Advances in Model-based Downstream Process Development

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

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

genehmigte

DISSERTATION

von
Dipl.-Math. techn. Tobias Hahn
aus Erfurt

"WE CAN ONLY SEE A SHORT DISTANCE AHEAD,
BUT WE CAN SEE PLENTY THERE THAT NEEDS TO BE DONE."

- ALAN TURING

Danksagung

Zusammenfassung

In der Biotechnologie dient das sogenannte Downstream Processing (DSP) dazu, eine Zielkomponente aus einer heterogenen Mischung zu gewinnen. Dabei hat es hat in der Regel den größten Kostenanteil an einem biopharmazeutischen Produktionsablauf. Insbesondere bei monoklonalen Antikörpern reduziert sich die industrielle DSP-Entwicklung üblicherweise auf eine Folge von chromatographischen Prozessen. Chromatographieprozesse sind robust, erreichen eine hohe Selektivität bei milden Bedingungen und werden allgemein von Regulierungsbehörden akzeptiert. Die wichtigste Klasse bei der Aufreinigung von Bioprodukten ist die Flüssig-Fest-Säulenchromatographie, bei der die Probe in einem flüssigen Laufmittel gelöst und durch eine stationäre Phase gepumpt wird. Letztere kann aus einer Packung von porösen Teilchen oder einem Monolithen bestehen. Die Moleküle in der mobilen Phase werden zunächst durch das Flüssigkeitsvolumen außerhalb der Teilchen bzw. in den Makroporen transportiert und diffundieren dann in das Mikroporensystem. Die physikalischen und/oder chemischen Eigenschaften der stationären Phase werden dabei so gewählt, dass einige Komponenten der injizierten Probe stärker als andere zurückgehalten werden. In dieser Dissertation liegt der Fokus auf Adsorptionschromatographie, bei der die Bestandteile der Probe auf der Oberfläche der festen Phase unterschiedlich gut adsorbieren. Beispiele von Adsorptionschromatographie umfassen Affinitäts-, Ionenaustausch- und hydrophobe Interaktionschromatographie.

Um den Prozessentwicklungsaufwand zu reduzieren, wurden in der Industrie so genannte Plattformprozesse entwickelt, bei denen für das jeweilige Produkt nur wenige Parameter einer festgelegten Folge von Chromatographieschritten optimiert werden. Die sich derzeit in Entwicklung befindlichen Nicht-Standard-Antikörper, wie Antibody-Drug-Conjugates, oder neue Arten von Biopharmazeutika, z.B. virusähnliche Partikel, passen allerdings zunehmend nicht mehr in das steife Gerüst der Plattformprozesse. Hier muss der volle Umfang der Freiheitsgrade berücksichtigt werden, um effektive und effiziente Chromatographieprozesse zu entwickeln. Derzeit werden hierfür Hochdurchsatzverfahren verwendet, um einen großen Teil des Parameterraums zu screenen. Dafür wird eine vergleichsweise hohe Probenmenge benötigt, die häufig in der frühen Prozessentwicklung nicht zur Verfügung steht. Aus diesem Grund gewinnen Simulationswerkzeuge zunehmend die Aufmerksamkeit der Pharma-Industrie, um optimale Prozessparameter zu bestimmen. Modellbasierte Prozessoptimierung macht aufwendige Screening-Experimente überflüssig, sobald ein Modell für den spezifischen Prozessschritt kalibriert wurde.

Der Stofftransport durch die Säule und Poren wird durch die Modellierung der fluiddynamischen Effekte abgebildet, während die Retention der Moleküle durch empirische oder mechanistische Modelle zur Adsorption/Desorption und/oder Reaktion beschrieben wird.

Solche modellbasierten Ansätze werden in der akademischen Forschung entwickelt und verwendet, der Technologietransfer in die Industrie ist aber nach wie vor begrenzt. Die Hauptursache ist die fehlende Integration modellbasierter Werkzeuge in Laborarbeitsabläufe oder, mit anderen Worten, der Übergang von etablierten experimentellen Verfahren zur Anwendung computergestützter Methoden. Obwohl viele Forschungsgruppen ihre eigenen Simulationswerkzeuge entwickelt und damit vielversprechende Ergebnisse erzeugt haben, sind nur wenige Simulatoren öffentlich verfügbar und keines der bisherigen Produkte kombiniert hohe Simulationsgeschwindigkeit mit Erweiterbarkeit und Benutzerfreundlichkeit. Um Simulationswerkzeuge erfolgreich zu integrieren, müssen die enthaltenen Modelle die Standardoperationen abdecken, Parameterschätzung und -optimierung aus routinemäßig erhobenen Daten vereinfachen und flexibel für neue Aufgaben angepasst werden können. Bislang gibt es noch keine wissenschaftliche oder kommerzielle Software, um andere Modi als Ionenaustausch-Chromatographie mechanistisch fundiert zu simulieren. Selbst die existierenden Werkzeuge benötigen Informationen, die oft in industriellen Fallstudien nicht verfügbar sind, insbesondere molare Eingangskonzentrationen aller Komponenten des Probenmaterials. Im Lichte der Qualität-by-Design-Initiative ist es weiterhin nicht ausreichend, ein beliebiges funktionierendes Modell zu finden, sondern es wird eines mit einer schlüssigen physikalischen Interpretation gefordert.

Das Ziel dieser Arbeit war es, eine leistungsfähige Toolbox zur Bioprozessentwicklung mit zugehörigen Methoden zu erstellen, die diese Lücke füllt. Sie soll exakte und plausible Simulationsergebnisse liefern, sodass kostspielige Experimente ersetzt, die Entwicklungszeit minimiert, Produktionsprozesse optimiert und das Prozessverständnis verbessert werden können. Entsprechend müssen in dieser Toolbox mechanistische Modelle für die wichtigsten chromatographischen Grundoperationen enthalten sein sowie Optimierungsverfahren für Parameterschätzung und -optimierung.

Mit diesen Zielen vor Augen wurde das Software-Paket „ChromX" für schnelle und einfache *in silico* DSP-Entwicklung entworfen und implementiert. Es stellt den Kern dieser Arbeit dar, in dem es die Simulation von sequentiellen und kontinuierlichen Mehrsäulenverfahren erlaubt sowie Parameterschätzung, -optimierung und -sampling, einschließlich der zugehörigen Fraktionierung. Die verwendeten numerischen Methoden umfassen Zeitschrittverfahren mit finiten Differenzen, u.a. klassische und Fraction-Step-θ-Verfahren in Verbindung mit adaptiver Zeitschrittbestimmung. Die Ortsdiskretisierung wurde mit finiten Elemente durchgeführt. Es stehen lineare und quadratische Elemente mit oder ohne Streamline-Upwind-Petrov-Galerkin-Stabilisierung zur Oszillationsunterdrückung zur Verfügung.

Diese Dissertation besteht aus neun Publikationen und Manuskripten, die Forschungsergebnisse zu verschiedenen Aspekten der Chromatographie-Modellierung, Modellkalibrierung und modellbasierten Prozessentwicklung wiedergeben.

Die erste Publikation untersucht die Verwendbarkeit von UV-Spektren mehrerer Proteinspezies und multivariater Statistik für Hochdurchsatz-Batch-Chromatographie. Aus den gewonnenen Daten konnten SMA-Isothermenparameter bei zwei pH-Werten und verschiedenen Ionenstärken bestimmt und damit mechanistische Einblicke in das kompetitive Bindungsverhalten gewonnen werden. Das Modell konnte die beobachteten starken Verdrängungseffekte genau wiedergeben, die für das schwächer bindende Protein zu einem Adsorptionsmaximum bei geringer Konzentrationen in der mobilen Phase führte.

Die zweite Publikation erklärt die Arbeitsweise der Software ChromX. ChromX soll nicht nur die Forschungstätigkeit von Experten unterstützen, sondern auch für Benutzer ohne Modellierungshintergrund, wie Laboranten und Studenten, bedienbar sein. ChromX wurde daher mit einer benutzerfreundlichen Oberfläche und einer umfassende Dokumentation ausgestattet. Eine klare Verfahrensvorschrift wurde für die Modellkalibrierung entwickelt und in einem Laborpraktikum für Studenten erfolgreich getestet. Die Publikation stellt die Ergebnisse des Praktikums im Wintersemester 2013/2014 vor, in dem ein monoklonaler Antikörper aus einer ternären Mischung aufgereinigt werden musste.

Ohne Vorwissen über das Adsorptionsverhalten der Komponenten ist die „inverse Methode" ein geeignetes Verfahren, welche die Parameter systematisch verändert, bis eine Übereinstimmung des gemessenen Chromatogramms mit der Modellvorhersage erreicht ist. Für deterministische Parameterschätzung und anschließende Prozessoptimierung ist dabei die Bestimmung von Ableitungen der Modellgleichungen nach den unbekannten Parametern von entscheidender Bedeutung. In der Vergangenheit wurden hierfür Approximation mit finiten Differenzen und automatische Differenzierung verwendet, die aber beide sehr rechenaufwendig sind. In der dritten Publikation wird stattdessen die „adjungierte Methode" für ein breites Modellspektrum aus der Chromatographie formuliert. Es konnte gezeigt werden, dass die Herleitung und Implementierung praktisch möglich ist, dass Vorwärtslöser leicht in adjungierte Löser umgewandelt werden können und signifikante Geschwindigkeitssteigerungen gegenüber früheren Ansätzen erreicht werden.

Während es das Ziel der Parameterschätzung ist, die Übereinstimmung von Simulation und Messung zu maximieren, beschäftigt sich das vierte Manuskript mit der „Optimalen Versuchsplanung" (engl. Optimal Experimental Design, OED), die zunächst die experimentellen Bedingungen mit dem höchsten Informationsgehalt bestimmt. Sie minimiert dabei die Unsicherheit der Parameter, gemessen als Funktion der Kovarianzmatrix. Mit diesem Ansatz werden die Parameter mit der höchsten statistischen Qualität mit der geringsten Anzahl an Versuchen gefunden. Hier wurde OED zum ersten Mal für Säulenchromatographie von Proteinen durchgeführt, einschließlich der näherungsweisen Berechnung von Kovarianzen und Konfidenzintervallen. Eine Fallstudie vergleicht Parameterschätzung aus herkömmlicher Versuchsplanung mit OED für das Modellprotein Glucoseoxidase auf Ionenaustausch-, hydrophober Interaktions- und Mixed-Mode-Chromatographie. Darüber hinaus wurde das klassische General-Rate-Modell mit salzabhängiger Porendiffusion erweitert. Die Ergebnisse unterstreichen die verbesserte Zuverlässigkeit mit OED bei potentiell viel geringerem Zeit- und Materialaufwand.

Die übrigen fünf Manuskripte sind Fallstudien zu Modellkalibrierung und Prozessoptimierung. Die erste Fallstudie behandelt modellbasierte konzertierte Prozessoptimierung von zwei aufeinander folgenden Ionenaustausch-Säulenchromatographieschritten. Die Vorteile wurden durch Vergleich mit einem sequentiellen Ansatz demonstriert, der die beiden Schritte nacheinander optimiert. Die *in silico* optimalen Bedingungen konnten im Laborversuch reproduziert werden. Weiterhin wurde ChromX erfolgreich zur Fehlerdiagnostik verwendet: Abweichungen von Messungen und Simulation konnten auf einen unvollständigen Pufferaustausch, d.h. einer unerwartet hohen Salzkonzentration beim Probenauftrag des zweiten Säulenlaufs, zurückgeführt werden.

Während alle bisherigen Experimente unter präzise definierten Bedingungen durchgeführt wurden, konzentrieren sich die übrigen vier Fallstudien auf praktische Prozessentwicklung. Hier sind die molaren Konzentrationen aller Spezies in der Probe in der Regel

unbekannt und können für komplexe Gemische nur mit hohem analytischen Aufwand bestimmt werden. Präparative Chromatographie wird deshalb normalerweise durch UV-Messungen gesteuert und analytische Chromatographieläufe bestimmen die Reinheit oft aus den UV-Peakflächen der Verunreinigungen. Während UV-Absorptionsmessungen eine solch wichtige Rolle in der Bioprozessentwicklung spielen, konnten sie bislang nicht direkt zur Modellierung verwendet werden. Die am häufigsten verwendeten Modelle setzen bekannte Massen- oder molare Konzentrationen voraus.

Die sechste Publikation beschreibt einen neuartigen Transformationsansatz, der die Modellierung anhand von UV-Absorptionsdaten als Konzentrationsmaß erlaubt und zeigt zudem, dass die unbekannten molaren Konzentrationen aus experimentellen Daten unter nichtlinearen Bedingungen bestimmt werden können. Eine Fallstudie demonstriert die mechanistische Modellierung eines industriell relevanten Chromatographieschrittes für ein komplexes Ausgangsmaterial, das mindestens 11 Verunreinigungen enthält. Das Modell konnte nur anhand von UV-Absorptionsdaten ohne Kenntnis der genauen Probenzusammensetzung erstellt werden.

Das siebte Manuskript geht noch einen Schritt weiter und beschäftigt sich mit einem Datensatz aus der industriellen Antikörperaufreinigung mit unbekannten Verunreinigungskonzentrationen und ohne unterscheidbare Peaks. Die Kombination der zuvor entwickelten Methode zur Modellkalibrierung basierend auf UV-Daten mit hochaufgelösten Fraktionsanalysen, erlaubte die Simulation und Optimierung dieses komplexen Trennproblems. Die optimalen Prozessparameter aus konventioneller Prozessentwicklung konnten *in silico* reproduziert werden.

Die achte Veröffentlichung unterstreicht, dass die modellbasierte Prozessentwicklung besonders wertvoll ist, wenn nur eine geringe Probenmenge zur Verfügung steht. Die Publikation stellt einen kombinierten Optimierungsansatz auf Basis von Hochdurchsatz-Kultivierungsexperimenten im Mikrolitermaßstab und Chromatographie-Modellierung vor. Eine Reihe von zufällig ausgewählten Upstream-Bedingungen wurde hierfür mit ChromX modelliert. Es konnte gezeigt werden, dass die Verunreinigungen in den verschiedenen Bedingungen identisch sind und sich nur in ihren relativen Anteilen unterscheiden. Eine nachgeschalteten multikriterielle Prozessoptimierung ermittelte die besten Upstream/Downstream-Bedingungen, die im Labor erfolgreich validiert wurden.

Im neunten und letzten Manuskript wird gezeigt, dass die Simulation und Optimierung von chromatographischen Verfahren mit ChromX nicht auf Festbetten von porösen Partikeln und Proteine bis Antikörpergröße beschränkt ist. Ein Aufreinigungsschritt für virusähnliche Partikel (engl. VLPs) mit Membranadsorber-Kapsulen konnte erfolgreich im Modell abgebildet werden. Zum ersten Mal konnte eine Prozessmodellierung unter minimalem experimentellen Aufwand für die Entwicklung eines selektiven, robusten und skalierbaren Prozesses für ein komplexes VLP-Trennproblem eingesetzt werden.

Abstract

In biotechnology, the so-called downstream processing (DSP) aims to purify a target component from a very heterogeneous mixture and constitutes the biggest expenditure of a biopharmaceutical production sequence. Especially for monoclonal antibodies, industrial DSP development is commonly based on a sequence of chromatographic processes. These offer high selectivity at mild conditions as well as high robustness and general acceptance by regulatory authorities. The most important class in bioseparation is liquid-solid column chromatography. Here, the sample is dissolved in a liquid mobile phase and pumped through a stationary phase consisting of a packed bed of porous particles or a porous monolith. The molecules in the mobile phase are transported through the film layer outside of the particles or micro pores, then enter the pore system and diffuse inside them. The physical or chemical properties of the stationary phase are chosen such that some components of the injected sample are retained more strongly than others. The focus of this thesis lies on adsorption chromatography where the components of the sample adsorb on the surface of the solid phase with different strengths. Examples of adsorption chromatography include affinity, ion-exchange and hydrophobic interaction chromatography.

To reduce the process development effort, so-called platform processes have been developed in industry with a fixed order of chromatographic steps and only few degrees of freedom to be optimized for the specific product. Currently, non-standard antibodies such as antibody-drug-conjugates or new types of biopharmaceuticals, e.g. virus-like particles, emerge that do not fit into platform processes anymore. Hence, the full amount of degrees of freedom has to be considered to find effective and efficient chromatography set-ups. Nowadays, high-throughput methods are used to screen a large part of the design space but require a comparably high amount of sample that is often not available for early stage process development. Thus, simulation tools are increasingly gaining the attention of the pharmaceutical industry in order to determine the optimal process parameters. Model-based process optimization makes elaborate screening experiments redundant once the model has been calibrated to the specific process step.

The mass transport through the column and pores is described by modeling the fluid dynamics effects, while the retention of the species is described by empirical or mechanistic models for adsorption/desorption and/or reaction. In most cases, one-dimensional convection-diffusion-reaction equations provide for a good agreement with measurements. Such model-based approaches are developed and used in academia but technology transfer to industry is still limited. The main reason is the lack of integration of novel tools into laboratory work-flows or in other words the transition from established experimental procedures to computer-aided process development. Although many research groups

developed their own simulation tools and published results generated with it, only few simulators are publicly available and none of the available tools combines high simulation speed with extensibility and usability. To integrate simulation tools successfully, the underlying models have to cover the usual operations and materials, simplify parameter estimation and optimization from routine data collection while maintaining high flexibility. To the author's knowledge, there is no academic or commercial software available to accurately simulate other modes than ion-exchange chromatography. Even these tools require information that are often not available in industrial settings, especially molar feed concentrations of all species in the sample material. Furthermore, it is not sufficient to find any kind of model, but one with a meaningful physical interpretation.

The objective of the presented research was to create a powerful bioprocess development toolbox with accompanying work flows. It should deliver accurate simulation results, allow to replace costly experimentation, enable comprehensive parameter studies for better process understanding, and facilitate optimization of production processes and development cycles. With these goals in mind, the software package ChromX was designed and implemented for fast and easy *in silico* downstream process optimization and constitutes the back-bone of this work. It supports sequential and continuous multi-column processes, parameter estimation, optimization, and sampling. Numerical methods include time marching with finite differences, e.g. classical and fractional step theta schemes, combined with adaptive time step determination. Space discretization is accomplished with finite elements: linear and quadratic elements are available with or without Streamline-Upwind-Petrov-Galerkin (SUPG) stabilization for oscillation suppression.

This thesis consists of nine publications and manuscripts that focus on different aspects of chromatography modeling, model calibration and model-based process development. The first four manuscripts present results generated with novel computational methods, the following five are case studies of model-calibration and process optimization.

The first paper explores the competitive binding behavior of proteins in high-throughput ion-exchange batch chromatography. UV spectra and multivariate statistics are employed to distinguish two protein species. The quality of the analytical data allowed to fit a Steric Mass Action model at two pH values and various ionic strengths. The model could accurately describe the observed strong displacement effects leading to an adsorption maximum for the weaker binding protein at low concentrations in the mobile phase. This behavior was predicted by theory but never observed as clear as in this publication.

The second paper explains the working principle of the ChromX software. ChromX is not only intended to support the research of experts but should also to be operable by users without modeling background such as laboratory technicians and students. ChromX was hence extended with a user-friendly interface and comprehensive documentation. A clear standard-operation procedure was developed for model calibration and successfully tested in laboratory courses for undergraduate students. The paper presents the results of the 2014 lab course, in which a monoclonal antibody was to be purified from a ternary mixture.

With no a priori knowledge about the components' behavior, curve fitting is a suitable option which alters parameters in a systematic fashion to achieve a match of the measured chromatogram and the model prediction. The determination of parameter derivatives

of liquid chromatography models is hence crucial for deterministic parameter estimation and subsequent process optimization. Finite difference approximations and automatic differentiations have been employed in the past but are computationally expensive.

In the third publication, the adjoint method is formulated for a wide range of models used in liquid chromatography. It could be shown that derivation and implementation is manageable and that forward solvers can be easily turned into adjoint solvers, achieving significant speed-ups over previous approaches. Furthermore, a new formulation of the Steric Mass Action model was developed that improves parameter identifiability. A case study with three model proteins demonstrated the effectiveness for both, parameter estimation and optimization.

While the objective of parameter estimation is to minimize the disagreement of simulation and measurement, the fourth manuscript deals with the technique of Optimal Experimental Design (OED) which takes one step backwards and first finds the experimental set-up with highest information content. It then minimizes the uncertainty of the parameters, measured as function of the covariance matrix. Using this approach, the parameters with the highest statistical quality are found using the least number of experiments. For the first time, OED was implemented for column chromatography of proteins, including the computation of approximate covariance matrices and confidence intervals.

A case study evaluates parameter estimation by conventional experimental design and OED for the model protein glucose oxidase on ion-exchange, hydrophobic interaction and mixed-mode resins. Using OED, the covariance ellipsoid shrunk after each additional experiment. Furthermore, the General Rate Model for the pore concentration was extended with salt-dependent diffusion. The results indicate that peak distortions that were previously attributed to different binding orientations are in fact caused by fluid dynamic effects.

The fifth manuscript performs model-based concerted process optimization of two consecutive ion-exchange column chromatography runs with model proteins. The overall optimum was superior to sequential optimization of the two consecutive steps and could be reproduced in lab experiments. Furthermore, ChromX was successfully employed for error diagnostics: deviations of measurement and simulation could be traced back to an incomplete buffer exchange leading to an unexpected high salt concentration during loading of the second column.

While all manuscripts so far were conducted under precisely defined conditions, the remaining four case studies focus on practical process development. Here, the molar concentrations of all species in the sample are usually unknown and can only be determined with high analytical effort for complex feedstocks. Therefore, preparative chromatography steps are usually controlled using UV measurements, and analytical chromatography runs often determine peak areas of impurities as a measure of overall purity. While UV absorption measurements play such an important role in bioprocess development, they could not be used directly for modeling chromatography steps as the commonly applied models rely on mass or molar concentrations.

The sixth manuscript describes a novel transformation approach that allows for modeling based on UV absorption data as measure of concentration and even proves that the unknown molar concentrations can be read from the experimental data under non-linear

conditions. A study presents mechanistic modeling of an industrially relevant chromatography setting with a crude feed stock including at least 11 impurities. An accurate model could be obtained just based on UV absorption data without any knowledge of the feed composition.

The seventh manuscript goes one step further and deals with a data set from industrial antibody purification with unknown impurity concentrations and no distinguishable peaks. Combining the previously developed method for model calibration based on UV data with high resolution fraction analyses allowed to simulate and optimize the complex separation problem. The optimal process parameters found by conventional process development could be reproduced *in silico*.

The eighth publication underlines that model-based process development is especially valuable when only a small sample volume is available. The paper presents a combined optimization approach based on high-throughput micro-scale cultivation experiments and chromatography modeling. A set of randomly picked upstream conditions was modeled using ChromX. It could be shown that the impurities resulting from the different conditions are in fact identical and only differ in their relative amounts. A multi-criteria downstream process optimization identified the best overall upstream-downstream conditions that were validated successfully in the laboratory.

The ninth and final manuscript demonstrates that the simulation and optimization with ChromX is not limited to packed beds of porous beads and proteins up to anti-body size: a purification step for virus-like particles (VLPs) could be successfully designed for membrane adsorber capsules. In contrast to column chromatography, the mobile is pumped through a cylinder ring formed by a spirally wound membrane. For the first time, process modeling enabled the *in silico* design of a selective, robust and scalable process with minimal experimental effort for a complex VLP feedstock.

Contents

1 | Introduction

The so-called downstream processing (DSP) aims to purify a target component from a very heterogeneous mixture and constitutes the biggest expenditure of a biopharmaceutical production sequence. Especially for monoclonal antibodies, industrial DSP development is commonly based on a sequence of chromatographic processes.

According to the International Union of pure and applied Chemistry (IUPAC), chromatography is defined as a "physical method of separation in which the components to be separated are distributed between two phases, one of which is stationary while the other moves in a definite direction." The term is derived from Greek *chroma* = color and *graphein* = writing. It was introduced in 1906 by the Russian botanist Michael S. Tswett (* 1872, † 1919), who published a paper on separating plant pigments extracted by hexane using a glass column packed with finely divided $CaCO_3$.

Chromatography can be further classified according to the states of the two phases:

		Stationary Phase	
		Solid	Liquid
Mobile Phase	Gas	Gas-Solid-Chrom.	Liquid-Solid-Chrom.
	Liquid	Liquid-Solid-Chrom.	Liquid-Liquid-Chrom.

In bioseparation, the most important class is liquid-solid column chromatography. Here, the stationary phase consists of a packed bed of porous particles or a monolith, held in a cylinder, made of glass, metal or plastics. For simplicity, a packed bed of porous particles will be assumed in the following.

As illustrated in Figure 1.1, the molecules in the mobile phase are transported through the fluid outside of the particles, and then enter the particle's pore system and diffuse inside the pores. The physical or chemical properties of the stationary phase are chosen such that some components of the injected sample are retained more strongly then others. In Size Exclusion Chromatography (SEC), the pore of the stationary phase have a precisely defined diameter. Large molecules cannot enter as deep as small molecules and leave the column earlier. In Partition Chromatography, differences in solubility in a stationary liquid (or liquid-supported solid) are used for separation. In this thesis, the focus lies on Adsorption Chromatography, where the components of the sample adsorb on the surface of the solid phase with different strengths. In case the molecule stays bound to the surface, a desorption step is employed to induce diffusion out of the particle.
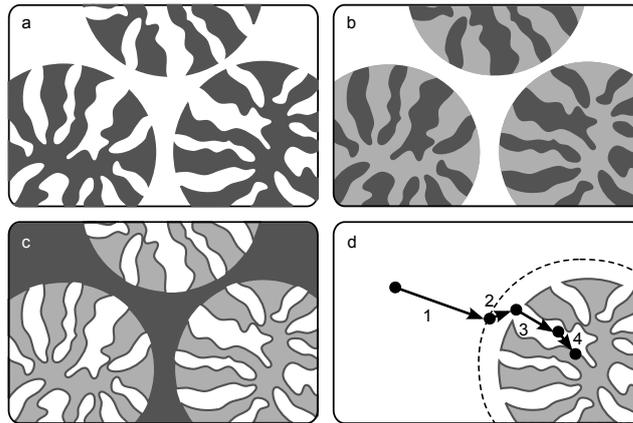
Figure 1.1: Structure of packed beds (a-c) and adsorption process (d). The total column volume is divided into the volume of the stationary phase (a), the volume of mobile phase in the pore system (b), and the interstitial volume of the mobile phase (c). Molecules are transported through the interstitial volume (d.1), pass a boundary layer (also referred to as film) around the beads (d.2), diffuse within the pore system (d.3), and adsorb onto the surface (d.4). [Chap. 3]
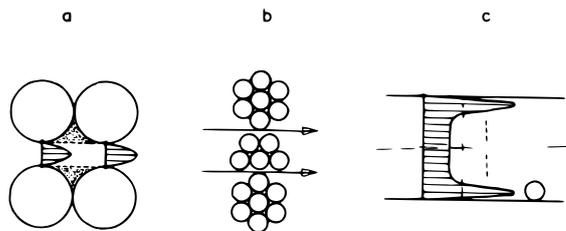
Figure 1.2: Fluid dynamic effects in a porous column. Adapted from [182].

Examples of adsorption chromatography include affinity chromatography which is based on the affinity of e.g. proteins to specific ligands such as enzymes. In ion-exchange chromatography, cationic or anionic groups are bonded to the resin surface and counter ions in the mobile phase will be attracted to the resin. Hydrophobic interaction chromatography uses differences in hydrophobicity of the components for separation.

1.1 Fundamentals of Chromatography Modeling and Simulation

The mass transport through the column and pores is described by modeling the fluid dynamics effects, while the retention of the species is described by empirical or mechanistic models for adsorption/desorption and/or reaction [72; 129].

The fluid dynamic effects include flow in the loops of the system and the flow through the porous column. Within the column, effects happen in different scales, examples include microscopic channels (Fig. 1.2a), mesoscopic packing non-idealities (Fig. 1.2b) and macroscopic wall effects (Fig. 1.2c). Modelling all these effects can become computationally infeasible and high-resolution measurements are necessary in order to distinguish the influences based on the recorded concentration distribution.

(a) 3D column geometry.

(b) 2D geometry, assuming radial symmetry.

(c) 1D geometry, assuming no wall effects.

(a) Convective transport.

(b) Convection and diffusion.

(c) Convection, diffusion and reaction.
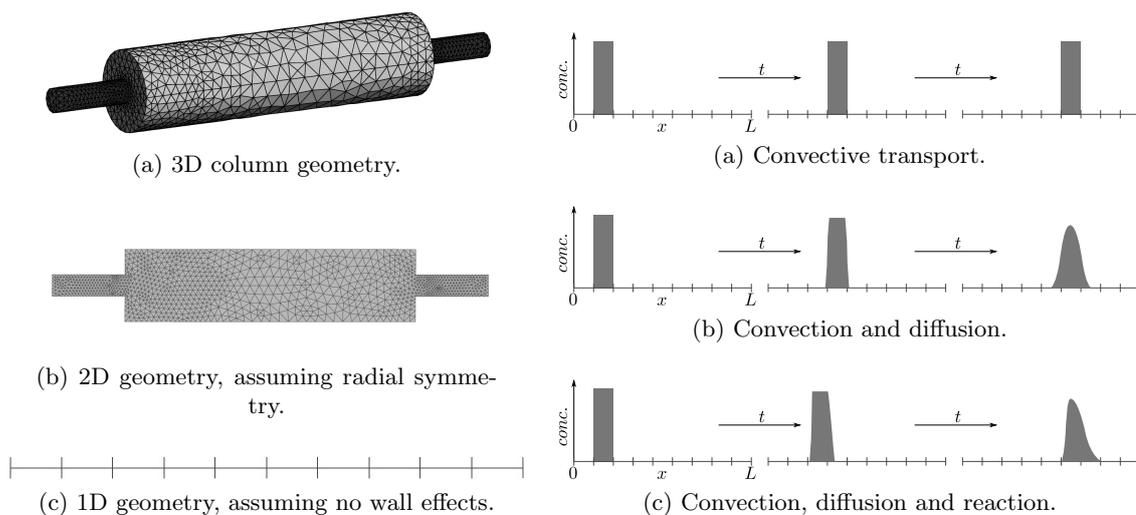
Figure 1.3: Complexity reduction from 3D to a 1D model.

Figure 1.4: Concentration transport in a chromatogaphic column.

A well packed column allows to assume radial symmetry and reduce the complexity from 3D to 2D (Fig. 1.3b). If the column is large enough for wall effects to be neglected, the injection of a rectangular pulse of a molecule that cannot enter the pore system will only lead to a symmetrically broadened peak at the outlet (Fig. 1.5) that can be modeled in 1D (Fig. 1.3c).

The transport in flow direction (Fig. 1.4a) depends on the pump speed and bed porosity (alias, *column porosity*). All diffusive effects in the interstitial volume leading to the broadening of an injected pulse are assumed to follow Fick's law of diffusion with a lumped *axial dispersion* coefficient (Fig. 1.4b). For smaller non-interacting molecules, additional effects are included, depending on the chosen model complexity. These include a *film transfer* coefficient that models the transition inhibition into and out of the pores by a boundary layer, and a *pore diffusion* coefficient that accounts for intra-particle diffusion. Diffusion on the surface is usually neglected [72]. Adsorption onto the surface is modeled by an *isotherm* equation (that describes the concentration of adsorbed species as a function of the concentration in the mobile phase (Fig. 1.6) and leads to unsymmetrical peak shapes (Fig. 1.4c).

1.1.1   Overview of Chromatography Models

A variety of models are employed depending on the phenomena occurring in the respective system, but most of them can be formulated using a common structure. The following summary follows the text books by Schmidt-Traub [129] and Guiochon *et al.* [72]. Both provide an extensive introduction to the topic, including all assumptions that apply to a particular simplification. The summary given below assumes familiarity with the notation of partial differential equations. A nomenclature is given in section B.1.1.

Current models for column chromatography describe the changes of concentrations in three volumes, the concentration c in the interstitial volume of the mobile phase, c_p in the pore volume of the mobile phase, and q the stationary phase concentration. Simplified models assume an equal concentration in c and c_p.
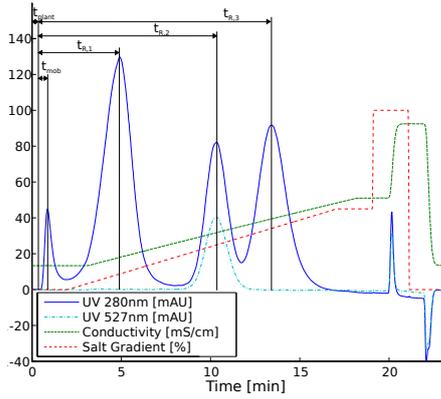
Figure 1.5: Detector signal over time with plant dead time t_{plant}, mobile phase residence time t_{mob} and components' retention times $t_{R,1}, t_{R,2}, t_{R,3}$. Adapted from [77].
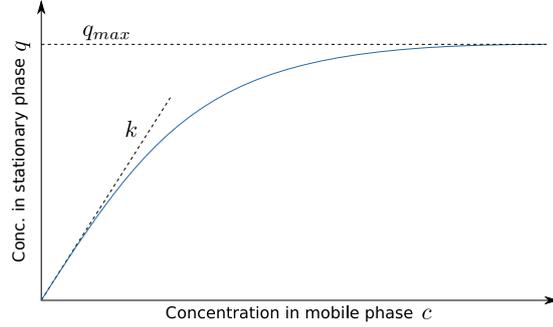
Figure 1.6: Adsorption isotherm saturation concentration q_{max} and slope k.

1.1.1.1   Equilibrium Dispersive Model

Typically, the concentration transport in the column is assumed to be independent of the cross-sectional position such that the interstitial concentration c only depends on the time $t \in [0, T]$ and axial position $x \in [0, L_{Col}]$ in a column of length L_{Col} and is influenced by convection, diffusion, and mass transfer into and out of the pore volume or stationary phase, e.g. the *Equilibrium Dispersive Model*

$$\frac{\partial c}{\partial t}(x,t) = \underbrace{-\frac{u(t)}{\varepsilon_{Tot}}\frac{\partial c}{\partial x}(x,t)}_{\text{Convection}} + \underbrace{D_{app}\frac{\partial^2 c}{\partial x^2}(x,t)}_{\text{Diffusion}} - \underbrace{\frac{1-\varepsilon_{Tot}}{\varepsilon_{Tot}}\frac{\partial q}{\partial t}(x,t)}_{\text{Mass transfer}} \qquad (1.1)$$

where u is the induced flow velocity, D_{app} the apparent dispersion coefficient, and ε_{Tot} the total porosity. Obviously, the interstitial flow velocity increases with lower porosity. The apparent dispersion coefficient covers all effects that lead to peak broadening in the column and the fraction in the mass transfer term accounts for the volume differences of the mobile and stationary phases.

Several assumptions were made to consider the velocity and axial dispersion as constants in space, the most fundamental ones being a uniformly packed column, an incompressible liquid, and a constant viscosity. This model is complemented by the Danckwerts boundary conditions at the inlet and outlet [39]. These model the inflow into dispersive systems more accurately and are widely used in chemical engineering. At the outlet, a zero gradient for the concentration is assumed.

$$\frac{\partial c}{\partial x}(0,t) = \frac{u(t)}{D_{app}}(c(0,t) - c_{in}(t)) \qquad (1.2)$$

$$\frac{\partial c}{\partial x}(L_{Col}, t) = 0 \qquad (1.3)$$

Here, c_{in} is the prescribed concentration at the inlet.

1.1.1.2   Transport Dispersive Model

Instead of a direct coupling with the stationary phase concentration q in Eq. (1.1), a concentration of the pore volume of the mobile phase c_p can be introduce to derive a *Transport Dispersive Model*:

$$
\frac{\partial c}{\partial t}(x,t) = \underbrace{-\frac{u(t)}{\varepsilon_{Col}}\frac{\partial c}{\partial x}(x,t)}_{\text{Convection}} + \underbrace{D_{ax}\frac{\partial^2 c}{\partial x^2}(x,t)}_{\text{Diffusion}}
$$
$$
\underbrace{-\frac{1-\varepsilon_{Col}}{\varepsilon_{Col}}\frac{3}{r_p}k_{eff}\left(c\left(x,t\right)-c_p\left(x,t\right)\right).}_{\text{Mass transfer}} \tag{1.4}
$$

This equation now considers the interstitial volume of the mobile phase and mass transfer into the pore volume only. The column porosity ε_{Col} substitutes total porosity. The previously used D_{app} is split into an axial dispersion coefficient D_{ax} and a term that includes the effective mass transfer coefficient k_{eff} and the bead radius r_p. k_{eff} accounts for film transfer and pore diffusion, $3/r_p$ results from the surface-to-volume ratio of spheres and is usually factored out of k_{eff} for packings of spherical beads.

1.1.1.3   Lumped Rate Model

The *Lumped Rate Model* for the concentration in the pore volume assumes the transferred mass from the volume to stay in the mobile phase, i.e. direct adsorption is not considered. Hence, the transferred amount is corrected by the particle porosity ε_{Bead}. As in Eq. (1.1), the concentration exchange with the stationary phase must consider the differences in volume, also expressed by ε_{Bead}.

$$
\frac{\partial c_p}{\partial t}(x,t) = \frac{3}{r_p}\frac{k_{eff}}{\varepsilon_{Bead}}\left(c\left(x,t\right)-c_p\left(x,t\right)\right) - \frac{1-\varepsilon_{Bead}}{\varepsilon_{Bead}}\frac{\partial q}{\partial t}(x,t) \tag{1.5}
$$

1.1.1.4   General Rate Model

A *General Rate Model* substitutes k_{eff} in Eq. (1.4) by a film transfer coefficient k_{film} and introduces a pore diffusion coefficient D_{pore}. To model the diffusion in the pores, the spherical particle is radially resolved, introducing a new dimension $r \in [0, r_p]$. Here, only the outside of the particle, $r = r_p$, is in contact with the interstitial volume. Again, direct adsorption onto the exterior of the particle is not considered. The concentration change at the boundary is described by

$$
\frac{\partial c_p}{\partial r}(x,r_p,t) = \frac{1}{D_{pore}}\frac{k_{film}}{\varepsilon_{Bead}}\left(c(x,t)-c_p\left(x,r_p,t\right)\right). \tag{1.6}
$$

The factor $3/r_p$ vanishes when introducing the new radial dimension and considering areal instead of volumetric concentration exchange. $1/D_{pore}$ is included to suppress pore diffusion limitation at the boundary and cancels with D_{pore} in Eq. (1.8).

For reasons of symmetry, the concentration gradient is expected to vanish in the center of the bead:

$$\frac{\partial c_p}{\partial r}(x, 0, t) = 0. \tag{1.7}$$

For the interior, the coupling with the stationary phase is described as before. In addition, a Fickian pore diffusion operator is employed that consists of two terms for spherical geometries.

$$\frac{\partial c_p}{\partial t}(x, r, t) = D_{pore}\left(\frac{\partial^2 c}{\partial r^2} + \frac{2}{r}\frac{\partial c}{\partial r}\right) - \frac{1 - \varepsilon_{Bead}}{\varepsilon_{Bead}}\frac{\partial q}{\partial t}(x, r, t) \tag{1.8}$$

The first term, $D_{pore}\frac{\partial^2 c}{\partial r^2}$, describes symmetrical diffusion with parameter D_{pore}. As the particles are spherical, diffusion actually is not symmetric, but radially dependent. The second term, $\frac{2D_{pore}}{r}\frac{\partial c}{\partial r}$, is a convection operator that transports the concentration towards the center of the bead to achieve the correct asymmetry. Details on the derivation can be found in the text books mentioned above.

1.1.1.5 Steric Mass Action Isotherm

An isotherm equation describes the concentration q_i of adsorbed species i as a function of the species' concentration in the mobile phase $c_{p,i}$. For ion-exchange chromatography of proteins, the steric mass action (SMA) isotherm introduced by Brooks and Cramer [24] is employed as a differential equation.

$$k_{kin,i}\frac{\partial q_i}{\partial t}(x, t) = k_{eq,i}\underbrace{\left(\Lambda - \sum_{j=1}^{n}(\nu_j + \sigma_j)q_j(x, t)\right)^{\nu_i}}_{\bar{q}_{salt}^{\nu_i}}c_{p,i}(x, t)$$
$$-c_{p,salt}^{\nu_i}(x, t)q_i(x, t) \tag{1.9}$$

The equation describes the rate of change of the adsorbed protein concentration with an adsorption-promoting term $\bar{q}_{salt}^{\nu}c_p$, and a desorption-promoting term $c_{p,salt}^{\nu}q$. It can reproduce the influence of the counter-ion concentration c_{salt} on the retention behavior of protein species using the proteins' *characteristic charges* ν_i. This value can be interpreted as a number of charged patches on the protein's surface involved in binding. The underlying assumption is a stoichiometric exchange, where an adsorbing protein P_i frees ν_i salt ions S. The index A in Eq. (1.10) depicts the adsorbed state.

$$P_i + \nu_i S_A \rightleftharpoons P_{A,i} + \nu_i S \tag{1.10}$$

Additionally, SMA considers material properties like the total *ionic capacity* Λ, which is the concentration of binding sites on the stationary phase, and *steric shielding* effects of the proteins covering an amount of binding sites σ greater than the actual number of sites it interacts with. Hence, the term in parentheses in Eq. (1.9) is equal to the concentration of available binding sites $\bar{q}_{salt}$ and provides for competition between all N proteins. k_{eq} is the protein's adsorption *equilibrium coefficient* and k_{kin} a *kinetic coefficient* for the rate of concentration exchange.

For comparison, the protein concentrations would be low in analytical high-performance liquid chromatography (HPLC), such that the sum is approximately zero. The concentration exchange also is assumed to be instantaneous, i.e. $k_{kin} = 0$. If in addition, an isocratic experiment with constant counter-ion concentration is performed, then Eq. (1.9) reduces to the linear isotherm equation

$$q_i(x,t) = k_{eq,i}\left(\frac{\Lambda}{c_{p,salt}}\right)^{\nu_i} c_{p,i}(x,t) = K_i c_{p,i}(x,t). \tag{1.11}$$

where q is directly proportional to c_p with the adsorption constant K.

As SMA accounts for changes in the counter-ion concentration, it is complemented by the additional balance equation:

$$q_{salt}(x,t) = \Lambda - \sum_{j=1}^{n} \nu_j q_j(x,t). \tag{1.12}$$

Note that a radial dimension r is added in case of a General Rate Model. As the total capacity is measured in mol/L, the unit of concentration cannot be chosen freely in multi-component settings. Although the SMA isotherm was developed for ion-exchange chromatography [24], it was also applied successfully to multimodal chromatography with mobile phase modifiers [165].

1.1.1.6   Multi-component Langmuir Isotherm

Compared to the SMA model, the Langmuir isotherm does not consider multi-point binding, i.e. $\nu = 1$.

$$k_{kin}\frac{\partial q}{\partial t}(x,t) = k_{eq}\left(\Lambda - \sum_{j=1}^{n}(1+\sigma_j)q_j(x,t)\right)c_p(x,t)$$
$$-c_{p,salt}(x,t)q(x,t). \tag{1.13}$$

In addition, there is no variable counter-ion concentration, such that one can divide by the constant $c_{p,salt}$ and include it into an equilibrium parameter k'_{eq} and kinetic parameter k'_{kin}.

$$k'_{kin}\frac{\partial q}{\partial t}(x,t) = k'_{eq}\left(\Lambda - \sum_{j=1}^{n}(1+\sigma_j)q_j(x,t)\right)c_p(x,t) - q(x,t). \tag{1.14}$$

Instead of a total capacity, Langmuir uses single-component maximum concentrations $q_{max} = \frac{\Lambda}{1+\sigma}$ equal to the component's concentration for a fully saturated adsorber. Placing Λ outside of the parentheses and including it into k''_{eq} yields

$$k'_{kin}\frac{\partial q}{\partial t}(x,t) = k''_{eq}\left(1 - \sum_{j=1}^{n}\frac{q_j(x,t)}{q_{max,j}}\right)c_p(x,t) - q(x,t). \tag{1.15}$$

(a) Adsorption kinetic.

(b) Adsorption isotherm.
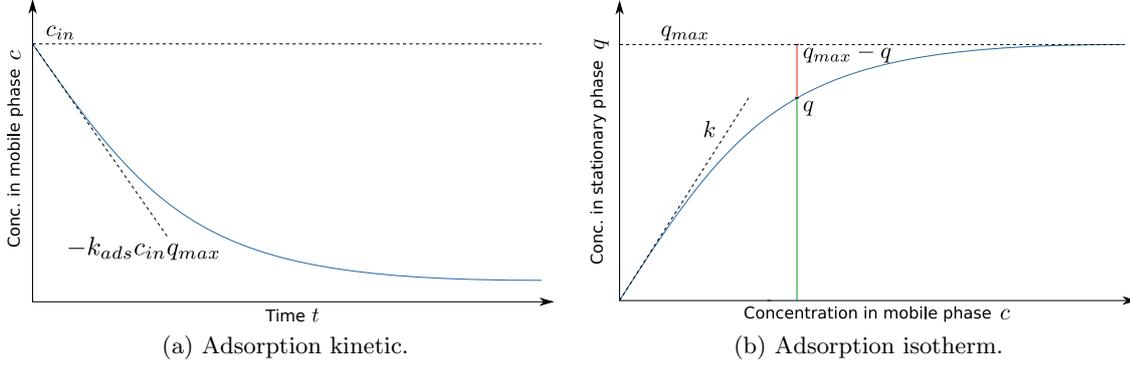
Figure 1.7: Characteristics of the Langmuir adsorption model.

For historical reasons, the adsorption equilibrium parameter for the Langmuir model is divided by q_{max}:

$$k'_{kin} \frac{\partial q}{\partial t}(x,t) = k_{eq,L} q_{max} \left(1 - \sum_{j=1}^{n} \frac{q_j(x,t)}{q_{max,j}}\right) c_p(x,t) - q(x,t), \qquad (1.16)$$

This is the final form of the multi-component kinetic Langmuir isotherm [129]. A plot of the single-component binding kinetics is shown in Figure 1.7a with $k_{ads} = k_{eq,L}/k'_{kin}$. The slope is given by

$$\left. \frac{\partial q}{\partial t}(x,t) \right|_{q \to 0, c \to c_{in}} = \frac{k_{eq,L} q_{max} c_{in}}{k'_{kin}} =: k_{ads} q_{max} c_{in}. \qquad (1.17)$$

.

For instantaneous mass transfer, $k_{kin} = 0$, and a single component, the well-known *equilibrium Langmuir isotherm* is obtained:

$$0 = k_{eq,L} q_{max} \left(1 - \frac{q(x,t)}{q_{max}}\right) c_p(x,t) - q(x,t) \qquad (1.18)$$

$$\Longleftrightarrow 0 = k_{eq,L} q_{max} c_p - k_{eq,L} q(x,t) c_p(x,t) - q(x,t) \qquad (1.19)$$

$$\Longleftrightarrow k_{eq,L} q_{max} c_p(x,t) = k_{eq,L} q(x,t) c_p(x,t) + q(x,t) \qquad (1.20)$$

$$\Longleftrightarrow k_{eq,L} q_{max} c_p(x,t) = [k_{eq,L} c_p(x,t) + 1] q(x,t) \qquad (1.21)$$

$$\Longleftrightarrow q(x,t) = q_{max} \frac{k_{eq,L} c_p(x,t)}{1 + k_{eq,L} c_p(x,t)}. \qquad (1.22)$$

As shown in Figure 1.7b, the ratio of occupied and free binding sites is proportional to the free protein concentration with factor $\frac{k}{q_{max}} = k_{eq,L}$.

Despite their strong assumptions, multi-component Langmuir isotherms are continued to be used to model protein binding (in contrast to stoichiometric ion exchange). This includes hydrophobic interaction chromatography [35] and affinity chromatography [179].

1.1.2 Characteristics of the Steric Mass Acion Model

The simulation of ion-exchange chromatography with the SMA model and variations of it will be a encountered throughput the following chapters. Some characteristics will be
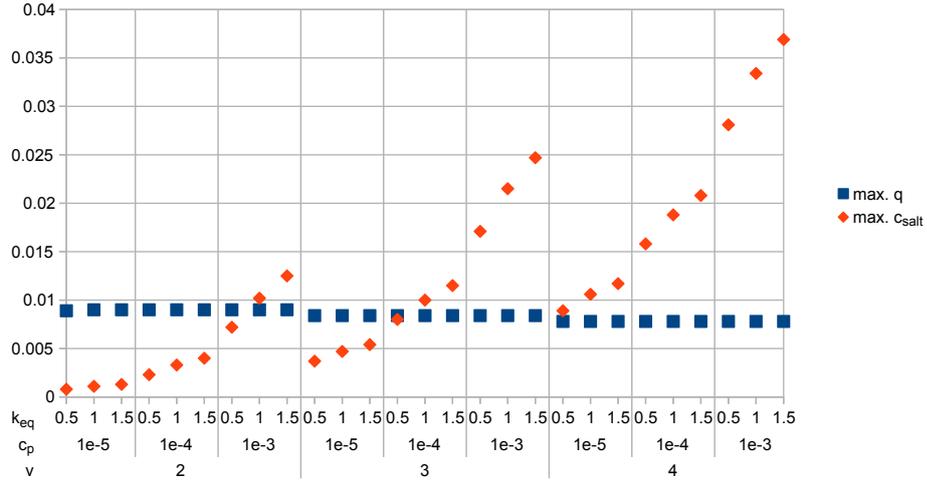
Figure 1.8: Single component maximum concentration q for fixed $\Lambda = 0.3$, $\sigma = 30$.
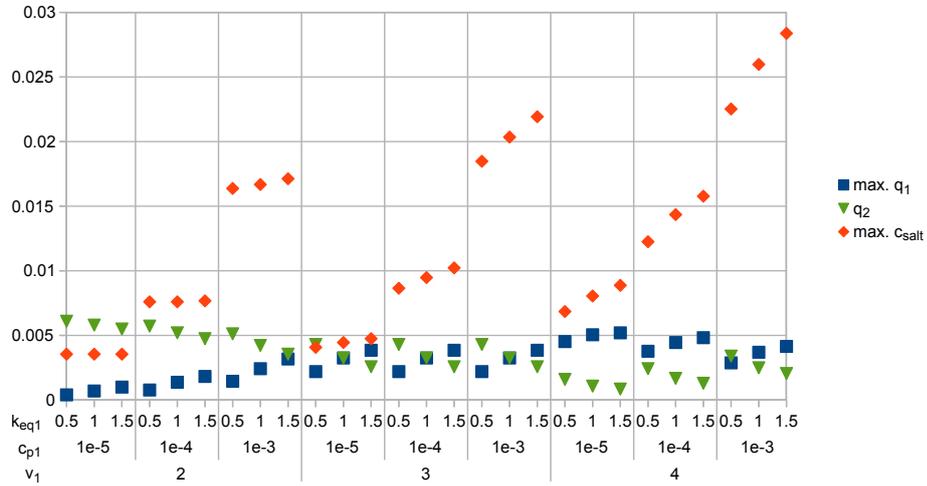
Figure 1.9: Two component loading, maximum concentration q_1 for fixed $\Lambda = 0.3$, $\sigma_1 = 30$.

discussed in the following.

The typical sequence of steps in bind-elute mode consists of loading the feed onto the equilibrated column, washing to remove weakly and non-binding contaminants, eluting the target component and re-equilibrating the column. For loading, the important parameters are feed concentration, flow speed and mobile phase salt concentration.

Suppose, flow rate and feed concentration are chosen in such a way that at least the beginning of the column is saturated at the end of the loading phase. In this regime, we can study the influence of running buffer salt concentration using the equilibrium equation $\left(\frac{\partial q}{\partial t} = 0 \right)$:

$$q_i = k_{eq,i} \left(\frac{\Lambda - \sum_{j=1}^{k}(\nu_j + \sigma_j)q_j}{c_{salt}} \right)^{\nu_i} c_{p,i}. \tag{1.23}$$

Obviously, in the single component case, $c_{salt} \to 0$ yields the highest stationary phase concentration. For given parameters, Equation (1.23) can be solved as fixed point problem $q = f(q)$, where the solution has a natural upper limit of $q_{max} = \frac{\Lambda}{\nu + \sigma}$. Figure 1.8 shows the
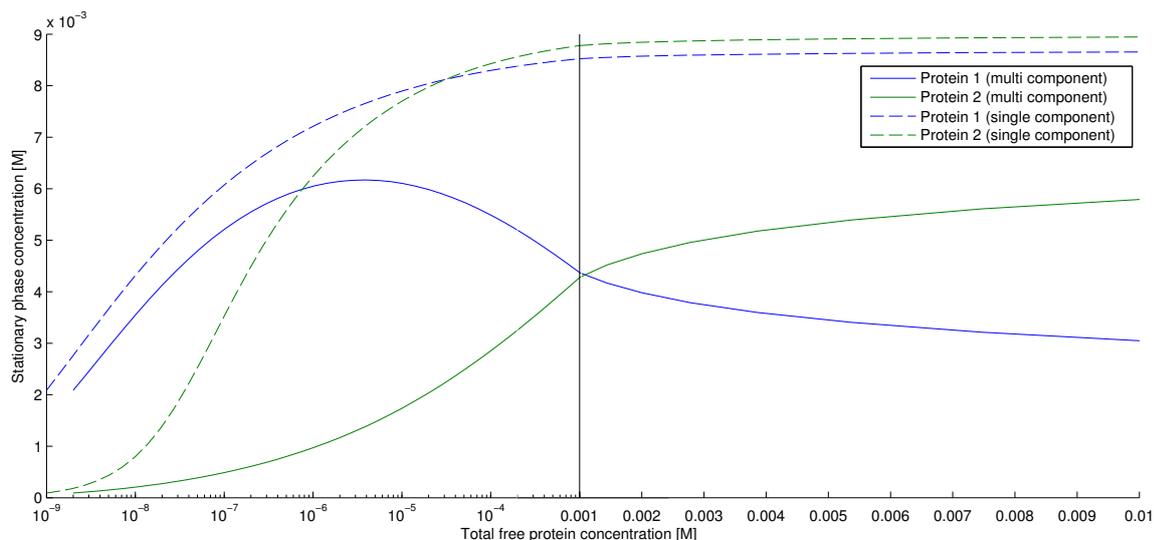
Figure 1.10: SMA isotherms for $k_{eq,1} = 0.5, \nu_1 = 4, k_{eq,2} = 3, \nu_2 = 3, \sigma_{1/2} = 30, \Lambda = 0.3\,M, c_{salt} = 5\,mM$ and equal free protein concentrations on logarithmic scale until 1 mM and linear scale therafter.

maximum attainable concentration q for fixed $\Lambda = 0.3$ and $\sigma = 30$ and the minimum c_{salt} that achieved this value. Lowering c_{salt} further yields no fixed point, meaning no improvement is possible without increasing the feed concentration. Obviously, the equilibrium constant has no effect on the maximum concentration, as long as the salt concentration is chosen sufficiently small. The higher the charge, the higher is the tolerable salt buffer concentration.

Suppose, there is an impurity with fixed parameters $\{k_{eq,2} = 1, \nu_2 = 3, \sigma_2 = 30\}$ and equal molar concentration in the feed. For variable component 1 parameters ($\Lambda = 0.3$, $\sigma_1 = 30$ fixed as before), the highest attainable concentration q_1 is shown in Figure 1.9. Naturally, a higher characteristic charge will provide for a higher concentration in the stationary phase. At comparable charges, a higher equilibrium constant has the same effect. The lower the characteristic charge, the lower must be the salt buffer concentration in order to allow binding. A higher feed concentration is only beneficial if the target component has a lower characteristic charge than the contaminant, as it decreases the influence of the different affinities. It is interesting to note that although having the higher characteristic charge, a target component with $\{k_{eq,1} = 0.5, \nu_1 = 4\}$ has a lower saturation concentration than the impurity in case of a high feed concentration. Here, the multi-component isotherm shows a non-intuitive shape as plotted in Figure 1.10.

Examining Eq. (1.9) shows that during washing no steady state can be assumed. The target component should stay bound to the adsorber, i.e. $q > 0$, what implies $c_p > 0$. With the feed concentration reduced to zero, c_p can only be greater than zero if the bound amount is decreasing, i.e. $\frac{\partial q}{\partial t} < 0$. Solving the partial differential equations in time is inevitable. This topic is covered in the following section.

The drop in feed concentration also changes the equilibrium state and may cause displacement of the weaker binding component by the stronger one as visible in Fig. 1.9 for $\nu_1 = 2$. For all combinations, it is visible that the salt concentration needs to drop when washing in order to achieve/keep the highest saturation. For removing weakly binding impurities,

a slightly higher level is beneficial. As shown in Fig. 1.9, the margin depends on the species' parameters as well as on the feed concentration.

1.1.3 Mathematical Solution of Column Chromatography Models

1.1.3.1 Analytical Solution

The textbook by Guiochon et al. [72] reproduces many results obtained for different transport and isotherm equations. In this section, only single-component solutions for an *ideal model* with equilibrium Langmuir isotherm will be stated, as used in the examples in Sec. B.1.3. The ideal model assumes no dispersion and is given by Eq. (1.1) with $D_{app} = 0$:

$$\frac{\partial c}{\partial t}(x,t) = -\frac{u(t)}{\varepsilon_{Tot}}\frac{\partial c}{\partial x}(x,t) - \frac{1-\varepsilon_{Tot}}{\varepsilon_{Tot}}\frac{\partial q}{\partial t}(x,t). \tag{1.24}$$

For the equilibrium Langmuir isotherm, we have

$$\frac{\partial q}{\partial t} = \frac{\partial}{\partial t}\left(q_{max}\frac{k_{eq,L}c}{1+k_{eq,L}c}\right) \tag{1.25}$$

$$= q_{max}\frac{k_{eq,L}}{(1+k_{eq,L}c)^2}\frac{\partial c}{\partial t}. \tag{1.26}$$

Insertion of Eq. (1.26) into Eq. (1.24) yields

$$\frac{\partial c}{\partial t} = -\frac{u(t)}{\varepsilon_{Tot}}\frac{\partial c}{\partial x} - \frac{1-\varepsilon_{Tot}}{\varepsilon_{Tot}}\frac{q_{max}k_{eq,L}}{(1+k_{eq,L}c_p)^2}\frac{\partial c}{\partial t} \tag{1.27}$$

$$\Longleftrightarrow \left(1+\frac{1-\varepsilon_{Tot}}{\varepsilon_{Tot}}\frac{q_{max}k_{eq,L}}{(1+k_{eq,L}c)^2}\right)\frac{\partial c}{\partial t} = -\frac{u(t)}{\varepsilon_{Tot}}\frac{\partial c}{\partial x} \tag{1.28}$$

$$\Longleftrightarrow \frac{\partial c}{\partial t} = -\frac{u(t)}{\varepsilon_{Tot}+(1-\varepsilon_{Tot})\frac{q_{max}k_{eq,L}}{(1+k_{eq,L}c)^2}}\frac{\partial c}{\partial x} \tag{1.29}$$

Hence, the migration speed in the column depends on the mobile phase concentration and is given by

$$u_c(t,c) = \frac{u(t)}{\varepsilon_{Tot}+(1-\varepsilon_{Tot})\frac{q_{max}k_{eq,L}}{(1+k_{eq,L}c)^2}}. \tag{1.30}$$

For low sample concentrations, with $1 \gg k_{eq,L}c$, a pulse moves approximately uniformly through the column with speed $u_0 = u_c(t,0)$. The retention time for a constant flow rate is then given by

$$t_R = \frac{L_{Col}}{u_0} = \frac{L_{Col}\left(\varepsilon_{Tot}+(1-\varepsilon_{Tot})q_{max}k_{eq,L}\right)}{u}. \tag{1.31}$$

For a pulse of significant width, the peak originates from the center of the pulse. Hence, half the injection time t_{inj} is added, which can be calculated from sample volume and volumetric flow:

$$t_{R,P} = \frac{L_{Col}}{u_0} = \frac{L_{Col}\left(\varepsilon_{Tot}+(1-\varepsilon_{Tot})q_{max}k_{eq,L}\right)}{u} + \frac{t_{inj}}{2}. \tag{1.32}$$

For higher concentrations, u_c increases non-linearly with the mobile-phase concentration, leading to *shock formation*. Higher concentrations move faster through the column than lower ones and build up at the front of the injected pulse. Physically, it is impossible for them to move faster than the pulse front. Adsorption cannot prevent this phenomenon, as the uptake rate decreases with increasing mobile phase concentration in case of the Langmuir isotherm. An asymmetrical peak with a steep front and a diffuse rear part is observed. Let c_S be a high mobile phase concentration, which does not change significantly through adsorption. Then, the Langmuir isotherm can be linearized to

$$\frac{\partial q}{\partial t} \quad = \quad \frac{\partial}{\partial t}\left(q_{max}\frac{k_{eq,L}c}{1+k_{eq,L}c_S}\right) \tag{1.33}$$

$$= \quad q_{max}\frac{k_{eq,L}}{1+k_{eq,L}c_S}\frac{\partial c}{\partial t} \tag{1.34}$$

and we obtain the retention time of the shock front:

$$t_{R,S} \quad = \quad \frac{L_{Col}\left(\varepsilon_{Tot} + (1-\varepsilon_{Tot})\frac{q_{max}k_{eq,L}}{1+k_{eq,L}c_S}\right)}{u}. \tag{1.35}$$

1.1.3.2   Numerical Solution

Typically, physical problems that are modeled by means of partial differential equations are solved by discretization of the equations in the domain of interest, e.g. using finite difference, finite element or finite volume methods on spatial grids. This results in linear or nonlinear systems of algebraic or ordinary differential equations that are typically solved by time-stepping schemes, non-linear and linear iterative solvers. To solve a problem using finite elements, a variational formulation of the problem must be given.

Weak Formulation   The solution of the equations above has to be very smooth, here twice continuously differentiable in space because of the diffusion operator, i.e.

$$c_i \in \mathcal{C}^2\left([0, L_{Col}]\right) \cap \mathcal{C}\left([0, T]\right). \tag{1.36}$$

As we observed before, concentration shock fronts may develop inside the column for low dispersion, which are not differentiable any more. We aim to relax this condition and also allow solutions that fulfil the equations only under the integral.

The residual $\mathcal{R}$ between strong solution c_i and weak solution $\tilde{c}_i$ in a less restrictive space $\mathcal{V}$ is supposed to be zero for a sufficiently large number of input data and is defined by inserting the weak solution into the original (strong) form, e.g.

$$\mathcal{R}\left(\tilde{c}_i\right) = \quad \frac{\partial \tilde{c}_i}{\partial t}\left(x, t\right) + u\frac{\partial \tilde{c}_i}{\partial x}\left(x, t\right) - D_{app}\frac{\partial^2 \tilde{c}_i}{\partial x^2}\left(x, t\right). \tag{1.37}$$

By requiring the residual to be orthogonal to the weak solution space $\mathcal{V}$, i.e. the integral of the residual multiplied by any function of that space is zero, we get an integral condition:

$$\int_{\Omega} w\mathcal{R}\left(\tilde{c}_i\right)\, \mathrm{d}x = 0, \quad \forall w \in \mathcal{V}. \tag{1.38}$$

Here, we followed the so-called Galerkin ansatz, where the test function w is chosen from the solution space $\mathcal{V}$. Other test spaces can be beneficial: the so-called Streamline-Upwind-Petrov-Galerkin method suppresses unnatural oscillations by choosing $w \in \mathcal{V} + \tau \nabla \mathcal{V}$, with a stabilization parameter τ depending on the Peclet number.

Inserting the residual equation above yields

$$0 = \int_\Omega w \frac{\partial c_i}{\partial t} + uw \frac{\partial c_i}{\partial x} - D_{app} w \frac{\partial^2 c_i}{\partial x^2} \, \mathrm{d}x. \tag{1.39}$$

Applying Green's formula allows to reduce the second order derivatives on the solution to first order by shifting it onto the test function

$$0 = \int_\Omega w \frac{\partial c_i}{\partial t} + uw \frac{\partial c_i}{\partial x} + D_{app} \frac{\partial w}{\partial x} \frac{\partial c_i}{\partial x} \, \mathrm{d}x + \int_\Gamma D_{app} w \frac{\partial c_i}{\partial x} \cdot n \, \mathrm{d}s. \tag{1.40}$$

Obviously the Danckwerts boundary conditions fit very well to the weak formulation of the mass transfer equation, inserting yields

$$0 = \int_\Omega w \frac{\partial c_i}{\partial t} + uw \frac{\partial c_i}{\partial x} + D_{app} \frac{\partial w}{\partial x} \frac{\partial c_i}{\partial x} \, \mathrm{d}x + \int_{\Gamma_{in}} uw \left(c_i \left(0, t \right) - c_{in,i} \left(t \right) \right) \cdot n \, \mathrm{d}s. \tag{1.41}$$

Discretization in Space   Because of the non-linearity introduced by the isotherm equation, there is no analytical solution to the whole multi-component setting. To solve the equations numerically, we approximate the solution on a computational grid.

Let $c_{i_h} \in \mathcal{V}_h$ be the approximate solution on this mesh and $\{\varphi_k\}$ a basis of the discrete space $\mathcal{V}_h$, we define

$$c_{i_h} \left(x, t \right) = \sum_k c_{i_k} \left(t \right) \varphi_k \left(x \right). \tag{1.42}$$

The basis functions can be defined in any dimension and are supposed to be always zero on all but one mesh element, furthermore the function takes the value 1 only at one point of this element.

In our case, the solution will not be showing strong oscillations on an element, for a reasonably fine mesh. It is hence sufficient to only look at linear or quadratic elements. Linear elements have two basis functions, each taking the value 1 at one of the nodes and 0 at the other as shown in Fig. 1.11a. Similarly, quadratic Finite Elements have a third basis function which takes the value one at the center point of the element, cf. Fig. 1.11b.

The discrete counterpart of the space of weighting/test functions $\mathcal{W}_h$ is likewise spanned by a basis $\{\psi_k\}$. In the usual Galerkin scheme this space is chosen equal to $\mathcal{V}_h$ or as a subspace of it, in our case it will be the finite dimensional subspace (because of discretization) of $H^1 \left(\Omega \right)$, the space of square-integrable functions with integrable first derivatives. The solution may have non-zero boundary values and is thus chosen to be in $H^1 \left(\Omega \right)$ as well.

(a) Linear basis functions.

(b) Quadratic basis functions.
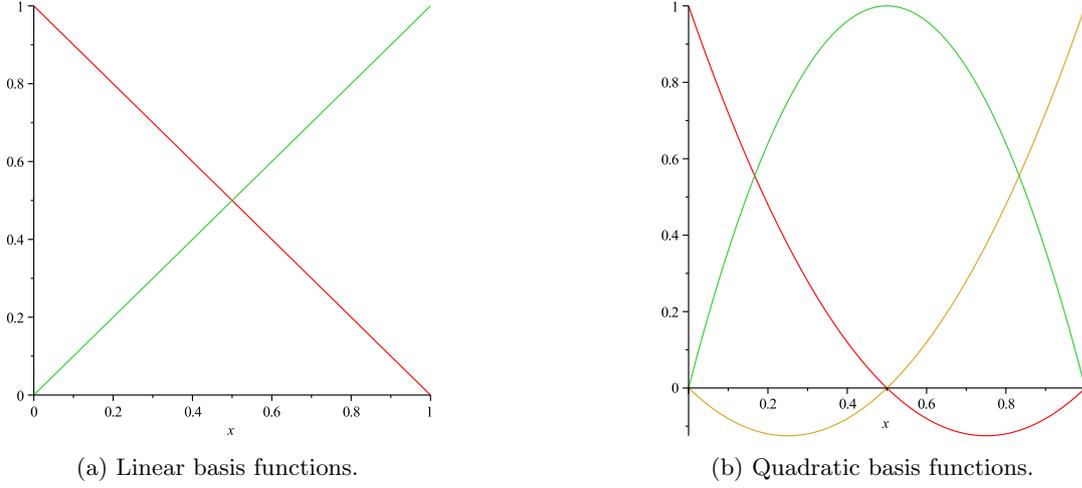
Figure 1.11: Possible basis functions of solution and test spaces.

Inserting the basis representation into the weak formulation yields

$$
\begin{aligned}
0 \ = \ & \int_\Omega \sum_k \psi_k \left(\frac{\partial}{\partial t} \sum_l c_{i_l} \varphi_l \right) + u \sum_k \psi_k \left(\frac{\partial}{\partial x} \sum_l c_{i_l} \varphi_l \right) \\
& + D_{app} \left(\frac{\partial}{\partial x} \sum_k \psi_k \right) \left(\frac{\partial}{\partial x} \sum_l c_{i_l} \varphi_l \right) \\
& - \int_{\Gamma_{in}} u c_{in} \sum_k \psi_k \cdot n \, \mathrm{d}s + \int_{\Gamma_{in}} u c_i \sum_k \psi_k \cdot n \, \mathrm{d}s \quad \forall \psi \in H^1(\Omega). \qquad (1.43)
\end{aligned}
$$

As all basis functions are only non-zero on a single element of the domain, we can re-write this as sum over local integrals.

$$
\begin{aligned}
0 \ = \ & \frac{\partial}{\partial t} \sum_{k,l} c_{i_k} \int_\Omega \psi_k \varphi_l \, \mathrm{d}x + u \left(\sum_{k,l} c_{i_l} \int_\Omega \psi_k \frac{\partial \varphi_l}{\partial x} \, \mathrm{d}x \right) \\
& + D_{app} \left(\sum_{k,l} c_{i_{kl}} \int_\Omega \frac{\partial \psi_k}{\partial x} \frac{\partial \varphi_l}{\partial x} \, \mathrm{d}x \right) \\
& - u c_{in} \sum_l \int_{\Gamma_{in}} \psi_k \cdot n \, \mathrm{d}s + u c_i \sum_l \int_{\Gamma_{out}} \psi_k \cdot n \, \mathrm{d}s. \qquad (1.44)
\end{aligned}
$$

The three types of integrals are usually referred to as

- Mass Matrix $M_{kl} = \int_\Omega \psi_k \varphi_l \, \mathrm{d}x$,

- Stiffness Matrix $A_{kl} = \int_\Omega \frac{\partial \psi_k}{\partial x} \frac{\partial \varphi_l}{\partial x} \, \mathrm{d}x$,

- Convection Matrix $C_{kl} = \int_\Omega \psi_k \frac{\partial \varphi_l}{\partial x} \, \mathrm{d}x$.

Using this notation, the formulation reduces to

$$
\begin{aligned}
0 \;=\;& M\frac{\partial c_i}{\partial t} + (u \cdot C + D_{app} \cdot A)\,c_i \\
& - u c_{in} \sum_l \int_{\Gamma_{in}} \psi_k \cdot n \,\mathrm{d}s + u c_i \sum_l \int_{\Gamma_{in}} \psi_k \cdot n \,\mathrm{d}s.
\end{aligned}
\tag{1.45}
$$

Depending on the choice of basis function the boundary integral will only consist of contributions of a few basis functions.

For the mass, stiffness and convection matrix the integrals will only dependent on the shape of the elements: Let h be the distance between the two mesh points of the element, then the basis functions of linear elements can be written as

$$
\varphi_1 \;= \psi_1 = \; \frac{h - x}{h},
\tag{1.46}
$$

$$
\varphi_2 \;= \psi_2 = \; \frac{x}{h},
\tag{1.47}
$$

and the integrals are

$$
\int_0^h \varphi_1 \psi_1 = \frac{h}{3}, \qquad \int_0^h \varphi_2 \psi_1 = \frac{h}{6},
\tag{1.48}
$$

$$
\int_0^h \varphi_1 \psi_2 = \frac{h}{6}, \qquad \int_0^h \varphi_2 \psi_2 = \frac{h}{3}.
\tag{1.49}
$$

For equidistant nodes, the mass matrix will have a tridiagonal structure given by

$$
\frac{h}{6}
\begin{pmatrix}
2 & 1 & & \\
1 & 4 & 1 & \\
  & 1 & 4 & \ddots \\
  &   & \ddots & \ddots
\end{pmatrix}.
\tag{1.50}
$$

The assembly of the stiffness and convection matrix is straight forward and thus omitted here.

When dealing with radial geometries, e.g. when simulating intra-bead concentrations with the General Rate Model or the flow through membrane capsules, we encounter the convection operator $\frac{1}{r}\frac{\partial c}{\partial r}$ that models the increasing flow rate because of the narrowing diameter. We either have a boundary condition for $r = 0$ or we are dealing with a cylinder ring, such that $r \neq 0$.

Formulating the discrete operator as before, we get

$$
C_{rad,k,l} \;=\; \int_\Omega \frac{1}{r}\psi_k \frac{\partial \varphi_l}{\partial r}\,\mathrm{d}r \quad \text{on } \Omega = [r_{inner}, r_{outer}].
\tag{1.51}
$$

This time, the boundaries of the integral do not vanish, when inserting the basis functions and integrating, thus we cannot write the tridiagonal structure directly. Also, because of r in the denominator, we get logarithmic terms. For equidistant nodes with distance h, the contributions of the two linear basis function on an element $[a, b]$ are

$$
\frac{1}{h}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} + \frac{\log b - \log a}{h^2}\begin{bmatrix} -b & b \\ a & -a \end{bmatrix}.
\tag{1.52}
$$

We can immediately start simulating the convection through a cylinder ring by assembling the matrix for a given inner and outer radius.

Discretization in Time   Naturally, the solution at a certain point in time only depends on previous but not future events. Hence, we can discretize the simulation space as well and compute the solution step-by-step. Any method for solving Ordinary Differential Equations (ODE) can be used here, but typically finite difference methods are employed.

Assuming, we want solve the initial value problem

$$\frac{\mathrm{d}c\left(t\right)}{\mathrm{d}t} = F\left(t, c(t)\right), \quad c\left(t^0\right) = c^0, \quad t \in \left(t^n, t^{n+1}\right), \tag{1.53}$$

then the forward difference approximation would read

$$\frac{c\left(t^{n+1}\right) - c\left(t^n\right)}{t^{n+1} - t^n} = F\left(t, c(t^n)\right). \tag{1.54}$$

In the realm of time stepping methods this is method is called forward Euler. The backward Euler scheme reads respectively

$$\frac{c\left(t^{n+1}\right) - c\left(t^n\right)}{t^{n+1} - t^n} = F\left(t, c(t^{n+1})\right). \tag{1.55}$$

Another often employed method is the Crank-Nicolson scheme which consists of half a forward and backward step:

$$\frac{c\left(t^{n+1}\right) - c\left(t^n\right)}{t^{n+1} - t^n} = \frac{1}{2}\left[F\left(t, c(t^{n+1})\right) + F\left(t, c(t^n)\right)\right]. \tag{1.56}$$

These three methods are often combined in the Standard θ-scheme

$$\frac{c\left(t^{n+1}\right) - c\left(t^n\right)}{t^{n+1} - t^n} = \theta F\left(t, c(t^{n+1})\right) + (1 - \theta) F\left(t, c(t^n)\right), \tag{1.57}$$

with the choices of $\theta = 0$, 0.5 and 1 leading to the three above mentioned methods.

As usual with the finite difference method, the discretization introduced an approximation error which is in the order of $\mathcal{O}\left(\Delta t\right)$ for the Euler schemes and $\mathcal{O}\left(\Delta t\right)^2$ for Crank-Nicolson.

Schemes which make use of unknown future values such as backward Euler and Crank-Nicolson are called implicit schemes. These schemes use iterative procedures to approximate these unknown values and thus allow larger time steps while on the other hand the cost per time step is high. Explicit schemes such as forward Euler cannot use large time steps for stability reasons, but do not need to solve a linear system. The choice of an optimal time stepping scheme depends on the problem. When a highly dynamic instationary process has to be resolved in detail, an explicit scheme is a good choice. Implicit schemes offer a high stability and large time steps but may smooth out dynamics.

1.1.4   Existing Simulation Software

In order to find optimal process parameters, modeling tools are increasingly gaining the attention of the pharmaceutical industry [78]. Although many research groups developed
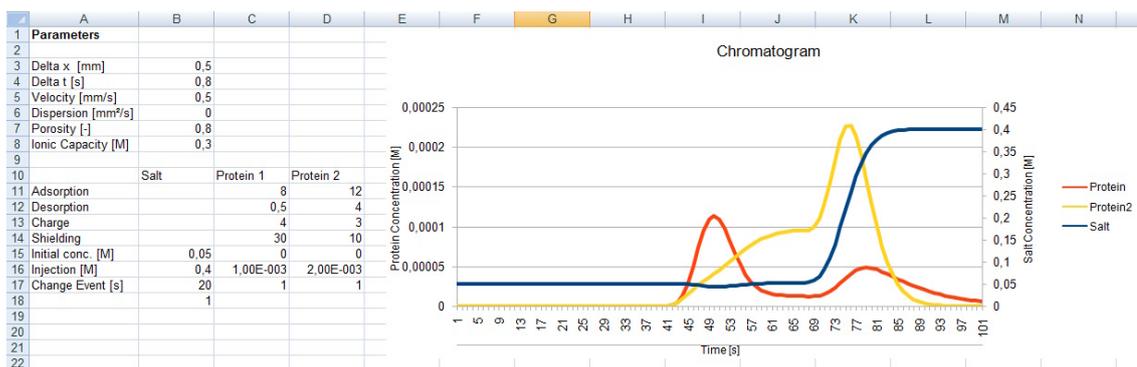
(a) Parameter entry and two-coponent chromatogram.
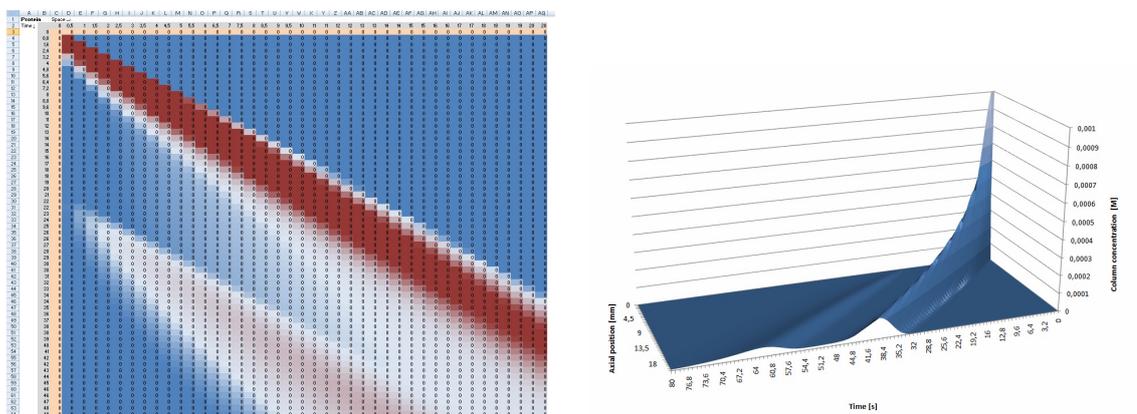
(b) Sheet with concentrations of Protein 1 in time and space.

(c) Surface plot over time and space for Protein 1 with injection in the right back corner and chromatogram along the front.

Figure 1.12: Excel implementation.

their own simulation tools and published results generated with it, only few simulators are publicly available. The six simulators found for nonlinear chromatography are: Aspen Chromatography [8], CADET [52], Chromulator [184], ChromWorks [32], gProms [156], and pcs [118]. Aspen, ChromWorks, and gProms are commercial Windows applications that cannot be downloaded directly. CADET is the most advanced noncommercial simulator in terms of numerical methods, but its C++ code must be compiled manually and does not offer a graphical interface. pcs is based on MATLAB and uses the comparably inflexible finite difference method. Summarizing, none of the available tools combines high simulation speed with extensibility and usability.

1.1.5   Design Considerations for a Novel Simulation Toolbox

Basically, even a spreadsheet can be sufficient for simple simulation cases. To demonstrate this, an Equilibrium Dispersive Model with two-component SMA isotherm was discretized with finite differences and implemented in Excel (Fig. 1.12). The limitations are obvious, the implementation of a different model requires a complete redevelopment.

Thereafter, the finite element approach presented in the previous section was implemented in MATLAB as explained in Appendix A. This simulation environment is highly flexible

but does not provide the fast simulation run times needed for parameter estimation by curve fitting or sampling of a large parameter space.

Short solving times require both powerful hardware and efficient parallel numerical methods compliant with multi-level parallelism of modern computing platforms. Here, the main challenge lies in fully utilizing the available computing power of emerging technologies like multi-core CPUs, graphics processing units (GPUs), multi-GPUs, and any other coprocessor-accelerated or heterogeneous platform without increasing the complexity of the software. Generic interfaces should allow to build solvers without having detailed information on the underlying platform. Likewise, an abstraction layer should allow for mapping partial differential equations into parallel software without dealing with the actual discretization of the equations. To fulfill this, strict modularization and clear interfaces between various modules are necessary. The implementation should utilize modern objected-oriented and template concepts to reduce complexity, ease maintainability of the code and also enable re-usability of specific parts for the extension of functionalities and features.

Summarizing, a new kind of simulation toolbox must be designed and implemented to arrive at a truly flexible multi-purpose software package that can be adapted to most user scenarios and most importantly to tackle the problems formulated in the following research proposal.

1.2 Research Proposal

Chromatography has been the core unit operation of bioseparation processes for decades and is becoming a more and more versatile purification tool. New or combined interaction modes offer high selectivity at mild conditions as well as high robustness and general acceptance by regulatory authorities. To reduce the process development effort, so-called platform processes have been developed in industry, with a fixed order of chromatographic steps and only few degrees of freedom which are optimized for the specific product. Currently, non-standard antibodies, such as antibody-drug-conjugates, or new types of biopharmaceuticals, e.g. virus-like particles, are emerging that do not fit into platform processes anymore. Hence, the full amount of degrees of freedoms has to be considered to find effective and efficient chromatography set-ups.

High-throughput methods can be used to screen a large portion of the design space, but require a comparably high amount of sample that is often not available for early stage process development. Model-based process development is highly attractive as it makes elaborate screening experiments redundant once the model has been calibrated to the specific process step. Model-based approaches are developed and used in academia, but technology transfer to industry is still limited. The main reason is the lack of integration of novel tools into laboratory work-flows or in other words the transition from established experimental procedures to applying model-based tools. To integrate simulation tools successfully, the underlying models have to cover the usual operations, simplify parameter estimation and optimization from routine data collection while maintaining high flexibility. In the light of the Quality-by-Design initiative, it is not enough to find any kind of model, but one with a meaningful physical interpretation.

The objective and ultimate goal of this research proposal is to create a powerful bioprocess development toolbox with accompanying work flows. It should delivery accurate simulation results, allowing to replace costly experimentation to minimize development time, enabling comprehensive parameter studies for better process understanding, and allowing for optimization of production processes and development cycles. In this toolbox, mechanistic models for the major relevant chromatographic unit operations shall be included. To the authors knowledge, there is no academic or commercial software available to accurately simulate other modes than ion-exchange chromatography. Even these tools require information that is often not available in industrial settings, especially molar feed concentrations of all species in the sample material. Preparative chromatography steps are usually controlled using the UV measurements, and analytical chromatography runs often determine peak areas of impurities as a measure of overall purity. Furthermore, several wavelengths can be recorded in parallel, providing additional information and avoiding detector saturation problems. It is hence highly desirable to exchange molar concentrations with UV absorbance as measure of concentration in the existing mechanistic models.

With no a priori knowledge about the components' behavior, the inverse method is a suitable option which alters parameters in a systematic fashion to achieve a match of the measured chromatogram and the model prediction. Based on defined experimental proceeding, the parameter estimation algorithm should allow for determining the optimal sequence and the process parameters of a chromatographic purification process.

While the objective of parameter estimation is to find parameters which minimize the disagreement of simulation and measurement, the technique of Optimal Experimental Design (OED) takes one step backwards and first finds the experimental set-up with highest information content and then minimizes the uncertainty of the parameters, measured as function of the covariance matrix. Using this approach, the parameters with the highest statistical quality are found using the least number of experiments. It shall be evaluated whether this technique can be applied to the current models of chromatography and whether the covariance matrix can be used as measure for the quality of the parameter estimates and whether the model equations are formulated in a numerically reasonable way.

The alternative, monte-carlo sampling, might become too time-consuming for complex simulations. On the other-hand, OED requires the determination of covariance matrices and performs optimization tasks on them. In order to do this, many parameter derivatives have to be calculated. Depending on the number of components and model parameters derivative computation can become equally time consuming in practical applications. As a sub-objective, this problem shall be solved to allow for the efficient use of OED.

All estimated parameters should be taken with a grain of salt, as even thermodynamically derived equations might not include all micro or nano scale phenomena. It is of utmost importance for this research proposal to constantly reflect on parameter values and their physical interpretation in order to gain deeper understanding of the fundamentals of chromatography.

1.3   Outline

The following Chapter 2 studies competitive multi-component ion-exchange adsorption in batch chromatography, validating the theoretical considerations from the introduction. Chaps. 3–5 discuss advanced computational methods for simulating and optimizing preparative column chromatography of proteins. The remaining chapters Chaps. 6–10 focus on model calibration and model-based process-development under various conditions.

2. **Deconvolution of High-throughput Multi-component Isotherms Using Multivariate Data Analysis of Protein Spectra**

P. Baumann*, T. Hahn*, T. Huuk*, A. Osberghaus, J. Hubbuch (* contributed equally)

This case study explores the usability of UV spectra and multivariate statistics in high-throughput batch chromatography, incorporating multiple protein species. The presented approach enables integration of the analytical setup in the batch chromatographic work flow, using a standard UV/VIS spectrophotometer. The quality of the analytical data allowed to fit steric mass action isotherms at various pH values and ionic strengths and get a mechanistic insight into the competitive binding behavior.

3. **Simulating and Optimizing Preparative Protein Chromatography with ChromX**

T. Hahn, T. Huuk, V. Heuveline, J. Hubbuch

In this paper, the working principle of the ChromX software is explained. ChromX is not only intended to support research of expert users but should also be operatable for users without modeling background such as laboratory technicians and students. ChromX was hence extended with a user-friendly interface and successfully used in laboratory courses for undergraduate students. The paper presents the results of the 2014 lab course, where a monoclonal antibody was to be purified from a ternary mixture.

4. **Adjoint-based Estimation and Optimization for Column Liquid Chromatography Models**

T. Hahn, A. Sommer, A. Osberghaus, V. Heuveline, J. Hubbuch

The determination of parameter derivatives for liquid chromatography models is crucial for deterministic process optimization and parameter estimation. Finite difference approximations and automatic differentiations have been employed in the past but are computationally expensive. In the manuscript, the adjoint method is formulated for a wide range of models used in liquid chromatography. We showed that derivation and implementation is manageable and that forward solvers can be easily turned into adjoint solvers, achieving significant speed-ups over finite differences.

5. **Optimal Experimental Design for the Determination of Isotherm Parameters of Glucose Oxidase using Mixed Mode Column Chromatography**

   T. Hahn, G. Wang, T. Huuk, V. Heuveline, J. Hubbuch

This manuscript presents the first study of Optimal Experimental Design (OED) for column chromatography of proteins, including the computation of approximate covariances and confidence intervals. A case study compares parameter estimation by conventional experimental design and OED for the model protein glucose oxidase on ion-exchange, hydrophobic interaction and mixed-mode. Furthermore, the classical General Rate Model was extended with salt-dependent pore diffusion. The results underline the improved reliability using OED at potentially much less time and material consumption.

Manuscript in preparation.

6. **Model-based Integrated Optimization and Evaluation of a Multi-step Ion Exchange Chromatography**

   T. Huuk, T. Hahn, A. Osberghaus, J. Hubbuch

This paper presents a case study on model-based concerted process optimization of two consecutive ion exchange chromatographies. The superiority is demonstrated by comparison to a sequential approach that optimizes the two steps consecutively. Verification is carried out with a set of three model proteins. The *in silico* optimum is reproduced in lab experiments and the modeling tool is successfully employed for error diagnostics.

Manuscript published in Separation and Purification Technology , 136, 207–222, 2014.

7. **UV Absorption-based Inverse Modelling of Protein Chromatography**

   T. Hahn, P. Baumann, T. Huuk, V. Heuveline, J. Hubbuch

UV absorption measurements play an important role in bioprocess development. But so far, absorption measurements could not be used for modeling chromatography steps as the commonly applied models rely on mass or molar concentration. This study presents mechanistic modeling of an industrially relevant chromatography setting without any knowledge of the feed composition. In our manuscript, we deal with a crude feed stock including at least 11 impurities that we model just based on UV absorption data.

Manuscript in press, to appear in Engineering in Life Sciences.

8. **Calibration-free Inverse Modeling of Ion-exchange Chromatography in Industrial Antibody Purification**

   T. Hahn, T. Huuk, A. Osberghaus, K. Doninger, S. Nath S. Hepbildikler, V. Heuveline,
   J. Hubbuch

In this manuscript, we deal with a data set from industrial antibody purification, with unknown contaminant concentrations and no distinguishable peaks. Combining the previously developed method for model calibration, that requires the identificaion of single-component absorption curves in a chromatogram, with high resolution fraction analyses allowed to simulate and optimize the complex separation problem. The optimal process parameters found by conventional process development could be reproduced *in silico*.

Manuscript in press, to appear in Engineering in Life Sciences.

9. **High-throughput Micro-scale Cultivations and Chromatography Modeling: Powerful Tools for Integrated Process Development**

P. Baumann, T. Hahn, J. Hubbuch

This paper presents a combined optimization approach based on high-throughput micro-scale cultivation experiments and chromatography modeling. The multi-criteria downstream process optimization for a set of randomly picked upstream conditions producing high yields was performed *in silico* using ChromX. It could be shown that the impurities in the various conditions are in fact identical. The *in silico*-optimized operational modes for product capturing were validated successfully in the lab.

Manuscript in press, to appear in Biotechnology and Bioengineering.

10. **Modeling and Simulation of Anion-exchange Membrane Chromatography for Purification of *Sf* 9 Insect Cell-derived Virus-like Particles**

C. Ladd Effio, T. Hahn, J. Seiler, S. Oelmeier, I. Asen, C. Silberer, L. Villain, J. Hubbuch

In this work, we present process modeling of an anion-exchange membrane capture step for virus-like particles. A mechanistic approach was implemented for radial-flow ion-exchange membrane chromatography using the lumped-rate model and Steric Mass Action model for the *in silico* optimization of a VLP capture step. For the first time, process modeling enabled the *in silico* design of a selective, robust and scalable process with minimal experimental effort for a complex VLP feedstock.

Manuscript submitted to Journal of Chromatography A.

2 | Deconvolution of High-throughput Multi-component Isotherms Using Multivariate Data Analysis of Protein Spectra

Pascal Baumann[1,*], Tobias Hahn[1,*], Thiemo Huuk[1,*], Anna Osberghaus[1], Jürgen Hubbuch[1]

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

[*] Contributed equally to this work

Abstract

Gaining a more profound understanding of biopharmaceutical downstream processes is a key demand of the Quality by Design (QbD) guideline. One of the most dominant approaches to gain process understanding is the extensive use of experimental high-throughput formats, such as batch chromatography on robotic liquid handling stations. Using these high-throughput experimental formats, the generation of numerous samples poses an enormous problem to subsequent analytical techniques.

Here, a high-throughput case study for batch chromatographic multi-component isotherms is presented. To debottleneck the subsequent analytics, a non-invasive technique using UV spectra and multivariate statistics was adapted to a batch chromatographic format. Using this approach, it was possible to integrate the entire analytical setup into the robotic work flow.

As a case study, batch isotherms for SP Sepharose FF and the model proteins cytochrome c and lysozyme at various pH values and ionic strengths were recorded. A successful examination of the quality of the analytical procedure compared to classical single wavelength photometry was carried out. To address the growing demand for a more profound process understanding, the experimental data were fitted to the steric mass action isotherm, getting a more detailed insight into the competitive binding behavior at various pH values and ionic strengths.

Practical application

This case study explores the usability of UV spectra and multivariate statistics to prevent an analytical bottleneck in high-throughput batch chromatography, incorporating multiple protein species. The presented approach enables integration of the analytical setup in the batch chromatographic work flow, using a standard UV/VIS spectrophotometer. The quality of the analytical data was sufficient to fit steric mass action isotherms at various pH values and ionic strengths and get a mechanistic insight into the competitive binding behavior.

2.1   Introduction

Nowadays, biopharmaceutical downstream process (DSP) development is mainly based on chromatographic separation techniques. Chromatography, as many other separation techniques for biopharmaceuticals, is influenced by numerous operational parameters that affect process performance. The growing need for a deeper mechanistic understanding of the technical process on a molecular level and the impact of the process on the product quality, demanded by the Quality by Design (QbD) approach, pose a great challenge to biopharmaceutical DSP [31; 86]. The requested process understanding can be generated by several approaches. One very obvious approach is the systematic experimental evaluation of the impact of different operational parameters on the downstream process performance.

A possibility to support the exploration of the design space, as basis for many statistical DoE approaches, is the use of high-throughput techniques [101]. These approaches enable miniaturization of the experimental systems and parallelization and automation of the experimental work flow allowing for full factorial experimental designs [15; 145; 176]. Therefore, high-throughput techniques enable a reduction of time and material consumption. In the context of chromatographic process development, there are several experimental systems adapted to robotic high-throughput experiments.

Batch chromatography can be implemented very easily on a liquid handling station using a defined adsorber volume provided in a 96-well plate. The batch chromatographic systems can be used for resin screening [101], batch bind-elute studies or the measurement of adsorption isotherms [35; 101; 142; 143] and kinetic data [36].

One major disadvantage of these experimental high-throughput formats is that the above-mentioned experimental bottleneck is not prevented but often only shifted to subsequent analytics. In the simplest case, when only single protein data have to be recorded, a photometric measurement within the robotic work flow is possible. This simple analytical approach is restricted to the quantification of a single protein species or several protein species with different exclusive absorption maxima beside 280 nm [12]. The major advantage of photometric assays is the high sample throughput, which can easily cope with the amount of samples generated in robotic high-throughput experiments and the non-invasive nature of the technique.

In pharmaceutical high-throughput process development, a simple photometric measurement will be unable to differentiate between several protein species. In this case, more

sophisticated analytics have to be performed. Such advanced analytics might be e.g., analytical high-performance liquid chromatography (HPLC) and product-specific immunoassays [123; 190]. The time requirement of these immunoassays is still high in comparison to that of the chromatographic experiments. HPLC assays are predominant in the characterization of the size and charge heterogeneity of a target protein. One major drawback of processing high-throughput experimental samples with HPLC assays is the low sample throughput of HPLC techniques. Despite the mentioned drawbacks, HPLC and other analytical techniques such as immunoassays bear the major disadvantage of an invasive nature.

An ideal analytical technique in the context of high-throughput experimental work flows would be a photometric assay that can be carried out in standard UV/VIS plate readers typically installed on robotic workstations. This assay has to be protein species-specific and quantitative. Such an assay, based on the measurement of UV spectra and subsequent multivariate statistics, was introduced by Hansen *et al.* in 2011 [79]. When using protein mid-UV (200 – 300 nm) spectra, the absorption of different protein species is strongly influenced by properties of peptide bonds and amino acid residues. A partial least squares - projection to latent structures (PLS) regression enables subsequent quantitative evaluation of mixtures of different protein species.

Here, we present a case study on the high-throughput generation of multi-component isotherm data. We combine a high-throughput method for the collection of multi-component isotherms in an automated batch format with a non-invasive protein-specific quantification method based on the measurement of mid-UV spectra. The collected data are subsequently fitted using the steric mass action isotherm including competitive protein binding introduced by Brooks and Cramer [24].

2.2 Materials and Methods

2.2.1 Materials

2.2.1.1 Disposables & Reaction Vessels

As a strong cation exchange adsorber, sulfopropyl (SP) sepharose fast flow (FF) provided by GE Healthcare Life Sciences, Sweden, was used. Binding experiments were carried out in 2 mL 96-well square deep well plates (VWR, Germany). Absorption and protein spectra measurements were carried out in 96-well flat bottom UV-Star microplates (Greiner Bio-One, Germany). Buffers were filtered using 0.2 µm cellulose acetate filters supplied by Sartorius, Germany.

2.2.1.2 Chemicals & Buffers

Binding experiments at pH 5 were carried out in 20 mM acetate buffer consisting of acetic acid (Merck, Germany) and sodium acetate (Sigma-Aldrich, USA). For experiments at pH 7, a 20 mM phosphate buffer consisting of di-sodium hydrogen phosphate and sodium

di-hydrogen phosphate (Sigma-Aldrich, USA) was used. The ionic strength of the applied buffers was adjusted to 42, 65, 90, 115, and 150 mM using sodium chloride (AppliChem, Germany). The model proteins lysozyme from chicken egg white and cytochrome c from equine heart were purchased from Sigma-Aldrich, USA. Calibrations for the multivariate data analysis model of the two-component system were carried out using a 2.4 mg/mL protein solution. For binding experiments, a 10 mg/mL stock solution with 70 % cytochrome c and 30 % lysozyme was used.

2.2.1.3   Instrumentation & Software

For pH adjustment of all buffers, a HI-3220 pH meter (Hanna Instruments, USA) was used. The instrument was calibrated using high-precision standards from Hanna Instruments (USA). For generation of equal amounts of adsorber per well, a Media Scout Resi Quot System (Atoll, Germany) was applied. For pressure adjustment, a vacuum pump with pressure regulation was used. For the batch isotherms, a Freedom EVO® 200 liquid handling station (Tecan, Germany) was used, operated with EvoWare 2.1. The system is equipped with eight fixed tips, a plate-moving arm and an orbital shaker (Tecan, Germany). A Rotanta 46RSC centrifuge (Hettich, Germany) and an Infinite M200 UV plate spectrophotometer (Tecan, Germany) are integrated in the system. The spectrometer was controlled by i-control 1.9 (Tecan, Germany). Data processing and creation of figures was performed in Matlab R2011a (MathWorks, USA).

2.2.2   Experimental Setup

2.2.2.1   Model Calibration and Validation

The multivariate data analysis calibration [79] was based on a four-level D-optimal onion design generated with MODDE (Umetrics, Sweden) with additional data points added at low concentration levels. The model included 7 mixing ratios (1:0; 2.5:1; 2:1; 1:1; 1:2; 1:2.5; 0:1) and 15 concentration levels (concentration levels correspond to the protein concentration of cytochrome c and lysozyme in total) from 0 to 1.2 g/L (each 5 concentrations in a range of 0 to 0.1, 0.1 to 0.5, and 0.5 to 1.2 g/L). For example, a mixing ratio of 2:1 means that the mixture contains 2/3 of cytochrome c and 1/3 of lysozyme. For a total concentration level of 1.2 mg/mL at a mixing ratio of 2:1, the solution contains 0.8 mg/mL cytochrome c and 0.4 mg/mL lysozyme. In total, 32 samples were used for model calibration.

8 samples were added as a test set for model validation. The cytochrome c and lysozyme stock solutions were pipetted in the desired mixing ratios and diluted with the respective buffers on the liquid handling station. All 40 samples were prepared as 1.8 mL solutions in 96 deep-well plates to avoid small pipetted volumes. 300 µL of each sample were transferred to 96-well flat bottom UV-Star microplates. Sample absorption spectra were measured in a range of 240 – 300 nm in 2 nm steps. The spectral data of the 32 samples mentioned above were used for model calibration in Matlab R2011a, using the PLS toolbox (Eigenvector Research, USA). 5 samples out of 32 were used as an internal cross validation. The regression model was then validated on the external test set of additional 8 samples and used for the concentration determination of unknown samples.

2.2.2.2   Generation of Equal Adsorber Volumina

Generation of equal 20.8 µL adsorber amounts was achieved using the Media Scout Resi Quot system (Atoll, Germany) described by Herrmann et al. [82]. The system was equipped with a pressure-controlled vacuum pump and the working pressure was set to 800 mbar. To remove the adsorber storage solution, the adsorber plaques were washed twice with deionized water and the applied binding buffer of the respective experiment. The equilibrated plaques were transferred into a 2 mL 96-well square deep well plate and suspended in 100 µL binding buffer. The plate was then stored until use on the liquid handling station. The outer wells were not used for isotherm experiments due to the largest variance in adsorber volume on the plate [148].

2.2.2.3   Isotherm Experiments

Isotherms covering 10 different starting concentrations c_{in} were generated on the liquid handling station. As only the inner 60 wells were used for the experiments, 6 isotherms could be created per sample plate. Each adsorber plate was used for one pH level including 5 different salt level isotherms (42, 65, 90, 115, and 150 mM) and one isotherm as duplicate for investigating the repeatability of isotherm data.

The investigated pH levels were pH 5 and 7. 10 mg/mL stock solutions of 70 % cytochrome c and 30 % lysozyme in the respective buffers were applied. The starting concentrations c_{in} of the isotherm experiments were set to 2, 2.5, 3, 3.5, 4, 4.5, 5, 6, 7, and 8 mg/mL with a final volume of 800 µl per well (including 100 µL adsorber storage buffer from plaque generation). The adsorber plate was closed with a lid by the robotic plate-moving arm and placed in the orbital shaker for 2 h. Kinetic studies for lysozyme [50] and a monoclonal antibody [15] have shown that an incubation time of 20 to 40 min is sufficient for reaching the binding equilibrium on SP Sepharose FF. Afterwards, the plate was centrifuged for 10 min at 1000 rpm in the Rotanta 46RSC centrifuge. 300 µL of the resulting supernatant was transferred to a 96-well flat bottom UV-Star microplate. As for model calibration, sample absorption spectra were measured in a range of 240–300 nm in 2 nm steps. The spectral data were processed using the previously calibrated regression model. Lysozyme and cytochrome c concentrations could be determined selectively. The validity of cytochrome c levels was additionally assured by comparison to 527 nm absorption measurements.

2.2.2.4   Isotherm Fitting

The purpose of isotherm fitting is to validate, whether the observed measurements follow the theoretical framework for protein adsorption in ion-exchange chromatography. The adsorption model applied is the semi-mechanistic steric mass action isotherm (SMA) introduced by Brooks and Cramer [24]. It incorporates effects of counter-ions on the retention behavior of proteins and includes characteristic charges of proteins ν_i. Additionally, steric shielding effects of proteins are considered as a parameter σ_i, representing sterically hindered binding sites without electrostatic interactions. As a final factor, the total ionic capacity Λ of the applied adsorber represents the total number of electrostatic binding

sites. The isothermal form of the SMA isotherm for a mixture of two proteins is shown in Eq. (2.1) and Eq. (2.2), with q_i and $c_{eq,i}$ being the concentration of the protein i adsorbed and in solution, respectively. The effective pore salt concentration is described by c_{salt}. $k_{eq,i}$ is the equilibrium constant of adsorption and desorption.

$$q_1 = k_{eq,1} \left(\frac{\Lambda - (\nu_1 + \sigma_1)q_1 - (\nu_2 + \sigma_2)q_2}{c_{salt}} \right)^{\nu_1} c_{eq,1} \qquad (2.1)$$

$$q_2 = k_{eq,2} \left(\frac{\Lambda - (\nu_2 + \sigma_2)q_2 - (\nu_1 + \sigma_1)q_1}{c_{salt}} \right)^{\nu_2} c_{eq,2} \qquad (2.2)$$

The SMA model allows for fitting isotherm data, including concurring binding behavior and varying salt concentrations. The fitting procedure was carried out in Matlab R2011a using the least squares data fitting function lsqcurvefit (trust-region-reflective algorithm). The estimated parameters include the equilibrium constant $k_{eq,i}$, the characteristic charge ν_i, and the steric shielding σ_i.

2.3   Results and Discussion

2.3.1   Model Generation & Experimental Performance

Partial least squares regression (PLS) is used to reduce data sets and finding significant variance for correlating several input variables (e.g. wavelengths) with output variables (e.g. concentrations). Input variables of similar information content are lumped as so-called latent variables (LVs) leading to a data reduction. The first LV carries the highest information content, whereas each additional LV added leads to less and less improvement of the model. At a certain number of LVs, further addition of LVs then leads to incorporation of measurement noise into the system which needs to be avoided.

For generation of the MVDA model, the optimal number of latent variables was found to be 5, yielding normally distributed residues for the cross validation of all 32 samples. This setup was found identical for both pH 5 and pH 7. The generated MVDA models were then applied to the external test set consisting of 8 samples. The maximal relative deviation in concentration was determined as 3.9 % for pH 5 and 5.3 % for pH 7. Hence, the model performed well for a set of samples that were not used for the calibration experiments.

Besides testing the performance of the MVDA model, also the repeatability of the experimental data was investigated. The duplicates of the isotherms in the presence of 90 mM ionic strength at pH 5 and pH 7 are shown in Supplementary Fig. B.2 for cytochrome c (left) and lysozyme (right). The protein bound to the adsorbent is plotted versus the residual protein in free solution for equilibrium conditions. The trends of the duplicates (cross/diamond) agree well for both investigated pH values and proteins. Also, the absolute values of the duplicates are in good agreement considering the experimental difficulties when working with low adsorbent and liquid volumes. Consequently, both the obtained MVDA model and the experimental data were of high quality and were used in the presented study.

2.3.2   Multi-component Isotherms

To prove the applicability of the MVDA model for real isotherm data in a mixture of proteins, the concentrations of cytochrome c derived from the MVDA are compared to the values of a selective 527 nm analytical wavelength (see Supplementary Fig. B.3). The data points from the MVDA model, plotted over the selective 527 nm wavelength measurements, are shown in a parity plot. It has to be noted that the MVDA model was calibrated in the UV range (240 – 300 nm) whereas the selective wavelength for cytochrome c is at a much higher wavelength in the VIS region. For pH 5 (Supplementary Fig. B.3A) as well as for pH 7 (Supplementary Fig. B.3B), the isotherm data agree well with coefficients of determination of the data points for all investigated ionic strengths (ISs) of 99.74 % for pH 5 and 99.34 % for pH 7. Thus, the MVDA model was applicable to this large data set for straightforward HTS applications.

The isotherm data for both lysozyme and cytochrome c are shown in Fig. 2.1 as 2D (left) and 3D (right) scatter plots. The experimental data are illustrated as markers and the fitted SMA model is shown as curves and planes. For all data sets, the binding of proteins decreases with increasing ionic strength (from 42 mM to 150 mM IS). For cytochrome c (red), both pH conditions (pH 5 – Fig. 2.1A and pH 7 – Fig. 2.1B) show a similar trend for all ionic strengths: The isotherm data increase for low starting concentrations c_{in} and decrease for higher values of c_{in}. However, the overall binding of cytochrome c at pH 5 (Fig. 2.1A) is slightly increased when compared to pH 7 (Fig. 2.1B). E.g., in the setups of 42 mM IS, the maximum binding capacity $q_{Cyt\,c}$ was determined to be 63.98 mg/mL for pH 5 and 50.13 mg/mL for pH 7. Again, for lysozyme (bottomblue), both pH conditions behave similarly for all ISs. In contrast to cytochrome c, the isotherm data increase consistently with the starting concentrations c_{in}. The overall binding behavior of lysozyme at pH 5 (Fig. 2.1A) is slightly decreased when compared to pH 7 (Fig. 2.1B). The maximum binding capacity q_{Lys} at 42 mM IS was determined to be 67.48 mg/mL for pH 5 and 74.77 mg/mL for pH 7. Although the initial amount of lysozyme in the mixture was much lower (30 %) compared to cytochrome c (70 %), the maxima of q_{Lys} exceed those of $q_{Cyt\,c}$.

The experimental results encountered follow the expected trends. As observed for all investigated setups, an increase in ionic strength causes a weakening of the electrostatic binding in ion exchange chromatography. Cytochrome c showed a strong increase in protein binding for low starting concentrations c_{in}, starting to decrease with higher values of c_{in}. This indicates a displacement of cytochrome c by lysozyme when the binding process approaches the maximal binding capacity of the adsorber. This assumption was confirmed by the trends encountered for lysozyme. Here, a continuous increase in protein binding was observed towards a maximum for the highest starting concentrations c_{in}. This displacement under all investigated conditions agrees with the isoelectric points of the two proteins being 10 to 10.5 for cytochrome c and 11.4 for lysozyme (compare data sheet Sigma-Aldrich) yielding a higher net charge for lysozyme and thus a stronger binding to the cation exchange resin. This also explains the enhancement of the displacement reaction for pH 7 compared to pH 5 due to cytochrome c being closer to its nominal pI.

2.3.3   SMA Data Fitting

For SMA data fitting, the total ionic capacity Λ for the adsorber plaques was estimated first. A packed 1 mL of SP Sepharose FF has a total ionic capacity of 800 mM [148].
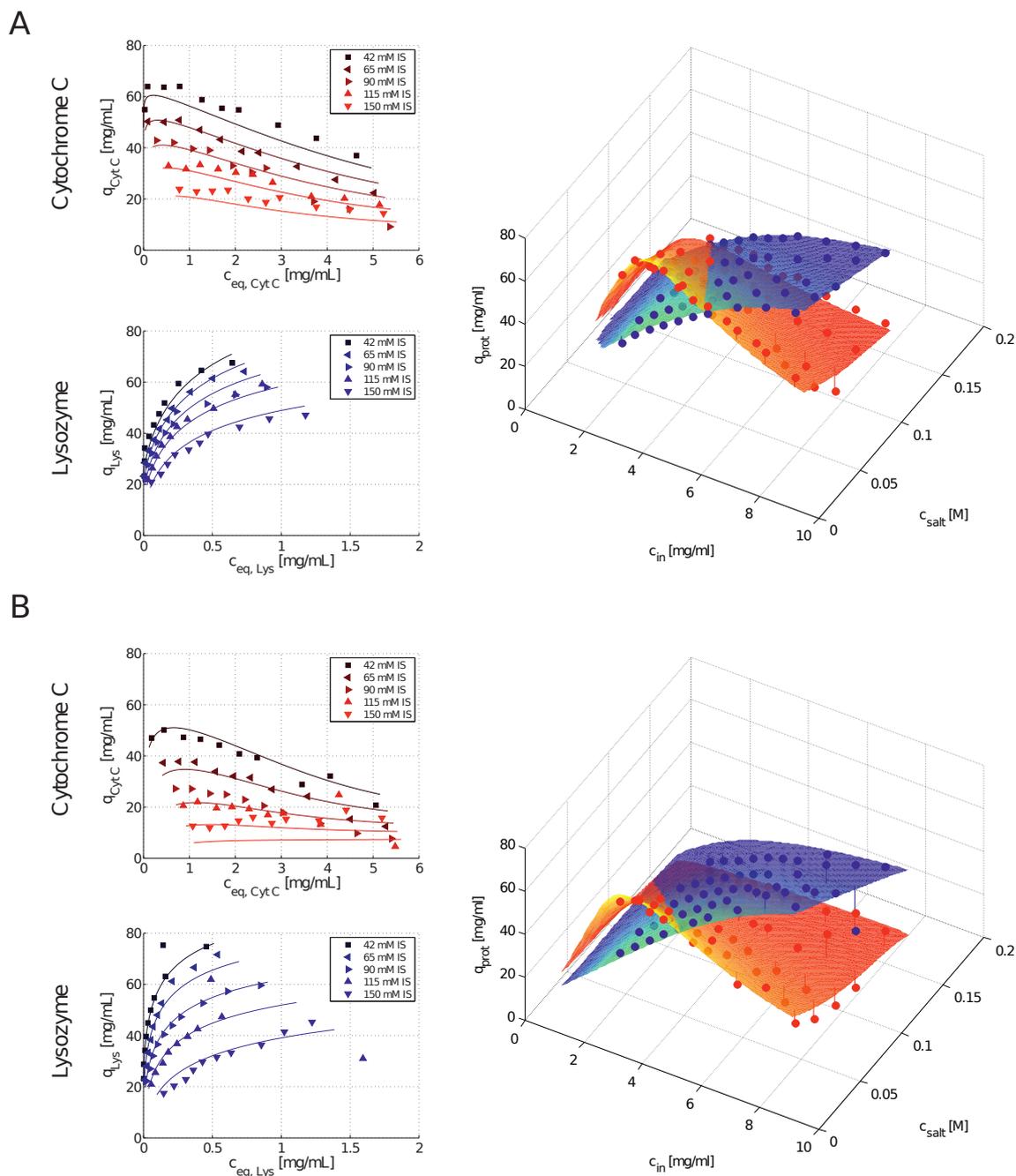
Figure 2.1: Isotherm data derived from the MVDA model-based equilibrium concentrations for both cytochrome c (red) and lysozyme (blue) and the corresponding SMA isotherm fitting curves. The experimental results at pH 5 (A) and pH 7 (B) are illustrated for all investigated ionic strengths in a range of 42 mM to 150 mM. The experimental data points are highlighted as markers and the SMA isotherm fitting curves are illustrated as solid lines for 2D plots and as planes for 3D plots.
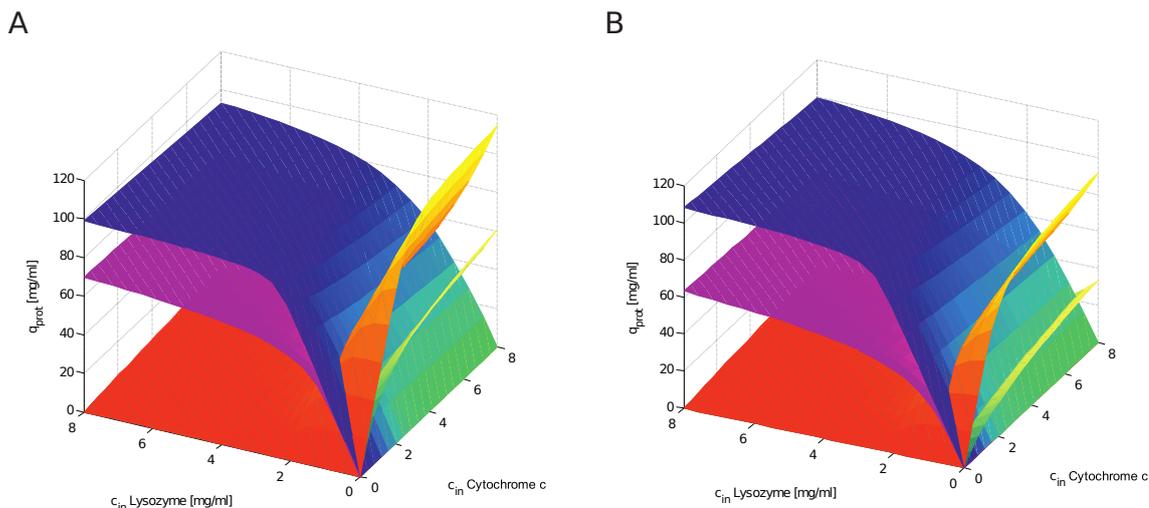
Figure 2.2: redicted SMA model for different concentrations and mixing ratios of cytochrome c and lysozyme at pH 5 (A) and pH 7 (B). Lysozyme is shown in blue (42 mM IS) and pink (150 mM IS), cytochrome c is illustrated in red (42 mM IS) and yellow (150 mM).

Table 2.1: SMA parameters of cytochrome c and Lysozyme at pH 5 and pH 7 determined by least squares fitting of the MVDA isotherm data.

Protein	pH	$k_{eq,i}$	ν_i	σ_i
Cytochrome c	5	0.010	9.955	45.775
Cytochrome c	7	0.040	4.827	36.877
Lysozyme	5	0.632	7.945	48.278
Lysozyme	7	0.454	5.834	51.871

The calculation of the equivalent Λ for the adsorber plaques being 0.504 mM followed the description by Hermann et al. using the packed-bed porosity factor [82]. The data fitting was performed in Matlab R2011a using q_i and $c_{eq,i}$ being the concentration of the protein i adsorbed and in solution inside the pore as input variables according to Eqs. (2.1), (2.2). The corresponding SMA fitting functions are illustrated in Fig. 2.2 by solid lines plotted combined with the respective MVDA data points. The fits were found to be of good agreement, the experimental isotherm points are matched well and the competitive binding trends were conserved.

The resulting SMA parameter sets for both proteins under the investigated conditions are listed in Table 2.1. The lysozyme equilibrium constants $k_{eq,i}$ are much higher (ranging from 0.454 to 0.632) compared to cytochrome c (0.010 to 0.040) for both pH 5 and pH 7. Comparable values and trends were reported by Gallant et al. for both proteins at pH 6 ($k_{eq,Lys} = 0.124$, $k_{eq,Cyt\,c} = 0.006$) [57]. As shown in Table 2.1, the characteristic charge ν_i of cytochrome c was determined to be 20 % higher than for lysozyme at pH 5 whereas for pH 7, the same trend is shown inversely. ν_i values given by Gallant et al. at pH 6 ($\nu_{Lys} = 5.95$, $\nu_{Cyt\,c} = 6.15$) as the center point between pH 5 and pH 7 indicated an identical charge characteristic for both proteins agreeing well with this inversion of ν_i from pH 5 to pH 7 shown in this study. Also, the absolute values for ν_i match well. The steric shielding parameters σ_i shown in Table 2.1 are similar for all proteins under all conditions ranging from 36.877 to 51.871. Those values are in accordance with Gallant et al. for cytochrome c (53.4) and Osberghaus et al. for lysozyme (29.7 – 36.8) and

cytochrome c (28.7 – 40.8) [148].

The overall model quality depends on the identifiability of isotherm parameters from the chosen experiments. To determine the range of the design space used for model calibration, the total concentration of occupied binding sites in equilibrium Λ_{bound} is calculated:

$$\Lambda_{bound} \quad = \quad \Lambda - \sum_i \left(\nu_i + \sigma_i \right) q_i \tag{2.3}$$

Pairs $(q_{Cyt\,c}, q_{Lys})$ can be taken from the experimental data in [mg/mL] and divided by the respective molecular weights. $MW_{Cyt\,c} = 12.38$ kDa and $MW_{Lys} = 14.30$ kDa were taken from the Sigma Aldrich data sheet and the SMA parameters for pH 5 and pH 7 are given in Table 2.1. For the lowest salt concentration and highest initial concentration at pH 5, Λ_{bound} adds up to 432.5 mM (166.7 mM cytochrome c, 265.8 mM lysozyme) and for pH 7 to 371.5 mM. For the lowest initial concentration, we obtain $\Lambda_{bound,pH5} = 338.7$ mM and $\Lambda_{bound,pH7} = 302.7$ mM. Comparing these values to the total ionic capacity of Λ of 504 mM, the loading is in a range of 60 % to 86 %.

This could be an explanation for the deviation of the equilibrium constants from literature values. As the linear range of the isotherms is not covered by the experiments, the 95 % confidence intervals for k_{eq} are the largest ($k_{eq,Cyt\,c,pH5}$ 94 %, $k_{eq,Cyt\,c,pH7}$ 233 %, $k_{eq,Lys,pH5}$ 68 %, $k_{eq,Cyt\,c,pH7}$ 187 %). However, it should be be noted that Gallant et al. were investigating cytochrome c and lysozyme at pH 6 and not at pH 5 and 7. Absolute values were thus not comparable but relative trends were conserved. The confidence intervals for the steric shielding parameter were the smallest ($\sigma_{Cyt\,c,pH5}$ 13 %, $\sigma_{Cyt\,c,pH7}$ 34 %, $\sigma_{Lys,pH5}$ 11 %, $\sigma_{Cyt\,c,pH7}$ 28 %) as they can be read from the concentration in the saturated state. For the characteristic charge we determined slightly larger intervals ($\nu_{Cyt\,c,pH5}$ 20 %, $\nu_{Cyt\,c,pH7}$ 42 %, $\nu_{Lys,pH5}$ 21 %, $\nu_{Cyt\,c,pH7}$ 41 %). As the parameters are still correlated, additional experiments could be done to minimize the uncertainty, e.g. following [10]. Nevertheless, it can be concluded that the SMA model is able to reproduce the observed binding behavior.

Finally, the fitted SMA model was plotted for different protein concentration levels c_{in} of both proteins for pH 5 (Fig. 2.2A) and pH 7 (Fig. 2.2B). Lysozyme is shown in blue (42 mM IS) and pink (150 mM IS), cytochrome c is illustrated in red (42 mM IS) and yellow (150 mM). Data points along the x and y axes (0 mg/mL lysozyme or cytochrome c) are the single component isotherms showing a linear slope for low concentrations and resulting in a maximum (q_{max}) at the adsorber saturation. The model predicts an almost constant maximal binding capacity at pH 5 and pH 7 at both ISs for lysozyme whereas the binding of cytochrome c is enhanced at pH 5 especially for 150 mM IS compared to pH 7. Furthermore the model in Fig. 2.2 predicts that already 2 mg/mL of lysozyme present in the mixture suppresses the binding of cytochrome c on the adsorber for both pH setups. These findings agree with the experimental data which indicated a diminished binding of cytochrome c (Fig. 2.1) though being the species of excess concentration in the mixture (70 % cytochrome c/ 30 % lysozyme).

2.3.4   SMA Parameters vs. Experimental Data

The equilibrium constant $k_{eq,i}$ as a factor describing the binding affinity of proteins under given conditions being over 10-fold higher for lysozyme compared to cytochrome c (compare Table 2.1) is in accordance with the isotherms given in Fig. 2.1. The displacement

of cytochrome c by lysozyme at higher starting concentrations c_{in} was discussed above. Surprisingly, the corresponding characteristic charge ν_i is not the crucial factor driving this phenomenon. E.g., for pH 5, ν_{Lys} of 7.945 is lower than $\nu_{Cyt\,c}$ of 9.955 but still the equilibrium is strongly shifted towards lysozyme. However, the above-discussed effects of increased displacement of cytochrome c at pH 7 (towards the pI of both proteins) agree with the change in characteristic charge from pH 5 to pH 7 which decreases strongly for cytochrome c from 9.955 to 4.827 whereas lysozyme is much less affected. In summary, the equilibrium constant was found to be a factor for indicating displacement phenomena whereas the characteristic charge shows the extent. The shielding factors σ_i being similar for both proteins were expected as lysozyme and cytochrome c are similar in size. Lysozyme as the slightly larger molecule ($MW_{Lys} = 14.30$ kDa, $MW_{Cyt\,c} = 12.38$ kDa) resulted in marginally higher σ_i values.

2.4 Concluding Remarks

It was demonstrated that multivariate data analysis (MVDA) of protein spectra is a straightforward method of determining accurate concentration levels for protein mixtures for HTS applications such as batch isotherms. The data points were of high quality and were thus usable for modeling purposes. Based on the MVDA data, accurate SMA fits were found that coincide with literature values.

Analyzing the presented amount of samples using a standard analytical method, such as analytical chromatography takes approximately 5 min per sample or above. Using the presented MVDA method, in contrast, 10-15 s per sample are sufficient. The time requirement per sample is one of the key aspects in the concept of QbD, as the bottleneck of HTS and exploring the design space is shifted to the analytics. Besides this, the major advantage over traditional analytical tools, its non-invasive nature, is very convenient in the context of performing experiments on robotic work stations, making sampling obsolete. The proposed methodology can be used for any multi-component mixture that do show differences in single-component absorption spectra. Brestrich *et al.* used the methodology for a distinction between antibody monomers, aggregates and lower molecular weight species though being spectrometrically similar [10; 23]. Such species can be distinguished due to an increase or decrease of exposed amino acid residues and by included scattering effects for larger aggregates. The analytical technology based on MVDA of protein spectra might become a standard tool for robotic work stations and in-line analytics[22; 23]. The technique of model calibration might be extended to more complex protein compositions and proteins which are not available as pure components.

Acknowledgments

3 | Simulating and Optimizing Preparative Protein Chromatography with ChromX

Tobias Hahn[1], Thiemo Huuk[1], Vincent Heuveline[2], Jürgen Hubbuch[1]

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

[2] Heidelberg University, Interdisciplinary Center for Scientific Computing, Engineering Mathematics and Computing Lab, Heidelberg, Germany

Abstract

Industrial purification of biomolecules is commonly based on a sequence of chromatographic processes, which are adapted slightly to new target components, as the time to market is crucial. To improve time and material efficiency, modeling is increasingly used to determine optimal operating conditions, thus providing new challenges for current and future bioengineers.

At the Karlsruhe Institute of Technology (KIT), mechanistic modeling of protein chromatography has long been part of the curriculum of the Bioengineering master's degree program, supported by exercises using simulation software. Emphasis lies on nonlinear preparative chromatography, where the result strongly depends on the sample concentration. For undergraduate students to gain hands-on experience in model-based optimization, a three-week, in-depth laboratory course was designed on the purification of a ternary mixture of proteins using

ion-exchange chromatography and mechanistic modeling. Students apply in-house software ChromX, which is made available for download, together with tutorials on numerics and practical applications.

This article presents the working principle of ChromX and results of the laboratory course for undergraduate students.

3.1   Introduction

In a biopharmaceutical production sequence, the biggest expenditure is associated with the purification of the target component from a very heterogeneous mixture, so-called downstream processing (DSP). Industrial DSP development, especially for monoclonal antibodies (mAb), is commonly based on a sequence of chromatographic processes, which are adapted slightly to new target components [115; 171].

In order to find optimal process parameters, modeling tools are increasingly gaining the attention of the pharmaceutical industry [78]. For using these tools, bioengineers are needed to acquire a new set of skills. Statistics, numerical mathematics, scientific computing, and data mining will play an increasing role in the future.

In the fourth semester of the bioengineering bachelor's degree program at the Karlsruhe Institute of Technology (KIT), students choose a three-week in-depth laboratory course in enzyme technology, biochemical engineering, or downstream processing. The DSP course applies cation-exchange chromatography to purify an antibody from a ternary protein mixture. Previously, a systematic trial-and-error approach was pursued for this purpose, which can be misleading as unexpected non-linear effects may occur. In contrast to analytical chromatography, where diluted samples are injected, industrial production processes are performed in the so-called nonlinear or preparative mode. The maximum binding capacity of the chromatographic medium is approached by injecting a large amount of concentrated feed solution. This leads to concentration-dependent *nonlinear* retention, as well as to competition of the species in the feed for binding sites.

Although many research groups developed their own simulation tools and published results generated with it, only few simulators are publicly available. The six simulators found for nonlinear chromatography are: Aspen Chromatography [8], CADET [52], Chromulator [184], ChromWorks [32], gProms [156], and pcs [118]. Aspen, ChromWorks, and gProms are commercial Windows applications that cannot be downloaded directly. They are not considered further because students should be able to use the software outside the university on their own. Nevertheless, Aspen has been used successfully by instructors to teach chromatography [200]. CADET is the most advanced noncommercial simulator in terms of numerical methods, but its C++ code must be compiled manually and does not offer a graphical interface. pcs is based on MATLAB, which usually is not taught in undergraduate programs. Moreover, MATLAB licenses become expensive once students graduate. Chromulator software was used in the KIT master's degree program in the past. As it became more troublesome to run the software on current versions of Microsoft Windows, we decided to extend our simulation code for research, ChromX [85], with a user-friendly interface and also to use it in undergraduate courses as a "black box". We are

convinced the separation processes will be designed almost entirely in a computer-assisted fashion in the future and, hence, want to provide our students with the opportunity to gain first experience.

ChromX for Microsoft Windows is available for download with step-by-step tutorials on the different phenomena occurring in liquid chromatography [98]. In addition, a set of MATLAB files explains the numerical methods. A short overview of models and their numerical solutions is given in the Chap. 1, the Supporting Information B.1 contain some exemplary exercises from the master's degree program. A comprehensive introduction to preparative and nonlinear chromatography is in the text book by Schmidt-Traub [129], and more in-depth analyses are presented in the book by Guiochon et al. [72].

3.2 Theory

3.2.1 Chromatography Models

In column liquid chromatography, the sample is dissolved in a liquid (mobile phase) and flows through a packed bed of porous particles or a monolithic column (stationary phase). For simplicity, a bed of uniformly sized spherical particles is assumed in the following sections. The physical or chemical properties of the stationary phase and the different components are utilized in a way that some components are retained more strongly than others. As illustrated in Figure 3.1, molecules are transported through the fluid outside of the particles, and then enter the particle's pore system and diffuse inside the pores. Adsorption and desorption are followed by diffusion out of the particle. The mass transport through the column and pores is described by modeling the fluid dynamics effects, while the retention of the species is described by empirical or mechanistic models for adsorption/desorption and/or reaction [72; 129].

For a large molecule that cannot enter the pore system, transport in flow direction depends on the pump speed and bed porosity (alias, *column porosity*). All diffusive effects in the interstitial volume leading to the broadening of an injected pulse are assumed to follow Fick's law of diffusion with a lumped *axial dispersion* coefficient. For smaller non-interacting molecules, additional effects are included, depending on the model complexity chosen. These include a *film transfer* coefficient that models the inhibition of transition into and out of the pores by the boundary layer, and a *pore diffusion* coefficient that accounts for intra-particle diffusion. Diffusion on the surface is usually neglected [72].

Adsorption onto the surface is modeled by an *isotherm* equation that describes the concentration of adsorbed protein as a function of the protein concentration in the mobile phase at constant temperature. For ion-exchange chromatography, the steric mass action (SMA) isotherm [24] was employed, as described in Chap. 1. It allows for modeling the influence of the counter-ion concentration in the mobile phase on the sorption behavior of proteins.
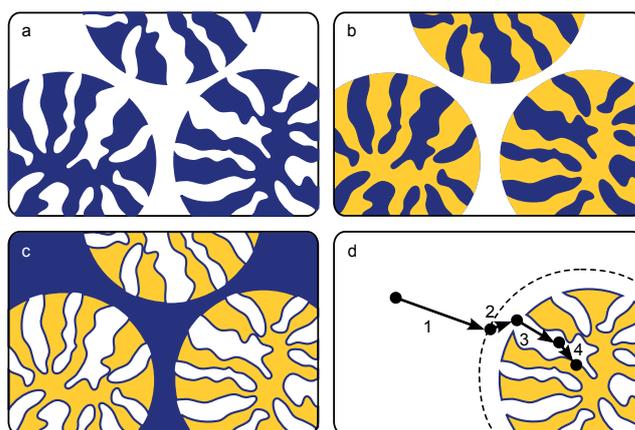
Figure 3.1: Structure of packed beds (a-c) and adsorption process (d). The total column volume is divided into the volume of the stationary phase (a), the volume of mobile phase in the pore system (b), and the interstitial volume of the mobile phase (c). Molecules are transported through the interstitial volume (d.1), pass a boundary layer (also referred to as film) around the beads (d.2), diffuse within the pore system (d.3), and adsorb onto the surface (d.4).

3.2.2   Numerical Solution

For each species, a system of partial differential equations must be solved: a convection-diffusion equation for the interstitial volume of the mobile phase, a diffusion equation for the pore volume, and an isotherm equation for adsorption/desorption. As all species are coupled via the isotherm in nonlinear chromatography, no closed-form solution is available and numerical solutions must be computed. When only linear chromatography is considered, specialized tools provide for a higher performance [21; 166; 175].

The tutorials distributed with ChromX describe the step-by-step solution of the equations with MATLAB and ChromX. First, convection and diffusion problems are solved in axial or radial flow systems. Pore diffusion models are added next, before adsorption and desorption are simulated.
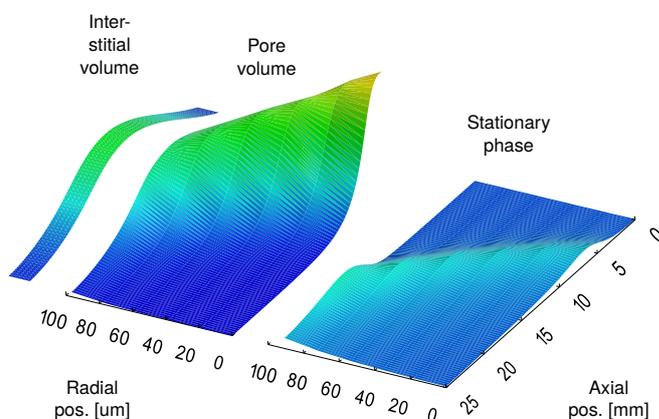
Figure 3.2: Qualitative visualization of intra-column and intra-particle concentrations during a step elution with ParaView.
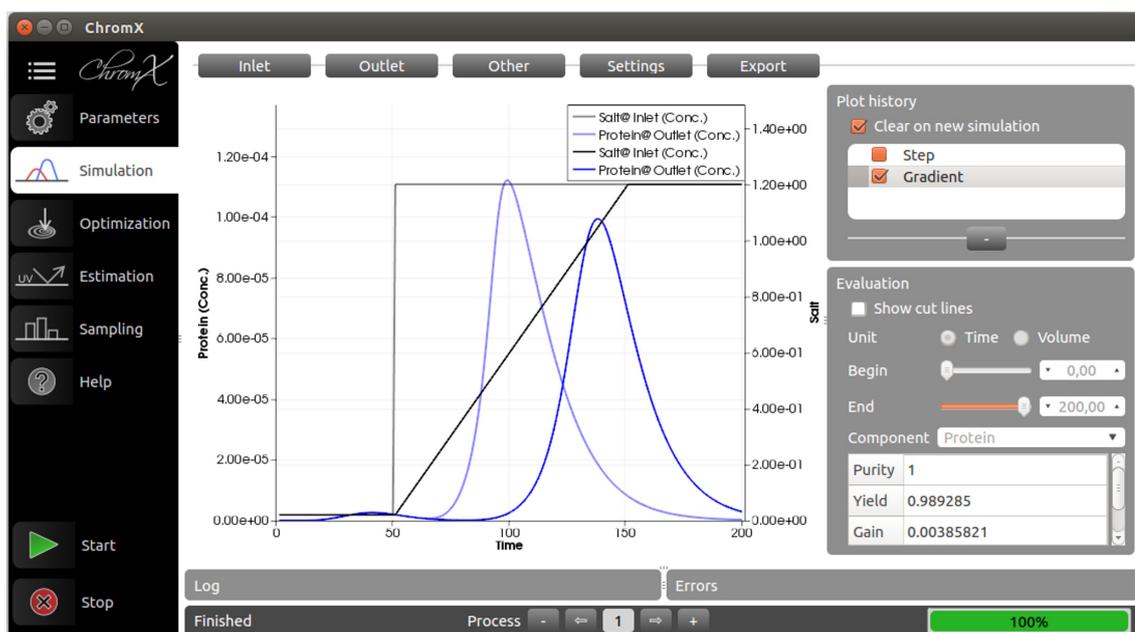
Figure 3.3: ChromX user interface: Comparison of peak shapes when switching from step (semi-transparent lines) to gradient elution (solid lines).

3.3 ChromX Features

The core of ChromX is the modular equation solver. The version used for teaching hides most of the parameters (types of finite elements, solver tolerances, matrix formats, hardware-specific settings, etc.) to simplify the program for new users. Fixed units were not prescribed for various reasons. Simple binding models, such as the Langmuir isotherm, can be used with any unit that is able to quantify concentrations, e.g., M, mg/mL or even absorption units [129]. Moreover, switching from CADET using meter as a standard unit of length or from Chromulator that uses non-dimensional formulations to other tools is facilitated. The embedded Help page provides information in the form "film transfer co-efficient k_f [length/time]" in the section on models; recommendations are given in the Supporting Information B.1.

Column and component-specific parameters can be edited easily. Complex injection and elution profiles are defined in terms of concentration events, e.g., a salt step, followed by a gradient and a final high-salt step. Proteins can be added, removed, and sorted in the interface. The number of proteins is only limited by the computer's memory.

A process setup can be cloned to simulate two or more chromatography runs in parallel with, e.g., different gradient lengths or slopes. This is done during parameter estimation, where several experiments are necessary to determine all isotherm parameters by chromatogram fitting. ChromX contains built-in heuristic and deterministic optimization routines for this task.

When all protein parameters are determined, optimization can be performed with respect to a chosen objective function. Alternatively, hypercube sampling [14] can be used to

study the dependence of the objective function on system parameters, such as batch, variations, etc.

ChromX includes extensive evaluation and export functions. How simulation results can be evaluated and compared easily is illustrated in Figure 3.2. First, a step elution is simulated with the SMA isotherm model and compared to a gradient elution. The results are plotted in a semi-transparent manner and each curve can be evaluated in terms of yield, purity, and peak area.

All results can be exported as a bitmap, vector graphic or spread sheet in XLS format. Intra-column time series can be exported in the VTK format and represented with, e.g., the open source visualization software ParaView [4] (Figure 3.3). This is explained in a dedicated tutorial [98].

3.4   Software Exercises

Software exercises accompany the lecture series on chromatography modeling and are intended to provide a deeper understanding of the mathematical operators and parameters in the differential equations and their effect on the chromatogram. System characteristics are to be pre-calculated with correlations for setups with and without dispersion and film kinetic effects. The calculated breakthrough and retention times are validated with the simulation software and Langmuir isotherm:

- Effects of a change in the feed concentration, i.e., switching from linear chromatography to preparative settings

- Axial dispersion affecting the shape of a breakthrough curve

- Shape changes for breakthroughs and isocratic elutions with different film kinetic effects and pore diffusion limitations

- Multi-component settings that demonstrate that superimposition of single-component simulations do not approximate a mixture in preparative chromatography.

Exemplary exercises and solutions are included in the Supporting Information B.1.

3.5   Laboratory Course Structure

In the DSP course, cation-exchange chromatography is used to purify an antibody from a ternary protein mixture. Students are in three groups of two students each. The first group performs process development with a statistical design of experiments, the second applies model-based process optimization, and the third conducts the necessary analytics for quality assessment. This sharing of tasks between process development and analytics mimics DSP development at pharmaceutical companies where different departments interact. For the mutual exchange of experience among the groups, daily progress meetings

are organized. The course starts with presentations on the theory, methods, and proposed proceeding, followed by two weeks of lab work. The course finishes with a poster presentation of the results and a preparation of a report of 25-30 pages. The total time students have to invest is estimated to be 120 hours. The only prerequisite is a course on biotechnological separation processes that introduced the principle of chromatography.

The objectives of the model-based process optimization group are to calibrate a model using tracer experiments, as well as step and gradient elutions, to perform simulations outside of the lab to optimize the separation of proteins, and to validate these simulation results. Excerpts from the lab assignments are given in Boxes 1 and 2. The lab course results are available as a case study, including all measurements and parameter files.

Box 1. Excerpts from Laboratory Assignments: Process Development with a Statistical Design of Experiments.

The students are to use an ion exchange step to purify the given feedstock, which contains the target protein (monoclonal antibody) and two other proteins as impurities. The concentration of the antibody is about 3.8 g/L. The concentrations of the impurities are unknown. The two impurities have theoretical isoelectric points of 7 and 11 and molecular weights of 17 kDa and 14 kDa. The first impurity has a chromophore group, which accounts for the dark color of the solution. The feedstock has a pH of about 8 and a conductivity of about 5 mS/cm. They are to reach the highest possible purity, yield, production rate, and concentration of the antibody. First, the type of ion exchanger and an appropriate adsorbent and buffer are to be selected. For regulatory reasons, the choice of buffers is to be limited to the use of simple buffer salts and NaCl. Both batch binding experiments and column chromatography may be performed. The process is to be carried out in a 1 mL column with a fast protein liquid chromatography (FPLC) system. The flow rate is to be chosen such that the contact time is roughly equivalent to 1 min (empirical value). All other process parameters, such as the elution mode (step or gradient), gradient length, ionic strengths, load volume, and pooling criteria, may be chosen freely. To optimize the development process, a statistical design of experiments (DoE) is to be made, with the parameters being varied by a software provided. The experiments are to be analyzed in cooperation with the analytics group. For evaluation of the DoE and in order to find the optimal process parameter combination, an objective function consisting of the goals listed above is to be found. This is to be done in agreement with the model-based development group.

Box 2. Excerpts from Laboratory Assignments: Model-based Process Development

The students are given the same separation task and the same feedstock as described in Box 1. However, the process is to be designed in a model-based manner. For the characterization of the column as well as of the feedstock, additional experiments will be required. The column selection is to be done by the first group. For this purpose, experiments are to be planned, which provide the basic information needed for modeling. Afterwards, the component-specific model parameters of the individual components of the feedstock are to be determined. With these values, process simulation is to be performed, considering the above recommendation for the flow rate. The students are to find a combination of elution mode (step or gradient), gradient length, ionic strengths, load volume, and pooling criteria, by means of which a high purity, yield, production rate, and concentration of the antibody can be achieved. Finally, a validation run is to be performed under the optimal conditions identified. The results are to be compared to those of the first group.

3.6 Laboratory Results

Initial experiments were conducted to determine system properties such as dead volumes, column and bead porosity, and the ionic capacity of chromatography resin. Afterwards, four experiments in the bind/elution mode were performed for parameter estimation, i.e., one step elution and two gradient elutions with a low sample volume, as well as one gradient elution with a large sample volume. The proceeding is described in detail in the Supporting Information B.1.

Exemplary results are shown in Figures 3.4a and 3.4b. A small breakthrough of nonbinding myoglobin and antibody was accurately modeled in the beginning. This also applied to the elution far beyond the detector limit in the experiment with a high sample volume.

Finally, the calibrated model was used for process optimization. The objective of the laboratory course was to find a compromise between the production rate and yield, while achieving a high purity. There is no unique solution to this problem. Students decided in favor of a very fast process, which concentrates the antibody at the same time (sample volume 3.6 mL, collected volume 2 mL). The step elution shown in Figure 3.4c reaches a purity above 99 % at 80 % yield. The results were subsequently validated experimentally in the lab.

3.7 Pedagogical Aspects

The DSP laboratory course for undergraduate students was taught for the third time in the 2014/2015 winter semester. Students appeared to learn preparative chromatography in more depth compared to the previously used trial-and-error approach. The course goals described above were met with this teaching format, although undergraduate students were familiarized with the models necessary for this particular task (SMA for ion-exchange chromatography) only. Because of the small number of students, statistical statements
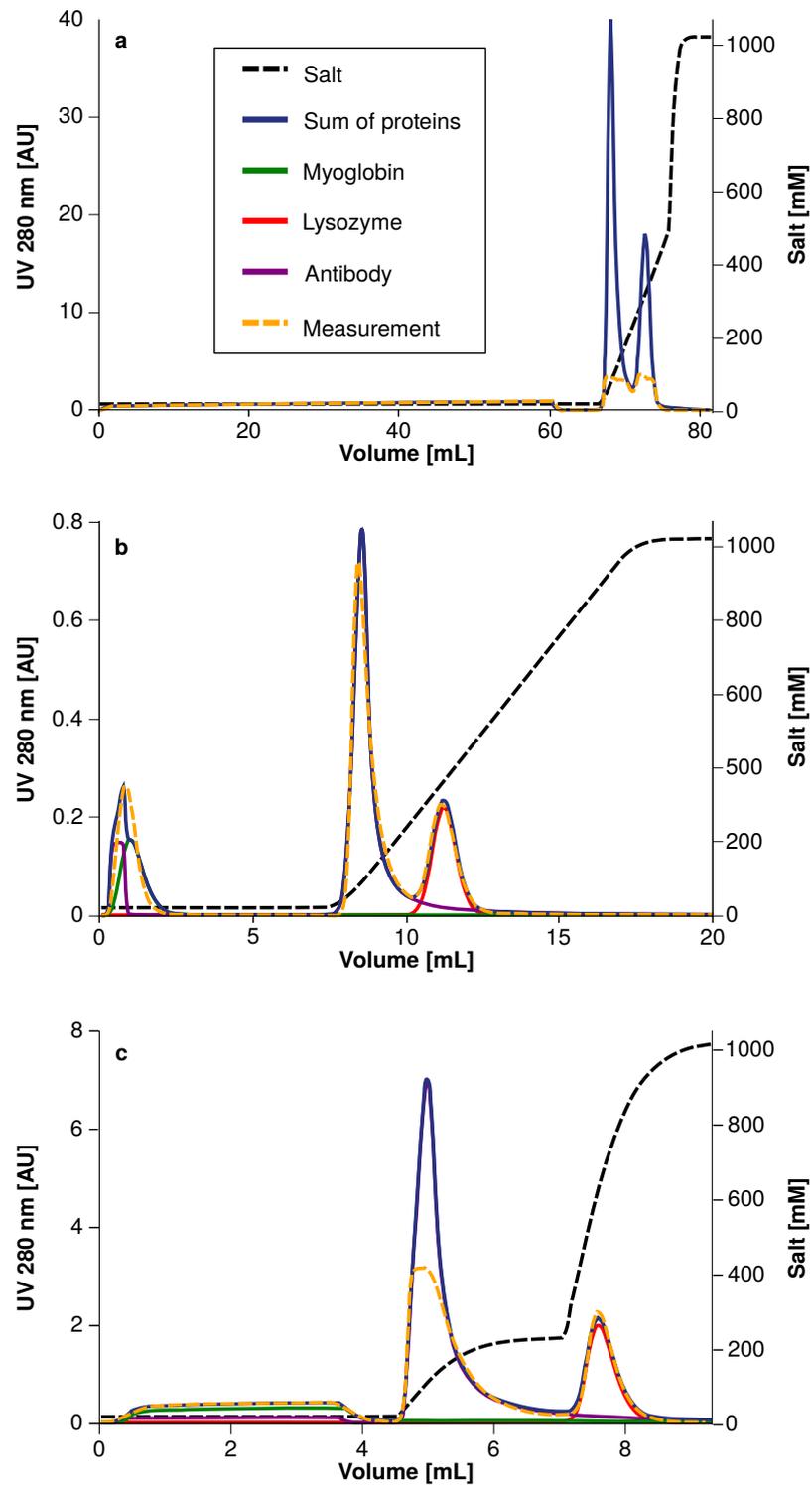
Figure 3.4: Comparison of simulated and measured chromatograms for the experiments with 60 mL sample volume and 20 - 520 mM salt gradient (a), 0.5 mL sample volume and 20 - 1020 mM salt gradient (b), and optimization result with 3.6 mL sample volume and 230 mM salt step elution (c). The UV detector was saturated at 3 AU such that the elution peaks in (a) and (c) were not completely recorded. Single component curves were omitted in (a) for reasons of clarity.

cannot be made, but statistics on the lecture series of the master's degree program are encouraging and given in the Supporting Information B.1.

Compared to conventional process optimization, ChromX provided an immediate feedback that might mislead students into trying setups without first anticipating results. An experiment that takes 30 minutes or longer has to be considered more carefully in order not to waste time and material. Hence, the grading scheme reflected the conception of the lab course: 20 % project plan presentation, 40 % practical work, 20 % report, and 20 % poster and poster presentation. For grading, the course of process design rather than the actual results were considered. Nevertheless, the solutions found by the model-based optimization group were almost identical to those of the group focusing on process optimization with a statistical design of experiments. The long runtime of the experiments and the daily progress meetings allowed for an extensive exchange among the groups and joint learning.

3.8   Conclusion

In pharmaceutical practice, preparative chromatography is of highest importance. To demonstrate the differences from analytical chromatography, a user-friendly interface was developed for use in a simulation toolbox for research and focused on models with a reasonable level of detail, i.e., where changes in model parameters directly related to peak properties. Easy visualization and export of results were other design criteria. This allowed us to establish an undergraduate laboratory course on the purification of a ternary protein mixture using mechanistic modeling.

The feedback from students was very positive. ChromX allowed them to study the influence of system and model parameters on the outcome of experiments. It also prepared them for the challenges lying ahead in industrial bioengineering.

Associated content

Supporting Information

ChromX for Microsoft Windows is available free of charge for academia [98]. On the ChromX webpage, we offer introductory material that explains the underlying models, their mathematical solution, and parameter estimation for MATLAB and ChromX. In this way, instructors can incorporate ChromX into their existing classroom and laboratory activities. ChromX will be further developed. We welcome any feedback and suggestions concerning the software, the tutorials as well as the case studies.

Mathematical background, working principle of ChromX and the MATLAB tutorials, software exercises with Langmuir isotherm, and experimental results of the laboratory course. This material is available free of charge via the Internet at http://pubs.acs.org.

Acknowledgment

4 | Adjoint-based Estimation and Optimization for Column Liquid Chromatography Models

Tobias Hahn[1], Anja Sommer[1], Anna Osberghaus[1], Vincent Heuveline[2], Jürgen Hubbuch[1]

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

[2] Heidelberg University, Interdisciplinary Center for Scientific Computing, Engineering Mathematics and Computing Lab, Heidelberg, Germany

Abstract

Simulation and optimization of chromatographic processes are continuously gaining practical importance, as they allow for faster and cheaper process development. Although a lot of effort has been put into developing numerical schemes for simulation, fast optimization and estimation algorithms also are of importance. To determine parameters for an a priori defined model, a suited approach is the inverse method that fits the measurement data to the model response.

This paper presents an adjoint method to compute model parameter derivatives for a wide range of differentiable liquid chromatography models and provides practical information for the implementation in a generic simulation framework by the example of ion-exchange chromatography.

The example shows that the approach is effective for parameter estimation of model proteins and superior to forward sensitivities in terms of computational effort. An optimization of peak separation in salt step elution demonstrates that the method is not restricted to inverse parameter estimation.

4.1 Introduction

Chromatography modeling has a long history in academia, as it fosters the understanding of the underlying physical and chemical processes. The potential of modeling for industrial

applications is enormous. It ensures time- and material-efficient process development - if and only if the parameters for a chosen model can be determined with low sample consumption in short time.

In column liquid chromatography the sample is dissolved in a liquid (mobile phase) and flows through a packed bed of porous particles or a monolithic column (stationary phase). The physical or chemical properties of the stationary phase and the different components are utilized such that some components are retained more strongly than others. The mass transport through the column is described by modeling the fluid dynamics, while the retention of the species is described by empirical or (semi-)mechanistical models for adsorption, reactions, etc. [129]. The most commonly employed models in liquid chromatography describe the mass transport in the column by so-called Convection-Diffusion-Reaction (CDR) equations, where the reaction term models phase transitions and eventually the retention of the species. If no deeper insight into the retention mechanism is available, Langmuir-type models are often employed as sub-models for the reaction part, introducing at least two unknown parameters per component. With no a priori knowledge about the components' behavior, the inverse method is an attractive option, which alters parameters in a systematic fashion to achieve a match of measured chromatogram and model prediction. For some sub-models, such as the Steric Mass Action (SMA) model for ion-exchange chromatography of proteins [24], model calibration protocols exist [24], which allow for determining the component-specific parameters in a consecutive fashion. In a comparative study [149], this approach and the inverse method were found to reach equal prediction quality such that the latter is recommended for fast process development. For other modes of protein chromatography, e.g. hydrophobic interaction (HIC) and mixed mode chromatography (MMC), different types of models are available. They are partly based on thermodynamic approaches that involve hardly measurable quantities, e.g. for HIC [132] or MMC [142]. In these settings, inverse parameter estimation is the only available option. If several models come into consideration, model discrimination can be performed that depends even more on fast parameter estimation techniques.

Optimization algorithms for minimizing the error between simulation and measurements can be divided into deterministic and heuristic methods. Deterministic algorithms, such as steepest descent or Newton's method, can achieve local quadratic convergence rates by using derivative information, but might get trapped in local optima. When starting near the global optimum, they constitute the methods of choice. The problem of local optima is due to the non-linearity of most retention models. Heuristic algorithms can leave local optima by random jumps, but require a much larger number of model evaluations. This class includes genetic algorithms [207] or simulated annealing [94]. A combination of deterministic and heuristic methods was studied in [207]. It represents a feasible approach in case of highly non-linear parameter dependencies.

To use a gradient-based deterministic algorithm like Newton's method, the gradients or *sensitivities* have to be computed for each unknown parameter. The *adjoint* or *backward sensitivity* method was presented earlier [88] for an ideal model with the Langmuir isotherm equation. This algorithm solves an adjoint equation backwards in time, allowing to compute any parameter sensitivity by an additional integration involving the *adjoint* or *dual solution*. Later, the approach was considered to be complicated, difficult to derive, and error-prone [53], mostly because of the need for additional model-specific derivatives. Instead a refined finite difference approach was proposed, which computes the needed

sensitivities by solving one additional problem in time per parameter. In a recent publication [158], this approach was applied to compute sensitivities with higher precision. Automatic differentiation was used to avoid manual errors, but adding such routines to an existing code requires thorough programming knowledge and particular insight into memory management when dealing with iterative solvers [80]. The achievable precision is higher compared to classical finite differences, but the computational effort is reported to be slower by a factor of 1.4 [158]. It is arguable, whether highest precision is necessary throughout the whole estimation or optimization process. Especially when starting the iteration, faster computation is more important than exact intermediate derivatives.

For systems with a large number of component-specific parameters (e.g. for the target molecule, variants, fragments, aggregates, etc.), heuristic methods as well as forward sensitivity approaches may be computationally expensive and intractable. We hence derived a general formulation of the adjoint method for differentiable liquid chromatography models in the form of CDR equations, demonstrating that the derivation is manageable for a wide range of scenarios. The "error-prone" derivative calculation may be accomplished reliably with computer algebra systems or even online derivative calculators. We will comment on practical implementation and in particular on re-usability of the existing code. A practical example shows that the approach is effective and far superior to forward sensitivities in terms of computational effort.

The approach will be demonstrated for two test cases of SMA parameter estimation, first using four chromatograms of cytochrome c in different gradient elutions and second using three chromatograms of a mixture of cytochrome c, lysozyme, and chymotrypsin in different gradient experiments. In this study, no multiple local minimums were observed, but long drawn-out minimums skew to the model parameters (Fig. 4.3a), depending on the formulation of the SMA model equation. This may result in saddle points which cause the same problems as local minimums for algorithms only considering first-order sensitivities. To avoid the calculation of second-order derivatives, the problem of locally vanishing derivatives is mitigated by choosing a suitable formulation of the kinetic SMA isotherm equations, such that first order derivatives are sufficient for successful parameter estimation.

The adjoint method can also be used in optimization algorithms, for example to determine optimal process parameters occurring in the boundary conditions. The gradients of the objective function with respect to the involved parameters (pump speed, buffer/sample concentrations, etc.) can be computed using the adjoint method in the same way. An example shows the peak separation optimization of the ternary mixture by determining the optimal salt concentration for an elution step.

4.2   Theory

4.2.1   Chromatography Models

A variety of models are employed depending on the phenomena occurring in the respective system, but most of them can be formulated using a common structure. Current models for column chromatography describe the changes of concentrations in three phases, the

mobile phase concentration c, the pore phase concentration (mobile phase within the porous particles) c_p, and the stationary phase concentration q. Details on the derivation and an overview of models can be found in [72; 129]. In the following section, we will discuss models with the general structure:

$$
\begin{align}
0 = \quad F_1(c, c_p) \quad &= F_{1,1}(c, c_p) \cdot c + F_{1,2}(c, c_p) \cdot c_p, \tag{4.1}\\
0 = \quad F_2(c, c_p, q) \quad &= F_{2,1}(c, c_p, q) \cdot c + F_{2,2}(c, c_p, q) \cdot c_p + F_{2,3}(c, c_p, q) \cdot q, \tag{4.2}\\
0 = \quad F_3(c_p, q) \quad &= F_{3,2}(c_p, q) \cdot c_p + F_{3,3}(c_p, q) \cdot q \tag{4.3}
\end{align}
$$

where the terms $F_{m,n}$ are general operators in phase equation $m \in \{1, 2, 3\}$ that occur as pre-factor to a phase concentration c, c_p, q indicated by $n \in \{1, 2, 3\}$. The concentration variables are vectors containing quantities for all considered components, e.g. $c = (c_0, c_1, \ldots)^T$.

Typically, the concentration transport in the column is assumed to be independent of the cross-sectional position, such that the interstitial concentration only depends on the time $t \in [0, T]$ and axial position $x \in [0, L]$ in a column of length L and is influenced by convection, diffusion, and interaction with the pore phase, e.g.

$$
\begin{align}
\frac{\partial c_i}{\partial t}(x, t) = \ & -u(t)\frac{\partial c_i}{\partial x}(x, t) + D_{ax}\frac{\partial^2 c_i}{\partial x^2}(x, t) \notag\\
& - \frac{1 - \varepsilon_c}{\varepsilon_c}\frac{3}{r_p}k_{eff}\left(c_i(x, t) - c_{p,i}(x, t)\right) \tag{4.4}
\end{align}
$$

where u is the interstitial velocity, D_{ax} the axial diffusion coefficient, $\frac{1-\varepsilon_c}{\varepsilon_c}\frac{3}{r_p}k_{eff}$ a coefficient modeling the effective mass transfer rate with respect to the adsorbent particle radius r_p and column porosity ε_c. Several assumptions were made to consider the velocity and axial dispersion as constant, the most fundamental ones are a uniformly packed column, an incompressible liquid and a constant viscosity. For a detailed discussion we refer to [72]. In the nomenclature of Eq. (4.1), the terms $F_{1,1}$ and $F_{1,2}$ are

$$
\begin{align}
F_{1,1}(t) \ &= \ \frac{\partial}{\partial t} + u(t)\frac{\partial}{\partial x} - D_{ax}\frac{\partial^2}{\partial x^2} + \frac{1 - \varepsilon_c}{\varepsilon_c}\frac{3}{r_p}k_{eff}, \tag{4.5}\\
F_{1,2} \ &= \ -\frac{1 - \varepsilon_c}{\varepsilon_c}\frac{3}{r_p}k_{eff} \tag{4.6}
\end{align}
$$

This model is complemented by Danckwerts boundary conditions at the inlet and Neumann conditions at the outlet

$$
\begin{align}
\frac{\partial c_i}{\partial x}(0, t) \ &= \ \frac{u(t)}{D_{ax}}\left(c_i(0, t) - c_{in,i}(t)\right), \tag{4.7}\\
\frac{\partial c_i}{\partial x}(L, t) \ &= \ 0. \tag{4.8}
\end{align}
$$

A typical example of a model for mass transfer into the adsorbent particle is the Transport Dispersive Model that uses (4.5) together with

$$
\begin{align}
F_{2,1}(c, c_p, q) \ &= \ -\frac{3}{\varepsilon_p r_p}k_{eff}, \tag{4.9}\\
F_{2,2}(c, c_p, q) \ &= \ \frac{\partial}{\partial t} + \frac{3}{\varepsilon_p r_p}k_{eff}, \tag{4.10}
\end{align}
$$

$$F_{2,3}(c, c_p, q) = \frac{1 - \varepsilon_p}{\varepsilon_p} \frac{\partial}{\partial t}. \tag{4.11}$$

Reduced models that do not include a pore phase, like the Equilibrium Dispersive Model

$$\frac{\partial c}{\partial t} = -\frac{\varepsilon_c}{\varepsilon_t} u \frac{\partial c}{\partial x} + D_{app} \frac{\partial^2 c}{\partial x^2} - \frac{1 - \varepsilon_t}{\varepsilon_t} \frac{\partial q}{\partial t} \tag{4.12}$$

also fit in this structure with $F_{1,2} = 1$, $F_{2,1} = 0$, $F_{2,2} = 1$, and $F_{2,3} = -\frac{1-\varepsilon_t}{\varepsilon_t} \frac{\partial}{\partial t}$. Here, the total porosity is given by $\varepsilon_t = \varepsilon_c + \varepsilon_p(1 - \varepsilon_c)$ and the apparent dispersion coefficient D_{app} is distinguished from the axial dispersion coefficient D_{ax}. D_{app} is only approximately constant for linear or nearly linear isotherms, but often used for nonlinear isotherms [129]. If the velocity and the axial or apparent dispersion coefficient depend on x, e.g. in radial flow chromatography [67], the operator $F_{1,1}$ is usually modeled to be independent of c and c_p, such that these models also fit in the structure.

The pore phase c_p might also include a radial dimension to cover particular effects, e.g. pore diffusion or blocking. An example of a radially dependent model is the following General Rate Model as formulated in [66]. Although it is not used below, it is included for completeness.

$$F_{2,2}(c, c_p, q) = \frac{\partial}{\partial t} - D_p \left(\frac{\partial^2}{\partial r^2} + \frac{2}{r} \frac{\partial}{\partial r} \right),$$

$$F_{2,3}(c, c_p, q) = +\frac{1 - \varepsilon_p}{\varepsilon_p} \frac{\partial}{\partial t}, \tag{4.13}$$

with boundary conditions

$$\frac{\partial c_p}{\partial r}(t, x, r_p) = \frac{D_p k_f}{\varepsilon_p}(c(t, x) - c_p(t, x, r_p)), \tag{4.14}$$

$$\frac{\partial c_p}{\partial r}(t, x, 0) = 0, \tag{4.15}$$

adding to $F_{2,1}$ and $F_{2,2}$. Here, ε_p is the particle porosity and D_p the pore diffusion coefficient.

The variety of retention sub-models is large, an often employed model in ion-exchange chromatography for proteins is the SMA model for components $i = 1, \ldots, n$

$$F_{3,2_i} = -k_{a,i}(x, t) \underbrace{\left(\Lambda - \sum_{j=1}^{n} (\nu_j + \sigma_j) q_j(x, t) \right)^{\nu_i}}_{\bar{q}_{Salt}}, \tag{4.16}$$

$$F_{3,3_i} = \frac{\partial}{\partial t} + k_{d,i}(x, t) c_{p,Salt}^{\nu_i}(x, t), \tag{4.17}$$

with

$$q_{Salt} = \Lambda - \sum_{j=1}^{n} \nu_j q_j(x, t) \tag{4.18}$$

$$= \bar{q}_{Salt} + \sum_{j=1}^{n} \sigma_j q_j(x, t). \tag{4.19}$$

where Λ is the ionic capacity of the adsorbent, ν, σ, k_a, and k_d are the characteristic charge, the steric shielding, the adsorption coefficient and desorption coefficient of the component, respectively. Of course, the stationary phase concentrations are also radially dependent when using e.g. the General Rate Model. In contrast to the previous operators, $F_{3,2}$ and $F_{3,3}$ are now non-linear because of the exponent ν.

For the SMA model this results in an equation system whose block structure resembles the one plotted in Figure 4.1. At this stage, the blocks still contain the space derivatives and might depend non-linearly on the concentrations. The right-hand side contains parts of the boundary conditions and other constants, in case of SMA the ionic capacity Λ.

For the HIC [132] and MMC [142] models, the structure looks similar. For isotherms of the Langmuir type, the bottom row simplifies as one of the operators $F_{3,2}$ or $F_{3,3}$ becomes linear:

- Equilibrium formulation [129]:

$$F_{3,2_i} = -k_{eq,i} q_{max,i}, \tag{4.20}$$

$$F_{3,3_i} = 1 + \sum_{j=1}^{n} k_{eq,j} c_{p,j}. \tag{4.21}$$

- Kinetic formulation [120]:

$$F_{3,2_i} = -k_{a,i} q_{max,i} \left(1 - \sum_{j=1}^{n} \frac{q_j}{q_{max,j}} \right), \tag{4.22}$$

$$F_{3,3_i} = \frac{\partial}{\partial t} + k_{d,i}. \tag{4.23}$$

4.2.2 Numerical Solution Schemes

4.2.2.1 Linearization

The non-linearities introduced by the retention sub-models are typically treated by the iterative solution of a linearized equation, e.g. using Newton's method or Picard iteration [159]. Picard iteration does not rely on the gradients $\frac{\partial F_1}{\partial c_i}$, $\frac{\partial F_1}{\partial c_{p,i}}$, $\frac{\partial F_2}{\partial c_i}$, etc., but becomes unstable for larger time steps. As both methods have to update the non-linear blocks, Newton's method is computationally favorable because of the faster convergence. Especially when using equilibrium isotherms ($\frac{\partial q}{\partial t} = 0$), the mass transfer per time step is not limited and the higher stability of Newton's method is essential.

Newton's method solves the system

$$
\begin{bmatrix}
\frac{\partial F_1(c,c_p)}{\partial c} & \frac{\partial F_1(c,c_p)}{\partial c_p} & 0 \\
\frac{\partial F_2(c,c_p,q)}{\partial c} & \frac{\partial F_2(c,c_p,q)}{\partial c_p} & \frac{\partial F_3(c,c_p,q)}{\partial q} \\
0 & \frac{\partial F_3(c_p,q)}{\partial c_p} & \frac{\partial F_3(c_p,q)}{\partial q}
\end{bmatrix}
\cdot
\begin{bmatrix}
c' - c \\
c_p' - c_p \\
q' - q
\end{bmatrix}
=
\begin{bmatrix}
-F_1(c,c_p) \\
-F_2(c,c_p,q) \\
-F_3(c_p,q)
\end{bmatrix}
\tag{4.24}
$$

4.2 THEORY

$$
\begin{bmatrix}
1 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 \\
0 & 1 & \ldots & 0 & 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & 1 & 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 \\
0 & 0 & \ldots & 0 & \square & 0 & \ldots & 0 & \square & 0 & \ldots & 0 \\
0 & 0 & \ldots & 0 & 0 & \square & \ldots & 0 & 0 & \square & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & 0 & 0 & 0 & \ldots & \square & 0 & 0 & \ldots & \square \\
0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 \\
0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 & 0 & 1 & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 1
\end{bmatrix}
\cdot \frac{\partial}{\partial t}
\begin{bmatrix}
c_{Salt} \\ c_1 \\ \vdots \\ c_k \\ c_{p,Salt} \\ c_{p,1} \\ \vdots \\ c_{p,k} \\ q_{Salt} \\ q_1 \\ \vdots \\ q_k
\end{bmatrix}
+
$$

$$
\begin{bmatrix}
\blacksquare & 0 & \ldots & 0 & \blacksquare & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 \\
0 & \blacksquare & \ldots & 0 & 0 & \blacksquare & \ldots & 0 & 0 & 0 & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & \blacksquare & 0 & 0 & \ldots & \blacksquare & 0 & 0 & \ldots & 0 \\
\blacksquare & 0 & \ldots & 0 & \blacksquare & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 \\
0 & \blacksquare & \ldots & 0 & 0 & \blacksquare & \ldots & 0 & 0 & 0 & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & \blacksquare & 0 & 0 & \ldots & \blacksquare & 0 & 0 & \ldots & 0 \\
0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 & \square & \square & \ldots & \square \\
0 & 0 & \ldots & 0 & \square & \square & \ldots & 0 & 0 & \square & \ldots & \square \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & 0 & \square & 0 & \ldots & \square & 0 & \square & \ldots & \square
\end{bmatrix}
\cdot
\begin{bmatrix}
c_{Salt} \\ c_1 \\ \vdots \\ c_k \\ c_{p,Salt} \\ c_{p,1} \\ \vdots \\ c_{p,k} \\ q_{Salt} \\ q_1 \\ \vdots \\ q_k
\end{bmatrix}
=
\begin{bmatrix}
BC_{in,Salt} \\ BC_{in,1} \\ \vdots \\ BC_{in,k} \\ 0 \\ 0 \\ \vdots \\ 0 \\ \Lambda \\ 0 \\ \vdots \\ 0
\end{bmatrix}
$$

Matrix types
◨ Diagonal
■ Band

Figure 4.1: Structure plot of the equation system. Matrices in the same block are colored identical to illustrate block-wise transposition compared to Figure 4.2. The entries of same-colored matrices in a block are not identical.

for $\left(c', c'_p, q'\right)^T$. The models presented for the interstitial and pore phases are linear in all concentration fields, such that non-linearities only occur in the blocks of the bottom row. The structure of the system will not change significantly, linear operators stay in place, only couplings hidden in the non-linearities are substituted:

$$
\begin{bmatrix}
F_{1,1}(c,c_p) & F_{1,2}(c,c_p) & \\
F_{2,1}(c,c_p,q) & F_{2,2}(c,c_p,q) & F_{2,3}(c,c_p,q) \\
0 & \frac{\partial(F_3(c_p,q))}{\partial c_p} & \frac{\partial(F_3(c_p,q))}{\partial q}
\end{bmatrix}
\cdot
\begin{bmatrix}
c' - c \\ c'_p - c_p \\ q' - q
\end{bmatrix}
=
\begin{bmatrix}
-F_1(c,c_p) \\ -F_2(c,c_p,q) \\ -F_3(c_p,q)
\end{bmatrix}
\quad (4.25)
$$

The additional effort is small. For the SMA model, the two derivatives are:

$$
\frac{\partial \left(F_{3,2}(c_p,q)c_{p,i} + F_{3,3}(c_p,q)q_i\right)}{\partial c_{p,j}} =
\begin{cases}
\nu k_d c_{Salt}^{\nu-1} q & \text{for } j = 0 \\
-k_a \bar{q}_{Salt}^{\nu} & \text{for } i = j > 0 \\
0 & \text{else}
\end{cases}
\quad (4.26)
$$

$$
\frac{\partial \left(F_{3,2}(c_p,q)c_{p,i} + F_{3,3}(c_p,q)q_i\right)}{\partial q_j} =
$$

$$
\begin{cases}
1 & \text{for } i = 0,\ j = 0 \\
\nu_j & \text{for } i = 0,\ j \neq 0 \\
-k_{a,i} c_{p,i} \nu_i \left(\nu_i + \sigma_i\right) \bar{q}_{Salt}^{\nu_i - 1} + k_d c_{Salt}^{\nu} & \text{for } i \neq 0,\ i = j \\
-k_{a,i} c_{p,i} \nu_i \left(\nu_j + \sigma_j\right) \bar{q}_{Salt}^{\nu_i - 1} & \text{for } i \neq 0,\ i \neq j
\end{cases}
\quad (4.27)
$$

4.2.2.2   Space Discretization

Discretization in space approximates the continuous concentrations by a finite number of nodal values on a computational mesh and the continuous operators are substituted for discrete versions. Using classical finite element notation, Equation (4.4) is written in a semi-discrete form as

$$
\begin{aligned}
M\frac{\partial \bar{c}_i}{\partial t}(x,t) &= \left(-uC + D_{ax}A - \frac{1-\varepsilon_c}{\varepsilon_c}\frac{3}{r_p}k_{eff}M\right)\bar{c}_i(x,t) \\
&\quad + \frac{1-\varepsilon_c}{\varepsilon_c}\frac{3}{r_p}k_{eff}M\bar{c}_{p,i}(x,t) + s
\end{aligned}
\tag{4.28}
$$

where C, A, and M are the convection, stiffness, and mass matrix, respectively, i.e. the discrete transport, diffusion, and mass storage operators. s includes the discretized sources and sinks. Hereinafter, it will not matter which discretization scheme is used. For details on implementing a certain scheme, the reader is referred to the following references in the context of chromatography: Finite differences [51], finite elements [89; 119], and finite volumes [196].

The different schemes will lead to different structures, mostly tridiagonal matrices for first-order approximations and 5-band matrices for second-order schemes. As there are no space derivatives in the retention model equations, the coupling with c_p is only point-wise, leading to diagonal matrices in the blocks of the bottom row. For finite differences this is obvious, as there are no operators to be treated and also for finite elements, the orthogonality of the basis functions leads to the same result when replacing the concentrations with their basis representations on $\{x_1, ..., x_d\} \subseteq [0, L]$:

$$
\sum_{l=1}^{d} \varphi_l(x_k)\frac{\partial q_{i_l}}{\partial t}(t)
$$

$$
= k_{a,i}\left(\sum_{l=1}^{d}\varphi_l(x_k)c_{p,i_l}(t)\right)\left(\Lambda - \sum_{j\geq 1}(\nu_j+\sigma_j)\sum_{l=1}^{d}\varphi_l(x_k)q_{j_l}(t)\right)^{\nu_i}
$$

$$
- k_{d,i}\left(\sum_{l=1}^{d}\varphi_l(x_k)q_{i_l}(t)\right)\left(\sum_{l=1}^{d}\varphi_l(x_k)c_{p,0_l}(t)\right)^{\nu_i}
\tag{4.29}
$$

$$
\stackrel{\psi_l(x_k)=\delta_{kl}}{\Longleftrightarrow} \frac{\partial q_{i_k}}{\partial t}(t) = k_{a,i}\left(c_{p,i_k}(t)\right)\left(\Lambda - \sum_{j\geq 1}(\nu_j+\sigma_j)q_{j_k}(t)\right)^{\nu_i}
$$

$$
- k_{d,i}\left(q_{i_k}(t)\right)\left(c_{p,0_k}(t)\right)^{\nu_i}
\tag{4.30}
$$

This results in an easy implementation: A retention model and its derivatives can be assembled in a vector of diagonal entries using point-wise vector operations.

The structure of the equation system after space discretization is given in Figure 4.1.

4.2.2.3   Time Discretization

For time discretization, implicit finite difference schemes are often employed because of their good stability and accuracy, e.g. the Crank-Nicolson or the fractional step θ

scheme [62]. Again, the block structure of the equation system does not change, only the matrix entries are different. For the following parameter estimation method, any of the above schemes can be chosen.

4.2.3 Parameter Estimation

The inverse method uses the chosen simulation scheme within an iterative optimization method. Starting from an initial guess, the model is evaluated and the quality of the result is assessed to obtain information on the location of a better parameter set. A standard approach is the method of least squares given by the objective function

$$R\left(p\right) = \int_0^T r\left(t; p\right) \, \mathrm{d}t = \int_0^T \left(\sum_i c_i(L, t; p) - m(t)\right)^2 \, \mathrm{d}t, \qquad (4.31)$$

the sum over all time steps of the squared difference between model prediction at the end of the column $c_i(L, t; p)$ and measured chromatogram $m(t)$. If parameters p are chosen close to the true unknown model parameters, the difference will be small and so will be the objective function. The transformation from the sum of molar concentrations to adsorption units often is sufficiently accurate when assuming a linear dependency, a more advanced technique is presented in [37].

As described in the introduction, gradient-based optimization schemes are more complicated in implementation, but can provide local quadratic convergence. They are the methods of choice to find the minimum of a reasonably well behaving function in short time. An overview of Newton methods is given in [44]. Important versions are the local Newton's method, also called Newton-Simpson, or in the scalar case, the Newton-Raphson method

$$R'\left(p^0\right) \Delta p \approx -R\left(p^0\right), \qquad (4.32)$$

and global Newton's method, which does not rely on a good initial guess p^0, but uses second order derivatives to compute the correction according to

$$R''\left(p^0\right) \Delta p \approx -R'\left(p^0\right). \qquad (4.33)$$

With second-order derivative information, saddle points can be distinguished from minimums and quadratic convergence is achieved. This often is not the case with the local Newton's method.

Second-order derivatives in general are hard to compute, such that a quasi Newton method might be employed, which uses an approximation of the Hessian R'' in order to improve the performance at the cost of little additional computation.

4.2.4 Derivative Computation

The partial derivatives occurring in R' can be approximated directly (forward method) e.g. by finite differences: $\frac{\partial R}{\partial p} \approx \frac{R(p+\varepsilon)-R(p)}{\varepsilon}$. This requires one additional gradient evaluation per parameter. In the setting of kinetic SMA (Eqs. (4.16)-(4.18)), there are four

parameters to be estimated for each component, leading to a high number of additional simulations.

Another approach is to explicitly evaluate the derivatives $\frac{dR}{dp} = \frac{\partial R}{\partial c}\frac{dc}{dp} + \frac{\partial R}{\partial p}$, though the terms $\frac{dc}{dp}$ are very tedious to derive. This approach was studied in [137] for a two-component Transport Dispersive/SMA model, but requires an increased implementation effort.

The adjoint or backward method represents a highly attractive compromise. The needed amount of additional computations is in the order of one model evaluation, while only the partial derivatives $\frac{\partial F}{\partial p}$ of the model are needed and not the total derivatives of the concentrations.

The algorithm is derived by applying a classical Lagrangian multiplier formulation to the problem. It is of importance whether the equations have already been discretized in time or not. As discussed in [113], calculating the adjoint of the discretization (AD) results in a gradient that is exact for the discrete objective function leading to complete convergence. Discretization of the adjoint (DA) is consistent with adaptive time stepping, but requires more knowledge to formulate correct boundary conditions for the adjoint equation. In the following section, we will reproduce the AD approach of [27], because it can be implemented easily. In case of adaptive solvers, we refer to the ADDA method presented in [113], which avoids the obstacles associated with DA.

The optimization problem to be solved is given by

$$
\begin{aligned}
\min_p R\,(p) &= \min_p \int_0^T r(p,t)\,\mathrm{d}t \\
F\,(c,\dot{c},t;p) &= 0 \\
G\,(c(t=0);p) &= 0
\end{aligned}
\tag{4.34}
$$

where F contains the space-discrete model equations with incorporated boundary conditions. This allows to treat boundary parameters like the feed concentration just as any other parameter. G contains initial conditions that might also depend on a variable parameter.

The derivation of the algorithm is straightforward. Following [27], the Lagrange multipliers λ and μ are introduced to form the augmented objective

$$
\mathcal{L}\,(c;p,\lambda) = R\,(p) - \int_0^T \left(\lambda^T F\,(c,\dot{c},t;p)\right)\,\mathrm{d}t + \mu^T G\,(c(t=0);p)
\tag{4.35}
$$

Since $F = G = 0$ by construction, the gradient of the Lagrange functional is equivalent to the gradient of the objective. Inserting Eq. (4.31) and differentiating yields

$$
\begin{aligned}
\frac{\mathrm{d}R}{\mathrm{d}p} = \frac{\mathrm{d}\mathcal{L}}{\mathrm{d}p} &= \int_0^T \left[\frac{\partial r}{\partial c}\frac{\mathrm{d}c}{\mathrm{d}p} + \frac{\partial r}{\partial p} + \lambda^T\left(\frac{\partial F}{\partial c}\frac{\mathrm{d}c}{\mathrm{d}p} + \frac{\partial F}{\partial \dot{c}}\frac{\mathrm{d}\dot{c}}{\mathrm{d}p} + \frac{\partial F}{\partial p}\right)\right]\,\mathrm{d}t \\
&\quad + \mu^T\left(\frac{\partial G}{\partial c(t=0)}\frac{\mathrm{d}c(t=0)}{\mathrm{d}p} + \frac{\partial G}{\partial p}\right).
\end{aligned}
\tag{4.36}
$$

First, time derivatives $\dot{c}$ are eliminated by integration by parts

$$
\int_0^T \lambda^T \frac{\partial F}{\partial \dot{c}}\frac{\mathrm{d}\dot{c}}{\mathrm{d}p}\,\mathrm{d}t = -\int_0^T \left(\dot{\lambda}^T\frac{\partial F}{\partial \dot{c}} + \lambda^T\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial F}{\partial \dot{c}}\right)\frac{\mathrm{d}c}{\mathrm{d}p}\,\mathrm{d}t + \left[\lambda^T\frac{\partial F}{\partial \dot{c}}\frac{\mathrm{d}c}{\mathrm{d}p}\right]_0^T
\tag{4.37}
$$

For the presented chromatography models, $\frac{\partial F}{\partial \dot{c}}$ is constant such that $\frac{d}{dt}\frac{\partial F}{\partial \dot{c}}$ equals zero and can be removed. Inserting the remaining parts into Eq. (4.36) and collecting terms in $\frac{dc}{dp}$ and $\frac{dc(t=0)}{dp}$ yields:

$$
\begin{aligned}
\frac{d\mathcal{L}}{dp} = & \int_0^T \left(\left(\frac{\partial r}{\partial c} + \lambda^T \frac{\partial F}{\partial c} - \dot{\lambda}^T \frac{\partial F}{\partial \dot{c}} \right) \frac{dc}{dp} + \frac{\partial r}{\partial p} + \lambda^T \frac{\partial F}{\partial p} \right) dt + \left[\lambda^T \frac{\partial F}{\partial \dot{c}} \frac{dc}{dp} \right]_{t=T} \\
& - \left(\left[\lambda^T \frac{\partial F}{\partial \dot{c}} \right]_{t=0} - \mu^T \frac{\partial G}{\partial c(t=0)} \right) \frac{dc(t=0)}{dp} + \mu^T \frac{\partial G}{\partial p}
\end{aligned}
\tag{4.38}
$$

Because of $F = G = 0$, $\lambda(t)$ and μ can be chosen freely, allowing to cancel the second term in the second row and first term in the bottom row when choosing $\lambda(T) = 0$ and $\mu^T = \left[\lambda^T \frac{\partial F}{\partial \dot{c}} \right]_{t=0} \left(\frac{\partial G}{\partial c(t=0)} \right)^{-1}$. Setting $\lambda(t)$ such that

$$
\frac{\partial r}{\partial c} + \lambda^T \frac{\partial F}{\partial c} - \dot{\lambda}^T \frac{\partial F}{\partial \dot{c}} = 0
\tag{4.39}
$$

eliminates all terms containing the complicated $\frac{dc}{dp}$. We are left with

$$
\frac{dR}{dp} = \frac{d\mathcal{L}}{dp} = \int_0^T \left(\frac{\partial r}{\partial p} + \lambda^T \frac{\partial F}{\partial p} \right) dt + \left[\lambda^T \frac{\partial F}{\partial \dot{c}} \right]_{t=0} \frac{\partial G}{\partial c(t=0)}^{-1} \frac{\partial G}{\partial p}
\tag{4.40}
$$

The algorithm to compute the derivatives needed by the optimization scheme is then

1. Solve the model equations $F(\dot{c}, c, t; p) = 0$ from $t = 0, \ldots, T$.

2. Determine $\lambda(t)$ for $t = T, \ldots, 0$ with $\lambda(T) = 0$ by solving Eq. (4.39).

3. Compute the derivative according to Eq. (4.40).

For the second step, we require $\frac{\partial F}{\partial \dot{c}}$, $\frac{\partial F}{\partial c}$, and $\frac{\partial r}{\partial c}$. The first two terms are already needed for the Newton iteration during the model evaluation and $\frac{\partial r}{\partial c}$ is simply $2 \left(\sum_i c_i(L, t; p) - m(t) \right)$. This term can be stored for later use in the calculation of the objective's value at the end of step 1. For solving the equation backwards in time, the structure of the sub-matrices in $\frac{\partial F}{\partial c}$ as well as the individual sub-blocks have to be transposed (change of block indexes during assembly). As the blocks of originally linear operators do not change, this can be done by re-using the already assembled matrices from the forward problem. The changes are illustrated in Figure 4.2. Furthermore, the problem to be solved is linear in λ, such that it should be faster to compute than the forward problem, as no Newton iteration is needed.

For the third step, the partial derivatives $\frac{\partial r}{\partial p}$, $\frac{\partial F}{\partial p}$, and are $\frac{\partial G}{\partial p}$ are needed. As a result of this step, the method is sometimes called error-prone, but nowadays, derivatives can be computed reliably with computer algebra systems or even online derivative calculators. When performing parameter estimation, the objective and initial conditions do not depend on model parameters. When performing e.g. load optimization with respect to initial salt

$$
\begin{bmatrix}
1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\
0 & 1 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & 1 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 & \blacksquare & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 & 0 & \blacksquare & \cdots & 0 & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & 0 & 0 & 0 & \cdots & \blacksquare & 0 & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 & \blacksquare & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 & 0 & \blacksquare & \cdots & 0 & 0 & 1 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & 0 & 0 & 0 & \cdots & \blacksquare & 0 & 0 & \cdots & 1
\end{bmatrix}
\cdot \frac{\partial}{\partial t}
\begin{bmatrix}
\lambda_{1,0} \\ \lambda_{1,1} \\ \vdots \\ \lambda_{1,k} \\ \lambda_{2,0} \\ \lambda_{2,1} \\ \vdots \\ \lambda_{2,k} \\ \lambda_{3,0} \\ \lambda_{3,1} \\ \vdots \\ \lambda_{3,k}
\end{bmatrix}
-
$$

$$
\begin{bmatrix}
\blacksquare & 0 & \cdots & 0 & \blacksquare & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\
0 & \blacksquare & \cdots & 0 & 0 & \blacksquare & \cdots & 0 & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \blacksquare & 0 & 0 & \cdots & \blacksquare & 0 & 0 & \cdots & 0 \\
\blacksquare & 0 & \cdots & 0 & \blacksquare & 0 & \cdots & 0 & \blacksquare & \blacksquare & \cdots & \blacksquare \\
0 & \blacksquare & \cdots & 0 & 0 & \blacksquare & \cdots & 0 & \blacksquare & \blacksquare & \cdots & \blacksquare \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \blacksquare & 0 & 0 & \cdots & \blacksquare & \blacksquare & \blacksquare & \cdots & \blacksquare \\
0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \blacksquare & \blacksquare & \cdots & \blacksquare \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \blacksquare & \blacksquare & \cdots & \blacksquare
\end{bmatrix}
\cdot
\begin{bmatrix}
\lambda_{1,0} \\ \lambda_{1,1} \\ \vdots \\ \lambda_{1,k} \\ \lambda_{2,0} \\ \lambda_{2,1} \\ \vdots \\ \lambda_{2,k} \\ \lambda_{3,0} \\ \lambda_{3,1} \\ \vdots \\ \lambda_{3,k}
\end{bmatrix}
=
\begin{bmatrix}
\frac{\partial r}{\partial c_{Salt}} \\ \frac{\partial r}{\partial c_1} \\ \vdots \\ \frac{\partial r}{\partial c_k} \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0
\end{bmatrix}
$$

Matrix types: ◣ Diagonal, ■ Band

Figure 4.2: Structure plot of the adjoint equation system. Diagonal matrices in the upper row are identical to the blocks of the same color in Figure 4.1. Band matrices in the second row are transposed versions of the matrices of the same color in Figure 4.1. Diagonal matrices in the bottom row have no relation to previously assembled matrices.

concentration, $\frac{\partial G}{\partial p}$ will be non-zero. For SMA, we obtain

$$
\frac{\partial F_{3,i}}{\partial k_{a,j}} = \begin{cases} -c_{p,i}\bar{q}_{Salt}^{\nu_i} & \text{for } i > 0,\ j = i, \\ 0 & \text{else,} \end{cases} \tag{4.41}
$$

$$
\frac{\partial F_{3,i}}{\partial k_{d,j}} = \begin{cases} q_i c_{Salt}^{\nu_i} & \text{for } i > 0,\ j = i, \\ 0 & \text{else,} \end{cases} \tag{4.42}
$$

$$
\frac{\partial F_{3,i}}{\partial \sigma_j} = \begin{cases} k_{a,i}c_{p,i}\nu_i \bar{q}_{Salt}^{\nu_i - 1} q_j & \text{for } i > 0, \\ 0 & \text{else,} \end{cases} \tag{4.43}
$$

$$
\frac{\partial F_{3,i}}{\partial \nu_j} = \begin{cases} k_{a,i}c_{p,i}\nu_i \bar{q}_{Salt}^{\nu_i - 1} q_j & \text{for } i,j > 0,\ j \neq i, \\ \left(k_{a,i}c_{p,i}\bar{q}_{Salt}^{\nu_i}\left(\ln(\bar{q}_{Salt}) - \frac{\nu_i \cdot q_i}{\bar{q}_{Salt}}\right) \\ + k_{d,i}q_i c_{Salt}^{\nu_i}\ln(c_{Salt})\right) & \text{for } i,j > 0,\ j = i, \\ q_j & \text{for } i = 0, \\ 0 & \text{else.} \end{cases} \tag{4.44}
$$

As mentioned above, these matrices consist of diagonal elements only and can be assembled quickly using vectorized operations. Of course, the results of the forward simulation are needed here, resulting in a higher memory consumption. For 5 components and 1000 time steps on a 100 node grid, only 12 MB and possibly a few GB for larger General Rate Model computations will be needed. This is no challenge for current workstations.

In fact, the time integration is just a sum of dot products $\lambda^T \frac{\partial F}{\partial p}$. Together with the assembly, this is much cheaper than a time step of the forward or backward problem in terms of computation.

In total, we obtain a scheme which provides the derivatives of the objective with respect to a high number of parameters in approximately the time of a model evaluation and allows for local quadratic convergence sufficiently near the minimum.

Computing second-order derivatives using an adjoint approach certainly is possible. Different approaches are covered in [151]. They require a minimum number of $2n_p + 1$ additional problems to be solved, with n_p being the number of unknown parameters. These approaches are appropriate for highly non-linear models, otherwise they are computationally inferior to local or Quasi Newton Methods.

4.3   Materials and Methods

To compare the adjoint methods to a finite difference-based forward method, chromatographic experiments with model proteins were conducted to determine SMA parameters from a set of obtained chromatograms. Furthermore, a suitable formulation of the SMA isotherm equation is chosen, which improves the estimation of model parameters.

4.3.1   Materials

Cytochrome c (bovine heart), lysozyme (chicken egg white), chymotrypsin (bovine pancreas), sodium citrate, and dextran 670000 were purchased from Sigma (St. Louis, MO, USA). Sodium chloride was provided by Merck KGaA (Darmstadt, Germany). Acetone and ethanol (99.8 %) were purchased from Carl Roth GmbH + Co.KG (Karlsruhe, Germany). For all experiments, a pre-packed POROS 50 HS 1ml cation exchange column was used. The resin was purchased from Applied Biosystems (Carlsbad, CA, USA) and the packing was done by Atoll (Weingarten, Germany). The experiments were performed on an AKTApurifier10 manufactured by GE Healthcare (Buckinghamshire, United Kingdom) with a constant flow rate of 0.838 ml/min (0.868 mm/s). Tubing connections and all other LC system parameters were kept constant. UV measurements were done at 280 nm and 527 nm. Primary analyses and documentation of the chromatograms were performed with the control software Unicorn from GE Healthcare. All further data analysis was performed with Matlab R2012a from Mathworks (Natick, ME, USA). Numerical experiments were performed on a Intel Core i7 workstation at 2.70 GHz with 8 GB RAM, running Ubuntu Linux 12.04 LTS.

4.3.2   Methods

For all experiments two buffer solutions were prepared, a 50 mM sodium citrate solution as low-salt (or running) buffer and a 50 mM sodium citrate + 1 M sodium chloride solution as high-salt buffer, both at pH 5.

Table 4.1: Column and system parameters.

Parameter	Symbol	Value
Total column volume	V_c	1 ml
Column length	l_c	20 mm
Pump flow	u_p	0.838 ml/min
System dead volume	V_{dead}	0.12 ml
Bed porosity	ε_c	0.32
Particle porosity	ε_p	0.54
Interstitial velocity	u	0.87 mm/s
Axial dispersion	D_{ax}	0.0016 mm/s
Ionic capacity	Λ	0.29 M

4.3.2.1 Determination of Column and System Parameters

The parameters for the Equilibrium Dispersive (4.12) and Transport Dispersive Model (4.5)-(4.11), excluding k_{eff}, were determined directly, since these are induced by the system, column packing or adsorber material and constant for all species. The system dead volume, including the distance to the UV sensor, was determined by examining the UV absorption of an acetone pulse without the column. The column porosity or void fraction can be determined by using a non-binding tracer that does not penetrate the pores (here, dextran), while for the particle porosity a non-binding tracer that penetrates the pores is needed (acetone). The axial dispersion value was calculated from the HETP value and the ionic capacity was determined by acid-base titration: The column was flushed with a 0.5M HCl solution until a constant UV and conductivity signal appeared. Afterwards, the column was washed with UPW until a constant UV and conductivity baseline was reached. Thereafter, the column was titrated at a flow of 100 cm/h with a 0.01 M NaOH solution until an increase in the conductivity signal was recorded. From the Na-ion concentration of the titrant and the volume of the applied titrant, the concentration of exchangeable ions was calculated. Results are shown in Table 4.1.

4.3.2.2 Calibration Experiments

Two protein solutions were prepared: 0.6 mM cytochrome c solution and 0.6 mM protein mixture solution, consisting of 0.2 mM lysozyme, 0.2 mM cytochrome c, and 0.2 mM chymotrypsin. All proteins were dissolved in low-salt buffer. Every run was performed at a constant flow of 0.838 ml/min.

After 2 column volumes (CV) of equilibration with low-salt buffer, $25\,\mu l$ of the sample were injected. 2 CV after injection, the gradient began and was fixed to a duration of 15 CV. The final NaCl concentration was kept for another 2 CV, until a 1 M high-salt buffer step of 2 CV completed the procedure.

For cytochrome c, four gradients with a final NaCl concentration of 0.15 M, 0.35 M, 0.55 M, and 0.75 M were performed, the protein mixture was run with gradients up to 0.25 M, 0.45 M, and 0.65 M in final concentration.

4.3.2.3   Algorithmic Implementation

Below, the Equilibrium Dispersive and Transport Dispersive Models are used in combination with the Steric Mass Action Model. Space discretization was performed using linear finite elements on a grid of 100 cells of the same size. Time integration for the forward and backward problem was computed with MATLAB's `ode15s` solver, a scheme for stiff ordinary differential equations accepting Jacobian matrices as derived for Newton's method, Eq. (4.24). The maximum step size of was set to 1 s to compute the least-squares error with sufficient precision. For estimation and optimization using the adjoint-based derivatives, a Newton-Simpson scheme was implemented.

4.3.2.4   Parameter Estimation Method for the SMA Model

First, the quality of the SMA parameter estimation for cytochrome c from four gradient experiments was evaluated against MATLAB's standard routine `lsqnonlin` for least-squares optimization, approximating the parameter sensitivities using a finite difference approximation.

The steric shielding parameter σ can be neglected in this setting, as concentrations near the saturation capacity did not occur at any point in the experiments, i.e. $\Lambda \gg (\sigma + \nu)q$ in Eq. (4.16). With the low concentrations used, the parameter is hardly identifiable and was set to a reasonable value of 30 for all components, as in [150].

As shown in [100], the peak position for low feed concentrations is only determined by the characteristic charge parameter ν and the equilibrium parameter $k_{eq} = \frac{k_a}{k_d}$, while the peak width is influenced by k_d. Starting from the kinetic formulation of the SMA model in Eqs. (4.16), (4.17) and estimating k_a ad k_d, a change in one of the parameters will simultaneously affect the peak height and retention time in the simulation. While a not perfectly matching peak height still guarantees an overlap of measurement and simulated chromatogram, a wrong retention time will usually cause a larger least-squares error. As soon as a good fit of retention time is found, k_a and k_d will have to change at the same time to keep k_{eq} constant, while matching the width. A gradient-based estimation only considers the change of residual value separately for every parameter, such that it hits a saddle point $\frac{\partial R(p)}{\partial k_a} \approx \frac{\partial R(p)}{\partial k_d} \approx 0$, if the improvement is skew to the parameter axes. The effect is shown for the cytochrome c case in Figure 4.3a.

To incorporate the search direction $\frac{k_a}{k_d}$, the SMA equation is divided by k_d such that the right-hand side only contains the retention time determining k_{eq} and ν and the left-hand side contains a kinetic parameter $1/k_d$ adjusting peak width:

$$\frac{1}{k_d}\frac{\partial q}{\partial t} = \frac{k_a}{k_d}\bar{q}_{Salt}^{\nu}c_p - c_{p,Salt}^{\nu}q. \tag{4.45}$$

Consideration of k_d or $\frac{1}{k_d}$ only affects the scale of the contour plot in Fig. 4.3b that still shows a long drawn-out minimum, but now perfectly aligned with the k_d-axis.
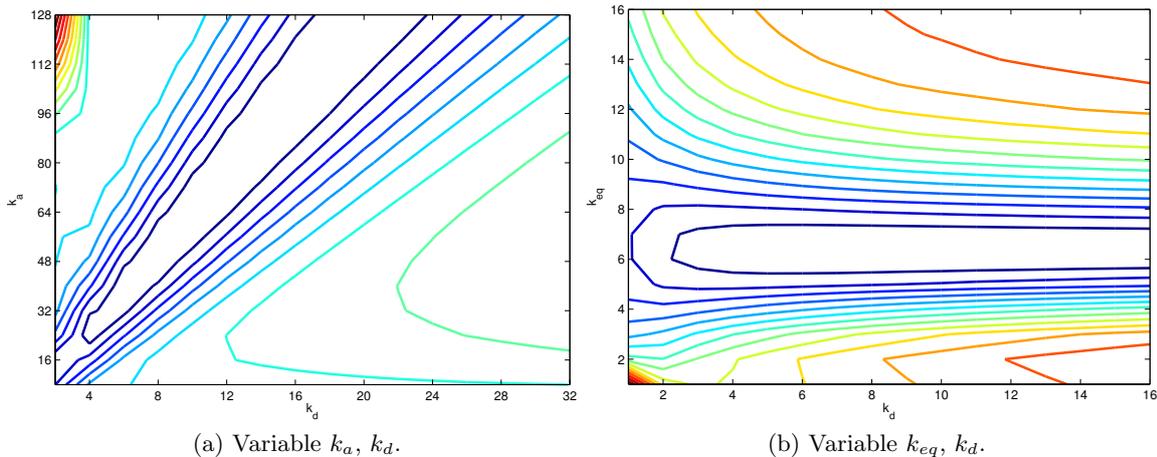
(a) Variable k_a, k_d.   (b) Variable k_{eq}, k_d.

Figure 4.3: Least squares error for 0.35 M gradient of the cytochrome c test case with fixed $\nu = 2.667$, $\sigma = 30$.

4.4 Results and Discussion

4.4.1 Initial Experiments

Figure 4.4 presents exemplary chromatograms for cytochrome c with an elution gradient concentration of 0.55 M and the mixture of chymotrypsin, cytochrome c, and lysozyme for a gradient with a final concentration of 0.45 M. The proteins' concentration is measured via their UV adsorption at the column outlet. All proteins show absorption at a wavelength of 280 nm as indicated in blue, cytochrome c absorbs also at 527 nm, shown in light blue. The salt gradient as prescribed at the inlet is given in percent in red and the measured conductivity at the outlet in green. The protein mixture chromatogram in Figure 4.4b shows three main peaks. From the isoelectric points of the proteins and the absorption of cytochrome c at 527 nm, it is obvious that chymotrypsin (pI 8.75) elutes first, cytochrome c (pI 9.59) second, and lysozyme (pI 9.6) last. The small pre-peak is caused by impurities or variants of chymotrypsin.

4.4.2 Parameter Estimation

For the test case with cytochrome c only, an Equilibrium Dispersive Model, Eq. (4.12), is employed, such that only the SMA parameters are to be estimated.

Estimation of parameters for cytochrome c of the canonical kinetic SMA formulation in Eq. (4.16)-(4.18) starting from the initial guess $\{k_a, k_d, \nu\} = \{10, 10, 3\}$ gave no improvement at all (data not shown). Starting from $\{10, 10, 4\}$, the result for both algorithms is plotted in Figure 4.5, showing a very poor fit. The finite difference approximation converged to $\{10.65, 8.78, 4.03\}$ and the adjoint method to $\{10.06, 7.91, 3.74\}$. Both methods match the back slope of the peak and then get stuck, unable to increase the steepness.

Using the adapted version of the SMA model Eq. (4.45), curve fitting results in a parameter set which matches peak height and retention time. For the adjoint method, a simple

(a) Experimental results for cytochrome c with a gradient to 0.55 M.

(b) Experimental results for the protein mixture (cytochrome c, chymotrypsin, and lysozyme) with a gradient to 0.45 M.
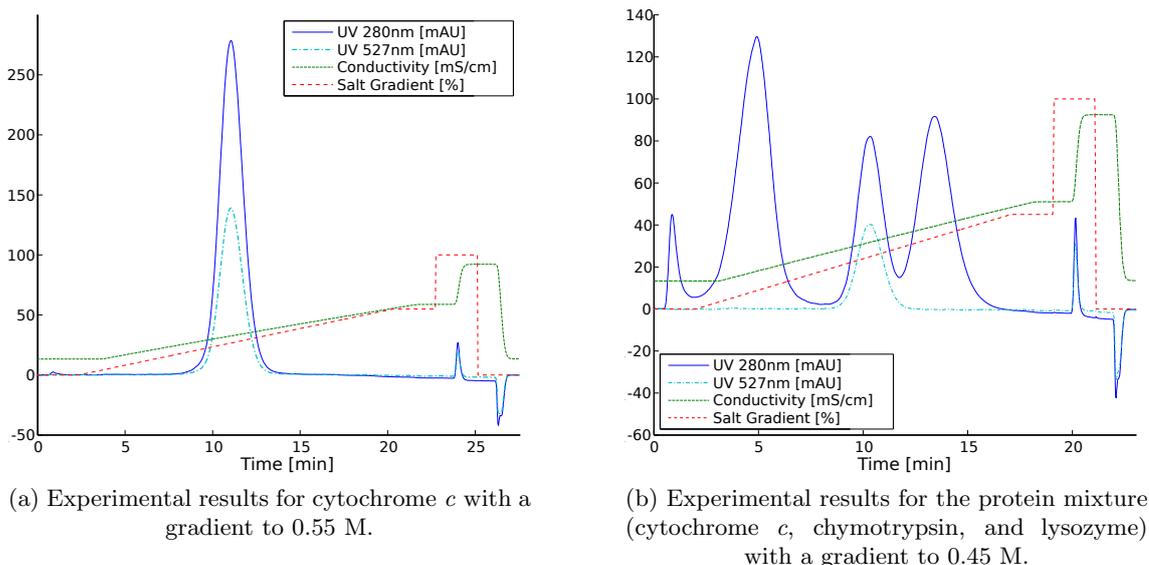
Figure 4.4: Example results of calibration experiments.

Newton-Simpson scheme with constant step size $\alpha = \frac{0.3}{\|\Delta p\|}$ has been used. With increasing residual error, the step size was lowered to 50 % of its original value. The results in Figure 4.6 show differences in agreement between simulation and measurement for both algorithms, starting from $\{k_{eq}, k_d, \nu\} = \{1, 10, 3\}$. When comparing residual errors, the adjoint method shows a slightly better result: $4.17 \cdot 10^{-10}$ compared to the finite difference residual error of $1.04 \cdot 10^{-9}$. The set of parameters differs accordingly, as shown in Table 4.2. The finite difference result matched the peak heights of the last three experiments very well, but failed to match the first experiment with the same quality. In contrast to this, the adjoint result was very good in the first experiment, but the peak height was slightly off for the other chromatograms.

The fact that both parameter sets show a good agreement for the majority of the experiments indicates that their objective function contributions do not change significantly for small perturbations in the model parameters. This is confirmed by the long drawn-out minimum in Fig. 4.3 and the fact that MATLAB's `DiffMinChange` parameter, which controls the perturbation Δp, had to be increased to at least 0.01 to achieve any convergence towards the minimum. For final model calibration, additional experiments have to be included to increase the certainty of the estimate by minimizing the associated confidence intervals.

The adjoint method reaches the minimum after 15 iterations (15 forward and 15 backward evaluations), as shown in Figure 4.7. Afterwards, the gradient is very small and the step width has to be decreased strongly to find another improvement. At the same time, the error of `lsqnonlin` is ten times larger and after 15 iterations (60 forward evaluations) still four times as large as the Newton/adjoint method. The step size determination of `lsqnonlin` obviously is challenged by the small derivatives and unable to lower ν in favor of a larger k_{eq}.

The fact that both algorithms take slightly different paths towards the optimum is due to the parameter derivatives being partially dependent on the other coefficients. Furthermore, the divergence after the first iteration suggests that the parameter derivatives are
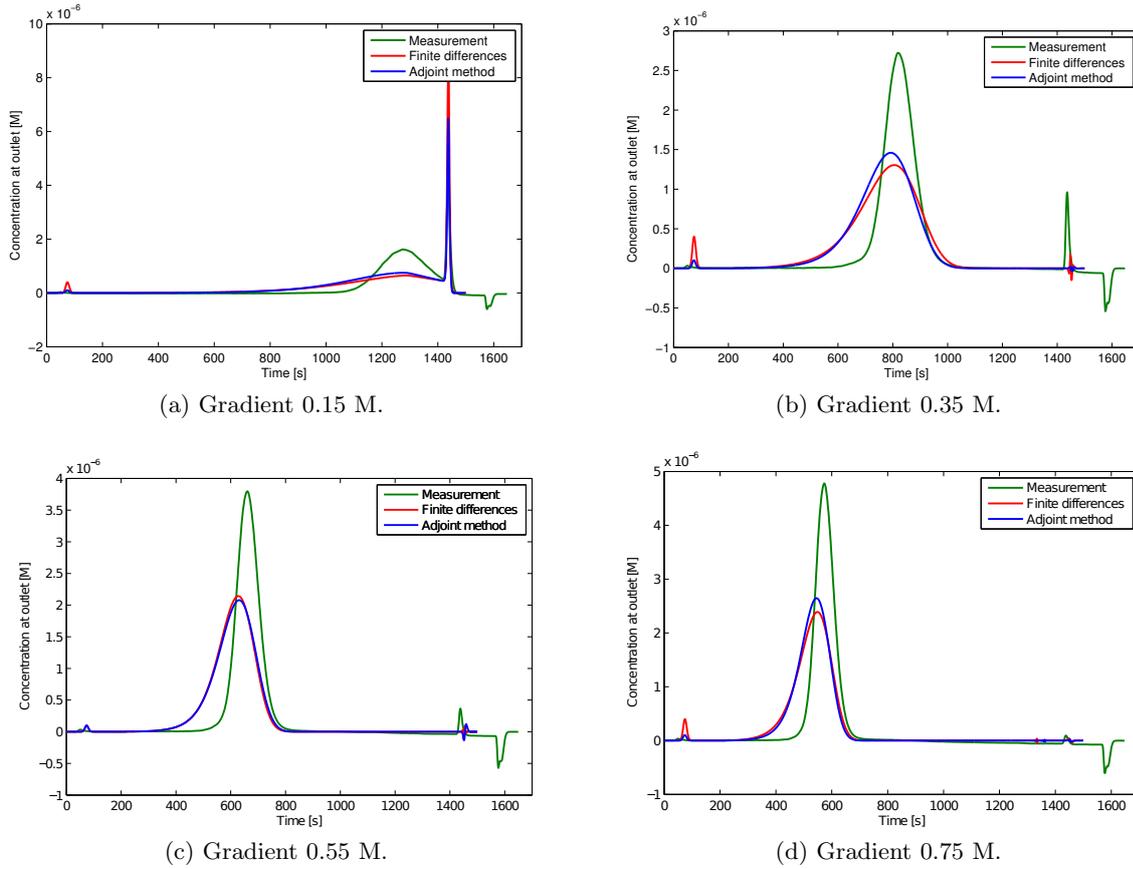
(a) Gradient 0.15 M.

(b) Gradient 0.35 M.

(c) Gradient 0.55 M.

(d) Gradient 0.75 M.

Figure 4.5: Parameter estimation for cytochrome c using SMA formulation (4.16)-(4.18).

Table 4.2: Parameter estimation results for cytochrome c.

Parameter	lsqnonlin	Adjoint method
k_{eq}	4.259	6.072
k_d	22.73	8.382
ν	3.628	2.667
σ	30	30

sufficiently non-linear, such that the secant with $\Delta p \geq 0.01$ differs distinctly from the adjoint derivative. Curvature information from the second derivative might have been beneficial.

A forward and backward solution at $k_{eq} = 4.56$, $k_d = 1.52$ and $\nu = 3.38$ of the 0.35 mM NaCl gradient experiment is plotted in Figure 4.8. The forward solution of the interstitial concentration of cytochrome c shows the injection at the inlet from time $t = 0$ and the development of the mobile phase concentration until reaching the outlet. The dual solution shows the current difference between simulation and measurement at the outlet and follows its origin back in time. Its magnitude has no physical interpretation.

For the ternary mixture, the Transport Dispersive Model, Eq. (4.11), was used, which includes an additional effective mass transfer coefficient k_{eff}. It only occurs in the linear blocks, such that the derivatives of the model with respect to this parameter are easy to calculate. In order to estimate all four parameters (σ left out again) for each of the

(a) Gradient 0.15 M.   (b) Gradient 0.35 M.

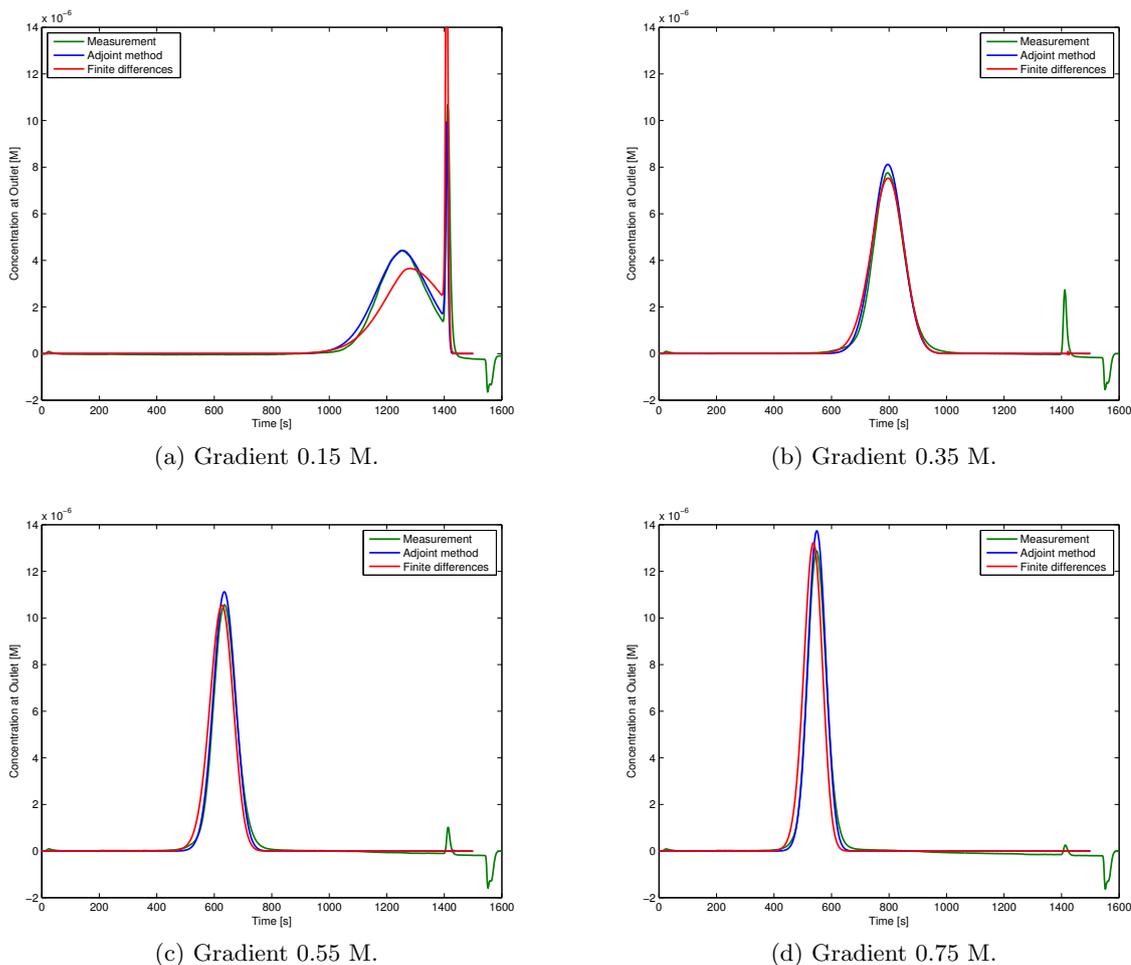(c) Gradient 0.55 M.   (d) Gradient 0.75 M.

Figure 4.6: Parameter estimation results for cytochrome c using SMA formulation (4.45).

three proteins, the individual points of the chromatogram had to be taken as separate measurements by leaving out the time integration. Otherwise, the equation system for calculating the update would be under-determined and Newton's method could not be applied. This affects the results such that instead of the sum of least squares, the maximum single error plays a more important role, in this case, the peak of lysozyme in the first experiment at 20 ml (Fig. 4.9a).

The estimated parameters are given in Table 4.3. All proteins started with $k_{eff} = 0.03$, $k_d = 10$, and $\nu = 3$. To distribute the initial peaks over time, the equilibrium parameter was chosen to be 1 for chymotrypsin, 5 for cytochrome c, and 10 for lysozyme. The plots in Figure 4.9 show that curves fit well in this setting. Although chymotrypsin would have been better modeled as two independent components because of the impurities, the single-component result also fits the measurement well enough to demonstrate the effectiveness of the method. Comparing the cytochrome c parameters to the single protein case, the kinetic parameter k_d increases to compensate the diffusive effects introduced by k_{eff}. The other two are still in the order of the former estimate and the overall fit is the best of the three proteins. Lysozyme, like chymotrypsin, shows small deviations in the peak shape, but the high peak at 20 ml in the first experiment is recovered perfectly.
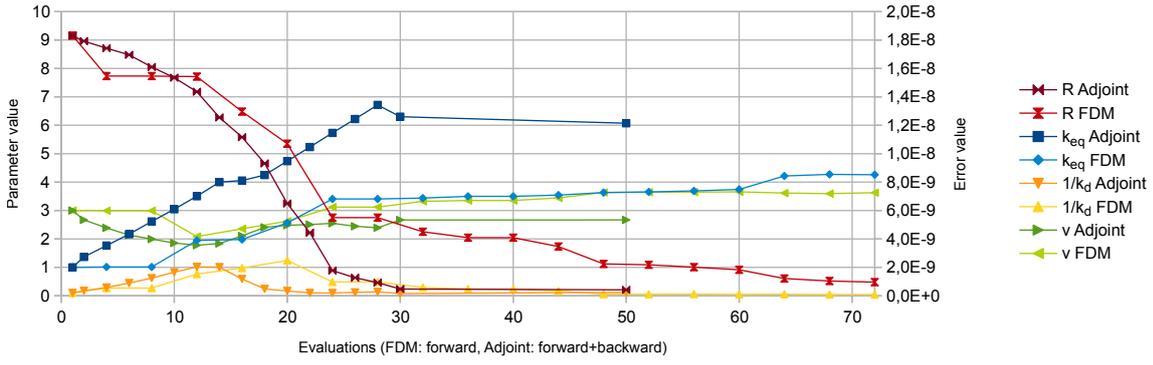
Figure 4.7: Comparison of the estimation procedure for cytochrome c using adapted SMA formulation and gradients computed by the adjoint and finite difference method (FDM). $\frac{1}{k_d}$ is plotted as this term occurs naturally in Eq. (4.45) and was used for estimation.
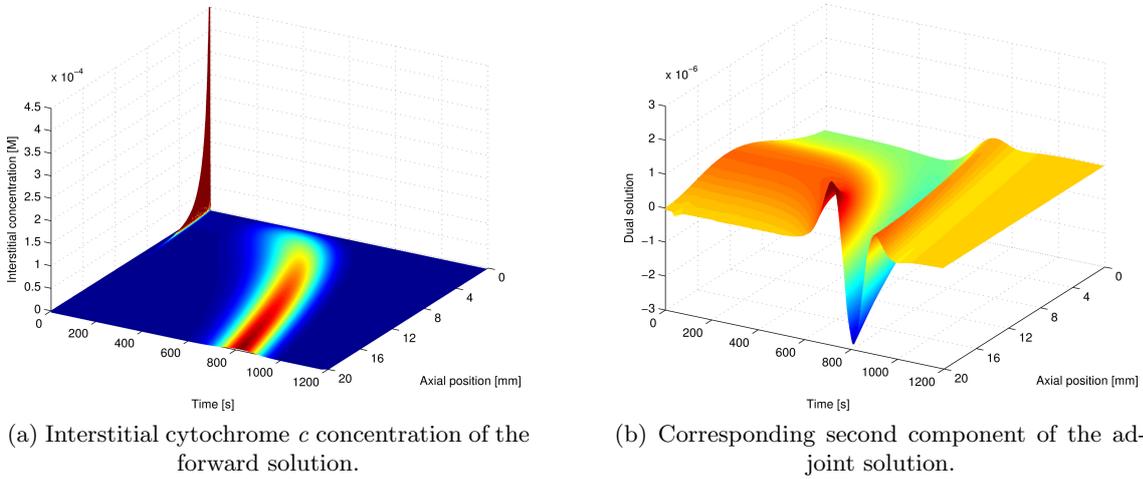

(a) Interstitial cytochrome c concentration of the forward solution.

(b) Corresponding second component of the adjoint solution.

Figure 4.8: Intermediate solution of the cytochrome c parameter estimation.

4.4.3   Optimization

By changing the objective function, the same method can be applied for optimization. To separate the peaks of the three proteins, a possible objective function is defined by minimizing the mutual overlap:

$$R\left(p\right) = \int_0^T r\left(t; p\right) \, \mathrm{d}t = \int_0^T \sum_{i,j \geq 1, i \neq j} c_i(L, t; p) \cdot c_j(L, t; p) \, \mathrm{d}t. \tag{4.46}$$

In this case, the gradients needed for the adjoint problem are given by

$$\frac{\partial r(t; p)}{\partial c_i} = \sum_{i \neq j \geq 1} c_j(L, t; p). \tag{4.47}$$

For simplicity, assuming the previously used gradient is to be replaced by a step elution, the derivative with respect to the induced salt concentration is non-zero for the CDR equation of salt only, Eq. (4.11). More specifically, it adds only to the first entry of the
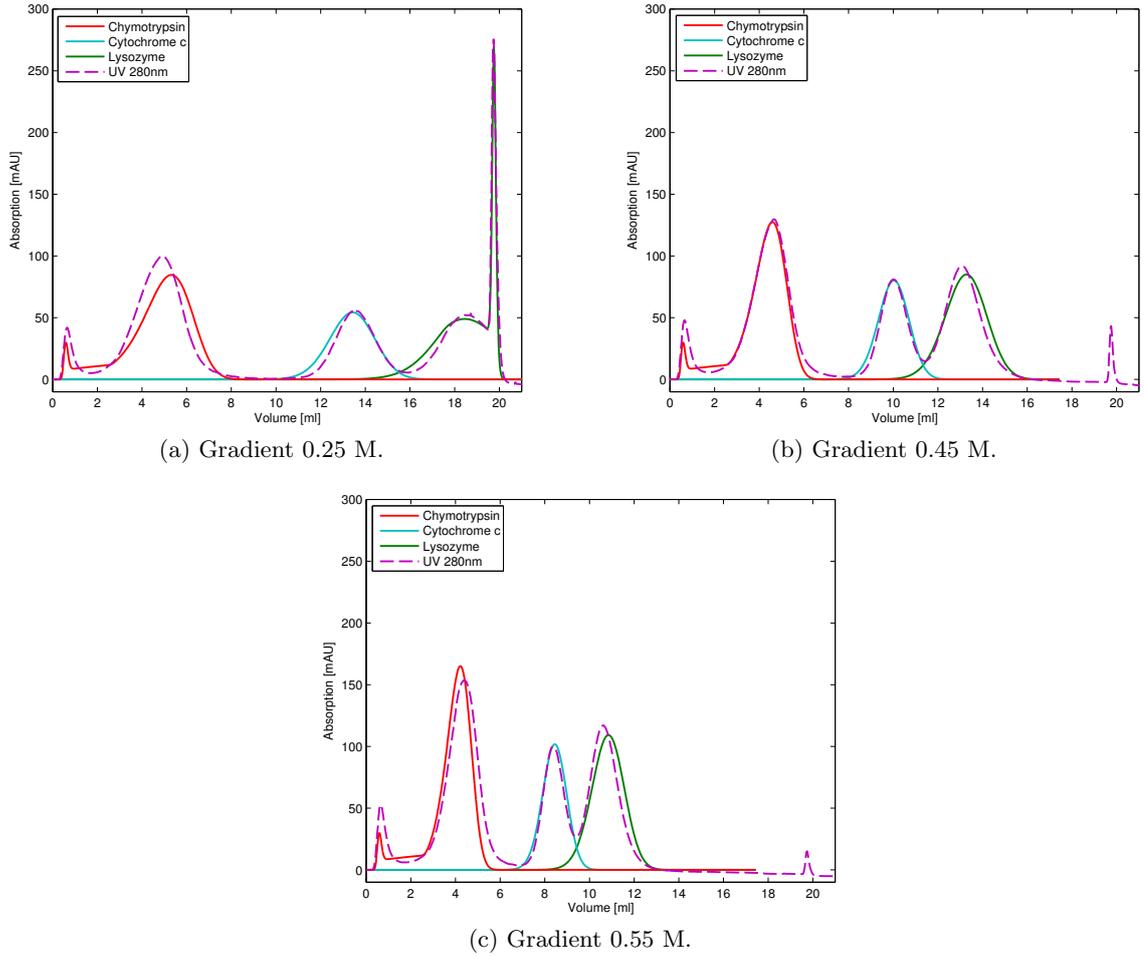
(a) Gradient 0.25 M.

(b) Gradient 0.45 M.

(c) Gradient 0.55 M.

Figure 4.9: Parameter estimation results for the protein mixture.

right-hand side vector in Fig. 4.1:

$$\frac{\partial F_{1,1_{Salt}}}{\partial c_{in,Salt}} = \begin{cases} -u & \text{for } x = 0, \\ 0 & \text{else.} \end{cases} \tag{4.48}$$

When optimizing the step concentration, the derivative only is non-zero in the time interval of the step. Initiating the step 2 CV after injection and keeping the level for another 2 CV before switching to the 1 M high-salt buffer, the optimal step concentration is found to be 0.31 M. The resulting chromatogram is plotted in Fig. 4.10a. Starting from the initial guess of 0.1 M, the iteration converged after three steps. With an initial guess of 0.5 M, thirteen iterations were needed as visualized in Fig. 4.10b.

The programming effort to adapt the estimation method to optimization is low. The forward and backward solvers do not change and the Newton-Simpson optimizer can be re-used. Only the constant right-hand side of the adjoint problem differs because of the new objective. The parameter gradient with respect to the inlet concentration has to be defined as shown above.

Table 4.3: Parameter estimation results for the protein mixture.

Parameter	Chymotrypsin	Cytochrome c	Lysozyme
k_{eff}	0.019	0.030	0.025
k_{eq}	0.113	6.920	19.56
k_d	71.43	22.73	8.850
ν	3.1681	2.467	3.107
σ	30	30	30

(a) Optimized step 0.31 M.
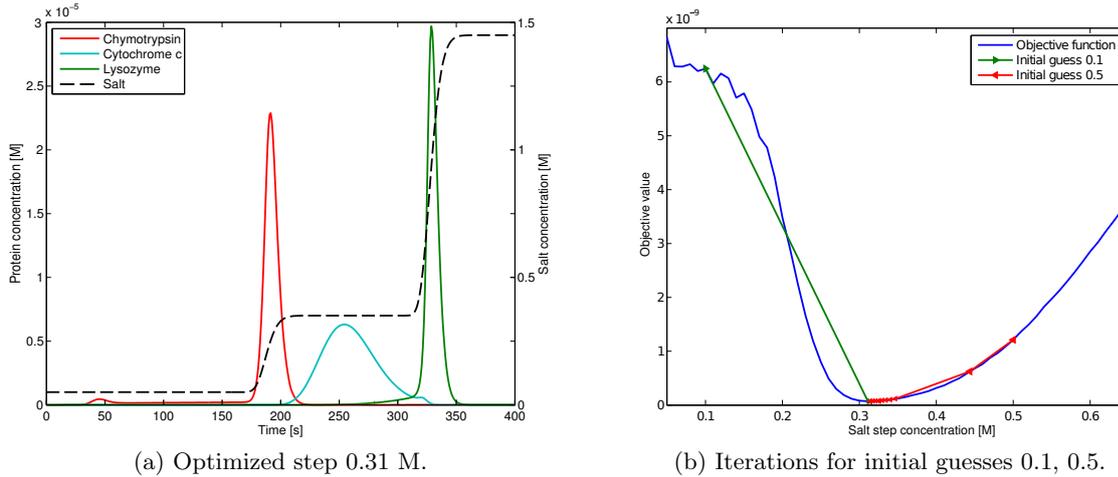

(b) Iterations for initial guesses 0.1, 0.5.

Figure 4.10: Optimization results for the protein mixture.

4.4.4 Algorithm Performance

For the first fifteen forward simulations of the cytochrome c case, the average solution time in MATLAB was 3.46 s. The backward problem was slightly slower with an average of 3.92 seconds, which is due to the higher amount of memory operations during assembly. Because of the short computation time, the linearity is not beneficial. Integrating the final derivatives took 0.28 s per derivative, resulting in a total speed-up of 1.7. Using vectorization and customized solvers for the forward and backward problem an additional speed-up is to be expected.

For the protein mixture, an average forward solution took 18.1 s, a backward solution 8.6 s, and assembling all derivatives 2.91 s. Solving 13 forward problems for determining the 12 derivatives needed in this case takes approximately 7.95 times longer when using the adjoint method.

An overview of the runtime measurements is given in Table 4.4.

4.5 Conclusions and Outlook

The high demand for effective and inexpensive production procedures fosters the need for fast model-based optimization methods. Model-specific parameters can be determined conveniently from experiments using an inverse method that alters the model parameters

Table 4.4: Average runtime per estimation iteration. The FDM solves 4 forward problems for cytochrome c and 13 forward problems for the mixture, respectively.

Routine	Cytochrome c 3 derivatives	Ternary mixture 12 derivatives
Adjoint forward	3.46 s	18.1 s
Adjoint backward	3.92 s	8.6 s
Adjoint integration	0.28 s	2.9 s
Adjoint total	7.66 s	29.6 s
FDM total	13.05 s	235.2 s
Speed up	1.70	7.95

in an iterative fashion to match model prediction and measurements. Within this work the adjoint method presented for the calculation of derivatives with respect to parameters in differential equations in [27] was applied in the setting of liquid chromatography modeling. The focus was placed on demonstrating that adjoint sensitivities can be calculated with reasonable effort by re-using existing codes to a large extent. An example implementation demonstrated that the adjoint method is far more efficient than finite difference approximations of the gradients and sufficiently consistent in combination with well-formulated model equations. For the widely used SMA model, a parameter transformation helped to avoid saddle points. In total, a speed-up of 8 could be achieved for estimating SMA parameters of a ternary mixture.

The costs are reduced to transposing the block structure and sub-blocks of already existing matrices and calculating only additional derivatives in the form of diagonal matrices. Exchanging the objective function to conduct optimization, estimation or even optimal experimental design only affects the constant part on the right-hand side. A more robust solver with a wider convergence range might be a useful supplement when dealing with stronger non-linearities in retention models and should be subject of further research.

If no simple transformation is known, a preconditioned optimization algorithm could be employed or second-order derivatives could be computed via the adjoint method. Both approaches should be considered in the future.

In the context of the Quality by Design strategy of the Food and Drug Administration (FDA) [86] guiding the development of new biopharmaceutical processes, the computation of parameter sensitivities allows for additional risk assessment. The most critical process parameters can be identified from the parameter sensitivities under the chosen process conditions. For example, a low sensitivity to contaminant concentrations in the sample indicates a high tolerance to batch variances. To study all parameters at once, a high number of sensitivities have to be computed. This application should be studied in the future.

Acknowledgments

5 | Optimal Experimental Design for the Determination of Isotherm Parameters of Glucose Oxidase using Mixed Mode Column Chromatography

Tobias Hahn[1], Gang Wang[1], Thiemo Huuk[1], Vincent Heuveline[2], Jürgen Hubbuch[1]

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

[2] Heidelberg University, Interdisciplinary Center for Scientific Computing, Engineering Mathematics and Computing Lab, Heidelberg, Germany

Abstract

Nowadays' models in chromatography include many parameters that are hard to measure experimentally, e.g. thermodynamic activity coefficients in models for hydrophobic interaction and mixed mode chromatography. In practical applications, it is hard to predict, how many and which experiments to conduct for model calibration.

Optimal Experimental Design (OED) first determines the experimental set-up with highest probable information content. Using this approach, the model parameters with the highest statistical quality are found using the least number of experiments. This is especially valuable when low sample volume is available for screening.

We present OED for column chromatography of proteins including the computation of approximate covariances and confidence intervals. A case study with the model protein glucose oxidase investigates identifiability of adsorption isotherm parameters on three resins by GE Healthcare: Capto™ Q, Phenyl, and adhere. An extended General Rate Model was developed to account for salt-dependent diffusion phenomena.

The results underline the improved reliability using OED at potentially much less time and material consumption compared to manual experimental design. From the covariance matrices it is further possible to gain insight into the models' ability to describe the protein's sorption behavior for the various resins.

5.1   Introduction

Mixed mode resins offer multiple binding interactions and are increasingly used in the purification of antibodies [69; 195], oligosaccharides [139] and many other molecules. By adjusting the operational conditions the binding modes can be reduced to a single interaction [83; 102] or used together to effect the separation. The mixed mode chromatography (MMC) ligands often include features of ion-exchange chromatography (IEC) and of hydrophobic interaction chromatography (HIC) ligands. They are highly suitable for application in pH-tunable hydrophobic [95] or high salt conditions [90], and for capturing specific proteins under physiological conditions [70].

As for other chromatographic methods, the fundamentals of mixed mode chromatography can be understood and improved by creating reliable models [33]. The Steric Mass Action model(SMA) [24] was employed successfully for mixed mode chromatography [97]. Mollerup's thermodynamic framework [131; 132] increased the understanding of IEC as well as HIC [134] and the optimisation of purification processes [133]. Based on this, Nfor *et al.* developed a model for protein adsorption in IEC/HIC mixed mode chromatography [142]. These models include many parameters, which are hard to measure experimentally, for example thermodynamic activity or interaction coefficients. Especially in multi component settings, it is hard to predict how many experiments to conduct for model calibration and which ones will have the highest information content. Conventional Design of Experiments (DoE) approaches will most probably include unnecessary experiments that do not add certainty to the parameter estimates.

As mathematical models are available that describe the dynamics of the system, sensitivity analysis allows to find optimal expriments for model calibration. While the objective of parameter estimation is to find model parameters which minimize the disagreement of simulation and measurement, Optimal Experimental Design (OED) [9] takes one step backward and first determines the experimental set-up with highest probable information content. Here, the objective is to minimize the uncertainty of the parameters measured via the parameter covariance matrix. Using this approach, the model parameters with the highest statistical quality are found using the least number of experiments. This is especially valuable when low sample volume is available for screening.

The technique was described for nonlinear ordinary differential equations by Schlöder *et al.* and applied to the reaction of urethane [11]. The same methodology was used by Arellano-Garcia for SO_2 oxidation [7]. A fed-batch reactor and the Lotka Volterra fishing problem were approached with OED by Tenen *et al.* [180]. In the context of chromatography, Wozny *et al.* [10] investigated the optimal determination of SMA parameters for β-lactoglobulin using static batch experiments

Here, we present OED for column chromatography of proteins, including the computation of approximate covariances and confidence intervals. A case study compares parameter estimation with manually chosen experimental conditions and OED for the model protein glucose oxidase on IEC, HIC and MMC resins to gain a deeper understanding of the differences in sorption behavior. With Capto™ adhere, GE Healthcare provides a mixed-mode adsorbent that chemically combines the anion-exchange adsorbent Capto™ Q, and the hydrophobic interaction adsorbent Capto™ Phenyl and is hence chosen for this task.

5.2 Theory

5.2.1 Adsorption Models

The employed models for ion-exchange, hydrophobic interaction, and mixed mode chromatography are structurally very similar and presented in the following sections

5.2.1.1 Ion-exchange Chromatography

A protein molecule unit P in solution is assumed to bind to ν ligands L, exchanging ν units of salt counter-ions SL:

$$P + \nu SL \rightleftharpoons PL + \nu S. \tag{5.1}$$

The Steric Mass Action model [24] considers that the number of available ligands is further reduced by steric shielding effects that are mostly caused by protein-protein repulsion [110]. The multi-component equilibrium formulation of the mobile phase concentrations c and stationary phase concentrations q of m proteins and salt is given by

$$\frac{q_i}{c_i} \;=\; k_{eq,i} \left(\frac{\overline{q}_{salt}}{c_{salt}} \right)^{\nu_i} \quad i = 1, \ldots, m, \tag{5.2}$$

where $\overline{q}_{salt}$ is the concentration of available ligands, given by

$$\overline{q}_{salt} \;=\; \Lambda_{IEC} - \sum_{i=1}^{m} \left(\nu_i + \sigma_i \right) q_i. \tag{5.3}$$

Here, Λ_{IEC} is the total concentration of binding sites in the stationary phase and σ_i accounts for the steric shielding of protein i.

The concentration of counter-ions in the stationary phase is analogously given by

$$q_{salt} \;=\; \Lambda_{IEC} - \sum_{i=1}^{m} \nu_i q_i. \tag{5.4}$$

Jackobsson *et al.* [87] introduced binding kinetic rate constants k_{ads} and k_{des} with $k_{eq} = k_{ads}/k_{des}$, here given in the formulation with $k_{kin} = k_{des}^{-1}$:

$$k_{kin,i} \frac{\partial q_i}{\partial t} \;=\; k_{eq,i} \overline{q}_{salt}^{\nu_i} c_i - c_{salt}^{\nu} q_i \quad i = 1, \ldots, n. \tag{5.5}$$

5.2.1.2 Hydrophobic Interaction Chromatography

Mollerup established a model based on fundamental chemical thermodynamics [132]. In contrast to IEC, a protein molecule unit P in solution is assumed to bind to n ligands L, forming a protein ligand complex:

$$P + nL \rightleftharpoons PL_n. \tag{5.6}$$

Here, the equilibrium formulation is written in terms of the activity coefficients of the species

$$
\frac{q_i}{c_i} \;=\; k_{eq,i} \left(\frac{\overline{q}_L}{c_v}\right)^{n_i} \frac{\gamma_P \gamma_L}{\gamma_{PN_n}} \quad i = 1, \ldots, m, \tag{5.7}
$$

where γ_L and γ_{PN_n} are assumed to be unity and the molarity of the solution in the pore volume c_v is constant. The protein solute activity coefficient is modeled as

$$
\gamma_{P,i} \;=\; \exp\left(K_{s,i} c_{salt} + K_{p,i} c_i\right) \gamma_{P,i}^{\infty,w} \quad i = 1, \ldots, m, \tag{5.8}
$$

where K_s and K_p are constant interaction parameters and $\gamma_P^{\infty,w}$ is the activity coefficient at infinite dilution.

Similarly to SMA, the concentration of available ligands $\overline{q}_L$, given by

$$
\overline{q}_L \;=\; \Lambda_{HIC} - \sum_{j=1}^{m} \left(n_j + s_j\right) q_j, \tag{5.9}
$$

with a steric shielding coefficient s.

In the classical form [132], $\Lambda_{HIC}^{n_i}$ is lumped into k_{eq} such that the loading is described with the help of the single-component saturation concentrations $q_{max,i} = \Lambda_{HIC}/(n_i + s_i)$:

$$
\overline{q}_L \;=\; \Lambda_{HIC} \overline{q}'_L = \Lambda_{HIC} \left(1 - \sum_{i=1}^{m} \frac{q_i}{q_{max,i}}\right) \quad i = 1, \ldots, m, \tag{5.10}
$$

In the kinetic formulation, c_v and $\gamma_P^{\infty,w}$ are lumped in the equilibrium constant:

$$
k_{kin,i} \frac{\partial q_i}{\partial t} \;=\; k_{eq,i} \overline{q}_{L,i}'^{n_i} \exp\left(K_{s,i} c_{salt} + K_{p,i} c_i\right) c_i - q_i. \tag{5.11}
$$

5.2.1.3 Mixed Mode Chromatography

Nfor *et al.* developed a model for protein adsorption in IEC/HIC mixed mode chromatography by assuming both adsorption modes to happen at the same time [142]:

$$
P + \nu SL + nL \Longleftrightarrow PL_n + \nu S. \tag{5.12}
$$

The equilibrium formulation was derived to be

$$
\frac{q_i}{c_i} \;=\; k_{eq,i} \left(\frac{\overline{q}_{salt}}{c_{salt}}\right)^{\nu_i} \left(\frac{\overline{q}_L}{c_v}\right)^{n_i} \gamma_p \tag{5.13}
$$

with the additional counter-ion balance known from SMA, Eq. (5.4).

The kinetic form is likewise:

$$
\begin{aligned}
k_{kin,i} \frac{\partial q_i}{\partial t} \;=\;& k_{eq,i} \overline{q}_{salt}^{\nu_i} \overline{q}_L^{n_i} \exp\left(K_{s,i} c_{salt} + K_{p,i} c_i\right) c_i \\
& - c_{salt}^{\nu_i} q_i.
\end{aligned} \tag{5.14}
$$

The similarity to Eqs. (5.5) and (5.11) is apparent. When switching hydrophobic effects off, i.e. $n = K_s = K_p = 0$, Eq. (5.14) reduces to Eq. (5.5). For $\nu = 0$, we recover Eq. (5.11).

5.2.2 Column Model

The General Rate Model (GRM) [72] is employed to describe the macroscopic protein transport through the column. The systems are of Convection Diffusion Reaction (CDR) type. Eq. (5.15) describes the rate of change of a concentration $c_i(x,t)$ in the interstitial phase of a column with length L to consist of convective mass transport in space with the average interstitial velocity of the fluid u, peak broadening effects that are modeled as dispersion in axial direction with respect to a coefficient D_{ax}, and transition from the interstitial concentration into the particle pore concentration $c_{p,i}$ which depends on the porosity of the bed ε_b, the radius of adsorber particles r_p and a component-specific film transfer coefficient $k_{film,i}$. The model is complemented with Danckwerts boundary conditions, Eqs. (5.15),(5.17).

$$\frac{\partial c_i}{\partial t} = -u(t)\frac{\partial c_i}{\partial x} + D_{ax}\frac{\partial^2 c_i}{\partial x^2} - \frac{1-\varepsilon_b}{\varepsilon_b}k_{film,i}\frac{3}{r_p}\left(c_i - c_{p,i}\right) \tag{5.15}$$

$$\frac{\partial c_i}{\partial x}(0,t) = \frac{u(t)}{D_{ax}}\left(c_i(0,t) - c_{in,i}(t)\right) \tag{5.16}$$

$$\frac{\partial c_i}{\partial x}(L,t) = 0 \tag{5.17}$$

The GRM, Eq. (5.18), introduces a radial dimension $r \in [0, r_p]$ for the particles and a component-specific pore diffusion coefficient D_p to model diffusion-driven mass transfer in the pore system.

$$\frac{\partial c_{p,i}}{\partial t} = \begin{cases} \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 D_{p,i}\frac{\partial c_{p,i}}{\partial r}\right) - \frac{1-\varepsilon_p}{\varepsilon_p}\frac{\partial q_i}{\partial t} & \text{for } r \in (0, r_p) \\ \frac{k_{film,i}}{\varepsilon_p D_{p,i}}\left(c_i - c_{p,i}\right) & \text{for } r = r_p, \\ 0 & \text{for } r = 0. \end{cases} \tag{5.18}$$

The high salt concentrations used in the experiments influence the kinetic effects. It was impossible to find $D_{p,i}$ values that lead to sufficiently well fitting simulated chromatograms for experiments with salt concentrations above 1 M (data not shown). Similarly, a strong increase of the $HETP$ value (height-equivalent of a theoretical plate) starting at 1 M salt was reported for HIC resins [135] that could be modeled with an exponential function of the form

$$HETP = a + b\exp(c \cdot c_{salt}). \tag{5.19}$$

We also observed different HETP values for pulse injections of small tracers dissolved in low and high salt buffer, such that the adsorption related terms in the HETP cannot be the cause of the effect. The most influential term remaining in the HETP value for a GRM is pore diffusivity [72]. Assuming the other contributors to be constant, the relation is given by $HETP = c_1 + c_2 \cdot D_{p,i}^{-1}$ with two constants c_1, c_2 [72]. Combining this with the HETP correlation of [135] and rearranging, we formulate the salt concentration-dependent pore diffusion coefficient with three constants D_{p0}, D_{p1}, D_{p2} as

$$D_p(c_{salt}) = D_{p0}\frac{1 + D_{p1}}{1 + D_{p1} \cdot \exp\left(D_{p2} \cdot c_{salt}\right)}. \tag{5.20}$$

The formulation in this paper is chosen such that $D_p(0) = D_{p0}$ and values of different species can be compared easily. For parameter estimation, we work the numerically favorable equation

$$D_p(c_{salt}) \quad = \quad \frac{D'_{p0}}{1 + D_{p1} \cdot \exp(D_{p2} \cdot c_{salt})} \tag{5.21}$$

with $D'_{p0} = D_{p0}(1 + D_{p1})$ in order to avoid rounding errors in the division.

5.2.3 Model Calibration

Estimation of an unknown parameter set $\hat{\theta}$ solves the least squares optimization problem

$$\hat{\theta} \quad = \quad \arg\min_{\theta} \sum_{k=1}^{N_{exp}} \sum_j \left(\frac{c_{meas,k}(t_j) - c_{sim,k}(t_j; \theta)}{\sigma_k} \right)^2, \tag{5.22}$$

where $c_{meas,k}(t_j)$ and $c_{sim,k}(t_j)$ are the measured and simulated sum signals of experiment k at the column outlet at point in time t_j. σ_k is the variance of the measurement error of the respective experiment. Assuming the noise of the parameter estimates to be zero-mean Gaussian and isotropic, solving the least squares problem is identical to Maximum Likelihood estimation.

5.2.4 Statistical Analysis

As a measure of parameter certainty, an estimate of the covariance matrix is used. The diagonal elements of the covariance matrix represent the variance of the estimates. Obtaining the exact covariance matrix is a complex mathematical problem that leads most studies to use a Jacobian matrix [11] or a Fisher information matrix (FIM) instead [183; 198]. The asymptotic variance of the maximum likelihood estimator is given by the Cramer-Rao lower bound [38; 63], the reciprocal of the Fisher information:

$$Cov(\theta) \quad \geq \quad FIM(\theta)^{-1} = \left(\sum_{k=1}^{N_{exp}} FIM_k(\theta) \right)^{-1}. \tag{5.23}$$

with the Fisher information matrix of each experiment k being defined with the parameter sensitivities of a Function $F_k(t)$ [160] :

$$FIM_k(\theta) \quad = \quad \frac{1}{\sigma_k{}^2} \int_T \frac{\partial F_k}{\partial \theta_j} \frac{\partial F_k}{\partial \theta_j}^T dt \tag{5.24}$$

In our case, two function F are of interest, the model $F_k(t) = c_{sim,k}(t)$ and the least squares error $F_k = (c_{meas,k}(t) - c_{sim,k}(t))^2$.

5.2.4.1 Confidence Intervals

The confidence intervals contain the true parameter of interest with an a prioi defined probability. If not stated otherwise, confidence intervals with a probability of 95 % are

calculated. Given an estimate of the covariance matrix created form $F_k = (c_{meas,k}(t) - c_{sim,k}(t))^2$, a confidence intervals for θ_j is given by

$$\theta_j \quad \pm \quad c\sqrt{diag\left(Cov\left(\theta\right)\right)} \tag{5.25}$$

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

5.2.4.2   Optimal Experimental Design

The purpose of optimal experimental design is to identify the process set-ups that facilitate parameter estimation. Hence, the parameter sensitivity of the model is to be maximized. Here, we use $F_k(t) = c_{sim,k}(t)$ to determine the parameter covariance matrix of the designed experiments and aim to minimize it [38; 63].

For this purpose, a scalar function of the covariance matrix has to be defined. A discussion of proposed functions is given in [180]. In the following, the D-criterion is used that minimizes the determinant of the covariance matrix and thus the volume of the confidence ellipsoid.
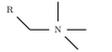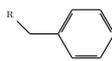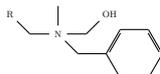
$$D(Cov) \quad = \quad \ln\det\left(Cov\right) \tag{5.26}$$

5.3   Materials and Methods

5.3.1   Chromatographic Instrumentation

The chromatographic experiments were carried out using an ÄKTApurifier 10 fast protein liquid chromatography (FPLC) system equipped with pump P-903, UV (10mm path length), conductivity and pH monitor UPC-900, an autosampler A-905 and a fraction collector Frac-950 (all GE Healthcare, Little Chalfont, Buckinghamshire, UK). The instrument was controlled with UNICORN 5.31 (GE Healthcare, Little Chalfont, Buckinghamshire, UK).

Table 5.1: Overview of chromatography resins.

Capto™	Ligand	Structure	Binding
Q	Quaternary amine		ionic
Phenyl	Phenyl		hydrophobic
adhere	N-benzyl-n-methyl ethanolamine		ionic, hydrophobic, hydrogen bonds

5.3.2   Adsorbers, Buffers, and Protein

All resins were acquired pre-packed by ATOLL as 1 ml miniChrom columns with dimension $5\,cm \times 0.2\,cm^2$. The three different resins used are shown in Tab. 5.1. Capto™ Q is a strong

anion exchange resin and Capto™ Phenyl is a hydrophobic interaction media. They are used for capture or intermediate purification steps of proteins. Capto™ adhere combines the features of both Capto™ Q and Capto™ Phenyl. All columns were stored in 20% ethanol at 4 °C.

20 mM 1-methylpiperazine buffers (Sigma-Aldrich, St. Louis, MO, USA) with 0 and 4 M NaCl (Merck, Darmstadt, Germany) were used at pH 4.5 for all experiments. The different salt profiles were mixed from these two buffers. To regenerate the single mode columns, the concentration of sodium chloride was increased to 4 M in case of Capto™ Q and decreased to 0 M in case of Capto™ Phenyl. Because of the hydrophobic and ionic features of Capto™ adhere, the pH was reduced to 3 in a regeneration step by applying 50 mM NaH_2PO_4 (VWR, Darmstadt, Germany) to induce the charge repulsion effect. For cleaning-in-place 1 M NaOH (Merck, Darmstadt, Germany) was used. All buffers were $0.22 \mu m$-filtrated and degassed by sonification.

Glucose Oxidase (from Aspergillus niger, no. G7141, Sigma-Aldrich, St. Louis, MO, USA) was used as model protein. The protein was prepared using the respective running buffer and $0.22 \mu m$-filtrated prior to usage.

5.3.3   System Characterization

To determine the system characteristics, experiments with tracer substances were carried out. The flow rate was kept at 0.2 ml/min. The UV signal at 300 nm and conductivity signal were recorded.

The FPLC dead volume was determined by tracer injections without a column attached to the system. All other data were corrected with respect to this dead volume. The column void volume was calculated from injections of a pore-penetrating, non-interacting (acetone) and a non-pore-penetrating, non-interacting tracer (dextrane 2000 kDa). The total ionic capacity was determined by acid-base titration.

5.3.4   Initial Model Calibration

Initial model calibration was performed with manually designed experiments. Thereafter, optimal experimental design was repeated until the the D-criterion value did not change significantly. If not stated otherwise, the sample concentration and volume were kept constant at 0.06 mM and 0.5 ml for low salt binding, 0.02 mM and 1 ml for high salt binding, because of solubility constraints.

In case of Capto™ Q, three bind-elute runs were performed: two gradient and one step elution. The gradients were ran from 0 M to 0.20 M NaCl, and 0 M to 4 M NaCl; the step height was set to 0.1 M. The gradients and steps were initiated after 6 ml and the length of the gradients was chosen to be 20 ml.

For Capto™ Phenyl, three step elutions with varying binding buffer salt concentrations were performed: 4 M, 2 M, and 1 M NaCl. All steps went down to 0 M NaCl and were initiated after 6, 6, and 25 ml.

In case of Capto™ adhere, three experiments with low-salt binding and three with high-salt binding were performed, two steps and one gradient in each case. The steps had final concentrations of 1 and 2 M NaCl and were again initiated after 6 ml. As before, the gradients started at 6 ml, had a length of 20 ml and were ran from 0 to 2 M, and 4 to 0 M NaCl.

Table 5.2: Measured (top) and calculated (bottom) column parameters.

Parameter	Symbol	Unit	Proceeding	Capto™ Q	Capto™ Phenyl	Capto™ adhere
Length	L	mm	From manufacturer	50	50	50
Volume	V	ml	From manufacturer	1	1	1
Bead radius	r_p	mm	From manufacturer	0.045	0.0375	0.0375
System dead volume	V_d	ml	Acetone pulse without column	0.078	0.078	0.078
Retention volume acetone	V_{RetAc}	ml	Acetone pulse with column	0.9	1.07	1.02
Retention volume dextran	V_{RetDex}	ml	Dextran pulse with column	0.42	0.5	0.47
HETP	$HETP$	mm	UNICORN peak integration	0.366	0.283	0.182
Volume of HCl	V_{HCl}	ml	Acid/base titration	24.2		13.28
Molarity of HCl	c_{HCl}	M	Manually controlled	0.01		0.01
Flow rate	u	$\frac{mm}{s}$	Manually controlled	0.167	0.167	0.167
Total column porosity	ε_t		$V_{RetAc} - V_d$	0.822	0.992	0.942
Bed porosity	ε_b		$V_{RetDex} - V_d$	0.342	0.422	0.392
Particle porosity	ε_p		$\frac{\varepsilon_t - \varepsilon_c}{1 - \varepsilon_c}$	0.729	0.986	0.905
Interstitial flow	u_{int}	$\frac{mm}{s}$	u/ε_b	0.487	0.395	0.426
Axial dispersion	D_{ax}	$\frac{mm^2}{s}$	$u_{int}/2 \cdot HETP$	0.089	0.056	0.036
Ionic capacity	Λ	M	$\frac{c_{HCl} \cdot V_{HCl}}{V_c(1 - \varepsilon_t)}$	1.36		2.29

5.3.5   Numerical Methods

The simulations were performed using the in-house software package ChromX [75]. A finite element method with linear Streamline-Upwind-Petrov-Galerkin elements was used here. The discretization in time is performed with the fractional step θ-scheme. The non-linearity of the equation system introduced by the isotherm was treated with Picard iteration. The resulting linear systems are solved by LU factorization.

The estimation process was performed sequentially. First, the measurements of a single run were used for model calibration. After every completed estimation, the next experiment was added to monitor the change of the confidence intervals for manually chosen experimental set-ups. ChromX currently uses forward finite differences to compute the parameter covariance matrix and confidence intervals.

The optimal experimental design procedure was implemented in MATLAB® R2014a that was coupled to ChromX for chromatogram generation. For highest accuracy, the sensitivities were directly calculated by differentiating the model equations. All equations are

Table 5.3: Initial calibration of pore diffusion models.

Species	Capto™	$k_{film}[mm/s]$	$D_{p0}\left[mm^2/s\right]$	D_{p1}	$D_{p2}\left[M^{-1}\right]$
Glucose oxidase	Q	$7.38\cdot10^{-4}\pm18\%$	$3.37\cdot10^{-6}\pm12\%$	-	-
	Phenyl	$2.72\cdot10^{-4}\pm1.5\%$	$2.70\cdot10^{-6}\pm6.5\%$	$0.05\pm45\%$	$2.45\pm11\%$
	adhere	$5.61\cdot10^{-4}\pm20\%$	$1.29\cdot10^{-6}\pm41\%$	$0.16\pm143\%$	$1.04\pm31\%$
Salt	Q	$5.68\cdot10^{-3}\pm187\%$	$2.61\cdot10^{-5}\pm76\%$	-	-
	Phenyl	$1.43\cdot10^{-4}\pm3.9\%$	$3.75\cdot10^{-6}\pm9.3\%$	$0.07\pm65\%$	$0.96\pm17\%$
	adhere	$1.15\cdot10^{-4}\pm12\%$	$1.24\cdot10^{-5}\pm91\%$	$13.9\pm110\%$	$0.58\pm42\%$

Table 5.4: Initial isotherm parameter estimation results.

Capto™	$k_{kin}[sM^{-1}]$	k_{eq}	ν	σ	$q_{max,HIC}[M]$
Q	$3.94\cdot10^{-5}\pm28\%$	$2.21\cdot10^{-6}\pm25\%$	$4.79\pm1.2\%$	$143\pm375\%$	-
Phenyl	$1.00\pm432\%$	$54.8\pm5.6\%$	-	-	$0.51\pm62k\%$
adhere	$0.21\pm45\%$	$0.16\pm104\%$	$2.54\pm3.2\%$	$283\pm198\%$	-

Capto™	n	s	$k_s[M^{-1}]$	$k_p[M^{-1}]$
Phenyl	$0.14\pm62k\%$	-	$3.71\pm0.9\%$	$-32.9\pm20\%$
adhere	$0.25\pm463\%$	$291\pm1875\%$	$3.68\pm3.6\%$	$-8.01\pm76\%$

sufficiently smooth to fulfill the Schwarz integrability condition. Hence, time and parameter derivative can be interchanged and the sensitivities can be integrated over time with the same numerical methods.

5.4   Results and Discussion

5.4.1   Column Characterization

The dead and void volumes given in Table 5.2 were determined from the pulse injections responses of acetone and dextran. The bead size of Capto™ Q is slightly larger, leading to a higher axial dispersion. Furthermore, its structure seems to differ as indicated by the significantly lower bead porosity compared to the other adsorbents. Capto™ Q also features dextran surface extenders which shall increase the binding capacity. Indeed, the volume of HCl needed for titration was almost twice as large as for Capto™ adhere. However, its lower porosity results in a lower ionic capacity per solid volume.

5.4.2   Initial Parameter Estimation

As described in section 5.2.2, different pore diffusion equations were employed to model the measured chromatograms. For HIC and MMC, the GRM with salt-dependent pore diffusion was used.

The estimated GRM and isotherm parameters are listed in Tables 5.3 and 5.4. The resulting plots are displayed in Figure 5.1 and show very good agreement with the measurements. The measurements have not been post-processed and include buffer effects. The simulated sum signal consists of the protein's UV trace and a linear contribution of the salt concentration.
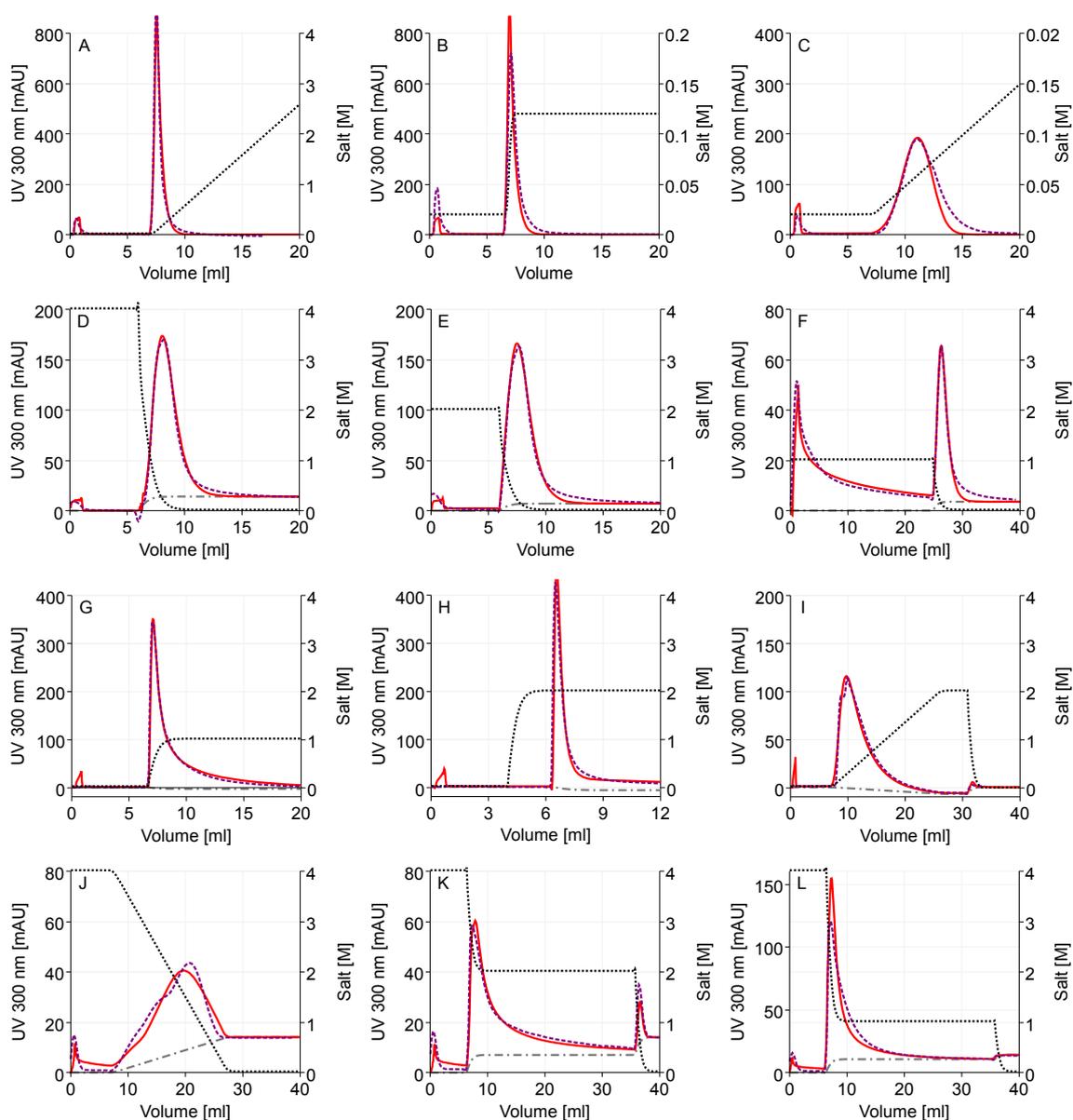
Figure 5.1: Comparison of measured chromatogram (- -) and simulated sum signal (−) for the salt elutions (· · ·) used for initial calibration of the IEC (A-C), HIC (D-F) and MMC (G-L) model. Measurements have not been post-processed and include buffer effects. The simulated sum signal consists of the protein's UV trace and include a baseline shift (-··-).

It is interesting to note, that the two gradients on Capto™ adhere show shoulders and almost double peaks. As the peak shape is identical in the other recorded wavelengths 280 and 457 nm, it is unlikely that impurities are the cause of this. Similar phenomena have been reported in the literature: Karger *et al.* described two peaks resulting from the injection of papain through HPLC and explained this two-peak phenomenon with a metastable state of adsorbed protein after conformational change [96]. McNay *et al.* demonstrated the partial unfolding of lysozyme adsorbed on hydrophobic surfaces with nuclear magnetic and isotope-exchange techniques [127]. By applying an empirical approach, Jungbauer *et al.* studied protein conformational change during hydrophobic interaction chromatography depending on ligand type and salt concentration in the mobile phase [93].

From the parameter estimates, it is unlikely that glucose oxidase unfolds during high salt binding on the mixed mode resin. The characteristic parameter n is rather small. Instead, we could observed the forming of shoulders in the simulation by varying the pore diffusion parameters. The intra-column concentration show that during elution, part of the desorbed protein starts to diffuse deeper into the particles, supported by the faster diffusion rate due to decreasing salt concentration. This phenomenon continues as long as a concentration gradient in the pores exits and the mobile phase salt concentration still allows for binding. Numerical errors are unlikely, the simulations were performed on a fine grid of 50 nodes in the radial dimension and did not change after further refinement.

5.4.3   Confidence Estimates

The 95% confidence intervals were determined from the respective global covariance matrix including all fluid dynamic and isotherm parameter. They reveal that the remaining uncertainty in the linear isotherm parameter estimates for glucose oxidase is still up to $\pm 30\%$, $\pm 20\%$ for the film transfer and pore diffusion coefficients, and the steric shielding is completely undetermined. The influence of the parameter σ is only visible in the non-linear range of the isotherm. It is safe to assume that the injected sample amount was not sufficient to reach it. Furthermore, the correlation matrix contains some high entries which means that the respective parameters correlate with each other and the experiments did not contain information that only relates to a single one, e.g. $(k_{kin}, k_{eq}) = 0.45$, $(k_{kin}, \nu) = -0.71$ and $(k_{eq}, \nu) = -0.67$.

For HIC, only k_{eq} and some of the GRM parameters could be determined with reasonable certainty. As for IEC, the nonlinear range was not reached, n and q_{max} are undetermined with confidence intervals $\pm 62,000\%$. Similar results were obtained for MMC. σ, n, and s, the parameters that are only influential in the nonlinear range have the highest uncertainty.

It is interesting to note, that the estimated k_p values are negative for both, HIC and MMC. The parameter k_s and k_p were introduced by Mollerup depending on protein, salt and pH [133]. The estimated values $k_s > 1$ indicates that the water-protein interactions are stronger than the salt-protein interactions under the given circumstances. The value $k_p < 0$ implies that the protein-protein interactions are stronger than the water-protein interactions.

5.4.4   Optimal Experimental Design

In order to reduce the parameter uncertainty, three experiments were designed consecutively for each resin.

For IEC, the proposed bind/elution conditions were

Exp. 4:   a step from 0.0419 M to 4.02 M,

Exp. 5:   a gradient from 0.02 M to 3.05 M over 20 CV and

Exp. 6:   a gradient from 0.1642 to 0.4959 M over 20 CV.

Table 5.5: Final calibration of pore diffusion models.

Species	Capto™	$k_{film}[mm/s]$	$D_{p0}\,[mm^2/s]$	D_{p1}	$D_{p2}\,[M^{-1}]$
Glucose Oxidase	Q	$6.15\cdot10^{-3}\pm116\%$	$1.60\cdot10^{-6}\pm6.0\%$	-	-
	Phenyl	$2.43\cdot10^{-4}\pm1.1\%$	$2.51\cdot10^{-6}\pm21\%$	$0.03\pm12\%$	$2.45\pm0.6\%$
	adhere	$3.74\cdot10^{-4}\pm10\%$	$1.26\cdot10^{-6}\pm66\%$	$0.18\pm156\%$	$0.94\pm45\%$
Salt	Q	$5.70\cdot10^{-3}\pm19\%$	$3.23\cdot10^{-5}\pm24\%$	-	-
	Phenyl	$1.79\cdot10^{-4}\pm3.7\%$	$3.88\cdot10^{-6}\pm19\%$	$0.06\pm18\%$	$0.96\pm9.2\%$
	adhere	$1.52\cdot10^{-4}\pm14\%$	$6.62\cdot10^{-5}\pm1308\%$	$3.23\pm1233\%$	$0.66\pm129\%$

Table 5.6: Final isotherm parameter estimation results.

Capto™	$k_{kin}[sM^{-1}]$	k_{eq}	ν	σ	$q_{max}[M]$
Q	$1.02\cdot10^{-4}\pm15\%$	$5.60\cdot10^{-6}\pm12\%$	$4.81\pm0.6\%$	$48.4\pm370\%$	-
Phenyl	$1.00\pm1462\%$	$50.4\pm7.8\%$	-	-	$0.13\pm83\%$
adhere	$0.15\pm63\%$	$0.15\pm232\%$	$2.90\pm9\%$	$291\pm83\%$	-

Capto™	n	s	$k_s[M^{-1}]$	$k_p[M^{-1}]$
Phenyl	$2.68\pm109\%$	-	$3.58\pm1.1\%$	$3577\pm422\%$
adhere	$0.24\pm984\%$	$348\pm171\%$	$3.80\pm9.8\%$	$-20.2\pm51\%$

Over the course of OED, the logarithm of the determinant reduced as from

Exp. 1–3: $\ln\left(\det\left(Cov\right)\right) = -123$ to

Exp. 1–4: $\ln\left(\det\left(Cov\right)\right) = -127$,

Exp. 1–5: $\ln\left(\det\left(Cov\right)\right) = -130.5$, and finally

Exp. 1–6: $\ln\left(\det\left(Cov\right)\right) = -130.9$.

The last experiment was not able to improve the objective value significantly indicating that the model cannot be improved under the given constraints.

All parameters are in the same order of magnitude. The film transfer values now attain the maximum and kinetics are even faster. The charge value changed only slightly and steric shielding increased by 25%. While the confidence intervals could be narrowed significantly, not all correlations improved the same. $(k_{kin}, \nu) = -0.19$ and $(k_{eq}, \nu) = -0.19$ improved clearly, but $(k_{kin}, k_{eq}) = 0.40$ did not. A very shallow salt gradient could have improved the certainty further, but was not proposed by OED. A narrower design space would have been beneficial, avoiding unnecessary high salt concentrations.

For HIC, most confidence intervals improved significantly. The optimally designed experiments were two gradients from 4 M to 0.02 M and 3.04 M to 0.23 M, and a step from 3.1 M to 0.47 M. The k_p value, which is now positive is still not well determined and could be removed from the equation when staying within the explored design space. The k_{kin} value of 1 is at the natural upper bound but the fit would have improved further for higher values. This indicates that the desorption term of the kinetic isotherm equation should be smaller. The cause could be further salt concentration dependencies or a hysteresis of adsorption and desorption kinetics. Further studies are necessary to investigate the kinetics of HIC binding under different salt concentrations.
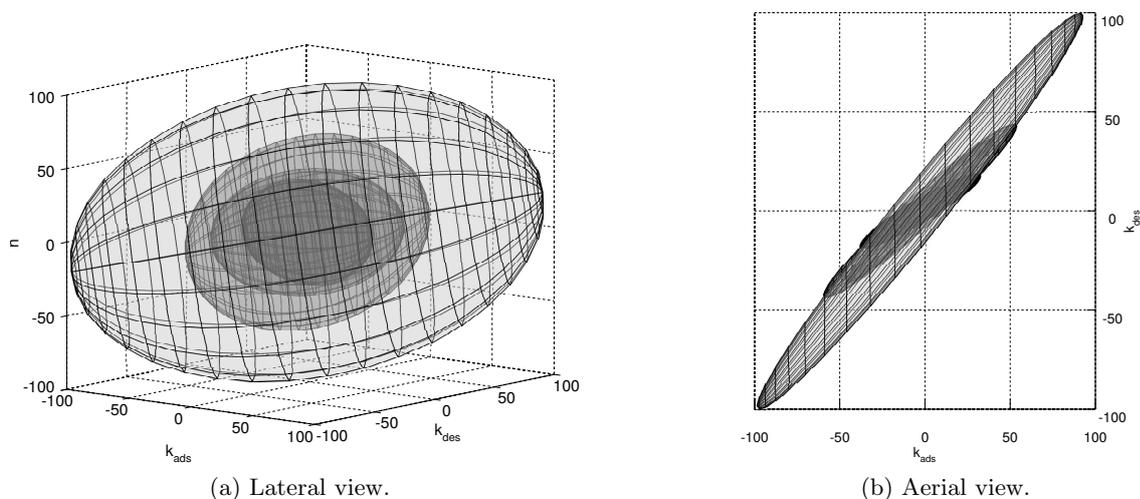
(a) Lateral view.   (b) Aerial view.

Figure 5.2: Ill-posed model equations lead to a suboptimal shape of the confidence ellipsoid. HIC isotherm parameters n, $k_{ads} = k_{eq}/k_{kin}$ and $k_{des} = k_{kin}^{-1}$ were estimated with the optimum being located skew to the parameter axes which show the remaining size of the confidence interval in percent compared to the first estimation.

Plotting the confidence ellipsoids can be beneficial when assessing the quality of the estimates. To demonstrate this, we performed parameter estimation for the HIC model again, but with the (k_{ads}, k_{des}) formulation instead of (k_{eq}, k_{kin}). In a previous study for SMA [77], it was shown that deterministic solvers fail when the direction of decent is skew to the parameter axes. For the HIC model parameters k_{ads}, k_{des}, and n, we obtained the ellipsoids plotted in Fig. 5.2 for Exp. 1–3, 1–4, 1–5, and 1–6. While the lateral view shows the relative shrinking of confidence intervals, the aerial view shows a similar result as in [77] for SMA. The deterministic optimizer was not able to move away from the initial estimate and the confidence ellipsoid is long drawn skew to the axes. From the plot it would have been obvious that the ratio k_{ads}/k_{des} must be considered to minimize the volume of the ellipsoid.

Completely different results were obtained for MMC. Unfortunately, some of the confidence intervals became larger after applying OED. This does not necessarily mean, that the actual values are worse only that the fit is not as sensitive anymore to small changes in the parameter values. The three additional runs were all gradient elutions from 0.17 M to 4.02 M, 0.22 M to 1.84 M and 3.5 M to 0.65 M.

The correlation matrix shows three clusters (Fig. 5.3). The GRM parameters of glucose oxidase correlate strongly with each other, the same applies to salt. This could be caused by the high salt concentrations of the first two OED gradients that lead to a significant flow-through peak which do not seem to be sensitive to changes in the GRM parameters. The third cluster spans all isotherm parameters.

While the fit is certainly very good and even shoulders could be simulated, only k_s and ν have acceptable confidence intervals. For comparison, the average error reported in [142] for protein parameters determined from batch isotherms are $k_{eq} > 25\,\%$, $\nu > 10\,\%$, $n > 40\,\%$, $k_p > 40\,\%$, $k_s > 25\,\%$, $q_{max} > 15\,\%$.

Figure 5.3: Visualization of the correlation matrix of the MMC model parameters after applying OED. The range of correlation coefficients is $[-1, 1]$ with higher absolute values indicating stronger correlations.

5.4.5 Interpretation of Isotherm Parameters

Many SMA parameter sets have been reported in literature, e.g. [76; 85; 149]. The order of magnitude of the parameters found here is reasonable: k_{kin} is small in order to generate a steep elution peak front and $\nu \approx 5$ means that the 160 kDa molecule binds to approximately 5 ligands at once. The k_{eq} values given here are taken with respect to the adsorber skeleton. To interpret the magnitude easier, approximate values per column volume can be calculated. This is achieved via a transformation of the equilibrium isotherm:

$$
\begin{aligned}
q &= k_{eq} \left(\Lambda - (\nu + \sigma) \, q \right)^{\nu} c_s^{-\nu} c \\
&= k_{eq} \left(\frac{1 - \varepsilon_t}{1 - \varepsilon_t} \Lambda - (\nu + \sigma) \, q \right)^{\nu} c_s^{-\nu} c \\
&= k_{eq} \left(1 - \varepsilon_t \right)^{-\nu} \left((1 - \varepsilon_t) \Lambda - (1 - \varepsilon_t) (\nu + \sigma) \, q \right)^{\nu} c_s^{-\nu} c \\
&=: k_{eq,CV} \left(\Lambda_{CV} - (\nu + \sigma_{CV}) \, q \right)^{\nu} c_s^{-\nu} c.
\end{aligned}
$$

We obtain a reasonable value of

$$k_{eq,CV} = k_{eq} \left(1 - \varepsilon_t\right)^{-\nu} = 0.023$$

that fits to the immediate desorption observed with slightly increased salt concentration. The steric shielding parameter is unfortunately undetermined.

The GRM parameters of the HIC model are the best in terms of certainty. The same applies to the non-linear parameters n and q_{max}. The usually lower capacity of HIC resins seems to be beneficial in this case. However, the high uncertainty in the kinetic parameter combined with the tendency to attain values greater than one indicates that the model is not perfectly formulated.

The confidence and correlation analysis of the MMC model parameters shows that the nonlinear parameters could not be determined under the given constraints. Nevertheless, OED was able to improve the remaing correlations in a way that GRM and isotherm parameters mostly correlate among themselves.

5.5   Conclusions

The resin Capto™ adhere offers the ion-exchange and hydrophobic interaction modes of Capto™ Q and Capto™ phenyl but the employed model protein glucose oxidase shows a very unique binding behavior. Models based on the SMA, Mollerups's HIC, and Nfor's MMC model were calibrated initially by three (IEC and HIC) or six (MMC) chromatograms obtained under manually selected conditions. As none of the models supports a change in pH, only elution by varying of the salt concentration could be applied.

Based on these initial models, new experimental set-ups were designed by minimizing the confidence ellipsoid according to the D-optimality criterion Eq. (5.26). After sequentially conducting the new experiments , the size of the confidence ellipsoid could be reduced. The results underline the potential of OED to reduce time and material consumption compared to DoEs, where additional experiments do not necessarily improve the confidence.

To evaluate the reliability of estimates, approximate confidence intervals were calculated for each estimated parameter after including the measurements of a new experiment. After three OED runs, IEC and HIC parameter estimates became more reliable, MMC still showed strong correlations and uncertainties.

The model parameters were analyzed to gain information on the adsorption of glucose oxidase on Capto™ Q, phenyl and adhere under the investigated experimental conditions. In the literature, only non-negative estimates for the mixed-mode isotherm parameter k_p were reported, when modeling the adsorption between Capto™ adhere and proteins with similar pI and size as glucose oxidase [142]. Here, we obtained highly negative values for both MMC and HIC, indicating that protein-protein interactions dominate over water-protein interactions. This results should be subject of further studies.

Furthermore, peak distortions were observed during HIC and MMC elution. On the one hand, these observations match with descriptions of partial unfolding of proteins during

interaction with hydrophobic ligands in the literature [93; 127]. But this phenomenon is currently not taken into account when modeling HIC and MMC mechanistically. On the other hand, the characteristic HIC parameter n is too small to justify the assumption of unfolding and we could trigger the forming of shoulders by introducing salt-dependent pore diffusion.

In summary, mixed mode chromatography offers a broad operating window under high and low salt conditions. The OED approach was used successfully to improve the reliability of isotherm parameter estimates in two of three cases investigated, IEX and HIC. For MMC it became obvious that an excellent fit does not necessarily imply a well-calibrated model. In future, the OED approach could be used for automated process development while enhancing the reliability and robustness of downstream-processing when implementing the Quality by Design approach. However, expert knowledge is still necessary to define a reasonable parameter space. Otherwise, OED might propose infeasible or unsuitable experiments at the boundaries.

Acknowledgments

6 | Model-based Integrated Optimization and Evaluation of a Multi-step Ion Exchange Chromatography

Thiemo Huuk[1], Tobias Hahn[1], Anna Osberghaus[1], Jürgen Hubbuch[1]

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

Abstract

Current approaches to downstream process development in the biopharmaceutical industry are commonly based on a combination of platform technology, high-throughput experimentation, and 'rules of thumb'. These empirical strategies conflict with demands for a mechanistic process understanding and a rational definition of design space, issued by the *Quality by Design* approach (QbD). Model-based process simulation and optimization are options for implementation of QbD. A model-based process optimization approach has to consider the complexity of biopharmaceutical downstream processes, especially the interactions of multiple chromatographic operations.

We present a case study on model-based concerted process optimization of two consecutive ion exchange chromatographies (Poros 50HS and Q Sepharose FF). Our optimization approach includes a process flowsheet optimization, the shape of the salt gradient, and the boundaries of fraction collection for both columns. The superiority of the presented concerted process optimization approach is demonstrated by comparison to a sequential approach that optimizes the two ion exchange chromatographies (IEX) consecutively. Verification is carried out with a set of three model proteins (cytochrome c, chymotrypsin, ribonuclease A).

The *in silico* optimum is reproduced in lab experiments and the modeling tool is successfully employed for the identification and characterization of critical process parameters (CPP).

Nomenclature

Abbreviation	Unit	Definition
$c_{p,i}$	M	concentration of protein i in the pores of the adsorber
$c_{p,salt}$	M	salt concentration in the pores of the adsorber
c_i	M	protein concentration i in the interstitial phase
c_{in}	M	protein concentration i at the column inlet
c_{out}	M	protein concentration i at the column outlet
D_{ax}	$mm^2 s^{-1}$	axial dispersion coefficient
ε_b		voidage of the bed
ε_p		particle voidage
ε_t		total voidage of the bed
$HETP_{Dex}$	mm	height equivalent of a theoretical plate calculated from dextran injection
$k_{ads,i}$		adsorption coefficient of protein i in the SMA isotherm
$k_{des,i}$		desorption coefficient of protein i in the SMA isotherm
$k_{eq,i}$		$k_{ads,i} \cdot k_{des,i}^{-1}$
$k_{eff,i}$	$mm^2 s^{-1}$	effective mass transfer coefficient of protein i
$k_{kin,i}$		$k_{des,i}^{-1}$
L	mm	length of the column
Λ	M	total ionic capacity per adsorber skeleton volume
ν_i		characteristic charge of protein i in SMA isotherm
q_i	M	protein concentration i on the adsorber skeleton phase
r_p	mm	radius of adsorber particles
σ_i		steric shielding coefficient of protein i in the SMA isotherm
t	s	time
$u(t)$	$mm s^{-1}$	interstitial velocity of the fluid
x	mm	axial position in the column

6.1   Introduction

Nowadays, the majority of downstream processes (DSP) for biopharmaceuticals are based on multiple chromatographic and non-chromatographic separation techniques. The use of chromatography is by large due to its mild conditions, the diversity of possible interaction modes, and the long history of chromatographic operations in industry and regulatory authorities. Aside from chromatography, non-chromatographic techniques such as micro- or ultra-filtration or pH conditioning operations are necessary to prepare the process solutions for a chromatographic operation or to ensure bacterial or viral safety and to remove insoluble particles [115; 171]. Industrial DSP development, especially for monoclonal antibodies (mAb), is commonly based on platform processes which are slightly adapted to new entities [40; 115; 171], high-throughput experimentation or 'rules of thumb' [144]. On

the one hand, platform approaches ensure fast process development and reduce the time to market. On the other hand, it has to be assumed that the strongly reduced exploration of design space during adaptation of platform processes to new entities is prone to lead to suboptimal processes [140; 203]. In addition, the limited mechanistic knowledge might account for an increased number of batch failures [31]. Model-based process understanding and optimization are sensible and important additions to platform processes [78; 81]. The integration of modeling tools into process development is an essential part of the strategy for implementation of the *Quality by Design* (QbD) approach [31; 78; 86]. Consequently, modeling tools are increasingly gaining the attention of the pharmaceutical industry [20; 41; 105].

Most experimental and modeling approaches focus on single chromatographic operations [35; 40; 68; 148; 150], or optimize consecutive operations in a sequential manner [71], or with strong simplifications such as non-mechanistic peak shapes (triangles instead of peaks) [154; 155] or short-cut methods [124]. However, industrial downstream processes (DSP) are commonly based on chromatographic and non-chromatographic operations, which are arranged in a sequential manner to meet the ambitious purity requirements [171; 191]. An approach that only focuses on single-unit operations will probably be unable to identify the global process optimum with respect to yield, purity or economic considerations [81; 141]. To meet the growing requirements of industry to integrate modeling tools into industrial DSP, the commonly used modeling approach has to be extended from single chromatographic operations to concerted multi-step optimization tasks.

In 2012, Helling et al. [81] presented a chromatographic sequence of hydrophobic interaction chromatography (HIC) and cation exchange chromatography (CEX) to justify the need for integrated optimization of a two-column sequence. The presented approach is restricted to optimization of the fraction boundaries and disregards optimization of the salt elution gradients or a process flowsheet optimization. Furthermore, the approach only mentions one set of isotherm parameters to predict the retention behavior of a whole crude antibody feedstock.

Nfor et al. [141] presented a model-based optimization of a multi-step downstream sequence, focusing on the optimal arrangement of several different chromatographic modi. The approach is restricted to a partial fraction of possible operational variables (linear flow, gradient length, loading factor). The implications of intermediate operations in between the chromatographic operations are only considered under economic aspects.

An approach disregarding the dynamics of interaction between the single-column processes (e.g. salt gradient, flow, fractionation, etc.) for all unit operations in a concerted manner including a process flowsheet optimization will most probably be unable to identify the global process optimum [140; 203].

In this manuscript, an integrated modeling approach for two consecutive ion exchange operations and an intermediate buffer exchange is presented. This approach is capable to identify the global process optimum for the separation of several protein species and is demonstrated by a case study combining cation and anion exchange chromatography (CEX: Poros 50HS, AEX: Q Sepharose FF). A ternary mixture of chymotrypsin (Chy), cytochrome c (Cyt), and ribonuclease A (RibA) is chosen as model system for the separation problem. This set of model proteins covers a narrow pI range of 8.7, 9.5, and

9.6, respectively, constituting a great challenge to the presented separation task. Inverse calibration of a lumped rate model for chromatography, combined with steric mass action isotherm, is carried out on linear gradient elution data and breakthrough curves [148]. First the column parameters are determined experimentally (e.g. voidage). Afterwards, the isotherm parameters are estimated iteratively by minimizing the discrepancy between the experimental and simulated chromatograms. Alternatively, the isotherm parameters could also be determined experimentally from isocratic, gradient, and frontal experiments and correlations (e.g. [152; 170]). In this case study we decided to use the model-based inverse calibration due to its better compatibility with industrial process development workflows, the waiver of correlations, and the faster process development. For a comparison of the inverse calibration and correlation approach we refer to [148]. The calibration is done separately for both chromatographic operations. A subsequent *in silico* optimization is used to identify the global process optimum for an objective function with respect to cytochrome c purity, yield, and a minimal volume of the fraction collected from the final column. The optimization includes the shape of the salt gradient and the boundaries of fraction collection. In addition, the optimization accounts for an arbitrary order of the two IEX (flowsheet optimization), namely an AEX→CEX and a CEX→AEX process flowsheet. The optimization of all parameters for both columns is carried out in a concerted manner. The advantage of this concerted optimization approach is demonstrated by a comparison with a conventional sequential optimization of the two IEX. The use of three linear gradient elutions and one breakthrough curve for model calibration reduces the sample consumption for this case study, as compared to classical approaches using *Design of Experiments* (DoE). This case study bases on a two-column process, but the presented modeling approach is extendable to more complex arrangements of unit operations and larger sets of protein species.

6.2   Theory

6.2.1   Transport-dispersive Model

The transport-dispersive model (TDM) [129] depicted in Eqs. (6.1),(6.2) is used to model the macroscopic mass transport through the column. The system is of convection-diffusion-reaction (CDR) type. The rate of change of a protein concentration $c_i(x,t)$ in the interstitial phase consists of convective mass transport in space with respect to the average interstitial velocity $u(t)$ of the flowing fluid. Peak broadening effects are modeled as dispersion in space with respect to a lumped coefficient D_{ax}, and transition from the interstitial concentration into the particle pore concentration $c_{p,i}(x,t)$ depending on the voidage of the bed ε_b, the radius of adsorber particles r_p, and an effective transfer coefficient $k_{eff,i}$. The model is one-dimensional, such that the concentrations depend on the axial position in the column $x \in [0, L]$ and time. Hence, the axial dispersion coefficient D_{ax} can be calculated from the height equivalent of a theoretical plate ($HETP_{Dex}$) of a dextran injection and the interstitial velocity ($D_{ax} = HETP_{Dex} \cdot u(t)/2$) [129]. The approach to determine D_{ax} from the $HETP$ stems from the van Deemter equation and works well, but correlation for film and pore diffusion coefficients (e.g. [91]) often do not describe the behavior for high and low concentrations at the same time. Hence we rather estimate the effective mass transfer coefficient k_{eff} or set it constant if the axial dispersion is dominant and we do not see a change in peak shape.

The second equation models the accumulation of mass in the pore phase $c_{p,i}$ and stationary phase q_i (protein binding to the adsorber) depending on the particle voidage ε_p. The model is complemented with Danckwerts boundary conditions, Eqs. (6.3),(6.4), and an isotherm equation modeling the stationary phase concentration q_i.

$$\frac{\partial c_i(x,t)}{\partial t} = -u(t)\frac{\partial c_i(x,t)}{\partial x} + D_{ax}\frac{\partial^2 c_i(x,t)}{\partial x^2}$$
$$-\frac{1-\varepsilon_b}{\varepsilon_b}k_{eff,i}\frac{3}{r_p}\left(c_i(x,t) - c_{p,i}(x,t)\right) \quad (6.1)$$

$$\varepsilon_p\frac{\partial c_{p,i}(x,t)}{\partial t} + (1-\varepsilon_p)\frac{\partial q_i(x,t)}{\partial t} = k_{eff,i}\frac{3}{r_p}\left(c_i(x,t) - c_{p,i}(x,t)\right) \quad (6.2)$$

$$\frac{\partial c_i}{\partial x}(0,t) = \frac{u(t)}{D_{ax}}\left(c_i(0,t) - c_{in,i}(t)\right) \quad (6.3)$$

$$\frac{\partial c_i}{\partial x}(L,t) = 0 \quad (6.4)$$

6.2.2 Steric Mass Action Isotherm

The steric mass action isotherm (SMA) introduced by Brooks and Cramer in 1992 [24] is a commonly used semi-mechanistic isotherm in ion exchange chromatography, involving one or more macromolecules with steric hindrance, in this case of proteins. It is capable to reproduce the influence of counter ions on the retention behavior of protein species, using the proteins' characteristic charges ν_i. Besides this, it considers column properties like the total ionic capacity Λ and steric shielding effects σ_i of the proteins, blocking an amount of binding sites greater than the actual number of sites it interacts with. The kinetic SMA isotherm is given in Eq. (6.5), with q_i and $c_{p,i}$ being the concentration of the protein i adsorbed and in solution, respectively. $c_{p,salt}$ is the salt concentration of the solution. $k_{ads,i}$ and $k_{des,i}$ are the adsorption and desorption coefficients.

$$\frac{\partial q_i(x,t)}{\partial t} = k_{ads,i}\underbrace{\left(\Lambda - \sum_{j=1}^{k}(\nu_j + \sigma_j)q_j(x,t)\right)^{\nu_i}}_{\bar{q}_{salt}(x,t)}c_{p,i}(x,t)$$
$$-k_{des,i}c_{p,salt}^{\nu_i}(x,t)q_i(x,t) \quad (6.5)$$

This formulation is not well suited for inverse parameter estimation as the change of k_{ads} or k_{des} always affects peak height and retention at the same time. In order to alter peak height unimpededly, the isotherm equation was modified as shown in Eq. (6.6). In this isotherm equation, parameters for kinetic effects ($k_{kin} = 1/k_{des}$) and equilibrium ($k_{eq} = k_{ads}/k_{des}$) are separated, such that a change in k_{kin} strongly affects peak height, while the retention time is preserved to a large extent; vice versa for k_{eq} [77].

$$\underbrace{\frac{1}{k_{des,i}}}_{k_{kin,i}}\frac{\partial q_i(x,t)}{\partial t} = \underbrace{\frac{k_{ads,i}}{k_{des,i}}}_{k_{eq,i}}\bar{q}_{salt}^{\nu_i}(x,t)c_{p,i}(x,t) - c_{salt}^{\nu_i}(x,t)q_i(x,t) \quad (6.6)$$

According to our experience, correlations for determining the linear isotherm parameters from the gradient elution results [152; 208] only determine the characteristic charge parameter well. A previous study [148] showed that this parameter can be found accurately

with chromatogram fitting as well, and that estimation of k_{eq} is inevitable for simulation. For an elaborate discussion on the applicability of correlations we refer to the Handbook of Process Chromatography by Hagel, Jagschies and Sofer [73].

6.2.3 Numerical Solution

Following the method of lines, the equation system is first discretized in space on given nodes, using the finite element method (FEM). FEM is a highly versatile method with strong mathematical foundation and well suited for CDR equations. The solution procedure starts with the weak formulation, incorporating the boundary conditions and representing the variables with basis functions from the respective spaces. A Galerkin ansatz was used here, choosing basis and test functions from the same spaces, specifically first- and second-order polynomials. The discretization in time is performed with the Crank-Nicolson method, a semi-implicit procedure providing second-order accuracy. Finally, the non-linearity of the equation system introduced by the isotherm must be treated with an iterative procedure, here Picard iteration or Newton's Method. The resulting linear systems are solved depending on their dimension with a direct method (e.g. LU factorization) or iterative method (e.g. GMRES). For a general overview of numerical methods for PDEs, we refer to [159].

6.2.4 Estimation and Optimization

Because of the system's non-linearity, finding a global optimum for parameter estimation and subsequently for process optimization is challenging. Deterministic methods cannot leave local minimums and only provide quadratic convergence when starting sufficiently near the optimum. Heuristic methods such as simulated annealing, genetic or evolutionary algorithms cover a larger search space by performing random jumps, but give no guarantee on the convergence rate. A common approach is to start with a heuristic method to identify candidates for global optima and find the final solution using a deterministic algorithm. For parameter estimation, the objective is constituted by curve fitting, such that the goal function is given by the sum of square errors between simulation and measurement data. The simulation uses molar concentrations. Therefore, the measured UV traces are converted from absorption units to molar concentrations using the known molar protein amounts injected to the column and the recorded peak areas. For process optimization, the found parameter estimates are fixed and a system parameter such as salt concentration is used as optimization variable. In this case study, the objective consists of the product-related quality attributes, namely loss, purity, and dilution of the target component cytochrome c.

6.3 Materials

6.3.1 Chromatographic Instrumentation

The chromatographic experiments were carried out using an ÄKTApurifier 10 fast protein liquid chromatography (FPLC) equipped with Pump P-903, UV (10mm path length), conductivity and pH monitor UPC-900, an autosampler A-900 and a fraction collector Frac-950 (all GE Healthcare, Little Chalfont, Buckinghamshire, UK). The instrument was controlled with UNICORN 5.10 (GE Healthcare, Little Chalfont, Buckinghamshire, UK).

6.3.2 Adsorbers, Buffers, and Proteins

For ion exchange chromatography (IEX), a 1mL column (8x20mm) prepacked with Poros 50HS strong cation exchange adsorber (Applied Biosystems, Carlsbad, CA, USA, column packing by Atoll, Weingarten, Germany) was used. A second 1mL column (7x25mm) prepacked with the strong anion exchange resin Q Sepharose FF (QFF) was supplied by GE Healthcare (Carlsbad, CA, USA). Between the runs, the columns were stored in a bacteriostatic solution. After storage, the columns were pre-charged by a prolonged equilibration with low and high salt buffer. For cation exchange chromatography, a 50mM sodium citrate buffer (Merck, Darmstadt, Germany) with 0 and 1M additional NaCl (Merck, Darmstadt, Germany) was used at pH 5.0. For the anion exchange experiments, a 50mM 1-methylpiperazine buffer (Sigma, St. Louis, MO, USA) supplemented with 0 or 1M NaCl was used at pH 9.8. All solutions were prepared using ultra-pure water (UPW) (arium pro UV, Sartorius, Göttingen, Germany). Buffers were $0.22\mu m$-filtrated and degassed by sonification.

Lyophilized α-chymotrypsin (bovine pancreas, no. C4129), ribonuclease A (bovine pancreas, no. R5503), and cytochrome c (bovine heart, no. 30398) were used as model proteins (all from Sigma, St. Louis, MO, USA). All protein solutions were prepared using the respective low salt buffer and were $0.22\mu m$-filtrated prior to usage.

6.3.3 Software

Isotherm parameter estimation, chromatogram simulation, and process optimization were carried out using the in-house developed software ChromX. ChromX provides numerical tools for solving various kinds of chromatography models, including the model combination of TDM and SMA. The discretization scheme chosen was a linear finite element ansatz in space using 100 equidistant nodes, which provided a high simulation speed and maintained accuracy as compared to runs with quadratic elements and a higher number of computational nodes. The aforementioned Crank-Nicolson scheme was used for time-discretization with the time step set constant to one second. A Picard iteration was employed for the non-linearity together with UMFPACK as linear solver. ChromX was used for a first evaluation of results, the final chromatograms were exported as comma-separated values and plotted with Matlab R2012b (The Mathworks, Natick, ME, USA) and CorelDRAW X5 (Corel, Ottawa, Canada). For parameter estimation and process optimization, ChromX offers interfaces to various libraries. The implementations used here were *levmar* [117] as deterministic method, based on the Levenberg-Marquardt algorithm (LMA), and the genetic algorithm optimizer *GAlib* [197] as heuristic counter-part (GA).

6.4 Methods

6.4.1 Extra Column Effects

The chromatographic system and the two columns were characterized with $25\mu L$ tracer injections at a linear flow of 100cm/h. This corresponds to a volumetric flow of 0.838 and

$0.641mL/min$ for the CEX and AEX, respectively. 1%(v/v) acetone (Merck, Darmstadt, Germany) injections were used to determine the system dead volume. As pore-penetrating, not interacting tracer, 1M NaCl and acetone were applied to the Poros 50HS and QFF column, respectively. The determined dead volumes calculated from the 280nm and conductivity signals were used for the correction of the measured raw data. $25\mu L$ injections of a filtrated 10g/L dextran 2000kDa solution (Sigma, St. Louis, MO, USA) onto the two columns were used to determine the volume of the inter-particle space. Furthermore, the dextran signals at 215nm were used to calculate the axial dispersion coefficient using UNICORN.

6.4.2   Column Titration

Acid-base titration was carried out to determine the total ionic capacity Λ of the two columns. In brief, the CEX column was flushed with a 0.5M HCl solution until a constant UV and conductivity signal was achieved. Afterwards, the column was washed with UPW until a constant UV and conductivity baseline was reached. After that, the column was titrated at a flow of 100cm/h with a 0.01M NaOH solution until an increase in conductivity signal was recorded. From the Na-ion concentration of the titrant and the volume of the applied titrant, the total number of exchangeable ions was calculated. In the procedure for the AEX, the HCl and NaOH solutions were exchanged.

6.4.3   Linear Gradient Elution

Linear gradient elution (LGE) data were used for isotherm parameter estimation. The whole method was carried out at a linear flow of $100cm/h$. After a two-column volume (CV) equilibration with the respective low salt buffer, $25\mu L$ of a 0.6mM pure protein solution was automatically injected onto the column. Ribonuclease A was applied in a concentration of 1.2mM due to its lower extinction coefficient. Unbound protein was removed from the column during a 2CV flushing with low salt buffer. Afterwards, 15CV linear gradient elutions with a final NaCl concentration of 500, 600 and 700mM for the CEX and 300, 400 and 500mM for the AEX were carried out. The final NaCl concentrations for ribonuclease A on the AEX were 200, 300 and 400mM, and for cytochrome c on the CEX 400, 500, 600mM. A subsequent 2CV wash with the final gradient concentration buffer and an additional 2CV regeneration step using the respective buffer with 1M NaCl were applied. The 280nm traces plotted over the volume were exported with UNICORN in Microsoft Excel format.

6.4.4   Breakthrough Curves

Breakthrough curves (BC) were used for the estimation of the shielding parameter σ in the SMA isotherm. For the BC, a 0.6mM protein or 1.2mM ribonuclease A solution in low salt buffer, respectively, was applied to the equilibrated column with a SuperLoop (GE Healthcare, Little Chalfont, Buckinghamshire, UK). The flow was chosen to $100cm/h$. The 280nm traces plotted over the volume were exported with UNICORN in Microsoft Excel format.

6.4.5   Buffer Exchange

A buffer exchange was carried out between the CEX and AEX step during the experimental process evaluation. For buffer exchange purposes, a VivaSpin 20 centrifugal concentrator (Sartorius, Göttingen, Germany) equipped with a 3kDa molecular weight cut-off (MWCO) polyethersulfone (PES) membrane was used. The fraction collected from CEX was twofold concentrated and adjusted with the target buffer. This procedure was repeated three times to exchange the buffer.

6.4.6   Parameter Estimation

The SMA isotherm parameters k_{kin}, k_{eq}, and the charge ν were estimated from the three linear gradient elution data using the software ChromX. The gained set of parameters was completed by the estimation of the shielding parameter σ from the breakthrough curves. Due to the appearance of multiple protein signals for ribonuclease A, ribonuclease was treated as two protein species, represented by two different sets of isotherm parameters (RibA1 and RibA2). For the AEX, a blank substraction due to a salt-induced baseline drift of the buffer was carried out. For estimation, the GA and LMA algorithms were used. The 280nm traces plotted over the volume, exported with UNICORN, have been imported into ChromX respecting the FPLCs dead volume.

6.4.7   *In silico* Process Optimization

The reviewed model process consists of one CEX and one AEX in arbitrary order. A ternary protein mixture containing 0.6mM of cytochrome c and chymotrypsin and 1.2mM ribonuclease A is applied to the first column *in silico*. The initial gradient condition, the gradient length and slope are *in silico* optimized for both IEX. In addition, the start and end of fraction collection are optimized *in silico* for both columns. The fraction collected from the first column is *in silico* desalted, doubled, and injected to the second column. Within the concerted optimization approach, the process parameters are optimized for both columns at once, just evaluating the fraction collected from the second column with respect to an objective function. The order of CEX and AEX is kept arbitrary. The concerted optimization approach is illustrated in Figs. 6.4 A and 6.5 A.

To evaluate whether the concerted approach is superior to conventional sequential optimization, the CEX→AEX process optimization is repeated using a sequential optimization. In the sequential approach, the first column is optimized with respect to the objective function. Afterwards, the fraction from the first column is virtually injected to the second column for its optimization. The sequential optimization approach is illustrated in Fig. 6.6 A. The parameters to be optimized and their ranges are illustrated in Fig. 6.1. The overlapping ranges for the start and end concentration of the salt gradient allow linear gradient and isocratic elution. The lower limit of the starting time for fraction collection is zero. Therefore, a flow-through operation is allowed besides bind-elute.

The objective function used in all approaches intends to achieve a great molar purity and a great molar yield of the target protein cytochrome c and a low dilution of cytochrome c
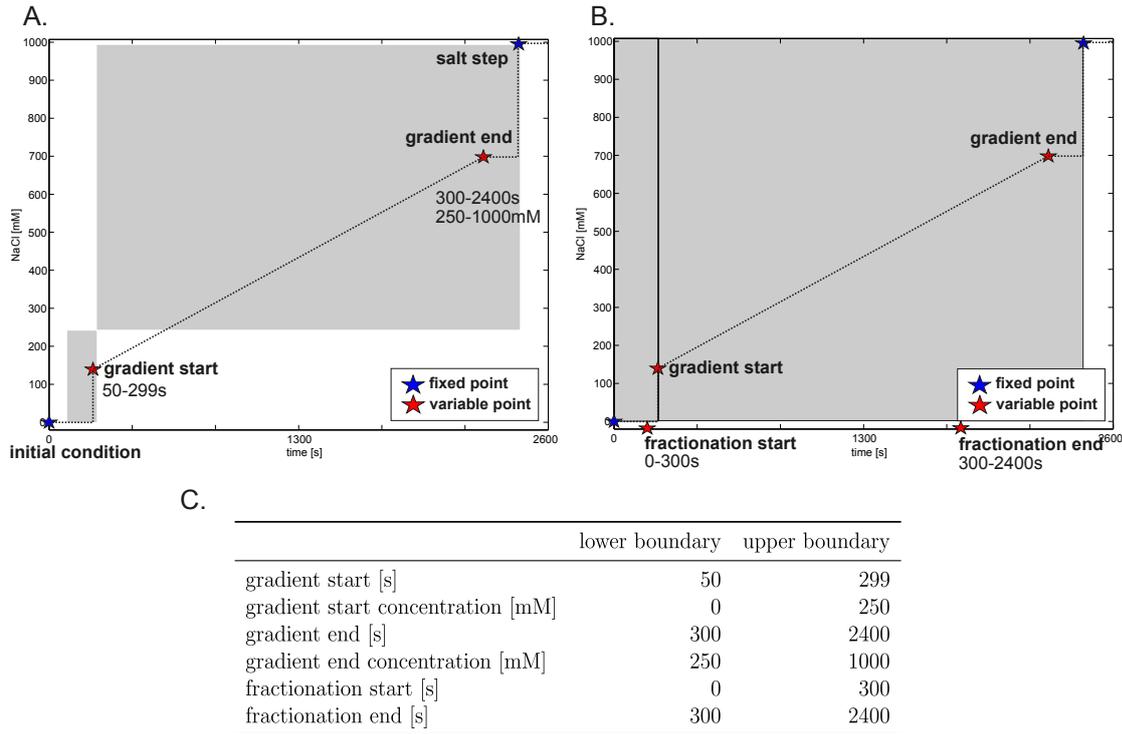
Figure 6.1: Parameters with ranges for the optimization of the IEX salt gradient (A) and the boundaries for the collection of fractions (B) in AEX and CEX. The gray boxes indicate the ranges of the specified parameters. Table C lists the boundaries for parameter optimization.

in the evaluated fraction. With weighting factors, the objective function is given by

$$purity[mol/mol] \in [0, 1]$$
$$yield[mol/mol] \in [0, 1]$$
$$fraction[s] \in [0, 2400]$$

$$\min_{\bar{p}} \left(\underbrace{(1 - purity)}_{impurity} + 0.5 \cdot \underbrace{(1 - yield)}_{loss} + 0.01 \cdot \frac{fraction}{mL} \right) \qquad (6.7)$$

with $\bar{p}$ depicting the variable parameter set. The terms refer to cytochrome c in the chosen fraction. The weighting terms were introduced to compensate the different magnitudes of the three factors. The range of time, respectively volume values is much larger then purity and yield, such that it has been weighted with 0.01. The model describes concentration over time, therefore the boundaries for fraction collection are given in seconds. Irrespective of this calculation, the objective of 'fraction' is calculated using volumes instead of times, therefore the fraction size within the objective function is given in mL. The first column is run in duplicate and the fractions are pooled and directed to the second column. The factor 0.5 for yield equalizes this intermediate pooling. Due to its insensitivity to local minimums of the objective function, the GA was used for process optimization.

Table 6.1: Voidages are calculated from tracer injections. The total ionic capacity is measured by acid-base titration. The axial dispersion coefficient is calculated from dextran injections.

		Poros 50HS	Q Sepharose FF
bed voidage	ε_b	0.341	0.315
particle voidage	ε_p	0.449	0.851
total voidage	ε_t	0.638	0.898
total ionic capacity per adsorber volume$/M$	Λ	0.165	2.117
axial dispersion coefficient$/\frac{mm^2}{s}$	D_{ax}	0.135	0.176

6.4.8 Experimental Process Evaluation

The best process in the *in silico* optimization was reproduced in lab experiments on the Äkta FPLC, namely the concerted optimized sequence of CEX→AEX. The ternary protein mixture was applied to the CEX and the optimized salt gradient and peak fractionation were executed. The collected fractions from two identical runs were pooled and the buffer was exchanged to the low salt AEX buffer. Afterwards, the sample was applied to the AEX, executing the optimized salt gradient and fraction collection. The performances of the experimental systems were evaluated with respect to the data predicted in the modeling approach.

6.5 Results

6.5.1 System Characterization

The FPLC dead volume of $70\mu L$ was determined by tracer injections without a column attached to the system. All other data were corrected with respect to this dead volume. The column voidages were calculated from injections of a pore-penetrating, non-interacting and a non-pore-penetrating, non-interacting tracer. The total ionic capacity was determined by acid-base titration. The calculated voidages and capacities are given in Tab. 6.1.

6.5.2 Parameter Estimation

Estimation of the isotherm parameters was carried out using the Levenberg-Marquart (LMA) and genetic algorithm (GA). First, the kinetic, equilibrium and charge parameters were estimated from the three linear gradient elution (LGE) data. Afterwards, the shielding parameter was estimated from the three LGE and the breakthrough curve (BC). After estimation of the shielding parameter, a comparison of the model response and the experimental data revealed that for the presented case study, film diffusion has a negligible impact on the model quality. Therefore, k_{eff} was set to $r_p \cdot 3^{-1}$ (ref. Eq. (6.1)).

The measured data and the model responses from parameter estimation are given in Figs. 6.2 and 6.3 for Poros 50HS and Q Sepharose FF, respectively. The estimated parameters for both columns are summarized in Tab. 6.2.
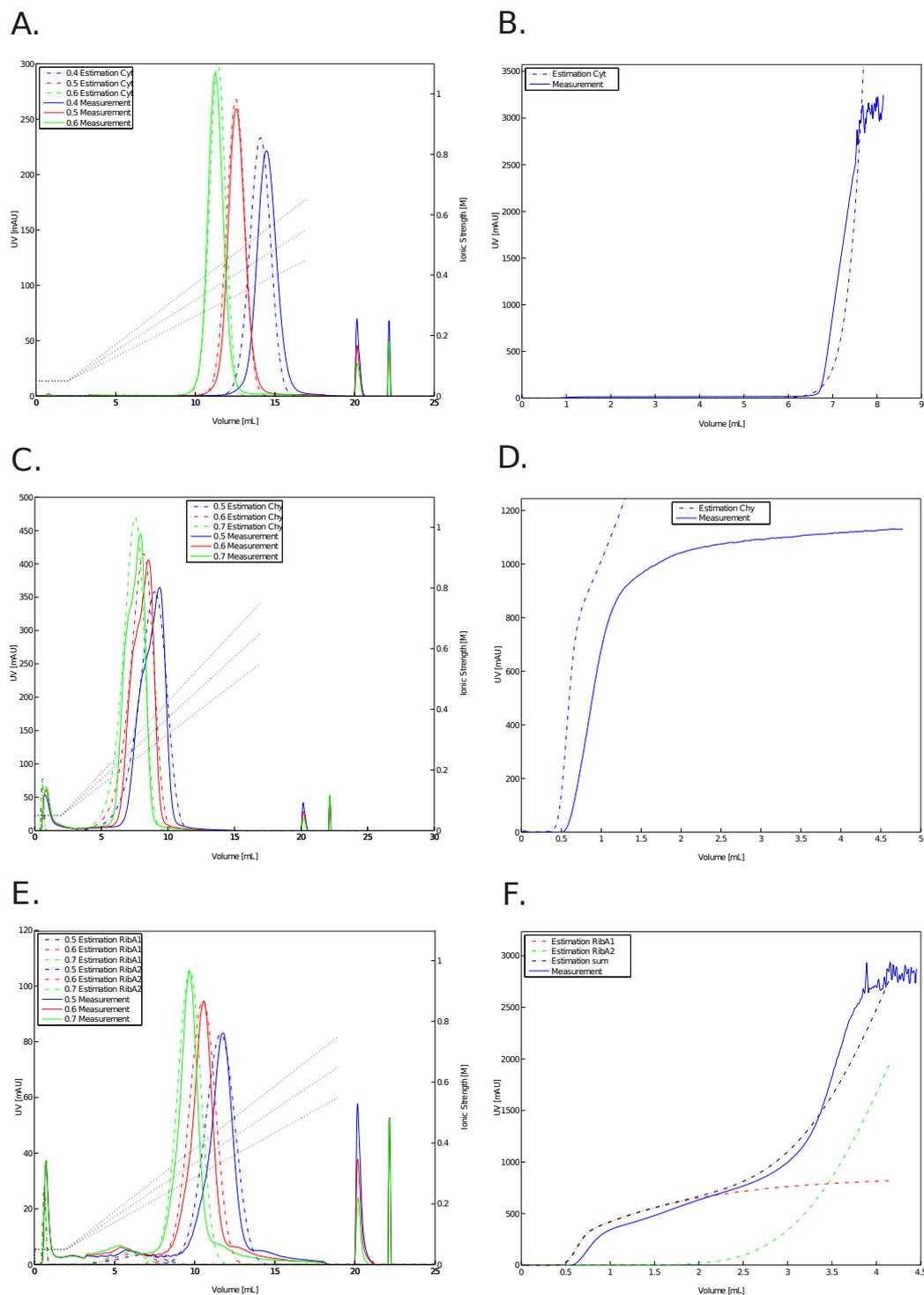
Figure 6.2: Results of the estimation of the isotherm parameters for Poros 50HS. A., C., and E. show the measured (solid lines) and estimated (dashed lines) data for the linear gradient elution experiments. B., D., and F. present the corresponding breakthrough curve data. A. and B. represent the cytochrome c data, C. and D. the chymotrypsin data, and E. and F. the ones for ribonuclease A.
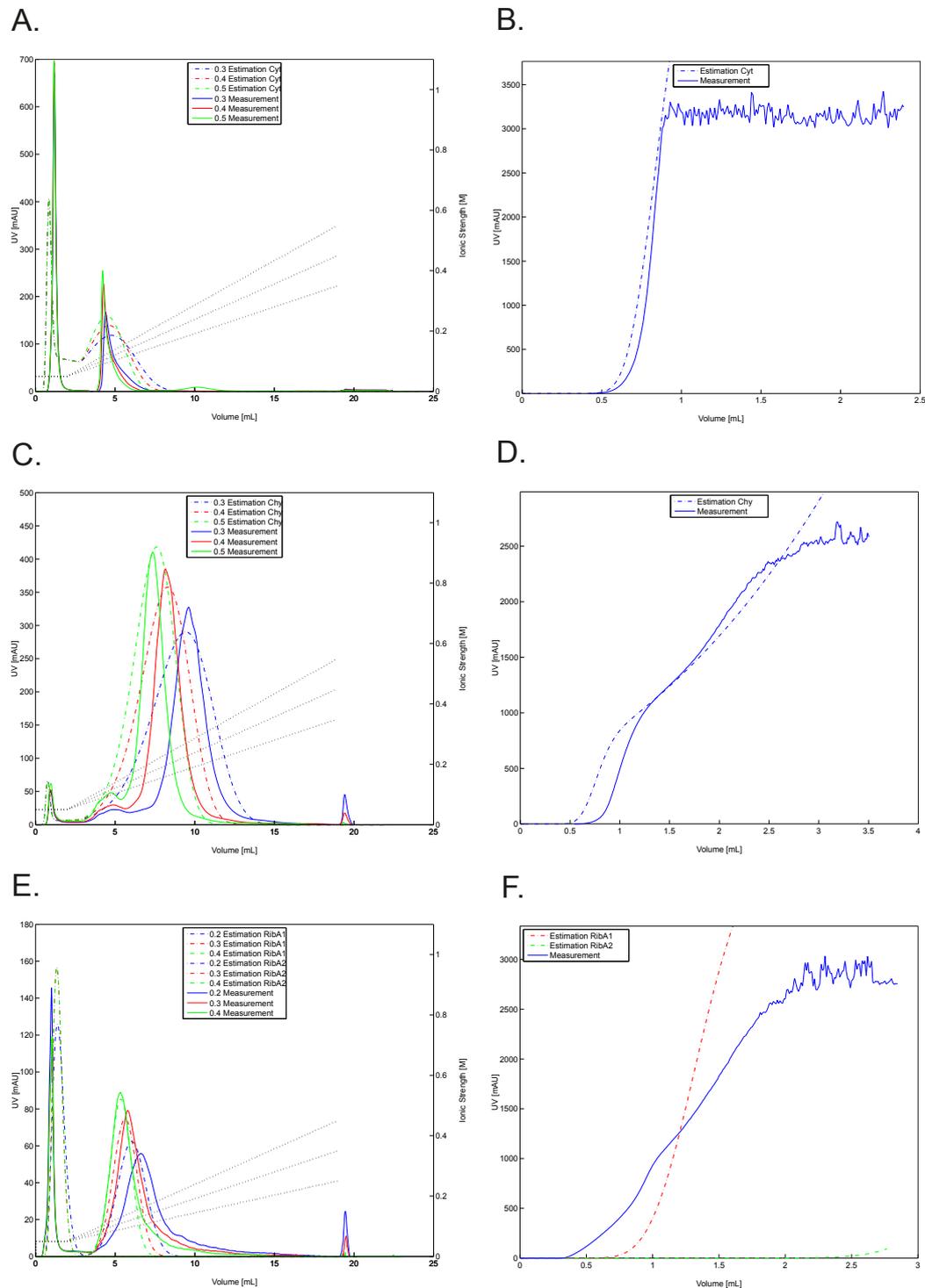
Figure 6.3: Results of the estimation of the isotherm parameters for Q Sepharose FF. A., C., and E. show the measured (solid lines) and estimated (dashed lines) data for the linear gradient elution experiments. B., D., and F. present the corresponding breakthrough curve data. A. and B. represent the cytochrome c data, C. and D. the chymotrypsin data, and E. and F. the ones for ribonuclease A.

Table 6.2: Estimated isotherm parameters for cytochrome c (Cyt), chymotrypsin (Chy) and the two ribonuclease A components (RibA1 and RibA2) on the Poros 50HS and the Q Sepharose FF column.

Poros 50HS		Cyt	Chy	RibA1	RibA2
kinetic	k_{kin}	0.054	0.049	0.1	0.203
equilibrium	k_{eq}	497.327	32.213	3.128	55.527
charge	ν	4.447	4.575	3.673	3.128
shielding	σ	5.001	0.1	53.245	2.964

Q Sepharose FF		Cyt	Chy	RibA1	RibA2
kinetic	k_{kin}	0.3	0.465	2	0.067
equilibrium	k_{eq}	0.009	0.033	3.722	0.22
charge	ν	2.25	2.61	0.066	1.545
shielding	σ	1	16.131	40.441	15.842

6.5.3   *In silico* Process Optimization

6.5.3.1   Concerted Flowsheet Optimization

After characterization of the Poros 50HS and Q Sepharose FF column with respect to the retention behavior of the model proteins, given in the estimation section, the isotherm parameters were used for *in silico* optimization of the process sequence. In this sequence, the salt gradients' starting point in time and concentrations and the gradient slopes and lengths were optimized within predefined ranges for both columns (for details ref. to Fig. 6.1). In addition, the boundaries of fraction collection were optimized for both columns. The fraction collected from the final column was evaluated with respect to the objective function given in Eq. (6.7). The results of concerted optimization after 2300 iterations using a genetic algorithm are given in Figs. 6.4 and 6.5 for the two process flowsheets CEX→AEX and AEX→CEX, respectively. Subfigure A illustrates the optimization approach, Table B presents the optimized parameters. The chromatograms of the optimized IEX are given in C and D for the first and second column, respectively. The gray boxes indicate the boundaries for fraction collection. E presents the protein amounts injected to the first column and the ones collected from the first and second column. Table F summarizes the purity, yield, and fraction volumes for both IEX and gives the calculated objectives.

The concerted optimization of the CEX→AEX process flowsheet leads to a steep gradient starting at 65s at 196mM NaCl and ending at 627s at 992mM (Fig. 6.4 C). In the optimized fraction collected from the CEX, chymotrypsin and ribonuclease A1 can be separated from the target protein cytochrome *c*. Ribonuclease A2 can only be partly separated from the target protein. The fraction collected from the CEX is virtually buffer exchanged to the buffer of the AEX, the protein amount is doubled, and the sample is injected to the AEX. The optimized parameters for the AEX represent a flat gradient starting at 237s at 68mM NaCl and ending at 2378s at 394mM (Fig. 6.4 D). The fraction collection from the AEX starts with the first increase of the cytochrome *c* trace and stops fractionation before the elution of the remaining contaminant ribonuclease A2. The cytochrome *c* purity can be

increased from initially 25% to nearly 90%, accepting a loss of about 45%. The objective function which has to be minimized starts at a value of 0.75 for the initial sample, decreases after the CEX to 0.53, and finally gives a value of 0.36.

The concerted optimization of the AEX→CEX process flowsheet leads to a gradient starting at 151s at 93mM NaCl and ending at 1522s at 571mM (Fig. 6.5 C). In the optimized fraction collected from the AEX, chymotrypsin and ribonucease A2 can be partly separated from the target protein cytochrome *c*. Ribonuclease A1 cannot be separated from the target protein. The fraction collected from the AEX is virtually buffer exchanged to the buffer of the CEX, the protein amount is doubled, and the sample is injected to the CEX. The optimized parameters for the CEX represent a steep gradient starting at 83s at 91mM NaCl and ending at 551s at 979mM (Fig. 6.5 D). The boundaries for fraction collection from the CEX are capable to separate chymotrypsin and ribonuclease A1 almost completely from the target protein. The amount of residual ribonuclease A2 remains almost constant. The cytochrome *c* purity can be increased from initially 25% to nearly 66%, accepting a loss of about 31%. The objective function which has to be minimized starts at a value of 0.75 for the initial sample, falls in quality after the AEX to 0.78, and finally gives a value of 0.60.

Comparing the two process flowsheet options, the CEX→AEX sequence leads to an objective of 0.36 and the AEX→CEX sequence to an objective of 0.60. Therefore, the CEX→AEX sequence is superior to the alternative IEX arrangement based on the given objective function. In the following Section 6.5.3.2, the found process optimum for the CEX→AEX sequence using a concerted optimization approach is compared to the same IEX arrangement, however using a sequential optimization approach.

6.5.3.2 Sequential Process Optimization

In the sequential process optimization approach given in Fig. 6.6 A, the CEX is optimized separately. Afterwards, the optimal fraction collected from the CEX is subjected to the AEX optimization.

The sequential optimization of the CEX→AEX process flowsheet was carried out for 2300 iterations on the CEX and the same number of iterations for the AEX to ensure comparability to the concerted optimizations with 2300 iterations for the whole process.

The optimization leads to a steep gradient starting at 50s at 195mM NaCl and ending at 724s at 915mM (Fig. 6.6 C). In the optimized fraction collected from the CEX, chymotrypsin and ribonuclease A1 can be separated from the target protein cytochrome *c*. Ribonuclease A2 can only be partly separated from the target protein. The fraction collected from the CEX is virtually buffer exchanged to the buffer of the AEX, the protein amount is doubled, and the sample is injected to the AEX. The optimized parameters for the AEX represent a flat gradient starting at 51s at 172mM NaCl and ending at 2279s at 234mM (Fig. 6.6 D). The fraction collection from the AEX exhibits the same purity as the fraction collected from the CEX (60%), but a lower yield of the target protein. The cytochrome *c* purity can be increased from initially 25% to nearly 60%, accepting a loss of about 27%. The objective function which has to be minimized starts at a value of 0.75 for the initial sample, decreases after the CEX to 0.46, and finally gives a value of 0.59. The decline of
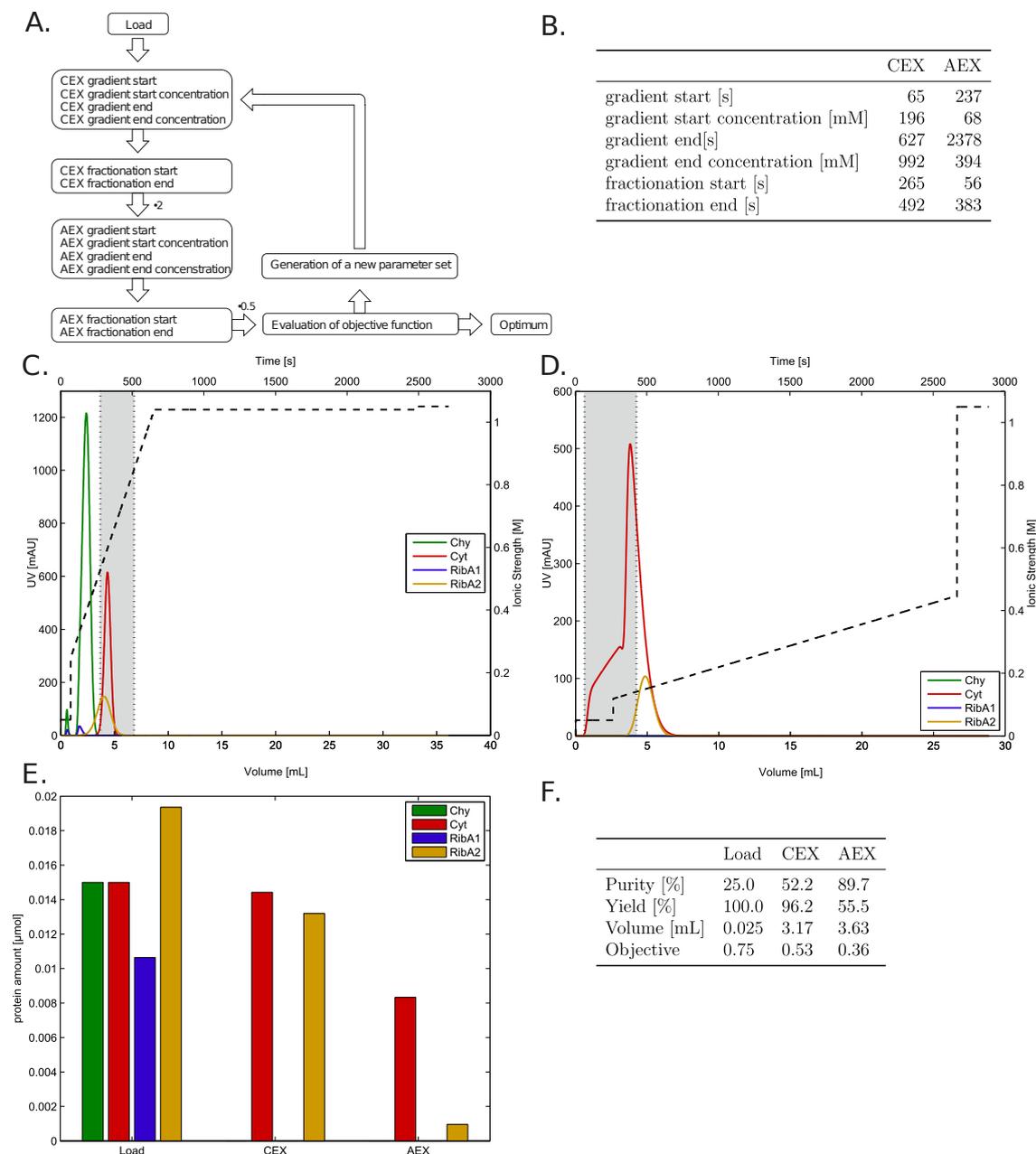
Figure 6.4: Results of the concerted *in silico* optimization of Poros 50HS (CEX) and Q Sepharose FF (AEX). The optimization principle is given in A. Table B summarizes the results of the parameter optimization. C. and D. present the corresponding chromatograms of the CEX and AEX, respectively. The gray boxes illustrate the boundaries of fraction collection. E. indicates the protein amounts injected to the first column and the ones collected from each column. Table F. summarizes the outcome of the process optimization.
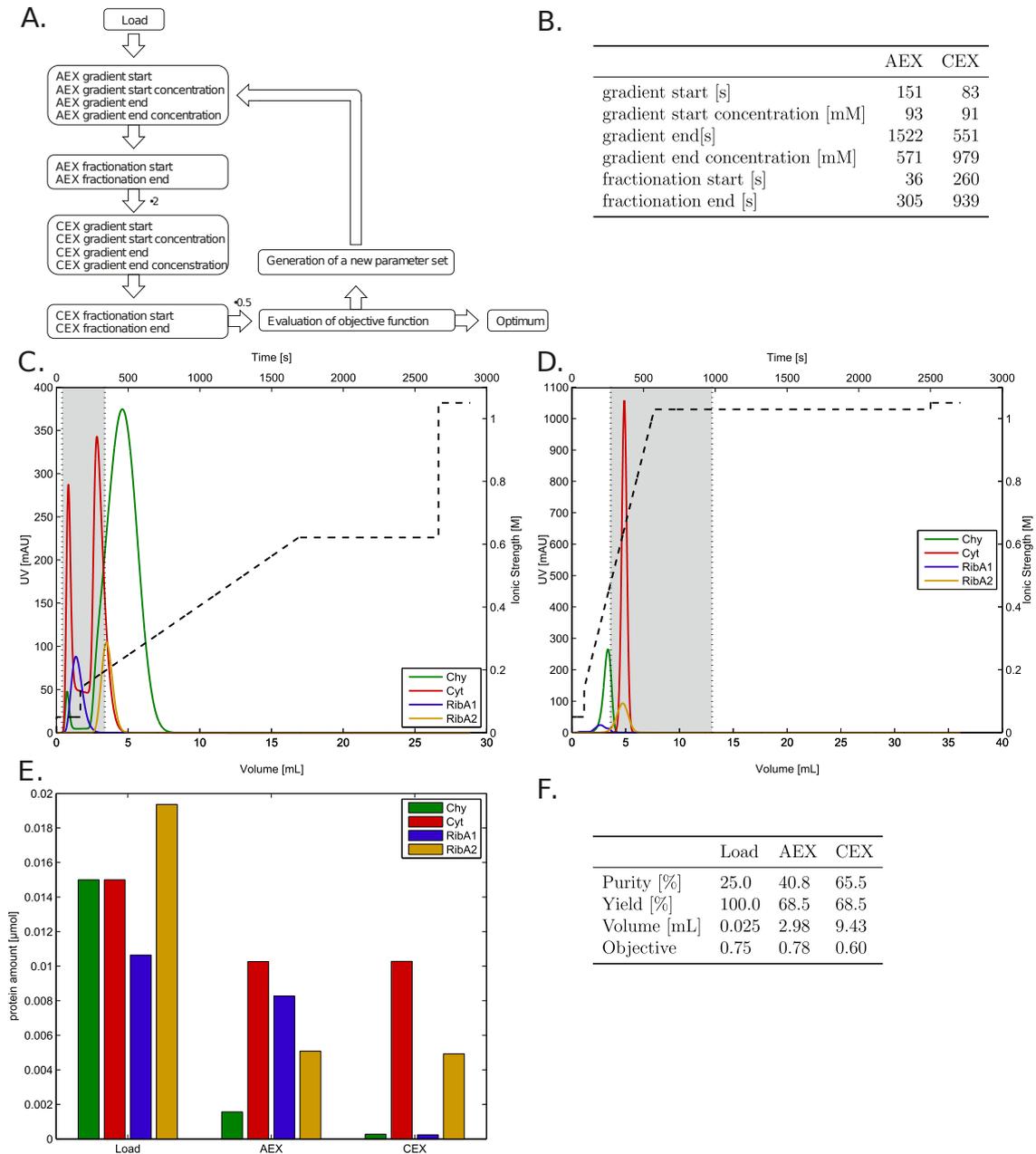
A.

	AEX	CEX
gradient start [s]	151	83
gradient start concentration [mM]	93	91
gradient end[s]	1522	551
gradient end concentration [mM]	571	979
fractionation start [s]	36	260
fractionation end [s]	305	939

	Load	AEX	CEX
Purity [%]	25.0	40.8	65.5
Yield [%]	100.0	68.5	68.5
Volume [mL]	0.025	2.98	9.43
Objective	0.75	0.78	0.60

Figure 6.5: Results of the concerted *in silico* optimization of Q Sepharose FF (AEX) and Poros 50HS (CEX). The optimization principle is given in A. Table B summarizes the results of the parameter optimization. C. and D. present the corresponding chromatograms of the AEX and CEX, respectively. The gray boxes illustrate the boundaries of fraction collection. E. indicates the protein amounts injected to the first column and the ones collected from each column. Table F. summarizes the outcome of the process optimization.

process performance from CEX to AEX is due to the failure to enhance the product purity on the final column, while accepting a loss of product. From a practical point of view, the process would have been stopped after the initial CEX, but this case study focuses on the comparison of a concerted and a sequential multi-column process optimization.

In comparison with the concerted optimization approach of the CEX→AEX process flowsheet which gave an objective of 0.36, the sequential optimization shows an objective of 0.59. Therefore, the concerted optimization approach achieved a superior process in comparison to the sequential approach, based on the given objective function.

6.5.4   Experimental Process Evaluation

To validate the best *in silico* chromatographic sequence, namely Poros 50HS→ Q Sepharose FF, this process is reproduced in lab experiments. In the *in silico* process, the buffer is exchanged virtually, the salt concentration is adjusted to the initial condition of the QFF, and the protein amount is doubled. In the experimental evaluation, the Poros operation is carried out twice, the two fractions are pooled and the buffer is exchanged using a UF/DF spin filter. Afterwards, the sample is injected to the QFF column. Fig. 6.7 presents the experimental results of the reproduced sequence. The blue line corresponds to the 280nm protein sum signal. The red 527nm signal equates to cytochrome *c*. The ratio of high salt buffer applied to the column entrance and the conductivity trace recorded at the column outlet are given by the dashed and dotted black curves. Fig. 6.7 A presents the experimental results of the Poros 50HS run, corresponding to the *in silico* optimum given in Fig. 6.4 C. In the experimental chromatogram, there is a minor flow-through fraction of proteins apart from cytochrome *c*. The *in silico* optimum reveals a fraction of chymotrypsin and ribonuclease A1 being apparent in the flow-through. The *in silico* optimization predicts the majority of proteins, eluting within two peaks in the increasing salt gradient. The first one contains chymotrypsin, ribonuclease A1 and a minor fraction of ribonuclease A2. The second peak, which corresponds to the *in silico* collected fraction, equates to the residual ribonuclease A2 and the target component cytochrome *c*. The experimental validation of this prediction is given in Fig. 6.7 A. In the validation, the two peaks are less resolved than in the simulation. The experimental peak size and the distribution of the target component cytochrome *c* is in accordance with the model prediction. The boundaries for fraction collection cover the cytochrome *c* peak.
The buffer of the collected fraction is exchanged with a spin filter and the doubled protein amount is injected to the Q Sepharose FF column. The *in silico* and experimental chromatograms are given in Figs. 6.4 D and 6.7 B, respectively. The model predicts that about half of the cytochrome *c* flows through the column without binding. The residual cytochrome *c* co-elutes with the remaining contaminant ribonuclease A2 at the beginning of the salt gradient. The experimental validation revealed greater deviations from the model prediction, but the overall peak composition still matches. Cytochrome *c* elutes in the flow-through and at the beginning of the salt gradient. The amount of protein besides cytochrome in the second peak is still predominant. The elevated baseline in Fig. 6.7 B results from the 1-methylpiperazine buffer and the increased salt concentration. However, the peak resolution and shape do not match to the model predictions acceptably.
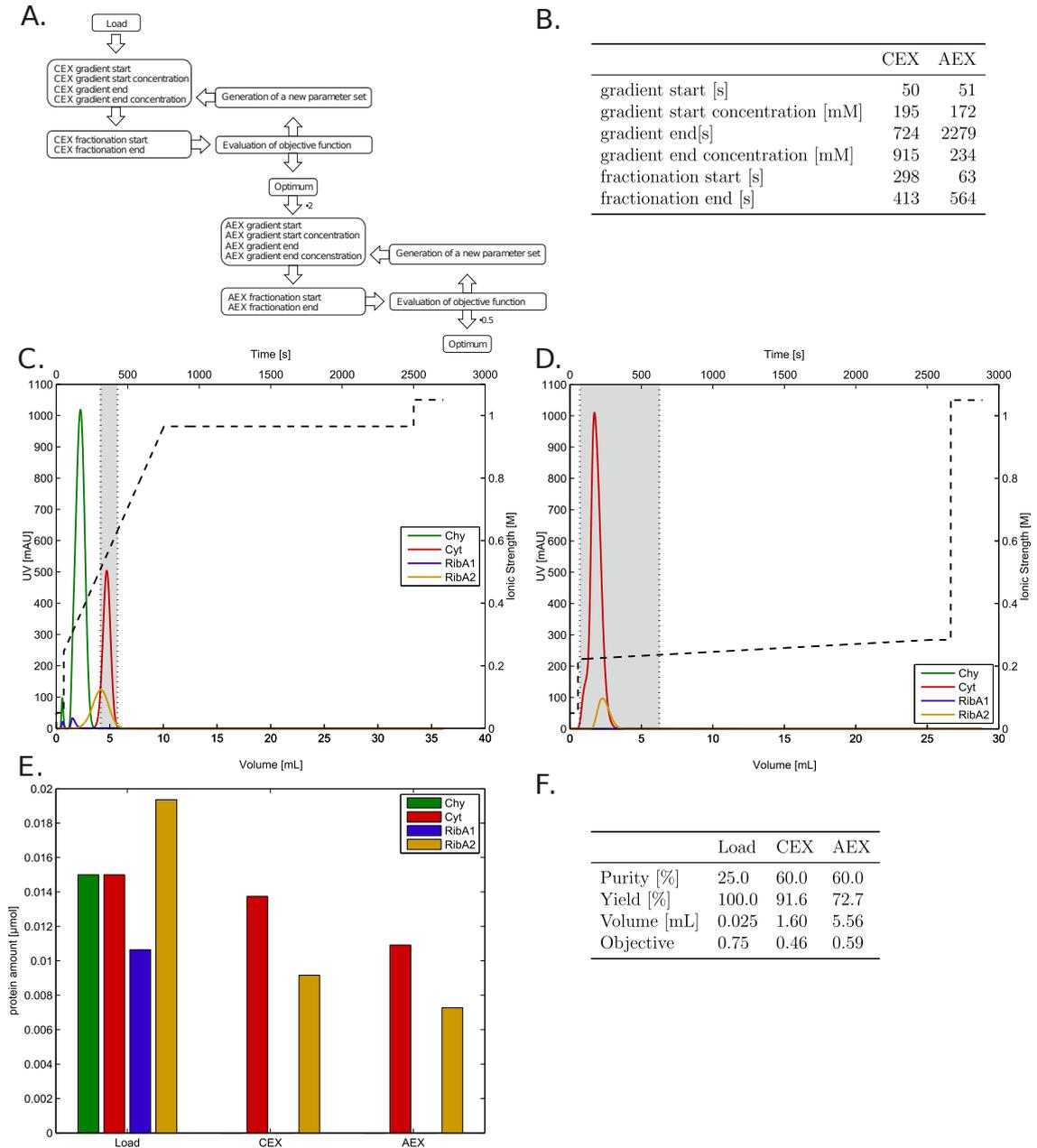
Figure 6.6: Results of the sequential *in silico* optimization of Poros 50HS (CEX) and Q Sepharose FF (AEX). The optimization principle is given in A. Table B summarizes the results of the parameter optimization. C. and D. present the corresponding chromatograms of the CEX and AEX, respectively. The gray boxes illustrate the boundaries of fraction collection. E. indicates the protein amounts injected to the first column and the ones collected from each column. Table F. quantifies the outcome of the process optimization.
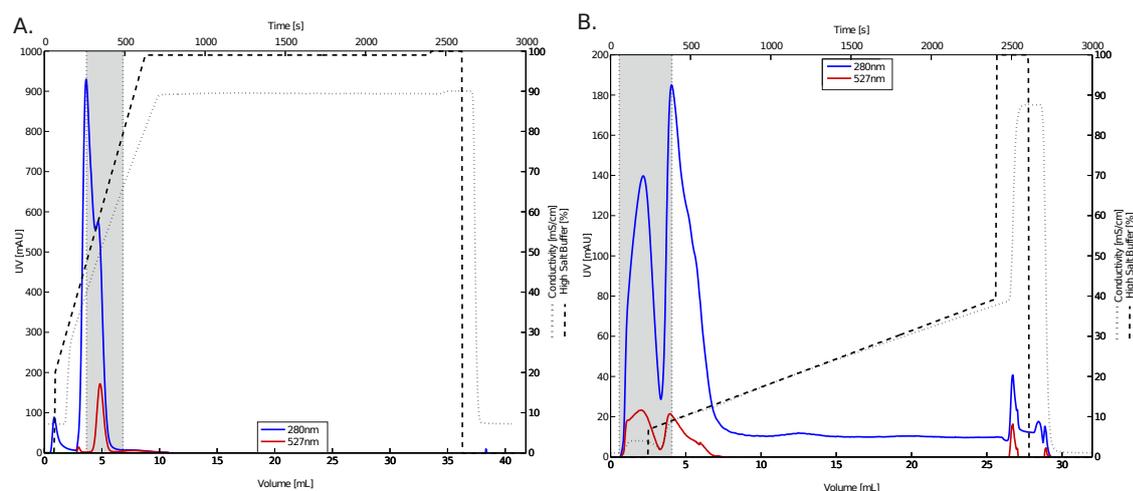
Figure 6.7: Experimental validation of the *in silico* optimized process. A presents the Poros 50HS and subfigure B the Q Sepharose FF chromatogram. The blue 280nm signal corresponds to the sum of proteins, the red 527nm signal is cytochrome *c* specific. The dashed and dotted black curves represent the ratio of high salt buffer applied to the column inlet and the conductivity trace, detected at the column outlet. The gray boxes indicate the boundaries for fraction collection.

6.5.5   Model-based Error Analysis

To evaluate the reasons for this mismatch between experimental data and the model prediction in AEX, the conductivity traces in Fig. 6.7 B were examined in more detail. Fig. 6.8 A presents the ratio of high salt buffer at column inlet (dashed line) and the conductivity trace recorded at the outlet (dotted line) in the experimental validation. The dotted conductivity trace exhibits an increase during the sample injection from about one to four mS/cm. The *in silico* optimized AEX step given in Fig. 6.4 D, does not account for this elevation of conductivity. To correct the model with respect to the experimental reality, the salt concentration of the sample injected to the Q Sepharose FF column, was *in silico* estimated and adjusted to $30mM$ NaCl. The resulting model prediction of the ionic strength traces is given in Fig. 6.8 B. It is obvious that the introduction of $30mM$ NaCl to the injected sample, leads to a correction of the predicted traces. The effect of the modified salt concentration to the *in silico* chromatogram is given in Fig. 6.8 C. For clarity, only the two remaining major proteins, cytochrome *c* and ribonuclease A2 are shown. The appearance of two poor-resolved peaks, both containing cytochrome *c* and just the latter one containing ribonuclease A2, matches the model prediction better.

6.6   Discussion

The model-based concerted optimization showed the superiority of the process flowsheet combining CEX→AEX as compared to the AEX→CEX process. Within the CEX→AEX process, the Poros 50HS exhibits a steep salt gradient. The majority of contaminants elute in the flow-through after sample injection and at the beginning of the salt gradient. The optimized boundaries for fraction collection match the peak of the target component cytochrome *c*. The collected fraction contains residual ribonuclease A2, which is difficult
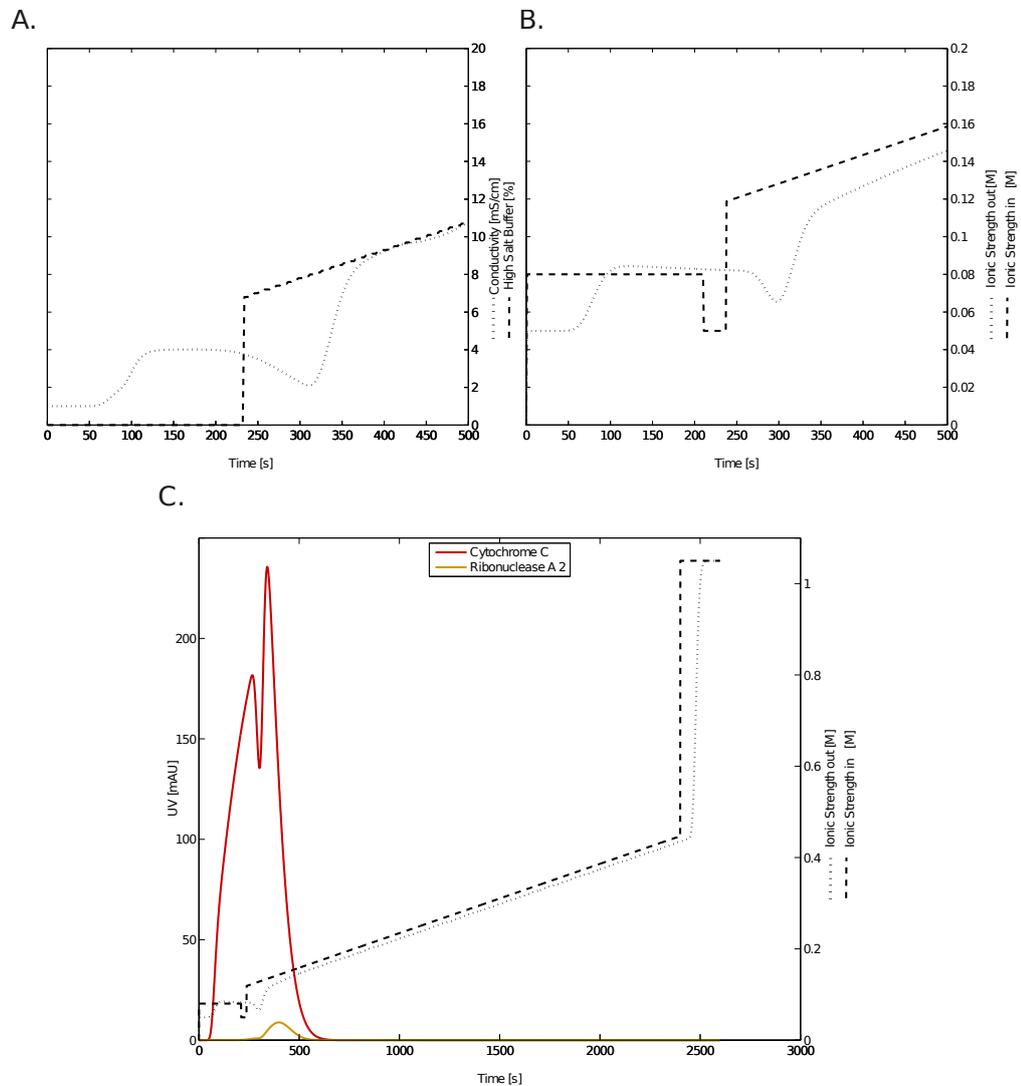
Figure 6.8: Evaluation of an elevated ionic strength of the protein sample, injected to the Q Sepharose FF column. Subfigure A presents the experimental traces for the ratio of high salt buffer applied to the column inlet (dashed line) and the conductivity trace recorded at the column outlet (dotted line). Subfigure B shows the effect of an elevated ionic strength of the injected protein sample. The dashed line shows the *in silico* ionic strength at the column entrance, the dotted line the *in silico* ionic strength at the column outlet. Subfigure C presents the effect of an elevated ionic strength of the injected sample to the chromatogram recorded *in silico* from Q Sepharose FF.

to separate from the target protein due to similar physico-chemical properties. In the subsequent AEX, the cytochrome c purity can be increased from 52% to nearly 90%, accepting a loss of half of the target species. After the final AEX, the achieved objective is 0.36 in comparison to the inferior AEX→CEX process with an objective of 0.6. The optimized process flowsheet is strongly influenced by the composition of the objective function and the introduced weighting factors. If protein purity or a high yield is the more important attribute, this can be considered by the modification of weighting factors.

To prove the necessity of a concerted process optimization, we reproduced the best column arrangement found in the concerted optimization within a sequential process optimization. In the sequential approach, the CEX was optimized independently and afterwards, the found optimum was used in a subsequent optimization of the AEX (for more details, ref. to Fig. 6.6 A).

In comparison to the concerted CEX→AEX process with a purity of 52%, a yield of 96%, and an objective of 0.53 after the initial CEX, the sequential optimization results in a purity of 60%, a yield of 92%, and an objective of 0.46 after the CEX. Therefore, the optimization of the CEX solely is superior to the intermediate result of the concerted process optimization. This finding coincides with the general expectation, because the concerted optimization intends to achieve a minimal objective after the final AEX, disregarding the quality of intermediate results like the fraction collected from the CEX. After the final AEX, the concerted optimization leads to a purity close to 90%, a yield of 56%, and an objective of 0.36. In contrast, the sequential optimization results in a final purity of 60%, a yield of 73%, and an objective of 0.59. Therefore, the result of the concerted optimization for the final AEX is superior to the one found for the sequential optimization. In addition, the objective for the concerted optimization is also superior to the intermediate result of the sequential optimization. This finding underlines the impressive capability of the concerted process optimization approach to avoid an over- or under-optimization of single intermediate states, as occurred for the sequential optimization. The latter one over-optimizes the intermediate CEX, indeed achieving an excellent intermediate result, but being clearly inferior when looking at the overall process performance.

The number of 2300 iterations in the optimization algorithm was kept constant for the two concerted processes. For the sequential process optimization, 2300 iterations per single IEX were used. Therefore, comparability of the found optima should be given with respect to an equal computational effort. However, it remains unclear whether the late termination of fraction collection in the sequential AEX (ref. Fig. 6.6 D) in comparison with the concerted CEX→AEX process (ref. Fig. 6.4 D) is coincidental.

To verify the *in silico* process optimization, the best process, namely the concerted optimized sequence of CEX→AEX, is reproduced in lab experiments.

The experimental reproduction of the *in silico* optimum for the CEX matches the model-derived data. The model predicts the flow-through and the two peaks within the salt gradient accurately. Minor differences occur with respect to the resolution of the two major peaks.

The *in silico* optimization of the subsequent Q Sepharose FF leads to a flat salt gradient elution profile. The contaminants exhibit a slightly greater affinity to the resin. About half of the target component cytochrome c elutes in the flow-through during sample injection.

The flow-through is collected as target fraction. The residual cytochrome c and the major contaminant ribonuclease A2 co-elute at the begin of the salt gradient. The co-eluting proteins are discharged due to the requirement of cytochrome purity in the objective function. The experimental validation of the optimized AEX operation reveals aberrations from the model prediction. In the experimental chromatograms, the flow-through peak and the one eluting at the beginning of the salt gradient achieve a better peak separation. In conjunction with this finding, there are differences in the overall peak shape.

We attribute the accurate prediction of the CEX and the difficulties with the AEX to the poorer model parameter estimation for the QFF in contrast to the Poros 50HS runs. The assumption that e.g. ribonuclease A consists of just two different species simplifies the more heterogeneous composition of ribonuclease species. This simplification has a negative impact on the quality of the estimation. In analytical chromatographic runs of the different model proteins, in AEX there are more protein peaks visible (data not shown). In addition, effects of propagation of uncertainty have to be considered. In practice, there are minor experimental process uncertainties apparent during the Poros 50HS run. These inaccuracies, e.g. in the collection of the target fraction, will propagate during the second column. To account for this finding, a term considering the process robustness or insensitivity against minor process changes might be introduced into the objective function in future experiments.

In addition to this general difficulty of considering the propagation of uncertainties [20; 148], we review the model-based approach to analyze the impact of potential contributors to the inaccurate fit of the experimental to the *in silico* data. The conductivity traces from the AEX validation experiments revealed an increase from one to four mS/cm during the sample injection(ref. Fig 6.7 B). The *in silico* process optimization does not exhibit this increase in conductivity during sample injection (ref. Fig. 6.4 D), because it assumes a complete buffer exchange in between the two IEX. To prove that the incomplete exchange of buffer during the experimental process validation is the major contributor to the uncertainty of the modeling approach, we estimated and adjusted the protein sample injected to the AEX *in silico* to a NaCl concentration of $30mM$. Fig. 6.8 presents the consequence of an elevated salt concentration of the sample injected to the AEX onto the AEX chromatogram. The increased salt concentration shifts the elution profile of cytochrome c in a manner that the *in silico* chromatogram and the one obtained during the experimental validation achieve an excellent degree of similarity. This finding indicates that the application of model-based tools within protein purification tasks is not only restricted to process development, but also capable of identifying critical process parameters or critical unit operations. Such an approach could probably support a QbD-driven approval of pharmaceuticals [20; 35; 204]. In the presented *in silico* optimized model process, the imperfectness of buffer exchange in between the two IEX seems to be the most critical operation to be considered. In comparison to classical process development using DoE, the overall sample consumption is reduced. For the three linear gradient elutions, $25\mu L$ of a $0.6mM$ protein solution were required per gradient ($1.2mM$ for ribonuclease A). The breakthrough curves (BC) constitute the major contributor to the overall protein consumption, as volumes from 5 to $10mL$ protein solution per BC were required. The BC is only necessary to estimate the shielding parameter σ within the SMA isotherm, so BCs are only required if the model has to extrapolate to the non-linear adsorption range of the isotherm. Otherwise, the protein consumption can be reduced extremely.

In addition, the time requirement can probably be minimized in contrast to classical DoE approaches to process optimization. An additional advantage of model-based process optimization in contrast to DoE approaches is that the lab experiments and parameter estimation have to be carried out just once. Testing, for example, different objective functions or column arrangements only demands multiple *in silico* process optimization cycles. In contrast, classical DoE approaches would require additional lab experiments, e.g. to incorporate the different possibilities of salt gradient profiles.

The model system used in this case study consists of two pure proteins (cytochrome c and chymotrypsin) and ribonuclease A which was treated as a two-component mixture. Beside the estimation of two isotherm parameter sets from one chromatogram, the presented modeling approach is also capable of predicting more heterogeneous protein mixtures. The latter case is just dependent on adequate analytics for fraction analysis.

In this study we used the SMA isotherm to model the interaction of the proteins with the adsorber surface. The use of a 'simpler isotherm' such as Langmuir is not practical because of the need to respect salt gradients. The SMA isotherm would reduce to Langmuir isotherm for $c_{salt} = const.$ and $\nu = 1$. According to Parente and Wetlaufer [152] $\nu = 1$ implies that the retention volumes of the species would have to fulfill $V_r \cdot (c_{gradient,end} - c_{gradient,begin}) = const.$ This is not the case in our results. Furthermore, some of the breakthrough experiments show a short steep ascent followed by a slower rise (Figs. 6.2 F, 6.3 D). This behavior could only be modeled with slow kinetics (steep ascent) and non-zero shielding parameter (slower rise). Equilibrium models or a pure stoichiometric displacement model would not have been sufficient.

6.7   Conclusion

We presented an approach to a model-based integrated downstream process optimization, considering a flowsheet optimization, the salt gradient elution profile, and the boundaries for fraction collection for two subsequent IEX steps. The modeling and process optimization approach was successfully applied to the task of global optimization of the chromatographic operations. The systematic comparison of sequential and concerted process optimization proved that a concerted process optimization approach prevents over- or under-optimization of single-unit operations. The time and effort to calibrate a model for concerted or two models for sequential optimization is identical.

The drawback of this concerted optimization approach is the impossibility to do the computation for the two columns in parallel, e.g. by the distribution across multiple computers, to reduce the overall time consumption.

Difficulties with the intermediate operation of exchanging the buffer system highlighted the need for an integration of such operations into the modeling approach. In future, intermediate operations such as ultra-/diafiltration will be integrated into our modeling approach, e.g. by inclusion of black-box models in between the chromatographic operations.

In addition, we applied our modeling tool to the evaluation of the impact of an imperfect buffer exchange operation in between the two IEX. The presented methodology might be

a useful tool in QbD-based approvals of biopharmaceuticals, especially for identification and characterization of critical process parameters.

The presented optimization approach using an in-house developed software toolbox is extendable to greater process sequences, including commonly used platforms in DSP. Furthermore, our modeling approach and the implemented software are capable of handling larger numbers of protein species (e.g. mAb heterogeneity or sets of host cell proteins) and are not restricted to small-scale chromatographic columns. Concluding, model-based optimization is open to coping with industrial downstream process development.

Acknowledgment

7 | UV Absorption-based Inverse Modelling of Protein Chromatography

Tobias Hahn[1], Pascal Baumann[1], Thiemo Huuk[1], Vincent Heuveline[2], Jürgen Hubbuch[1]

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

[2] Heidelberg University, Interdisciplinary Center for Scientific Computing, Engineering Mathematics and Computing Lab, Heidelberg, Germany

Abstract

UV absorbance measurements play an important role in bioprocess development. Yield and purity are often evaluated in terms of peak percentages in analytical size exclusion chromatography or ion-exchange chromatography. Also, industrial chromatography steps are usually controlled based on UV data with pooling decisions according to absorbance thresholds.

Model-based process development would make elaborate screening experiments redundant, once the model has been calibrated to the specific process step. So far, absorbance measurements could not be used directly for modelling chromatography steps as the commonly applied models rely on mass or molar concentration. This study presents mechanistic modelling of an industrially relevant chromatography setting without any knowledge of the feed composition. The model equations were re-written to employ boundary conditions in UV absorbance units, the absorption coefficients were shifted into the isotherm, and standard parameter estimation procedures could be applied. An anion-exchange chromatography case study of a target protein expressed in *Escherichia coli* and eleven lumped impurity peaks demonstrated practical applicability. The target protein concentration in the feed material was estimated from chromatograms. Using this method, initially unknown feed concentrations can be determined a posteriori for ion-exchange and multi-modal chromatography from single-component absorbance curves.

Practical Application

This study explores the feasibility of modelling ion-exchange chromatography without knowledge of feed composition in terms of molar or mass concentration. This is especially valuable in early-stage process development when no information is available on the impurities. It was shown that all model parameters can be determined uniquely from single-component elution curves. Here, the concentration of Cherry-tagged glutathione-S-transferase within a crude feedstock could be determined from chromatograms at a particular wavelength.

7.1   Introduction

Industrial downstream processing (DSP) faces the challenge of efficiently purifying a product out of a very heterogeneous mixture. The purification sequence is commonly based on platform processes that are only slightly adapted to new target components to accelerate process development [115; 171]. Prior to this, high-throughput screening methods are often used to find promising initial conditions for platform processing [19]. Mechanistic modelling is a favorable alternative, provided the model parameters can be determined with less effort, allowing identification of optimal process parameters *in silico*.

The common models in liquid chromatography describe the mass transport in the column by so-called Convection Diffusion Reaction (CDR) equations, where the reaction term models phase transitions and, eventually, the retention of the species. With no a priori knowledge about the components' behavior, the inverse method is a suitable option which alters parameters in a systematic fashion to achieve a match of the recorded chromatogram and the model prediction.

The potential of mathematical modelling and numerical optimization of chromatography has already been demonstrated in academic set-ups for mixtures of small molecules [1; 53], model proteins [150] and antibodies [40; 99], and also for industrial process steps [35; 133]. All applications have in common, that molar concentrations in the feed were known and sensor calibrations existed for all components.

Cornel et al. [37] determined absorption coefficients for a two-component mixture directly from the simulated concentration curves by choosing the best fitting values in each iteration of the estimation procedure. No sensor calibration was necessary, but mass concentrations in the feed were known a priori.

In the following sections, we describe a mechanistic modelling approach for an industrially relevant chromatography setting that does not require prior knowledge of the feed composition in terms of molar or mass concentrations. We re-write the model equation for boundary conditions in UV absorbance units and aim at determining the unknown feed concentrations a posteriori by taking advantage of the particular structure of stoichiometric exchange models. Standard parameter estimation procedures can be applied if single-component absorption curves are available.

A case study based on an anion-exchange chromatographic (AEX) process step (Q Sepharose FF, GE Healthcare) demonstrates the applicability. The mixture fed into AEX is a crude feed stock of *Escherichia Coli SE 1*, including the Cherry-tagged enzyme Glutathione-S-Transferase as the product.

Table 7.1: Measured column parameters.

Parameter	Symbol	Value	Unit	Proceeding
Diameter	d	7	mm	From manufacturer
Length	L	25	mm	From manufacturer
Bead radius	r_p	0.045	mm	From manufacturer
System dead volume	V_d	0.07	ml	Acetone injection without column
Retention volume Acetone	V_{RetAc}	0.96	ml	Acetone peak injection with column
Retention volume dextran	V_{RetDex}	0.34	ml	Dextran peak injection with column
Standard deviation of dextran	σ_{Dex}	0.029	ml	Äkta peak integration
Volume of HCl	V_{HCl}	1.48	ml	Acid/base titration
Molarity of HCl	c_{HCl}	0.01	M	Manually controlled
Flow rate	u	0.2	$\frac{mm}{s}$	Manually controlled

7.2   Materials and Methods

7.2.1   Column Parameter Determination

To model the mass transport in the chromatography system, the column properties listed in Tables 7.1 and 7.2 must be determined by pulse injections of non-interacting tracer molecules [129]. A 1 ml column (effective volume 0.962 ml), with Q Sepharose Fast Flow Resin (GE Healthcare, Freiburg, Germany) was analyzed firstly with an 1 % acetone (Merck, Darmstadt, Germany) pulse and secondly with a dextran pulse from *leuconostoc spp.* MW 2,000,000 (Sigma Aldrich, Steinheim, Germany) using an Äkta Purifier system (GE Healthcare, Little Chalfont, UK) controlled with Unicorn 5.2 (GE Healthcare, Uppsala, Sweden) to determine the essential system parameters [129] presented in Table 7.1. Acid-base titration was carried out to determine the total ionic capacity: the column was flushed with a 0.5 M NaOH solution (Merck, Darmstadt, Germany) until a constant UV and conductivity signal was achieved. Afterwards, the column was washed with ultra pure water until a constant UV and conductivity baseline was reached. Then, the column was titrated at a flow of 0.64 ml/min with a 0.01 M HCl solution (Merck, Darmstadt, Germany) until an increase in the conductivity signal was recorded. From the Cl-ion concentration of the titrant and the volume of the applied titrant, the total number of exchangeable ions was calculated. All chemicals used were obtained in highest quality.

With this set of parameters, all system-specific parameters occurring in the mathematical model can be fixed as given in Table 7.2.

7.2.2   Sample Production

The applied sample consisted of an *Escherichia Coli SE 1* lysate, including Cherry-tagged Glutathione-S-Transferase (GST) as a product. The Cherry-tag, which can be fused to any

Parameter	Symbol	Value	Unit	Proceeding
Volume	V	0.962	ml	$\frac{1}{4} \cdot \pi \cdot d^2 \cdot L$
Fluid volume	V_f	0.89	ml	$V_{RetAc} - V_d$
Interstitial volume	V_{int}	0.27	ml	$V_{RetDex} - V_d$
Total column porosity	ε_t	0.925		$\frac{V_f}{V}$
Interstitial porosity	ε_b	0.28		$\frac{V_{int}}{V}$
Particle porosity	ε_p	0.896		$\frac{V_f - V_{int}}{V - V_{int}}$
Interstitial flow	u_{int}	0.714	$\frac{mm}{s}$	$\frac{u}{\varepsilon_b}$
Axial dispersion	D_{ax}	0.1	$\frac{mm^2}{s}$	$\frac{1}{2} \cdot L \cdot u_{int} \cdot \left(\frac{\sigma_{Dex}}{V_{int}} \right)^2$
Ionic capacity	Λ	0.22	M	$\frac{c_{HCl} \cdot V_{HCl}}{V(1-\varepsilon_b)(1-\varepsilon_p)}$

Table 7.2: Calculated column parameters.

target protein, allows for straightforward product analytics by VIS absorbance measurements [12]. The cultivation was performed for 24 h in 800 ml standard TB (terrific broth) medium at 37 °C and 180 rpm rotational speed in 2.5 l Tunair flasks (Sigma Aldrich). Cell disruption was performed by sonication of the cell pellet in 20 ml of 50 mM Tris buffer (pH 8), including 1X Sigma FAST protease inhibitor (Sigma Aldrich), in a Branson Digital Sonifier (70 % pulse amplitude, 10 X 15 s pulse duration, 30 s resting on ice between pulses). The lysate was centrifuged at 12000 rpm for 60 min at 10 °C using a 5810 R centrifuge (Eppendorf, Germany) followed by a second clarification step using 0.2 µm sterile PES filters (VWR, Germany). Finally, the permeate was 10 times diluted in 50 mM Tris buffer (pH 8).

7.2.3   Sample Characterisation

For comparison, the product concentration was determined in the Caliper LabChip GX II capillary gel electrophoresis system with LabChip GX 3.1 software (Perkin Elmer, Hopkinton, USA). The HT Protein Express and Pico LabChip was run with the HT Protein Express LabChip reagent kit using the HT Protein Express 200 assay. Cherry-tagged GST was identified using the sample ladder from the reagent kit. The product was quantified by peak-baseline integration of the fluorescence signals (Fig. 7.1) and scaling to an external lysozyme protein standard of 1 mg/ml.

7.2.4   Bind-elute Experiments

The component-specific isotherm parameters are determined from bind-elute experiments. The general approach is identical to concentration-based parameter estimation [40; 99; 150]. A 50 mM Tris-HCl buffer, pH 8 was employed as the mobile phase during binding and the same buffer supplemented with 1 M NaCl was used for elution. Different salt gradients were generated from these two buffers. After the 0.5 ml sample was injected (12.4 ml for the breakthrough experiment), the column was washed with low-salt buffer for 3 ml of 50 mM Tris-HCl buffer, before initiating linear gradient (0 to 1 M NaCl) elution over 5, 10, 15 and 20 ml. On gradient completion the columns were irrigated with 2 ml of 50 mM Tris-HCl + 1 M NaCl before re-equilibrating with 5 ml of 50 mM Tris-HCl buffer. The linear phase linear velocity employed was 0.2 mm/s throughout.
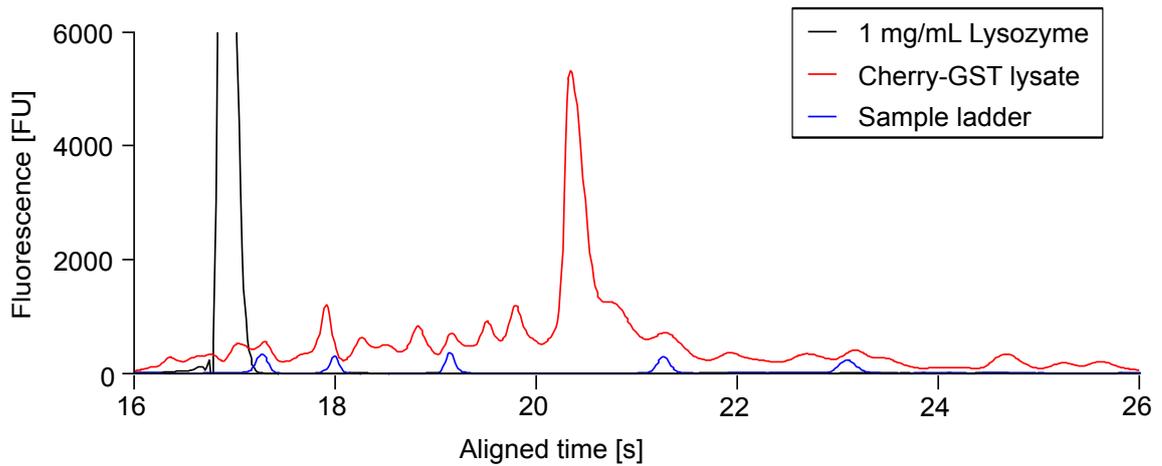
Figure 7.1: Capillary gel electrophoresis analysis of feed material.

7.2.5 Parameter Estimation

In general, estimation of an unknown parameter set $\overline{p}$ solves the least squares optimization problem

$$\min_{\overline{p}} \sum_{j} \left(m(t_j) - \sum_{i \geq 1} c_i(L, t_j; \overline{p}) \cdot a_i \right)^2, \tag{7.1}$$

where $m(t_j)$ is the measured chromatogram value at time t_j, typically given in milli absorbance units [mAU]. This measurement might also contain noise, which can be neglected when assuming the noise to be zero-mean Gaussian and isotropic.

$c_i(L, t)$ is the simulated mass or molar concentrations at the outlet of the column with length L. The transformation into absorbance units is performed with a scaling factor a_i. According to Beer's law, the absorption coefficient a_i consists of an extinction coefficient and UV cell path length. It is unknown in this case.

7.2.6 Chromatography Model

The Transport Dispersive Model (TDM) [129] in Eqs. (7.2) and (7.3) is used to model the macroscopic protein transport through the column. For simplicity, the component-specific internal and external diffusion effects are lumped in an effective mass transfer coefficient.

The system is of Convection Diffusion Reaction (CDR) type. Eq. (7.2) describes the rate of change of the concentration $c_i(x, t)$ of component i in the interstitial volume of a column with length L, which consists of convective mass transport in space with the average interstitial velocity of the fluid u. Peak broadening effects are modelled as dispersion in axial direction with respect to a coefficient D_{ax}. The exchange between the interstitial concentration and the particle pore concentration $c_{p,i}(x, t)$ depends on the porosity of the bed ε_b, the radius of adsorber particles r_p, and a component-specific effective mass transfer coefficient $k_{eff,i}$. The model is one-dimensional in space, such that the concentrations

depend on the axial position in the column and time. Equation (7.3) models the accumulation of mass in the pore volume $c_{p,i}$ and stationary phase q_i as a function of the particle porosity ε_p. The model is complemented by Danckwerts boundary conditions Eqs. (7.4), (7.5), including the applied inlet concentration $c_{in,i}$, and an isotherm equation modelling the stationary phase concentration q_i.

$$\frac{\partial c_i}{\partial t} = -u(t)\frac{\partial c_i}{\partial x} + D_{ax}\frac{\partial^2 c_i}{\partial x^2} - \frac{1-\varepsilon_b}{\varepsilon_b}k_{eff,i}\frac{3}{r_p}\left(c_i - c_{p,i}\right) \tag{7.2}$$

$$\frac{\partial c_{p,i}}{\partial t} = -\frac{1-\varepsilon_p}{\varepsilon_p}\frac{\partial q_i}{\partial t} + k_{eff,i}\frac{3}{\varepsilon_p r_p}\left(c_i - c_{p,i}\right) \tag{7.3}$$

$$\frac{\partial c_i}{\partial x}\left(0,t\right) = \frac{u(t)}{D_{ax}}\left(c_i\left(0,t\right) - c_{in,i}\left(t\right)\right) \tag{7.4}$$

$$\frac{\partial c_i}{\partial x}\left(L,t\right) = 0 \tag{7.5}$$

The steric mass action isotherm (SMA) [24] is a commonly used semi-mechanistic isotherm in ion-exchange chromatography. It is capable of reproducing the influence of counter ions on the retention behavior of protein species using the proteins' characteristic charges ν_i. In addition, it considers adsorber properties such as the total ionic capacity Λ and steric shielding effects σ_i of the protein covering an amount of binding sites, greater than the actual number of sites it interacts with. The kinetic SMA isotherm is given in Eq. (7.6) for k proteins, with q_i and $c_{p,i}$ being the concentration of the protein $i \in \{1,\ldots,k\}$ adsorbed and in solution, respectively. $c_{p,salt}$ is the salt concentration of the solution. $k_{ads,i}$ and $k_{des,i}$ are the constants of the adsorption and desorption rate.

$$\frac{\partial q_i}{\partial t} = k_{ads,i}\left(\Lambda - \sum_{j=1}^{k}(\nu_j + \sigma_j)q_j\right)^{\nu_i}c_{p,i} - k_{des,i}c_{p,salt}^{\nu_i}q_i \tag{7.6}$$

$$q_{salt} = \Lambda - \sum_{j=1}^{k}\nu_j q_j \tag{7.7}$$

The model is chosen because of its capability of simulating the whole chromatographic process, including elution, by changing the induced salt concentration at the inlet. The model is based on molar concentrations, such that the boundary conditions of the TDM must be set in terms of molarities.

7.2.7   Transformation

As the SMA model is based on molar concentrations [M], interstitial and pore volume concentrations must be given in [M] as well. This also applies to the boundary conditions. Here, the exact molar concentrations in the feed are unknown, as are the scaling factors for UV absorbance. We will re-write the equations to directly incorporate UV absorbance values.

First, the injected protein concentrations are transformed into absorbance values:

$$c'_{in}[mAU] = a\,[mAU/M] \cdot c_{in}[M]. \tag{7.8}$$

These can be determined later from the respective peak area in the chromatogram.

The equations for the interstitial and pore volume as well as the boundary conditions are linear in c. These can be multiplied by a to obtain

$$a_i \frac{\partial c_i}{\partial t} = a_i \left[-u \frac{\partial c_i}{\partial x} + D_{ax} \frac{\partial^2 c_i}{\partial x^2} - \frac{1 - \varepsilon_c}{\varepsilon_c} k_f \left(c_i - c_{p,i} \right) \right] \tag{7.9}$$

$$\Longleftrightarrow \frac{\partial c_i'}{\partial t} = -u \frac{\partial c_i'}{\partial x} + D_{ax} \frac{\partial^2 c_i'}{\partial x^2} - \frac{1 - \varepsilon_c}{\varepsilon_c} k_f \left(c_i' - c_{p,i}' \right) \tag{7.10}$$

$$\frac{\partial c_i'}{\partial x} \left(0, t \right) = \frac{u}{D_{ax}} \left(c_i' \left(0, t \right) - c_{in,i}' \left(t \right) \right) \tag{7.11}$$

$$\frac{\partial c_i'}{\partial x} \left(L, t \right) = 0 \tag{7.12}$$

We obtain equations for $c_i' = a_i \cdot c_i$ that use $c_{p,i}' = a_i \cdot c_{p,i}$. This is calculated from the scaled lumped rate model

$$\varepsilon_p \frac{\partial c_{p,i}'}{\partial t} + \left(1 - \varepsilon_p \right) \frac{\partial q_i'}{\partial t} = k_{eff,i} \frac{3}{r_p} \left(c_i' - c_{p,i}' \right). \tag{7.13}$$

Again, we require an equation for $q_i' = a_i \cdot q_i$. Scaling the kinetic SMA formulation yields

$$a_i \frac{\partial q_i}{\partial t} = a_i \left[k_{ads,i} \left(\Lambda - \sum_{j=1}^{k} \left(\nu_j + \sigma_j \right) q_j \right)^{\nu_i} c_{p,i} - k_{des,i} c_s^{\nu_i} q_i \right] \tag{7.14}$$

$$\Longleftrightarrow \frac{\partial q_i'}{\partial t} = k_{ads,i} \left(\Lambda - \sum_{j=1}^{k} \left(\nu_j + \sigma_j \right) q_j \right)^{\nu_i} c_{p,i}' - k_{des,i} c_s^{\nu_i} q_i' \tag{7.15}$$

$$\Longleftrightarrow \frac{\partial q_i'}{\partial t} = k_{ads,i} \left(\Lambda - \sum_{j=1}^{k} \left(\nu_j + \sigma_j \right) \frac{a_j}{a_j} q_j \right)^{\nu_i} c_{p,i}' - k_{des,i} c_s^{\nu_i} q_i' \tag{7.16}$$

$$\Longleftrightarrow \frac{\partial q_i'}{\partial t} = k_{ads,i} \left(\Lambda - \sum_{j=1}^{k} \frac{\nu_j + \sigma_j}{a_j} q_j' \right)^{\nu_i} c_{p,i}' - k_{des,i} c_s^{\nu_i} q_i'. \tag{7.17}$$

Here, we multiplied by $\frac{a_j}{a_j}$ in step (7.15)→(7.16) to transform the remaining q into q'. Essentially, we shifted the unknown scaling factor from the least-squares problem (7.1) into the isotherm.

We are left with Eq. (7.7), which can be altered to include q' as above:

$$q_{salt} = \Lambda - \sum_{j=1}^{k} \nu_j q_j \tag{7.18}$$

$$= \Lambda - \sum_{j=1}^{k} \nu_j \frac{a_j}{a_j} q_j \tag{7.19}$$

$$= \Lambda - \sum_{j=1}^{k} \frac{\nu_j}{a_j} q'_j. \tag{7.20}$$

The transformation procedure is also applicable to other isotherms with stoichiometric exchange, e.g. the mixed-mode isotherm in [142].

Binding models of Langmuir type and isotherms without an additional equation for counterions, e.g. for hydrophobic interaction chromatography (HIC) [133], can be treated as above, but due to the missing second equation, a will remain hidden within other constants. In case of the kinetic Langmuir isotherm, we obtain

$$\frac{\partial q_i}{\partial t} = k_{ads,i} q_{max,i} \left(1 - \sum_{j=1}^{k} \frac{q_j}{q_{max,j}} \right) c_{p,i} - k_{des,i} q_i \tag{7.21}$$

$$\Longleftrightarrow \frac{\partial q'_i}{\partial t} = k'_{ads,i} q'_{max,i} \left(1 - \sum_{j=1}^{k} \frac{q'_j}{q'_{max,j}} \right) c'_{p,i} - k_{des,i} q'_i, \tag{7.22}$$

with $k'_{ads,i} = k_{ads,i}/a_i$ and $q'_{max,i} = q_{max,i} \cdot a_i$. These parameters can be used for UV-based modelling but not for determining absorption coefficients and molar concentrations.

Other convection-diffusion models, such as the general rate model [129] or models of radial flow chromatography [65], are also linear in the concentration variables and can be treated as the TDM above.

7.2.8   Uniqueness

In this section, it is shown that the transformation does not affect the parameter determination.

The linear range of the isotherm is uninfluenced by the transformation. For $\Lambda \gg \sum_{j=1}^{k} \frac{\nu_j + \sigma_j}{a_j} q'_j$, we obtain

$$\frac{\partial q'_i}{\partial t} \approx k_{ads,i} \Lambda^{\nu_i} c'_{p,i} - k_{des,i} c_s^{\nu_i} q'_i. \tag{7.23}$$

Consequently, all methods to determine the linear SMA parameters k_{ads}, k_{des}, and ν can be employed here. It has been shown that the characteristic charge ν and equilibrium coefficient, defined as $k_{eq} = \frac{k_{ads}}{k_{des}}$, determine the retention time in gradient elution [169]. At least two gradients with different lengths and/or slopes are necessary to uniquely determine the two values. The kinetic parameter k_{des} can be identified from the peak shape [99]. This method is also applicable to multi-component settings, as no protein-protein interactions are assumed to happen in the linear range.

The non-linear parameter σ is typically determined by a frontal experiment or from batch isotherms. Both methods rely on determining the saturation capacity $q_{max} = \frac{\Lambda}{\nu + \sigma}$ and calculating

$$\sigma = \frac{\Lambda}{q_{max}} - \nu. \tag{7.24}$$

If the [mAU] equivalent q'_{max} was determined, the equation becomes

$$\sigma = \frac{\Lambda \cdot a}{q'_{max}} - \nu. \tag{7.25}$$

With known ν and a, σ can be uniquely determined. Alternatively, the steric factor can be identified from the peak shape in non-linear chromatography [56], again with known ν and a. This method allows for including additional steric shielding effects in multi-component settings.

To determine a, one of the original methods for identifying the characteristic charge can be used. It relied on measuring the increase in conductivity caused by freed counter-ions [24]. In the UV-based case, the amount of freed counter-ions can be determined from the second isotherm equation (7.20):

$$\int_0^L \sum_{j=1}^k \frac{\nu_j}{a_j} q'_j \, \mathrm{d}x. \tag{7.26}$$

If ν has been determined, e.g. from gradient elutions as above, a can be identified uniquely from the increase in conductivity in single-component adsorption. In non-linear multi-component settings, the increase of conductivity will be visible in the chromatogram. The a-dependent locally varying counter-ion concentration in the pores due to adsorbing proteins will lead to a different adsorption behaviour.

Summarizing, every parameter plays a distinct role and can be determined from single-component absorbance curves. If only the chromatogram is available in multi-component settings, the parameters might be correlated.

7.2.9   Numerical Solution

The numerical simulation is performed using the in-house software package ChromX [98]. Following the method of lines, the equation system is first discretized in space using the Finite Element Method (FEM). A Streamline-Upwind-Petrov-Galerkin (SUPG) ansatz was used here with linear basis and test functions. The discretisation in time is performed with the fractional step θ-scheme, a semi-implicit procedure providing second-order accuracy [62]. Finally, the non-linearity of the equation system introduced by the isotherm must be treated with an iterative procedure, here, Picard iteration. The resulting linear systems are solved by LU factorization.

A variety of algorithms is available for the solution of the optimisation problem in Eq. (7.1). We employed a heuristic method, the genetic algorithm implementation GAlib [197], and a deterministic Levenberg-Marquardt implementation CMinpack [45]. Genetic algorithms prevent local minimums by performing random jumps and, hence, explore a larger area of the search space. The result of the genetic algorithm is then refined with the deterministic algorithm. To support this, we divide the kinetic isotherm by k_{des} and use the formulation with equilibrium coefficient k_{eq}. Working with k_{ads} would require to always change k_{des} at the same time to keep the retention time constant. This is unsuitable for the deterministic algorithm that only uses first derivatives [77].
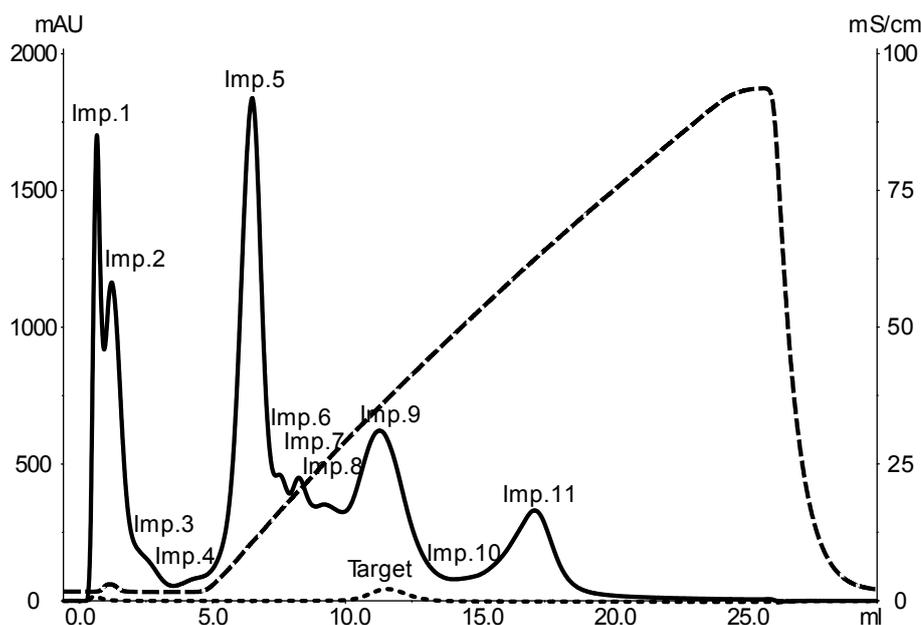
Figure 7.2: Result of the 20 ml gradient elution: UV 280 nm (solid line), UV 536 nm (dotted line) and conductivity (dashed line). 11 impurities were identified with peak maxima at 0.65 ml (imp. 1), 1.21 ml (imp. 2), 1.96 ml (imp. 3), 5 ml (imp. 4), 6.5 ml (imp. 5), 7.5 ml (imp. 6), 8.23 ml (imp. 7), 9.21 ml (imp. 8), 11.28 (imp. 9), 15.52 ml (imp. 10), 17.11 ml (imp. 11). The target component is clearly visible in the 536 nm signal, with the peak maximum at 11.54 ml.

7.3　Results and Discussion

7.3.1　Bind-elute Experiments

Five experiments in bind/elution mode were performed. Figure 7.2 shows the result obtained with a 20 ml gradient. Several impurity peaks could be resolved. The first one is a breakthrough at 0.65 ml. The second peak occurs slightly later at 1.21 ml, followed by two shoulders (impurities 3 and 4) and a high peak at 6.50 ml. The signal continues with three lower peaks (imp. 6, 7, 8), followed by a larger one at 11.28 ml that is also visible at 536 nm. It is identified to be the target component. The fact that this peak's maximum is reached 0.26 ml earlier at 280 nm leads to the assumption that a impurity (imp. 9) is eluting slightly before. A small shoulder (imp. 10) and a final peak at 17.1 ml (imp. 11) complete the elution profile.

7.3.2　Protein Parameter Estimation

The components' peak areas were determined with Unicorn peak integration from the 280 nm signal of the 20 ml gradient chromatogram. The resulting areas in mAU·ml were divided by the sample volume of 0.5 ml to obtain the inlet absorbance values for modelling c'_{in} in mAU. Hence, the estimated absorption coefficients a refer to 280 nm. The same bounds of integration were used for the 300 nm signal. No additional simulation was necessary, the simulated 280 nm peaks were scaled according to the ratio of 300 nm and
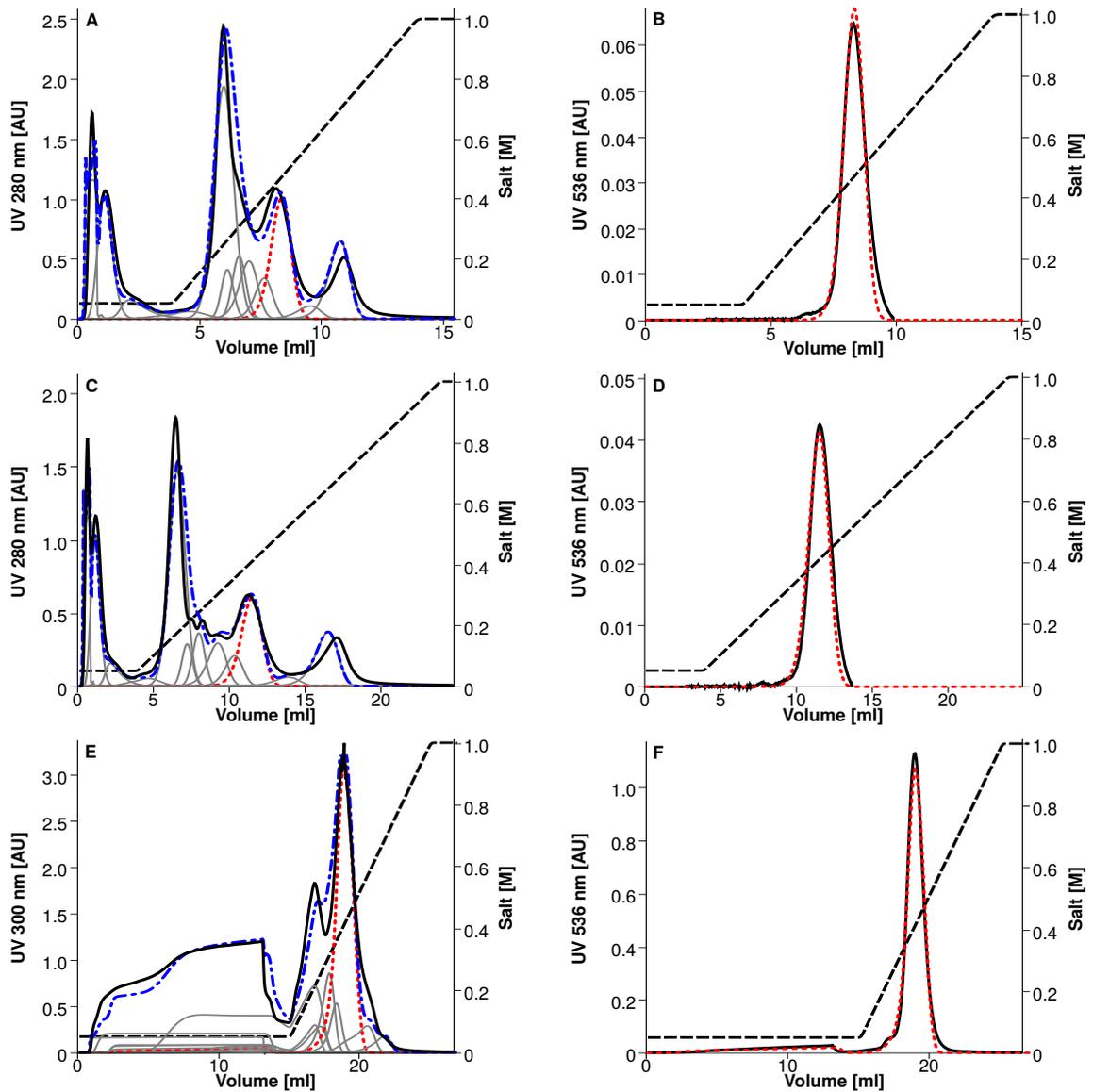
Figure 7.3: Comparison of measured and simulated chromatograms with UV signals (solid lines), conductivity measurement (dashed lines), simulated Cherry-GST absorbance (dotted lines), impurity traces (light solid lines), and sum of simulated proteins (dot-dashed lines). Plots A and B show the 10 ml gradient elution, C and D the 20 ml gradient elution, and E and F the breakthrough experiment.

Table 7.3: Estimated component-specific model parameters.

Component	$k_{eff}/10^{-3}$	k_{des}^{-1}	k_{eq}	ν	σ	$a/10^8$
Impurity 1	0.010	-	-	-	-	-
Impurity 2	2.090	1	0.004	0.066	157.1	0.600
Impurity 3	5.006	0	3.939	1.293	102.5	0.056
Impurity 4	5.586	0.050	22.46	0.566	6.325	0.130
Impurity 5	15.00	0.151	32.36	0.866	1.895	41.37
Impurity 6	13.00	0	21.54	1.890	0.001	0.150
Impurity 7	12.08	0	32.22	1.930	1.702	13.24
Impurity 8	7.077	0.066	75.83	4.135	0	1.967
Impurity 9	7.100	0.070	175.8	4.289	0.022	42.91
Cherry-tagged GST	9.271	0.208	334.0	4.000	0.035	0.865
Impurity 10	14.98	0.424	3084.3	5.350	2.241	2.396
Impurity 11	7.131	0.235	152227	7.968	0.025	30.00

280 nm peak areas. The 300 nm signal was used for modeling the breakthrough experiment. Here, the 280 nm signal was incomplete due to sensor saturation.

As the genetic algorithm performs random jumps, admissible parameter ranges have to be set. First estimates of ν and k_{eq} were obtained from the correlation of retention times in gradient elution as in [171] for all binding species. The resulting large equilibrium parameter values in the order of 10^6 fit well to the observed retention. k_{des} has to reside in the natural range $k_{des}^{-1} \in [0, 1]$ and the upper limit for k_{eff} is given by $3 \cdot k_{eff}/r_p = 1$. The steric factor was assumed to be in a range of $\sigma \in [0, 200]$ as it scales approximately with the molecular weight [100] and we expect HCPs of 100 kDa and above. The range of absorption coefficients was chosen large, $a \in [10^5, 10^{10}]$. Eventually, the curve fitting was refined using the Levenberg-Marquardt algorithm.

Selected results are presented in Figure 7.3. The left column shows the simulated components, their sum, and the chromatograms at 280 nm for the 10 and 20 ml gradient as well as the 300 nm chromatogram of the breakthrough experiment. The right column shows the same curves at 536 nm. The 15 ml gradient result looks very similar. In the 5 ml gradient chromatogram, highly overlapping impurities lead to only three distinct peaks (data not shown). The corresponding model parameters are given in Table 7.3. As the first peak is not retained, isotherm parameters could not be determined. The second and third impurity are only slightly retained and the correlation from [171] cannot be used. Furthermore, there might be other parameter combinations of k_{eq} and ν that lead to the same retention volume, and combinations of σ and a with the same amount of occupied ligands. Certainty can be increased by using different low-salt buffer concentrations or including samples with different impurity ratios. But parameter determination for components in the flow-through is not in the focus of this study. The other linear parameters lead to good agreement of simulation and measurement in the first and second row of Fig. 7.3. As the species do not interact, the four linear parameters can be determined well from the four gradient experiments. The experiment with 12.4 ml sample volume shows good agreement as well, in particular in UV 536 nm. Thanks to this visibility of Cherry-tagged GST in UV 536 nm, a single-component adsorption curve is available that allows for an accurate estimation of the absorption coefficient.

7.3.3 Capillary Gel Electrophoresis

The target protein can be identified easily in the capillary gel electrophoresis result because of its high fluorescence value. The determined concentration was $3.73 \cdot 10^5$ M, resulting in an absorption coefficient of $7.86 \cdot 10^7$ mAU/M at 280 nm. This is smaller than the estimated value by approximately 9 %. The estimate of a is very good, considering the number of interacting species.

Although capillary gel electrophoresis identified even more species, the lumped peaks found by Unicorn peak integration were sufficient to model the elution behavior of the protein components at all investigated wavelengths.

7.4 Concluding Remarks

This study demonstrates that mechanistic modelling can be applied to an anion-exchange step of a crude feed stock, even if the molar concentrations of the feed components are unknown. The model equations were re-written to define injection with respect to the peak areas determined from chromatograms at a chosen wavelengths. The unknown absorption coefficients that scale molar concentration to absorbance units then occur in the isotherm equation. The counter-ion balance of stoichiometric exchange models can be used for estimating these factors using the inverse method. For the steric mass action model, it was shown theoretically that this additional parameter can be uniquely determined in single-component settings.

In a multi-component case study, the molar concentration of the target protein, Cherry-tagged GST, estimated by chromatogram fitting was only 9 % less than the value measured by capillary gel electrophoresis. Here, a single-component absorbance curve was available through the absorbance of the Cherry-tag in UV 536 nm. For the other components, a correlation of the steric shielding factor and the absorption coefficient persists. Only the total counter-ion concentration on the adsorber surface is measurable, but not the exact amounts displaced per-component.

Additional reliability can be achieved by including samples with different impurity proportions or fraction analyses that only need to provide peak percentages in one of the observed wavelengths. In preparative chromatography process development, these fraction analyses are performed on a regular basis, e.g. with size-exclusion chromatography [99] or ion-exchange HPLC [35], such that no additional experiments are required.

Acknowledgment

8 | Calibration-free Inverse Modeling of Ion-exchange Chromatography in Industrial Antibody Purification

Tobias Hahn[1], Thiemo Huuk[1], Anna Osberghaus[1], Katharina Doninger[2], Susanne Nath[2], Stefan Hepbildikler[2], Vincent Heuveline[3], Jürgen Hubbuch[1]

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

[2] Pharmaceutical Biotech Production, Roche Diagnostics GmbH, Penzberg, Germany

[3] Heidelberg University, Interdisciplinary Center for Scientific Computing, Engineering Mathematics and Computing Lab, Heidelberg, Germany

Abstract

The identification of optimal process parameters for the isolation of a target component from multi component mixtures is especially challenging in industrial applications. With constantly increasing time to market pressure, screening a large parameter space is not feasible and Design-Of-Experiment approaches with few experiments might fail due to dynamic and nonlinear reactions to small parameter changes.

Model-based optimization can determine optimal operating conditions, once the model has been calibrated to the specific process step. In this work, parameters for the Steric Mass Action model are estimated for the target protein and three impurities of an industrial antibody cation-exchange purification step using only chromatograms at different wavelengths and additional fraction analyses with size exclusion chromatography. Information on the molar or mass concentrations in the feed are not available. The model-based optimization results coincide with conventional chromatogram-based optimization.

Practical Application

This paper demonstrates that model-based optimization can be applied to an industrial cation-exchange step, even if the molar concentrations of the feed components are unknown. Based on a few experiments, which might be already available from previous screenings, the model can be calibrated. For differentiating components eluting hidden in the sum signal, single components elution curves must be generated for peak fitting. This is accomplished by SEC fraction analyses at the same wavelength as the chromatogram.

8.1 Introduction

Biopharmaceutical products currently constitute one of the fastest growing markets for the pharmaceutical industry [104]. Industrial downstream processing (DSP) faces the challenge of efficiently purifying a product out of a very heterogeneous mixture. Especially for monoclonal antibodies (mAb), the purification sequence is commonly based on platform processes that are only slightly adapted to new target components [115; 171]. This approach ensures fast process development and reduces the time to market, but it is probable that the reduced exploration of design space leads to suboptimal processes [140]. Process understanding is hence a sensible and important addition to the use of platform processes [81]. As the integration of modeling tools into process development is also an essential part of the strategy for the implementation of the Quality by Design (QbD) approach [31; 86], modeling tools are increasingly gaining the attention of the pharmaceutical industry. It ensures time and material-efficient process optimization as long as the model parameters can be determined reliably with less effort.

For some sorption sub-models, such as the Steric Mass Action (SMA) model for ion-exchange chromatography of proteins [24], model calibration protocols for pure components exist [24], which allow for determining the component-specific parameters in a consecutive fashion. With no a priori knowledge about the components' behavior within mixtures, the inverse method is a suited option, which alters parameters in a systematic fashion to achieve a match of measured chromatogram and model prediction. In a comparative study [149], the direct approach and the inverse method were found to reach equal prediction quality such that the latter is recommended for fast process development.

The potential of mathematical modeling and numerical optimization of chromatography has been demonstrated in academic set-ups for mixtures of small molecules [1; 53], model proteins [150] and antibodies [40; 99], as well as for industrial process steps [35; 133]. All applications have in common, that sensor calibrations existed for all components, allowing to transform UV absorption values into mass or molar concentrations. This paper discusses the challenges when applying mechanistic modeling and numerical optimization to a data-set of industrial preparative chromatography without knowledge of feed composition in terms of molar concentrations. Additional experiments and analyses are avoided by employing a modeling approach that estimates the absorption coefficients from the recorded UV data.

The case study bases on a cation-exchange chromatographic (CEX) process step (Poros 50 HS resin). The antibody mixture fed into CEX is a protein A eluate after low pH

incubation and conditioning. The mixture contains several antibody variants differing in size and charge, that all elute in a common peak. Fraction collection and analysis has to be performed to obtain information on the location of the impurities. Using size exclusion chromatography, the targeted monomer and three impurities can be distinguished, two high molecular weight (HMW) and one low molecular weight (LMW) species. The found model parameters were used for *in silico* optimization and eventually coincide with conventional optimization by chromatogram evaluation.

8.2 Materials and Methods

Initial experiments were conducted to determine system properties, thereafter six experiments in bind/elution mode were performed. The following sections describe the experimental set-up, data processing and parameter estimation.

8.2.1 Column Parameter Determination

A $20\,ml$ column, with Poros 50 HS (Applied Biosystems, Carlsbad, CA, USA) resin was analyzed first with an acetone pulse and second with dextran using an Äkta Avant system (GE Healthcare, Little Chalfont, UK) to determine the essential model parameters as described in [129] and presented in Table 8.1. Acid-base-titration was carried out to determine the total ionic capacity: The column was flushed with a 0.5 M HCl solution until a constant UV and conductivity signal was achieved. Afterwards the column was washed with ultrapure water until a constant UV and conductivity baseline was reached. After that the column was titrated at a flow of 100 cm/h with a 0.01 M NaOH solution until an increase in conductivity signal was recorded. From the Na-ion concentration of the titrant and the volume of the applied titrant, the total number of exchangeable ions was calculated. All chemicals used were obtained highest quality.

With this set of parameters all system specific parameters occurring in the mathematical model can be fixed as also given in Table 8.1.

8.2.2 Bind-elute Experiments

As system liquid a 50 mM acetate buffer was used at pH 4.95, for elution a high-salt buffer of 50 mM acetate and 750 mM NaCl was employed. The different salt profiles in the following were mixed from these two buffers.

A sample volume of 81.3 ml = 4.2 column volumes (CV) was injected for each experiment. First, the gradient experiment was ran from 50 mM (0 % high salt buffer) to 550 mM (66 % high-salt buffer), with the gradient starting at 8.4 CV and ending at 18.1 CV. As the elution peak reached its maximum at a total salt concentration of 210 mM, five step elutions with concentrations 190, 200, 210, 220 and 230 mM were performed for optimization. Each step was induced at 7 CV and was followed by a wash with 100 % high-salt buffer as soon as the 280 nm UV signal fell below a threshold of 100 mAU.

Table 8.1: Measured (top) and calculated (bottom) column parameters.

Parameter	Symbol	Value	Unit	Proceeding
Diameter	d	10	mm	From manufacturer
Length	L	246	mm	Manual measurement
Bead radius	r_p	0.025	mm	From manufacturer
System dead volume	V_d	1.98	ml	Acetone injection without column
Retention volume Acetone	V_{RetAc}	16.94	ml	Acetone peak injection with column
Retention volume dextran	V_{RetDex}	10.29	ml	Dextran peak injection with column
Standard deviation of dextran	σ_{Dex}	0.161	ml	Äkta peak integration
Volume of $NaOH$	V_{NaOH}	130.2	ml	Acid/base titration
Molarity of $NaOH$	c_{NaOH}	0.01	M	Manually controlled
Flow rate	u	0.69	$\frac{mm}{s}$	Manually controlled
Volume	V	19.32	ml	$\frac{1}{4} \cdot \pi \cdot d^2 \cdot L$
Fluid volume	V_f	14.96	ml	$V_{RetAc} - V_d$
Interstitial volume	V_{int}	6.43	ml	$V_{RetDex} - V_d$
Total column porosity	ε_t	0.77		$\frac{V_f}{V}$
Interstitial porosity	ε_b	0.43		$\frac{V_{int}}{V}$
Particle porosity	ε_p	0.6		$\frac{V_f - V_{int}}{V - V_{int}}$
Interstitial flow	u_{int}	1.61	$\frac{mm}{s}$	$\frac{u}{\varepsilon_b}$
Axial dispersion	D_{ax}	0.124	$\frac{mm^2}{s}$	$\frac{1}{2} \cdot L \cdot u_{int} \cdot \left(\frac{\sigma_{Dex}}{V_{int}}\right)^2$
Ionic capacity	Λ	0.3	M	$\frac{c_{NaOH} \cdot V_{NaOH}}{V_c(1-\varepsilon_b)(1-\varepsilon_p)}$

The elution peaks of all experiments were captured in 3 ml samples and analyzed by SEC to determine the relative contribution of the two HMW species, the monomer, and the LMW species to the sum signal. In the case of the step elutions, the peak occurring during the subsequent wash was also fractionated and analyzed. The same analysis was performed for the feed material.

As the gradient elution resulted in a blunt top (Figure 8.2) and not a distinct peak, a longer gradient elution was performed to check for charge variants.

8.2.3 Mathematical Model

The UV absorbance based model as developed in [74] is used in the following. The General Rate Model (GRM) [129] models the macroscopic protein transport through the column. The system is of Convection Diffusion Reaction (CDR) type. Eq. (8.1) describes the rate of change of a concentration $c_i(x,t)$, measured in [M] for salt and [mAU] for proteins, in the interstitial volume of a column with length L to consist of convective mass transport in space with the average interstitial velocity of the fluid u, peak broadening effects that are modeled as dispersion in axial direction with respect to a coefficient D_{ax}, and transition from the interstitial concentration into the particle pore concentration $c_{p,i}(x,r,t)$ which

depends on the porosity of the bed ε_b, the radius of adsorber particles r_p and a component-specific film transfer coefficient $k_{film,i}$. The model is complemented with Danckwerts boundary conditions, Eqs. (8.2),(8.3). Equations (8.4)-(8.6) model the accumulation of mass in the pore volume $c_{p,i}$ and stationary phase q_i depending on the particle porosity ε_p and the component-specific pore-diffusion coefficient D_p.

$$\frac{\partial c_i}{\partial t} = -u(t)\frac{\partial c_i}{\partial x} + D_{ax}\frac{\partial^2 c_i}{\partial x^2} - \frac{1-\varepsilon_b}{\varepsilon_b}k_{film,i}\frac{3}{r_p}\left(c_i - c_{p,i}\right) \tag{8.1}$$

$$\frac{\partial c_i}{\partial x}\left(0,t\right) = \frac{u(t)}{D_{ax}}\left(c_i\left(0,t\right) - c_{in,i}\left(t\right)\right) \tag{8.2}$$

$$\frac{\partial c_i}{\partial x}\left(L,t\right) = 0 \tag{8.3}$$

$$\frac{\partial c_p}{\partial t}\left(x,r,t\right) = \frac{1-\varepsilon_p}{\varepsilon_p}\frac{\partial q}{\partial t}\left(x,r,t\right) + \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 D_p\frac{\partial c_p}{\partial r}\left(x,r,t\right)\right) \tag{8.4}$$

$$\frac{\partial c_p}{\partial r}\left(x,r_p,t\right) = \frac{k_{film}}{\varepsilon_p D_p}\left(c(x,t) - c_p\left(x,r_p,t\right)\right) \tag{8.5}$$

$$\frac{\partial c_p}{\partial r}\left(x,0,t\right) = 0 \tag{8.6}$$

The steric mass action isotherm (SMA) [24], modified by [74], is a commonly used semi-mechanistic isotherm in ion-exchange chromatography. It is capable of reproducing the influence of counter ions on the retention behavior of protein species using the proteins' characteristic charges ν_i. In addition, it considers adsorbent properties such as the total ionic capacity Λ and steric shielding effects σ_i of the protein i covering an amount of binding sites, greater than the actual number of sites it interacts with. The UV absorbance-based kinetic SMA isotherm is given in Eq. (8.7), with q_i and $c_{p,i}$ being the concentration of the protein i adsorbed and in solution respectively. $c_{p,salt}$ is the salt concentration of the solution. $k_{eq,i}$ and $k_{des,i}$ are the constants of equilibrium and desorption rate, and a_i the absorption coefficient that scales molar concentrations to absorbance units, according to Lambert-Beer law. The factors a_i consist of extinction coefficient and UV cell path length, where the extinction coefficients are unknown for the impurities and cannot be determined easily as the components are not available in pure form.

$$k_{des,i}^{-1}\frac{\partial q_i}{\partial t} = k_{eq,i}\left(\Lambda - \sum_{j=1}^{k}\frac{\nu_j + \sigma_j}{a_j}q_j\right)^{\nu_i}c_{p,i} - c_{p,salt}^{\nu_i}q_i \tag{8.7}$$

$$q_{salt} = \Lambda - \sum_{j=1}^{k}\frac{\nu_j}{a_j}q_j \tag{8.8}$$

The model is chosen because of its capability of simulating the whole chromatographic process, including elution, by changing the induced salt concentration at the inlet.

8.2.4 Numerical Solution

The numerical simulation is performed using the in-house software package ChromX. Following the method of lines, the equation system is first discretized in space on given nodes,

using the Finite Element Method (FEM). FEM is a highly versatile method with strong mathematical foundation and well suited for CDR equations. The solutions procedure starts with the weak formulation, incorporating the boundary conditions and representing the variables with basis functions from the respective spaces. A Streamline-Upwind-Petrov-Galerkin (SUPG) method was used here with linear basis and test functions. The discretization in time is performed with the fractional step θ-scheme, a semi-implicit procedure providing second-order accuracy [62]. Finally, the non-linearity of the equation system introduced by the isotherm must be treated with an iterative procedure, here, Picard iteration. The resulting linear systems are solved by LU factorization.

8.2.5   Parameter Estimation

Estimation of an unknown parameter set $\overline{p}$ solves the least squares optimization problem

$$\min_{\overline{p}} \sum_j \left(m(t_j) - \sum_{i \geq 1} c_i(L, t_j; \overline{p}) \right)^2 , \tag{8.9}$$

where $m(t_j)$ is the measured chromatogram value at point in time t_j. The measurements might also contain noise, which can be neglected when assuming the noise to be zero-mean Gaussian and isotropic. The concentration c_i is simulated in absorbance units. Linear scaling to molar concentrations was checked with concentrated feed samples.

The sensor signal is approximately proportional to the feed concentration up to the saturation limit in UV 280, as suggested by Lambert-Beer law. Furthermore, UV 300 nm captures approximately four times higher concentrations than UV 280 nm (data not shown). The chromatograms were also corrected by the influence of the salt concentration in the mobile phase.

A variety of algorithms is available for the solution of this optimization problem. In this work, we employed a heuristic method based on a genetic algorithm (GAlib ver. 2.4.7 [197]). Genetic algorithms avoid local minimums by performing random jumps and hence explore a larger area of the search space.

As the UV sensor is quickly saturated at 280 nm, we consider additional absorbance measurements at 260 and 300 nm to capture the complete peak shape. In addition, we include size-exclusion chromatography (SEC) analyses of fractions at a common wavelength, such that the contributions to the UV sum signal can be quantified. This data enters the least squares problem Eq. (8.9): m is extended with additional entries of the contributions for each fraction and species and c is extended accordingly with entries only including those simulated concentrations that belong to the respective contribution.

The estimation procedure using SEC data relies on pseudo absorbance profiles $m_i(t_j)$ for all components i that are generated by multiplying the UV signal with the component's fraction percentage in the corresponding time interval. Figure 8.1 shows the pseudo profiles for the step elution with lowest salt concentration. A HMW1 peak is visible during the high-salt step following elution, while LMWs are only visible at the beginning of the elution peak.
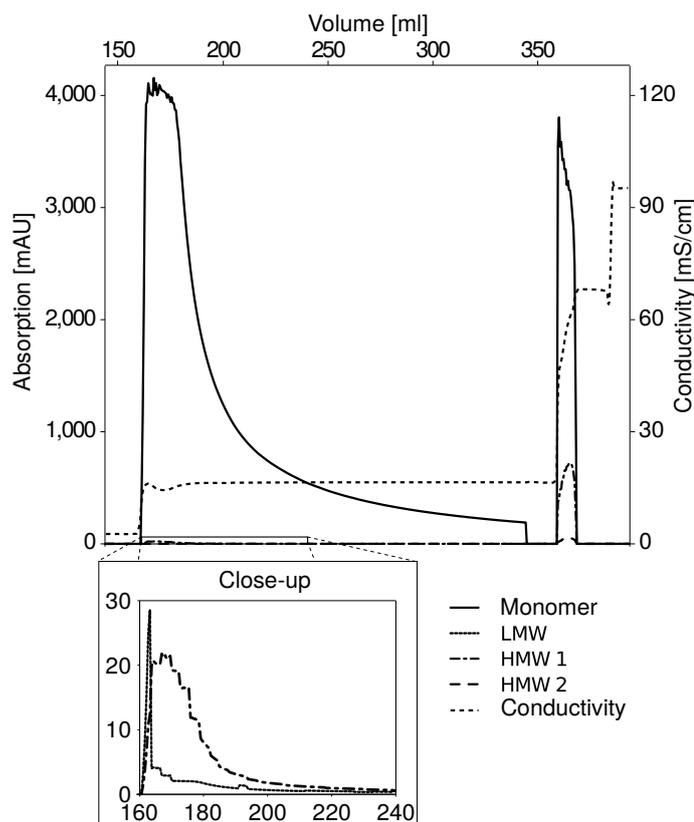
Figure 8.1: Single-component absorption profiles at UV 280 nm for the 190 mM salt step elution. The curves are generated by multiplying the chromatogram by component ratios found by fraction analysis with SEC.

8.3 Results and Discussion

8.3.1 Bind-elute Experiments

The feed analysis by SEC showed 96.79 % monomer, 2.95 % + 0.17 % species with higher molecular weight (HMW 1 and HMW 2) and 0.09 % species with lower molecular weight (LMW) at 280 nm. Further IEC analysis revealed, that the exact composition of the monomer is 15.4 % acidic, 63.3 % main and 21.3 % basic variant. An example of a step elution result is presented in Figure 8.1. It shows a steep front, long tailing and a considerable peak after the final high salt step. The gradient elution result in Figure 8.2 shows a blunt top, indicating that the monomer variants elute differently.

8.3.2 Reference Optimum

When not using model-based optimization in combination with reliable feed concentrations or alternative assays, the optimum has to be defined using only UV absorbance data. Purity is defined as average SEC target peak area of involved fractions divided by total absorbance and yield is similarly defined as ratio of collected peak area by total area. Because of process performance requirements, only step elution scenarios are considered
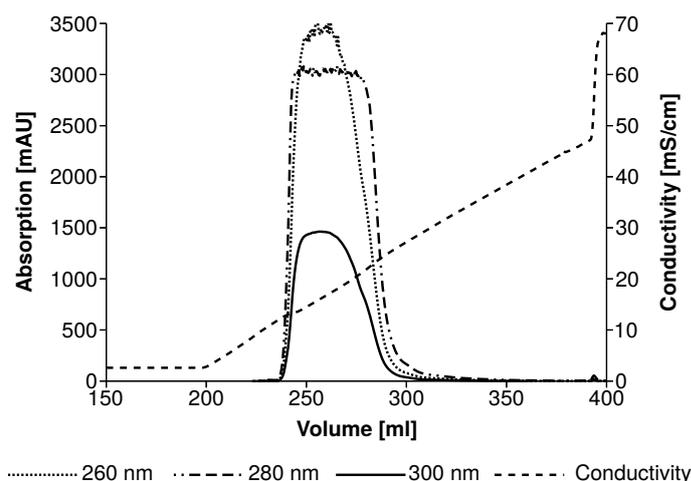
Figure 8.2: Gradient elution peak recorded at UV 260, 280 and 300 nm. Only the 300 nm signal shows the entire top. The peak shape indicates that the monomer consists of differently eluting charge variants.
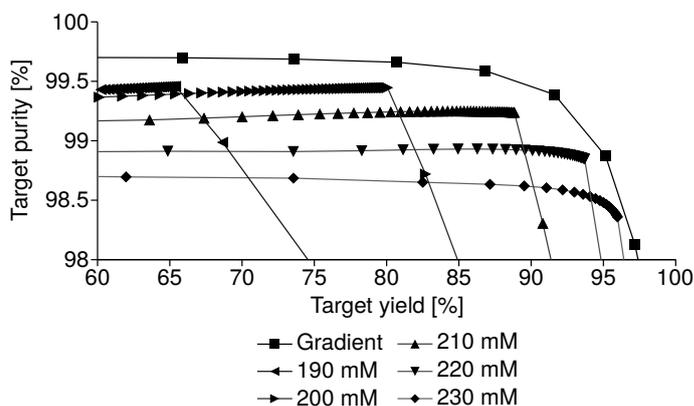

Figure 8.3: Monomer purity over yield when collecting fractions consecutively. Higher steps lead to co-eluting contaminants and lower purity. Lower steps allow to separate contaminants but do not collect all of the target component.

that achieve at least 99 % purity and 80 % yield with fraction size smaller than 5 column volumes.

Figure 8.3 shows the development of monomer purity over yield when starting to collect from the first fraction of the elution peak and successively adding the other fractions. Based on this data, the 200 and 210 mM steps perform best. Gradient elutions are undesired in the final process and were not considered. When collecting the first 33 fractions (= 99 ml ≈ 5 CV), the 210 mM step achieves a purity of 99.3 % and 86.5 % yield. The 200 mM step achieves a slightly better purity of 99.5 %, but only 80 % yield. The 220 mM step reaches a higher yield with fewer fractions but does not attain the desired purity.

8.3.3 Protein Parameter Estimation

The estimation algorithm employed is a genetic algorithm, to quickly cover a large search space. The inlet absorbance values are set for all components using the known peak area
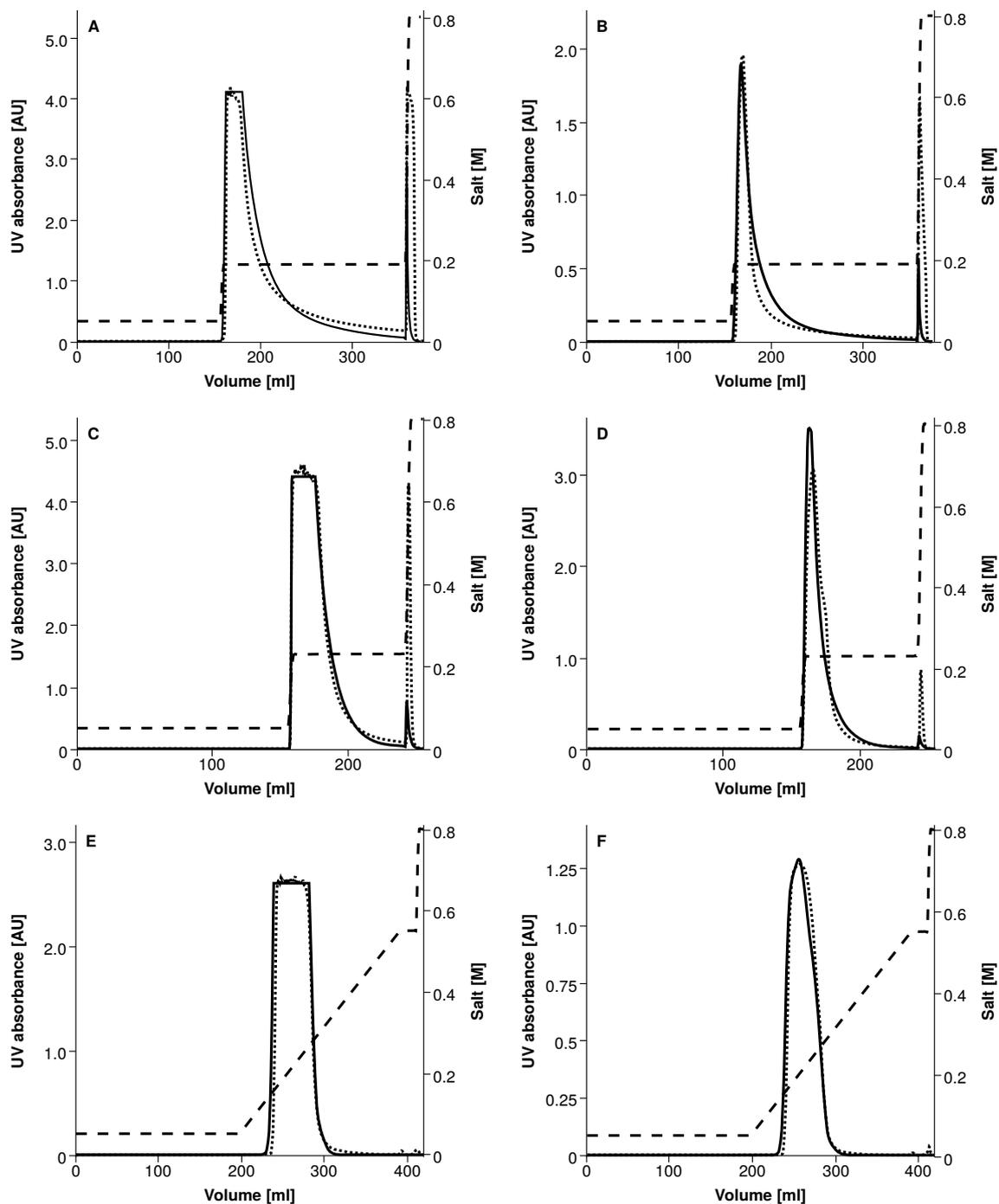
Figure 8.4: Comparison of measured chromatogram (···) and simulated sum signal (−) for the salt elutions (- -) used for model calibration.

Table 8.2: Estimated model parameters.

Component	k_{film}	D_p	k_{des}^{-1}	k_{eq}	ν	σ	a
Salt	0.0083	$7.00{\cdot}10^{-04}$	-	-	-	-	-
LMW	0.0083	$3.49{\cdot}10^{-06}$	$1.61{\cdot}10^{-2}$	1.70	3.32	65.0	$4.4{\cdot}10^{7}$
Monomer Acidic	0.0083	$2.10{\cdot}10^{-05}$	$9.00{\cdot}10^{-5}$	1.98	3.05	75.3	$8.22{\cdot}10^{7}$
Monomer Main	0.0083	$5.34{\cdot}10^{-06}$	$1.90{\cdot}10^{-3}$	1.42	5.21	75.3	$8.22{\cdot}10^{7}$
Monomer Basic	0.0083	$1.57{\cdot}10^{-04}$	$6.08{\cdot}10^{-4}$	1.91	6.90	75.3	$8.22{\cdot}10^{7}$
HMW 1	0.0083	$3.60{\cdot}10^{-06}$	$3.50{\cdot}10^{-6}$	2.40	8.23	210	$2.14{\cdot}10^{8}$
HMW 2	0.0083	$7.72{\cdot}10^{-06}$	$3.90{\cdot}10^{-5}$	5.70	5.20	287	$2.10{\cdot}10^{8}$

at 300 nm, the scaling factor from 300 to 280 nm and the results from SEC and IEC analysis of the feed at 280 nm.

From the available data, the gradient and the highest and lowest salt steps were used for estimation, as the peak shape did not change much for the intermediate steps.

Absorbance-based modelling is able to determine the non-linear parameters σ_i and a_i uniquely from single-component curves [74]. These are available for all impurities, but not for the monomer charge variants. As the variants are indistinguishable in SEC, we can assume them to have equal steric shielding coefficients and absorption coefficients.

Curve fitting finished with a very good match of measurement and simulation, considering the complex elution behavior with long tailing step elutions and blunt gradient top. The result is shown in Figure 8.4.

The found parameter set in Table 8.2 is reasonable, the characteristic charges ascend with the molecule size and in the monomer case with charge variant type. Steric shielding factors and absorption coefficients ascend approximately with molecule size as expected. Sorption kinetics and film diffusion are fast, as indicated by the steep elution fronts.

8.3.4 Optimization

Again, the genetic algorithm was employed to determine the optimal salt step height together with the fractionation boundaries. The optimization of load conditions was performed with traditional high-throughput screening beforehand. Hence, comparison to a non-model-based approach in this particular scale was not possible.

Results of the genetic algorithm are plotted in Figure 8.5. Although the model was calibrated only with the gradient and the 190 mM and 230 mM steps, the results resemble the findings in section 8.3.2 closely: only step concentrations below 210 mM allow a purity above 99 % and only step concentrations below 200 mM achieve high yields above 95 %. The yield values are slightly higher in the model-based optimization results as some parts of the reference elution peaks were not analyzed by SEC and could not be considered when calculating the reference optimum (cf. Fig. 8.1).

A compromise between yield and purity is found at approximately 200 mM, closely followed by 210 mM, just like in the reference analysis. The admissible volume of 100 ml would be fully exploited by a 195 to 200 mM step. A 210 mM step is predicted to achieve a yield of 91 % at the required purity of 99 % using only 60 ml.
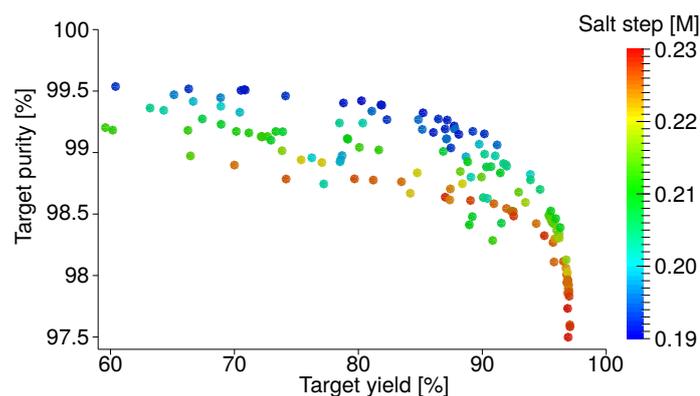
Figure 8.5: Intermediate values of salt step optimization with a genetic algorithm. Step concentrations below 210 mM allow a purity above 99 %. Highest yield values are achieved by step concentrations above 200 mM.

8.4   Concluding Remarks

The benefit of model based optimization compared to the traditional approach is foremost given by the fast adaptation to changes in the feed composition and the possibilities of further optimization. For example, salt profiles can be constructed consisting of several steps in combination with a gradient.

The reduction of the number of experiments in this case is not significant, as conventional optimization was straight forward and the model-based optimum would have to be verified by an additional experiment.

Fractionation itself is a necessity for both approaches, where further studies should examine if good results can be obtained with lower resolution.

Acknowledgment

9 | High-throughput Micro-scale Cultivations and Chromatography Modeling: Powerful Tools for Integrated Process Development

Pascal Baumann[1], Tobias Hahn[1], Jürgen Hubbuch[1]

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

Abstract

Upstream processes are rather complex to design and the productivity of cells under suitable cultivation conditions is hard to predict. The method of choice for examining the design space is to execute high-throughput cultivation screenings in micro-scale format. Various predictive *in silico* models have been developed for many downstream processes, leading to a reduction of time and material costs. This paper presents a combined optimization approach based on high-throughput micro-scale cultivation experiments and chromatography modeling. The overall optimized system must not necessarily be the one with highest product titers, but the one resulting in an overall superior process performance in up- and downstream.

The methodology is presented in a case study for the Cherry-tagged enzyme Glutathione-S-Transferase from *Escherichia coli* SE1. The Cherry-Tag™ (Delphi Genetics, Belgium) which can be fused to any target protein allows for direct product analytics by simple VIS absorption measurements. High-throughput cultivations were carried out in a 48-well format in a BioLector® micro-scale cultivation system (m2p-Labs, Germany). The downstream process optimization for a set of randomly picked upstream conditions producing high yields was performed *in silico* using a chromatography modeling software developed in-house (ChromX). The suggested *in silico*-optimized operational modes for product capturing were validated subsequently. The overall best system was chosen based on a combination of excellent up- and downstream performance.
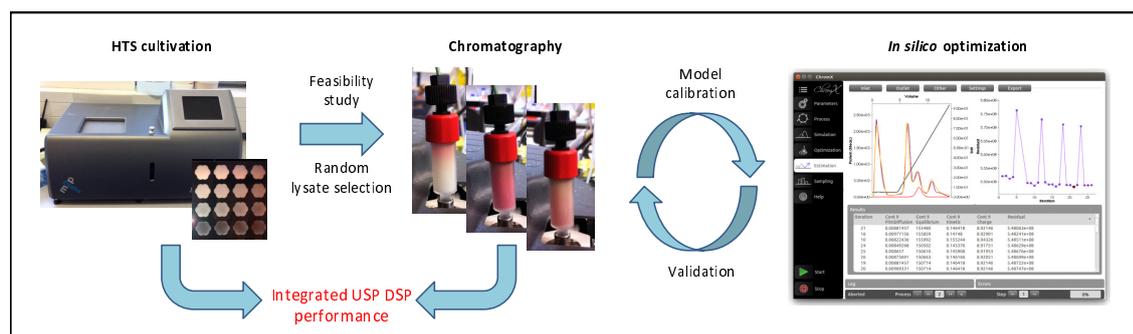
Figure 9.1: Schematic overview of the integrated optimization approach for bioprocess development based on high-throughput micro-scale cultivations and lysate characterization by chromatography modeling. Different lysates from the micro-scale cultivations are investigated in low column volume chromatography for model calibration. The created models are then used for *in silico* optimization of the purification process. The overall best system is chosen based on the overall optimal performance in up- and downstream.

9.1 Introduction

Although the complexity of the protein purification procedure is strongly dependent on the cultivation conditions, up- and downstream processes are mostly optimized separately. When linking those two process parts, the best cultivation must not necessarily be the one leading to highest yields and titers, but the one resulting in an overall superior process performance. For product formation, it is important to have tools to screen large numbers of organisms and production conditions. The subsequent downstream screening needs to be capable of scoring the large number of produced feedstocks in terms of their applicability for downstream process development (DSP).

Cultivation conditions for the effective production of biopharmaceuticals are hard to predict. The chosen cultivation conditions strongly influence product formation and can significantly differ from standard growth conditions. For human interferon-α2, human interferon-γ, and the interferon-induced murine protein Mx, it was shown that lowered temperatures lead to an increase in soluble protein during product formation in *E. coli* [164]. Similar results were shown for the production of Vitreoscilla globin under oxygen limitations [47]. Different levels of benzoylformate decarboxylase were obtained using various culture media components in *E. coli* [116]. These results, however, cannot be transferred directly when investigating different strains and products, and *in silico* model implementation is limited. Thus, high-throughput process development (HTPD) is the method of choice for screening a large pool of conditions in upstream process development (USPD) on the micro-scale. Different miniaturized high-throughput cultivation systems were described, mostly resulting in a strong decrease in accessible process information [18]. One approach uses small-scale vessels which imitate the industrial scale fermenters and can be run in parallel [5; 111; 157]. However, the highest capabilities for high-throughput experiments are reached by cultivations in microtiter plates (MTPs). A major disadvantage is the loss of process information and process control, leading to optimization results that might not be suited for scale up. An approach linking the advantages of high-throughput experimentation in microtiter plates with those of continuous process control is the Bio-Lector® technology (m2p-Labs, Germany). Cultivations are carried out in continuously

shaken 48-well plates on the sub-milliliter scale. For better mixing and oxygen uptake, microtiter plates with baffled wells (FlowerPlates®) were developed [55]. For continuous monitoring of cell growth by scattered light measurements, the plates are equipped with an optical bottom, making sampling obsolete [103; 106].

Besides the effects on product formation, influences of cultivation conditions on the contaminant levels are not considered in most cases. E.g. dissolved oxygen levels [125; 174], temperature changes [25; 206], cultivation media [112; 173], as well as the state of the cell in the cell cycle [49; 64] are highly influential on e.g. host cell protein levels. Consequently the initial purity of the product will differ largely.

The above-mentioned variations of the initial purity and concentrations of critical contaminants play a crucial role in downstream process development. Hence, cultivations should not only be scored by the product titers, but also by the contaminant profiles of respective upstream conditions. Besides HTPD, a large number of predictive *in silico* models have been developed for many downstream unit operations, leading to a reduction of time and material costs. Mathematical modeling and numerical optimization of chromatography were applied successfully for model proteins [150], antibodies [40; 99] as well as for industrial processes [35; 133]. All simulation procedures mentioned rely on known molar concentrations of the biomolecules involved. However, models can also be based on absorption data using an inverse method [37]. The major advantage of this procedure is that data from chromatograms can be used directly for model development. No sensor calibration is needed, which is why this method appears to be an ideal tool for comparing a large number of feedstocks within a short period of time.

This paper presents an integrated optimization approach based on high-throughput micro-scale cultivations and lysate characterization by chromatography modeling using chromatogram fitting (Fig. 9.1). The methodology is presented in a case study for the Cherry-tagged enzyme Glutathione-S-Transferase from *Escherichia coli* SE1. The 11 kDa Cherry-Tag™ (Delphi Genetics, Belgium) is fused to the target protein and can be detected by 414 nm absorption measurements [42; 43]. The upstream screening for varying cultivation conditions, such as temperature and induction conditions, was performed in a BioLector® system (m2p-Labs, Germany). A set of randomly picked upstream conditions producing high yields was then characterized and optimized *in silico* using a chromatography modeling software developed in-house (ChromX). The operational modes for process optimization suggested were validated subsequently. Instead of choosing the cultivation leading to highest product concentrations, the ideal system was chosen based on a combination of overall excellent up- and downstream performance.

9.2 Materials & Methods

9.2.1 Mathematical Methods

9.2.1.1 Chromatogram Fitting

In the following section, the governing equations of the simulations are presented. $\overline{p}$ describes the solution of the least squares optimization problem given in Eq. (9.1):

$$\arg\min_{\overline{p}} \sum_j \left(m(t_j) - \sum_{i \geq 1} c_i(L, t_j; \overline{p}) \cdot a_i \right)^2 \tag{9.1}$$

with $m(t_j)$ being the absorption value derived from the chromatogram at point in time t_j, c_i being the protein concentration i at the end of the column of length L, and a_i describing the absorption factor. The algorithms used for solving the above-mentioned problem were a heuristic method based on a genetic algorithm [197] and a deterministic method based on Levenberg-Marquardt implementation [45]. Genetic algorithms prevent local minima by performing random jumps for exploring a larger area in the search space. The deterministic algorithm served for result refinement.

9.2.1.2   Steric Mass-action (SMA) Model

As an adsorption model for ion-exchange chromatography, the semi-mechanistic steric mass-action isotherm (SMA) [24] was applied allowing for chromatography simulations including varying salt concentrations. It covers effects of counter-ions on the retention behavior of proteins, including the proteins' characteristic charges ν_i. Adsorbent properties, such as the total ionic capacity Λ, are also considered. Steric shielding effects caused by proteins covering binding sites without electrostatic interactions are included as a factor σ_i. The kinetic form of the SMA isotherm is shown in Eq. (9.2) for k proteins, with q_i and $c_{p,i}$ being the concentration of the protein i adsorbed and in solution inside the pore, respectively. $c_{p,salt}$ describes the effective pore salt concentration. $k_{ads,i}$ and $k_{des,i}$ are the constants of adsorption and desorption, respectively. q_{salt} in Eq. (9.3) describes the number of salt ions still attached to the adsorbent surface and is defined as the difference between the total ionic capacity Λ and the number of areas blocked by electrostatic adsorption $\sum_{j=1}^{k} \nu_j q_j$. The areas blocked due to steric shielding still possess the counter-ions and, hence, are not taken into account.

$$\frac{\partial q_i}{\partial t} = k_{ads,i} \left(\Lambda - \sum_{j=1}^{k} (\nu_j + \sigma_j) q_j \right)^{\nu_i} c_{p,i} - k_{des,i} c_{p,salt}^{\nu_i} q_i \tag{9.2}$$

$$q_{salt} = \Lambda - \sum_{j=1}^{k} \nu_j q_j \tag{9.3}$$

For small sample loadings ranging from 0 to 5% of the maximum binding capacity of the applied adsorbent resin as in this study the ionic capacity predominates and the occupied adsorption slots can be neglected. This results in a quasi-linear form of the SMA isotherm, meaning that the peak shape is not influenced by the concentration, as shown in Eq. (9.4). When dividing Eq. (9.4) by $k_{des,i}$, Eq. (9.5) results in a kinetic term $\frac{1}{k_{des,i}}$ and an equilibrium term $k_{eq,i}$ which equals $\frac{k_{ads,i}}{k_{des,i}}$.

$$\frac{\partial q_i}{\partial t} \approx k_{ads,i} \Lambda^{\nu_i} c_{p,i} - k_{des,i} c_{p,salt}^{\nu_i} q_i \tag{9.4}$$

$$\frac{1}{k_{des,i}} \frac{\partial q_i}{\partial t} \approx k_{eq,i} \Lambda^{\nu_i} c_{p,i} - c_{p,salt}^{\nu_i} q_i \tag{9.5}$$

In contrast, for setups at the column capacity limit besides the 4 gradients presented in this study a frontal experiment (breakthrough experiment) is mandatory. The estimation procedure then includes further parameters to be determined, namely the steric shielding factor σ and the absorption factors. However, considering the limited sample volume derived from high-throughput cultivations such frontal experiments were not possible.

Nevertheless, for the determination of critical contaminants the simplified methodology works well for pre-selecting feasible lysates for scale up applications.

9.2.1.3 Transport Dispersive Model (TDM)

The Transport Dispersive Model (TDM) [129] in Eqs. (9.6) and (9.7) describes macroscopic protein transport inside the chromatography column and considers the interstitial volume of the mobile phase, as well as the mass transfer into the pore volume. Film and pore diffusion effects are lumped in an effective mass transfer coefficient $k_{eff,i}$ for simplification. The system is of Convection Diffusion Reaction (CDR) type. Eq. (9.6) describes the rate of change of the concentration $c_i(x,t)$ of component i in the inter-particle phase of a column with length L. The first term of Eq. (9.6) refers to the convective transport which is dominated by the inter-particle velocity of the fluid u. The middle term describes hydrodynamic dispersion in axial direction using the dispersion coefficient D_{ax}. The last term refers to the transition of molecules from the interstitial into the particle pore phase c_i - $c_{p,i}$ which depends on the porosity of the bed ε_b, the radius of adsorbent particles r_p, and a component-specific lumped effective mass transfer coefficient $k_{eff,i}$.

Eq. (9.7) models the accumulation and distribution of mass inside the pores. It is composed of the protein pore concentration in the liquid phase $c_{p,i}$ and the protein concentration bound to the stationary phase q_i. Particle porosity ε_p is one of the influencing factors.

For the model, Danckwerts boundary conditions in Eqs. (9.8) and (9.9) were used describing that the concentration at the inlet is influenced by diffusion and back mixing whereas the concentration at the outlet is unaffected by such effects. The adsorption mechanism was described using the SMA isotherm model shown above in Eqs. (9.2) and (9.3).

$$\frac{\partial c_i}{\partial t} = -u\frac{\partial c_i}{\partial x} + D_{ax}\frac{\partial^2 c_i}{\partial x^2} - \frac{1-\varepsilon_b}{\varepsilon_b}k_{eff,i}\frac{3}{r_p}\left(c_i - c_{p,i}\right) \tag{9.6}$$

$$\frac{\partial c_{p,i}}{\partial t} = k_{eff,i}\frac{3}{r_p\varepsilon_p}\left(c_i - c_{p,i}\right) - \frac{1-\varepsilon_p}{\varepsilon_p}\frac{\partial q_i}{\partial t} \tag{9.7}$$

$$\frac{\partial c_i}{\partial x}\left(0,t\right) = \frac{u}{D_{ax}}\left(c_i\left(0,t\right) - c_{in,i}\left(t\right)\right) \tag{9.8}$$

$$\frac{\partial c_i}{\partial x}\left(L,t\right) = 0 \tag{9.9}$$

The model was simulated one-dimensionally in space. Consequently, the concentrations depend on the axial position in the column, only. As a quasi-linear form of the SMA isotherm is used, all concentration terms c_i and q_i in Eqs. (9.3) - (9.9) can be transformed into absorptions using the absorption factors.

9.2.1.4 Pareto Optimization

Yields and purities of the elution fractions were investigated for all possible product fractionation setups. Finding a global optimum for a multi-objective problem can be accomplished by Pareto optimization. Single objective problems mostly result in single solutions. For multi-objective problems, by contrast, infinite Pareto optimal solutions exist, as the single objectives interfere. The so-called Pareto front describes the feasible system points of all possible single parameter combinations.

9.2.2   Materials

9.2.2.1   Disposables & Reaction Vessels

The nucleic acid sequence of Glutathione-S-Transferase (GST) derived from *S. japonicum* was inserted into the pSCherry vector of the Cherry™ Express T7 protein expression kit (Delphi Genetics, Belgium). The storage of cells was realized in a Cryobank™ strain maintenance kit for microorganisms (Mast Diagnostica, Germany). Micro-scale cultivations were carried out in 48-well FlowerPlates® covered with adhesive sterile sealing foil (m2p-Labs, Germany). Pre-cultures were cultivated in 1 L baffled flasks (Schott, Germany). Cell harvest and removal of cell debris after disruption was carried out in 15 mL centrifugal tubes (Falcon, Germany). Absorption measurements were carried out in 96-well flat bottom UV-Star® half-area micro-plates (Greiner Bio-One, Germany). Anion-exchange chromatography experiments were conducted in a '1 mL' HiTrap™ Q FF column with effective volume of 0.962 mL (GE Healthcare Life Sciences, Sweden). For filtration of samples, 0.2 μm polyethersulfone sterile filters were applied (VWR, Germany). Capillary gel electrophoresis (GX II) was performed in an HT Protein Express & Pico LabChip® (Perkin Elmer, USA). Sample preparation for GX II was conducted in skirted 96-well twin.tec® PCR plates (Eppendorf, Germany).

9.2.2.2   Chemicals & Buffers

As a cultivation medium, standard Terrific Broth (TB) medium was used, including 12 g/L wheat peptone for microbiology (Fluka, Germany), 24 g/L bacteriological yeast extract (Amresco, USA), 5 g/L glycerol bidistilled 99.5%, 17 mM potassium dihydrogen phosphate, and 7 mM di-potassium hydrogen phosphate (VWR, Germany). Isopropyl-β-D-thiogalactopyranosid (IPTG) from a 0.25 M stock solution (VWR, Germany) was applied for the induction of the pSCherry's T7 promotor. As a lysis buffer for cell disruption, the 20 mM Tris-HCl AEX running buffer was used (VWR, Germany) adjusted to pH 8 with hydrochloric acid (Merck, Germany). 1X SigmaFAST™ Protease Inhibitor (Sigma Aldrich, Germany) was added for prevention of protease degradation. The binding buffer for the 1 mL HiTrap™ Q FF column was composed of 20 mM Tris-HCl at pH 8, as described above. The elution buffer consisted of the same buffer, including 1 M NaCl (Merck, Germany). For column regeneration, a 0.1 M sodium hydroxide solution (Sigma Aldrich, Germany) was applied. The pH adjustment of all buffers was done by titration using sodium hydroxide (Merck, Germany). A 1% acetone solution (Merck, Germany) and a 10 g/L MW 2,000,000 dextran solution from *leuconostoc spp.* (Sigma Aldrich) were applied to determine porosities and dispersion properties. The total ionic capacity was determined by acid-base titration using 0.5 M HCl solution and 0.01 M NaOH solution (Merck, Germany). For the capillary gel electrophoresis (GX II) experiments, an HT protein express reagent kit (Perkin Elmer, USA) was used. 1 mg/mL Lysozyme solution served as an internal concentration standard (Hampton research, USA).

9.2.2.3   Instrumentation & Software

Micro-scale cultivations were performed in a BioLector® MB micro-scale fermentation system equipped with the BioLection® HMI & analysis software (m2p-Labs, Germany).

Lab-scale fermentations were carried out in a MaxQ™ 6000 incubator (Thermo scientific, USA). Cell harvest and clarification after disruption were carried out in a 5810 R centrifuge (Eppendorf, Germany). The cells were disrupted using a Digital Sonifier® 450 (Branson Ultrasonic Corporation, USA). Absorption measurements were conducted in an Infinite M200 Reader controlled with I-control 1.9 (Tecan, Germany). For the chromatography procedures, an ÄKTA™ Purifier system (GE Healthcare Life Sciences, Sweden) was used, which was equipped with a pump P-900, mixer M-925, UV-detector UV-900, motor valve INV-907, pH and conductivity monitoring unit pH/C-900, and a fraction collector Frac-950 unit. The FPLC system was controlled using Unicorn 5.2 (GE Healthcare Life Sciences, Sweden). The pH of all buffers was adjusted with an HI-3220 pH meter (Hanna Instruments, USA). Random sampling, data processing, and creation of figures were performed in Matlab® R2011a (MathWorks, USA). Simulations and modeling were performed using the software developed in-house, called ChromX (Karlsruhe Institute of Technology, Germany). Capillary gel electrophoresis (GX II) was carried out in a Caliper LabChip® GX II using the LabChip® GX 3.1 software (Perkin Elmer, USA).

9.2.3   Experimental Setup

9.2.3.1   Micro-scale Cultivations

For the micro-scale cultivations, an equally treated pre-culture was used. A 250 mL TB medium cultivation was prepared by inoculating with an *E. coli* SE1 Cherry-GST cryo culture from the Cryobank™ strain maintenance kit in a 1 L baffled shake flask. The cultivation was performed in the MaxQ™ 6000 incubator at 170 rpm and 37 °C. After 17 h (end of the exponential growth), the cells were diluted in sterile TB medium to a starting $OD_{600\ nm}$ for the BioLector® experiments of 0.1 AU. Each well of the 48-well FlowerPlate® was then filled with 1 mL of diluted pre-culture. The micro-scale cultivation plate was subsequently sealed with an adhesive gas-permeable sterile membrane.

All cultivations were performed at a BioLector® shaking speed of 600 rpm. As the different cultivation processes were performed in quadruplicates, 12 different conditions could be screened per plate in total. Per 48-well FlowerPlate®, three inducer concentrations (IPTG concentrations of 0.1, 0.5, and 1 mM) and four induction times ($OD_{600\ nm}$ of 1, 2, 4, and 8) were investigated. As the BioLector® system measures scattered light signals instead of optical densities at 600 nm $OD_{600\ nm}$, a correlation function was used. For 600 rpm shaking frequency, the linear correlation slope of 17.26 SU/AU was determined. As soon as the scattered light signal exceeded the pre-defined induction threshold value, the respective well was induced by adding IPTG. The above-mentioned procedure was repeated for three different cultivation temperatures of 27, 32, and 37 °C as a full factorial experimental design. The cultivation procedure was stopped at 12 h after induction by transferring the cell broth to a 2 mL 96-well square deep-well plate and by centrifuging at 4000 rpm and 10 °C for 30 min. After discarding the supernatants, the cell pellets were frozen at -20 °C. The frozen cells from 1 mL cultivations were resuspended in 1.25 mL of lysis buffer and the quadruplicates were pooled. The cell disruption for the 5 mL suspensions was performed in 15 mL centrifugal tubes in the Branson Digital Sonifier® 450 using a 1/8" tapered micro-tip sonication probe at 30% maximal power output and a total treatment time of 90 s (6 × 15 s pulse on and 30 s pulse off for prevention of heat

Table 9.1: Column- and system-specific parameters

Parameter	Symbol	Value
Diameter	d	7 mm
Length	L	25 mm
Bead radius	r_p	0.045 mm
System dead volume	V_d	0.07 mL
Retention volume acetone	V_{RetAc}	0.96 mL
Retention volume dextran	V_{RetDex}	0.34 mL
Standard deviation of dextran	σ_{Dex}	0.029 mL
Volume of HCl	V_{HCl}	1.48 mL
Molarity of HCl	c_{HCl}	0.01 M
Flow rate	u	0.2 $\frac{mm}{s}$

denaturation). The sonication probe was fixed 0.5 cm above the centrifugal tube bottom. Lysate clarification was accomplished by centrifugation at 12000 rpm and 10 °C for 60 min in a 5810 R centrifuge. The supernatants of all cultivations were then analyzed using the Cherry-Tag™ analytics at 414 nm. 414 nm Cherry-GST signals could be correlated directly to concentrations via the previously determined absorption coefficient of 2.213 mg/(mL·AU$_{414\ nm}$) (data not shown).

9.2.3.2   Random Sampling Procedure

As a feasibility study, 5 upstream conditions were chosen randomly for ChromX lysate characterization, instead of choosing the cultivations of highest titers. First, the *'Threshold Criterion'* was defined: Conditions resulting in Cherry-GST concentrations below 0.4 mg/mL were deleted for economic reasons and replaced by a concentration of 0 mg/mL. All conditions were then scaled between 0 (= 0 mg/mL) and 1 (= maximal titer investigated), which is referred to as the *'Probability Criterion'*. The random sampling procedure of 5 cultivation conditions was then performed in Matlab® R2011a. The probabilities of being chosen were based on the *'Probability Criterion'*. This means that cultivations of lower product concentrations have lower probabilities of being picked, but can still be chosen for downstream investigation.

9.2.3.3   Model Parameter Determination

For the ChromX model, different chromatography column- and system- specific parameters were determined (compare *'Mathematical Methods'* section). The procedure followed the description of [129]. The bed porosity ε_b and the axial dispersion D_{ax} of the '1 mL' HiTrap™ Q FF column were determined with a 20 µL injection of 10 g/L MW 2,000,000 dextran solution from *leuconostoc spp.*. The large dextran molecules are excluded from the pores and, thus, only access the interstitial area of the chromatography column. For the particle porosity ε_p, a 20 µL injection of 1% acetone solution was performed. In contrast to this, the small acetone molecules have access to all pores and are a measure for the overall porosity of the system. The system's dead volume V_d was determined by acetone injection without connecting the column to the FPLC system.

Table 9.2: Model parameters derived from the column- and system-specific parameters

Parameter	Symbol	Value	Proceeding
Volume	V	0.962 mL	$\frac{1}{4} \cdot \pi \cdot d^2 \cdot L$
Fluid volume	V_f	0.89 mL	$V_{RetAc} - V_d$
Interstitial volume	V_{int}	0.27 mL	$V_{RetDex} - V_d$
Total column porosity	ε_t	0.925	$\frac{V_f}{V}$
Interstitial porosity	ε_b	0.28	$\frac{V_{int}}{V}$
Particle porosity	ε_p	0.896	$\frac{V_f - V_{int}}{V - V_{int}}$
Interstitial flow	u_{int}	0.714 $\frac{mm}{s}$	$\frac{u}{\varepsilon_b}$
Axial dispersion	D_{ax}	0.1 $\frac{mm^2}{s}$	$\frac{1}{2} \cdot L \cdot u_{int} \cdot \left(\frac{\sigma_{Dex}}{V_{int}} \right)^2$
Ionic capacity	Λ	0.22 M	$\frac{c_{HCl} \cdot V_{HCl}}{V(1-\varepsilon_b)(1-\varepsilon_p)}$

For the total ionic capacity Λ, an acid-base titration of the ion-exchange column was carried out. For this purpose, the column was flushed with a 0.5 M NaOH solution until the UV and conductivity signal was constant and all functional groups were blocked with OH^- ions. Following a wash with ultrapure water, a titration with 0.01 M HCl solution followed at a flow of 0.2 mm/s. The amount of acid needed until the conductivity signal increased was recorded. From the number of Cl^- ions replacing the OH^- ions, the total number of binding sites was determined. All other column properties were taken from the manufacturer. All given and determined parameters are listed in Table 9.1. The calculated model parameters and the underlying proceedings are shown in Table 9.2.

9.2.3.4   ChromX Model Calibration

For ChromX model calibration, 4 different gradient experiments were conducted for each cell lysate selected in the 'Random Sampling Procedure'. The number of gradients should ideally at least be equal to the number of estimated parameters (here: k_{kin}, k_{eq}, k_{eff}, ν). Each sample was 4 times diluted in 20 mM Tris binding buffer pH 8, with a 1 mL injection being ensured for each gradient. After sample application, the column was washed for 2 CVs with binding buffer. The elution was conducted by increasing the NaCl concentration up to 1 M applying 5, 10, 15, and 20 CV gradients. Including steep and shallow slopes helps the created model better predict steps and shallow gradients during the *in silico* optimization procedure. The high-salt stage was kept constant for another 2 CVs, followed by a re-equilibration step for 5 CVs. The resulting chromatograms from Unicorn 5.2 were then exported and implemented in the ChromX model. The areas of the contaminant peaks in each 20 CV run (the longest gradient shows the most 'visible' peaks) were used as the 'absorption factors' in the ChromX model.

Finally, the boundary conditions of all estimated parameters were set (namely for k_{kin}, k_{eq}, k_{eff}, ν) to accelerate the estimation process and to exclude non-logical solutions as e.g. negative values for the equilibrium binding coefficient k_{eq}. For initial simulations, a genetic algorithm was used to prevent local minima in the search space, followed by a deterministic Levenberg-Marquardt algorithm for result refinement. To determine the amount of protein loaded onto the chromatography column the fractions taken from the 20 CV gradients were measured in the LabChip® GX II device as duplicates using the

HT Protein Express LabChip®️ kit. The sample and chip preparation procedures were performed as described in the manufacturer's protocol for the HT Protein Express Assay (*www.bioneer.co.kr*). The results were analyzed by peak baseline integration using the HT Protein Express 200 assay in the LabChip®️ GX 3.1 software.

9.2.3.5 Pareto Optimization

The ChromX model of each cell lysate selected in the *'Random Sampling Procedure'* was used for an *in silico* optimization procedure. The operational setup proposed for product purification was subsequently validated using the ÄKTA™ Purifier system for all 5 selected samples. The best system of all was chosen based on a combination of excellent up- and downstream performance using a Pareto optimization method. The overall recovery Rec_i was defined as the initial titer from the cultivation $c_{init,i}$ multiplied by the yield Y_i in the downstream process of a sample cultivated under condition i, as shown in Eq. (9.10). The product recoveries were determined *in silico* by integration of the Cherry-GST- specific 414 nm signal. The purity Pur_i was based on integrating the 280 nm signals in ChromX. The 280 nm signal of a purified Cherry-GST sample served as a 100% purity reference $A_{280nm,Cherry-GST,i}$ and was compared to the actual 280 nm signal of the elution fraction $A_{280nm,Elu,i}$, as shown in Eq. (9.11).

$$Rec_i = c_{init,i}Y_i \tag{9.10}$$

$$Pur_i = \frac{A_{280nm,Cherry-GST,i}}{A_{280nm,Elu,i}} \tag{9.11}$$

An *in silico* fractionation was performed to scan all product fractionation possibilities. The resulting pooled recoveries and purities were then plotted in a Pareto optimization diagram, resulting in a Pareto front of optimal operational conditions for each upstream setup. The comparison of those fronts results in a global Pareto front of all selected cultivation conditions.

9.3 Results & Discussion

9.3.1 Micro-scale Cultivations & Random Sampling

All cultivations were performed at a BioLector®️ shaking speed of 600 rpm. Incubation temperatures of 27, 32, and 37 °C were investigated. On each 48-well FlowerPlate®️, four different induction times at cell densities of $OD_{600\ nm}$ 1, 2, 4, and 8 AU and three inducer concentrations of 0.1, 0.5, and 1 mM IPTG were investigated as quadruplicates. The product titers were determined after cell disruption by absorption measurements of the cleared lysates at 414 nm. Those data were converted into concentrations using the Cherry-GST extinction coefficient. The results of the micro-scale cultivation screening is shown in Fig. 9.2.

The maximal Cherry-GST titer of 0.85 mg/mL was determined for the 37 °C setup applying a late induction at $OD_{600\ nm}$ of 8 with 0.1 mg/mL of IPTG. Generally, a pronounced trend of increased soluble product formation towards late induction was observed for the
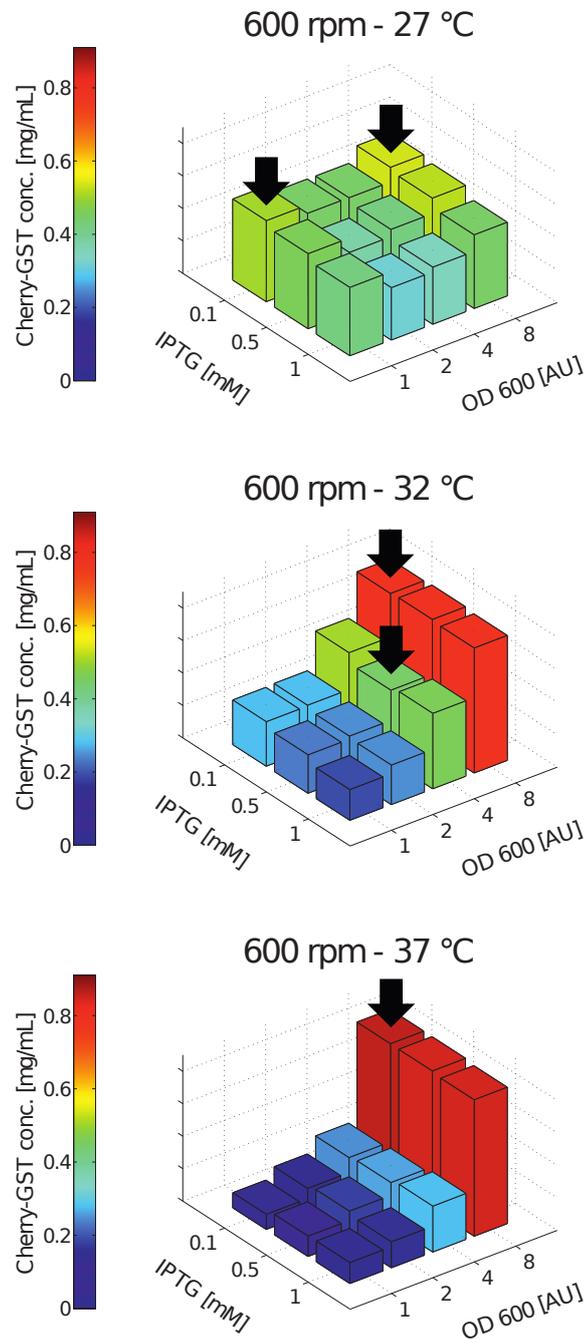
Figure 9.2: Product titers of soluble Cherry-GST [mg/mL] in *E. coli* SE1 cultivated at 600 rpm shaker frequency under different cultivation conditions. The investigated temperatures were 27, 32, and 37 °C. Additionally, three inducer concentrations (IPTG concentrations of 0.1, 0.5, and 1 mM) and four induction times (OD$_{600\ nm}$ of 1, 2, 4, and 8) were investigated. The sample selection of the *'Random Sampling Procedure'* is indicated by black arrows.
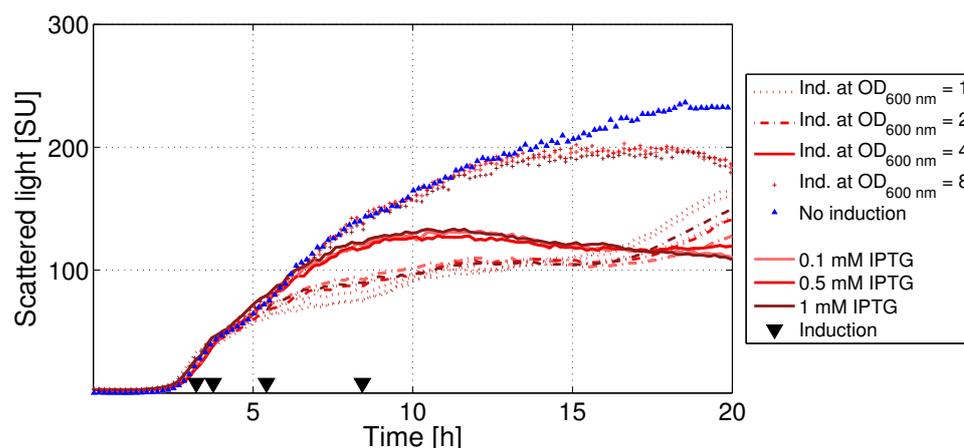
Figure 9.3: BioLector® growth curves of *E. coli* SE1 cultivated at 600 rpm shaker frequency under different cultivation conditions. Bacterial growth is shown for the 37 °C cultivation at three inducer concentrations (IPTG concentrations of 0.1, 0.5, and 1 mM) and four induction times ($OD_{600\,nm}$ of 1, 2, 4, and 8).

setups of 32 and 37 °C. Here, cells induced at $OD_{600\,nm}$ of 1 resulted in product titers of only 0.09 to 0.27 mg/mL. In contrast to this, product titers during induction at $OD_{600\,nm}$ of 8 were much higher, ranging from 0.77 to 0.85 mg/mL. Total protein extraction under denaturing conditions revealed that early induced cells resulted in inclusion bodies (data not shown). This implies that an early induction of the *E. coli* SE1 T7-promoter leads to an overproduction of Cherry-GST and, thus, to high amounts of insoluble and incorrectly folded proteins at 32 and 37 °C. The 27 °C setups showed a different trend. Here, almost constant Cherry-GST levels were obtained for all different induction setups in the range from 0.32 to 0.54 mg/mL. As *E. coli* is a mesophilic organism, lowered temperatures result in a downward regulation of the metabolism and, thus, in slowly and correctly folded Cherry-GST. However, the expression levels are far below those determined at higher cultivation temperatures. For all investigated setups, the IPTG concentration did not show any major influence on soluble Cherry-GST formation (Fig. 9.2). Also the growth curves of different inducer concentrations at 37°C (Fig. 9.3) do not deviate largely. In contrast to this, growth curves for different induction times and cultivation temperatures (data not shown) differed significantly, indicating large variances in contaminant levels of the respective cell lysates. Those variations in contaminant concentrations can be related to adaption of *E. coli* to changes in temperature, cell cycle and metabolism. Changes in cultivation temperature induce e.g. the formation of heat or cold shock proteins, temperature-adapted enzymes, and alternative transport mechanism proteins [25; 172; 199; 202; 206]. Additionally, different induction conditions resulted in a change of the cell cycle and overburden metabolism [49; 64]. It was therefore suggested that the overall system optimized for up- and downstream processing must not necessarily be the one with highest product titers. As a feasibility study, 5 upstream conditions were chosen randomly in Matlab® according to the description of the *'Random Sampling Procedure'* in the experimental section. The selected setups (Fig. 9.2 - arrows) included:

- 27 °C - $OD_{600\,nm}$ of 1 AU - 0.1 mM IPTG

- 27 °C - $OD_{600\,nm}$ of 8 AU - 0.1 mM IPTG

- 32 °C - $OD_{600\ nm}$ of 4 AU - 0.5 mM IPTG

- 32 °C - $OD_{600\ nm}$ of 8 AU - 0.1 mM IPTG

- 37 °C - $OD_{600\ nm}$ of 8 AU - 0.1 mM IPTG

9.3.2   ChromX Model Calibration

For the selected lysates from the micro-scale cultivations 5, 10, 15, and 20 CV gradients were performed in the '1 mL' HiTrap™ Q FF column for ChromX model calibration. To evaluate whether the assumption of low sample loading is valid in the presented setup the collected fractions derived from the 20 CV gradients of all setups were analyzed using quantitative capillary gel electrophoresis. It was shown that the amount of protein bound to the '1 mL' HiTrap™ Q FF column ranged from 0.2 to 1% of the nominal maximum binding capacity (120 mg HSA per mL chromatography resin - Product information sheet *www.sigmaaldrich.com*). Thus, the simplification of the SMA model was justified.
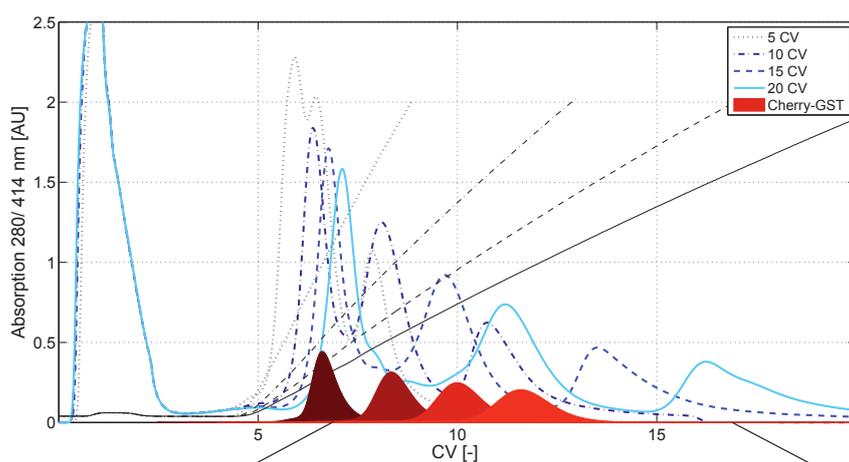
As an example, the chromatograms of the calibration experiments are shown for the 32 °C cultivation induced at $OD_{600\ nm} = 8$ with 0.1 mM IPTG in Fig. 9.4A. Flowthrough of non-binding and weakly binding contaminants can be observed during the first 3 CVs. The number of detectable peaks during the salt gradient elution starting after 5 CVs increases with the length of the gradient. Also the elution of Cherry-GST, which is shown as a red area, is shifted to higher column volumes and results in an increased product peak broadening. The 20 CV gradient served as a measure of the number of modeled contaminant peaks, which was identified to be 10 for all setups. The zero baseline peak integration in Unicorn resulted in the peak areas shown in Table 9.3. Those areas were used as scaling factors for the ChromX model. Obviously, there is no systematic trend of one cultivation condition towards higher or lower peak areas. This means that the previously determined differences in the growth curves due to altered metabolism were confirmed by ion-exchange chromatograms.

A magnification of the product elution window of the 20 CV gradient for all selected cultivation conditions is shown in Fig. 9.4B. Note that for comparison, the signals are scaled to an equal product peak area. The $UV_{280\ nm}$ peak maximum after 11.3 CVs is not equivalent to the Cherry-GST $VIS_{414\ nm}$ peak maximum after 11.7 CVs, indicating a closely eluting contaminant. The levels of this critical contaminant (CC) strongly differ within the 5

Table 9.3: Areas of all *'visible'* peaks of the 20 CV gradients determined by zero baseline integration in Unicorn. The 414 nm peak was used for product-specific analysis.

|	27 °C $OD_{600\ nm}$ 1 Peak area $[mAU \cdot mL]$	27 °C $OD_{600\ nm}$ 8 Peak area $[mAU \cdot mL]$	32 °C $OD_{600\ nm}$ 4 Peak area $[mAU \cdot mL]$	32 °C $OD_{600\ nm}$ 8 Peak area $[mAU \cdot mL]$	37 °C $OD_{600\ nm}$ 8 Peak area $[mAU \cdot mL]$
Peak 1	1369	711	1455	1049	441
Peak 2	1149	1569	1041	1374	2071
Peak 3	178	210	150	183	309
Peak 4	1661	1651	1042	1374	1594
Peak 5	278	309	189	230	274
Peak 6	198	240	166	205	373
Peak 7	211	256	204	287	376
Peak 8	890	1047	935	1425	1798
Peak 9	574	768	480	688	918
Peak 10	399	548	373	486	653
Peak$_{414\ nm}$	259	260	199	356	459
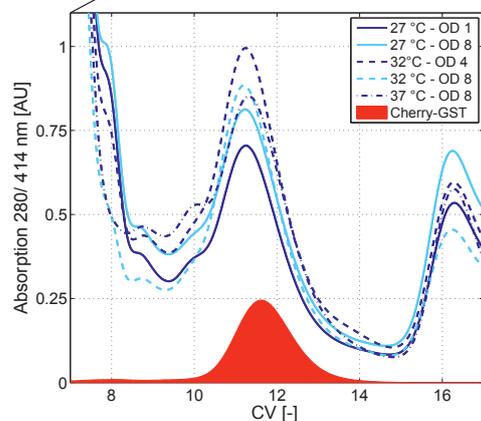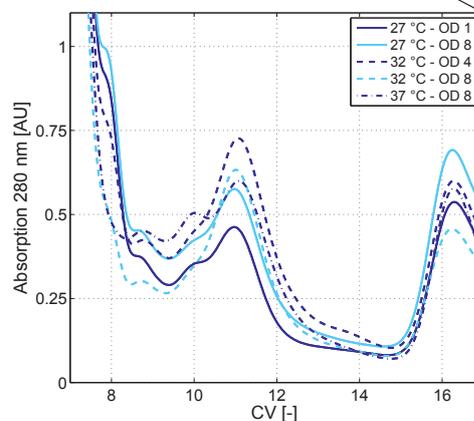
Figure 9.4: Chromatograms of the ChromX model calibration experiments. The $UV_{280\ nm}$ signals are shown as blue lines. Product peaks are indicated as red areas below. The salt concentrations are illustrated as black lines. A: 5, 10, 15, and 20 CV gradients of the lysate derived from the 32 °C cultivation, induced at $OD_{600\ nm} = 8$ with 0.1 mM IPTG. B: Magnification of the product elution window for the 20 CV gradient for all selected cultivation conditions. C: Magnification of the 280 nm signal of the product elution window for the 20 CV gradient for all selected cultivation conditions after subtraction of

Table 9.4: Converging SMA parameters of the target and the 10 contaminants for the 5 selected lysates.

Component	k_{eff}	k_{kin}	k_{eq}	ν
Contaminant 1	0.00049	0.000	0.0	0.000
Contaminant 2	0.00559	0.000	22.5	0.566
Contaminant 3	0.01200	0.000	24.7	1.156
Contaminant 4	0.01300	0.001	21.5	1.890
Contaminant 5	0.01208	0.050	26.2	3.930
Contaminant 6	0.01208	0.020	70.2	4.300
Contaminant 7	0.00808	0.090	243.2	5.670
Contaminant 8	0.00153	0.100	1623.4	6.159
Target	0.00128	0.000	757.1	6.589
Contaminant 9	0.01548	0.125	152227.0	8.968

investigated lysates as shown in Fig. 9.4C illustrating the overall 280 nm signal subtracted by the 280 nm signal of Cherry-GST. The 27 °C cultivation induced at $OD_{600\ nm} = 1$ with 0.1 mM IPTG reveals the lowest CC levels, resulting in an $UV_{280\ nm}$ peak maximum of 0.45 AU. In contrast to this, the 32 °C cultivation induced at $OD_{600\ nm} = 4$ with 0.5 mM IPTG reveals the highest CC levels with an $UV_{280\ nm}$ peak maximum of 0.72 AU. The quantitative variances in CC levels can also be seen in the areas of peak 8 in Table 9.3.

The chromatograms were exported from Unicorn 5.2 and implemented in the ChromX simulation. Including the scaling factors and the model parameters (Table 9.2), the model parameter estimates were obtained using first a heuristic and then a deterministic solver as described in the experimental section. The experimental and simulated data for the 32 °C cultivation induced at $OD_{600\ nm} = 8$ with 0.1 mM IPTG are shown for the 10 CV and 20 CV gradients in Fig. 9.5. The chromatogram derived from Unicorn (blue dashed line) agrees very well with the $UV_{280\ nm}$ sum signal in ChromX for all cultivation setups and gradients. Hence, the calibration was successful and the models could be used for *in silico* optimization studies. The single contaminant peaks 1 to 10 (grey lines) were resolved decoupled from the $UV_{280\ nm}$ sum signal, and single SMA parameters were estimated. All USP setups consisted of the identical number of 10 contaminant peaks, as was discussed before. As the contaminant levels of all investigated setups differed strongly, regions of overlapping peaks resulted in shifted peak maxima when comparing different cell lysates. For that reason it was first assumed that those shifted peaks incorporated highly distinct protein species and separate models were created for all 5 cell lysates, incorporating the 4 calibration gradients each. However, the SMA parameter sets of all different lysates converged towards similar values as shown in Table 9.4.

Thus, ChromX revealed that all USP setups consisted of 10 resemblant contaminant peaks which differed in their concentration levels only. To investigate this assumption, the same modeling experiments were carried out in one batch for all lysates, forcing the model to result in identical parameters for all characterized feed stocks. This simulation was found to have a performance equivalent to that of the above-mentioned setup to model all 5 setups in a decoupled way in a much shorter time frame. This even makes an investigation all conditions of micro-scale cultivations feasible as a full factorial design instead of performing a random sampling as shown in this study. For the selected cell systems and cultivation setups, however, it might also be possible to investigate new or different contaminant peaks. Then, the decoupled simulation methodology must be chosen.
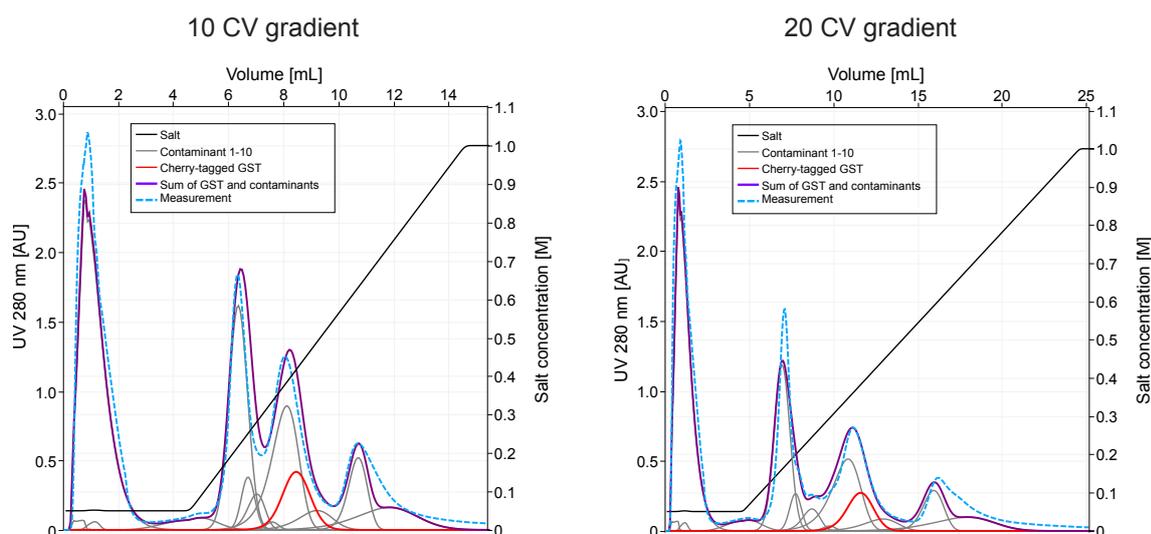
Figure 9.5: 10 CV and 20 CV gradients of the experimental (blue dashed lines) and ChromX-simulated (purple solid line) chromatograms are shown for the 32 °C cultivation induced at $OD_{600\ nm} = 8$ with 0.1 mM IPTG. The single contaminant peaks (gray lines) and the product peak (red line) are indicated below the $UV_{280\ nm}$ sum signal.

9.3.3   ChromX Optimization

Using the created models for the 5 lysates, an *in silico* optimization was performed with the help of the ChromX software. The emphasis of the optimization function was laid on yield and purity, resulting in identical purification setups for all samples investigated: After sample application, a 1 CV low-salt wash and a step gradient to 0.21 M NaCl followed, which was kept constant for 5 CVs. Afterwards, the salt concentration was shifted in a step to 0.37 M, followed by a 5 CV gradient to 0.47 M NaCl. In a final high-salt wash and re-equilibration step, the column was prepared for the next chromatography run.

The proposed optimal purification setups were subsequently validated as shown in Fig. 9.6 for the experimental and simulated data of the 32 °C cultivation, induced at $OD_{600\ nm} = 8$ with 0.1 mM IPTG. Again, the chromatogram derived from Unicorn (blue dashed line) agrees very well with the the $UV_{280\ nm}$ sum signal in ChromX for all cultivation setups. Thus, the chromatogram was correctly predicted by performing 4 different shallow and steep calibration gradients only.

9.3.4   Integrated Up- & Downstream Performance

The integrated up- and downstream performance was investigated using a Pareto optimization method. The single contaminant peaks 1 to 10 (gray lines) and the product peak (red line) from the optimized purification model were used to calculate the purities Pur_i and recoveries Rec_i for all possible peak fractionation setups, as was described in the experimental section. The product recoveries and purities strongly differed depending on the fractionation boundaries. The *in silico* procedure of scanning all product fractionation possibilities resulted in the Pareto fronts for the 5 lysates shown in Fig. 9.7. The overlaps of those optimal system points resulted in a global Pareto front of all lysates, which
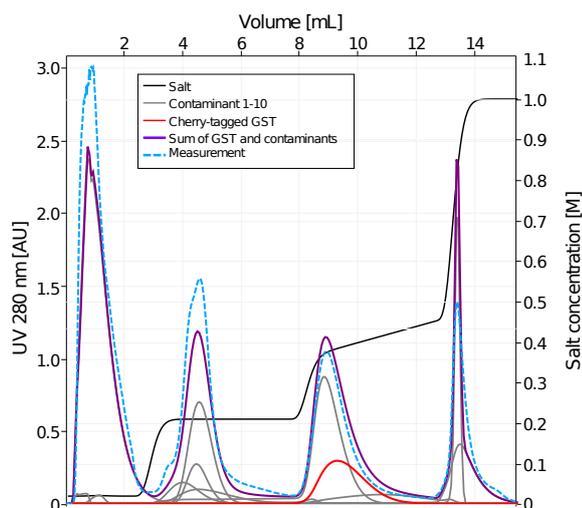
Figure 9.6: Optimized purification setup of the experimental (blue dashed lines) and ChromX-simulated (purple solid line) chromatograms shown for the 32 °C cultivation induced at $OD_{600\ nm} = 8$ with 0.1 mM IPTG. The single contaminant peaks (gray lines) and the product peak (red line) are indicated below the $UV_{280\ nm}$ sum signal.

describes the feasible conditions. The optimum recovery was found for the cultivation at 37 °C induced at $OD_{600\ nm} = 8$ with 0.1 mM IPTG (green triangles in Fig. 9.7). This result is quite intuitive, as most parts of the product peak are fractionated for this setup and the cultivation of highest initial upstream product titers performs best. However, the optimum purity was found for the cultivation at 27 °C induced at $OD_{600\ nm} = 1$ with 0.1 mM IPTG resulting in a purity of up to 65% (dark blue circles in Fig. 9.7). For all other cultivation setups, a purity above 58% could not be reached. The lowest process performance was identified for the cultivation at 32 °C induced at $OD_{600\ nm} = 4$ with 0.5 mM IPTG (dark red stars in Fig. 9.7). As discussed, the global Pareto front describes a set of optimal system points which have to be selected based on the value of the product.

It becomes obvious that the cultivation conditions are of crucial importance to downstream processing and the overall process performance due to different contaminant levels and species. The critical contaminants (CCs) which elute close to the product are the factors having the highest influence.

9.4   Conclusions

The optimization approach based on high-throughput micro-scale cultivation experiments, chromatography modeling, and Pareto optimization was found to be an elegant tool for integrated up- and downstream process optimization. Micro-scale cultivations in the Bio-Lector® system helped find feasible cultivation conditions. Besides variations of product titers, large differences in the growth curves were observed, indicating distinct contaminant levels. Instead of selecting cultivations of highest Cherry-GST titers, a feasibility study applying random sampling was conducted. Two criteria were specified: To eliminate all conditions of marginal product concentrations, a *'Threshold Criterion'* was given. Sampling from the resulting conditions was then performed randomly, with the selection
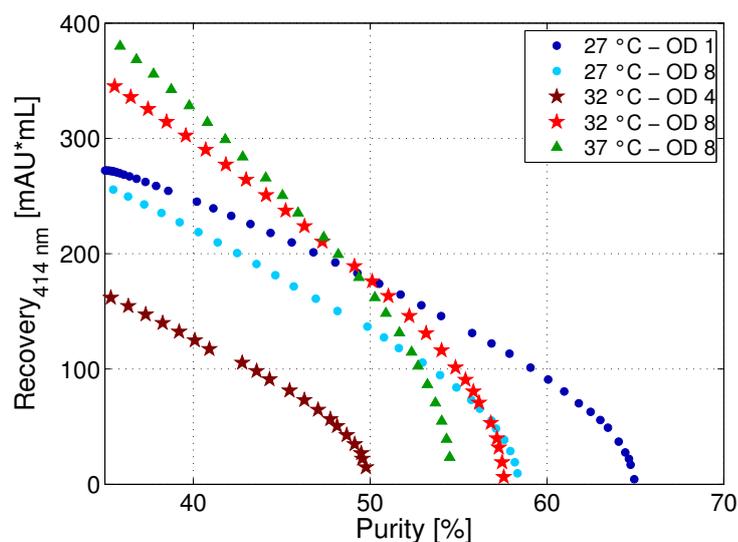
Figure 9.7: Pareto optimization based on different fractionation setups based on the purities and recoveries of the 5 selected lysates in the pooled elution fractions. The Pareto fronts of the 27 °C setups are shown as blue circles, the 32 °C setups as red and orange stars, and the 37 °C setup as a green triangle.

probability being based on the product titers (*'Probability Criterion'*). For the 5 chosen lysates, chromatography models were successfully created using 4 calibration gradients in an AEX chromatography column. ChromX revealed that all investigated USP setups consisted of 10 contaminant peaks at different concentration levels only. One coupled model with scaled contaminant levels and identical SMA parameter sets for all investigated lysates was found to perform similarly to decoupled models allowing for distinct SMA parameters sets for the five selected lysates. This even allows for an investigation of all conditions of the micro-scale cultivations as a full factorial design instead of performing a random sampling as shown in this study. However, for cell systems where new contaminant peaks are investigated, induced by different cultivation conditions, the decoupled simulation methodology is the method of choice. An *in silico* optimization of the AEX purification process based on the created models was performed and successfully validated. It was shown that the shallow and steep gradients from the calibration runs could correctly predict steps as well as gradients. As a final step, the integrated up- and downstream performance was determined by a Pareto optimization based on product recovery (initial titer · yield) and purity. The Pareto fronts of the selected lysates resulted in a global set of optimal system points which need to be set based on the value of the product.

In summary, the integrated approach showed large potential for concerted up- and downstream optimization and the upstream conditions were found to result in strong variations in the levels of critical contaminants (CC). The procedure now needs to be tested for further systems, e.g. for eucaryotic organims like *P. pastoris*.

10 | Modeling and Simulation of Anion-exchange Membrane Chromatography for Purification of Sf9 Insect Cell-derived Virus-like Particles

Christopher Ladd Effio[1], Tobias Hahn[1], Julia Seiler[1], Stefan Oelmeier[1,2], Iris Asen[3], Christine Silberer[3], Louis Villain[4], Jürgen Hubbuch[1]

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

[2] Boehringer Ingelheim Pharma GmbH & Co. KG, Biberach, Germany

[3] Diarect AG, Freiburg, Germany

[4] Sartorius Stedim Biotech, Göttingen, Germany

Abstract

Recombinant protein-based virus-like particles (VLPs) are steadily gaining in importance as innovative vaccines against cancer and infectious diseases. VLPs carry no replicative viral genetic information and are produced as large protein assemblies in recombinant expression systems. Multiple VLPs are currently evaluated in clinical phases requiring a straightforward and rational process design. To date, there is no generic platform process available for the purification of VLPs. Numerous process steps are needed to separate undesired host cell impurities. In order to accelerate and simplify VLP downstream processing, there is a demand for novel development approaches, technologies, and purification tools. Membrane adsorbers have been identified as promising stationary phases for the processing of bionanoparticles due to their large pore sizes. In this work, we present the potential of two strategies for designing VLP processes following the basic tenet of 'quality by design': High-throughput experimentation and process modeling of an anion-exchange membrane capture step. Automated membrane screenings allowed the identification of optimal VLP binding conditions yielding a dynamic binding capacity of 5.7 mg/mL for

human B19 parvovirus-like particles derived from *Spodoptera frugiperda Sf* 9 insect cells. A mechanistic approach was implemented for radial ion-exchange membrane chromatography using the lumped-rate model and steric mass action model for the *in silico* optimization of a VLP capture step. For the first time, process modeling enabled the *in silico* design of a selective, robust and scalable process with minimal experimental effort for a complex VLP feedstock. The optimized anion-exchange membrane chromatography process resulted in a protein purity of 81.5%, a DNA clearance of 99.2%, and a VLP recovery of 59%.

10.1   Introduction

Virus-like particles (VLPs) are highly ordered protein assemblies mimicking the structure of viruses. Recent progress in the development of VLP-based vaccines against highly pathogenic viruses and organisms such as Ebola [130], Malaria [2] or Influenza [54] highlights the importance of this novel molecule class for the global health system. Ensuring a fast and economic availability of VLP vaccines requires a rapid and robust manufacturing process. VLPs are currently produced in recombinant systems such as bacteria, yeast, insect or plant cells necessitating subsequent downstream process steps for the separation of host-cell impurities to minimize the risk of side effects [108]. The most prevalent unit operation for the purification of biopharmaceutical products is ion-exchange chromatography. Traditional chromatography media are based on adsorber particles with pore sizes in the nanometer region causing mass transfer limitations and low binding capacities for bionanoparticles such as viruses or VLPs [108; 205]. Among others, membrane adsorbers have been developed and constantly advanced for the purification of large biomolecules in bind-and-elute mode [126; 146; 178; 192–194]. Membrane adsorbers are composed of multiple flat sheets of functionalized membranes with pore sizes in the range of micrometers. An increase in the specific surface in such flat sheets can be achieved by grafting hydrogels on membranes to embed ligands [187]. According to Nestola et al. hydrogel-grafted membranes lead to higher recoveries for virus purifications than directly grafted membranes [138]. The geometries of membrane adsorbers are usually either stacked flat sheet or spiral wound modules with axial or radial flow. Regarding separation performance parameters such as resolution, peak width, and dynamic binding capacity, there are no outstanding differences between axial and radial flow geometries [17; 181]. However, radial flow geometries are especially suited for large-scale applications by virtue of lower pressure drops as demonstrated by Besselink et al. [17]: Wide, short axial flow columns can be similarly replaced with low-pressure narrow, tall radial-flow columns without a significant change in performance. The combination of both radial flow geometry and convection-driven mass transport in membrane adsorbers allows high throughputs and productivities for biopharmaceutical products [147; 186].

In recent years, the process development for biologics has been relying predominantly on empirical studies. Minimizing the risks of manufacturing failures and lot-to-lot variations of biological products has been the key motivation for the FDA to forward the approaches 'quality by design' (QbD) and 'process analytical technology' (PAT) [162]. One major contribution to and important tool for process control and design is mechanistic modeling. Multiple thermodynamic and hydrodynamic models have been developed and steadily extended for chromatographic separations of proteins [34; 59; 60; 74; 77; 78; 132]. Recently, novel hydrodynamic models were established for membrane adsorbers including axial and

radial flow and validated in various protein case studies [48; 187]. In contrast, there have been only few publications on modeling and simulation of VLP processes. Vicente et al. demonstrated the applicability of the steric mass action model [24] for purified rotavirus-like particles with a mean diameter of 80 nm on stacked flat sheet anion-exchange (AEX) membrane adsorbers [194]. However, the measured and simulated chromatograms for the complex VLP feedstock differed significantly and up to now there has not been any case study reporting on the *in silico* optimization of a VLP separation. The difficulty for *in silico* simulations and optimizations of complex biological feedstocks lies in modeling the mass transfer of components of unknown size and concentration. An approach to dealing with this was previously reported by Hahn et al. [76] and Baumann et al. [13]. In both case studies, UV absorption-based modeling was implemented to simulate the elution profile of complex biological feedstock components using UV peak areas instead of molar concentrations.

In this work, we present a case study for the purification of human B19 parvo-VLPs derived from *Spotoptera frugiperda Sf* 9 insect cells applying both high-throughput experimentation and UV absorption-based chromatography modeling. Human B19 parvo-VLPs have a size of 25 to 30 nm and can be assembled in insect cells by expressing the major capsid protein VP2. The particles are currently evaluated in clinical phase studies as vaccine candidates against diseases attributed to parvovirus infections such as fifth disease in children, anaemia, and hydrops fetalis [16; 29]. The developed membrane process allowed the separation of major contaminants such as host cell proteins and DNA in one step and can be predicted and controlled by *in silico* process simulations.

10.2   Materials & Methods

10.2.1   Disposables

For screening purposes on a liquid handling station, 1 mL-96 deep well polypropylene plates (Thermo Fisher Scientific Inc., Waltham, Massachusetts, USA) and 96-well half area polystyrene flat-bottom plates by Greiner Bio-One (Kremsmünster, Austria) were used. Fraction collection on a FPLC system was done with 2 mL-96 deep well polypropylene plates (Nalge Nunc International, Rochester, USA). Sterile filtration of buffer solutions was conducted with 0.2 μm cellulose acetate filters (Sartorius AG, Goettingen, Germany). Buffer exchange was performed with PD-10 desalting columns (GE Healthcare, Uppsala, Sweden).

10.2.2   Chemicals & Buffers

Bis-Tris, Bis-Tris propane and dextran T2000 were purchased from Sigma-Aldrich (St. Louis, USA). Guanidine hydrochloride from AppliChem GmbH (Darmstadt, Germany) and dithiothreitol from Amresco (Solon, USA) were used for the preparation of UHPLC samples. All other chemicals were purchased from Merck KGaA (Darmstadt, Germany). For automated batch membrane chromatography experiments, a low and high pH multi-component buffer with a buffer capacity of 10 mM, composed of Bis-Tris, Bis-Tris propane

and Tris was used to obtain a linear pH gradient between pH 7.5 and pH 9.0. The buffer recipe was calculated according to Kröner and Hubbuch [107]. 10% (v/v) glycerol and 0.5% (v/v) polysorbate 80 were added to all ion-exchange buffers to prevent aggregation and secondary interactions of VLPs [168]. FPLC runs were performed with 20 mM Tris pH 8.5, 0.5% (v/v) polysorbate 80 and 10% (v/v) glycerol (buffer A) and 20 mM Tris pH 8.5, 1 M NaCl, 0.5% (v/v) polysorbate 80 and 10% (v/v) glycerol (buffer B). Membranes were regenerated with 1 M NaOH and 3 M guanidine hydrochloride and stored in 20% (v/v) ethanol in 20 mM Tris pH 8.5.

10.2.3 Human B19 Parvovirus-like Particles

Human B19 parvo-VLPs composed of the major viral capsid protein VP2 were derived from *Sf*9 insect cells. Insect cell pellets were provided by Diarect AG (Freiburg, Germany) and lysed and clarified as described previously by sonication, centrifugation, and filtration [109]. Prior to membrane experiments, buffer exchange was performed using PD-10 desalting columns (GE Healthcare, Uppsala, Sweden).

10.2.4 Anion-exchange Membrane Chromatography

All AEX chromatography experiments were performed with Sartobind® Q membranes from Sartorius AG (Goettingen, Germany). Binding studies for VLPs were conducted in 96-well membrane plates on a robotic liquid handling system. Membrane characterization and salt elution studies with clarified VLP feedstocks were done with 3 mL Sartobind® nano Q membrane capsules. Experiments with membrane capsules were performed on an ÄKTA-purifier 10 fast protein liquid chromatography (FPLC) equipped with a pump module P-900 (up to 10 mL/min), UV (10 mm path length) monitor (UV-900), a conductivity monitor, and a fraction collector Frac-950 (GE Healthcare, Uppsala, Sweden). The instrument was controlled using the software UNICORN 5.31 (GE Healthcare, Uppsala, Sweden).

10.2.4.1 Hydrodynamic & Thermodynamic Models

FPLC experiments were conducted with radial spiral-wound membrane adsorbers as displayed in Fig. 10.1. A general rate model for radial flow chromatography has been developed by Gu et al. [67] considering radial dispersion, convection, and pore diffusion of a component i in radial direction x using differential mass balances. The continuity equation for the component i with a lumped effective mass transfer coefficient $k_{eff,i}$ incorporating film and pore diffusion is given by Eq. (10.1):

$$\frac{\partial c_i}{\partial t} = \left(\pm u(x) + \frac{D_{ax}}{x} \right) \frac{\partial c_i}{\partial x} + D_{ax} \frac{\partial^2 c_i}{\partial x^2} - \frac{(1 - \varepsilon_{pores})}{\varepsilon_{pores}} k_{eff,i} \left(c_i - c_{p,i} \right) \qquad i = 1, \dots, N$$

(10.1)

The algebraic sign of the convective transport term $u(x)$ is positive for inward flow which was the case for the membrane capsules used in this work. In Eq.(10.1), the first term describes the convective transport, the second term the hydrodynamic dispersion, and the
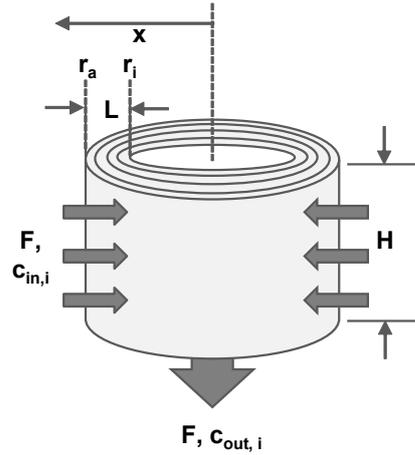
Figure 10.1: Schematic illustration of a spiral wound membrane adsorber with inward radial flow F, membrane thickness L, membrane height H, outer radius r_a and inner radius r_i.

last term refers to the transport of component i between bulk and membrane pore volume $c_i - c_{p,i}$ for the membrane porosity ε_{pores}. Danckwerts boundary conditions [39; 189] can be used to describe the effect of diffusion-dispersion phenomena on the concentration at the column inlet $x = r_a$ (*outer radius*)(Eq. (10.2)). In contrast, the outlet concentration at $x = r_i$ (*inner radius*) for inward flow is unaffected by these phenomena (Eq. (10.3)).

$$x = r_a : \qquad \frac{\partial c_i}{\partial x}(t, x = r_a) = \frac{u(x = r_a)}{D_{ax}}\left(c_i(t, x = r_a) - c_{in,i}(t)\right) \qquad (10.2)$$

$$x = r_i : \qquad \frac{\partial c_i}{\partial x}(t, x = r_i) = 0 \qquad (10.3)$$

As described previously by Van Beijeren et al. [187], the frontal area A_{front} for a spiral wound geometry depends on the radial position x and the membrane height H (Eq. (10.4)):

$$A_{front} = 2\pi x H \qquad (10.4)$$

Thus, for inward flow with the volumetric flow rate F, the linear interstitial velocity u increases towards the center of the membrane adsorber (Eq. (10.5)):

$$u(x) = \frac{F}{\varepsilon_{Pores} 2\pi x H} \qquad (10.5)$$

Most membrane adsorbers include a grafted hydrogel layer with embedded ligands to obtain a higher specific surface area [178]. Such hydrogels may exclude large molecules such as dextran T2000 and viruses [178; 192]. The hydrogel mass transfer can be modeled by incorporating a hydrogel porositiy $\varepsilon_{hydrogel}$. Hence, the mass balance for the stationary phase considers accumulation and mass transport inside the membrane pores (Eq. (10.6)):

$$\frac{\partial c_{p,i}}{\partial t} = \frac{k_{eff,i}}{\varepsilon_{Hydrogel}}\left(c_i - c_{p,i}\right) - \frac{(1 - \varepsilon_{Hydrogel})}{\varepsilon_{Hydrogel}}\frac{\partial q_i}{\partial t} \qquad i = 1, \dots, N \qquad (10.6)$$

The mass transport of component i in the liquid phase is described by the film transfer coefficient $k_{eff,i}$, the pore concentration $c_{p,i}$, the hydrogel porosity $\varepsilon_{Hydrogel}$, and the concentration bound to the stationary phase q_i. The adsorption term can be described for

ion-exchange (IEX) chromatography with the steric mass action (SMA) model developed by Brooks and Cramer [24]. The semi-mechanistic isotherm considers several effects for IEX adsorption and desorption processes:

- Concentration of available binding sites given by the total ionic capacity Λ

- Shielding of ligands by large macromolecules i described by the shielding factor σ_i

- Number of macromolecules' binding sites described by the characteristic charge ν_i

- Ratio of adsorption $k_{ads,i}$ and desorption constant $k_{des,i}$ depicted by the equilibrium term $k_{eq,i} = \frac{k_{ads,i}}{k_{des,i}}$

- Changing salt concentrations $c_{p,salt}$ in the pore volume

The rate of change of the concentration in the stationary phase q_i can be modeled with the kinetic form of the SMA isotherm (Eq. (10.7)) [77]:

$$\frac{1}{k_{des,i}} \frac{\partial q_i}{\partial t} = k_{eq,i} \left(\Lambda - \sum_{j=1}^{N} (\nu_j + \sigma_j) q_j \right)^{\nu_i} c_{p,i} - c_{p,salt}^{\nu_i} q_i \qquad i = 1, \ldots, N \qquad (10.7)$$

Eq. (10.8) describes the competition of salt ions and other components for binding sites in the stationary phase:

$$q_{salt} = \Lambda - \sum_{j=1}^{N} \nu_j q_j \qquad (10.8)$$

For estimating SMA and hydrodynamic parameters of complex feedstocks with multiple components of unknown concentration, the absorption coefficient a_i is used. The factor scales the molar concentrations required for modeling mass transport in the stationary phase to absorbance units. Thus, Eq. (10.7) can be written as (Eq. (10.9)):

$$\frac{1}{k_{des,i}} \frac{\partial q_i}{\partial t} = k_{eq,i} \left(\Lambda - \sum_{j=1}^{N} \frac{(\nu_j + \sigma_j)}{a_j} q_j \right)^{\nu_i} c_{p,i} - c_{p,salt}^{\nu_i} q_i \qquad i = 1, \ldots, N \qquad (10.9)$$

According to Baumann et al. [13], the total ionic capacity predominates, and occupied ligands can be neglected in the case of small sample loadings which do not approach the maximum binding capacity of the stationary phase. This makes it feasible to estimate most SMA and hydrodynamic parameters directly from chromatograms by UV absorption-based modeling as demonstrated previously [13; 74; 76].

Parameter estimation was performed as described previously by chromatogram fitting using a genetic algorithm [85]. Both, parameter estimation and *in silico* optimization of a VLP capture process, were conducted with the chromatography software ChromX [75].

10.2.4.2   Automated Batch Binding Screenings

Optimal binding conditions for VLPs on the AEX membranes were determined by miniaturized and automated microbatch experiments using Sartobind® Q 96-well plates with

a membrane area of 0.7 cm^2. The high-throughput screening procedure followed the same routine for adsorption isotherms and binding screenings: Equilibration, centrifugation, sample preparation, incubation, loading, centrifugation and preparation of analytical samples. Membranes were equilibrated with 2 mL buffer per well at the desired pH and ionic strength. Centrifugation was performed at 1000 rpm for 10 min. Sample preparation was conducted by mixing stock solutions of buffers and VLP feedstocks at different dilutions in 96-DeepWell plates. Systems were incubated for 10 min on an orbital shaker at 1000 rpm. Subsequently, 300 μL sample volume was added to each well of the 96-well membrane plate. Thereafter, samples were centrifuged and VLPs in the flow-through were tracked by RP-UHPLC as described below. The adsorption isotherm screening was performed under buffer conditions similar to the initial cell lysis step at 50mM Tris pH 7.4, 50 mM NaCl, 10% (v/v) glycerol and 0.5% (v/v) polysorbate 80. Data was fitted to the Langmuir isotherm. Eq. (10.10) [35; 84] describes the correlation for the equilibrium concentration in the mobile phase $c_{eq,i}$ and the concentration in the stationary phase q_i assuming a saturation capacity q_S and an equilibrium constant $k_{eq,i}$.

$$q_i = \frac{q_s k_{eq,L,i} c_{eq,i}}{1 + k_{eq,L,i} c_{eq,i}} \tag{10.10}$$

Subsequent to the determination of VLP adsorption isotherms, the goal was to find optimal buffer conditions for the binding step due to possible displacement effects by other components in the VLP feedstock. Four stock solutions with and without VLP feedstock were prepared at pH 7.5 and pH 9 with 15 and 120 mM NaCl using the multicomponent buffer described under Section 10.2.2. Buffers were mixed in the appropriate volume ratio to obtain a Design of Experiments (DoE) set-up with four pH values and eight ionic strengths with triplicate determinations. Method and analysis were the same as for the isotherm determination. To obtain optimal binding conditions, the starting VLP concentration in each well was chosen from the non-linear region of the adsorption isotherm.

10.2.4.3   System & Membrane Characterization

Modeling of hydrodynamic and thermodynamic mass transfer in the membrane capsule and FPLC system requires knowledge of several membrane and system characteristics. The determination of system dead volumes V_d, membrane porosity ε_{pores} and hydrogel porosity $\varepsilon_{hydogel}$ was done by injecting small and large tracer molecules: 1% v/v acetone, 1 M NaCl, and 10 mg/mL dextran T2000. 100 μL of the tracer substances were injected via an ÄKTA sample loop with a twofold loop overfill and tracked using UV and conductivity detectors. Acetone and NaCl were used to determine the system dead volume from sample loop to UV and conductivity detector, respectively. Dextran T2000 served as large tracer component for the determination of the interstitial volume V_{int}, the membrane porosity ε_{pores} and the hydrogel porosity $\varepsilon_{hydogel}$. The ionic capacity of the anion-exchange membrane was obtained by performing a titration with 0.5 M NaOH and 0.01 M HCl according to standard protocols [85; 149] (Eq. (10.11)):

$$\Lambda = \frac{c_{HCl} V_{HCl}}{V_M (1 - \varepsilon_t)} \tag{10.11}$$

All tracer runs and acid-base titrations were performed in at least five copies. Membrane characteristics such as membrane thickness L, membrane height H, outer radius r_a, and

Table 10.1: Measured membrane parameters

Parameter	Symbol	Value	Proceeding
Membrane volume	V_M	3 mL	From manufacturer
Membrane thickness	L	8 mm	From manufacturer
Outer radius	r_a	11.25 mm	From manufacturer
Inner radius	r_i	3.25 mm	From manufacturer
Membrane height	H	8 mm	From manufacturer
Total porosity	ε_t	0.798	From manufacturer
Flow rate	F	3 mL/min	Manually controlled
System dead volume	V_d	0.25 mL	Acetone injection without membrane
Retention volume acetone	V_{RetAc}	3.7 mL	Acetone injection with membrane
Retention volume dextran	V_{RetDex}	2.86 mL	Dextran injection with membrane
Retention volume NaCl	V_{NaCl}	3.42 mL	NaCl injection with membrane
Volume of HCl	V_{HCl}	12.8 mL	Acid-base titration
Molarity of HCl	c_{HCl}	0.01 M	Manually controlled

inner radius r_i are illustrated in Fig. 10.1. An overview of all system & membrane parameters is given in Table 10.1 providing corresponding equations for the calculated values.

Chromatograms of dextran T2000 and NaCl pulse experiments were used to estimate the hydrodynamic parameters, i.e. the axial dispersion D_{ax} and the film transfer coefficient $k_{eff,NaCl}$ by chromatogram fitting. All calculated membrane parameters are specified in Table 10.2.

10.2.4.4   Determination of Dynamic Binding Capacity

An important parameter for selecting and evaluating chromatography media is the dynamic binding capacity (DBC). The DBC of the AEX membrane adsorbers was determined for VLPs by overloading the membrane with 10 mL VLP feedstock at a VLP concentration $c_{VLP,0}$ of 2.7 mg/mL, and optimum buffer conditions, derived from the batch binding studies (20 mM Tris pH 8.5, 15 mM NaCl, 0.5% (v/v) polysorbate 80, 10% (v/v) glycerol). The breakthrough experiment was conducted at a flow rate of 3 mL/min on the ÄKTApurifier 10 FPLC. Fractions were collected and analysed by reversed-phase (RP)-UHPLC. The DBC was calculated determining the applied volume at 10% breakthrough $V_{10\%}$ from the fitted breakthrough curve (Eq. (10.12)):

$$DBC = c_{VLP,0} \frac{V_{10\%} - V_{RetDex}}{V_M} \tag{10.12}$$

10.2.4.5   Calibration of Steric Mass Action Model

For the simulation and *in silico* optimization of an AEX membrane chromatography step, a component-specific calibration of the SMA model was required. Experiments were performed with 3 mL Sartobind® Q membrane adsorbers on the ÄKTApurifier 10 FPLC. The

Table 10.2: Calculated membrane parameters

Parameter	Symbol	Value	Proceeding
Fluid volume	V_f	3.4 mL	$V_{RetAc} - V_d$
Void volume capsule	V_{voidC}	1.006 mL	$V_f - \varepsilon_t V_M$
Interstitial volume	V_{int}	2.56 mL	$V_{RetDex} - V_d - V_{voidC}$
Membrane porosity	ε_{pores}	0.518	$\frac{V_{int}}{V_M}$
Hydrogel porosity	$\varepsilon_{hydrogel}$	0.581	$\frac{\varepsilon_t - \varepsilon_{pores}}{1 - \varepsilon_{pores}}$
Axial dispersion	D_{ax}	$0.0374\,\mathrm{mm}^2/\mathrm{s}$	Chromatogram fitting of dextran peak
Film transfer coefficient NaCl	$k_{eff,NaCl}$	$0.1829\,\mathrm{mm}^2/\mathrm{s}$	Chromatogram fitting of NaCl peak
Ionic capacity	Λ	0.211 M	$\frac{c_{HCl} V_{HCl}}{V_M (1 - \varepsilon_t)}$
Volume applied at 10% breakthrough	$V_{10\%}$	9.2 mL	Interpolated from fraction analysis
Dynamic binding capacity for VLPs	DBC	5.7 mg/mL	$c_{VLP,0} \frac{V_{10\%} - V_{RetDex}}{V_M}$

composition of buffer A was 20 mM Tris pH 8.5, 0.5% (v/v) polysorbate 80, 10% (v/v) glycerol. For elution, buffer A was supplemented with 1 M NaCl to create buffer B. Regeneration buffer was 1 M NaOH and storage buffer 20% EtOH in 20 mM Tris pH 8.5. Membranes were equilibrated with 5 CV of 1.5% B prior to sample injection. 1 mL VLP feedstock was injected in each run and fractions were collected during the whole run to track the VLP elution. Three salt gradients with gradient slopes of 10 CV, 20 CV and 30 CV from 1.5% to 100% B were performed to generate chromatograms for the SMA parameter estimation. Applying the peak integration tool of the ÄKTA software UNICORN 5.31, the VLP feedstock was divided into multiple subcomponents. Each UV peak was integrated for the 30 CV gradient run and the ratio of peak area and load volume was set as concentration for the UV-based modeling [74] of each component. Estimation of the SMA parameters $k_{eq,i}$, ν_i and the film transfer coefficients $k_{eff,i}$ was done by chromatogram fitting using a genetic algorithm.

10.2.4.6 Optimization of VLP Purification

Given a defined set of hydrodynamic and thermodynamic parameters, it is possible to simulate chromatograms of a multicomponent system and to optimize the separation of a target component *in silico*. In this work, this was attempted by varying the salt concentrations of a three-step elution process and simulating the effect on purity (P) and yield (Y). Salt concentrations of the first and second salt step were varied between 0.2 and 0.5 M NaCl, while the third salt step elution was at 1 M NaCl. The least-square optimization was performed for an objective fraction volume of 3 CV. The solution of the optimization problem p_{opt} is defined as (Eq. (10.13)):

$$p_{opt} = \arg\min_p \left(\underbrace{(1 - P(p))}_{impurity} + \underbrace{(1 - Y(p))}_{loss} \right) \qquad (10.13)$$

The proposed optimal salt concentrations for the step elution process were evaluated for a loading of 1 and 6.5 mL and flow rates of 1 CV/min, 2 CV/min, and 3 CV/min.

10.2.5   Size-exclusion chromatography

Polishing of VLPs captured by an optimized salt step elution by AEX membrane chromatography was performed on a Superose® 6 Increase 10/300 column (GE Healthcare, Uppsala, Sweden). 500 μL of a 2 mg/mL VLP solution was injected at a flow rate of 0.5 mL/min using PBS pH 7.4 as mobile phase buffer. Fractions were analyzed by RP-UHPLC and capillary gel electrophoresis to evaluate the final purity of the samples.

10.2.6   Analytical methods

10.2.6.1   VLP Quantification

Quantification of VP2-VLPs was conducted as described previously by RP-UHPLC [109]. In brief, a Waters Acquity BEH C4 column was run in acetonitrile gradient elution mode on an UltiMate® 3000 RSLC x 2 Dual system (Thermo Fisher Scientific, Waltham, Massachusetts, USA). Samples were pre-denatured with a denaturing buffer containing 8 M guanidine hydrochloride and 60 mM DTT to avoid column blocking by aggregation and carryover. Final concentrations were determined by comparing peak areas with external VLP reference standards of known concentration.

10.2.6.2   Capillary Gel Electrophoresis

The final protein purity of processed VLPs was assessed by a LabChip® GX II capillary gel electrophoresis device separating proteins by size. Sample and chip preparation procedures were performed as described in the manufacturer's protocol for the HT Protein Express Assay [26]. Product quantification was based on peak-baseline integration and the protein purity (P) was determined as the ratio of product concentration c_{VP2} to the total protein concentration c_{Total} (Eq. (10.14)):

$$P = \frac{c_{VP2}}{c_{Total}} \tag{10.14}$$

10.2.6.3   DNA Quantification

DNA quantification was performed using the PicoGreen® dsDNA assay kit (Invitrogen, Paisley, United Kingdom) in 96-well plates according to the manufacturer's instructions.

10.2.6.4   Transmission Electron Microscopy

FPLC fractions and purified VLP samples were further analyzed by transmission electron microscopy to evaluate polydispersity and morphology of VLPs. As described previously [109], samples were applied on carbon grids, washed with ultrapure water, and stained with uranyl acetate prior to analysis on a Philips CM 200 FEG/ST transmission electron microscope at 200 kV.
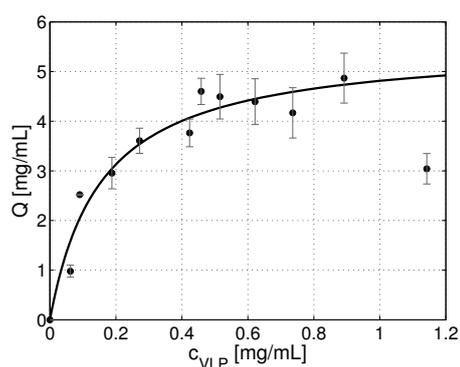
(a) VLP adsorption isotherm (50 mM Tris pH 7.4, 50 mM NaCl).
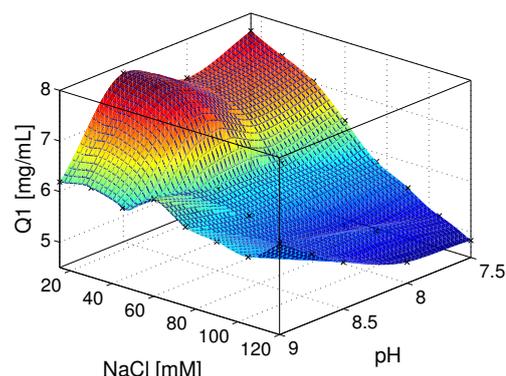
(b) Binding capacity of VLPs at varying buffer conditions.

Figure 10.2: Batch adsorption of human B19 parvo-VLPs on Sartobind® Q membranes. Experiments were performed with *Spodoptera frugiperda Sf* 9 insect cell lysate containing VLPs composed of the major capsid protein VP2. The VLP concentration in the stationary phase Q is fitted using the Langmuir isotherm (solid line). The binding screening for varying pH values and NaCl concentrations was conducted with a multicomponent buffer system at a VLP concentration of 1 mg/mL. Each data point represents the mean value of triplicate binding experiments.

10.3 Results & Discussion

10.3.1 Optimization of VLP Binding by High-throughput Experimentation

Optimal binding conditions for human B19 parvo-VLPs on AEX membranes were determined by varying pH and ionic strength of the binding buffer in automated high-throughput microbatch screenings. In a first step, the binding at pH 7.4 and 50 mM NaCl was investigated for increasing VLP concentrations. In Fig. 10.2a, the VLP concentration in the stationary phase Q is plotted against the concentration in the mobile phase c_{VLP} with standard deviations illustrated by error bars. The values were fitted using the Langmuir isotherm obtaining a saturation capacity q_S of 5.6 mg VLP per mL membrane volume. The curve shows a similar pattern as usual one-component adsorption isotherms with a linear region for low concentrations and a saturation region for high VLP concentrations. Nevertheless, there is a strong decrease in capacity for higher VLP concentrations such as 1.15 mg/mL. It must be taken into account that the VLP feedstock used for the experiments is a complex multicomponent mixture containing multiple host cell proteins (HCPs), baculoviruses, and nucleic acids (DNA). AEX ligands strongly bind DNA at pH-values higher than pH 4.0, potentially causing the displacement of bound VLPs at high loadings [58]. Besides multicomponent displacement effects, the binding of biomolecules on IEX stationary phases is also influenced by secondary interactions such as hydrophobic adsorption and desorption processes [30; 167]. Therefore, both pH and NaCl concentration were varied using a DoE approach to identify optimum binding conditions for VLPs. For a bind-and-elute procedure, the pH working range was chosen beyond the predicted isoelectric point of the major capsid protein VP2 at pH 6.5 [163]. Moreover, to minimize the risk of VLP disassembly during the IEX process, the pH working range did not exceed a pH of 9.0 [163].

(a) Simulation of axial dispersion of dextran T2000.
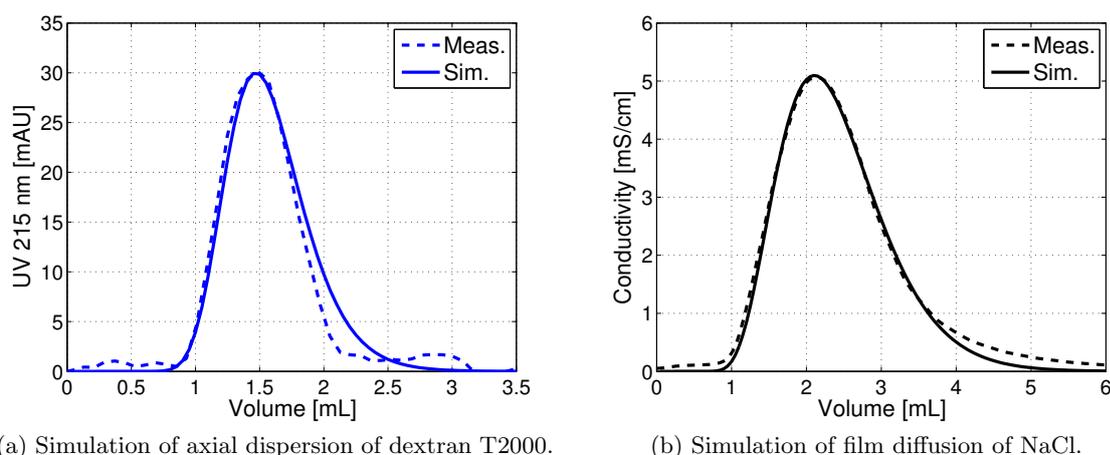

(b) Simulation of film diffusion of NaCl.

Figure 10.3: Comparison of measured (dotted lines) and simulated chromatograms (solid lines) for pulse experiments on a 3 mL Sartobind® Q membrane capsule with 100 μL of the tracer molecules dextran T 2000 (10 mg/mL) and NaCl (1 M). System and capsule void volumes are subtracted from each chromatogram.

Fig. 10.2b displays the correlation of the VLP concentration in the stationary phase $Q1$ (1 mg/mL VLP in the mobile phase) with both NaCl concentrations and pH. It can be seen that the affinity for VLPs initially increases with pH and the maximum loading concentration is obtained at pH 8.5 and 15 mM NaCl. The bound amount of VLPs at pH 7.5 and 120 mM NaCl is almost 40% lower than at the optimum condition. Interestingly, the highest pH does not come along with the highest $Q1$-value. At pH 9 and 15 mM NaCl, the loading mass is 19% lower than at pH 8.5. However, $Q1$ at pH 9.0 hardly decreases with increasing salt concentrations in contrast to pH 7.5, 8 and pH 8.5.

In general, there is a clear trend of decreased binding at increasing NaCl concentrations. Low NaCl concentrations below 15 mM lead to precipitation of VLPs, and datapoints below 15 mM NaCl are therefore not included in Fig. 10.2b. The stabilizing effect of NaCl has been reported before for various VLPs [121; 163]. The optimal binding buffer at pH 8.5 and 15 mM NaCl probably incorporates a negative VLP net charge, a low surface hydrophobicity, and a stable VLP conformation [163]. A slight pH increase to pH 9.0 implies a more negative net charge of the capsid protein VP2, but might coincide with conformational changes or even disassembly of VLPs as reported recently for hepatitis B surface antigen VLPs by Yang et al. [210].

10.3.2   Calibration of Hydrodynamic & Steric Mass Action Model

Subsequent to setting the buffer conditions for the AEX binding step by the DoE approach, elution conditions were optimized by process modeling and simulation. For modeling of the hydrodynamics in the membrane capsule, a dextran pulse was used to determine the axial dispersion coefficient D_{ax} by chromatogram fitting. In Fig. 10.3a, the UV absorption at a wave length of 215 nm is plotted against the mobile phase volume for a 10 mg/mL dextran T2000 pulse injection on a 3 mL Sartobind® Q membrane adsorber. The chromatogram shows an overlay of measured (*dotted line*) and simulated (*solid line*) UV signal. The simulation was possible by estimating the axial dispersion coefficient under consideration

of the system and capsule dead volume. All calculated membrane-specific parameters are summarized in Table 10.2.

An axial dispersion coefficient D_{ax} of 0.0374 mm^2/s was estimated which is quite low in comparison to the dispersion in agarose-based chromatography columns [85]. Thus, the determined D_{ax}-value already implies a small peak-broadening effect by axial dispersion. In order to model the ion-exchange on the membrane surface, the film diffusion coefficient of NaCl was estimated by chromatogram fitting of a 1 M NaCl pulse injection. In Fig. 10.3b, the conductivity signal is plotted against the mobile phase volume. In contrast to the dextran peak, the measured NaCl peak (*dotted line*) shows a noticeable tailing and a longer retention volume. The NaCl peak was simulated (*solid line*) using the obtained D_{ax} value from the dextran chromatogram fitting and an estimated film transfer coefficient $k_{eff,NaCl}$ of 0.1829 mm/s. The differences in the retention volume of dextran T2000 and NaCl confirm the exclusion effect of large molecules from hydrogel pores reported by Vicente et al. [192] and Tatarova et al. [178]. Both dispersion and film diffusion could be modeled using the continuity equation for radial flow with a lumped diffusion coefficient and standard tracer molecules for chromatographic stationary phases.

The SMA model was calibrated by determining the ionic capacity of the membrane (Table 10.2) and estimating component-specific SMA parameters by chromatogram fitting. In a first step, the dynamic binding of the AEX membrane for VLPs was determined by performing a breakthrough curve with the VLP feedstock. Knowledge of the dynamic binding capacity is essential to assess whether the requirements for UV absorption-based modeling are fulfilled in subsequent calibration runs (see Section 10.2.4.1). Fig. 10.4a shows the chromatogram of a breakthrough experiment: The UV absorption at 280 nm is plotted (*black solid line*) against the mobile-phase volume. The ratio of mobile-phase VLP concentration c to initial VLP concentration c_0 is illustrated by green bars with standard deviations for the analytics represented by grey error bars. The VLP breakthrough starts at 8.9 mL mobile-phase volume and the determined dynamic binding capacity at 10% breakthrough is 5.7 mg VLP per mL membrane adsorber (Table 10.2). The UV signal shows a first increase at about 3.7 min and a saddle point at approximately 9.2 mL followed by the same sharp increase as for the VLP concentration in the analyzed fractions. The sharp breakthrough of VLPs indicates a low dispersion of VLPs through the membrane pores as observed in the dextran T2000 pulse experiments for another large molecule. The achieved DBC of 5.7 mg/mL outperforms to the best of our knowledge all so far reported capacities of chromatographic matrices for VLPs [108].

VLP concentrations exceeding the starting concentration c_0 suggest a displacement effect of bound VLPs by other feedstock components with higher ligand affinities. Without the knowledge of the exact feedstock composition, modeling of the breakthrough of each component is not possible [76]. Therefore, model calibration and process simulation was only performed for low loadings with a total volume of 1 mL and a VLP mass of 2 mg corresponding to approximately 11% of the DBC. Figs. 10.4b–10.4d show the obtained UV 280 chromatograms for three different salt gradient lengths of 10, 20 and 30 CV from 0.015 M to 1 M NaCl. UV signal and VLP concentrations are again plotted by black solid lines and green bars, respectively. The conductivity is displayed by a dotted grey line. The chromatograms indicate the separation of multiple components by the salt gradient elution with increasing resolution for longer gradients. The most outstanding peak with the highest UV absorption at 280 nm elutes at a salt concentration between 0.3 and 0.5 M

(a) VLP breakthrough curve.

(b) 10 CV salt gradient.
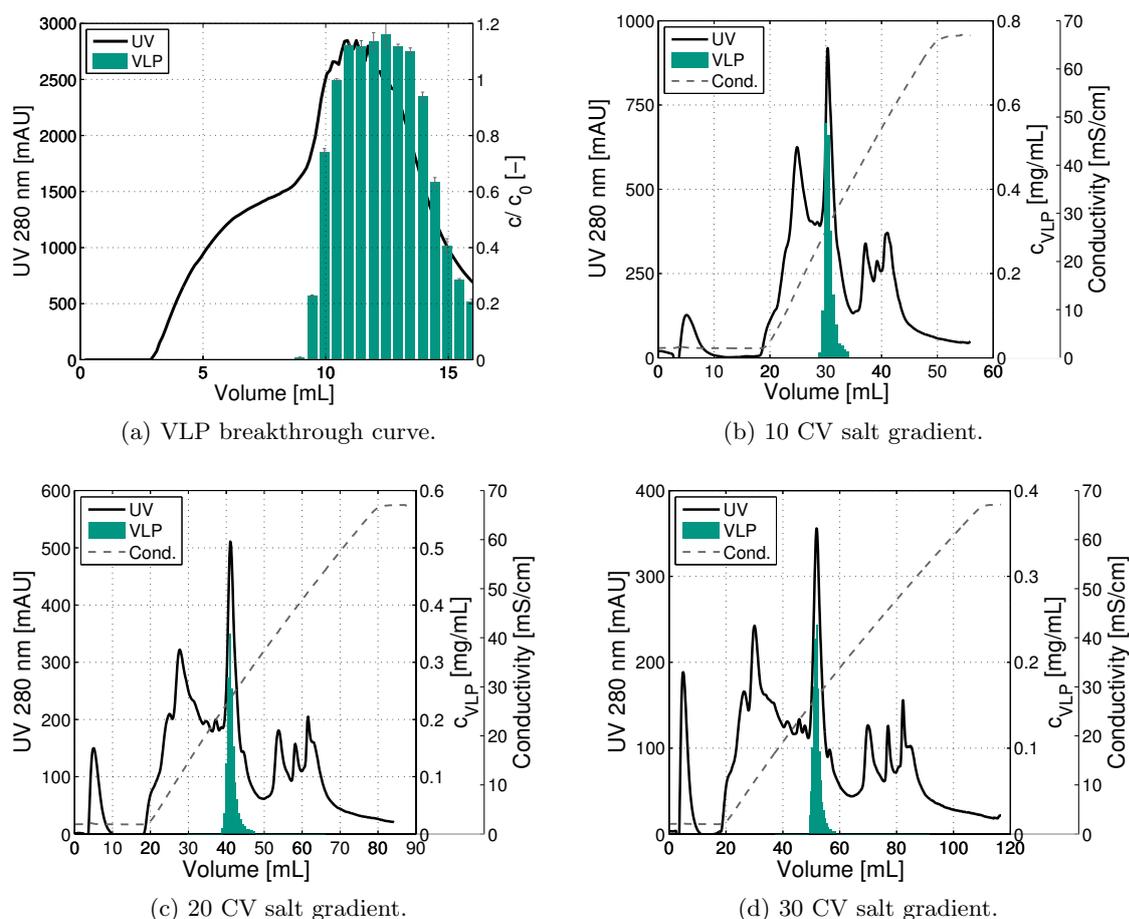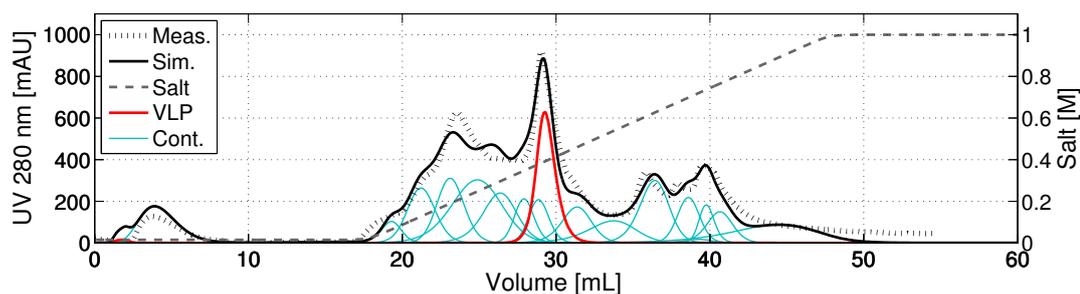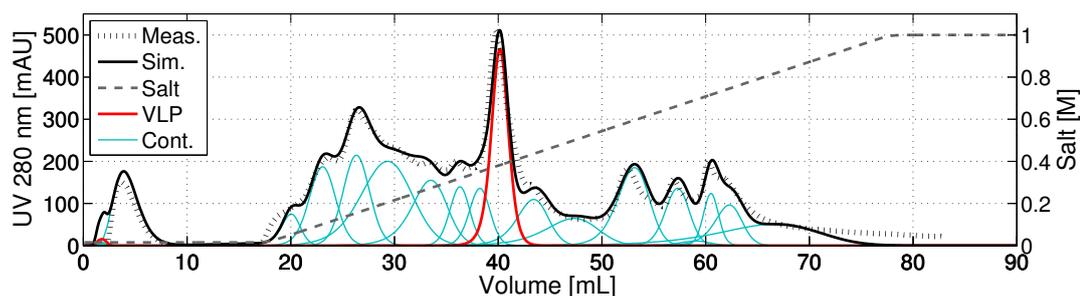
(c) 20 CV salt gradient.

(d) 30 CV salt gradient.

Figure 10.4: UV chromatograms of bind-and-elute experiments with a clarified human B19 parvo-VLP feedstock on a 3 mL Sartobind® Q membrane capsule. The experiments were conducted at a VLP concentration of 2.7 mg/mL (breakthrough experiment) and 2 mg/mL, respectively. The solid lines display the UV absorption signal at 280 nm, the dotted lines show the conductivity signals, and the green bars represent the VLP concentration in all fractions. The sample volumes were 10 mL for the breakthrough experiment and 1 mL for the salt gradient runs. Three different salt gradient slopes were applied with 10, 20 and 30 CV gradients from 15 mM to 1 M NaCl.

NaCl in between two other main peak groups. The fraction analysis of all gradient runs demonstrates that VLPs elute exclusively in the region of this major peak. As expected, VLPs get more diluted for lower salt gradient slopes. Although VLPs successfully bind and elute on the anion exchange membrane, the total recovery was only 61±3% for all experiments. The remaining amount of VLPs can only be washed out of the membrane with 1 M NaOH. Such product loss in an IEX bind-and-elute mode of up to 70% [185; 211] is often observed for VLPs and requires development and evaluation of novel chromatographic matrices and ligands. Conformational changes during adsorption and desorption processes might also elicit the product loss [210]. However, since the VLP recovery was nearly the same for both salt gradient and step elution runs in the evaluated calibration and optimization dataset, all chromatograms could be used for model calibration and process simulation. As described in Section 10.2.4.5, the UV chromatogram peaks were divided into 17 subgroups and used for SMA parameter estimation by chromatogram fit-

(a) Simulation of 10 CV salt gradient elution profile.

(b) Simulation of 20 CV salt gradient elution profile.
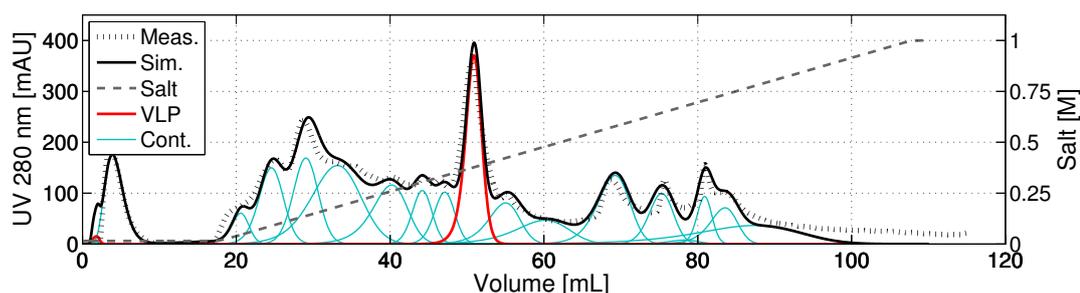
(c) Simulation of 30 CV salt gradient elution profile.

Figure 10.5: Comparison of measured (dotted lines) and simulated UV chromatograms (solid lines) for the complex VLP feedstock on the AEX membrane capsule. The feedstock was divided into 16 contaminant species (turquoise lines) and the target component (red lines). All component-specific lumped rate and SMA model parameters are provided as supplementary material in Tab. B.3. System and capsule void volumes are subtracted in each chromatogram.

ting. Figs. 10.5a–10.5c illustrate chromatograms of measured (dotted line) and simulated (solid line) elution of VLP feestock components for the 10, 20 and 30 CV salt gradients. The simulated VLP elution is displayed by a red line, while contaminant peaks are shown as turquoise lines. Estimated SMA parameters, measurement factors, and film diffusion coefficients are provided as supplementary material in Tab. B.3. The simulated chromatograms describe the measured elution with good accordance regarding peak shapes, areas, heights and retention volumes. An even higher accordance of measured and simulated chromatograms might of course be obtained by increasing the number of components for the complex feedstock. On the other side, the number of components strongly affects the modeling effort, provides few information without an extra feedstock characterization procedure, and should thus not be too high.

(a) Model-based prediction of VLP purity for vary-
ing salt step elutions.

(b) Model-based prediction of combined VLP
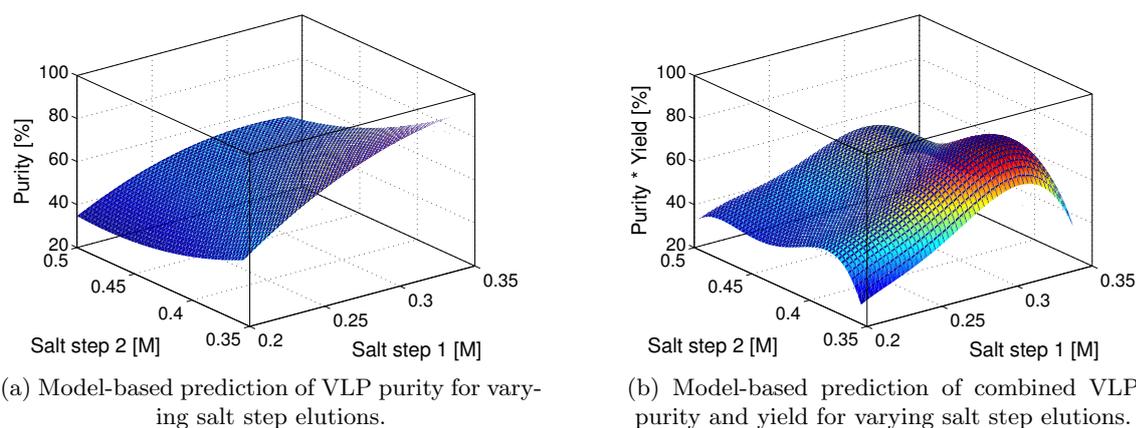purity and yield for varying salt step elutions.

Figure 10.6: *In silico* designed correlations of the model-based VLP purity and yield with
regards to salt concentrations in a three-salt step elution process with a Sartobind® Q
membrane adsorber. The calculations are solely based on the modeled contaminant groups
and component-specific SMA parameters.

10.3.3   In Silico Optimization

Knowledge of hydrodynamic and SMA parameters finally allowed the identification of
optimal elution conditions for the investigated separation problem. Using the simulation
software ChromX, the salt concentrations of a three-salt step elution process were varied
to evaluate the effect on purity and yield. Figs. 10.6a and 10.6b illustrate by 3D plots the
correlation of VLP purity and purity times yield with the salt concentrations of the first
and second salt step.

The first salt step elution is applied to separate impurities with a lower ligand affinity,
while the second step elutes both VLPs and impurities with a similar net charge. Thus,
purities and yields are only calculated for VLPs eluting in the second salt step for a fraction
volume of 3 CV. Fig. 10.6a shows that the purity increases for higher salt concentrations
in the first salt step and lower salt concentrations in the second salt step elution. This
correlation can be expected since there are sweet spots for the elution of both VLPs and
contaminants. The highest and lowest simulated purities obtained were 91% and 30%,
respectively. However, the VLP yield shows an opposite trend and increases with lower
salt concentrations of salt step 1 and higher salt concentrations of salt step 2, which
must be considered for the process design. Fig. 10.6b demonstrates that a high purity
and yield can only be reached in a narrow region of salt concentrations. A high salt
concentration of salt step 1 causes VLP elution in the first salt step while a lower salt
concentration of salt step 2 leads to VLP elution in the third salt step. The sweet spot
for an optimized capturing of VLPs is 0.3±0.02 M in salt step 1 and 0.38±0.02 M in salt
step 2. It should be noted that given purities are UV-based and calculated exclusively
for those contaminant groups included in the model calibration dataset. Moreover, the
model-based VLP yield is determined for the eluting fraction and does not account for the
product loss caused by irreversible binding or conformational change as described above.
The obtained correlations represent a design space for the AEX membrane processing of
VLPs that can be used to optimize a capture step, to scale the separation as well as
to understand and forecast the effect of process fluctuations following the basic tenet of

(a) Optimized step elution (1 mL sample volume).

(b) Measurement and fraction analysis of optimized step elution (1 mL sample volume).

(c) Optimized step elution (6.5 mL sample volume).

(d) Measurement and fraction analysis of optimized step elution (6.5 mL sample volume)
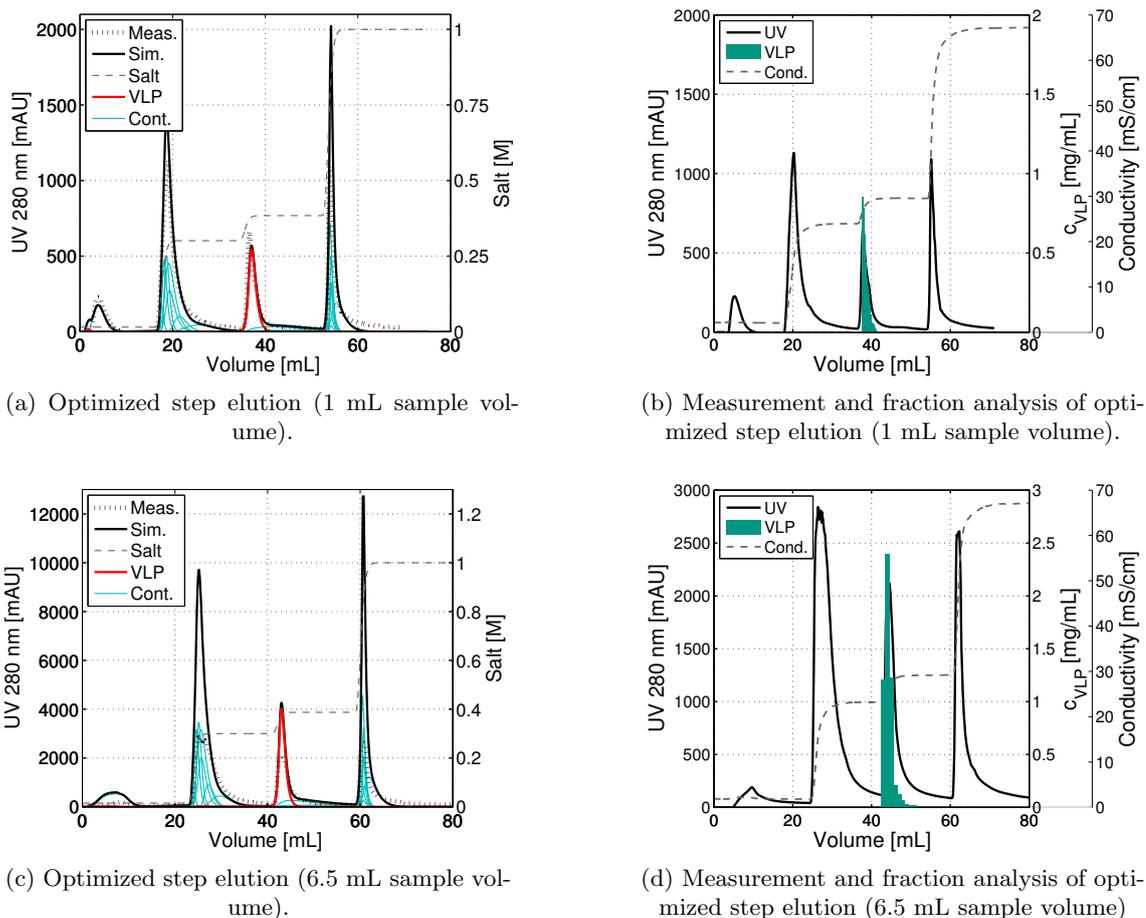
Figure 10.7: Comparison of measured (dotted lines) and simulated UV chromatograms (solid lines) for the optimized VLP capture process on the AEX membrane capsule at varying load volumes (1 mL and 6.5 mL VLP feedstock). The salt concentrations were 0.301 M NaCl for step 1, 0.384 M NaCl for step 2, and 1 M NaCl for the high-salt step. The upper limit of the UV detector is reached at values between 2500 and 3000 mAU. Product tracking by fraction analysis is shown in figures (b) and (d) displaying VLP concentrations as green bars.

'quality by design' (QbD). To evaluate the feasibility of the proposed salt step process, simulated and measured data were compared and are shown in Figs. 10.7a–10.7d. The salt concentrations are 0.301 M NaCl (salt step 1), 0.384 M NaCl (salt step 2), and 1 M NaCl (salt step 3) predicting a VLP purity of 88.6% in the second salt step. Fig. 10.7a and Fig. 10.7c display the measured (dotted lines) and simulated (solid black lines) UV signals of the purification process for a loading of 1 mL and a loading of 6.5 mL with VLPs peaks in red and contaminant peaks in turquoise. Fig. 10.7b and Fig. 10.7d show the measured UV absorption (solid black lines) and the VLP concentrations (green bars) in all collected fractions. For both loadings, VLPs eluted exclusively in the second salt step. The total yield of 59±2% was almost the same as in the calibration salt gradient runs indicating that no VLPs were lost by elution in the first or third salt step. The obtained protein purity for the pooled VLP fractions of the second salt step was 81.5±2% as determined by capillary gel electrophoresis (Fig. 10.10). The chromatograms illustrate that most contaminants elute in the first and third salt step, although two contaminant groups have similar elution

(a) Step elution at 6 mL/min.
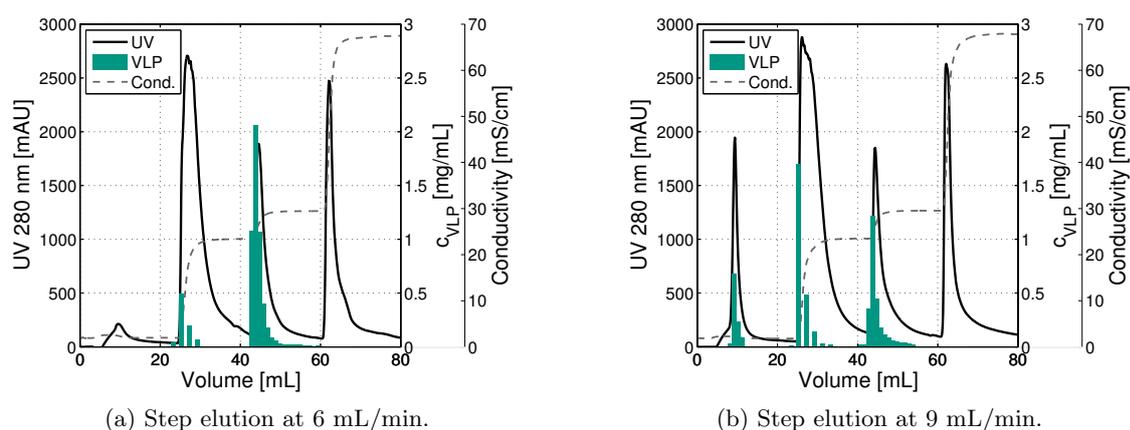

(b) Step elution at 9 mL/min.

Figure 10.8: Effect of the flow rate on the performance of the optimized salt step elution procedure of the VLP feedstock for a 3 mL Sartobind® Q membrane adsorber and a loading of 6.5 mL VLP feedstock. The solid lines display the UV absorption signal at 280 nm, the dotted lines show the conductivity signals, and the green bars represent the VLP concentration in all fractions.

characteristics as the target component and can hardly be separated. Even with longer salt gradients or other salt concentration profiles, a separation of these contaminants cannot be achieved without changing stationary and/ or mobile phase (data not shown). The chromatogram in Fig. 10.7c shows the same peak elution profile as in Fig. 10.7a, but with higher peak areas and heights. It must be taken into account that the UV detection limit lies close to 2700 mAU which leads to different peak heights at high loadings for measured and simulated chromatograms. Interestingly, the chromatogram and VLP concentrations in the second salt step demonstrate that the proposed *in silico* optimization procedure can also be used to depict the separation for higher loadings (76% of DBC) than applied in the model calibration dataset. This suggests a quite steep SMA isotherm indicating that in this case, non-linear effects could be neglected for simulating higher loadings of complex feedstocks.

10.3.4   Impact of Flow Rate

A major advantage of membranes in comparison to conventional chromatography resins is the rapid mass transport lacking intraparticle diffusion limitations [147]. In order to investigate the impact of flow rate on the bind-and-elute VLP process, the optimized salt step elution was ran at a flow rate of 2 CV/min and 3 CV/min. Fig. 10.8a and Fig. 10.8b show the obtained UV chromatograms and VLP concentrations.

In contrast to the operation at 1 CV/min, VLPs were not only detected in the second salt elution step, but also in fractions of salt step 1 and for 3 CV/min even in fractions of the flow-through. This demonstrates that an increase in flow rate drastically changes the performance and capacity of the AEX membrane process. The observed variations might be ascribed to kinetic and diffusive limitations of large biomolecules at higher flow rates. Similar phenomena have been reported by Rao [161] for the protein thyroglobulin applied on IEX membranes showing decreased DBC at higher flow rates. The findings point out

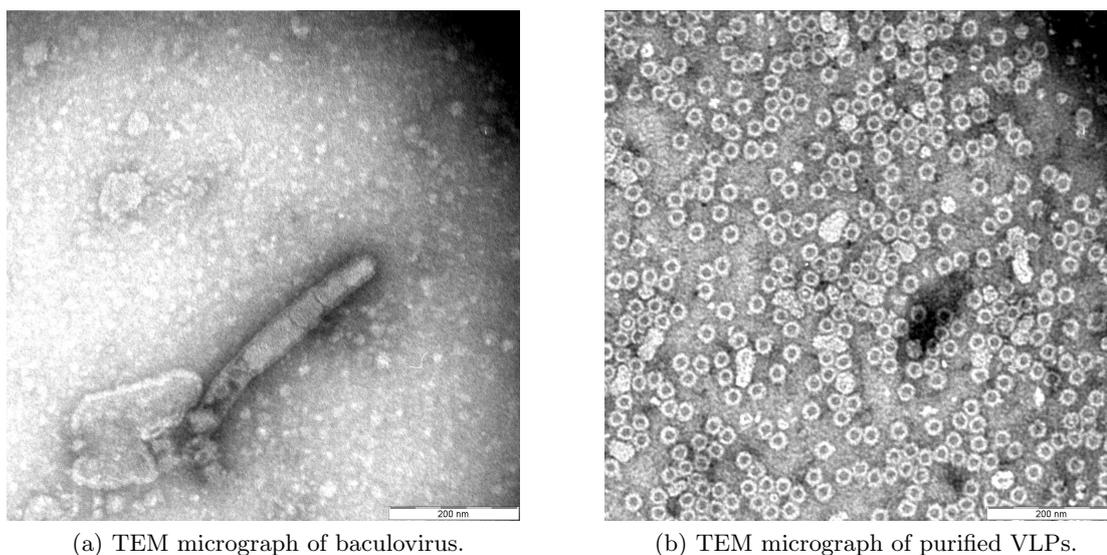(a) TEM micrograph of baculovirus.   (b) TEM micrograph of purified VLPs.

Figure 10.9: Analysis of process samples by transmission electron microscopy (TEM) and capillary gel electrophoresis. Electropherograms of process samples were normed to the target peak area.

that hydrogel and ligand design on membranes require further optimization by industry and academia to obtain higher binding capacities and faster downstream processes for large biomolecules.

10.3.5   Polishing & Process Evaluation

The designed AEX membrane process yielded a protein purity of 81.5±2 %, a DNA concentration of 8±1 ng DNA per 100μg VLP (99.2% clearance), and a VLP recovery of 59±2%. Particle morphology and homogeneity were assessed by TEM analysis. Fig. 10.9a shows the micrograph of an FPLC fraction collected at a retention volume of 28 mL. A rod-shaped particle is visible with a size of 50 x 300 nm. Comparing the particles with TEM micrographs from Vicente et al. [193] and Yang et al. [209], both size and morphology indicate the presence of baculoviruses in the first salt step of the optimized elution process. In contrast, there were not any baculoviruses observed on the micrographs of the pooled VLP fractions shown in Fig. 10.9b. Fig. 10.9b displays a high number of homogeneous icosahedral particles with a diameter of 25 nm characteristic of human B19 parvo-VLPs [29; 61; 163]. This demonstrates that the VLP structure was not affected by the bind-and-elute membrane process.

To further increase the purity of the processed VLPs, pooled fractions from the membrane step were applied on an SEC column. The UV chromatogram of the SEC step is provided as supplementary material, Fig. B.4. The protein purity after the SEC step was 97.9±1 % with a step yield of 84±2 % separating VLP aggregates and smaller biomolecules. Fig. 10.10 shows the electropherogram of the VLP feedstock (black dotted line), the membrane-processed VLPs (blue line), and the polished VLPs (red dotted line). The electropherogram underlines that the majority of contaminants is removed by the optimized AEX membrane process with only three residual protein impurity peaks
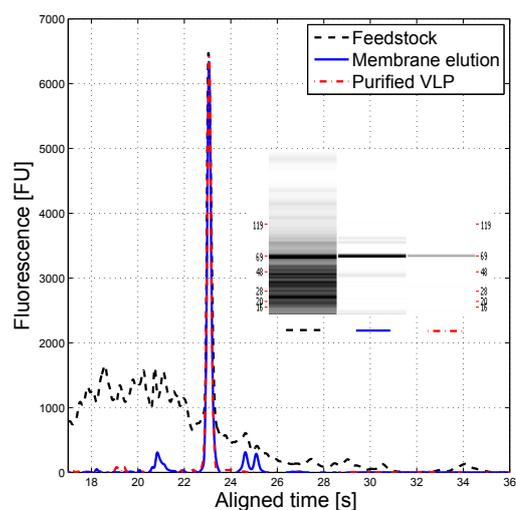
Figure 10.10: Capillary gel electropherogram.

left. Using the AEX membrane capture step and the SEC polishing procedure, all major protein impurities were successfully separated from the target component. The final DNA concentration of the purified VLPs was 0.5 ng DNA per 100 μg VLP (99.9% clearance).

10.4 Conclusion & Outlook

The aim of this study was to design a bind-and-elute membrane process for VLPs derived from Sf9 insect cells using rational development tools. For the first time, a proof of concept for simulating and optimizing the chromatographic separation of VLPs with a radial lumped rate model and the SMA isotherm was demonstrated. Automated high-throughput screenings in 96-well format allowed the identification of suitable binding conditions yielding a dynamic binding capacity of 5.7 mg VLP per mL Q membrane. *In silico* process modeling, simulation, and optimization were performed for a radial AEX membrane capsule to control and forecast the elution of VLPs and impurities. Using UV absorption-based modeling, the complex feedstock was divided into 17 subcomponents, and a design space for the elution conditions of the target component was established. An *in silico* optimized process with three salt steps enabled the separation of VLPs from the majority of baculoviruses, HCPs and DNA. The simulated and experimental purities were 88.6% and 81.5±2%, respectively. Residual contaminants were cleared by a subsequent SEC polishing procedure. A downside of the developed process was, however, the product loss (60% recovery) and the decrease in performance at increased flow rates. Both effects require a better understanding of the mass transport phenomena on the membrane for a more generic process model incorporating potential conformational VLP changes. Further membrane optimizations should focus on customizing ligand and stationary phase design for large biomolecules to increase the productivity of membrane-based VLP processes. Both high-through experimentation and chromatography modeling have proven to be valuable rational design tools for process development of VLPs.

Acknowledgements

Nomenclature

a_i	Absorption coefficient
D_{ax}	Axial dispersion coefficient
ε_{Pores}	Membrane porosity
$\varepsilon_{Hydrogel}$	Stationary phase porosity
ε_t	Total porosity
H	Membrane height
k_{eff}	Effective film/pore transfer coefficient
k_{eq}	Adsorption equilibrium coefficient
$k_{eq,L}$	Adsorption equilibrium coeff. for Langmuir isotherm
L	Membrane thickness
Λ	Stationary phase ionic capacity
ν	Characteristic charge for SMA isotherm
q	Stationary phase concentration of protein
q_s	Single-component saturation conc. for Langmuir isotherm
q_{salt}	Stationary phase concentration of salt
r_a	Outer radius of spiral wound membrane
r_i	Inner radius of spiral wound membrane
σ	Steric shielding coefficient for SMA isotherm
t	Time dimension
u	Interstitial mobile phase velocity
x	Space dimension

11 | Conclusion

In the course of this thesis, the software ChromX was developed as a highly versatile toolbox. It was employed for research (Chaps. 5, 6, 7, 9, 10), teaching (Chap. 3) and industrial application (Chap. 8). Compared to previous approaches, its modular and extensible structure allows for exploring new interaction modes (Chap. 5) as well as new formats and materials (Chap. 10) . ChromX reads measurements from standard lab equipment like the UNICORN chromatogram format and processes offline fraction analyzes. The software does not only set new standards in terms of usability (Chap. 3) but also performance. State-of-the-art mathematical methods were employed in all layers of the software: multi-core and many-core optimized linear algebra operations, direct and iterative linear solvers, stabilized space-discretization methods to cope with steep concentration gradients as well as adaptive time stepping. On top of the simulation core, adjoint-based gradient computation proved to be much more efficient than all previous attempts (Chap. 4). The method of Optimal Experimental Design (Chap. 5) solves several problems at once. It provides a measure of model quality, shows correlations between parameters and suggests experiments that will lead to higher reliability. So far, computationally expensive Monte-Carlo sampling was necessary that could only be applied to rather simple simulations with few components and parameters.

On the modeling side, competitive multi-component adsorption was studied using batch chromatography. The experimental results showed complex, non-intuitive isotherm shapes. It could be demonstrated that the Steric Mass Action (SMA) model is capable of reproducing the experimental results including displacement of the weaker binding species causing an adsorption maximum at low concentrations (Chap. 2).

In column chromatography, SMA was found to be a suitable model for both, anion and cation exchange chromatography (Chap. 6). Here, correct assumptions of the salt concentration in the binding buffer were of high importance. An incomplete buffer exchange in between the two column runs of a concerted optimization lead to deviations of experimental result and simulation. In this case study, ChromX was used as a tool for error diagnostic in order to find the actual salt concentration after buffer exchange.

In case of hydrophobic interaction and mixed-mode chromatography, the existing models had to be extended in order to match experimental results in column chromatography (Chap. 5). Both required the re-formulation as differential equation with a kinetic rate constant. For high concentration differences between low and high-salt buffer, a salt-dependent pore diffusion parameter had to be introduced.

While the molar concentrations of the species in the sample are always known for studies with a small number of model proteins, practical applications would require the purification and analysis of all the components. The method presented in Chap. 7 allows for

transforming any adsorption model to use UV adsorption as measure of concentration. It could be shown theoretically that this transformation is invertible for Langmuir-type and hydrophobic interaction models. In case of ion-exchange and mixed-mode chromatography, it was proven that the additional counter-ion balance equation for the ion-exchange ligands even allows for determining the absorbance coefficient in the Lambert-Beer equation and, hence, determine molar concentrations a posteriori. The UV-based modeling could be successfully applied to a capture step with a crude feedstock (Chap. 7) and an industrial antibody polishing step for the removal of aggregates and fragments (Chap. 8).

Similarly, ChromX was employed as analytical tool to characterize feedstocks with unknown molar concentrations obtained from different cultivation conditions (Chap. 9). It could be shown that the impurities in all feedstocks have the same SMA parameters. Hence, they are most certainly identical and only differ in their respective concentrations. Even the adsorption of highly charged nucleic acids could be accurately modeled with SMA parameters outside the usually reported range.

In Chap. 10, it could be demonstrated that simulation and optimization of chromatographic methods with ChromX is not limited to packed beds of porous media and proteins up to anti-body size: a purification step for virus-like particles could be successfully designed for membrane adsorber capsules.

12 | Outlook

While this thesis focused on simulating and optimizing single and sequentially coupled chromatography steps, the developed simulation framework could also be used for continuous processes. This type of operation is most often encountered in capture steps which are designed to extract the target component from host cell culture fluid (HCCF) and concentrate it. The main requirements are high productivity while achieving a high yield at the same time; purity is of minor importance in this first step. Typically the column is loaded to 1% of break through and a safety factor of -10% is applied to the loading time to avoid decreasing yield. Clearly, the resin capacity is not fully utilized in this strategy which lead to the exploration of Periodic Counter Current (PCC) chromatography as capture step for monoclonal antibody (mAb) production [128]. The employed Protein A affinity resin contributes significantly to the material costs in mAb manufacturing such that an increase in productivity can lead to considerable savings.

Counter current chromatography has been explored e.g. in [114], where particles move against a counter-currently flowing liquid. The achievable capacity utilization is high but the setup is inherently complex. In PCC, multiple columns are switching positions at periodic time intervals opposite to flow direction. In contrast to true counter current chromatography, the columns contain a fixed bed and switching is achieved by a system of valves. Given enough columns in series, this approach is an approximation to ideal counter current chromatography. PCC has been studied theoretically for mostly linear adsorption isotherms of small molecules in [136; 177] and experimentally in [28].

In biotechnological practice, three or four column systems are considered, as developed on the basis of an ÄKTA explorer system (GE healtcare, Uppsala, Sweden) [128]. The setup is considered to require simpler hardware and to be easier to operate [128] than other types of continuous chromatography given in [92]. With this setup, recent studies demonstrated a reduction in resin volume and buffer consumption by 40% [122] and that investment costs of a pilot-scale semi-continuous chromatography system would be payed back by the savings obtained by eight proof-of-concept batches [153]. The integration with a perfusion reactor was shown in [201], that ran uninterrupted for 30 days without performance decline. These three studies applied affinity chromatography as capture step, where the target component binds strongly to the adsorbent and yield is naturally high.

When not using affinity chromatography as capture step, the presence of strong binding impurities poses a challenge. The column wash, that is applied to remove weak binding impurities, can lead to significant product loss when aiming for high productivity. Still, ion-exchange capture is often used for enzyme purification, e.g. in [46], and recently even considered for mAb capture [3]. PCC chromatography might be able to solve this problem

Parameter	Symbol	Value
Total column volume	V_c	1 ml
Column length	l_c	20 mm
Bed porosity	ε_c	0.32
Particle porosity	ε_p	0.54
Particle radius	r_p	0.045 mm
Axial dispersion	D_{ax}	0.1 mm/s
Ionic capacity	Λ	0.3 M

Table 12.1: Column parameters

as it captures the flow-through while washing. Both, yield and productivity could be improved with this type of operation.

A three-column PCC system is operated in six steps. Every other step, another column is placed in loading position, i.e. receiving the HCCF feed. In the first of the two loading steps, the column is loaded just before breakthrough occurs. In the second step of loading, the second column can capture the flow through, while the first column is loaded until saturation. In step 3, the former first column is washed, with the flush being captured by the former third column. In this way, no product is lost until collection during elution in step 4. Column one is also regenerated in step 4 and assumes the role of capturing wash and breakthrough of the other columns in step 5 and 6. The cycle continues with all columns taking turns in being loaded, washed/eluted and capturing flow through.

A schematic diagram in Figure 12.1 shows the operation including initial start up with exemplary stationary phase concentrations for a strong (red) and weak binding species (orange) over the column length at the end of each step. The column parameters are given in 12.1 and were taken from a Poros 50 HS cation-exchanger resin in a 1ml column of 20 mm bed height.

The intra-column concentrations were simulated with ChromX for all three columns in parallel.

Further studies should explore, whether design charts similar to [28] can be found for non-affinity chromatography that allow an easy transition from batch-wise to continuous processing and whether displacement in overloaded conditions can be translated into increased selectivity.
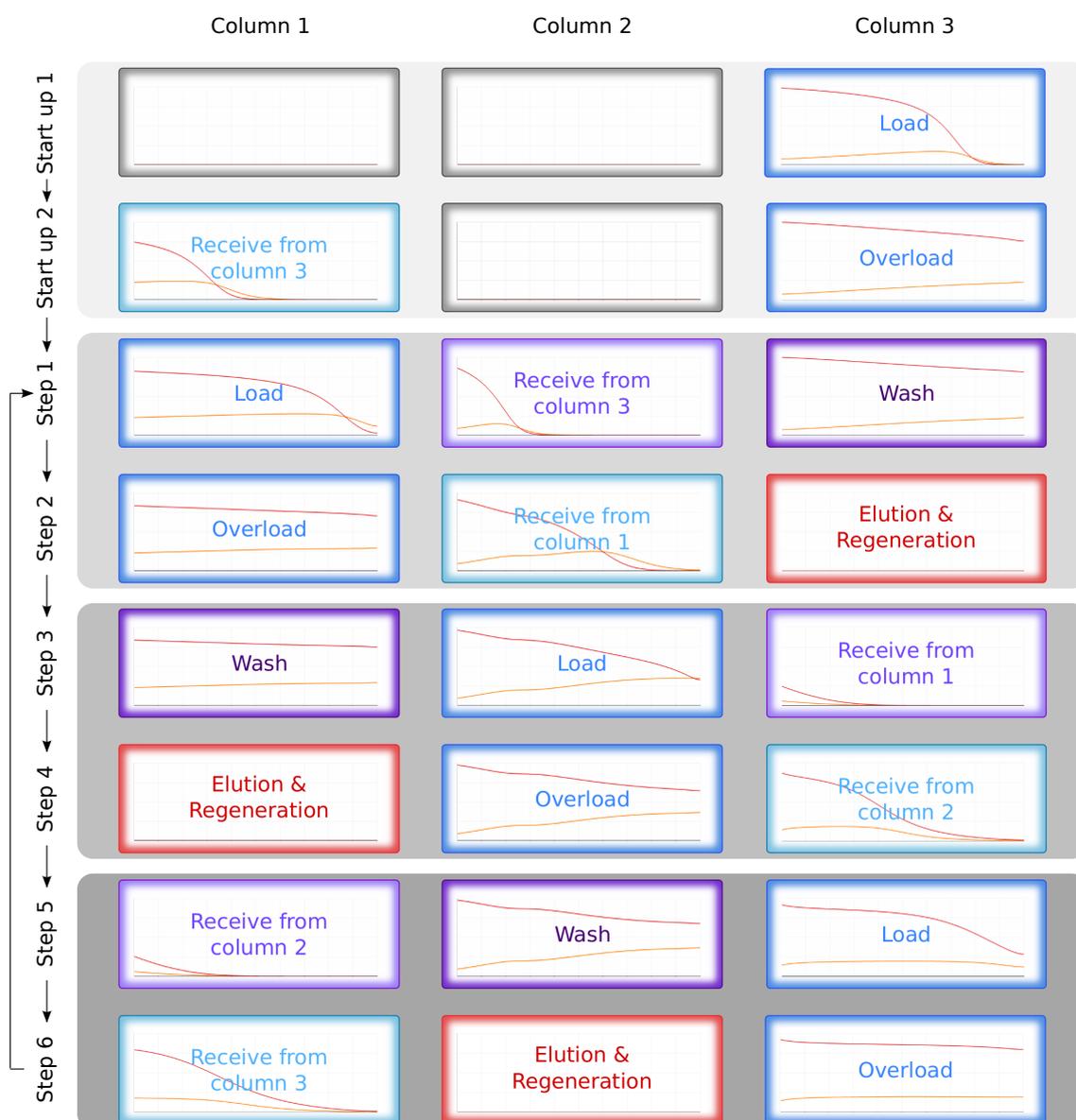
Figure 12.1: Schematic chart of three-column PCC chromatography cycle. After two start up stages, feed is directed to column one, which is loaded until the product breaks though (Step 1). The load is continued with column two capturing the product containing flow through (Step 2). Subsequently column one is washed, eluted and regenerated, while feed is directed to column two (Step 3,4). As washing column one might flush out product, the flow through is captured by column three (Step 3), which thereafter also captures the breakthrough of column 2 (Step 4), before being loaded by the feed stream itself (Step 5,6). The plots show quantitative stationary phase concentrations for two example species over column length, with inlet on the left side.

A | Comparison of MATLAB and ChromX Simulations

A.1 MATLAB Simulation

The following sections describe how to implement the finite element approach given in Sec. 1.1.3 in MATLAB.

A.1.1 Convection

Using the Dirichlet boundary condition first, we have to solve

$$
\begin{aligned}
M\frac{\partial c_i}{\partial t}\left(x,t\right) &= \left(-u \cdot C\right)c_i\left(x,t\right), \\
c_i\left(0,t\right) &= c_{in,i}\left(t\right).
\end{aligned}
$$

Employing MATLAB's solver for ordinary differential equations, we have to set $\frac{\partial c_i}{\partial t}$ in each time step.

At the inlet we prescribe two Gaussian peaks with mean 2 and 3 and standard deviations 0.6 and 0.5. We collect all necessary parameters in a parameter structure 'par':

```
par = struct (...
    'len', 10,...          % Length of column [mm]
    'gridsize',101,...     % Number of nodes in computational grid
    'duration',20,...      % Duration of simulation [s]
    'velocity',  1,...     % Velocity [mm/s]
    'components',2,...     % Number of components
    'exp',[2,3],...        % Mean of Gaussian peak [s]
    'stdev',[0.6,0.5]);    % Standard deviation [s]
```

Choosing the unit of length $[mm]$ and unit of time $[s]$ is arbitrary, the equations could be solved in $[m]$ and $[min]$ the same way, as long as the velocity is specified accordingly in $[length/time]$ and the diffusion coefficient later in $[length^2/time]$.

The following code for the inlet function produces the Gaussian peaks

```matlab
function value = inlet(t,c)
    value = 1/(par.stdev(c)*sqrt(2*pi))*...
    exp(-1/2*((t-par.exp(c))/par.stdev(c)).^2);
end
```

Next, we prepare the tridiagonal matrices C and M on a grid with equidistant nodes. We can use the 'diag' function to write the values directly on the diagonals. We just have to correct the first and last row

```matlab
dim = par.gridsize;                         % Number of grid nodes
h = par.len/(dim-1);                        % Length of each cell
M = h/6*  diag(ones(dim-1,1),1) +...        % Upper diagonal
    h/6*4*diag(ones(dim,1)) +...            % Main diagonal
    h/6*  diag(ones(dim-1,1),-1);           % Lower diagnal
M(dim,dim) = h/3;                           % Correct last entry
M(1,1) =      1;                            % Prepare Dirichlet BC
M(1,2) =      0;
```

Setting the first row to contain only a '1' allows us to set the value according to the Dirichlet Boundary condition.

For C we get similarly

```matlab
C = 1/2*diag(ones(dim-1,1), 1) + ...
      0*diag(ones(dim   ,1))    + ...
   -1/2*diag(ones(dim-1,1),-1);
C(1,1)      = -1/2;
C(dim,dim) =  1/2;
```

Our two species do not interact, we could solve the problems one after another. To prepare the code for additional concentrations, we write the species below each other in a common vector and use 'reshape' to split it up in the model function:

```matlab
% Get zero vector of correct size
c0 = zeros(par.components * dim ,1);
```

```matlab
% Matlab's ODE solver takes the model and solves until duration
[time,conc] = ode15s(@model,[0 par.duration],c0,options);

function dcdt = model(time,conc)                  % Model function
  conc = reshape(conc,[],par.components);         % Comp. k in conc(:,k)
  dcdt = 0*conc;                                  % Set dc/dt to zero
  for j=1:par.components
    dif = inlet(time)-conc(1,j);                  % Diff. to Dirichlet
    conc(1,j) = inlet(time,j);                    % Assume Dirichlet value
    dcdt(:,j) = M\(-par.velocity*C*conc(:,j));    % M*dc/dt = -vCc
    dcdt(1,j) = dif;                              % Set desired diff.
  end
  dcdt = reshape(dcdt,[],1);                       % Reshape column vector
end
```

Here, we first determined the difference between current solution at node 0 and the expected boundary value. In theory, this should be zero, but to be on the safe side we set the boundary value manually in the solution before writing down the equation. The first node of the time derivative $\frac{\partial c}{\partial t}$ is set to the previously determined difference to correct possible deviations.

The result should look as follows, the Gaussian peaks are only transported through the column without changing the shape. The time needed is given by column length divided by flow rate, here 10 s.
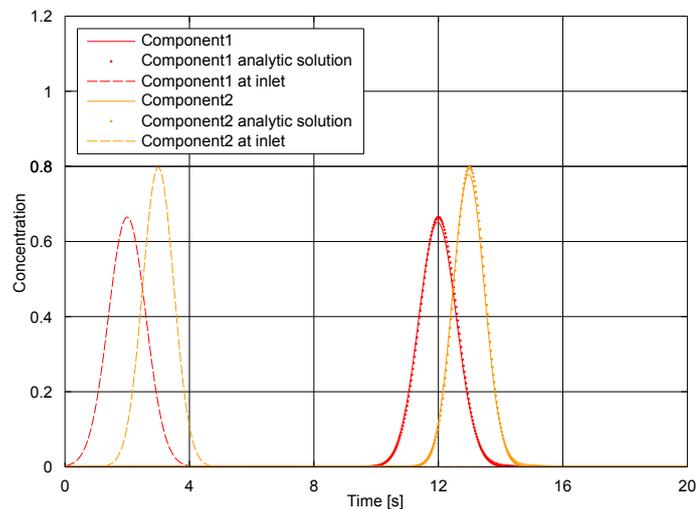
A.1.2   Convection in Radial Geometries

When dealing with radial geometries, e.g. when simulating intra-bead concentrations with the General Rate Model or the flow through membrane capsules, we encounter the convection operator

$$\frac{u}{r}\frac{\partial c}{\partial r}$$

that models the increasing flow rate because of the narrowing diameter. We either have a boundary condition for $r = 0$ or we are dealing with a cylinder ring, such that $r \neq 0$.

Formulating the discrete operator as before, we get

$$C_{rad,k,l} = \int_\Omega \frac{1}{r} \psi_k \frac{\partial \varphi_l}{\partial r} \, \mathrm{d}r \quad \text{on } \Omega = [r_{inner}, r_{outer}] \, .$$

This time, the boundaries of the integral do not vanish, when inserting the basis functions and integrating, thus we cannot write the tridiagonal structure directly. Also, because of r in the denominator, we get logarithmic terms.

Looping over all equidistant nodes and summing up the element contributions can be implemented as follows:

```
for i=1:dim−1
  a = par.innerRad + (i−1)*h;
  b = par.innerRad +     i*h;
  C(i,   i)   = C(i,   i)    +(h−b*log(b/a))/h^2;
  C(i,   i+1) = C(i,   i+1) −(h−b*log(b/a))/h^2;
  C(i+1,i)    = C(i+1,i)     −(h−a*log(b/a))/h^2;
  C(i+1,i+1) = C(i+1,i+1) +(h−a*log(b/a))/h^2;
end
```

We can immediately start simulating the convection through a cylinder ring by assembling the matrix for a given inner and outer radius. Setting a negative flow rate allows to simulate inward flow and, hence, switches the role of inlet and outlet.

A.1.3   Convection-Diffusion

To implement diffusion, we have to assemble the discrete diffusion operator as tridiagonal matrix and introduce a parameter D_{app} for the apparent dispersion:

```
A = −1/h*diag(ones(dim−1,1), 1) + ...
      2/h*diag(ones(dim,1)) + ...
     −1/h*diag(ones(dim−1,1),−1);
A(1,1)       = 1/h;
A(dim,dim) = 1/h;
```

To include it in the model, we simply add the operator.

```
dcdt(:,j) = M\−(par.velocity*C + par.dapp(j)*A)*conc(:,j)
```

A.1.4   Convection-Diffusion with Danckwerts Boundary Conditions

Danckwerts boundary conditions are used to incorporate diffusion in the tube when simulating dispersed plug flow. Changing from Dirichlet to Danckwerts Boundary conditions means, we must not modify the first row of M and evaluate the boundary integral, which only acts on the first node:

```matlab
function dcdt = model(time,conc)              % Model function
  conc = reshape(conc,[],par.components);     % Comp. k in conc(:,k)
  for i=1:par.components
    dcdt(:,i) = (-par.velocity*C + par.dapp(j)*A)* ...
                conc(:,i);                     % Convection, diffusion
    dcdt(1,i) = dcdt(1,i)-par.velocity*(...    % Boundary condition
                conc(1,i)-inlet(time));
    dcdt(:,i) = M\dcdt(:,i)                     % Mass matrix
  end
  dcdt = reshape(dcdt,[],1);                    % Reshape column vector
end
```

A.1.5   Convection-Diffusion with Pulse Injection

To simulate a step and compare the result with ChromX, we introduce the variables *low*, *high* and *event* to parametrize the step function and change the inlet function to

```matlab
function infun = inlet(t,c)
  %INLET Step function
  for l=1:size(t,1)
     if t(l)<par.event
       infun(l) = par.low;
     else
       infun(l) = par.high;
     end
  end
end
```

We also assume, that the system is initialized with the low concentration

```matlab
c0=par.low*ones(par.components * dim ,1);
```

Remark on Stability   Unnatural oscillations may occur depending on the problem and its discretization. Lowering the apparent dispersion coefficient to 1e-4 in the MATLAB step simulation will show this immediately. To smooth out the remaining oscillations for $D_{app} = 0$, the Implicit Euler scheme could be chosen. Implicit schemes are more stable, but computing a time step takes longer compared to explicit schemes.

ChromX applies Streamline-Upwind-Petrov-Galerkin stabilization, which allows to work with a much higher velocity/diffusion ratio. To trigger the same behavior D_{app} has to be set equal to zero.

A.1.6   Lumped Rate Model

The additional equation

$$\varepsilon_p \frac{\partial c_p}{\partial t}(x,t) + (1 - \varepsilon_p)\frac{\partial q}{\partial t}(x,t) = \frac{3}{r_p}k_{eff}\left(c(x,t) - c_p(x,t)\right)$$

is free of space derivatives. Hence, the discrete weak formulation includes only mass Matrices

$$\varepsilon_p M \frac{\partial c_p}{\partial t}(x,t) + (1 - \varepsilon_p) M \frac{\partial q}{\partial t}(x,t) = \frac{3}{r_p}k_{eff} M\left(c(x,t) - c_p(x,t)\right).$$

Event those can be left out completely, as M is invertible (symmetric, positive-definite) we can multiply with M^{-1}. The coupling is then only point-wise.

To extend the example "Convection-Diffusion with Pulse Injection" with a lumped rate model, we have to introduce the new parameters

```
'dax', 0.01,...              % Axial dispersion [mm^2/s]
'epsbed',  0.35,...          % Column porosity
'keff',  [0.015,0.005],...   % Effective film transfer [mm^2/s]
'rad', 0.045,...             % Particle radius [mm]
```

and reserve twice the amount of solution values to include c_p.

```
c0=zeros(2 * par.components * dim ,1);
```

Again, we use the reshape command to split up the solution vector, now with one more dimension.

```
conc  =  reshape(conc,[],par.components,2);  %  c_k  in  conc(:,k,1)
```

The equation for the column concentration is extended with the term for the transition into the pore phase.

```
dcdt(:,j,1)  =  (-par.velocity*C  +  par.dax*A)*conc(:,j,1) -...
                (1-par.epsbed)/par.epsbed*par.keff(j)*3/par.rad *...
                M*(conc(:,j,1)-conc(:,j,2));
```

and the lumped rate model with $\varepsilon_p = 1$ is simply

```
dcdt(:,j,2)  =  par.keff(j)*3/par.rad*(conc(:,j,1)-conc(:,j,2));
```

A.1.7   General Rate

We first apply the chain rule to expand the right-hand side term

$$
\begin{aligned}
\varepsilon_p \frac{\partial c_p}{\partial t}(x,r,t) + (1 - \varepsilon_p)\frac{\partial q}{\partial t}(x,r,t) &= \varepsilon_p \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 D_{pore}\frac{\partial c_p}{\partial r}(x,r,t)\right) \\
&= D_{porei}\varepsilon_p \left(\frac{2}{r}\frac{\partial c_p}{\partial r}(x,r,t) + \frac{\partial^2 c_p}{\partial r^2}(x,r,t)\right)
\end{aligned}
$$

and recover a diffusion operator and the convection operator for radial geometries that we encountered before. Again, we need the discrete variational formulation, where we apply Green's formula to the diffusion term to remove the second order derivative and insert the boundary equations. On the fly, we can divide by ε_p:

$$
\begin{aligned}
M\frac{\partial c_p}{\partial t}(x,r,t) + \frac{(1 - \varepsilon_p)}{\varepsilon_p}M\frac{\partial q}{\partial t}(x,r,t) &= D_{pore}\left(2C_{rad} - A\right)c_p(x,r,t) + \int_{\Gamma} D_{pore}w\frac{\partial c_p}{\partial r}\cdot n\,\mathrm{d}s \\
&= D_{pore}\left(2C_{rad} - A\right)c_p(x,r,t) \\
&\quad + \int_{\Gamma_{in}} w\frac{k_{film}}{\varepsilon_p}\left(c(x,t) - c_p(x,r_p,t)\right)\cdot n\,\mathrm{d}s
\end{aligned}
$$

Apart from the new D_{pore} parameters, we need a radial grid and thus two sets of M, A and C/C_{rad} matrices. The assembly is identical to the axial grid and the code for C_{rad} can

re-used. We only have to be careful with $r = 0$, where the element matrix was undefined. For each component, we now have one axial solution c and for each axial point, a radial grid solution c_p . The initial vector size increases consequently:

```
c0=zeros(par.components * par.axgrid * (1 + par.radgrid),1);
```

Again, we use the reshape command to split up the solution vector, now with dimension incresed by one.

```
conc = reshape(conc,[] , par.components,1 + par.radgrid);
```

We recover $c_k(x)$ in $conc(x, k, 1)$ and $c_{p,k}(x, r)$ in $conc(x, k, 1+r)$. The whole radial grid solution at x is then $conc(x, k, 2 : end)$ in MATLAB notation. The coupling of c and c_p takes place at $r = r_p$, the outermost point of our radial grid. In the column equation, we can use $conc(x, k, end)$ to access the last element, further changes compared to the Lumped Rate Model are not needed.

Programming the General Rate Model for each axial cell follows the same pattern as for the column. We write down the equation for $\frac{\partial c_p}{\partial t}$, set the boundary conditions, and multiply with the inverse mass matrix.

```
for k=1:par.axgrid
    dcdt(k,j,2:end)  =...
        par.dpore(j)*(2*Crad−Arad)*squeeze(conc(k,j,2:end));

    % Boundary condition
    dcdt(k,j,end) = dcdt(k,j,end) + ...
        par.kfilm(j)*(conc(k,j,1)−conc(k,j,end));

    % Mass matrix
    dcdt(k,j,2:end) = Mrad\squeeze(dcdt(k,j,2:end));
end;
```

The squeeze command is necessary to consider the radial grid section as a vector.

A.1.8   Langmuir Model

The additional equation

$$\frac{\partial q}{\partial t}(x,t) = k_{ads} q_{max} \left(1 - \sum_{j=1}^{n} \frac{q_j(x,t)}{q_{max,j}} \right) c_p(x,t) - k_{des} q(x,t)$$

is free of space derivatives, and can be applied point-wise like the Lumped Rate model
(see above).

To extend the Lumped Rate model, we have to introduce the new parameters

```
'epspar',    0.8 ,...          % Particle  porosity
'ads',  [1 ,10] ,...           % Langmuir  adsorption  coefficient
'des',  [1 ,1] ,...            % Langmuir  desorption  coefficient
'qmax',  [0.5 ,0.01] ,...      % Langmuir  maximum  concentration
```

The particle porosity was previously set 1 to ignore adsorption. Now we have to introduce
it. The others are the three component-specific Langmuir parameters.

We have to reserve additional space for the concentration vector q

```
c0=zeros(3 * par.components * dim  ,1);
```

and adjust the reshape commands to incorporate the third phase

```
conc  = reshape(conc ,[] , par.components ,3);  % c_k  in  conc(:,k,1)
```

The equation for the Lumped Rate model is extended with the terms depending on the
particle porosity

```
dcdt (: , j ,2)  = −(1−par. epspar )/ par. epspar*dcdt (: , j ,3)  +  ...
    par. keff ( j )*3/ par. rad / par. epspar *( conc (: , j ,1)−conc (: , j ,2 ));
```

and the Lumped Rate model is implemented as follows

```
psum=ones( par. gridsize ,1);
for  k=1:par.components
    psum=psum−conc (: , k ,3)/ par.qmax(k);
end;
dcdt (: , j ,3)  = par. ads ( j )*par.qmax( j )*psum.* conc (: , j ,2)  −  ...
    par. des ( j )* conc (: , j ,3)
```

We compute the sum $\left(1 - \sum_{j=1}^{n} \frac{q_j(x,t)}{q_{max,j}}\right)$ first and then write down the Langmuir model
equation.

The code is inserted above the Lumped rate model, as Langmuir computes $\frac{\partial q}{\partial t}(x,t)$, what
is needed in the equation of the Lumped Rate model.

A.1.9   Steric Mass Action model

The SMA model introduces one system parameter and four component specific ones

```
'ads ' ,  [ 0 , 1 , 10 ] , . . .         % SMA  adsorption  coefficient
'des ' ,  [ 0 , 1 , 1 ] , . . .          % SMA  desorption  coefficient
'nu ' ,  [ 0 , 2 , 3 ] , . . .           % SMA  characteristic  charge
'sigma ' ,  [ 0 , 10 , 10 ] , . . .      % SMA  steric  shielding  coefficient
'ioncap ' ,  0 . 3 , . . .               % SMA  ionic  capacity  [M]
```

SMA models the counter-ion concentration, that is most often called 'salt' and has no SMA parameters. Here, it is going to be the first component, hence the 0 entries. The ionic capacity of the resin is determined in [M], consequently all feed concentrations have to be given in [M] as well. While we previously did not have to care about initial conditions, SMA assumes an equilibrated column, where all binding sites are covered with counterions, i.e. $q_{salt} = \Lambda$, and an initial salt concentration is already in place $c_{salt}(x,0) = c_{p,salt}(x,0) = c_{in,salt}(0)$.

To set this easily, we can split the initial value at time of definition and set desired values

```
c0=zeros(dim , par . components , 3 ) ;
c0 ( : , 1 , 1 : 2 )=par . cload ( 1 ) ;
c0 ( : , 1 , 3 )=par . ioncap ;
```

The equation for q_{salt} is originally algebraic. To use MATLAB's ODE solver, we differentiate it and obtain

$$\frac{\partial q_{salt}}{\partial t}(x,t) = -\sum_{j=1}^{n} \nu_j \frac{\partial q_j(x,t)}{\partial t}.$$

We have to compute all non-salt components first, to get $\frac{\partial q_j}{\partial t}$ and find $\frac{\partial q_{salt}}{\partial t}$. This can be achieved by iterating backwards over the components

```
for  j=par . components: −1:1
    . . .
end ;
```

To implement the two equations, we distinguish between salt and non-salt components and write down the equations like in the Langmuir case

```
if j~=1 % Non−salt components
    psum=ones(par.gridsize,1)*par.ioncap;
    for k=2:par.components
        psum=psum−(par.nu(j)+par.sigma(j))*conc(:,k,3);
    end
    dcdt(:,j,3) = par.ads(j)*power(psum,par.nu(j)).* conc(:,j,2) −...
                par.des(j)*conc(:,j,3).*power(conc(:,1,2),par.nu(j));
else % Salt
    for k=2:par.components
        dcdt(:,1,3)=dcdt(:,1,3)−par.nu(k)*dcdt(:,k,3);
    end
end;
```

We first compute the sum $\left(\Lambda - \sum_{j=2}^{n} \left(\nu_j + \sigma_j\right) q_j\right)$, where we start from component 2 to leave out the salt concentration. The rest reduces to writing down the equations.

A.2 ChromX Comparison

ChromX was designed for fast and easy in silico downstream process optimization. It supports sequential and continuous multi-column processes, including optimal order and cut times determination. Compared to MATLAB, the time marching with finite differences is an own development including

- Forward/backward Euler scheme,
- Crank-Nicolson,
- Fractional step theta scheme,
- Backward-difference formulas.

which can all be used with adaptive time step determination.

Like in MATLAB, the space discretization is achieved with finite elements. Additionally, linear and quadratic elements are available with or without Streamline-Upwind-Petrov-Galerkin (SUPG) stabilization. This method adds missing diffusion to the problem and suppresses oscillations. Figure A.1 shows a plot of the simulated concentrations of bovine serum albumin (BSA) along the membrane specified in Chap. 10 with and without SUPG.

As non-linear solvers, Newton's method and a fix-point iteration are available. The remaining linear systems are solved with a direct method (UMFPACK) or the iterative GMRES algorithm.

ChromX is based on the hardware-aware linear algebra layer of HiFlow³ [6] and runs on Linux, Windows and OS X.
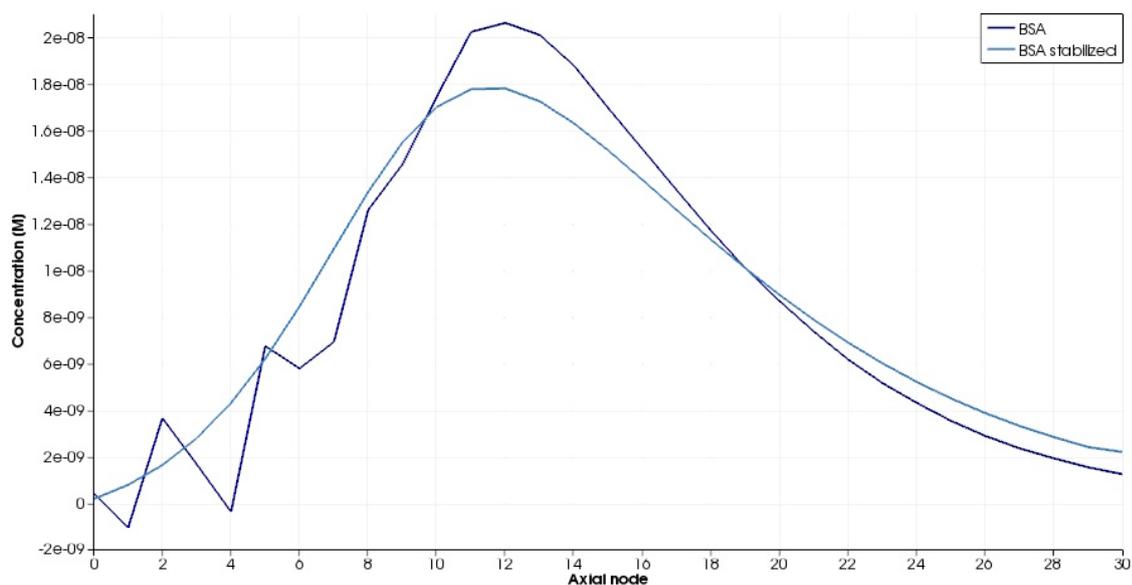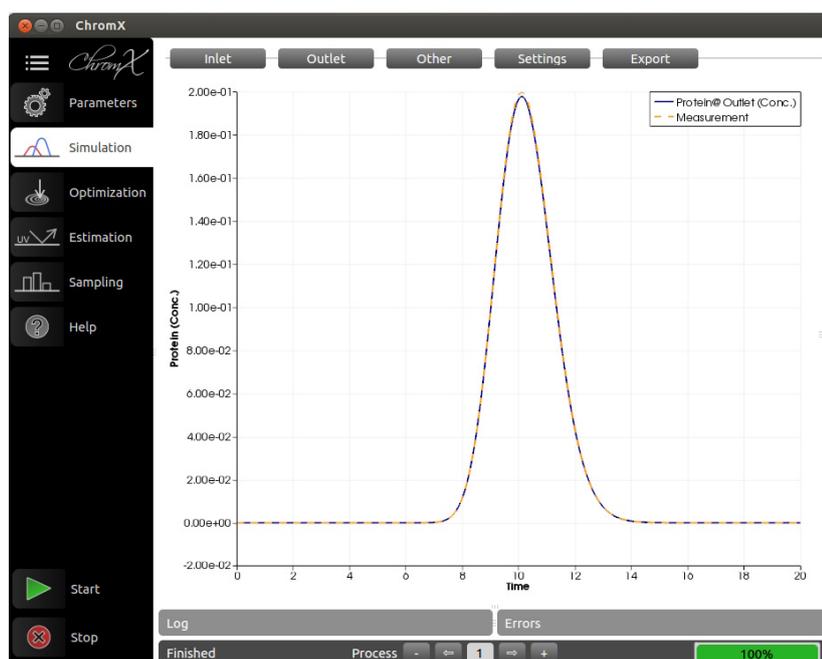
Figure A.1: Comparison of stabilized and unstabilized finite element solutions.

Convection-Diffusion

The MATLAB simulation result was exported as xls spreadsheet file and imported in ChromX as measurement. All experimental parameter settings have been transferred to ChromX. The result shows identical curves:
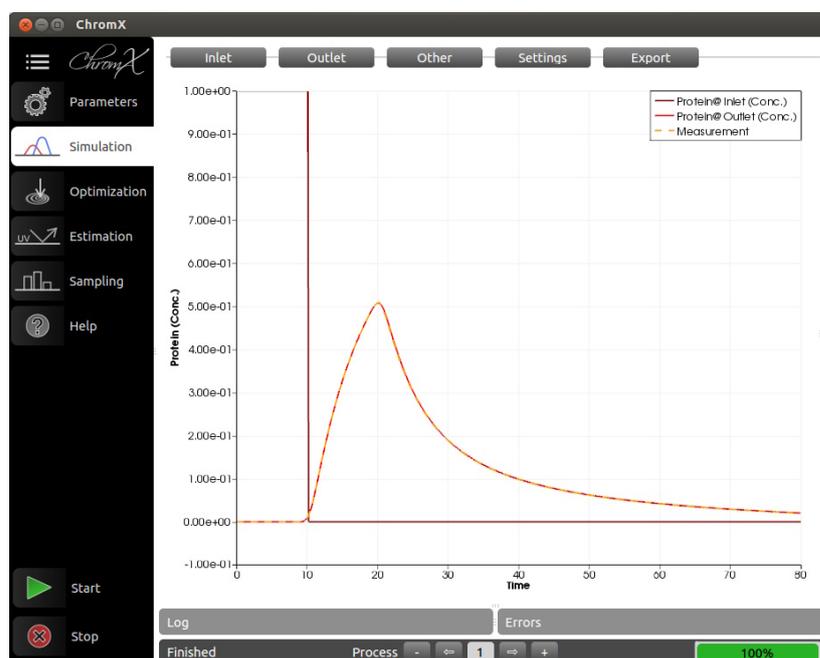
Lumped Rate Model

Again, the MATLAB simulation results were stored and loaded in ChromX. Identical parameter settings result in identical curves.
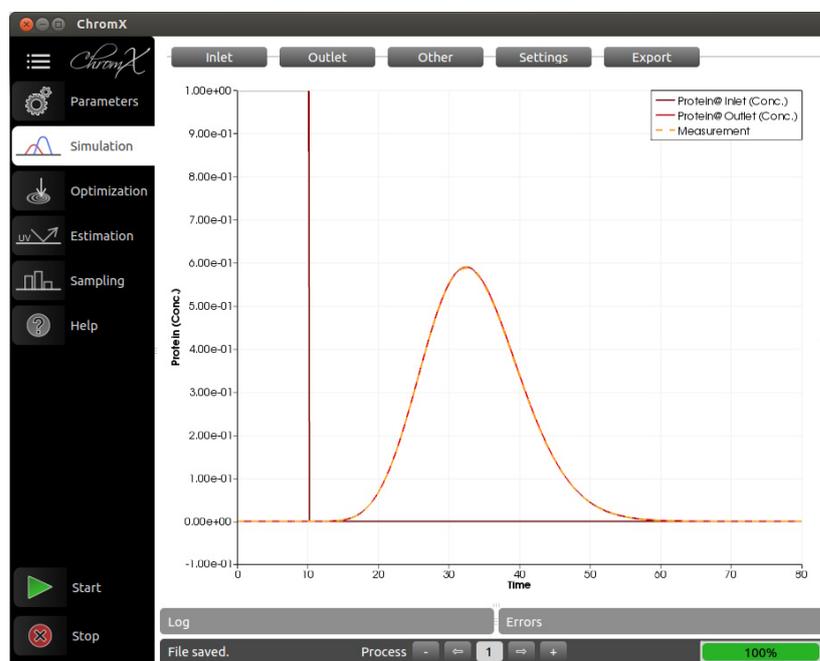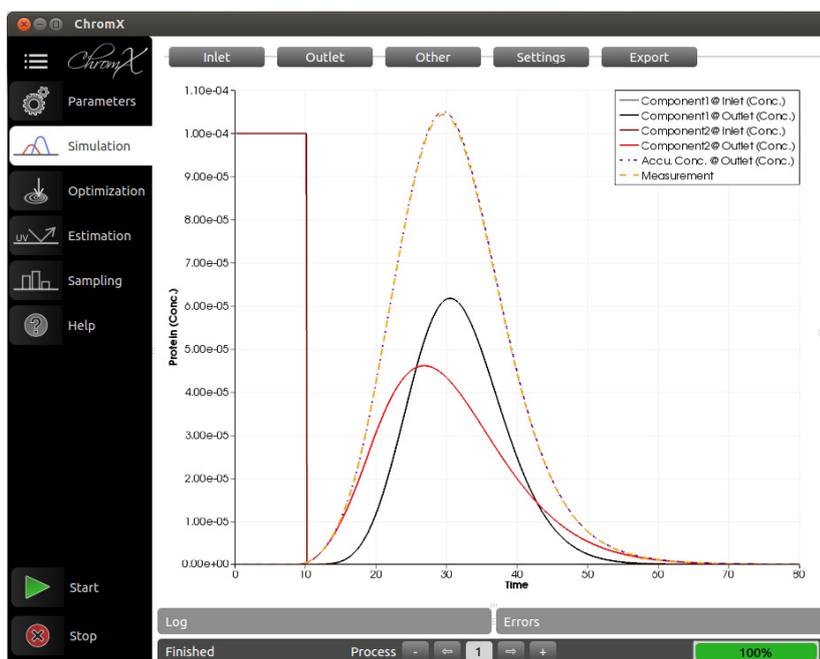

General Rate Model

The General Rate Model with slow pore diffusion generates long drawn-out peaks. The MATLAB scripts and ChromX show identical results:
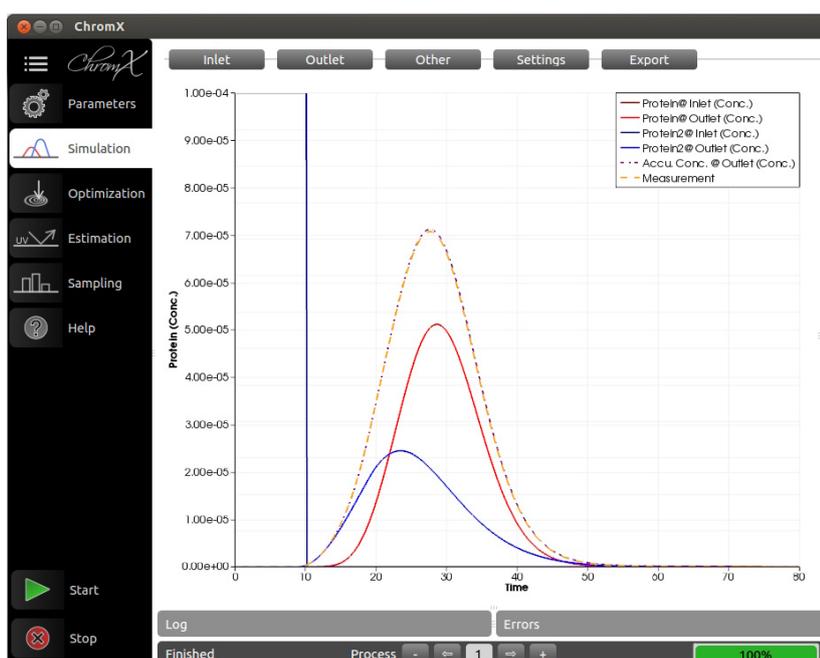
Langmuir Model

Here, the sum of the two-components was exported from MATLAB. The accumulated concentration in ChromX matches these measurements.

Steric Mass Action Model

Also for SMA, the accumulated concentration of the two-components matches the sum of the proteins as exported from MATLAB.

B | Supporting Information

B.1 Supporting Information for: Simulating and Optimizing Preparative Protein Chromatography with ChromX

Tobias Hahn[1], Thiemo Huuk[1], Vincent Heuveline[2], Jürgen Hubbuch[1]

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

[2] Heidelberg University, Interdisciplinary Center for Scientific Computing, Engineering Mathematics and Computing Lab, Heidelberg, Germany

B.1.1 List of Model Parameters

Symbol	Unit Type	Recommended Unit	Description
c	Conc.	M	Mobile phase concentration in interstitial volume
c_p	Conc.	M	Mobile phase concentration of protein in pore volume
$c_{p,s}$	Conc.	M	Mobile phase concentration of salt in pore volume
r_p	Length	mm	Particle radius of adsorbent
D_{app}	$\frac{\text{Length}^2}{\text{Time}}$	$\frac{\text{mm}^2}{\text{s}}$	Apparent dispersion coefficient
D_{ax}	$\frac{\text{Length}^2}{\text{Time}}$	$\frac{\text{mm}^2}{\text{s}}$	Axial dispersion coefficient
D_{pore}	$\frac{\text{Length}^2}{\text{Time}}$	$\frac{\text{mm}^2}{\text{s}}$	Pore diffusion coefficient
ε_{Col}	-	-	Column/bed porosity
ε_{Bead}	-	-	Stationary phase porosity
ε_{Tot}	-	-	Total porosity
k_{eff}	$\frac{\text{Length}}{\text{Time}}$	$\frac{\text{mm}}{\text{s}}$	Effective film/pore transfer coefficient
k_{eq}	-	-	Adsorption equilibrium coefficient
$k_{eq,L}$	Conc.$^{-1}$	M^{-1}	Adsorption equilibrium coeff. for Langmuir isotherm
k_{kin}	Time·Conc.$^\nu$	sM$^\nu$	Adsorption rate coefficient
L_{Col}	Length	mm	Column length

Λ	Conc.	M	Stationary phase ionic capacity
ν	-	-	Characteristic charge for SMA isotherm
q	Conc.	M	Stationary phase concentration of protein
q_{max}	Conc.	M	Single-component max. conc. for Langmuir isotherm
q_s	Conc.	M	Stationary phase concentration of salt
σ	-	-	Steric shielding coefficient for SMA isotherm
t	Time	s	Time dimension
u_{int}	$\frac{\text{Length}}{\text{Time}}$	$\frac{\text{mm}}{\text{s}}$	Interstitial mobile phase velocity
x	Length	mm	Space dimension

B.1.2   Experimental Parameter Determination

In the laboratory course, initial experiments were conducted to determine system properties. Afterwards, four experiments in the bind/elution mode were performed for parameter estimation, i.e. one step elution and two gradient elutions with a low sample volume and one gradient elution with a large sample volume.

First, the necessary system and column parameters were determined and verified by checking the agreement of the simulated salt elution profile and conductivity signal. Then, the isotherm parameters were determined by chromatogram fitting.

B.1.2.1   Dead Volumes

The system's dead volumes from the auto-sampler to the UV detector and to the conductivity detector were determined with an acetone pulse injection without column to be $V_{dead,cond} = 140\,\mu L$ and $V_{dead,UV} = 122\,\mu L$.

B.1.2.2   Column and Bead Dimensions

In this case study, a pre-packed SP Sepharose FF column (GE Healthcare, $L_{Col} = 25\,mm$, $V_{Col} = 0.962\,mL$) was used. The radius of the adsorber beads is $0.045\,mm$ according to the manufacturer.

B.1.2.3   Linear Flow Rate

The pump flow was set to $u_{Vol} = 0.962\,mL/min$. In ChromX, we have to specify the flow in distance/time. We use the units mm and s below, as they fit the process scale best. The linear flow rate can be calculated easily as shown below.

$$u\left[\frac{mm}{s}\right] = \frac{L_{col}}{V_{col}} \cdot \frac{u_{vol}\left[\frac{mL}{min}\right]}{60} = \frac{25}{0.962} \cdot \frac{0.962}{60}\frac{mm}{s} = 0.4167\frac{mm}{s} \qquad \text{(B.1)}$$

B.1.2.4   Porosities

$20\,\mu L$ pulse injections of 1 M NaCl (pore-penetrating) and 10 g/L dextran (2000 kDA, non-pore-penetrating) at the same flow rate were used to determine the porosities. The measured retention volumes were $V_{NaCl,rt} = 0.98\,mL$ and $V_{Dex,rt} = 0.38\,mL$. We first subtract the respective system dead volumes

- $V_{NaCl} = V_{NaCl,rt} - V_{dead,cond} = 0.84\,mL$,

- $V_{Dex} = V_{Dex,rt} - V_{dead,UV} = 0.258\,mL$,

and then calculate the porosities

- Total porosity: $\varepsilon_{tot} = \frac{V_{NaCl}}{V_{col}} = \frac{0.84}{0.962} = 0.873$,

- Column porosity: $\varepsilon_{col} = \frac{V_{Dex}}{V_{col}} = \frac{0.258}{0.962} = 0.268$,

- Bead porosity: $\varepsilon_{bead} = \frac{V_{NaCl}-V_{Dex}}{V_{col}-V_{Dex}} = 0.827$.

B.1.2.5   Axial Dispersion

The axial dispersion coefficient D_{ax} can be derived from the broadening of the dextran pulse. Moment analysis for the injection of an ideal Dirac pulse of a non-pore-penetrating, non-interacting tracer yields the following Eq. (B.2) for the parameter σ_{Dex} of the resulting Gaussian peak that can be solved for D_{ax}.[188]

$$\sigma_{Dex}^2 = 2 D_{ax} L_{Col} \left(\frac{\varepsilon_{Col}}{u}\right)^3. \tag{B.2}$$

Typically, the control software of the chromatography system includes peak analysis tools that calculate this value automatically or the height equivalent of a theoretical plate (HETP), which is defined as the rate of increase of the Gaussian peak profile per unit length, and can be written as in Eq. (B.3).[129; 188]

$$HETP = \frac{\sigma_{Dex}^2}{L_{Col}} \left(\frac{u}{\varepsilon_{Col}}\right)^2 = 2 D_{ax} \left(\frac{\varepsilon_{Col}}{u}\right). \tag{B.3}$$

Here, we obtained the value $HETP = 0.4798\,mm$. Using the linear flow rate and the column porosity from above, we can now calculate the axial dispersion coefficient

$$D_{ax} \quad = \quad HETP \cdot \frac{u}{2 \cdot \varepsilon_{col}} = 0.373\,mm^2/s. \tag{B.4}$$

Table B.2: Bind/elute experiments.

Elution Mode	Sample Volume (mL)	Gradient/Step Height (% Buffer B)	Length of Elution (CV)
Step	0.5	15	10
Gradient	0.5	50	10
Gradient	0.5	100	10
Gradient	60	50	10

B.1.2.6   Ionic Capacity

The total ionic capacity Λ of the packed resin was determined by acid-base titration.[85] The column is flushed with 0.5 M HCl, such that all ligands are saturated with H+ ions and then washed with ultrapure water. We then inject 0.1 M NaOH ($= c_{NaOH}$) solution to replace H+ by Na+ ions. The conductivity signal starts to increase at $V_{NaOH} = 3.27\,mL$. Consequently, $c_{NaOH} \times V_{NaOH} = 0.327\,mol$ are exchanged. To obtain the capacity of the stationary phase, we divide by its volume, expressed by total column volume and total porosity:

$$\Lambda \quad = \quad \frac{V_{NaOH} \cdot c_{NaOH}}{V_{col}\left(1 - \varepsilon_{tot}\right)} = 2.677\,M. \tag{B.5}$$

B.1.2.7   Bind/Elute Experiments

Three experiments with a low sample volume ($0.5\,mL$) and one with a breakthrough ($60\,mL$) were conducted with the mixture of antibody, lysozyme, and myoglobin. The used buffers were Buffer A ($20\,mM$ Bicine, $0\,mM$ NaCl, pH 8.2) and Buffer B ($20\,mM$ Bicine, $1000\,mM$ NaCl, pH 8.2). Elution was initiated 5.2 mL after the end of injection with the settings given in Table B.2, followed by washing with 100 % Buffer B after $9.5\,mL \approx 10$ column volumes (CV). As salt was not injected via the auto-sampler, but via a mixing chamber, an additional dead volume of $1.35\,mL$ had to be added to the event in ChromX. In case of steps, the additional dead volume is only $1.10\,mL$, if a pump wash was performed prior to the step.

The results were exported from the control software as XLS files, including volume, UV 280 nm, and conductivity data columns.

B.1.2.8   Component-specific Parameters

The column parameters were first checked by comparing the simulated salt elution profile with the recorded conductivity signal. Film transfer and pore diffusion parameters for the salt component were estimated.

The proteins' SMA parameters were determined by chromatogram fitting. First, the parameters that are active in the linear range of the isotherm were estimated from the

experiments with a low sample volume. Two gradient experiments are sufficient to de-
termine the characteristic charge and equilibrium parameter by their effect on retention
time.[152; 169] The kinetic parameter is responsible for additional peak broadening and
can also be determined from low-sample-volume experiments.[100] A step elution experi-
ment was included as well, as film transfer and pore diffusion parameters have a stronger
influence on the peak shape in this mode. The steric shielding parameter cannot be es-
timated from experiments with a low sample volume, as it occurs only in the sum of
Eq. (1.9) which is then close to zero. It was estimated from the experiment with a high
sample volume, while keeping the other parameters constant.

B.1.3 Master's Degree Program and Software Exercises

The lecture series on chromatography modeling is well-established in the curriculum of
the bioengineering master's degree program at KIT. It shifts the focus from finding a
workable solution in the lab to understanding the effects that lead to a certain peak shape
in the chromatogram. At first, the partial differential equations seem to have several
parameters, but in the course of the lecture series, the influences of void volumes and
diffusion effects become more obvious. Relating isotherm parameters to elution peak
shapes provides for a connection to practical laboratory experience. The software exercises
accompany the lecture series on chromatography modeling and are intended to provide
a deeper understanding of the mathematical operators and parameters in the differential
equations.

Surveys on the lecture series were conducted by the Executive Support Department of KIT,
Section III: Quality Management. In 2014, seven students responded to the survey on the
last day of class. From the 28 questions, one of the most important aspects in this context
was that the students recognized the importance of the lecture for further study (4.86/5,
standard deviation (SD) 0.38, 0=very low, 5=very high). They felt that difficult issues
could be presented, liked the practical examples, and were encouraged to work on their
own outside class (each 4.71/5, SD 0.49). One student commented that he/she particularly
liked the ChromX exercises. In comparison, 14 students performed the software exercises
with the Chromulator in 2011. They rated the practical examples and encouragement
to learn on their own with 3.93/5 (SD 0.92), and 4.43/5 (SD 0.65). Larger comparative
studies would be necessary to find the most effective way to use ChromX in the classroom.

In the following, we present exemplary exercises performed with the Langmuir isotherm.

B.1.3.1 Single-component Ideal Model

Analyze and simulate the following system with a single-component equilibrium Langmuir
isotherm:

- Column length = 25 mm,
- Column volume = 0.962 mL,
- Total porosity = 0.82,
- Flow rate = 1 mm/s,
- $D_{app} = 0.01$ mm^2/s,

- $k_{eq,L} = 1$ M^{-1},
- $q_{max} = 20$ M,
- $c(x, 0) = 0$ M,
- $c(0, 0\,s \leq t < 5\,s) = 0.001$ M,
- $c(0, t \geq 5\,s) = 0$ M.

Start a new ChromX session and set the model to *EquilibriumDispersive*, *NoPoreModel*, and *Langmuir*. For a better resolution, increase the *Axial Cells* to 200 and reduce the *Initial Step* to 0.1. Copy the column parameters from above and remove the Salt component. Set the Langmuir parameters as above and kinetics to zero to obtain the equilibrium model. Adjust the injection start concentration and time of the end of injection.

1. Do your observations agree with the theoretical retention time for an ideal model according to Eq. (1.32)?

2. Set the injection end to 200 s. Does the time of breakthrough agree with the retention time of a shock, as in Eq. (1.35)?

3. What is the sample concentration that generates a breakthrough at 80.5 s? Validate your result by simulation.

4. Reduce the injection end to 30 s. Describe the result.

5. Estimate when the rear of the peak will reach the base line. Explain deviations.

6. What happens when the capacity doubles and q_{max} increases to 40? Explain.

B.1.3.2   Competitve Adsorption and Displacement

Keep the simulation and column setup of the previous section and simulate two single-component experiments and one with both components:

- Component 1: $k_{eq,L} = 1$ M^{-1}, $q_{max} = 20$ M,

- Component 2: $k_{eq,L} = 3$ M^{-1}, $q_{max} = 10$ M,

Both components shall be injected for 20 s with a sample concentration of 0.05 M.

7. Compare the peaks. Explain the result.

8. Increase the injection time to 200 s to simulate a two-component breakthrough experiment. Explain the behavior of the first component.

9. What happens, when adding a third component with $k_{eq,L} = 5$ M^{-1}, $q_{max} = 10$ M?

B.1.3.3   Proposed Solutions

1. The equation correctly predicts the peak maximum at 113 s (Fig. B.1a).

2. Yes, the inflection point of the breakthrough at 0.0005 M is exactly at 110.5 s (Fig. B.1a).

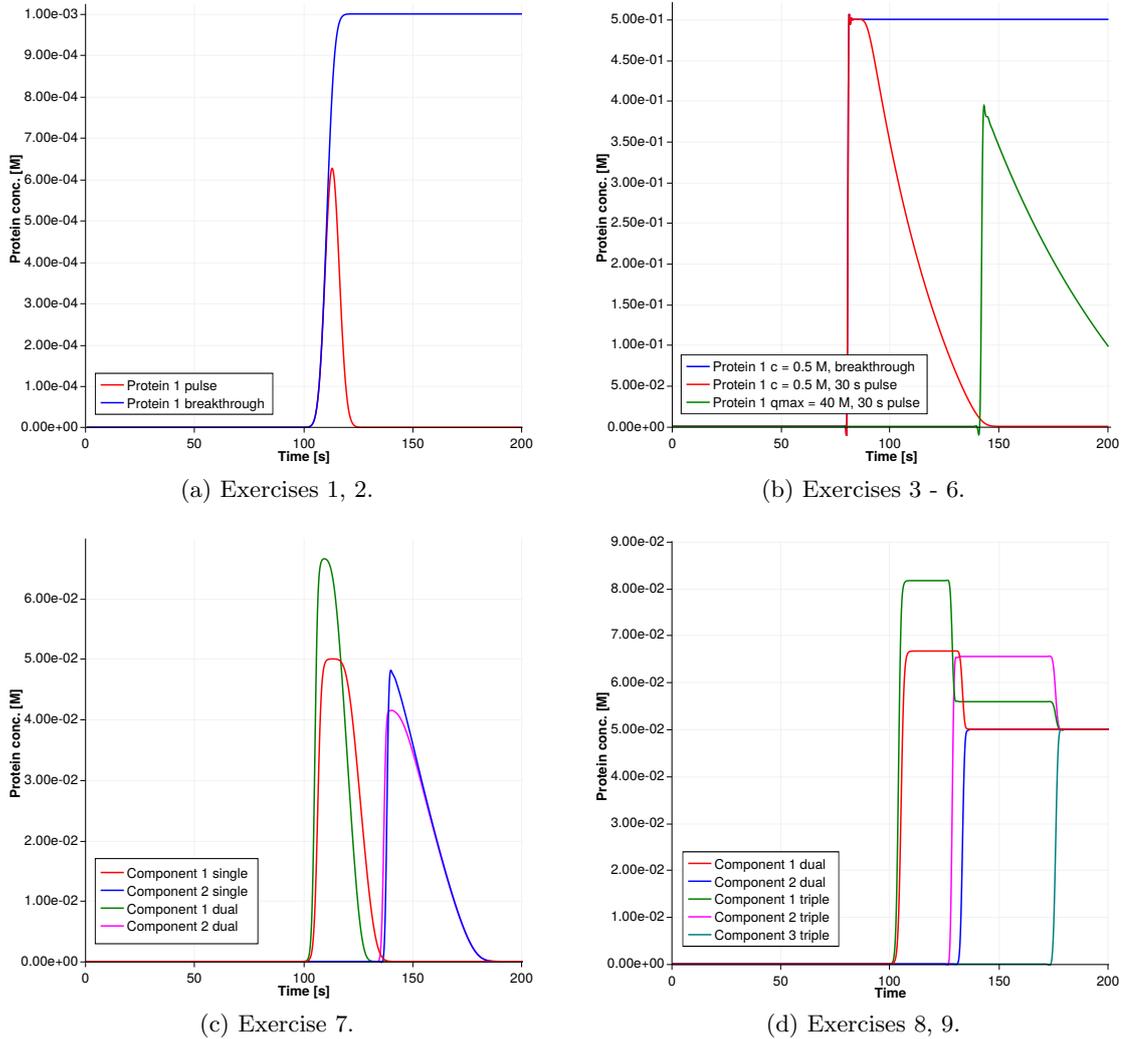3. The concentration is 0.5 M according to Eq. (1.35) (Fig. B.1b).

(a) Exercises 1, 2.

(b) Exercises 3 - 6.

(c) Exercise 7.

(d) Exercises 8, 9.

Figure B.1: Plots of simulation results

4. The shock front stays at 80.5 s. We observe a short plateau and a diffuse rear. The peak ends at 145 s (Fig. B.1b).

5. The rear of the pulse is to migrate according to Eq. (1.32) (with t_{inj} instead of $t_{inj}/2$) and reach the base line at 140.5 s. Because of the non-zero dispersion, additional broadening occurs. The value of 0.01 mm^2/s leads to an approximate broadening of 0.05 mm/s in each axial direction. The contact time of $140.5\,s - 30\,s = 110.5\,s$ leads to a total broadening of 5.525 mm. At an interstitial flow rate of $u/\varepsilon_{Tot} = 1/0.82$ mm/s, this equals 4.5 s.

6. Theoretically, the shock retention time is 110 s and 215.5 s for a symmetrical peak. We observe a shock at 143 s with a height of 0.4 M. Obviously, the intra-column concentrations were not sufficient to develop the whole shock (Fig. B.1b).

7. In preparative chromatography, the system response to multi-component feedstocks is not just a superposition of the single-component breakthroughs or peaks. Component 1 competes with Component 2 for binding sites, while having the smaller $k_{eq,L}$ value. It does not adsorb as much as in the single-component case and, thus,

migrates faster through the column, resulting in a slightly earlier and higher peak. Component 2 also moves slightly faster, as it cannot bind with the same amount in the beginning. Again, the concentration migrates faster. As Component 2 follows the even faster moving Component 1, binding sites are constantly freed at the front of the band of Component 2, leading to a less strong shock build-up and a smoother top (Fig. B.1c).

8. Because of its smaller $k_{eq,L}$, Component 1 adsorbs more slowly and the concentration front migrates faster. The following Component 2 partly displaces the first component and the desorbed concentration accumulates in a faster-moving plateau. Because of the nonlinear adsorption behavior described by the Langmuir isotherm, the additional concentration cannot fully re-adsorb.

9. The effect is increased in the three-component setting. Component 1 is displaced even more strongly and also Component 2 shows this behavior. Component 3 having the largest $k_{eq,L}$, follows the others and adsorbs in the three-component equilibrium state.

B.2   Supporting Information for: Deconvolution of High-throughput Multi-component Isotherms Using Multivariate Data Analysis of Protein Spectra

Pascal Baumann[1,*], Tobias Hahn[1,*], Thiemo Huuk[1,*], Anna Osberghaus[1], Jürgen Hubbuch[1]

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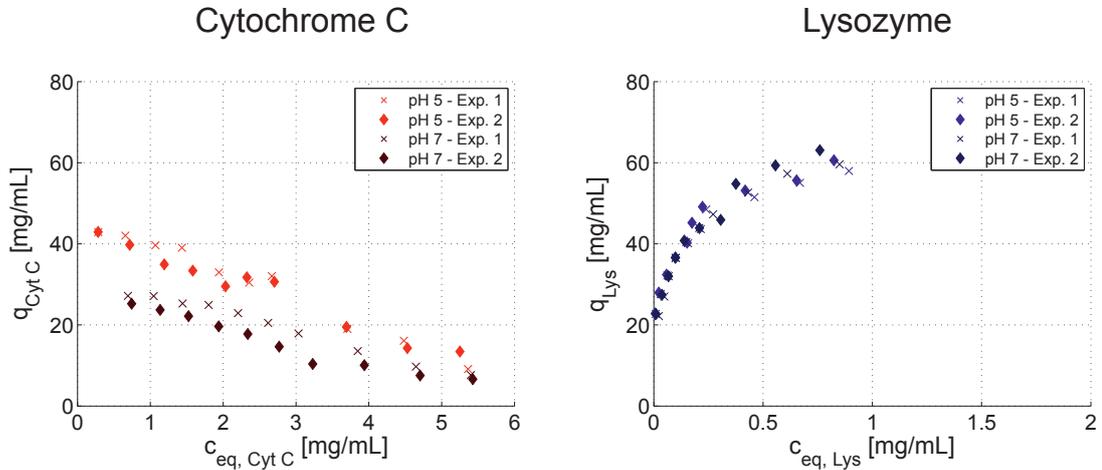
[*] Contributed equally to this work

Figure B.2: Comparison of isotherm data for determination of the experimental robustness and model performance. Duplicates of the isotherm experiments performed at 90 mM ionic strength for pH 5 and pH 7. The duplicates are indicated as diamonds and crosses, respectively. The results for cytochrome c are shown in red (left) and for lysozyme in blue (right).
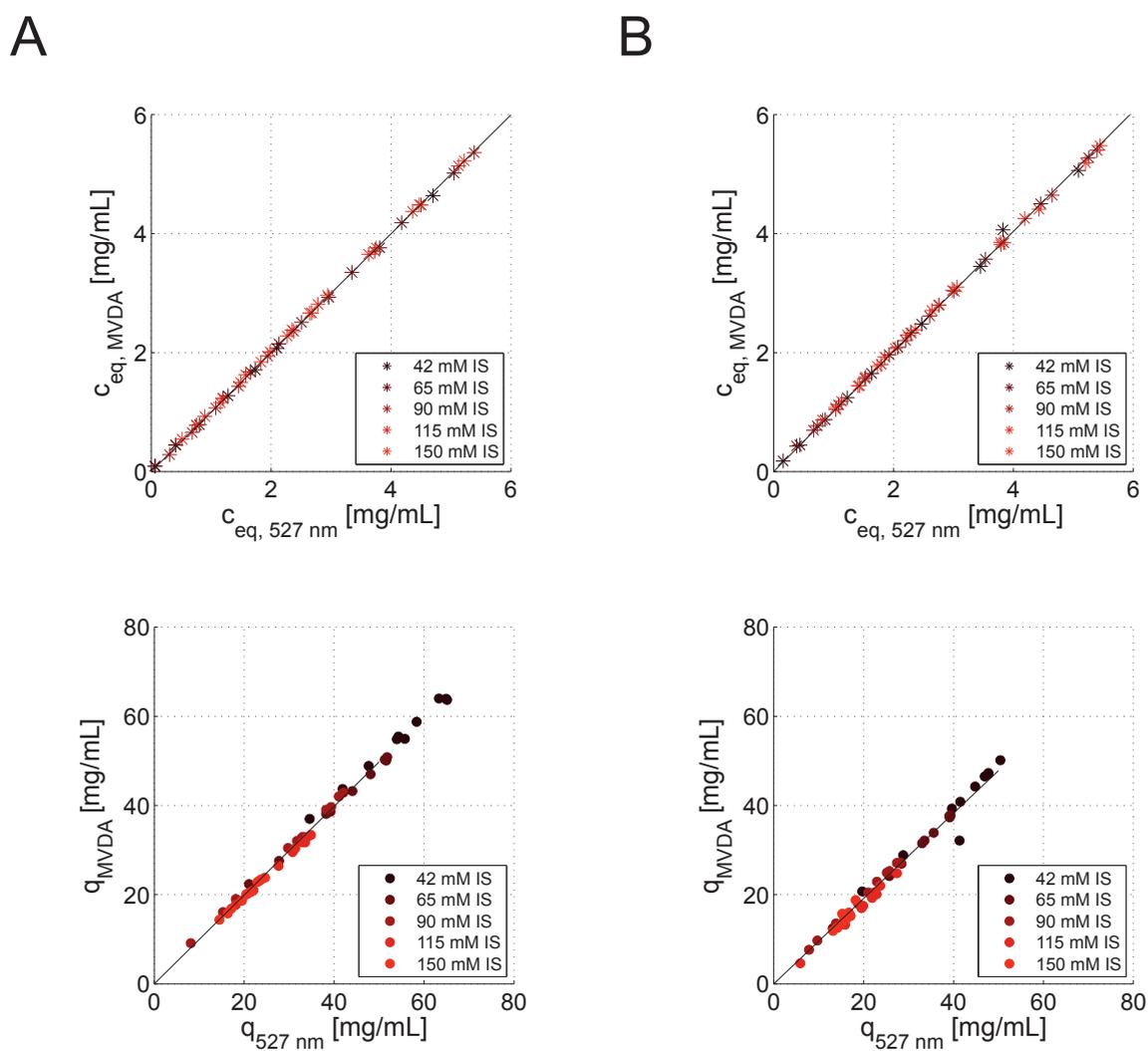
Figure B.3: Comparison of cytochrome c isotherm data points at pH 5 (A) and pH 7 (B) derived from the MVDA model and the selective 527 nm wavelength as a secondary analytics for validation. The agreement of the different data point is shown in parity plots for the equilibrium concentration of cytochrome in solution and bound to the adsorbent.

B.3 Supporting Information for: Modeling and Simulation of Anion-exchange Membrane Chromatography for the Purification of *Sf*9 Insect Cell-derived Virus-like Particles

Christopher Ladd Effio[1], Tobias Hahn[1], Julia Seiler[1], Stefan Oelmeier[1,2], Iris Asen[3], Christine Silberer[3], Louis Villain[4], Jürgen Hubbuch[1]

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

[2] Boehringer Ingelheim Pharma GmbH & Co. KG, Biberach, Germany

[3] Diarect AG, Freiburg, Germany

[4] Sartorius Stedim Biotech, Göttingen, Germany

Table B.3 and Figure B.4 are provided as supplementary data to section 10.3.2 and 10.3.5 of the above stated article. Table B.3 summarizes UV peak areas and estimated SMA parameters of VLP feedstock subcomponents (*Spodoptera frugiperda Sf*9 insect cell lysate) for a Sartobind® Q membrane adsorber. Fig. B.4 shows the UV chromatogram of the SEC polishing procedure for human B19 parvo-VLPs subsequent to an anion-exchange membrane chromatography step.

Table B.3: UV peak areas and converged SMA parameters for a *Spodoptera frugiperda Sf*9 insect cell lysate containing human B19 parvovirus-like particles

| Component	$Area_{280nm}[mAU*mL]$	k_{eff}	k_{eq}	ν
Contaminant 1	527	0.430	0.300	0.790
Contaminant 2	202	0.662	0.161	2.112
Contaminant 3	606	0.390	0.647	2.245
Contaminant 4	687	0.652	1.827	2.667
Contaminant 5	1260	0.202	4.999	2.742
Contaminant 6	695	0.150	21	7.817
Contaminant 7	339	0.980	90	7.955
Contaminant 8	343	0.820	318	9.000
Target	1111	0.162	6.21E+13	34.320
Contaminant 9	426	0.216	2.65E+04	12.571
Contaminant 10	427	0.161	8.00E+03	8.600
Contaminant 11	722	0.337	4.80E+06	13.548
Contaminant 12	396	0.990	2.80E+07	13,690
Contaminant 13	40	0.280	1.20E+10	18.548
Contaminant 14	228	0.990	2.23E+15	28.260
Contaminant 15	328	0.199	5.43E+15	28.050
Contaminant 16	655	0.140	1.35E+05	7.060
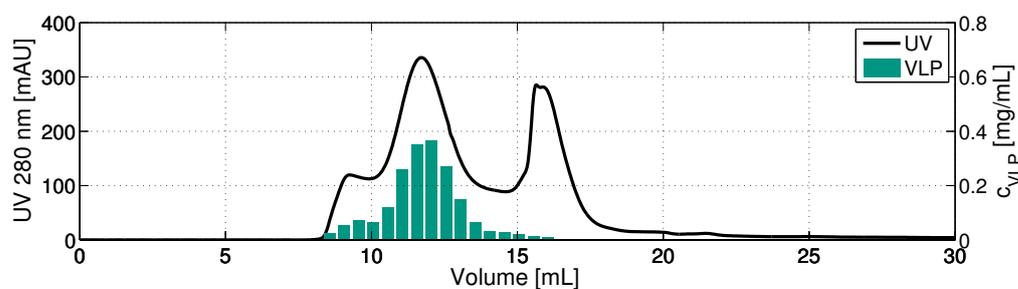
Figure B.4: UV chromatogram of a size-exclusion chromatography run with human B19 parvovirus-like particles captured from *Spotoptera frugiperda Sf* 9 insect cells by anion-exchange membrane chromatography. 500 μL of a 2 mg/mL VP2-VLP solution were injected on a Superose® 6 Increase 10/300 column at a flow rate of 0.5 mL/min. PBS pH 7.4 was used as mobile phase buffer. The solid black line shows the UV absorption at 280 nm while the green bars represent the VLP concentration in the collected fractions.

Bibliography

1. ABEL, S.; ERDEM, G.; AMANULLAH, M.; MORARI, M.; MAZZOTTI, M.; AND MORBIDELLI, M., 2005. Optimizing control of simulated moving beds–experimental implementation. *Journal of chromatography. A*, 1092, 1 (2005), 2–16. doi:10.1016/j. chroma.2005.04.101.

2. AGNANDJI, S. T.; LELL, B.; FERNANDES, J. F.; ABOSSOLO, B. P.; KABWENDE, A. L.; ADEGNIKA, A. A.; MORDMÜLLER, B.; ISSIFOU, S.; KREMSNER, P. G.; LOEMBE, M. M.; SACARLAL, J.; AIDE, P.; MADRID, L.; LANASPA, M.; MAND-JATE, S.; APONTE, J. J.; BULO, H.; NHAMA, A.; MACETE, E.; ALONSO, P.; ABDULLA, S.; SALIM, N.; MTORO, A. T.; MUTANI, P.; TANNER, M.; MAVERE, C.; MWANGOKA, G.; LWENO, O.; JUMA, O. A.; SHEKALAGHE, S.; TINTO, H.; D'ALESSANDRO, U.; SORGHO, H.; VALEA, I.; OUÉDRAOGO, J. B.; LOMPO, P.; DIALLO, S.; TRAORE, O.; BASSOLE, A.; DAO, E.; HAMEL, M. J.; KARIUKI, S.; ONEKO, M.; ODERO, C.; OTIENO, K.; AWINO, N.; MUTURI-KIOI, V.; OMOTO, J.; LASERSON, K. F.; SLUTSKER, L.; OTIENO, W.; OTIENO, L.; OTSYULA, N.; GONDI, S.; OTIENO, A.; OGUTU, B.; OCHOLA, J.; ONYANGO, I.; OYIEKO, J.; NJUGUNA, P.; CHILENGI, R.; AKOO, P.; KERUBO, C.; MAINGI, C.; OLOTU, A.; BEJON, P.; MARSH, K.; MWABINGU, G.; GITAKA, J.; OWUSU-AGYEI, S.; AS-ANTE, K. P.; BOAHEN, O.; DOSOO, D.; ADJEI, G.; ADENIJI, E.; YAWSON, A. K.; KAYAN, K.; CHANDRAMOHAN, D.; GREENWOOD, B.; LUSINGU, J.; GESASE, S.; MALABEJA, A.; ABDUL, O.; MAHENDE, C.; LIHELUKA, E.; LEMNGE, M.; THE-ANDER, T. G.; DRAKELEY, C.; MBWANA, J.; ANSONG, D.; AGBENYEGA, T.; ADJEI, S.; BOATENG, H. O.; RETTIG, T.; BAWA, J.; SYLVERKEN, J.; SAM-BIAN, D.; SARFO, A.; AGYEKUM, A.; MARTINSON, F.; HOFFMAN, I.; MVALO, T.; KAMTHUNZI, P.; NKOMO, R.; TEMBO, T.; TSIDYA, G. T. M.; KILEMBE, J.; CHAWINGA, C.; BALLOU, W. R.; COHEN, J.; GUERRA, Y.; JONGERT, E.; LAPIERRE, D.; LEACH, A.; LIEVENS, M.; OFORI-ANYINAM, O.; OLIVIER, A.; VEKEMANS, J.; KASLOW, D.; LEBOULLEUX, D.; SAVARESE, B.; AND SCHELLEN-BERG, D., 2014. Efficacy and Safety of the RTS,S/AS01 Malaria Vaccine during 18 Months after Vaccination: A Phase 3 Randomized, Controlled Trial in Chil-dren and Young Infants at 11 African Sites. *PLoS Medicine*, 11, 7 (2014). doi: 10.1371/journal.pmed.1001685.

3. AHAMED, T.; CHILAMKURTHI, S.; NFOR, B. K.; VERHAERT, P. D. E. M.; VAN DEDEM, G. W. K.; VAN DER WIELEN, L. A. M.; EPPINK, M. H. M.; VAN DE SANDT, E. J. A. X.; AND OTTENS, M., 2008. Selection of pH-related parameters in ion-exchange chromatography using pH-gradient operations. *Journal of Chromatog-raphy A*, 1194, 1 (2008), 22–29. doi:10.1016/j.chroma.2007.11.111.

4. AHRENS, J.; GEVECI, B.; AND LAW, C., 2005. 36 - ParaView: An End-User Tool for Large-Data Visualization. In *Visualization Handbook* (Ed. C. D. H. R. JOHNSON), 717 – LXXII. Butterworth-Heinemann, Burlington. ISBN 978-0-12-387582-2. doi: 10.1016/B978-012387582-2/50038-1.

5. ALTENBACH-REHM, J.; NELL, C.; ARNOLD, M.; AND WEUSTER-BOTZ, D., 1999. Parallel Bubble Columns with Fed-Batch Technique for Microbial Process Development on a Small Scale. *Chemical Engineering & Technology*, 22, 12 (Dec. 1999), 1051–1058. doi:10.1002/(SICI)1521-4125(199912)22:12<1051::AID-CEAT1051>3.0. CO;2-C.

6. ANZT, H.; AUGUSTIN, W.; BAUMANN, M.; GENGENBACH, T.; HAHN, T.; HELFRICH-SCHKARBANENKO, A.; HEUVELINE, V.; KETELAER, E.; LUKARSKI, D.; NESTLER, A.; RITTERBUSCH, S.; RONNAS, S.; SCHICK, M.; SCHMIDTOBREICK, M.; SUBRAMANIAN, C.; WEISS, J.-P.; WILHELM, F.; AND WLOTZKA, M., 2012. HiFlow3: A Hardware-Aware Parallel Finite Element Package. In *Tools for High Performance Computing 2011* (Eds. H. BRUNST; M. S. MÜLLER; W. E. NAGEL; AND M. M. RESCH), 139–151. Springer Berlin Heidelberg.

7. ARELLANO-GARCIA, H.; SCHÖNEBERGER, J.; AND KÖRKEL, S., 2007. Optimale Versuchsplanung in der Chemischen Verfahrenstechnik. *Chemie-Ingenieur-Technik*, 79, 10 (2007), 1625–1638. doi:10.1002/cite.200700110.

8. ASPEN TECHNOLOGY. Aspen Chromatography. http://www.aspentech.com/ products/aspen-chromatography.aspx. Accessed May 2015.

9. ATKINSON, A.; DONEV, A.; AND TOBIAS, R., 2010. *Optimum Experimental Designs, With SAS*. Oxford University Press. ISBN 13: 9780199296606. doi: 10.1198/jasa.2008.s258.

10. BARZ, T.; LÖFFLER, V.; ARELLANO-GARCIA, H.; AND WOZNY, G., 2010. Optimal determination of steric mass action model parameters for β-lactoglobulin using static batch experiments. *Journal of Chromatography A*, 1217, 26 (2010), 4267–4277. doi: 10.1016/j.chroma.2010.04.001.

11. BAUER, I.; BOCK, H. G.; KÖRKEL, S.; AND SCHLÖDER, J. P., 2000. Numerical methods for optimum experimental design in DAE systems. *Journal of Computational and Applied Mathematics*, 120, 1-2 (2000), 1–25. doi:10.1016/S0377-0427(00)00300-9.

12. BAUMANN, P.; BLUTHARDT, N.; RENNER, S.; BURGHARDT, H.; OSBERGHAUS, A.; AND HUBBUCH, J., 2015. Integrated development of up- and downstream processes supported by the Cherry-Tag™ for real-time tracking of stability and solubility of proteins. *Journal of Biotechnology*, 200 (2015), 27–37. doi:10.1016/j.jbiotec.2015.02. 024.

13. BAUMANN, P.; HAHN, T.; AND HUBBUCH, J., 2015. High-throughput micro-scale cultivations and chromatography modeling: Powerful tools for integrated process development. *Biotechnology and Bioengineering*, (2015). doi:10.1002/bit.25630.

14. BEACHKOFSKI, B. K. AND GRANDHI, R. V., 2002. Improved Distributed Hypercube Sampling. In *AIAA Paper*, 1–8. doi:10.2514/6.2002-1274.

15. BENSCH, M.; SCHULZE WIERLING, P.; VON LIERES, E.; AND HUBBUCH, J., 2005. High throughput screening of chromatographic phases for rapid process development. *Chemical Engineering and Technology*, 28, 11 (2005), 1274–1284. doi:10.1002/ceat. 200500153.

16. BERNSTEIN, D. I.; EL SAHLY, H. M.; KEITEL, W. A.; WOLFF, M.; SIMONE, G.; SEGAWA, C.; WONG, S.; SHELLY, D.; YOUNG, N. S.; AND DEMPSEY, W., 2011. Safety and immunogenicity of a candidate parvovirus B19 vaccine. *Vaccine*, 29, 43 (2011), 7357–7363. doi:10.1016/j.vaccine.2011.07.080.

17. BESSELINK, T.; VAN DER PADT, A.; JANSSEN, A. E. M.; AND BOOM, R. M., 2013. Are axial and radial flow chromatography different? *Journal of Chromatography A*, 1271 (2013), 105–114. doi:10.1016/j.chroma.2012.11.027.

18. BETTS, J. I. AND BAGANZ, F., 2006. Miniature bioreactors: current practices and future opportunities. *Microbial cell factories*, 5 (2006), 21. doi:10.1186/1475-2859-5-21.

19. BHAMBURE, R.; KUMAR, K.; AND RATHORE, A., 2011. High throughput process development for biopharmaceutical drug substances. *Trends in Biotechnology*, 29, 3 (2011), 127–135. doi:10.1016/j.tibtech.2010.12.001.

20. BORG, N.; WESTERBERG, K.; ANDERSSON, N.; VON LIERES, E.; AND NILSSON, B., 2013. Effects of uncertainties in experimental conditions on the estimation of adsorption model parameters in preparative chromatography. *Computers and Chemical Engineering*, 55 (2013), 148–157. doi:10.1016/j.compchemeng.2013.04.013.

21. BOSWELL, P. G.; STOLL, D. R.; CARR, P. W.; NAGEL, M. L.; VITHA, M. F.; AND MABBOTT, G. A., 2013. An advanced, interactive, high-performance liquid chromatography simulator and instructor resources. *Journal of Chemical Education*, 90, 2 (2013), 198–202. doi:10.1021/ed300117b.

22. BRESTRICH, N.; BRISKOT, T.; OSBERGHAUS, A.; AND HUBBUCH, J., 2014. A tool for selective inline quantification of co-eluting proteins in chromatography using spectral analysis and partial least squares regression. *Biotechnology and Bioengineering*, 111, 7 (2014), 1365–1373. doi:10.1002/bit.25194.

23. BRESTRICH, N.; SANDEN, A.; KRAFT, A.; MCCANN, K.; BERTOLINI, J.; AND HUBBUCH, J., 2015. Advances in inline quantification of co-eluting proteins in chromatography: Process-data-based model calibration and application towards real-life separation issues. *Biotechnology and Bioengineering*, 112, 7 (Jul. 2015). doi: 10.1002/bit.25546.

24. BROOKS, C. A. AND CRAMER, S. M., 1992. Steric mass-action ion exchange: Displacement profiles and induced salt gradients. *AIChE Journal*, 38, 12 (Dec. 1992), 1969–1978. doi:10.1002/aic.690381212.

25. BUDDE, I.; STEIL, L.; SCHARF, C.; VÖLKER, U.; AND BREMER, E., 2006. Adaptation of Bacillus subtilis to growth at low temperature: A combined transcriptomic and proteomic appraisal. *Microbiology*, 152, 3 (2006), 831–853. doi: 10.1099/mic.0.28530-0.

26. CALIPER LIFE SCIENCES, 2010. HT Protein Express LabChip Â® Kit , Version 2 LabChip GXII User Guide.

27. Cao, Y.; Li, S.; Petzold, L.; and Serban, R., 2003. Adjoint Sensitivity Analysis for Differential-Algebraic Equations: The Adjoint DAE System and Its Numerical Solution. *SIAM Journal on Scientific Computing*, 24, 3 (2003), 1076–1089. doi: 10.1137/S1064827501380630.

28. Carta, G. and Perez-Almodovar, E. X., 2010. Productivity Considerations and Design Charts for Biomolecule Capture with Periodic Countercurrent Adsorption Systems. *Separation Science and Technology*, 45, 2 (2010), 149–154. doi:10.1080/01496390903423865.

29. Chandramouli, S.; Medina-Selby, A.; Coit, D.; Schaefer, M.; Spencer, T.; Brito, L. A.; Zhang, P.; Otten, G.; Mandl, C. W.; Mason, P. W.; Dormitzer, P. R.; and Settembre, E. C., 2013. Generation of a parvovirus B19 vaccine candidate. *Vaccine*, 31, 37 (Aug. 2013), 3872–3878. doi:10.1016/j.vaccine.2013.06.062.

30. Chen, W. Y.; Liu, Z. C.; Lin, P. H.; Fang, C. I.; and Yamamoto, S., 2007. The hydrophobic interactions of the ion-exchanger resin ligands with proteins at high salt concentrations by adsorption isotherms and isothermal titration calorimetry. *Separation and Purification Technology*, 54, 2 (2007), 212–219. doi:10.1016/j.seppur.2006.09.008.

31. Chhatre, S.; Farid, S. S.; Coffman, J.; Bird, P.; Newcombe, A. R.; and Titchener-Hooker, N. J., 2011. How implementation of quality by design and advances in biochemical engineering are enabling efficient bioprocess development and manufacture. *Journal of Chemical Technology and Biotechnology*, 86, 9 (2011), 1125–1129. doi:10.1002/jctb.2628.

32. ChromWorks. ChromWorks. http://chromworks.com. Accessed May 2015.

33. Chung, W. K.; Freed, A. S.; Holstein, M. a.; McCallum, S. a.; and Cramer, S. M., 2010. Evaluation of protein adsorption and preferred binding regions in multimodal chromatography using NMR. *Proceedings of the National Academy of Sciences of the United States of America*, 107, 39 (2010), 16811–16816. doi:10.1073/pnas.1002347107.

34. Close, E. J.; Salm, J. R.; Bracewell, D. G.; and Sorensen, E., 2014. A model based approach for identifying robust operating conditions for industrial chromatography with process variability. *Chemical Engineering Science*, 116 (2014), 284–295. doi:10.1016/j.ces.2014.03.010.

35. Close, E. J.; Salm, J. R.; Bracewell, D. G.; and Sorensen, E., 2014. Modelling of industrial biopharmaceutical multicomponent chromatography. *Chemical Engineering Research and Design*, 92, 7 (2014), 1304–1314. doi:10.1016/j.cherd.2013.10.022.

36. Coffman, J. L.; Kramarczyk, J. F.; and Kelley, B. D., 2008. High-throughput screening of chromatographic separations: I. method development and column modeling. *Biotechnology and Bioengineering*, 100, 4 (2008), 605–618. doi:10.1002/bit.21904.

37. CORNEL, J.; TARAFDER, A.; KATSUO, S.; AND MAZZOTTI, M., 2010. The direct inverse method: A novel approach to estimate adsorption isotherm parameters. *Journal of Chromatography A*, 1217, 12 (2010), 1934–1941. doi:DOI10.1016/j.chroma.2010.01.063.

38. CRAMER, H., 1957. *Mathematical Methods of Statistics.* Princeton University Press. ISBN 9780691005478. doi:10.2307/2280199.

39. DANCKWERTS, P., 1953. Continuous flow systems. *Chemical Engineering Science*, 2, 1 (Feb. 1953), 1–13. doi:10.1016/0009-2509(53)80001-1.

40. DEGERMAN, M.; JAKOBSSON, N.; AND NILSSON, B., 2006. Constrained optimization of a preparative ion-exchange step for antibody purification. *Journal of Chromatography A*, 1113, 1-2 (Apr. 2006), 92–100. doi:10.1016/j.chroma.2006.01.121.

41. DEGERMAN, M.; WESTERBERG, K.; AND NILSSON, B., 2009. A Model-based approach to determine the design space of preparative Chromatography. *Chemical Engineering and Technology*, 32, 8 (2009), 1195–1202. doi:10.1002/ceat.200900102.

42. DELPHI GENETICS SA, 2009. Cherry Codon Kit. http://www.delphigenetics.com/upload/manuals/CherryCodonmanual-v3.0.pdf.

43. DELPHI GENETICS SA, 2009. Cherry Express Kit. http://www.delphigenetics.com/upload/manuals/CherryExpressmanual-v3.0.pdf.

44. DEUFLHARD, P., 2005. *Newton methods for nonlinear problems: Affine invariance and adaptive algorithms.* Springer Series in Computational Mathematics. Springer. ISBN 978-3-642-23898-7. doi:10.1007/978-3-642-23899-4.

45. DEVERNAY, F. C/C++ Minpack, ver. 1.3.2. http://devernay.free.fr/hacks/cminpack/.

46. DICKINSON, F. M.; HART, G. J.; AND KITSON, T. M., 1981. The use of pH-gradient ion-exchange chromatography to separate sheep liver cytoplasmic aldehyde dehydrogenase from mitochondrial enzyme contamination, and observations on the interaction between the pure cytoplasmic enzyme and disulfiram. *The Biochemical journal*, 199, 3 (1981), 573–579.

47. DIKSHIT, K. L.; DIKSHIT, R. P.; AND WEBSTER, D. A., 1990. Study of Vitreoscilla globin(vgb) gene expression and promoter activity in E. Coli through transcriptional fusion. *Nucleic Acids Research*, 18, 14 (1990), 4149–4155. doi:10.1093/nar/18.14.4149.

48. DIMARTINO, S.; BOI, C.; AND SARTI, G. C., 2011. A validated model for the simulation of protein purification through affinity membrane chromatography. *Journal of Chromatography A*, 1218, 13 (2011), 1677–1690. doi:10.1016/j.chroma.2010.11.056.

49. DÜRRSCHMID, K.; REISCHER, H.; SCHMIDT-HECK, W.; HREBICEK, T.; GUTHKE, R.; RIZZI, A.; AND BAYER, K., 2008. Monitoring of transcriptome and proteome profiles to investigate the cellular response of E. coli towards recombinant protein expression under defined chemostat conditions. *Journal of Biotechnology*, 135, 1 (2008), 34–44. doi:10.1016/j.jbiotec.2008.02.013.

50. DZIENNIK, S. R.; BELCHER, E. B.; BARKER, G. A.; AND LENHOFF, A. M., 2005. Effects of ionic strength on lysozyme uptake rates in cation exchangers. I: Uptake in SP sepharose FF. *Biotechnology and Bioengineering*, 91, 2 (Jul. 2005), 139–153. doi:10.1002/bit.20503.

51. ENMARK, M.; ARNELL, R.; FORSSÉN, P.; SAMUELSSON, J.; KACZMARSKI, K.; AND FORNSTEDT, T., 2011. A systematic investigation of algorithm impact in preparative chromatography with experimental verifications. *Journal of Chromatography A*, 1218, 5 (2011), 662–672. doi:10.1016/j.chroma.2010.11.029.

52. FORSCHUNGSZENTRUM JÜLICH. CADET. https://github.com/modsim/CADET. Accessed May 2015.

53. FORSSÉN, P.; ARNELL, R.; AND FORNSTEDT, T., 2006. An improved algorithm for solving inverse problems in liquid chromatography. *Computers & Chemical Engineering*, 30, 9 (Jul. 2006), 1381–1391. doi:10.1016/j.compchemeng.2006.03.004.

54. FRIES, L. F.; SMITH, G. E.; AND GLENN, G. M., 2013. A Recombinant Viruslike Particle Influenza A (H7N9) Vaccine. *New England Journal of Medicine*, 369 (2013), 2564–2566. doi:10.1056/NEJMc1313186.

55. FUNKE, M.; DIEDERICHS, S.; KENSY, F.; MÜLLER, C.; AND BÜCHS, J., 2009. The baffled microtiter plate: Increased oxygen transfer and improved online monitoring in small scale fermentations. *Biotechnology and Bioengineering*, 103, 6 (2009), 1118–1128. doi:10.1002/bit.22341.

56. GALLANT, S. R.; KUNDU, A.; AND CRAMER, S. M., 1995. Modeling non-linear elution of proteins in ion-exchange chromatography. *Journal of Chromatography A*, 702, 1-2 (May 1995), 125–142. doi:10.1016/0021-9673(94)00992-I.

57. GALLANT, S. R.; KUNDU, A.; AND CRAMER, S. M., 1995. Optimization of step gradient separations: Consideration of nonlinear adsorption. *Biotechnology and Bioengineering*, 47, 3 (1995), 355–372. doi:10.1002/bit.260470310.

58. GEIGERT, J., 2004. *The challenge of CMC regulatory compliance for biopharmaceuticals.* Springer Science & Business Media, New York. ISBN 1461348048, 9781461348047. doi:10.1007/978-1-4614-6916-2.

59. GERONTAS, S.; ASPLUND, M.; HJORTH, R.; AND BRACEWELL, D. G., 2010. Integration of scale-down experimentation and general rate modelling to predict manufacturing scale chromatographic separations. *Journal of Chromatography A*, 1217, 44 (2010), 6917–6926. doi:10.1016/j.chroma.2010.08.063.

60. GERONTAS, S.; SHAPIRO, M. S.; AND BRACEWELL, D. G., 2013. Chromatography modelling to describe protein adsorption at bead level. *Journal of Chromatography A*, 1284 (2013), 44–52. doi:10.1016/j.chroma.2013.01.102.

61. GILBERT, L.; TOIVOLA, J.; WHITE, D.; IHALAINEN, T.; SMITH, W.; LINDHOLM, L.; VUENTO, M.; AND OKER-BLOM, C., 2005. Molecular and structural characterization of fluorescent human parvovirus B19 virus-like particles. *Biochemical and Biophysical Research Communications*, 331, 2 (2005), 527–535. doi:10.1016/j.bbrc.2005.03.208.

62. GLOWINSKI, R., 2003. Finite element methods for incompressible viscous flow. In *Numerical Methods for Fluids (Part 3)* (Eds. P. CIARLET AND J. LIONS), vol. 9 of *Handbook of Numerical Analysis*, 3 – 1176. Elsevier. doi:10.1016/S1570-8659(03)09003-3.

63. GOODWIN, GRAHAM C, PAYNE, R. L., 1977. *Dynamic System Identification: Experiment Design and Data Analysis.* Academic Press. ISBN 9780080956459.

64. GRÜNENFELDER, B.; RUMMEL, G.; VOHRADSKY, J.; RÖDER, D.; LANGEN, H.; AND JENAL, U., 2001. Proteomic analysis of the bacterial cell cycle. *Proceedings of the National Academy of Sciences of the United States of America*, 98, 8 (2001), 4681–4686. doi:10.1073/pnas.071538098.

65. GU, T., 1995. Multicomponent Radial Flow Chromatography. In *Mathematical Modeling and Scale-up of Liquid Chromatography*, 102–110. Springer Berlin Heidelberg. ISBN 978-3-642-79543-5.

66. GU, T.; TSAI, G.-J.; AND TSAO, G. T., 1990. New approach to a general nonlinear multicomponent chromatography model. *AIChE Journal*, 36, 5 (May 1990), 784–788. doi:10.1002/aic.690360517.

67. GU, T.; TSAI, G.-J.; AND TSAO, G. T., 1991. A theoretical study of multicomponent radial flow chromatography. *Chemical Engineering Science*, 46, 5–6 (1991), 1279–1288. doi:10.1016/0009-2509(91)85055-3.

68. GUÉLAT, B.; STRÖHLEIN, G.; LATTUADA, M.; DELEGRANGE, L.; VALAX, P.; AND MORBIDELLI, M., 2012. Simulation model for overloaded monoclonal antibody variants separations in ion-exchange chromatography. *Journal of Chromatography A*, 1253 (2012), 32–43. doi:10.1016/j.chroma.2012.06.081.

69. GUERRIER, L.; FLAYEUX, I.; AND BOSCHETTI, E., 2001. A dual-mode approach to the selective separation of antibodies and their fragments. *Journal of Chromatography B: Biomedical Sciences and Applications*, 755, 1-2 (2001), 37–46. doi:10.1016/S0378-4347(00)00598-3.

70. GUERRIER, L.; GIROT, P.; SCHWARTZ, W.; AND BOSCHETTI, E., 2000. New method for the selective capture of antibodies under physiolgical conditions. *Bioseparation*, 9, 4 (2000), 211–221. doi:10.1023/A:1008170226665.

71. GUIOCHON, G. AND BEAVER, L. A., 2011. Separation science is the key to successful biopharmaceuticals. *Journal of Chromatography A*, 1218, 49 (2011), 8836–8858. doi:10.1016/j.chroma.2011.09.008.

72. GUIOCHON, G.; SHIRAZI, D. G.; FELINGER, A.; AND KATTI, A. M., 2006. *Fundamentals of Preparative and Nonlinear Chromatography.* Academic Press. ISBN 9780123705372.

73. HAGEL, L.; JAGSCHIES, G.; AND SOFER, G., 2008. *Handbook of Process Chromatography.* Elsevier Academic Press, Amsterdam, 2. ed. edn. ISBN 9780123740236. doi:10.1016/B978-012374023-6.50010-5.

74. HAHN, T.; BAUMANN, P.; HUUK, T.; HEUVELINE, V.; AND HUBBUCH, J., 2015. UV absorption-based inverse modeling of protein chromatography. *Engineering in Life Sciences*, (May 2015). doi:10.1002/elsc.201400247.

75. HAHN, T.; HUUK, T.; HEUVELINE, V.; AND HUBBUCH, J., 2015. Simulating and Optimizing Preparative Protein Chromatography with ChromX. *Journal of Chemical Education*, 0 (2015), 150623161654001. doi:10.1021/ed500854a.

76. HAHN, T.; HUUK, T.; OSBERGHAUS, A.; DONINGER, K.; NATH, S.; HEPBILDIKLER, S.; HEUVELINE, V.; AND HUBBUCH, J., 2015. Calibration-free inverse modeling of ion-exchange chromatography in industrial antibody purification. *Engineering in Life Sciences*, (2015). doi:10.1002/elsc.201400248.

77. HAHN, T.; SOMMER, A.; OSBERGHAUS, A.; HEUVELINE, V.; AND HUBBUCH, J., 2014. Adjoint-based estimation and optimization for column liquid chromatography models. *Computers and Chemical Engineering*, 64 (2014), 41–54. doi:10.1016/j.compchemeng.2014.01.013.

78. HANKE, A. T. AND OTTENS, M., 2014. Purifying biopharmaceuticals: Knowledge-based chromatographic process development. *Trends in Biotechnology*, 32, 4 (2014), 210–220. doi:10.1016/j.tibtech.2014.02.001.

79. 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 (2011), 2661–2669. doi:10.1002/bit.23229.

80. HASCOËT, L. AND DAUVERGNE, B., 2008. Adjoints of large simulation codes through Automatic Differentiation. *Revue européenne de mécanique numérique*, 17, 1-2 (2008), 63–86. doi:10.3166/remn.17.63-86.

81. HELLING, C.; BORRMANN, C.; AND STRUBE, J., 2012. Optimal Integration of Directly Combined Hydrophobic Interaction and Ion Exchange Chromatography Purification Processes. *Chemical Engineering and Technology*, 35, 10 (2012), 1786–1796. doi:10.1002/ceat.201200043.

82. HERRMANN, T.; SCHRÖDER, M.; AND HUBBUCH, J., 2006. Generation of equally sized particle plaques using solid-liquid suspensions. *Biotechnology Progress*, 22, 3 (2006), 914–918. doi:10.1021/bp050296i.

83. HOU, Y. AND CRAMER, S. M., 2011. Evaluation of selectivity in multimodal anion exchange systems: A priori prediction of protein retention and examination of mobile phase modifier effects. *Journal of Chromatography A*, 1218, 43 (2011), 7813–7820. doi:10.1016/j.chroma.2011.08.080.

84. HUBBUCH, J.; LINDEN, T.; KNIEPS, E.; LJUNGLÖF, A.; THÖMMES, J.; AND KULA, M. R., 2003. Mechanism and kinetics of protein transport in chromatographic media studied by confocal laser scanning microscopy: Part I. The interplay of sorbent structure and fluid phase conditions. *Journal of Chromatography A*, 1021, 1-2 (2003), 93–104. doi:10.1016/j.chroma.2003.08.112.

85. HUUK, T. C.; HAHN, T.; OSBERGHAUS, A.; AND HUBBUCH, J., 2014. Model-based integrated optimization and evaluation of a multi-step ion exchange chromatography. *Separation and Purification Technology*, 136 (Nov. 2014), 207–222. doi:10.1016/j.seppur.2014.09.012.

86. ICH, 2011. ICH-Endorsed Guide for ICH Q8/Q9/Q10 Implementation. In *International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use*.

87. JAKOBSSON, N.; KARLSSON, D.; AXELSSON, J. P.; ZACCHI, G.; AND NILSSON, B., 2005. Using computer simulation to assist in the robustness analysis of an ion-exchange chromatography step. *Journal of Chromatography A*, 1063, 1-2 (2005), 99–109. doi:10.1016/j.chroma.2004.11.067.

88. JAMES, F.; SEPÚLVEDA, M.; CHARTON, F.; QUINONES, I.; AND GUIOCHON, G., 1999. Determination of binary competitive equilibrium isotherms from the individual chromatographic band profiles. *Chemical Engineering Science*, 54, 11 (1999), 1677–1696. doi:10.1016/S0009-2509(98)00539-9.

89. JAVEED, S.; QAMAR, S.; SEIDEL-MORGENSTERN, A.; AND WARNECKE, G., 2011. A discontinuous Galerkin method to solve chromatographic models. *Journal of Chromatography A*, 1218, 40 (2011), 7137–7146. doi:10.1016/j.chroma.2011.08.005.

90. JOHANSSON, B. L.; BELEW, M.; ERIKSSON, S.; GLAD, G.; LIND, O.; MALOISEL, J. L.; AND NORRMAN, N., 2003. Preparation and characterization of prototypes for multi-modal separation media aimed for capture of negatively charged biomolecules at high salt conditions. *Journal of Chromatography A*, 1016, 1 (2003), 21–33. doi:10.1016/S0021-9673(03)01140-3.

91. JUNGBAUER, A., 1996. Insights into the chromatography of proteins provided by mathematical modeling. *Current Opinion in Biotechnology*, 7, 2 (1996), 210–218. doi:10.1016/S0958-1669(96)80015-2.

92. JUNGBAUER, A., 2013. Continuous downstream processing of biopharmaceuticals. *Trends in Biotechnology*, 31, 8 (2013), 479–492. doi:10.1016/j.tibtech.2013.05.011.

93. JUNGBAUER, A.; MACHOLD, C.; AND HAHN, R., 2005. Hydrophobic interaction chromatography of proteins: III. Unfolding of proteins upon adsorption. *Journal of Chromatography A*, 1079 (2005), 221–228.

94. KACZMARSKI, K. AND ANTOS, D., 2006. Use of simulated annealing for optimization of chromatographic separations. *Acta Chromatographica*, , 17 (2006), 20–45.

95. KALLBERG, K.; BECKER, K.; AND BÜLOW, L., 2011. Application of a pH responsive multimodal hydrophobic interaction chromatography medium for the analysis of glycosylated proteins. *Journal of Chromatography A*, 1218, 5 (2011), 678–683. doi:10.1016/j.chroma.2010.11.080.

96. KARGER, B. L. AND BLANCO, R., 1989. The effect of on-column structural changes of proteins on their HPLC behavior. *Talanta*, 36, 1-2 (Jan. 1989), 243–248. doi:10.1016/0039-9140(89)80102-X.

97. KARKOV, H. S.; SEJERGAARD, L.; AND CRAMER, S. M., 2013. Methods development in multimodal chromatography with mobile phase modifiers using the steric mass action model. *Journal of Chromatography A*, 1318 (2013), 149–155. doi: 10.1016/j.chroma.2013.10.004.

98. KARLSRUHE INSTITUTE OF TECHNOLOGY (KIT). ChromX. http://www.chromx.org. Accessed Mar 2015.

99. KARLSSON, D.; JAKOBSSON, N.; AXELSSON, A.; AND NILSSON, B., 2004. Model-based optimization of a preparative ion-exchange step for antibody purification. *Journal of Chromatography A*, 1055, 1-2 (2004), 29–39. doi:10.1016/j.chroma.2004.08.151.

100. KARLSSON, D.; JAKOBSSON, N.; BRINK, K.-J.; AXELSSON, A.; AND NILSSON, B., 2004. Methodologies for model calibration to assist the design of a preparative ion-exchange step for antibody purification. *Journal of Chromatography A*, 1033, 1 (Apr. 2004), 71–82. doi:10.1016/j.chroma.2003.12.072.

101. KELLEY, B. D.; SWITZER, M.; BASTEK, P.; KRAMARCZYK, J. F.; MOLNAR, K.; YU, T.; AND COFFMAN, J., 2008. High-throughput screening of chromatographic separations: IV. Ion-exchange. *Biotechnology and Bioengineering*, 100, 5 (Aug. 2008), 950–963. doi:10.1002/bit.21905.

102. KENNEDY, L. a.; KOPACIEWICZ, W.; AND REGNIER, F. E., 1986. Multimodal liquid chromatography columns for the separation of proteins in either the anion-exchange or hydrophobic-interaction mode. *Journal of Chromatography A*, 359 (Jan. 1986), 73–84. doi:10.1016/0021-9673(86)80063-2.

103. KENSY, F.; ZANG, E.; FAULHAMMER, C.; TAN, R.-K.; AND BÜCHS, J., 2009. Validation of a high-throughput fermentation system based on online monitoring of biomass and fluorescence in continuously shaken microtiter plates. *Microbial cell factories*, 8 (2009), 31. doi:10.1186/1475-2859-8-31.

104. KNÄBLEIN, J., 2013. *Twenty Thousand Years of Biotech - From "Traditional" to "Modern Biotechnology"*, 1–38. Wiley-VCH Verlag GmbH & Co. KGaA. ISBN 9783527669417. doi:10.1002/9783527669417.ch1.

105. KONTORAVDI, C.; SAMSATLI, N. J.; AND SHAH, N., 2013. Development and design of bio-pharmaceutical processes. *Current Opinion in Chemical Engineering*, 137 (2013), 1–7. doi:10.1016/j.coche.2013.09.007.

106. KOTTMEIER, K.; WEBER, J.; MÜLLER, C.; BLEY, T.; AND BÜCHS, J., 2009. Asymmetric division of Hansenula polymorpha reflected by a drop of light scatter intensity measured in batch microtiter plate cultivations at phosphate limitation. *Biotechnology and Bioengineering*, 104, 3 (2009), 554–561. doi:10.1002/bit.22410.

107. KRÖNER, F. AND HUBBUCH, J., 2013. Systematic generation of buffer systems for pH gradient ion exchange chromatography and their application. *Journal of Chromatography A*, 1285 (2013), 78–87. doi:10.1016/j.chroma.2013.02.017.

108. LADD EFFIO, C. AND HUBBUCH, J., 2015. Next generation vaccines and vectors: Designing downstream processes for recombinant protein-based virus-like particles. *Biotechnology journal*, 10, 5 (2015), 715–727. doi:10.1002/biot.201400392.

109. LADD EFFIO, C.; WENGER, L.; ÖTES, O.; OELMEIER, S. A.; KNEUSEL, R.; AND HUBBUCH, J., 2015. Downstream processing of virus-like particles: Single-stage and multi-stage aqueous two-phase extraction. *Journal of Chromatography A*, 1383 (2015), 35–46. doi:10.1016/j.chroma.2015.01.007.

110. LADIWALA, A.; REGE, K.; BRENEMAN, C. M.; AND CRAMER, S. M., 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 (2005), 11710–11715. doi:10.1073/pnas.0408769102.

111. LAMPING, S.; ZHANG, H.; ALLEN, B.; AND AYAZI SHAMLOU, P., 2003. Design of a prototype miniature bioreactor for high throughput automated bioprocessing. *Chemical Engineering Science*, 58, 3-6 (Feb. 2003), 747–758. doi:10.1016/S0009-2509(02)00604-8.

112. LEE, K.; LEE, H. G.; PI, K.; AND CHOI, Y. J., 2008. The effect of low pH on protein expression by the probiotic bacterium Lactobacillus reuteri. *Proteomics*, 8, 8 (2008), 1624–1630. doi:10.1002/pmic.200700663.

113. LI, S. AND PETZOLD, L., 2004. Adjoint sensitivity analysis for time-dependent partial differential equations with adaptive mesh refinement. *Journal of Computational Physics*, 198, 1 (2004), 310–325. doi:10.1016/j.jcp.2003.01.001.

114. LIAPIS, A. I. AND RIPPIN, D. W. T., 1979. The simulation of binary adsorption in continuous countercurrent operation and a comparison with other operating modes. *AIChE Journal*, 25, 3 (May 1979), 455–460. doi:10.1002/aic.690250310.

115. LIU, H. F.; MA, J.; WINTER, C.; AND BAYER, R., 2010. Recovery and purification process development for monoclonal antibody production. *mAbs*, 2, 5 (2010), 480–499. doi:10.4161/mabs.2.5.12645.

116. LOSEN, M.; FRÖLICH, B.; POHL, M.; AND BÜCHS, J., 2004. Effect of Oxygen Limitation and Medium Composition on *Escherichia coli* Fermentation in Shake-Flask Cultures. *Biotechnology Progress*, 20, 4 (2004), 1062–1068. doi:10.1021/bp034282t.

117. LOURAKIS, M., 2004. levmar: Levenberg-Marquardt nonlinear least squares algorithms in C/C++. http://www.ics.forth.gr/~lourakis/levmar/.

118. LUND UNIVERSITY. pcs. http://www.chemeng.lth.se/pcs. Accessed May 2015.

119. MA, Z. AND GUIOCHON, G., 1991. Application of orthogonal collocation on finite elements in the simulation of non-linear chromatography. *Computers & Chemical Engineering*, 15, 6 (Jun. 1991), 415–426. doi:10.1016/0098-1354(91)87019-6.

120. MA, Z.; WHITLEY, R. D.; AND WANG, N.-H. L., 1996. Pore and surface diffusion in multicomponent adsorption and liquid chromatography systems. *AIChE Journal*, 42, 5 (May 1996), 1244–1262. doi:10.1002/aic.690420507.

121. MACH, H.; VOLKIN, D. B.; TROUTMAN, R. D.; WANG, B.; LUO, Z.; JANSEN, K. U.; AND SHI, L., 2006. Disassembly and reassembly of yeast-derived recombinant human papillomavirus virus-like particles (HPV VLPs). *Journal of Pharmaceutical Sciences*, 95, 10 (2006), 2195–2206. doi:10.1002/jps.20696.

122. Mahajan, E.; George, A.; and Wolk, B., 2012. Improving affinity chromatography resin efficiency using semi-continuous chromatography. *Journal of Chromatography A*, 1227 (2012), 154–162. doi:10.1016/j.chroma.2011.12.106.

123. Maiser, B.; Kröner, F.; Dismer, F.; Brenner-Weiss, G.; and Hubbuch, J., 2012. Isoform separation and binding site determination of mono-PEGylated lysozyme with pH gradient chromatography. *Journal of Chromatography A*, 1268 (2012), 102–108. doi:10.1016/j.chroma.2012.10.047.

124. Marek, W.; Muca, R.; Woś, S.; Piatkowski, W.; and Antos, D., 2013. Isolation of monoclonal antibody from a Chinese hamster ovary supernatant. II: Dynamics of the integrated separation on ion exchange and hydrophobic interaction chromatography media. *Journal of Chromatography A*, 1305 (2013), 64–75. doi:10.1016/j.chroma.2013.06.076.

125. Marino, M.; Hoffmann, T.; Schmid, R.; Möbitz, H.; and Jahn, D., 2000. Changes in protein synthesis during the adaptation of Bacillus subtilis to anaerobic growth conditions. *Microbiology*, 146 (Pt. 1) (2000), 97–105.

126. McNally, D. J.; Darling, D.; Farzaneh, F.; Levison, P. R.; and Slater, N. K. H., 2014. Optimised concentration and purification of retroviruses using membrane chromatography. *Journal of Chromatography A*, 1340 (2014), 24–32. doi:10.1016/j.chroma.2014.03.023.

127. McNay, J. L. and Fernandez, E. J., 1999. How does a protein unfold on a reversed-phase liquid chromatography surface? *Journal of Chromatography A*, 849, 1 (1999), 135–148. doi:10.1016/S0021-9673(99)00546-4.

128. Miao, F.; Jen, D.; Jones, T.; Vora, H.; Ramelmeier, A.; Lacki, K. M.; and Bryntesson, L. M., 2004. Feasibility study of periodic counter-current chromatography for the capture of monoclonal antibodies from crude harvest with MabSelect Protein A resin. In *Downstream GAb '04 Abstracts: Extended Reports from the 3rd International Symposium on Downstream Processing of Genetically Engineered Antibodies and Related Molecules*. GE Healthcare/Amersham Biosciences, 2005, Nice, France.

129. Michel, M.; Epping, A.; and Jupke, A., 2005. *Modeling and Determination of Model Parameters*, chap. 6, 215–312. Wiley-VCH Verlag GmbH & Co. KGaA. ISBN 9783527306435. doi:10.1002/3527603484.ch6.

130. Mohammadi, D., 2015. International community ramps up Ebola vaccine eff ort Effi cacy trials of candidate Ebola vaccines are to start this year following discussions at a. *The Lancet*, 384, 9955 (Nov. 2015), 1658–1659. doi:10.1016/S0140-6736(14)61788-8.

131. Mollerup, J. M., 2006. Applied thermodynamics: A new frontier for biotechnology. *Fluid Phase Equilibria*, 241, 1-2 (2006), 205–215. doi:10.1016/j.fluid.2005.12.037.

132. Mollerup, J. M., 2008. A review of the thermodynamics of protein association to ligands, protein adsorption, and adsorption isotherms. *Chemical Engineering and Technology*, 31, 6 (2008), 864–874. doi:10.1002/ceat.200800082.

133. MOLLERUP, J. M.; HANSEN, T. B.; KIDAL, S.; SEJERGAARD, L.; AND STABY, A., 2007. Development, modelling, optimisation and scale-up of chromatographic purification of a therapeutic protein. *Fluid Phase Equilibria*, 261, 1-2 (Dec. 2007), 133–139. doi:10.1016/j.fluid.2007.07.047.

134. MOLLERUP, J. M.; HANSEN, T. B.; KIDAL, S.; AND STABY, A., 2008. Quality by design-Thermodynamic modelling of chromatographic separation of proteins. *Journal of Chromatography A*, 1177, 2 (2008), 200–206. doi:10.1016/j.chroma.2007.08.059.

135. MUCA, R.; PIATKOWSKI, W.; AND ANTOS, D., 2009. Altering efficiency of hydrophobic interaction chromatography by combined salt and temperature effects. *Journal of Chromatography A*, 1216, 50 (2009), 8712–8721. doi:10.1016/j.chroma.2009.04.046.

136. NERETNIEKS, I., 1975. A simplified theoretical comparison of periodic and countercurrent adsorption. *Chemie Ingenieur Technik - CIT*, 47, 18 (1975), 773–773. doi:10.1002/cite.330471810.

137. NESTLER, A., 2012. *Parameteridentifikation und Optimale Versuchsplanung bei instationären partiellen Differentialgleichungen*. Ph.D. thesis, Karlsruhe Institute of Technology (KIT).

138. NESTOLA, P.; VILLAIN, L.; PEIXOTO, C.; MARTINS, D. L.; ALVES, P. M.; CARRONDO, M. J. T.; AND MOTA, J. P. B., 2014. Impact of grafting on the design of new membrane adsorbers for adenovirus purification. *Journal of Biotechnology*, 181 (2014), 1–11. doi:10.1016/j.jbiotec.2014.04.003.

139. NEVILLE, D. C. a.; DWEK, R. a.; AND BUTTERS, T. D., 2009. Development of a single column method for the separation of lipid- and protein-derived oligosaccharides. *Journal of Proteome Research*, 8, 2 (Feb. 2009), 681–687. doi:10.1021/pr800704t.

140. NFOR, B. K.; AHAMED, T.; VAN DEDEM, G. W. K.; VAN DER WIELEN, L. A. M.; VAN DE SANDT, E. J. A. X.; EPPINK, M. H. M.; AND OTTENS, M., 2008. Design strategies for integrated protein purification processes: Challenges, progress and outlook. *Journal of Chemical Technology and Biotechnology*, 83, 2 (2008), 124–132. doi:10.1002/jctb.1815.

141. NFOR, B. K.; AHAMED, T.; VAN DEDEM, G. W. K.; VERHAERT, P. D. E. M.; VAN DER WIELEN, L. A. M.; EPPINK, M. H. M.; VAN DE SANDT, E. J. A. X.; AND OTTENS, M., 2013. Model-based rational methodology for protein purification process synthesis. *Chemical Engineering Science*, 89 (2013), 185–195. doi:10.1016/j.ces.2012.11.034.

142. NFOR, B. K.; NOVERRAZ, M.; CHILAMKURTHI, S.; VERHAERT, P. D. E. M.; VAN DER WIELEN, L. A. M.; AND OTTENS, M., 2010. High-throughput isotherm determination and thermodynamic modeling of protein adsorption on mixed mode adsorbents. *Journal of Chromatography A*, 1217, 44 (2010), 6829–6850. doi:10.1016/j.chroma.2010.07.069.

143. NFOR, B. K.; RIPIĆ, J.; VAN DER PADT, A.; JACOBS, M.; AND OTTENS, M., 2012. Model-based high-throughput process development for chromatographic whey

proteins separation. *Biotechnology Journal*, 7, 10 (2012), 1221–1232. doi:10.1002/biot.201200191.

144. Nfor, B. K.; Verhaert, P. D. E. M.; van der Wielen, L. A. M.; Hubbuch, J.; and Ottens, M., 2009. Rational and systematic protein purification process development: the next generation. *Trends in Biotechnology*, 27, 12 (2009), 673–679. doi:10.1016/j.tibtech.2009.09.002.

145. Oelmeier, S. A.; Dismer, F.; and Hubbuch, J., 2010. Application of an aqueous two-phase systems high throughput screening method to evaluate mab HCP separation. *Biotechnology and bioengineering*, 108, 1 (2010), 69–81. doi:10.1002/bit.22900.

146. Opitz, L.; Lehmann, S.; Zimmermann, A.; Reichl, U.; and Wolff, M. W., 2007. Impact of adsorbents selection on capture efficiency of cell culture derived human influenza viruses. *Journal of Biotechnology*, 131, 3 (2007), 309–317. doi:10.1016/j.jbiotec.2007.07.723.

147. Orr, V.; Zhong, L.; Moo-Young, M.; and Chou, C. P., 2013. Recent advances in bioprocessing application of membrane chromatography. *Biotechnology Advances*, 31, 4 (2013), 450–465. doi:10.1016/j.biotechadv.2013.01.007.

148. Osberghaus, A.; Baumann, P.; Hepbildikler, S.; Nath, S.; Haindl, M.; von Lieres, E.; and Hubbuch, J., 2012. Detection, quantification, and propagation of uncertainty in high-throughput experimentation by Monte Carlo methods. *Chemical Engineering and Technology*, 35, 8 (2012), 1456–1464. doi:10.1002/ceat.201100610.

149. Osberghaus, A.; Hepbildikler, S.; Nath, S.; Haindl, M.; von Lieres, E.; and Hubbuch, J., 2012. Determination of parameters for the steric mass action model-A comparison between two approaches. *Journal of Chromatography A*, 1233 (2012), 54–65. doi:10.1016/j.chroma.2012.02.004.

150. Osberghaus, A.; Hepbildikler, S.; Nath, S.; Haindl, M.; von Lieres, E.; and Hubbuch, J., 2012. Optimizing a chromatographic three component separation: A comparison of mechanistic and empiric modeling approaches. *Journal of Chromatography A*, 1237 (2012), 86–95. doi:10.1016/j.chroma.2012.03.029.

151. Papadimitriou, D. I. and Giannakoglou, K. C., 2008. Direct, adjoint and mixed approaches for the computation of Hessian in airfoil design problems. *International Journal for Numerical Methods in Fluids*, 56, 10 (Apr. 2008), 1929–1943. doi:10.1002/fld.1584.

152. Parente, E. S. and Wetlaufer, D. B., 1986. Relationship between isocratic and gradient retention times in the high-performance ion-exchange chromatography of proteins. *Journal of Chromatography A*, 355, 1 (Jan. 1986), 29–40. doi:10.1016/S0021-9673(01)97301-7.

153. Pollock, J.; Bolton, G.; Coffman, J.; Ho, S. V.; Bracewell, D. G.; and Farid, S. S., 2013. Optimising the design and operation of semi-continuous affinity chromatography for clinical and commercial manufacture. *Journal of Chromatography A*, 1284 (2013), 17–27. doi:10.1016/j.chroma.2013.01.082.

154. POLYKARPOU, E. M.; DALBY, P. A.; AND PAPAGEORGIOU, L. G., 2012. A novel efficient optimisation system for purification process synthesis. *Biochemical Engineering Journal*, 67 (2012), 186–193. doi:10.1016/j.bej.2012.06.012.

155. POLYKARPOU, E. M.; DALBY, P. A.; AND PAPAGEORGIOU, L. G., 2012. An MILP formulation for the synthesis of protein purification processes. *Chemical Engineering Research and Design*, 90, 9 (2012), 1262–1270. doi:10.1016/j.cherd.2011.11.021.

156. PROCESS SYSTEMS ENTERPRISE. gProms. http://www.psenterprise.com/gproms.html. Accessed May 2015.

157. PUSKEILER, R.; KAUFMANN, K.; AND WEUSTER-BOTZ, D., 2005. Development, parallelization, and automation of a gas-inducing milliliter-scale bioreactor for high-throughput bioprocess design (HTBD). *Biotechnology and Bioengineering*, 89, 5 (Mar. 2005), 512–523. doi:10.1002/bit.20352.

158. PÜTTMANN, A.; SCHNITTERT, S.; NAUMANN, U.; AND VON LIERES, E., 2013. Fast and accurate parameter sensitivities for the general rate model of column liquid chromatography. *Computers & Chemical Engineering*, (2013). doi:10.1016/j.compchemeng.2013.04.021.

159. QUARTERONI, A. AND VALLI, A., 1994. *Numerical approximation of partial differential equations*. Springer; Berlin, Heidelberg, 1st ed. edn. ISBN 9783540852674. doi:10.1007/978-3-540-85268-1.

160. QURESHI, Z. H.; NG, T. S.; AND GOODWIN, G. C., 1980. Optimum experimental design for identification of distributed parameter systems. *International Journal of Control*, 31, 1 (1980), 21–29. doi:10.1080/00207178008961025.

161. RAO, C. S., 2001. Purification of large proteins using ion-exchange membranes. *Process Biochemistry*, 37, 3 (2001), 247–256. doi:10.1016/S0032-9592(01)00207-2.

162. RATHORE, A. S. AND WINKLE, H., 2009. Quality by design for biopharmaceuticals. *Nature biotechnology*, 27, 1 (2009), 26–34. doi:10.1038/nbt0109-26.

163. SÁNCHEZ-RODRÍGUEZ, S. P.; MÜNCH-ANGUIANO, L.; ECHEVERRÍA, O.; VÁZQUEZ-NIN, G.; MORA-PALE, M.; DORDICK, J. S.; AND BUSTOS-JAIMES, I., 2012. Human parvovirus B19 virus-like particles: In vitro assembly and stability. *Biochimie*, 94, 3 (2012), 870–878. doi:10.1016/j.biochi.2011.12.006.

164. SCHEIN, C. H. AND NOTEBORN, M. H. M., 1988. Formation of Soluble Recombinant Proteins in Escherichia Coli is Favored by Lower Growth Temperature. *Bio/Technology*, 6, 3 (Mar. 1988), 291–294. doi:10.1038/nbt0388-291.

165. SEJERGAARD, L.; KRARUP, J. K.; KARKOV, H. S.; BAGGE HAGEL, A. B.; AND CRAMER, S. M., 2014. Model based process development for the purification of a modified human growth hormone using multimodal chromatography. *Biotechnology Progress*, 30, 5 (2014), 1057–1064. doi:10.1002/btpr.1923.

166. SHALLIKER, R. A.; KAYILLO, S.; AND DENNIS, G. R., 2008. Optimizing Chromatographic Separation: An Experiment Using an HPLC Simulator. *Journal of Chemical Education*, 85, 9 (Sep. 2008), 1265. doi:10.1021/ed085p1265.

167. SHCHUKINA, O.; ZATIRAKHA, A.; SMOLENKOV, A.; NESTERENKO, P.; AND SHPIGUN, O., 2015. Anion exchangers with branched functional ion exchange layers of different hydrophilicity for ion chromatography. *Journal of Chromatography A*, (2015). doi:10.1016/j.chroma.2015.06.039.

168. SHI, L.; SANYAL, G.; NI, A.; LUO, Z.; DOSHNA, S.; WANG, B.; GRAHAM, T. L.; WANG, N.; AND VOLKIN, D. B., 2005. Stabilization of human papillomavirus virus-like particles by non-ionic surfactants. *Journal of Pharmaceutical Sciences*, 94, 7 (2005), 1538–1551. doi:10.1002/jps.20377.

169. SHUKLA, A. a.; BAE, S. S.; MOORE, J. a.; BARNTHOUSE, K. a.; AND CRAMER, S. M., 1998. Synthesis and Characterization of High-Affinity, Low Molecular Weight Displacers for Cation-Exchange Chromatography. *Industrial & Engineering Chemistry Research*, 37, 10 (1998), 4090–4098. doi:10.1021/ie9801756.

170. SHUKLA, A. a.; BARNTHOUSE, K. a.; BAE, S. S.; MOORE, J. a.; AND CRAMER, S. M., 1998. Structural characteristics of low-molecular-mass displacers for cation-exchange chromatography. *Journal of Chromatography A*, 814, 1-2 (Dec. 1998), 83–95. doi:10.1016/S0021-9673(98)00374-4.

171. SHUKLA, A. a.; HUBBARD, B.; TRESSEL, T.; GUHAN, S.; AND LOW, D., 2007. Downstream processing of monoclonal antibodies-Application of platform approaches. *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences*, 848, 1 (Mar. 2007), 28–39. doi:10.1016/j.jchromb.2006.09.026.

172. SMOOT, L. M.; SMOOT, J. C.; GRAHAM, M. R.; SOMERVILLE, G. a.; STURDEVANT, D. E.; MIGLIACCIO, C. A. L.; SYLVA, G. L.; AND MUSSER, J. M., 2001. Global differential gene expression in response to growth temperature alteration in group A Streptococcus. *Proceedings of the National Academy of Sciences*, 98, 18 (Aug. 2001), 10416–10421. doi:10.1073/pnas.191267598.

173. STANCIK, L. M.; STANCIK, D. M.; SCHMIDT, B.; BARNHART, M.; YONCHEVA, Y. N.; JOAN, L.; BARNHART, D. M.; AND SLONCZEWSKI, J. L., 2002. pH-Dependent Expression of Periplasmic Proteins and Amino Acid Catabolism in Escherichia coli pH-Dependent Expression of Periplasmic Proteins and Amino Acid Catabolism in Escherichia coli. *J Bacteriol*, 184 (2002), 4246–4258. doi:10.1128/JB.184.15.4246.

174. STARCK, J.; KÄLLENIUS, G.; MARKLUND, B.-I.; ANDERSSON, D. I.; AND AKERLUND, T., 2004. Comparative proteome analysis of Mycobacterium tuberculosis grown under aerobic and anaerobic conditions. *Microbiology*, 150, Pt. 11 (2004), 3821–3829. doi:10.1099/mic.0.27284-0.

175. STONE, D. C., 2007. Teaching Chromatography Using Virtual Laboratory Exercises. *Journal of Chemical Education*, 84, 9 (Sep. 2007), 1488–1495. doi:10.1021/ed084p1488.

176. SUSANTO, A.; KNIEPS-GRÜNHAGEN, E.; VON LIERES, E.; AND HUBBUCH, J., 2008. High Throughput Screening for the Design and Optimization of Chromatographic Processes: Assessment of Model Parameter Determination from High Throughput Compatible Data. *Chemical Engineering & Technology*, 31, 6 (2008), 1846–1855. doi:10.1002/ceat.200800457.

177. SVEDBERG, U., 1976. Numerical solution of multicolumn adsorption processes under periodic countercurrent operation. *Chemical Engineering Science*, 31, 5 (Jan. 1976), 345–353. doi:10.1016/0009-2509(76)80003-6.

178. TATÁROVÁ, I.; FÁBER, R.; DENOYEL, R.; AND POLAKOVIČ, M., 2009. Characterization of pore structure of a strong anion-exchange membrane adsorbent under different buffer and salt concentration conditions. *Journal of Chromatography A*, 1216, 6 (2009), 941–947. doi:10.1016/j.chroma.2008.12.018.

179. TEJEDA-MANSIR, A.; MONTESINOS, R. M.; AND GUZMÁN, R., 2001. Mathematical analysis of frontal affinity chromatography in particle and membrane configurations. *Journal of Biochemical and Biophysical Methods*, 49, 1-3 (2001), 1–28. doi:10.1016/ S0165-022X(01)00196-8.

180. TELEN, D.; LOGIST, F.; VAN DERLINDEN, E.; TACK, I.; AND VAN IMPE, J., 2012. Optimal experiment design for dynamic bioprocesses: A multi-objective approach. *Chemical Engineering Science*, 78 (Aug. 2012), 82–97. doi:10.1016/j.ces.2012.05.002.

181. THARAKAN, J. AND BELIZAIRE, M., 1995. Ligand efficiency in axial and radial flow immunoaffinity chromatography of factor IX. *Journal of Chromatography A*, 702, 1-2 (1995), 191–196. doi:10.1016/0021-9673(94)01141-Z.

182. TSOTSAS, E. AND SCHLÜNDER, E.-U., 1989. Strömungsungleichverteilung und axiale dispersion in festbetten. *Chemie Ingenieur Technik*, 61, 3 (1989), 243–246. doi: 10.1002/cite.330610313.

183. UCINSKI, D., 2004. *Optimal measurement methods for distributed parameter system identification*. CRC Press LLC. ISBN 9780203026786.

184. UNIVERSITY OF OHIO. Chromulator. http://www.ohio.edu/people/gu/CHROM. Accessed May 2015.

185. URBAS, L.; JARC, B. L.; BARUT, M.; ZOCHOWSKA, M.; CHROBOCZEK, J.; PIHLAR, B.; AND SZOLAJSKA, E., 2011. Purification of recombinant adenovirus type 3 dodecahedric virus-like particles for biomedical applications using short monolithic columns. *Journal of Chromatography A*, 1218, 17 (2011), 2451–2459. doi: 10.1016/j.chroma.2011.01.032.

186. VAN BEIJEREN, P.; KREIS, P.; AND ZEINER, T., 2012. Ion exchange membrane adsorption of bovine serum albumin-Impact of operating and buffer conditions on breakthrough curves. *Journal of Membrane Science*, 415-416 (2012), 568–576. doi: 10.1016/j.memsci.2012.05.051.

187. VAN BEIJEREN, P.; KREIS, P.; AND ZEINER, T., 2013. Development of a generic process model for membrane adsorption. *Computers & Chemical Engineering*, 53 (2013), 86–101. doi:10.1016/j.compchemeng.2013.03.005.

188. VAN DEEMTER, J.; ZUIDERWEG, F.; AND KLINKENBERG, A., 1995. Longitudinal diffusion and resistance to mass transfer as causes of nonideality in chromatography. *Chemical Engineering Science*, 50, 24 (Dec. 1995), 3869–3882. doi: 10.1016/0009-2509(96)81813-6.

189. VAN GENUCHTEN, M. T. AND PARKER, J. C., 1984. Boundary Conditions for Displacement Experiments through Short Laboratory Soil Columns1. *Soil Science Society of America Journal*, 48, 4 (1984), 703. doi:10.2136/sssaj1984.03615995004800040002x.

190. VAN LIEROP, M. J. C.; WIJNANDS, F.; LOKE, Y. W.; EMMER, P. M.; LUKASSEN, H. G. M.; BRAAT, D. D. M.; VAN DER MEER, A.; MOSSELMAN, S.; AND JOOSTEN, I., 2002. Detection of HLA-G by a specific sandwich ELISA using monoclonal antibodies G233 and 56B. *Molecular human reproduction*, 8, 8 (2002), 776–784. doi:10.1093/molehr/8.8.776.

191. VÁZQUEZ-REY, M. AND LANG, D. A., 2011. Aggregates in monoclonal antibody manufacturing processes. *Biotechnology and Bioengineering*, 108, 7 (2011), 1494–1508. doi:10.1002/bit.23155.

192. VICENTE, T.; FÁBER, R.; ALVES, P. M.; CARRONDO, M. J. T.; AND MOTA, J. P. B., 2011. Impact of ligand density on the optimization of ion-exchange membrane chromatography for viral vector purification. *Biotechnology and Bioengineering*, 108, 6 (2011), 1347–1359. doi:10.1002/bit.23058.

193. VICENTE, T.; PEIXOTO, C.; CARRONDO, M. J. T.; AND ALVES, P. M., 2009. Purification of recombinant baculoviruses for gene therapy using membrane processes. *Gene therapy*, 16, 6 (Jun. 2009), 766–775. doi:10.1038/gt.2009.33.

194. VICENTE, T.; SOUSA, M. F. Q.; PEIXOTO, C.; MOTA, J. P. B.; ALVES, P. M.; AND CARRONDO, M. J. T., 2008. Anion-exchange membrane chromatography for purification of rotavirus-like particles. *Journal of Membrane Science*, 311, 1-2 (2008), 270–283. doi:10.1016/j.memsci.2007.12.021.

195. VOITL, A.; MÜLLER-SPÄTH, T.; AND MORBIDELLI, M., 2010. Application of mixed mode resins for the purification of antibodies. *Journal of Chromatography A*, 1217, 37 (2010), 5753–5760. doi:10.1016/j.chroma.2010.06.047.

196. VON LIERES, E. AND ANDERSSON, J., 2010. A fast and accurate solver for the general rate model of column liquid chromatography. *Computers and Chemical Engineering*, 34, 8 (2010), 1180–1191. doi:10.1016/j.compchemeng.2010.03.008.

197. WALL, M., 1996. GAlib: A C++ library of genetic algorithm components. http://lancet.mit.edu/ga/.

198. WALTER, E. AND PRONZATO, L., 1997. *Identification of Parametric Models: from Experimental Data*. Springer-Verlag. ISBN 3540761195.

199. WANG, J.; ZHAO, C.; MENG, B.; XIE, J.; ZHOU, C.; CHEN, X.; ZHAO, K.; SHAO, J.; XUE, Y.; XU, N.; MA, Y.; AND LIU, S., 2007. The proteomic alterations of Thermoanaerobacter tengcongensis cultured at different temperatures. *Proteomics*, 7, 9 (2007), 1409–1419. doi:10.1002/pmic.200500226.

200. WANKAT, P. C., 2006. Using a commercial simulator to teach sorption separations. *Chemical Engineering Education*, 40, 3 (2006), 165.

201. WARIKOO, V.; GODAWAT, R.; BROWER, K.; JAIN, S.; CUMMINGS, D.; SIMONS, E.; JOHNSON, T.; WALTHER, J.; YU, M.; WRIGHT, B.; MCLARTY, J.; KAREY, K. P.; HWANG, C.; ZHOU, W.; RISKE, F.; AND KONSTANTINOV, K., 2012. Integrated continuous production of recombinant therapeutic proteins. *Biotechnology and Bioengineering*, 109, 12 (2012), 3018–3029. doi:10.1002/bit.24584.

202. WEINER, J.; ZIMMERMAN, C. U.; GÖHLMANN, H. W. H.; AND HERRMANN, R., 2003. Transcription profiles of the bacterium Mycoplasma pneumoniae grown at different temperatures. *Nucleic Acids Research*, 31, 21 (Nov. 2003), 6306–6320. doi:10.1093/nar/gkg841.

203. WESTERBERG, K.; BROBERG-HANSEN, E.; SEJERGAARD, L.; AND NILSSON, B., 2013. Model-based risk analysis of coupled process steps. *Biotechnology and Bioengineering*, 110, 9 (2013), 2462–2470. doi:10.1002/bit.24909.

204. WESTERBERG, K.; DEGERMAN, M.; AND NILSSON, B., 2010. Pooling control in variable preparative chromatography processes. *Bioprocess and Biosystems Engineering*, 33, 3 (2010), 375–382. doi:10.1007/s00449-009-0335-8.

205. WILLOUGHBY, N., 2009. Too big to bind? Will the purification of large and complex therapeutic targets spell the beginning of the end for column chromatography? *Journal of Chemical Technology and Biotechnology*, 84, 2 (2009), 145–150. doi:10.1002/jctb.2020.

206. WOO, P. C. Y.; LAU, S. K. P.; TSE, H.; TENG, J. L. L.; CURREEM, S. O. T.; TSANG, A. K. L.; FAN, R. Y. Y.; WONG, G. K. M.; HUANG, Y.; LOMAN, N. J.; SNYDER, L. A. S.; CAI, J. J.; HUANG, J.-D.; MAK, W.; PALLEN, M. J.; LOK, S.; AND YUEN, K.-Y., 2009. The Complete Genome and Proteome of *Laribacter hongkongensis* Reveal Potential Mechanisms for Adaptations to Different Temperatures and Habitats. *PLoS Genet*, 5, 3 (2009). doi:10.1371/journal.pgen.1000416.

207. XU, J.; ZHU, L.; XU, G.; YU, W.; AND RAY, A. K., 2013. Determination of competitive adsorption isotherm of enantiomers on preparative chromatographic columns using inverse method. *Journal of Chromatography A*, 1273 (2013), 49–56. doi:10.1016/j.chroma.2012.11.052.

208. YAMAMOTO, S., 2005. Electrostatic interaction chromatography process for protein separations: Impact of engineering analysis of biorecognition mechanism on process optimization. *Chemical Engineering and Technology*, 28, 11 (2005), 1387–1393. doi:10.1002/ceat.200500199.

209. YANG, D.-G.; CHUNG, Y.-C.; LAI, Y.-K.; LAI, C.-W.; LIU, H.-J.; AND HU, Y.-C., 2007. Avian influenza virus hemagglutinin display on baculovirus envelope: cytoplasmic domain affects virus properties and vaccine potential. *Molecular therapy: the journal of the American Society of Gene Therapy*, 15, 5 (2007), 989–996. doi:10.1038/sj.mt.6300236.

210. YANG, Y.; MENGRAN YU; ZHANG, S.; MA, G.; AND SU, Z., 2015. Adsorption of virus-like particles on ion exchange surface: Conformational changes at different pH detected by dual polarization interferometry. *Journal of Chromatography A*, (2015). doi:10.1016/j.chroma.2015.07.019.

211. Yu, M.; Li, Y.; Zhang, S.; Li, X.; Yang, Y.; Chen, Y.; Ma, G.; and Su, Z., 2014. Improving stability of virus-like particles by ion-exchange chromatographic supports with large pore size: Advantages of gigaporous media beyond enhanced binding capacity. *Journal of Chromatography A*, 1331 (2014), 69–79. doi:10.1016/j.chroma.2014.01.027.