. RSM predicted a maximal overall resolution for a gradient of 30 cv with constant 10 % high salt elution buffer, a medium resolution for the factors listed in the second column and a minimal resolution between the peaks for the factor values in the third column. As these predictions are partly based on a response surface extrapolation, the negative resolution value in the third column of table 3 is to be regarded as a tendency to a poor resolution. The experiments with factor settings from table 3 were performed based on the instructions given in section 3.2. Figure 7A shows the experimental results for the factor set in the first column of table 3. Based on the DoE-RSM approach the optimal salt gradient was predicted to be a 30 cv long step at 0.05 M NaCl. The RSM-based prediction failed, as obviously Ribonuclease A, cytochrome *c* and lysozyme elute simultaneously in the high salt wash step, what results in a resolution of 0.1 and not 9.46 (see table 3) The experimental result for the factors corresponding to the prediction for medium resolution shows in fact a resolution of 4.2, a value, that is close to the predicted resolution of 4.8 (figure 7B). The experimental results for

	ν	k_{ads}	k_{des}	k_{eq}	σ (fixed)
ribonuclease A	1.6	6.2	22.16	0.27	30
cytochrome <i>c</i>	2.8	4.6	17.16	0.27	30
lysozyme	3.4	1.6	11.78	0.14	30

TABLE 4. Table of SMA parameters determined with the inverse method

factor	maximal resolution	minimal resolution
Start	25.19	100
Length	28.37	30
Slope	2.58	0

TABLE 5. Predictions for factor sets inducing maximal and minimal overall resolution in the three component system. These predictions are based on **mechanistic modeling**.

the factor set that was predicted to result in a minimal resolution (third column in table 3) show a sum of resolutions of 2.8 (figure 7C). This result has a correct trend, as it is small but obviously it is not minimal.

4.2. Results of the mechanistic modeling approach. The monitored absorbance curves of the experiments planned with two onion designs (see tables 1 and 2) were employed to determine the SMA parameters by an inverse method. As the steric factor σ had negligible influence on the fitting result in the ranges of 20 to 40, it was fixed during the optimization of Eq. (9) to a reasonable value of 30. The estimated SMA parameters for all three components are given in table 4. These parameters provided the best fit between model response and chromatograms monitored at the Ettan LC system. The first column shows the estimated characteristic charges for the three proteins, ribonuclease A (1.6), cytochrome *c* (2.8) and lysozyme (3.4). The ascending order of these values correlates with the elution sequence. In addition, adsorption and desorption coefficients were estimated - their ratio can be summed up to the equilibrium coefficient (4th column of table 4). The parameters in table 4 are in the limits of experimentally determined SMA parameters and are therefore reasonable (compare for example with parameters in [24], [39] and [35]).

The completely calibrated mechanistic model was then employed to predict optimal gradient factors leading to a maximal or minimal sum of resolutions (see section 3.4.2). The predicted factors for maximal and minimal resolution are presented in table 5. Particularly the prediction for a minimal overall resolution differs from the RSM-based prediction (see the third column of table 3).

The optimal gradient, according to the mechanistic model, begins with a concentration of 25.19 % of the elution buffer. The slope of the gradient is very smooth. It requires the addition of 0.01 % of high salt buffer per cv to the elution buffer. The gradient ends after 28.37 cv with 27.77 % of the elution buffer. The mechanistic model proposes, similar to the DoE-RSM-prediction, a step-like gradient for the best resolution between proteins. The corresponding experimental results to the factors given in table 5 are shown in figure 8A. The peak predictions for ribonuclease A and lysozyme

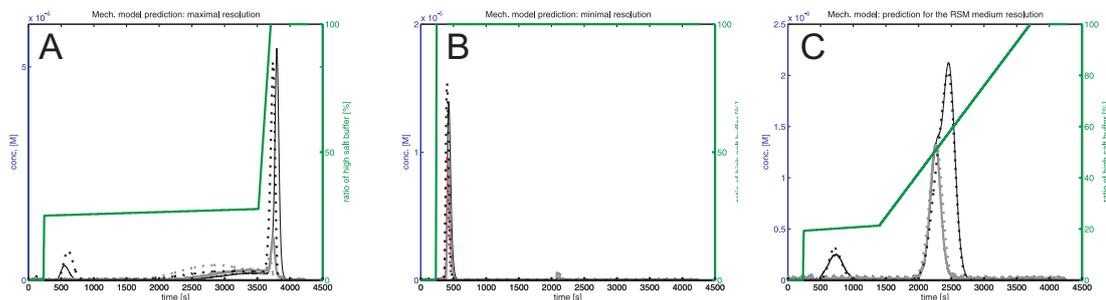


FIGURE 8. Results for the mechanistic model-based prediction for a maximal (subfigure A) and minimal resolution (subfigure B) for the separation of ribonuclease A, cytochrome *c* and lysozyme on SP Sepharose FF with a bilinear gradient. The continuous lines display the overall protein concentration (black) and the concentration of cytochrome *c* (grey). The dotted lines display the model-based prediction. Subfigure C shows the chromatogram for the RSM-predicted medium resolution superimposed with the highly accurate model prediction.

(first and third peak) are very accurate compared to the experimental data (continuous lines). The prediction for cytochrome *c* (peak in the center) is slightly shifted to reality. Nevertheless, the resolution between all protein peaks is high. The predicted factors for minimal resolution (2nd column of table 5) induce the chromatogram displayed in figure 8B. The resolution is obviously minimal, as salt concentration rises in a step from 0 to 100% and all proteins elute at once. In figure 8C again the experimental results for the RSM-predicted medium resolution are shown. In addition, the model response for this gradient is displayed with dotted lines. This figure gives a good example for the high predictivity of the mechanistic model.

Based on these encouraging results, a model-based prediction for a changed objective was employed: the optimization of the specific resolution between only cytochrome *c* and lysozyme. The optimal separation gradient with respect to the changed objective was predicted with the calibrated mechanistic model and then experimentally validated. Figure 9 shows the optimized gradient with the predicted chromatogram in dotted lines, whereas experimental data is given in continuous lines. Even though the prediction for cytochrome *c* was again slightly shifted, the favoured resolution between the cytochrome *c* and lysozyme peaks was high, due to this extraordinary gradient with a negative slope at its beginning. Very interesting is the fact that cytochrome *c* elutes in two parts - the major part of it elutes previously to the sharp bend in the gradient and the minor part afterwards. This can be explained by the fact that due to the low salt concentration at the end of the falling gradient, a certain proportion of protein molecules binds again to the column and is only eluted with the following rising salt concentration after the sharp bend in the elution gradient.

5. DISCUSSION

An empiric response surface modeling approach and a mechanistic modeling approach were compared and examined with respect to performance, predictivity and potential synergies considering the optimization of chromatographic separation processes. On the one hand, this comparison and evaluation was performed theoretically (see section 2), on the other hand a direct comparison of performance was achieved by applying both approaches to a case study: the optimization of bilinear

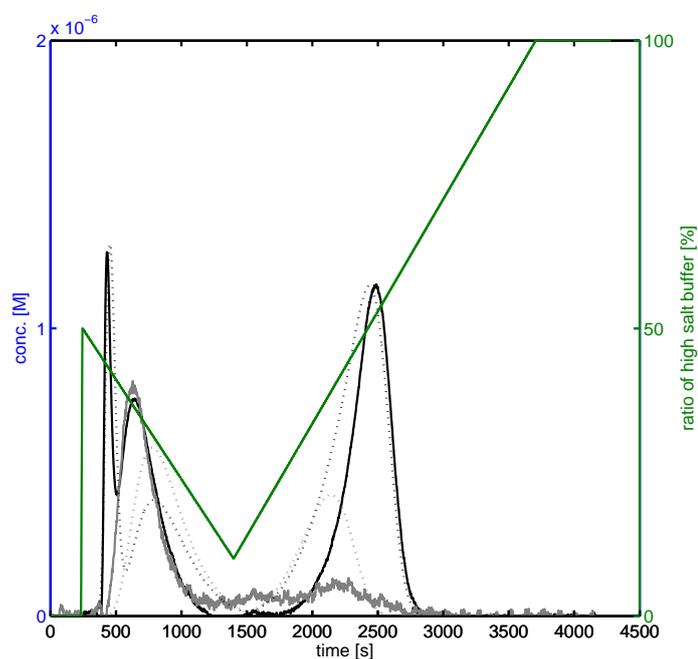


FIGURE 9. Result for a model-based optimization of the bilinear gradient considering a maximal resolution between cytochrome *c* and lysozyme. The predicted chromatogram (total protein conc.: black dotted line, cyt. *c* conc.: grey dotted line) and the experimental data (continuous lines) are superimposed.

elution gradients for the separation of a three-component mixture. Two of the components had a very similar *pI*, what increased the problem's complexity.

62 experiments were planned, based on two D-optimal onion designs (see tables 1 and 2). The high number of experiments was necessary, because the results of the first design (cmp. figure 4) showed the optimal settings to lie at the border of the design space. This situation should be possibly omitted with regard to higher predictiveness of the RSM-model. This demonstrates the difficult task of design space choice in the DoE-RSM-approach in many cases leading to multiple attempts. A quadratic model function with interaction terms was fitted to the results of both onion designs. This regression established a functional relationship between the objective function (sum of resolutions) and the gradient shape regulating factors **Start**, **Length** and **Slope**. The coefficient plot (see figure 5) revealed the most important factors and their influence on the objective function.

Based on this empiric model function and the modeling surfaces (see figure 6), the factor setups for different qualities of separation (maximal, medium and minimal resolution) were predicted (see table 3), the validation experiments were performed and results compared with the predictions (see figure 7). The analysis of the response surfaces showed that the examined system is significantly not robust, particularly close to the response surface's maximum. A small change in the shape of the bilinear gradient induces considerable effects on the separation. An example for this are the factors of the optimal gradient predictions of both approaches that are quite close to each other (compare the first columns of tables 3 and 5); however, the separation results differ significantly (figure 7A

and 8A). The predictivity analysis of the DoE-RSM-approach showed that the predictions had correct trends but were inaccurate, especially for extremal points. The DoE-RSM-based prediction for the maximal resolution of peaks failed. The same was true for the prediction with respect to a minimal resolution. An important reason for the failure of RSM in the prediction of the separation results, is the fact that the optimal factor set was located at the edge of the design space, where interpolation is more probable to fail due to the lesser number of reference points. The prediction for the minimal resolution probably failed, because extrapolations outside the ranges of the original design space are problematic for empiric RSM, as it only can predict continuous trends. In addition, a very important reason for the low predictivity is indicated by the low coefficient of determination $R^2 = 0.78$. Quadratic RSM can only handle up to quadratic complexity. The lack of fit shows that this system is definitely more complex. As the predictions were based on a model explaining only about 78 % of the variety in the experimental data, the probability to fail was increased.

The sorption parameters of the mechanistic model could be determined by the inverse method in this case study based on the DoE-planned datasets (see table 4). The calibrated chromatography model was employed for the numerical optimization of the separation problem. The elution gradient was optimized with respect to maximal and minimal overall resolution between the component peaks (see table 5). The validation experiments identified the mechanistic model to be successful and highly predictive (see figure 8). The optimal gradients were predicted correctly. While it seems to be obvious for experienced experimenters that an immediate step elution at 0.5 M gives no separation, this fact is not obvious for a model. The correct prediction in this case emphasizes the superiority of the introduced mechanistic modeling approach. The extrapolation of data beyond the borders of the underlying experimental design was possible, because the model is based on mechanistic processes in chromatography.

The prediction of cytochrome *c* data was slightly less accurate than the prediction of retention time and peak shape for the other components. This can be caused by protein-protein-interactions or other effects that were not considered in the modeling, like, for example pH-effects induced by the salt gradient. Without re-calibration, an optimal gradient for a high resolution between cytochrome *c* and lysozyme was calculated and the result showed again a high predictivity for extrapolated issues (see figure 9). This result could not have been so rapidly achieved with the DoE-RSM approach, as after the recalculation of the specific objective, the multivariate regression function would have needed re-calibration. Moreover, the optimal gradient was again located outside of the original design space, where the RSM-approach has a very low predictivity.

A separation optimization considering different pH conditions was no issue in this manuscript. Definitely a better resolution can be obtained at a different pH condition. Though there are a lot of promising approaches to this modeling issue in mechanistic chromatography models (see [40] and [41]), no approach is fully established. Thus, a comparison of modeling approaches including the optimization of pH conditions has not yet been made.

5.1. Conclusion and outlook. Two approaches for optimization in chromatography processes, a DoE-RSM approach and an approach based on mechanistic modeling, were to be compared based on their theoretical background and on their performance in a multicomponent separation process.

This comparison revealed advantages and disadvantages of both approaches. An advantage of the DoE-RSM-approach is the comfortable and quick calibration leading to reliable predictions with respect to simple correlations inside the design space. For example, correlations between salt concentration in the buffer at gradient start and the retention time of the first peak could be

predicted very accurately (data not shown). Another advantage of the DoE-RSM-approach is the easy identification of factor importance and influences on the objective as well as of the system's robustness.

Nevertheless, the DoE-RSM approach is significantly limited when dealing with complex chromatographic processes. The empiric multilinear model revealed a lack of fit and predictions with respect to an optimal elution gradient for separation failed due to this lack of fit and low robustness of nonlinear gradient elution processes. Results from extrapolation beyond the design space were not reliable as they showed large deviances to the correct results. This fact demands for re-calibration of the DoE-RSM-approach, whenever a new objective is chosen for optimization.

A disadvantage of mechanistic modeling, compared to simple screening methods and the DoE-RSM-approach, is the higher preliminary experimental effort with respect to model calibration and the need for efficient solution algorithms for the partial differential equation system. Nevertheless, it could be shown that the model could easily be calibrated based on the DoE-planned experiments. The predictions inside and beyond the design space were highly accurate and the optimization of the elution gradient was successful for various objectives. Re-calibration was not necessary. The knowledge gain with respect to the process was high, because all parameters are of mechanistic nature. Thus, the completely calibrated model could now be employed for similar separation problems.

The comparison of the two approaches for the optimization of chromatographic separation processes reveals synergies that could lead to new concepts of optimization. Based on these two approaches, an optimization could start with the DoE-RSM-based modeling, revealing factor importances and complexity of the problem. Additionally, this strategy allows for information on robustness issues, for first predictions concerning optimal factor settings and provides sufficient experiments for the calibration of the mechanistic model. The mechanistic model could be calibrated in the next step and be employed for accurate predictions on the process and for the handling of changing objectives as well as for quantitative robustness analysis and process monitoring. This concept will be applied and refined in ongoing research on various (industrial) processes.

REFERENCES

- [1] D. Fast, P. Culbreth, E. Sampson, *Clin. Chem.* 28 (1982) 3.
- [2] J. C. Berridge, *J. Chrom.* 485 (1989), pp. 3–12.
- [3] V. Drgan, D. Kotnik, M. Novič, *Anal. Chim. Acta* 705 (2011) 1-2, pp. 315–21.
- [4] J. E. Madden, N. Avdalovic, P. R. Haddad, J. Havel, *J. Chromatogr., A* 910 (2001) 1, pp. 173–9.
- [5] E. Marengo, E. Robotti, M. Bobba, M. C. Liparota, *Curr. Anal. Chem.* 2 (2006) 14, pp. 181–194.
- [6] T. Bolanča, Š. Cerjan-Stefanović, M. Luša, Š. Ukić, M. Rogošić, *Sep. Sci. Technol.* 45 (2010) 2, pp. 236–243.
- [7] K. Kaczmarek, D. Antos, *Acta Chromatogr.* 17 (2006), pp. 20–45.
- [8] A. Susanto, K. Treier, E. Knieps-Grünhagen, E. Von Lieres, J. Hubbuch, *Chem. Eng. Technol.* 32 (2009).
- [9] M. Max-Hansen, F. Ojala, D. Kifle, N. Borg, B. Nilsson, *J. Chromatogr., A* 1218 (2011) 51, pp. 9155–61.
- [10] F.D.A.U.S. Department of Health and Human Services (Ed.), *Guidance for Industry PAT - A Framework for Innovative Pharmaceutical Manufacturing and Quality Assurance*, 2004.
- [11] L. Pullan, *J. Liq. Chromatogr. Relat. Technol.* 11 (1988) 13.
- [12] W. Bachman, J. Stewart, *J. Chromatogr.* 481 (1989) 21.
- [13] M. H. J. Bergqvist, P. Kaufmann, *Lipids* 28 (1993) 7.
- [14] A. Nguyen, T. Aerts, D. Van Dam, P. De Deyn, *J. Chromatogr., B* 878 (2010).
- [15] S. L. C. Ferreira, R. E. Bruns, E. G. P. d. Silva, W. N. L. Dos Santos, C. M. Quintella, J. M. David, J. B. d. Andrade, M. C. Breikreitz, I. C. S. F. Jardim, B. B. Neto, *J. Chromatogr., A* 1158 (2007) 1-2.
- [16] D. Bylund, A. Bergens, S. Jacobsson, *Chromatographia* 44 (1997) 1-2.
- [17] P. Lebrun, B. Govaerts, B. Debrus, A. Ceccato, G. Caliaro, P. Hubert, B. Boulanger, *Chemom. Intell. Lab. Sys.* 91 (2008) 1.

-
- [18] G. Guiochon, *J. Chromatogr.*, A 965 (2002) 1-2.
- [19] B. Nfor, P. Verhaert, L. van der Wielen, J. Hubbuch, M. Ottens, *Trends Biotechnol.* 27 (2009) 12.
- [20] P. Jandera, J. Churáček, *J. Chromatogr.* 91 (1974), pp. 223–235.
- [21] J. Ståhlberg, *J. Chromatogr.*, A 855 (1999) 1, pp. 3–55.
- [22] J. Mollerup, *Chem. Eng. Technol.* 31 (2008) 6.
- [23] S. Gallant, A. Kundu, S. Cramer, *Biotechnol. Bioeng.* 47 (1995) 3.
- [24] S. Gallant, S. Vunnum, S. Cramer, *J. Chromatogr.*, A 725 (1996) 2.
- [25] V. Natarajan, B. Bequette, S. Cramer, *J. Chromatogr.*, A 876 (2000) 1-2.
- [26] N. Jakobsson, M. Degerman, E. Stenborg, B. Nilsson, *J. Chromatogr.*, A 1138 (2007) 1-2.
- [27] H. Kempe, A. Axelsson, B. Nilsson, G. Zacchi, *J. Chromatogr.*, A 846 (1999) 1-21-2.
- [28] E. von Lieres, J. Andersson, *Comput. Chem. Eng.* 34 (2010) 8.
- [29] C. Brooks, S. Cramer, *AIChE J.* 38 (1992).
- [30] N. Titchener-Hooker, S. Chan, D. G. Bracewell, *AIChE J.* 54 (2008) 4.
- [31] L. Pedersen, J. Mollerup, E. Hansen, A. Jungbauer, *J. Chromatogr.*, B 790 (2003) 1-2.
- [32] S. D. Gadam, G. Jayaraman, S. M. Cramer, *J. Chromatogr.*, A 630 (1993) 1-2.
- [33] P. Danckwerts, *Chem. Eng. Sci.* 2 (1953) 1.
- [34] A. A. Shukla, S. S. Bae, J. A. Moore, K. A. Barnhouse, S. M. Cramer, *Ind. Eng. Chem. Res.* 37 (1998) 1010.
- [35] A. Osberghaus, S. Hepbildikler, S. Nath, M. Haindl, E. v. Lieres, J. Hubbuch, *Journal of Chromatography A* 1233 (2012), pp. 54–65.
- [36] I.-M. Olsson, J. Gottfries, S. Wold, *Chemom. Intell. Lab. Sys.* 73 (2004) 1.
- [37] I.-M. Olsson, J. Gottfries, S. Wold, *J. Chemom.* 18 (2004) 12, pp. 548–557.
- [38] H. Vink, *J. Chromatogr.* 69 (1972).
- [39] A. Ladiwala, K. Rege, C. M. Breneman, S. M. Cramer, J. M. Prausnitz, *Proc. Natl. Acad. Sci. U. S. A.* 102 (2005) 33.
- [40] J. C. Bosma, J. A. Wesselingh, *AIChE J.* 44 (1998) 11, pp. 2399–2409.
- [41] T. Yang, M. C. Sundling, A. S. Freed, C. M. Breneman, S. M. Cramer, *Anal. Chem.* 79 (2007) 23, pp. 8927–39.

Model-integrated process development demonstrated on the optimization of a robotic cation exchange step

A. Osberghaus¹, K. Drechsel¹, S. Hansen¹, S. Hepbildikler², S. Nath², M. Haindl²,
E. von Lieres³ and J. Hubbuch^{1,*}

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

² : Pharmaceutical Biotech Production, Roche Diagnostics GmbH, Nonnenwald 2, 82377 Penzberg, Germany

³ : Institute of Bio- and Geosciences 1, Research Center Jülich, 52425 Jülich, Germany

* : Corresponding author. *E-mail-address*:juergen.hubbuch@kit.edu (J. Hubbuch)

ABSTRACT

A new concept for chromatography process development based on high-throughput data and mechanistic modeling will be presented in this paper. The concept is established in close cooperation between experimentation, modeling and model-based experimental design and allows for robustness analyses and upscale predictions. It will be demonstrated based on a case study: The optimization of a multicomponent separation (lysozyme, ribonuclease A and cytochrome c on SP Sepharose FFTM), subject to pH conditions and optimal settings for the shape of the elution gradient. Peak resolution and a precise prediction of retention times were chosen as performance variables in the case study to demonstrate the flexibility of the concept. It was shown that the concept of model-integrated process development is simple to perform from miniaturized scale on. The data, derived from model-based optimally designed experiments, provided sufficient information for process development, the model was calibrated and predictions for optimal separation setups as well as for the upscale showed a high precision. Consequently, the accumulation of data from high-throughput screenings can be used profitably for model-based process optimization and upscale predictions.

Keywords: scale-up, optimization, simulation, mathematical modelling, chromatography, downstream processing

1. INTRODUCTION

Chromatography is one of the main workhorses in bioseparation engineering and downstream processing of biomolecules. The development of optimal and robust chromatography processes is a central issue, because the costs for separations account for 50 - 80% of production costs [1]. In addition, the concept of Quality by Design (QbD), launched by the US Food and Drug Administration in 2004, pays increased attention to process robustness and reproducibility issues [2]. Optimization and robustness analyses in chromatography pose multivariate problems which are dependent on many factors directly influencing the retention times of proteins (e.g. pH conditions, resin, buffer). Thus, there has been a growing interest in approaches based on high-throughput experimentation, for example demonstrated by Wiehdahl et al. [3], and intelligent search algorithms like the genetic

algorithm employed by Susanto et al. [4], because these methods speed up the optimization and have the additional advantage of gain in knowledge on the examined system.

A highly knowledge-based approach is the simulation of chromatographic processes with mechanistic models. The application of mechanistic modelling in chromatography is on the rise due to extended computational power and increased memory capacities [5]. Optimal solutions for separation problems can be derived from mechanistic modeling, taking into account the physical processes on column and adsorbent particle level and leading to specific time-dependent concentration profiles for all calculated components at the column outlet. Thus, parameters of a separation problem can be changed and the effects on the chromatogram can be calculated *in silico*. Modeling has been successfully employed for the simulation and optimization of chromatography processes, for example in [6] and [7], as well as for robustness analyses and sensitivity analyses demonstrated by [8],[9].

The issue of efficient model calibration has been thoroughly addressed by [10], [11] and [12]. Concepts for model calibration including high-throughput batch binding experiments have been published in [13]. The establishment of these experiments on robotic platforms (see [14], [15]) allowed for an even quicker determination of static capacities and uptake kinetics. Considerable research has been devoted to the use of these data for chromatography modeling and scale-up predictions, for example in [16], [17] and [18]. Thus, along with an important study on model-based scale-up published recently by Gerontas et al. [19] a lot of strategies for an integration of modeling into process development have been established.

However, modeling based on data from high-throughput chromatography with RoboColumnsTM might be a promising way to integrate chromatography models even more efficiently into the first stages of process development. Based on RoboColumnsTM, reliable data for the scale-up can be obtained as was shown in [3]. So far, only a single study from Susanto et al. [20] uses data from RoboColumnsTM for modeling purposes. This is probably due to challenges in data from robotic systems, such as lower information densities in fractionation data and higher background noise; both effects leading to a more difficult identification and monitoring of proteins. Nevertheless, for successful process development it is of high priority to learn, how the accumulated data from high-throughput screenings can be used profitably for modeling and QbD concerns. New concepts for model calibration based on chromatography data from miniaturized scale are necessary. An efficient and seamless cooperation of different powerful process development tools, such as experimental design, modeling and the recently established methodology for selective protein quantification [21] might be a promising way to tackle optimization challenges.

In this paper a new concept for chromatography process development based on high-throughput data from RoboColumnsTM and mechanistic modeling will be presented. The concept is established through close cooperation between modeling and experimentation and allows for the optimization of chromatography steps and upscale predictions. The scheme for the concept, consisting of three basic steps, is shown in figure 1. In the first step of process development, pulse and frontal experiments are performed in order to characterize the column and the packed bed (porosity, dispersion, capacity). The results can be simultaneously used for the calibration of the mechanistic model equation that describes the flow through the column. In the second step of process development, frontal and gradient elution experiments support the decision on optimal pH conditions. These gradient elution experiments will be planned previously by mechanistic modeling, in order to contain increased information with respect to the calibration of the mechanistic model equation describing sorption

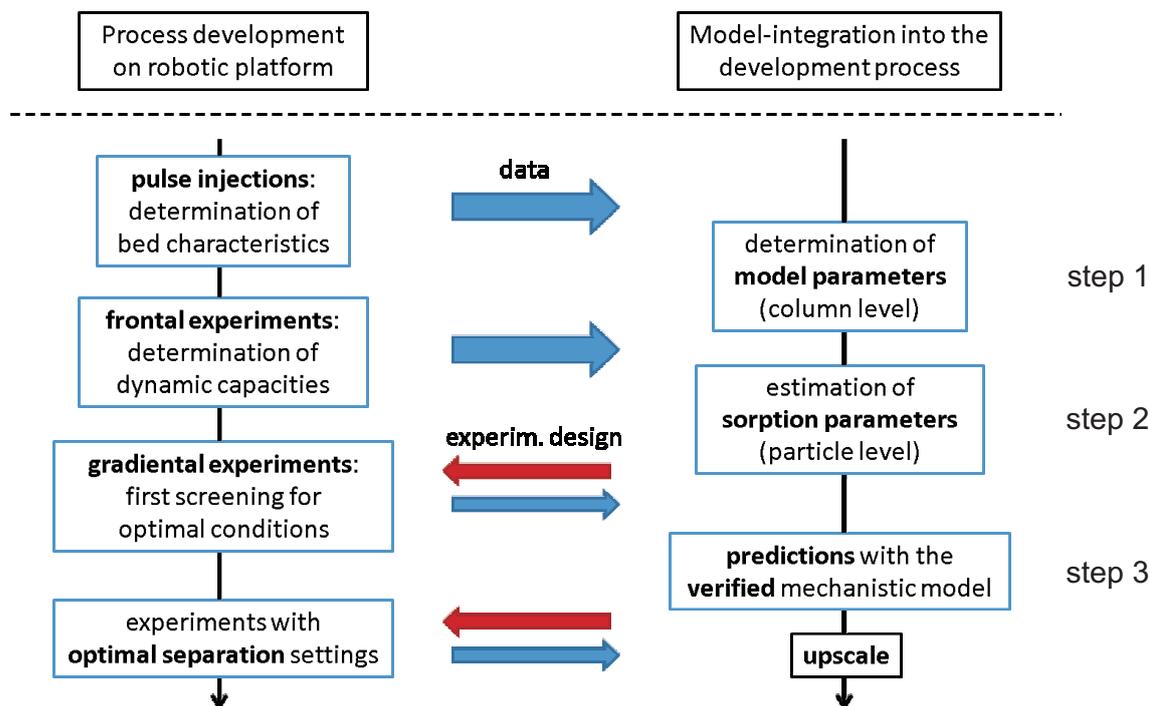


FIGURE 1. Concept for model integrated process development in three steps. Common experimentation in process development and the calibration of a mechanistic model for chromatography are tightly connected by model-based experimental design. This concept will be demonstrated along a case study - the optimization of a three component separation (ribonuclease A, cytochrome *c* and lysozyme on SP Sepharose FFTM)

processes in the particle. Based on the completely calibrated model, optimal separation gradients will be predicted *in silico* in the third step. Specific objectives, for example the resolution between the peaks or a certain retention behaviour, are chosen as performance variables for optimization. The planned experiments will be performed and model predictivity will be verified based on the experimental results by comparing peak shapes, resolutions and retention times. The verified model is then fit for upscale predictions.

This model-based integrated approach for process development has been developed stepwise as part of a general project on the chances and challenges of chromatography modeling based on high-throughput data. A survey on optimal parameter determination for model calibration on particle level provided an important basis [22]. This basis was complemented by a comparison of typical optimization approaches in chromatography separation [23]. The model-based integrated approach will be demonstrated based on a case study for development and optimization of a three component separation. A system of lysozyme, ribonuclease A and cytochrome *c* has to be separated on the adsorbent SP Sepharose FFTM subject to pH conditions and optimal settings for the shape of the elution gradient.

2. MATERIALS AND METHODS

2.1. Materials. Blue dextran 2,000,000, cytochrome *c* (horse heart), lysozyme (chicken egg white) and ribonuclease A (bovine pancreas) were purchased from Sigma (St. Louis, MO, USA). Acetic acid, 2-morpholinoethan sulfonic acid monohydrate (MES), sodium monobasic phosphate, sodium dibasic phosphate and sodium hydroxide for titration were purchased from Merck KGaA (Darmstadt, Germany). Acetone and ethanol (70 %, den.) were purchased from Carl Roth GmbH + Co.KG (Karlsruhe, Germany), MES Sodium Salt, ULTROL[®] grade from Calbiochem (Gibbstown, NJ, USA) and sodium chloride from AppliChem (Darmstadt, Germany). Three buffers were prepared, each with a concentration of 20 mM: pH 5 - acetate buffer, pH 6 MES buffer and pH 7 sodium phosphate buffer. To obtain high salt buffers for elution purposes, 0.5 M sodium chloride was added to the respective buffer.

2.2. Apparatus, columns & software. RoboColumns[™] for miniaturized chromatography on robotic platforms were purchased from Atoll (Weingarten, Germany) [0.01 m bed height, 0.005 m I.D.]. They were packed with SP Sepharose FF[™] (strong cation exchanger, mean particle diameter: 90 μm). The ionic capacity of the adsorbent (1200 mM) was determined with a syringe in triplicate by displacement measurements with acetic acid following furthermore the instructions in [24].

The miniaturized chromatography experiments were performed on a Freedom Evo 200 robotic platform purchased from Tecan (Crailsheim, Germany). The platform is equipped with a liquid handling arm for eightfold parallel pipetting, a plate gripper for device transport, a plate stacker module and a Te-Chrom device for RoboColumn[™] handling. The system dead volume considering column chromatography is negligible as the pipetting tips are placed directly on the filter above the adsorbent bed. The flow rate was 5 $\mu\text{L/s}$ (90 cm/h) in all experiments (except frontal experiments which were performed at 2.5 $\mu\text{L/s}$). The elution fractions with a volume of 67 μL each were collected into 96 well half-area plates from Greiner Bio-One (Kremsmuenster, Austria) and analyzed in the infinite M200 Reader from Tecan with two alternative protein quantification methods that are explained in section 2.3.4. The fraction volumes were monitored by 990-900 nm difference measurements. Freedom EVOware[®] 2.3 SP3 from Tecan (Crailsheim, Germany) was used for robotic control and Microsoft Office Excel 2007 (Redmond, WA, USA) was applied for data readout. All succeeding calculations, data manipulation, modeling and visualization were accomplished in Matlab[™] (The Mathworks, Natick, ME, USA).

2.3. Experimental methods.

2.3.1. Pulse injections and a shifting method for higher datapoint densities. For the determination of parameters characterizing the packed bed, pulse experiments with acetone (1 % in sodium phosphate buffer) and blue dextran 2,000,000 (5 mg/mL in sodium phosphate buffer) were performed on the robotic station. For a higher accuracy in retention time determination, the fractionation volume was the least possible what resulted in a high data point density. The fraction size of 67 μL is equivalent to a number of 2-4 drops from the column outlet. Nevertheless, data point density was not high enough to determine the retention time of very narrow peaks. Thus, two pulse experiments were performed successively and in the second experiment the fractionation times were shifted by 33.5 μL . All results were determined by measurements done in triplicate.

2.3.2. Frontal experiments. For the determination of dynamic bed capacities, frontal experiments were performed with lysozyme, cytochrome *c* and ribonuclease A at pH 5, pH 6 and pH 7 in the miniaturized system. After equilibrating the column for at least 5 column volumes (cv) with the respective buffer, protein solution ($0.4 \cdot 10^{-3}$ M) was injected with a volumetric flow rate of 2.5 $\mu\text{L/s}$

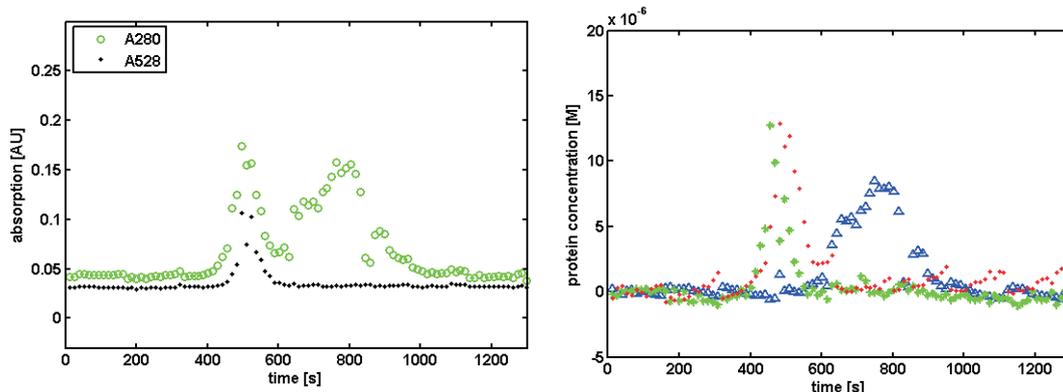


FIGURE 2. Comparison between protein quantification based on 280 nm absorption measurements and a new regression-based approach for high-throughput protein quantification [21]. Here, identical elution data are shown in comparison analysed by the two methods (elution of a threecomponent protein mixture consisting of lysozyme (triangles), ribonuclease A (stars) and cytochrome c (dots)).

until the breakthrough was complete (see instructions in [3]). The flowthrough was collected in fractions of $150 \mu\text{L}$.

2.3.3. Gradient elution experiments. All gradient elution experiments started with an equilibration phase lasting 5 cv, followed by an injection of $20 \mu\text{L}$ of the three-component protein mixture consisting of $0.2 \cdot 10^{-3} \text{ M}$ lysozyme, cytochrome c and ribonuclease A. The specific elution gradients were prepared beforehand by mixing equilibration and high salt buffer in parallelized mode in a way that every new concentration step corresponded to a new fraction. Due to the very small volumes of this premixed concentration steps ($67 \mu\text{L}$) the salt concentration gradients could be assumed to be linear in the mechanistic modeling. The gradients were defined by three design factors:

- salt concentration at gradient initiation
- salt concentration at gradient termination
- gradient volume [0 to 35 column volumes]

A 3 cv high salt wash step followed the procedure and the experiment was optionally terminated by preparing the columns for storage in a 20%-ethanol-solution.

2.3.4. Protein quantification. Protein concentrations in the collected fractions were analysed using two different methods:

- measurement of absorption at 280 nm and 528 nm (pulse and frontal experiments)
- equidistant measurement of absorption at 15 wavelengths between 240 and 300 nm (gradient elution experiments)

While the first method of absorption measurement is well established for overall protein quantification in a probe, the second approach was only recently developed [21]. It is based on the rapid quantification of single protein concentrations in a mixture by interpretation of spectral data with a precalibrated regression model and is suitable for high-throughput experimentation as protein quantification in a 96-well plate lasts merely 6 minutes. The precalibration of the regression model was performed by scanning spectra from training set data. The training set consisted of 32 solutions

with specific ratios of protein concentrations, designed by a four layer onion DoE (more details in [25]). 12 blank probes, containing only buffer, were added to calibrate the baseline location. Then, a multivariate regression model was fitted to the training set measurements with the MatlabTM routine `plsregress` considering four latent variables; the resulting regression coefficient matrix was used henceforth to determine protein concentrations in mixtures of unknown proportions. Figure 2 allows for a method comparison showing identical multicomponent elution data, evaluated using the two approaches. On the left-hand side the overall protein content in the fractions was determined by 280 nm absorption measurements, for the determination of cytochrome *c* by 528 nm absorption measurements. On the right-hand side single protein concentrations [lysozyme (triangles), ribonuclease A (stars) and cytochrome *c* (dots)] could be determined based on the regression model. With this method, the two peaks at elution beginning can clearly be discriminated and it is no longer necessary to perform the absorption data conversion using extinction coefficients.

2.4. Mathematical methods.

2.4.1. Introduction of the model equations. A mechanistic model imitates the physical processes that occur in the observed system and describes them based on a set of mathematical equations. Thus, typical rate models for chromatographic processes contain convective and diffusive flows through a compressed pile of particles on the column level and the imitation of mass transfer resistances and surface interactions on particle level. In IEC modeling the transition of components from column to particle level is commonly modeled assuming a film; the sorption of protein on the particle surface can be imitated by the steric-mass-action-(SMA) model, developed by [26] and commonly used for the modeling of salt gradient elutions in IEC, for example in [27], [28] and [29]. For the solution of the whole differential-algebraic equation system, Danckwerts' boundary conditions were applied [30]. For more details on rate models see [6].

In this case study, the decision concerning the most reasonable model complexity was taken in favor of the lumped transport-dispersive approach. Although axial dispersion and mass transfer to the particles can sometimes be neglected in protein chromatography, the size of RoboColumnsTM attaches importance to wall effects and mass transfer. Thus, the applied model includes convective, dispersive processes, mass transfer resistances and the SMA model for sorption kinetics.

The time- and position-dependent change of concentration on column level for the *i*-th component, $\partial c_i / \partial t$, is described by Eq. (1). The first term on the right-hand side of Eq. (1) describes the convective transport through the column, the second term the dispersive transport and the third term the transport through a film to the particle surface.

$$\frac{\partial c_i}{\partial t} = -u_{int} \frac{\partial c_i}{\partial x} + D_{ax} \frac{\partial^2 c_i}{\partial x^2} - \frac{1 - \varepsilon_c}{\varepsilon_c} \cdot \frac{3}{r_p} k_{eff,i} [c_i - c_{p,i}] \quad (1)$$

u_{int} denotes the interstitial velocity, ε_c the column porosity, r_p the particle radius and $k_{eff,i}$ the lumped film diffusion coefficient. D_{ax} displays the axial dispersion, more precisely, a combined effect of dispersion and diffusive processes, dispersion being eddies and all effects implied by three-dimensionality.

The time and position-dependent change of concentration on particle level for the *i*-th component, $\partial c_{p,i} / \partial t$, is analogously described by Eq. (2):

$$\frac{\partial c_{p,i}}{\partial t} = \frac{3}{\varepsilon_p r_p} k_{eff,i} [c_i - c_{p,i}] - \frac{1 - \varepsilon_p}{\varepsilon_p} \frac{\partial q_i}{\partial t} \quad (2)$$

with q_i denoting the concentration of particle-bound component i and ε_p the particle porosity, The first term on the right-hand side of Eq. (2) displays the mass transfer to particle surface and the second term describes ad- and desorption processes on particle level, i.e. the interaction between mobile and bound phase.

For the description of ad- and desorption processes, the SMA approach was embedded into the mechanistic model. In comparison to the Langmuir kinetic the SMA kinetic has the advantage of including the salt concentration into the sorption calculations. This advantage is crucial for the simulation of salt gradient elution experiments. The model equations for n components ($n = 1[\text{salt}] + \text{number of protein components}$) are given by

$$\frac{\partial q_i}{\partial t} = k_{ads,i} c_i \bar{q}_1^{\nu_i} - k_{des,i} c_1^{\nu_i} q_i \quad i > 1 \quad (3)$$

$$\Lambda = q_1 + \sum_{i=2}^n \nu_i q_i \quad (4)$$

$$\bar{q}_1 = q_1 - \sum_{i=2}^n \sigma_i q_i \quad (5)$$

Eq. (3) expresses the time dependent change of the concentration of surface bound component i ($\frac{\partial q_i}{\partial t}$). $k_{ads,i}$ denotes the adsorption rate and $k_{des,i}$ the desorption rate. The parameter Λ (ionic capacity of the adsorbent) limits the available binding places and displays the rivalry between salt concentration q_1 and the other bound components q_i , $2 \leq i \leq n$ with their specific characteristic charges ν_i . \bar{q}_1 , the concentration of bound salt ions available for exchange with the protein, is given by the total salt ion concentration q_1 less the shielded ions determined by the protein specific steric factors (σ_i) in Eq. (5). If the assumption of rapid equilibrium is valid ($\frac{\partial q_i}{\partial t} = 0$), Eqs. (3), (4) and (5) can be linked to the SMA isotherm:

$$c_i = \left(\frac{q_i}{k_{eq,i}} \right) \left(\frac{c_1}{\Lambda - \sum_{i=2}^n (\nu_i + \sigma_i) q_i} \right)^{\nu_i} \quad i > 1 \quad (6)$$

where the parameter $k_{eq,i}$ is the ratio of ad- and desorption coefficient.

SMA parameters for the mechanistic model can be determined based on data from gradient and breakthrough experiments (compare to [31]). The inverse method states a second, equally predictive approach and is directly based on process data and the mechanistic model (see [22]). A quick solution of the model equations is crucial, since the model has to be solved hundreds of times in model-based experimental design. The applied model, including convective, dispersive processes, mass transfer resistances and the SMA model for sorption kinetics, was solved in MATLAB on a Dual Core Processor with 2.81 GHz in approximately 10 seconds. The system of partial differential-algebraic equations was discretized in space using finite differences with 200 axial nodes.

The discretized right-hand side of the spatially discretized equation system was implemented as MatlabTM function. This function was integrated using the MatlabTM routine ode15s. An explicitly derived system Jacobian, i.e. the derivative of the right-hand side with respect to the state variables,

was implemented as a separate function for speeding up ode15s. More details on discretization and solution of the model equations are given in [32].

2.4.2. *The inverse method for SMA parameter estimation.* Let $c_i(t_j)$ be the chromatogram of component i monitored at the points in time $j = t_0, \dots, t_{end}$ and preprocessed to a concentration time series based on the partial least squares regression explained in section 2.3.4. Let $\hat{c}_i(t_j; \theta)$ be the solution of a mechanistic model for chromatography for component i at the same points in time, dependent on the parameter setting $\theta = \{\theta_{fix}, \theta_{grad,k}, \theta_{SMA}\}$ with θ_{fix} the previously determined set of model parameters, $\theta_{grad,k}$ the specific gradient volume for the k 'th elution gradient, $1 \leq k \leq N_G$ from a total number of N_G gradients, and θ_{SMA} the set of protein specific SMA parameters. These parameters were to be estimated for all three components, ribonuclease A, cytochrome c and lysozyme at the three different pH conditions, pH 5, pH 6 and pH 7. Aiming at a best fit between model response and chromatographic data, the SMA parameter estimation for three proteins at a specific pH condition based on an inverse method can be stated as an optimization of a sum of least squares:

$$res(\theta_{SMA}) = \sum_{k=1}^{N_G} \sum_{i=1}^3 \sum_{j=0}^{end} (\hat{c}_i(t_j; \theta_{fix}, \theta_{grad,k}, \theta_{SMA}) - c_i(t_j))^2 \quad (7)$$

and the resulting SMA parameter estimation minimizes $res(\theta_{SMA})$. The minimization of Eq. (7) was in all cases performed with the MatlabTM procedure `lsqnonlin`. A local optimizer was sufficient because reasonable initial values for the optimization were available from previous studies [22]. A multi-start strategy supported and assured the optimum, i.e. the estimated parameters. The estimation was employed simultaneously based on four double-determined chromatograms ($N_G = 8$) with optimal elution gradient volumes, that were planned based on a D-optimal experimental design in order to obtain most information with respect to the estimation of SMA parameters.

2.4.3. *D-optimal experimental design.* Let all controllable and uncontrollable factors influencing an experimental result be called *design factors*. For example, the length, the start and end concentration of a salt gradient are design factors in a chromatographic separation. Experimental planning allows for the determination of design factor sets that lead to experiments containing a maximum of information with respect to a specific objective, for example, for the estimation of most accurate SMA-parameters by an inverse method.

Definition 1. Let now $FI(\zeta)$ be the Fisher Information matrix based on the design factors $\zeta_1 \dots \zeta_n$. A defined **experimental design** $\zeta^* = \zeta_1^* \dots \zeta_n^*$ is called *D-optimal* if and only if

$$\det FI(\zeta^*) \in \max_{\zeta \in V^+} \det FI(\zeta) \quad (8)$$

with V^+ being the space of all possible experimental designs [33].

As the inverse of the Fisher Information matrix is equal to the covariance matrix under specific assumptions (see [34] and [35]), experimental designs minimizing the determinant of the covariance matrix are also called D-optimal experimental designs. A popular way to calculate covariance matrices without a demand for linearization assumptions are bootstrap methods based on Monte Carlo simulations. Based on the mechanistic model, a huge number of *in silico*-experiments are performed corresponding to a specific experimental design and afflicted with noise, that is characteristic for noise on chromatograms from miniaturized chromatography. Then the SMA-parameters are estimated based on these *in silico*-chromatograms (see Eq. (7)). The deviations and the covariance

matrix of these estimations can be calculated and the information contents of the specific experimental designs can be compared according to Eq. (8). Further details on this approach are given in [36] and [37].

In this Monte Carlo simulation study, chromatography experiments were simulated, then converted to fractionation data and a specific independent and normally distributed noise was added to the data points by the MatlabTM routine `nrand` to produce typical noise on chromatograms from miniaturized chromatography on robotic platforms. The applied noise had a standard deviation of 0.0005 mM. This deviation was derived beforehand by a characterization of noise in chromatograms that were determined according to the settings in section 2.3.

For the examination of information content dependent on elution gradient volume, at first the SMA parameter estimation based on a single gradient was examined ($N_G = 1$), i.e. the determinant of the covariance matrix was analyzed for all design factors fixed except for the single gradient elution volume, that varied in the simulations between 0 cv and 60 cv. For the examination of the optimal distribution of gradient volumes for an SMA parameter estimation, further simulations of two gradients in the range from 0 cv to 60 cv with different gradient volumes were analyzed with respect to their joint information content ($N_G = 2$), while SMA parameters were estimated simultaneously based on both chromatograms.

2.4.4. Optimization of the gradient elution shape based on mathematical modeling. A calibrated mechanistic model can be employed for optimizations or predictions with respect to specific objectives. To demonstrate the flexibility and predictivity of the applied chromatography model, three optimization objectives for the three component separation were defined beforehand:

- (1) objective **A**: minimal peak overlap (salt wash excluded)
- (2) objective **B**: minimal peak overlap (salt wash included)
- (3) objective **C**: lysozyme elutes 800 seconds after gradient start

The first objective is typical for separation problems, where high resolutions (= small peak overlap) between the peaks are required during elution time. The second objective includes the salt wash step in the optimization assuming that this might lead to better resolution results; the third objective was defined for a verification of the prediction of retention times. The specific optimal gradients with respect to the objectives were calculated numerically based on the MatlabTM routine `fminsearch` and optimization equations analogously to equations shown in [23].

3. RESULTS

3.1. Determination of parameters on column level. In the first step of the concept, (model) parameters on column scale are to be determined based on manufacturer information as well as pulse and frontal experiments performed in parallel and miniaturized on the robotic platform. Figure 3 shows example chromatograms of acetone and dextrane pulse injections derived from the shifting method for higher datapoint densities (see section 2.3.1). The shift in fraction data is discriminated by the color of the dots. Based on these results, the retention time of the single pulse tracer injections is determinable from the data with a precision of $\leq 16 \mu\text{L}$; the blue dextran peak has a retention time of about $67 \pm 16 \mu\text{L}$ and the acetone peak of about $180 \pm 16 \mu\text{L}$. Based on these retention times, the total, interstitial and particle porosities were calculated. A summary of the model parameters that were determined in this first step is given in table 1.

The column length (0.01 m) and the mean particle radius ($0.45 \cdot 10^{-6}$ m) are preset by the manufacturer. The lumped film diffusion coefficient was estimated to be $1.5 \cdot 10^{-6}$ m/s based on

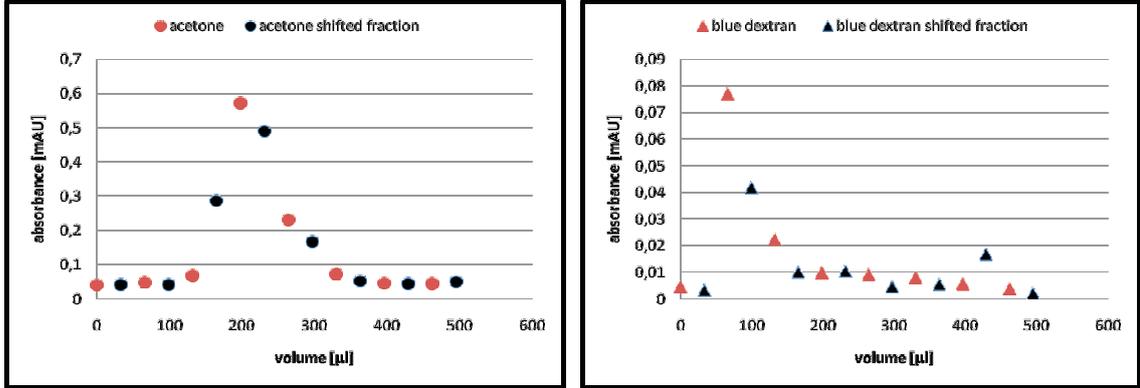


FIGURE 3. Example detection of acetone and blue dextran peaks from pulse injections in miniaturized chromatography (SP Sepharose FFTM, 1% acetone respectively 0.5 mg/mL blue dextran, velocity: 5 μ L/s). The different colors of dots display the shifting method that was applied for higher datapoint density.

parameter		determination
column length	0.01 m	manufacturer information
particle radius	$45 \cdot 10^{-6}$ m	manufacturer information
total porosity	0.85	pulse experiment
interstitial porosity	0.335	pulse experiment
particle porosity	0.9	pulse experiment
ionic capacity of adsorbent	1200 mM	displacement experiment
lumped film diffusion coefficient	$1.5 \cdot 10^{-6}$ m/s	estim. based on Äkta-data
axial dispersion	$1.57 \cdot 10^{-11}$ m ² /s	estim. based on Äkta-data

TABLE 1. Parameter values that were determined in the first step of model-integrated process development

previous experimental work in [22]. The calculated porosities $\varepsilon_t=0.85$, $\varepsilon_p=0.9$ and $\varepsilon_i=0.335$ are reasonable and lie in ranges given in literature, for example in [38]. The ionic capacity of the adsorbent (1200 mM) was measured beforehand with a displacement experiment and is also reasonable (see [39]). The axial dispersion could not be calculated based on the pulse experiments due to the still small number of data points defining the peaks. Thus, the value for axial dispersion ($1.57 \cdot 10^{-11}$ m²/s) was estimated based on previous determinations of this parameter on an Äkta system. In figure 4 example results for miniaturized frontal experiments are shown. Figure 4A shows two superimposed breakthrough curves for ribonuclease A (squares) and lysozyme (crosses) at pH 7. The breakthrough curve for ribonuclease A is clearly steeper and saturation is reached earlier. Figure 4B shows the superimposed breakthrough curves for lysozyme at pH 5 (squares) and pH 7 (crosses). The earlier start and saturation of the breakthrough curve at pH 5 is evident from the data. Repetitions of these experiments showed similar variances in the breakthrough curves as were observed in [3] (data not shown). Based on breakthrough data, dynamic capacities could be determined at 10 % (Q10) and 50 % (Q50) of the breakthrough for ribonuclease A, cytochrome *c*

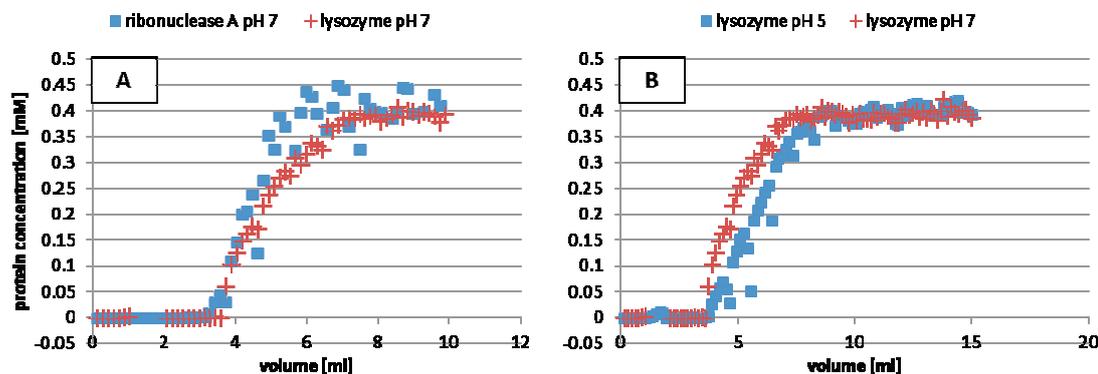


FIGURE 4. Figure 4A shows Breakthrough curves for ribonuclease A (squares) and lysozyme (crosses) at pH 7. Figure 4B shows breakthrough curves for lysozyme at pH 5 (squares) and pH 7 (crosses). The frontal experiments were performed parallelized and miniaturized on a robotic platform (init. conc.: $0.4 \cdot 10^{-3}$ M, pipetting velocity: $2.5 \mu\text{L/s}$).

	capacity [mg/ml _{res}]	ribonuclease A	cytochrome <i>c</i>	lysozyme
pH 5	Q10	116.45	100.8	114.46
	Q50	152.07	136.22	165.96
pH 6	Q10	133.58	130.03	108.73
	Q50	158.92	147.37	168.82
pH 7	Q10	104.12	128.8	110.15
	Q50	113.71	143.65	135.9

TABLE 2. Adsorbent capacities of SP Sepharose FF for ribonuclease A, cytochrome *c* and lysozyme at pH 5, 6 and 7 determined by breakthrough experiments on the robotic platform

and lysozyme at pH 5, 6 and 7. The capacities for all three proteins on SP Sepharose FFTM are given in table 2. A comparison between the capacities for ribonuclease A and lysozyme at pH 7 and for lysosyne at pH 5 and pH 7 respectively, determines quantitatively the capacity increase that was qualitatively noticed before in figure 4. Furthermore, a significant decrease of Q50-capacities from pH 6 to pH 7 could be determined based on the miniaturized data as well as the fact that the capacity for ribonuclease A on SP Sepharose FFTM is always smaller than the capacity for cytochrome *c* and lysozyme.

3.2. Estimation of parameters on particle level. In the second step of model-integrated process development, gradient elution experiments were performed. Based on this data and the frontal data from the first step, the aim was to achieve the decision for an optimal pH-range for the separation. In addition, the data was to be used for model calibration with respect to sorption-parameters (particle level). The Monte Carlo based approach described in section 2.4.3 was employed for the previous determination of an optimal experimental design. At first, the information content in

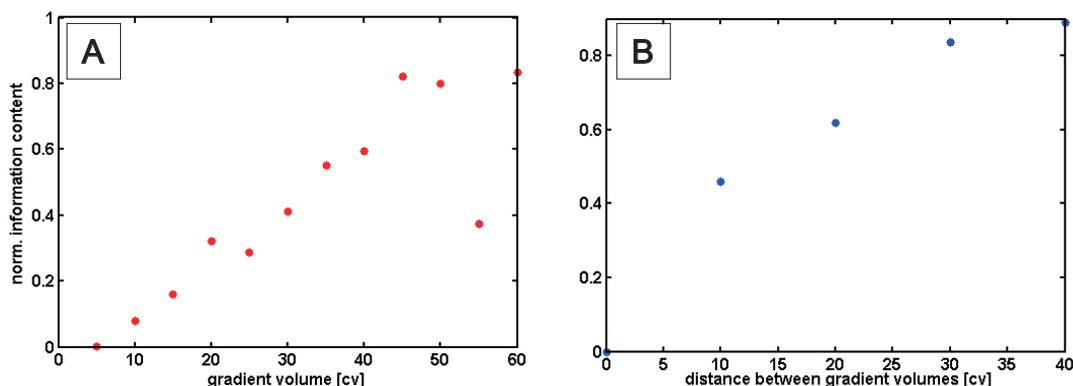


FIGURE 5. Figure 5A shows the effect of elution gradient volume on the normalized information content in fractionation data. Figure 5B shows a positive correlation between the normalized information content in two gradient elution experiments and the distance of elution volumes. In both figures, information content was calculated based on Monte Carlo simulations under the objective of precise SMA parameter estimation.

single gradient elution experiments of different gradient volume with respect to the estimation of parameters for the steric mass action (SMA) sorption model was calculated for elution volumes between 5 cv and 60 cv. Based on 1000 simulations in each case, the determinant of the inverse covariance matrix, i.e. the information content with respect to SMA parameter estimation, was derived. The information content was calculated according to the theory in section 2.4.3. However, the calculated information value is most meaningful in comparison to the information from other experimental designs. Thus, as the lowest possible information content of an experimental design can be assumed to be 0, the highest calculated value from the simulations was set to be 1 to simplify the comparison of designs. Figure 5A illustrates the effect of elution gradient volume on the normalized information content in fractionation data. The results show that the information content correlates linearly with increasing gradient volume, up to a gradient volume of 45 cv. For gradient volumes longer than 45 cv, the information content appears to stay at a constant level.

Secondly, as parameters for the SMA model were to be calculated based on at least two data sets, the influence of elution volume distribution had to be examined. Hence, one experiment with an elution volume of 10 cv and a second experiment in the range from 10 cv to 50 cv elution volume was simulated. Analogously to the previous Monte Carlo approach, the results were analyzed with respect to their joint information content ($N_G = 2$ in Eq. (7)).

Figure 5B shows a positive correlation between joint information content in two gradient elution experiments and the distance between the elution volumes in the simulations (from 0 to 40 cv). Similar results were derived when fixing the first gradient at 20 or 30 cv. This result implies that joint information content in gradient elution experiments increases with increased distance between the analysed elution gradient volumes. This increase appears to reach a saturation at a distance larger than 40 cv.

As the longest gradient that could be performed on the robotic platform had a volume of 35 cv and the overall number of gradients for a single pH-screening was set to four (for practical performance of double determinations with parallel pipetting of eight columns), gradient elution volumes of 5 cv,

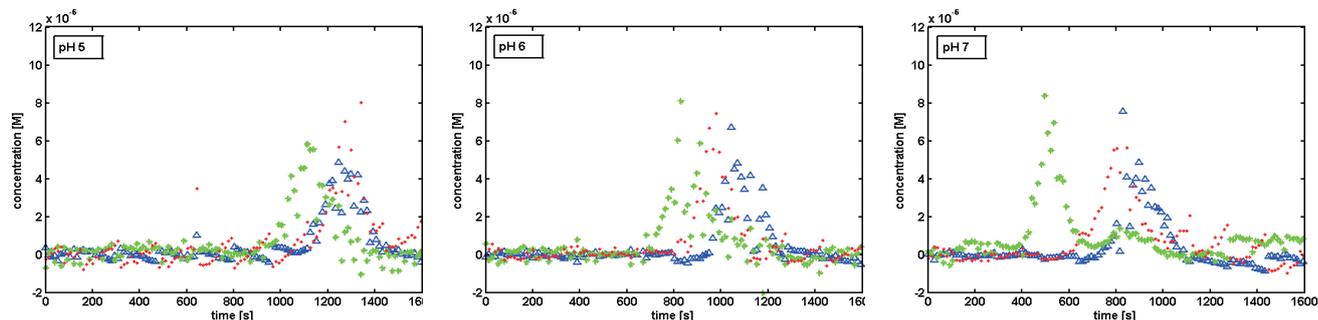


FIGURE 6. The results for a linear gradient elution of a three-component mixture with 35 cv elution gradient volume are shown for three pH conditions (SP Sepharose FFTM, pH-values of 5, 6 and 7 in ascending order from left-hand side to right-hand side). Ribonuclease A elutes firstly (stars), then cytochrome *c* (dots) and lysozyme (triangles) elute [init. conc.: $0.2 \cdot 10^{-2}$ M, volumetric flow rate: $5 \mu\text{L/s}$].

		ν	k_{eq}
pH 5	rib	5.35	0.08
	cyt	5.4	0.24
	lys	5.54	0.23
pH 6	rib	3.47	0.27
	cyt	3.9	0.25
	lys	3.82	0.53
pH 7	rib	2.51	0.17
	cyt	3.1	0.49
	lys	3.73	0.27

TABLE 3. Estimated values of the SMA sorption parameters based on an inverse method with data from optimally designed, miniaturized gradient elutions.

15 cv, 25 cv and 35 cv were chosen for the screening experiments, taking into account both effects derived from the optimal experimental design by choosing the longest possible elution gradient and the largest possible distances between elution volumes. In figure 6 three example chromatograms for the optimally designed gradient elution experiments with an elution gradient volume of 35 cv at pH 5, 6 and 7 are shown. The elution gradient started after 200 seconds and lasted for 1400 seconds, at which point 100% high salt buffer with a salt concentration of 0.5 M NaCl was reached.

At all examined pH conditions ribonuclease A elutes first (stars) followed by cytochrome *c* (dots) and lysozyme (triangles). At pH 5 cytochrome *c* and lysozyme elute almost simultaneously, at higher pH-values cytochrome *c* elutes slightly earlier than lysozyme. The overall resolution between the peaks at pH 5 is very low, but it increases with ascending pH-value while retention volumes decrease. These screening results together with the capacity measurements (see table 2) bring about the decision to tackle this separation problem at pH 7. Obviously, the complete separation of lysozyme and cytochrome *c* poses the major challenge in this separation problem and demands an optimization of the elution gradient shape. According to the concept of model-integrated process

development, SMA parameters were estimated using the inverse method based on the gradient elution experiments (see Eq. (7)). The results for this estimation are given in table 3. The parameter of steric shielding (σ) is not enlisted, as the estimation showed a broad statistical spread between 20 and 50. Numerical difficulties and broad statistical spread in the determination of σ have been reported before in [40] and [31]. In subsequent simulations the parameter of steric shielding was set to a value of 30, as obviously this parameter has a rather small influence on the specific separation problem in the examined design space. The estimated values for characteristic charge (ν) and equilibrium coefficient (k_{eq}) lie within reasonable limits compared with literature (e.g. [41], [42]). The characteristic charges decrease significantly with increasing pH-conditions and the characteristic charge from ribonuclease A is generally smaller than the characteristic charge of the two other proteins.

3.3. Predictions based on the mechanistic model. The third step of model-integrated process development tackles the model-based optimization of the separation gradient and the issue of model verification. The calibrated mechanistic model was employed for predictions of optimal gradient elution experiments with respect to the three objectives A, B and C of resolution maximization and predictivity control: minimal peak overlap (salt wash excluded/included) and the constraint on lysozyme to elute 800 seconds after gradient begin respectively. The results of the model-based optimization of the three-component separation on SP Sepharose FFTM with respect to the three objectives A, B and C are given in figure 7 on the left-hand side. The determined optimal gradients are given as a dark green line for each objective, the simulated corresponding chromatograms are depicted as continuous lines (1st peak: ribonuclease A, 2nd peak: cytochrome *c* and 3rd peak: lysozyme) and the calculated overlap is marked by the darker areas between the peaks. A comparison of the results for objective A and objective B (subfigures 7A and 7B on the left-hand side of figure 7) reveals that the resolution between the three peaks can be significantly improved by including the high salt wash step to the elution time. Then, lysozyme is eluting only in the high salt wash step. In addition, subfigure 7B shows that there exists no gradient in the design space for complete separation of cytochrome *c* and lysozyme. The numerically optimized gradient with respect to objective C is given in subfigure 7C on the left-hand side of figure 7. A gradient with the initial salt concentration of 0.1 M NaCl reaching the terminal salt concentration of 0.21 M NaCl in 740 seconds (ca. 18.5 cv) is predicted to set average retention time of lysozyme on 800 seconds after gradient initiation. Corresponding experimental results are shown on the right-hand side of figure 7 (ribonuclease A: stars, cytochrome *c*: dots and lysozyme: triangles). The experimental data matches the predictions with a high precision. Even details, for example the small peak from cytochrome *c* in the salt wash step (subfigure 7B), were predicted correctly. The prediction of retention time with respect to objective C had a precision of up to 2.2 seconds. The derivation of upscale predictions based on the calibrated and verified mechanistic model is a consequent and important step in the concept of model-integrated process development. For upscale predictions, the bed characterizing parameters of a 1 mLprepacked HiTrapTM column with SP Sepharose FFTM [0.025 m bed height, 0.007 m I.D.](GE Healthcare, Little Chalfont, Buckinghamshire, UK) were inserted into the mechanistic model (see [22]). In figure 8A the model-based upscale prediction of an optimal gradient considering objective B is shown. The prediction is already transformed to the usual measurement conditions on an Äkta system: continuous absorbance measurements at 280 and 528 nm. In figure 8B experimentally derived chromatograms for model verification corresponding to the prediction are given. Again, lysozyme and ribonuclease A are predicted very precisely; the

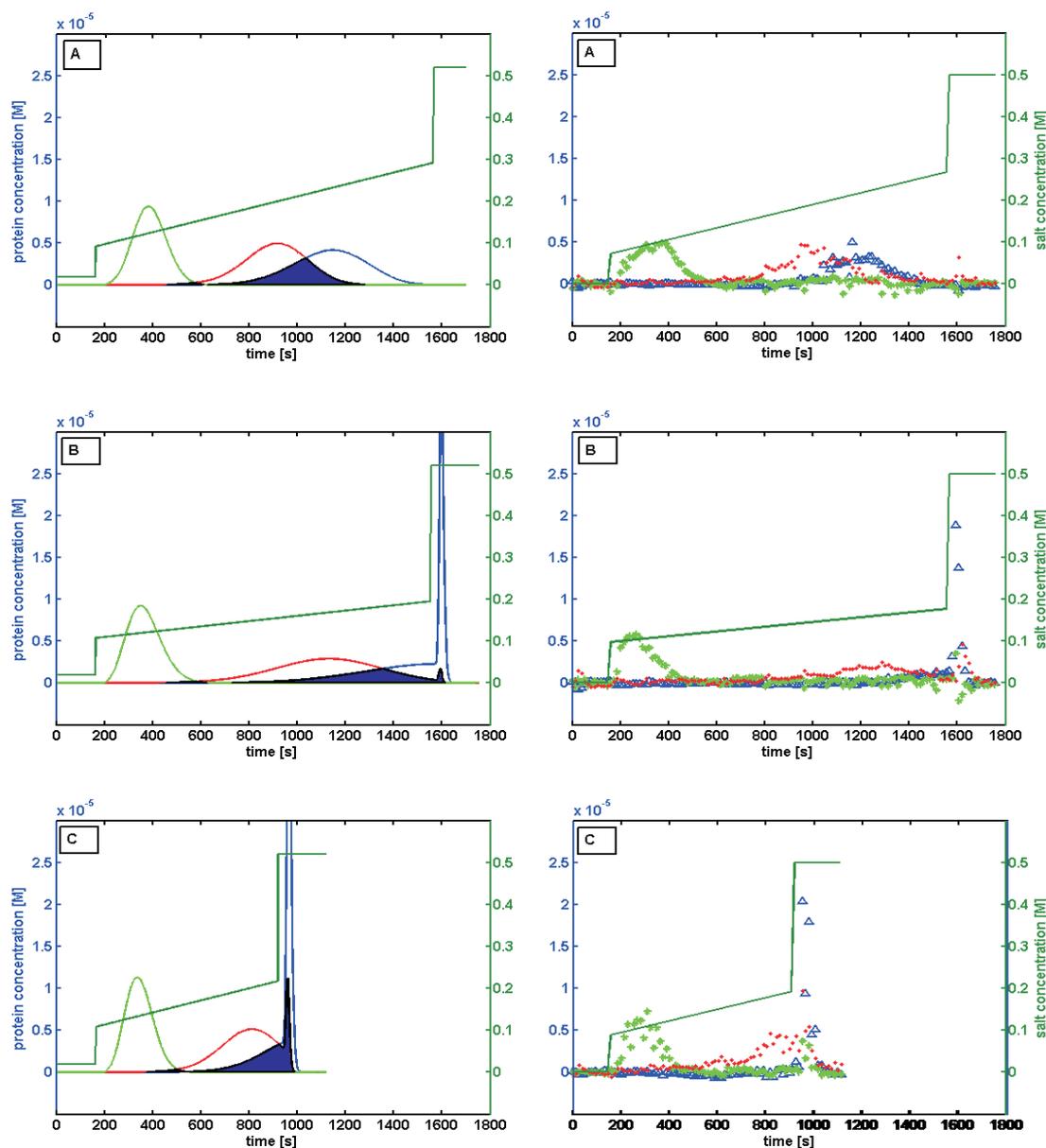


FIGURE 7. Model-based optimized gradients with respect to the objectives A, B and C (left-hand side) are compared to experimental data for model verification (right-hand side, 1st peak: ribonuclease A, 2nd peak: cytochrome *c* and 3rd peak: lysozyme). The predicted overlap is marked by the darker areas [init. conc. of proteins: $0.2 \cdot 10^{-2}$ M, volumetric flow rate: $5 \mu\text{L/s}$].

prediction for cytochrome *c* deviates slightly in retention time (3 min deviation in 60 min overall elution time). To quantitatively verify the performance of upscale predictions based on the model calibrated with miniaturized data, the SMA parameters for the model were estimated once again based directly on chromatograms from the 1 mL column on an Äkta system (see [23], ribonuclease A:

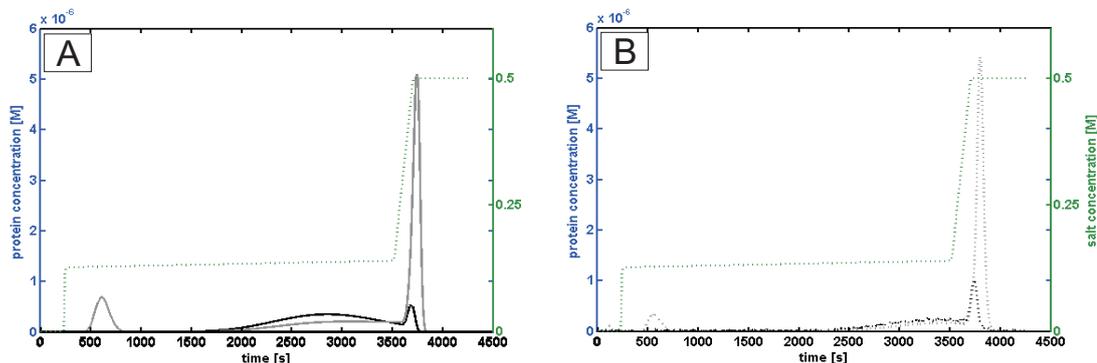


FIGURE 8. Model-based prediction for an optimal gradient (objective B, dotted line) in 1 mL scale based on model calibration with chromatograms from miniaturized chromatography (figure 8A, dark continuous peak: overall protein concentration, light continuous peak: cytochrome *c* concentration) is compared to experimental data (figure 8B, 1 mL column with SP sepharose FFTM, three-component separation performed on an Äkta system).

	optimal value RoboColumn	optimal value 1 mL column
gradient volume	18 cv	25 cv
start conc. NaCl	0.14 M	0.15 M
end conc. NaCl	0.16 M	0.16 M

TABLE 4. Comparison between the prediction for optimal separation gradients for the three-component separation (objective B) on 1 mL scale. The prediction on the left-hand side is based on a SMA parameter estimation with chromatograms from miniaturized chromatography; the prediction on the right-hand side on an SMA parameter estimation based on chromatograms from 1 mL scale on an Äkta system.

$\nu = 1.6$, $k_{eq} = 0.28$, cytochrome *c*: $\nu = 2.8$, $k_{eq} = 0.27$, lysozyme: $\nu = 3.4$, $k_{eq} = 0.14$). Consequently, an optimal gradient with respect to objective B could now be predicted with the newly calibrated model. On the left-hand side of table 4 the prediction for the optimal gradient based on SMA parameter estimations from miniaturized experiments is given in numbers (see figure 8). On the right-hand side the prediction based on SMA parameters derived from chromatograms from the 1 mL column on the Äkta system is shown. In both cases, the optimal gradient volumes as well as optimal initial and terminal salt concentrations were calculated. The comparison shows that the determined optimal initial and terminal salt concentrations are very similar. The predictions for the optimal elution gradient volume deviate slightly, but as the gradient is very flat, this deviation in gradient volume has a rather small influence on the peak resolutions.

4. DISCUSSION

The approach of model-integrated process development in three steps proved to be successful and flexible, leading to optimized separations and precise upscale predictions. In the first step of model-integrated process development reasonable porosities (see table 1) and reliable dynamic capacities

(see table 2) were calculated based on pulse and breakthrough experiments. A new method for higher data density in miniaturized column experiments was successfully introduced (see figure 3) and porosity determinations were comparable to the values derived by Susanto et al. [20], whose results are based on pulse experiments with a selfmade adapter on an Äkta system. Small differences in the values can be traced back to the use of different measurement approaches and discrepancies between prepacked column lots. Three column parameters (axial dispersion, lumped film diffusion coefficient and ionic capacity of the column) had to be determined by additional experiments, because the density of data points that could be achieved in miniaturized processes was still too low for reasonable calculations. The data quality of the breakthrough curves is slightly better compared to the curves determined in earlier work by [3]. Still, similar robotic-induced effects in the data, like concentration decreases due to fraction collection plate changes and significant data scattering for high protein concentrations were observed (figure 4).

In the second step of model-integrated process development, gradient elution experiments were performed. These experiments were planned by model-based optimal experimental design with respect to the estimation of sorption parameters. The experimental plan revealed an linear information increase for higher gradient volumes up to a elution gradient volume of 45 cv (figure 5A) and an increase in information content of two gradient volumes with increasing distance between their elution volumes. The linear correlation can be explained by a higher number of non-zero data-points in fractionation data, due to peak broadening for large elution volumes. However, for elution volumes larger than 45 cv, this broadening causes very shallow and noisy peaks; this might be the explanation for the stagnating information content beyond this limit. This fact probably also explains the observed saturation for elution volumes, with a larger distance than 40 cv (figure 5B). Still, it could be observed that larger distances between elution gradient volumes contain more information. This is probably due to the increase in variation in chromatographic data with increasing variation of the elution gradient.

The results from the gradient elution experiments met both targets of the second step of model-integrated process development: they allowed for a reliable decision on the optimal pH-range for the separation problem and for the estimation of SMA parameters by an inverse method (table 3), i.e. the calibration of the mechanistic model on particle level. The elution behavior of the three proteins was significantly dependent on the pH condition (figure 6). The same behaviour of ribonuclease A, cytochrome *c* and lysozyme was shown before in miniaturized and lab scale experiments, for example in [4] and [23].

As SMA parameters are highly system-dependent (adsorbent, porosities, pH, etc.), they cannot be compared directly with values from literature. Nevertheless, the parameter estimations in table 3 are reasonable estimations compared to publications from [42] and [41]. According to the theory of pH-induced binding site variations it is very likely that characteristic charges ν change with different pH-values (see [43] and [44]). Moreover, it was observed in [22] that for the three model proteins on SP Sepharose FFTM the order relation of characteristic charges correlates with elution order. Therefore, the closeness of agreement in the characteristic charges for lysozyme and cytochrome *c* already indicates the challenge of this separation problem, qualitatively observed in figure 6.

The third step of model-integrated process development tackles the optimization of the separation gradient based on the completely calibrated model and model verification. The mechanistic model was employed for predictions of optimal gradient elution experiments concerning three objectives of resolution maximization and retention time control. The flexibility of mechanistic modeling allowed for predictions for various objectives without the need of model recalibration, a high gain in comfort

compared to approaches based on empiric response-surface modeling [23]. Figure 7 revealed the high predictivity of mechanistic modeling even for small details and the important information that no gradient exists in this design space for complete separation of cytochrome *c* and lysozyme. Only by superimposing predictions and lab results, slight discrepancies between the predicted and actual cytochrome *c* peaks were detectable, which can probably be explained by an interim change of the cytochrome *c* lot - perhaps a recalibration would have been necessary, if higher precision was desired. The model-based upscale on a 1 mL column was successful (figure 8). The comparison of optimal gradient predictions for the 1 mL scale, on the one hand performed based on miniaturized data and on the other hand performed based on 1 mL data (see table 4) showed very close results and a high sensitivity of the separation problem with respect to initial and terminal salt concentration of the gradient.

Although this manuscript deals with a model system of three well known proteins, the establishment of the model-integrated process development for more complex industrial systems is possible with little additional effort. The presented method of high-throughput protein quantification is generally applicable to protein mixtures. Furthermore, high-throughput analytics of the fractions like it is described in [5] and [45] allow for the application of the inverse method for the determination of SMA parameters.

5. CONCLUSION

In this paper a new concept for chromatography process development based on high-throughput data and mechanistic modeling was presented. With the help of a case study, the three steps of this concept (see figure 1) were demonstrated and discussed. Data from high-throughput screening on a robotic platform could successfully be employed for model calibration and verification. All but three parameters (axial dispersion, lumped film diffusion coefficient and ionic capacity) have been determined or estimated based on minaturized and parallelized pulse and gradient elution experiments. Model integrated process development proved to be flexible with regard to various objectives for the optimization process and led to optimal settings with regard to adsorbent choice, pH condition and elution gradient shape in every case. The efficient and seamless cooperation of high-throughput screening, modeling and model-based experimental design proved to be successful leading to optimized separations and resonable upscale predictions. With regard to previously mentioned model-based robustness analyses and sensitivity studies, this concept opens a promising way to tackle these issues from the first stages of process development on. The concept of integrated modeling in combination with the introduced method of high-throughput protein quantification should now be challenged by industrial separation problems for improvement and validation.

REFERENCES

- [1] P. Knight, Chromatography: 1989 report, *Bio/technology* 7 (1989) 243–249.
- [2] F.D.A.U.S. Department of Health and Human Services (Ed.), *Guidance for Industry PAT - A Framework for Innovative Pharmaceutical Manufacturing and Quality Assurance*, 2004.
- [3] M. Wiendahl, P. S. Wierling, J. Nielsen, D. F. 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.
- [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 (2009) 140–154.

- [5] K. Beckley, P. Verhaert, L. van der Wielen, J. Hubbuch, M. Ottens, Rational and systematic protein purification process development: The next generation, *Trends in Biotechnology* 27 (12) (2009) 673 – 679.
- [6] G. Guiochon, Preparative liquid chromatography, *Journal of Chromatography A* 965 (1-2) (2002) 129–61.
- [7] S. Chan, N. Titchener-Hooker, D. Bracewell, E. Sorensen, A systematic approach for modeling chromatographic processes - application to protein purification, *AIChE Journal* 54 (4) (2008) 965 – 977.
- [8] N. Jakobsson, D. Karlsson, J. Axelsson, G. Zacchi, B. Nilsson, Using computer simulation to assist in the robustness analysis of an ion-exchange chromatography step, *Journal of Chromatography A* 1063 (1-21-2) (2005) 99–109.
- [9] M. Teeters, T. Benner, D. Bezila, H. Shen, A. Velayudhan, P. Alred, Predictive chromatographic simulations for the optimization of recovery and aggregate clearance during the capture of monoclonal antibodies, *Journal of Chromatography A* 1216 (2009) 6134–6140.
- [10] U. Altenhoener, M. Meurer, J. Strube, H. Schmidt-Traub, Parameter estimation for the simulation of liquid chromatography, *Journal of Chromatography A* 769 (1) (1997) 59 – 69.
- [11] H. Teoh, M. Turner, N. Titchener-Hooker, E. Sorensen, Experimental verification and optimisation of a detailed dynamic high performance liquid chromatography column model, *Computers & Chemical Engineering* 25 (4-6) (2001) 893–903.
- [12] P. Persson, P. Gustavsson, G. Zacchi, B. Nilsson, Aspects of estimating parameter dependencies in a detailed chromatography model based on frontal experiments, *Process Biochemistry* 41 (8) (2006) 1812 – 1821.
- [13] N. Tugcu, S. Cramer, K. Rege, Predicting column performance in displacement chromatography from high throughput screening batch experiments, *Separation Science and Technology* 38 (7) (2003) 1499–1517.
- [14] T. Herrmann, M. Schroeder, J. Hubbuch, Generation of equally sized particle plaques using solid-liquid suspensions, *Biotechnology Progress* 22 (3) (2006) 914–8.
- [15] M. Bensch, P. Schulze Wierling, E. von Lieres, J. Hubbuch, High throughput screening of chromatographic phases for rapid process development, *Chemical Engineering & Technology* 28 (11) (2005) 1274–1284.
- [16] J. Mollerup, T. Hansen, S. Kidal, L. Sejergaard, A. Staby, Development, modelling, optimisation and scale-up of chromatographic purification of a therapeutic protein, *Fluid Phase Equilibria* 261 (1-2) (2007) 133–139.
- [17] J. Coffman, J. Kramarczyk, B. Kelley, High-throughput screening of chromatographic separations: I. Method development and column modeling, *Biotechnology and Bioengineering* 100 (4) (2008) 605–18.
- [18] T. Bergander, K. Nilsson-Välímää, K. Öberg, K. Lacki, High-throughput process development: Determination of dynamic binding capacity using microtiter filter plates filled with chromatography resin, *Biotechnology Progress* 24 (2008) 632–9.
- [19] S. Gerontas, M. Asplund, R. Hjorth, D. Bracewell, Integration of scale-down experimentation and general rate modelling to predict manufacturing scale chromatographic separations, *Journal of Chromatography A* 1217 (44) (2010) 6917–26.
- [20] 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.
- [21] 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.
- [22] A. Osberghaus, S. Hepbildikler, S. Nath, M. Haindl, E. von Lieres, J. Hubbuch, Determination of parameters for the steric mass action model - a comparison between two approaches, *Journal of Chromatography A* 1233 (2012) 54–65.
- [23] A. Osberghaus, S. Hepbildikler, S. Nath, M. Haindl, E. von Lieres, J. Hubbuch, Optimizing a chromatographic three component separation: A comparison of mechanistic and empiric modeling approaches, *Journal of Chromatography A*, doi: 10.1016/j.chroma.2012.03.029.
- [24] S. Gallant, A. Kundu, S. Cramer, Modeling non-linear elution of proteins in ion-exchange chromatography, *Journal of Chromatography A* 702 (1-2) (1995) 125 – 142.
- [25] I. Olsson, J. Gottfries, S. Wold, D-optimal onion designs in statistical molecular design, *Chemometrics and Intelligent Laboratory Systems* 73 (1) (2004) 37 – 46.
- [26] C. Brooks, S. Cramer, Steric mass-action ion exchange: Displacement profiles and induced salt gradients, *AIChE Journal* 38 (1992) 1969 – 1978.

-
- [27] N. Titchener-Hooker, S. Chan, D. Bracewell, A systematic approach for modeling chromatographic processes: Application to protein purification, *AIChE Journal* 54 (4) (2008) 965–977.
- [28] L. Pedersen, J. Mollerup, E. Hansen, A. Jungbauer, Whey proteins as a model system for chromatographic separation of proteins, *Journal of Chromatography B* 790 (1-2) (2003) 161 – 173.
- [29] S. Gadam, G. Jayaraman, S. Cramer, Characterization of non-linear adsorption properties of dextran-based polyelectrolyte displacers in ion-exchange systems, *Journal of Chromatography A* 630 (1-2) (1993) 37 – 52.
- [30] P. Danckwerts, Continuous flow systems: Distribution of residence times, *Chemical Engineering Science* 2 (1) (1953) 1–13.
- [31] A. Shukla, S. Bae, J. Moore, K. Barnhouse, S. Cramer, Synthesis and characterization of high-affinity, low molecular weight displacers for cation-exchange chromatography, *Industrial & Engineering Chemistry Research* 37 (1010) (1998) 4090–4098.
- [32] E. von Lieres, J. Andersson, A fast and accurate solver for the general rate model of column liquid chromatography, *Computers & Chemical Engineering* 34 (8) (2010) 1180–1191.
- [33] H. Bandemer, A. Bellmann, *Statistische Versuchsplanung*, B.G. Teubner Verlagsgesellschaft, Leipzig, 1994.
- [34] K. Mardia, J. Kent, J. Bibby, *Multivariate Analysis*, Academic Press Inc. Ltd., London, 1979.
- [35] L. Fahrmeir, A. Hamerle, *Multivariate statistische Verfahren*, Walter de Gruyter & Co., Berlin, 1995.
- [36] B. Efron, R. Tibshirani, *An Introduction to the Bootstrap*, Chapman and Hall, CRC Press, London, Boca Raton, 1993.
- [37] M. Joshi, A. Seidel-Morgenstern, A. Kremling, Exploiting bootstrap method for quantifying parameter confidence intervals in dynamical systems, *Metabolic Engineering* 8 (2006) 447–455.
- [38] H. Schmidt-Traub (Ed.), *Preparative Chromatography*, Wiley-VCH, Weinheim, Germany, 2006.
- [39] C. A. Teske, E. von Lieres, M. Schröder, A. Ladiwala, S. M. Cramer, J. Hubbuch, Competitive adsorption of labeled and native protein in confocal laser scanning microscopy, *Biotechnology and Bioengineering* 95 (2006) 58–66.
- [40] H. Iyer, S. Tapper, P. Lester, B. Wolk, R. van Reis, Use of the steric mass action model in ion-exchange chromatographic process development, *Journal of Chromatography A* 832 (1999) 1–9.
- [41] S. Gallant, S. Cramer, Productivity and operating regimes in protein chromatography using low-molecular-mass displacers, *Journal of Chromatography A* 771 (1997) 9–22.
- [42] A. Ladiwala, K. Rege, C. Breneman, S. Cramer, J. Prausnitz, A priori prediction of adsorption isotherm parameters and chromatographic behavior in ion-exchange systems, *Proceedings of the National Academy of Sciences of the United States of America* 102 (33) (2005) 11710–11715.
- [43] T. Yang, M. Sundling, A. Freed, C. Breneman, S. Cramer, Prediction of pH-dependent chromatographic behavior in ion-exchange systems, *Analytical Chemistry* 79 (23) (2007) 8927–39.
- [44] F. Dismer, J. Hubbuch, 3D structure-based protein retention prediction for ion-exchange chromatography, *Journal of Chromatography A* 1217 (8) (2010) 1343–53.
- [45] T. Ahamed, B. Nfor, L. van der Wielen, P. Verhaert, G. van Dedem, M. Eppink, E. van de Sandt, M. Ottens, Omics tools in accelerated purification process development: Multidimensional fractionation and characterization of crude protein mixtures, *Journal of Bioscience and Bioengineering* 108 (1) (2009) S58–S58.

4 Conclusion & Outlook

In this thesis it was shown that mechanistic modeling in chromatography is a powerful tool for the optimization and characterization of separation processes. The superiority of this approach compared to screening approaches, search algorithms and empiric modeling, especially with regard to complex separations of multicomponent mixtures, could be demonstrated. The application of model-based Design of Experiments resulted into a more efficient concept of model calibration by inverse methods; the results revealed, that for the calibration of the model-embedded sorption equations (SMA model), the results from a few designed gradiental experiments are equally predictive than the established approach based on laborious experimental methods. All designed experiments were easily included into process development routines and only a few additional experiments had to be undertaken for the determination of model parameters like ionic capacity or axial dispersion. It could be demonstrated that the model can be calibrated based on miniaturized and parallelized high-throughput screening experiments from a robotic platform. Thus, despite of the lower data quality, even experiments from this very early step of process development are fit for model-integrated process optimization.

Already calibrated chromatography models were successfully employed for the optimization of multicomponent separation with various shapes of elution gradients. Furthermore it was shown that the model can be employed for robustness analyses and owns a high and precise predictivity not only in the primarily examined design space but also for upscale predictions.

During the work on this thesis, related issues arose, that could not be considered due to a restrictive time frame and the intention to focus on the main topic; the three most interesting of these issues will be mentioned in this outlook. Although there is a lot of research on sorption processes in chromatography, the mechanistic modeling of these processes should still be looked on in more detail. It would be of interest to know, how explicitly the SMA parameters really explain ad- and desorption in a mechanistical way, like it is assumed in the community, and how the effects of the experimental pH conditions could be integrated into this sorption model. Perhaps even new approaches have to be developed in order to do this, because molecular dynamic simulations definitely show changes in protein confirmation and binding positions for different pH conditions. The connection between mechanistic modeling and molecular dynamic simulations could therefore lead to an even more predictive and mechanistic model for sorption processes.

With regard to the findings in this thesis, experimental design should more intensely be applied in the field of chromatography. The issue of designing experiments, that contain maximal information with respect to specific objectives, let it be model calibration or the optimization of a separation step, should be adressed thoroughly in various relevant applications to reveal the convincing advantages. Furthermore, a consequent use of Experimental design would lead to interesting strategies for process control, for example for the optimal choice of fractionation time steps.

A third issue of interest poses the expansion of mechanistic models for chromatography to other chromatography modi. As explained in the introduction, in pharmaceutical industry mostly two or three orthogonal chromatography steps are used for purification. A model allowing for the simulation and predictions for a complete series of chromatography

steps of different modi would be a valuable tool in designing industrial multistep processes. This model should also allow for uncertainty propagation analyses. Then, based on inverse methods, failed experiments could be examined concerning the reason of the failure, what would lead to a real improvement in (biopharmaceutical) process development and control.

5 Abbreviations and Symbols

abbreviation	unit	definition
cv		column volume
DoE		design of experiments
DWP		deep well plate
HTS		high-throughput screening
IEC		ion exchange chromatography
LiHA		liquid handling of the robotic station
PDAE		partial differential and algebraic equations
RSM		response surface modeling
SMA		steric mass action
$c_{a,s}$	M	salt concentration at gradient begin
$c_{e,s}$	M	salt concentration at gradient end
c_i	mg ml ⁻¹	concentration of component i in the mobile phase
$c_{i,0}$	mg ml ⁻¹	concentration of component i at column inlet
$c_{i,p}$	M	concentration of component i in the pores (particle level)
$c_{i,sim}$	mg ml ⁻¹	simulated concentration of component i
c_{prot}	mg ml ⁻¹	concentration of protein stock solution
c_{salt}	M	salt concentration
cp		objective value/response at a DoE-center point
D_{ax}	m ² s ⁻¹	axial dispersion
f_{RSM}		response surface model function
K_{eq}	ml mg ⁻¹	equilibrium coefficient (Langmuir model)
$k_{i,eff}$	ms ⁻¹	effective film transfer coefficient for component i
$k_{i,ads}$	s mM ^{-ν}	adsorption coefficient of component i (SMA model)
$k_{i,des}$	s mM ^{-ν}	desorption coefficient of component i (SMA model)
$k_{i,eq}$		equilibrium coefficient of component i (SMA model)
L_C	m	column length
Q		quotient for the comparison of relative standard deviations from simulated and experimental results
q_i	M	concentration of component i on adsorbent particle surface
$q_{i,sim}$	mg ml ⁻¹	simulated concentration of component i on adsorbent particle surface
q_{max}	mg ml ⁻¹	saturation coefficient (Langmuir model)
R^2		coefficient of determination
r_p	m	particle radius
rep		reproducibility
$res_{P1,P2}$		resolution between peaks belonging to the proteins P1 and P2
u_{int}	ms ⁻¹	interstitial flow velocity
V_0	ml	breakthrough volume at 10% of a nonretarded tracer
V_{ads}	ml	adsorbent volume
V_B	ml	breakthrough volume at 10% of a protein component

V_{buff}	ml	volume of buffer solution
V_C	m^3s^{-1}	column volume
V_d	ml	dead volume
V_G	ml	gradient volume
V_{prot}	ml	protein solution volume
V_R	ml	retention volume
V_{stor}	ml	storage buffer volume
V_{sup}	ml	supernatant volume
β		phase ratio
ε_{ads}	ml	independent and normally distributed error on the adsorbent plaque volumes
ε_c		column porosity
ε_{cprot}	ml	independent and normally distributed error on the concentration of the protein stock solution
ε_p		particle porosity
ε_{pipp}	ml	independent and normally distributed error on the LiHA pipetting
ε_{stor}	ml	independent and normally distributed error on the pipetting of storage buffer
ε_t		total column porosity
Λ	M	ionic capacity
μ_P	ml	first central moment of the peak belonging to protein P
ν_i		characteristic charge of component i (SMA model)
σ_i		steric factor of component i (SMA model)
σ_P		squareroot of the second central moment of the peak belonging to protein P
$\sigma_{rel,ads}$		relative standard deviation of the adsorbent plaque volume
$\sigma_{rel,cprot}$		relative standard deviation of the protein stock solution concentration
$\sigma_{rel,meas}$		relative standard deviations from the results of the experimental binding studies
$\sigma_{rel,pipp}$		relative standard deviation of the LiHA pipetting
$\sigma_{rel,sim}$		relative standard deviations of the results of the simulated binding studies
$\sigma_{rel,stor}$		relative standard deviation of the storage buffer pipetting
θ_{est}		parameters that will be estimated in the least squares optimization solving the inverse problem
θ_{fix}		parameters that are fixed during the least squares optimization solving the inverse problem
θ_{grad}		parameters/factors that describe the unique shape of a elution gradient

References

- ALTENHOENER, U., MEURER, M., STRUBE, J. and SCHMIDT-TRAUB, H. (1997). *Parameter estimation for the simulation of liquid chromatography*. Journal of Chromatography A, 769(1):59 .
- ANDRZEJEWSKA, A., KACZMARSKI, K. and GUIOCHON, G. (2009). *Theoretical study of the accuracy of the pulse method, frontal analysis, and frontal analysis by characteristic points for the determination of single component adsorption isotherms*. Journal of Chromatography A, 1216(7):1067.
- BACHMAN, W. and STEWART, J. (1989). *Optimization in photochemical reaction detection - application to High-Performance Liquid Chromatography-Photolysis-Electrochemical detection*. Journal of Chromatography, 481(21):121.
- BAK, H., THOMAS, O. and ABILDSKOV, J. (2007). *Lumped parameter model for prediction of initial breakthrough profiles for the chromatographic capture of antibodies from a complex feedstock*. Journal of Chromatography B, 848(1):131.
- BANDEMER, H. and BELLMANN, A. (1994). *Statistische Versuchsplanung*. B.G. Teubner Verlagsgesellschaft, Leipzig.
- BARZ, T., LOEFFLER, V., ARELLANO-GARCIA, H. and WOZNY, G. (2009). *Optimal determination of SMA-parameters for the adsorption isotherms of β -lactoglobulin under non-isocratic conditions*. AIChE Journal, 38:1969 .
- BECKLEY, K., VERHAERT, P., VAN DER WIELEN, L., HUBBUCH, J. and OTTENS, M. (2009). *Rational and systematic protein purification process development: The next generation*. Trends in Biotechnology, 27(12):673 .
- BEREZKIN, V. (1989). *Biography of Mikhail Semenovitch Tswett and translation of Tswett's preliminary communication on a new category of adsorption phenomena*. Chemical Reviews, 89(2):279.
- BERGQVIST, M. H. J. and KAUFMANN, P. (1993). *A multivariate optimization of triacylglycerol analysis by high-performance liquid chromatography*. Lipids, 28(7):667.
- BROOKS, C. and CRAMER, S. (1992). *Steric Mass-Action Ion Exchange: Displacement Profiles and Induced Salt Gradients*. AIChE Journal, 38:1969 .
- BUCHACHER, A. and IBERER, G. (2006). *Purification of intravenous immunoglobulin G from human plasma-aspects of yield and virus safety*. Biotechnology Journal, 1(2):148.
- BYLUND, D., BERGENS, A. and JACOBSSON, S. (1997). *Optimisation of chromatographic separations by use of a chromatographic response function, empirical modelling and multivariate analysis*. Chromatographia, 44(1-2):74.
- CHAN, S., TITCHENER-HOOKER, N., BRACEWELL, D. and SORENSEN, E. (2008). *A systematic approach for modeling chromatographic processes - application to protein purification*. AIChE Journal, 54(4):965 .

- CURLING, J. (2007). *History of Chromatography: Process Chromatography: Five Decades of Innovation*. Biopharm International.
- DISMER, F. and HUBBUCH, J. (2010). *3D structure-based protein retention prediction for ion-exchange chromatography*. Journal of Chromatography A, 1217(8):1343.
- EFRON, B. and TIBSHIRANI, R. (1993). *An Introduction to the Bootstrap*. Chapman and Hall, CRC Press, London, Boca Raton.
- FAHRMEIR, L. and HAMERLE, A. (1995). *Multivariate statistische Verfahren*. Walter de Gruyter & Co., Berlin.
- FAST, D., CULBRETH, P. and SAMPSON, E. (1982). *Multivariate and univariate optimization studies of liquid-chromatographic separation of steroid mixtures*. Clinical Chemistry, 28(3):444.
- F.D.A.U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES (editor) (2004). *Guidance for Industry PAT - A Framework for Innovative Pharmaceutical Manufacturing and Quality Assurance*.
- FERREIRA, S., BRUNS, R., DA SILVA, E., DOS SANTOS, W., QUINTELLA, C., DAVID, J., DE ANDRADE, J., BREITKREITZ, M., JARDIM, I. and NETO, B. (2007). *Statistical designs and response surface techniques for the optimization of chromatographic systems*. Journal of Chromatography A, 1158(1-2):2.
- GADAM, S., JAYARAMAN, G. and CRAMER, S. (1993). *Characterization of non-linear adsorption properties of dextran-based polyelectrolyte displacers in ion-exchange systems*. Journal of Chromatography A, 630(1-2):37 .
- GALLANT, S. (2004). *Modeling ion-exchange adsorption of proteins in a spherical particle*. Journal of Chromatography A, 1028(2):189.
- GALLANT, S., KUNDU, A. and CRAMER, S. (1995). *Optimization of step gradient separations: Consideration of nonlinear adsorption*. Biotechnology and Bioengineering, 47(3):355.
- GALLANT, S., VUNNUM, S. and CRAMER, S. (1996). *Optimization of preparative ion-exchange chromatography of proteins: Linear gradient separations*. Journal of Chromatography A, 725(2):295.
- GILES, C., SMITH, D. and HUITSON, A. (1974). *A general treatment and classification of the solute adsorption isotherm*. Journal of Colloid and Interface Science, 47(3):755 .
- GRITTI, F. and GUIOCHON, G. (2011). *Multi-location peak parking method: An important new tool for the study of mass transfer kinetics in liquid chromatography*. Journal of Chromatography A, 1218(7):896.
- GUIOCHON, G., FELINGER, A., SHIRAZI, D. and KATTI, A. (2006). *Fundamentals of Preparative and Nonlinear Chromatography*. Elsevier Academic Press, San Diego, USA.

REFERENCES

- HANSEN, S. K., SKIBSTED, E., STABY, A. and HUBBUCH, J. (2011). *A label-free methodology for selective protein quantification by means of absorption measurements*. Biotechnology and Bioengineering, 108(11):2661.
- HJERTÉN, S. (1964). *The preparation of agarose spheres for chromatography of molecules and particles*. Biochimica et Biophysica Acta, 79:393.
- JOSHI, M., SEIDEL-MORGENSTERN, A. and KREMLING, A. (2006). *Exploiting bootstrap method for quantifying parameter confidence intervals in dynamical systems*. Metabolic Engineering, 8:447.
- KELLEY, B. (2007). *Very large scale monoclonal antibody purification: the case for conventional unit operations*. Biotechnology Progress, 23(5):995.
- LANGMUIR, I. (1916). *The constitution and fundamental properties of solids and liquids: I. solids*. Journal of the American Chemical Society, 38(111):2221.
- LEBRUN, P., GOVAERTS, B., DEBRUS, B., CECCATO, A., CALIARO, G., HUBERT, P. and BOULANGER, B. (2008). *Development of a new predictive modelling technique to find with confidence equivalence zone and design space of chromatographic analytical methods*. Chemometrics and Intelligent Laboratory Systems, 91(1):4 . Selected papers presented at the Chemometrics Congress.
- LIERES, E. v. and ANDERSSON, J. (2010). *A fast and accurate solver for the general rate model of column liquid chromatography*. Computers & Chemical Engineering, 34(8):1180.
- MARDIA, K., KENT, J. and BIBBY, J. (1979). *Multivariate Analysis*. Academic Press Inc. Ltd., London.
- MICHAELIS, L. and MENTEN, M. (1913). *Die Kinetik der Invertinwirkung*. Biochemische Zeitschrift, 49(333-369):352.
- NATARAJAN, V., BEQUETTE, B. and CRAMER, S. (2000). *Optimization of ion-exchange displacement separations: I. Validation of an iterative scheme and its use as a methods development tool*. Journal of Chromatography A, 876(1-2):51.
- NGUYEN, A., AERTS, T., VAN DAM, D. and DE DEYN, P. (2010). *Biogenic amines and their metabolites in mouse brain tissue: development, optimization and validation of an analytical HPLC method*. Journal of Chromatography B, 878:3003.
- OSBERGHAUS, A., DRECHSEL, K., HANSEN, S., HEPBILDKLER, S., NATH, S., HAINDL, M., VON LIERES, E. and HUBBUCH, J. (2012a). *Model-integrated process development demonstrated on the optimization of a robotic cation exchange step*. Chemical Engineering Science, accepted.
- OSBERGHAUS, A., HEPBILDKLER, S., NATH, S., HAINDL, M., VON LIERES, E. and HUBBUCH, J. (2012b). *Determination of parameters for the steric mass action model - A comparison between two approaches*. Journal of Chromatography A, 1233:54.

- OSBERGHAUS, A., HEPBILDIKLER, S., NATH, S., HAINDL, M., VON LIERES, E. and HUBBUCH, J. (2012c). *Optimizing a chromatographic three component separation: A comparison of mechanistic and empiric modeling approaches*. Journal of Chromatography A, doi: 10.1016/j.chroma.2012.03.029.
- PORATH, J. and FLODIN, P. (1959). *Gel Filtration: A Method for Desalting and Group Separation*. Nature, 183:1657.
- PULLAN, L. (1988). *An Efficient Strategy for Optimization of Isocratic Mobile Phase Conditions for HPLC Separation of a Complex Mixture*. Journal of Liquid Chromatography & Related Technologies, 11(13):2697.
- SCHMIDT-TRAUB, H. (editor) (2006). *Preparative Chromatography*. Wiley-VCH, Weinheim, Germany.
- SHUKLA, A., BAE, S., MOORE, J., BARNTHOUSE, K. and CRAMER, S. (1998). *Synthesis and characterization of high-affinity, low molecular weight displacers for cation-exchange chromatography*. Industrial & Engineering Chemistry Research, 37(1010):4090.
- SOBER, H. and PETERSON, E. (1954). *Chromatography of proteins on cellulose ion-exchangers*. Journal of the American Chemical Society, 76(1):1711.
- STABY, A., JENSEN, R., BENSCH, M., KRARUP, J., NIELSEN, J., LUND, M., KIDAL, S., HANSEN, T. and JENSEN, I. (2007). *Comparison of chromatographic ion-exchange resins VI. Weak anion-exchange resins*. Journal of Chromatography A, 1164:82.
- SUSANTO, A., HERRMANN, T., VON LIERES, E. and HUBBUCH, J. (2007). *Investigation of pore diffusion hindrance of monoclonal antibody in hydrophobic interaction chromatography using confocal laser scanning microscopy*. Journal of Chromatography A, 1149(2):178.
- SUSANTO, A., TREIER, K., KNEIPS-GRÜNHAGEN, E., VON LIERES, E. and HUBBUCH, J. (2009). *High Throughput Screening for the Design and Optimization of Chromatographic Processes: Automated Optimization of Chromatographic Phase Systems*. Chemical Engineering & Technology, 32:140.
- TESKE, C., VON LIERES, E., SCHRÖDER, M., LADIWALA, A., CHRAMER, S. and HUBBUCH, J. (2009). *High Throughput Screening for the Design and Optimization of Chromatographic Processes: Automated Optimization of Chromatographic Phase Systems*. Chemical Engineering & Technology, 32:140.
- WALSH, G. (2003). *Pharmaceutical biotechnology products approved within the European Union*. European Journal of Pharmaceutics and Biopharmaceutics, 55(1):3.

Akademischer Lebenslauf

Anna Osberghaus (geb. Siudak)



Aktuelle Dienstadresse

Anna Osberghaus
KIT Campus Süd
Institut für Bio- und Lebensmitteltechnik
Bereich IV: Molekul. Aufarbeitung von Bioprodukten
Engler-Bunte-Ring 1, 76131 Karlsruhe

Telefon: +49 721 608-46239
Fax: +49 721 608-46240
Email: anna.osberghaus@kit.edu

Persönliche Daten

Geburtsdatum/-ort: 2.Sept. 1982, Bocholt, Deutschland
Familienstand: verheiratet

Akademischer Werdegang

- 05.2008 - jetzt **Doktorandin** am Institut für Bio- und Lebensmitteltechnik;
Karlsruhe Institut für Technologie

Thema der Doktorarbeit: „*Optimierung einer chromatographischen Multikomponententrennung „in silico“ unter der Verwendung von High-Throughput-Screening-Daten*“
- 2007 **Diplom in Mathematik** (Sehr Gut), Universität Oldenburg;

Thema der Diplomarbeit: „*Robuste Parameterschätzung, Modelldiskriminierung und optimale Versuchsplanung am Beispiel von In-vitro-Datensätzen zur Benzaldehydlyase*“
- 02.2006 - 07.2006 **Jagiellonen-Universität Krakau, Polen**
Erasmus-gefördertes Studium der Mathematik und Soziologie
- 10.2002 - 12.2007 **Carl-von-Ossietzky Universität Oldenburg**
Studium der Mathematik
(Schwerpunkt: Anwendung in den Lebenswissenschaften)

Akademische Tätigkeiten

- 05.2008 - jetzt **Wissenschaftliche Mitarbeiterin**, Karlsruhe Institut für Technologie, Institut für Bio- und Lebensmitteltechnik (u.a. Praktikumsleitung und Betreuung von Diplomanden)
- 02.2008 - 04.2008 **Wissenschaftliche Mitarbeiterin**, Forschungszentrum Jülich, Institut für Biotechnologie 2
- 05.2007 - 12.2007 **Wissenschaftliche Hilfskraft**, Forschungszentrum Jülich, Institut für Biotechnologie 2
Forschungsprojekt zur Modellierung enzymatischer Vorgänge
- 03.2005 - 04.2005 **Praktikum**, Landesamt für Daten und Statistik in Düsseldorf

Curriculum Vitae

Anna Osberghaus (née Siudak)

Address

Anna Osberghaus
KIT Campus Süd
Institute of Process Engineering in Life Sciences
Section IV: Biomolecular Separation Engineering
Engler-Bunte-Ring 1, 76131 Karlsruhe



Tel: +49 721 608-46239
Fax: +49 721 608-46240
E-mail: anna.osberghaus@kit.edu

Education

- 05.2008 - now **Karlsruhe Institute for Technology, Karlsruhe, Germany**
pHD-student at the Institute of Process Engineering in Life Sciences;

pHD Thesis: „*Optimization of a chromatographic multi-component separation "in silico" using HTS-data*“
- 2007 **Diploma in Mathematics (1,0)**, Oldenburg University;

diploma thesis: „*Robust parameter estimation, model discrimination and optimal experimental design using in-vitro data of benzaldehyde lyase as a case study*“
- 02.2006 - 07.2006 **Jagiellonian University, Poland, Cracow**
studies in applied mathematics and sociology funded by the EU
- 10.2002 - 12.2007 **Carl-von-Ossietsky-University, Germany, Oldenburg**
studies in mathematics
(focus: applications in life sciences)

Work experience

- 05.2008 - now **Scientific Assistant**, Karlsruhe Institute for Technology,
Institute of Process Engineering in Life Sciences (i.a.project work and supervision of student projects)
- 02.2008 - 04.2008 **Scientific Assistant**, Research Center Jülich, Institute for Biotechnology 2
- 05.2007 - 12.2007 **Graduate Assistant**, Research Center Jülich, Institute for Biotechnology 2
Project: Modeling of enzymatic activity
- 10.2006 - 04.2007 **Student Assistant** at the Department of Mathematics, University of Oldenburg
- 03.2005 - 04.2005 **Internship**, State Office for data handling and statistics Düsseldorf

Scientific Publications

Talks

Anna Osberghaus, Tobias Hahn, Vincent Heuveline, Jürgen Hubbuch

Modellbasierte Fehlerursachendiagnostik für flüssigchromatographische Anwendungen

Ideenwettbewerb BW, Stuttgart, Germany, 16.-18. January, 2012

Anna Osberghaus, Pascal Baumann, Jürgen Hubbuch

Propagation of error: Uncertainty analyses in high throughput experimentation

ECAB Conference, Berlin, Germany, 25.-29. September, 2011

Anna Osberghaus, Stefan Hepbildikler, Hans Rogl, Susanne Nath, Eric von Lieres, Jürgen Hubbuch

Combination of HTE and Mechanistic Modeling for a Rapid and Highly Predictive Optimization of an Ion Exchange Chromatography Step

HTPD Conference, Cracow, Poland, 5.-7. October, 2010

Anna Siudak, Eric von Lieres, Christine Müller

Mathematical modelling – a short overview of benefits and shortcomings based on enzyme catalysis data as a case study

KIT PhD Symposium, Karlsruhe, Germany, 17. March, 2009

Posters

Anna Osberghaus, Tobias Hahn, Vincent Heuveline, Jürgen Hubbuch

Model based diagnostics and control of processes in ion exchange chromatography

DECHEMA Conference, Potsdam, Germany, 30. May – 1. June, 2011

Anna Siudak, Eric von Lieres, Susanne Nath, Jürgen Hubbuch

Unavoidable variance trends in data obtained by experiments on robotic platforms – how to consider such effects in model parameter estimation

PROCESSNET, Mannheim, Germany, 8-10. September, 2009

Eric von Lieres, Patrick Diederich, Anna Siudak, Arthur Susanto, Katrin Treier, Jürgen Hubbuch

Optimization of High Throughput Process Optimization

Recovery XIII, Quebec, Canada, 22. – 27. June, 2008

Papers

Anna Osberghaus, Katharinja Drechsel, Sigrid Hansen, Stefan Hepbildikler, Susanne Nath, Markus Haindl, Eric von Lieres, Jürgen Hubbuch

Model-integrated process development demonstrated on the optimization of a robotic cation exchange step

accepted by Chemical Engineering Science (2012)

Anna Osberghaus, Pascal Baumann, Stefan Hepbildikler, Susanne Nath, Markus Haindl, Eric von Lieres, Jürgen Hubbuch

Detection, quantification and propagation of uncertainty in high throughput experimentation by Monte-Carlo methods

accepted by Chemical Engineering Technology (2012)

Anna Osberghaus, Stefan Hepbildikler, Susanne Nath, Markus Haindl, Eric von Lieres, Jürgen Hubbuch

Determination of parameters for the steric mass action model - A comparison between experimental and modeling approaches

Journal of Chromatography A, 1233 (2012) 54-65

Anna Osberghaus, Stefan Hepbildikler, Susanne Nath, Markus Haindl, Eric von Lieres, Jürgen Hubbuch

Optimizing a chromatographic three component separation: A comparison of mechanistic and empiric modeling approaches

Journal of Chromatography A (2012), doi:10.1016/j.chroma.2012.03.029

Anna Siudak, Eric von Lieres, Christine Müller

Estimation, model discrimination, and experimental design for implicitly given nonlinear models of enzyme catalyzed chemical reactions

Mathematica Slovaca 59 (2009) 593-610

Diploma thesis

Anna Siudak

Robuste Parameterschätzung, Modelldiskriminierung und optimale Versuchsplanung am Beispiel von In-vitro-Datensätzen zur Benzaldehydlyase

Universität Oldenburg, 2007