ADVANCING DOWNSTREAM PROCESS DEVELOPMENT

MECHANISTIC MODELING AND ARTIFICIAL INTELLIGENCE

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
M.Sc. Gang Wang
aus Zhejiang (China)

Danksagung

In den letzten Jahren, die dieser wissenschaftlichen Arbeit gewidmet waren, standen viele Menschen an meiner Seite, die mich fachlich und menschlich entscheidend beeinflusst haben. Meinen besonderen Dank möchte ich folgenden Personen ausdrücken.

Prof. Jürgen Hubbuch danke ich für die herzliche Aufnahme in seine Gruppe, die Einführung in die wissenschaftliche Community, das entgegengebrachte Vertrauen, die vielen Freiheiten und Möglichkeiten, sich weiterzuentwickeln. Prof. Matthias Franzreb möchte ich für das Interesse an meiner Arbeit sowie die Übernahme des Korreferats danken.

Der Europäischen Union danke ich für die Finanzierung meiner Forschung im Rahmen des Projektes "Horizon 2020".

Der Arbeitsgruppe des MABs danke ich zutiefst für die Zusammenarbeit, die fachlichen Diskussionen auf einem sehr hohen Niveau, die Ehrlichkeit und Offenheit, sowie den emotionalen Beistand.

Besonders bedanke ich mich bei meinen fleißigen studentischen Helfern: Till Briskot, Roxana Disela, Christian Fischer, Fabian Görlich und Alexandros Papadopoulos. Margret Meixner, Marion Krenz, Iris Perner-Nochta und Michael Wörner für die administrative Unterstützung. Unserem Ingenieurteam Susanna Suhm, Cathrin Dürr, Nicolai Bluthardt, Stefanie Limbrunner, Kristina Schleining und Birgit Roser für die unschlagbare Labororganisation. Tobias Hahn, Pascal Baumann, Frank Hämmerling und Josefine Morgenstern waren für mich große Vorbilder. Es war für mich eine große Ehre, mit euch arbeiten zu dürfen. Steffen Großhans, Josefine Morgenstern, Matthias Rüdt, Johannes Winderl danke ich für die geglückten und auch die weniger geglückten gemeinsamen Projekte. Meinem Bürokollegen Johannes Winderl danke ich für die perfekte Arbeitsatmosphäre, sowie die fachliche, weltpolitische und wirtschaftwissenschaftliche Diskussionen und Anstöße. Susanna Suhm, Steffen Großhans, Anna Wöll, Johannes Winderl und Sarah Gretzinger für die lustigen und lockeren Momente.

Ich möchte meinen Eltern Aihong und Wei danken für eure unendliche Unterstützung, für eure Liebe und Vertrauen.

Bei meiner Frau Yingying und den Kleinen - Vincent und Mila, kann ich mich gar nicht genug bedanken für euren Rückhalt, Liebe und Toleranz gegenüber meiner etwas miserablen Work-Life-Balance.

"Until I began to learn to draw,
I was never much interested in looking at art."

- Richard P. Feynman

Zusammenfassung

Die aktuellen technologischen Fortschritte treffen auf globale Trends wie Bevölkerungswachstum, steigende Lebensstandards und erhöhte Lebenserwartung, weswegen die Nachfrage nach Biopharmazeutika höher denn je ist. Als eine der am schnellsten wachsenden Branchen konzentriert sich die biotechnologische Industrie auf die Entwicklung und Herstellung sogenannter Biologika, also neuartiger biologischer Moleküle, welche dazu beitragen, Krankheiten zu heilen und damit Leben zu erhalten. Einer der wichtigsten Schritte in der Herstellung von Biologika ist das sogenannte Downstream-Processing (DSP), das darauf abzielt, einen Wirkstoff biologischen Ursprungs aufzureinigen. Da ein Großteil der gesamten Entwicklungs- und Produktionskosten auf das DSP zurückzuführen ist, steht dieser Bereich unter besonderem Preisdruck. Zahlreiche Ereignisse aus jüngster Vergangenheit und absehbarer Zukunft wie z.B. der Aufstieg von Biosimilars aufgrund des Patentablaufs von Blockbuster Biopharmazeutika, der steigende Titer im Upstream-Processing (Stammkonstruktion, -screening und -kultivierung), die Konstruktion neuartiger, strukturell hoch komplexer Moleküle und die Kürzung des öffentlichen Gesundheitshaushalts verlangen bahnbrechende, kosteneffiziente Technologien. Die Entwicklung eines neuen Medikaments erfordert aktuell eine Vorlaufzeit von bis zu zehn Jahren, und droht bis zur endgültigen Genehmigung durch Arzneimittelbehörde (mit einer Wahrscheinlichkeit von ca. $90\,\%$) zu scheitern. Es ist im Interesse der Patienten, der Gesundheitsbehörden und der Hersteller, die Entwicklungszeit von Biopharmazeutika zu verkürzen.

Derzeit wird die Entwicklung von Aufreinigungsprozessen hauptsächlich auf Basis von Plattformprozessen und Hochdurchsatzexperimentieren (HTE) durchgeführt. Als Plattformprozesse werden Prozesse bezeichnet, welche aus einer starren Abfolge von Operationseinheiten bestehen, die zur Aufreinigung von monoklonalen Antikörpern (mAbs) mit ähnlichen chemischen Eigenschaften entwickelt wurden. Plattformprozesse müssen für die Aufreinigung eines bestimmten mAb nur geringfügig angepasst werden. Sie stellen jedoch einen Kompromiss zwischen der Verringerung des Entwicklungsaufwands und der Einschränkung der Freiheitsgrade in der Prozessoptimierung dar. Zur Aufreinigung neuartiger Biologika, wie z.B. nicht standardmäßiger Antikörper, konjugierter Proteine und virusähnlicher Partikeln, ist die starre Sequenz von Operationseinheiten in Plattformprozessen ungenügend. Im aktuellen Arbeitsablauf der Industrie wurden Hochdurchsatztechnologien in Kombination mit dem Konzept *Design-of-Experiments* (DoE) eingesetzt, um einen größeren Spielraum durch Miniaturisierung, Parallelisierung und Automatisierung der Experimente abzudecken. Aufgrund der empirischen Natur von HTE ist ein systematisches Screening möglicher Lösungskandidaten und eine vergleichsweise hohe Probenmenge erforderlich. Daher sind Ansätze basierend auf HTE häufig nicht in der frühen Phase der Medikamentenentwicklung

einsetzbar. Der Ansatz der mechanistischen Modellierung im DSP nutzt Prozesswissen hergeleitet aus physikalischen Grundlagen und bietet die Möglichkeit der Material- und Zeiteinsparnis. Die üblicherweise verwendeten mechanistischen Modelle sind Differential-gleichungen, die unter Berücksichtigung von Massenbilanzen und/oder Grundsätzen der Thermodynamik hergeleitet werden, und beschreiben die Fluiddynamik und andere beob-achtete Phänomene. Somit entspricht der Ansatz der mechanistischen Modellierung der Forderung der *Quality-by-Design* (QbD) Initiative der US-amerikanischen *Food and Drug Administration* (FDA), dass die Entwicklung und Herstellung von Medikamenten durch wissenschaftliche Erkenntnisse unterstützt werden sollen. Als Folge findet kommerzialisierte Simulationssoftware, welche die komplizierte Berechnung der Modelle durchführt, immer mehr Anwendungen in der biopharmazeutischen Industrie.

Die Hauptaufgaben der mechanistischen Modellierung bilden die Bereiche Modellbildung, Modellkalibrierung und Modellanwendung. In allen drei Bereichen verhindern bestehende Problemstellungen einerseits den Einsatz modellbasierter Ansätze und andererseits die Entwicklung und Umsetzung neuer Anwendungen. In Bezug auf die Modellbildung ist die *Steric-Mass-Action* (SMA) Isotherme ein Beispiel für ein erfolgreiches Modell, welches zu einem Konsens in Wissenschaft und Industrie über die Fähigkeit der mechanistischen Modellierung zur Beschreibung von Ionenaustauschchromatographie (IEX) geführt hat. Zur Abdeckung anderer wichtiger Proteinaufreinigungsprozesse, wie der hydrophoben Interakti-onschromatographie (HIC) oder Polyethylenglycol (PEG) induzierter Präzipitation, fehlen jedoch weitgehend akzeptierte Modelle, die ausreichende Genauigkeit bieten. Im Bereich der Modellkalibrierung ist die gängige Praxis zur Bestimmung unbekannter Modellparameter die Durchführung analytischer Kalkulationen oder numerischer Methoden wie z.B. der Ausgleichungsrechnung. Der Nachteil dieser traditionellen Methoden besteht darin, dass diese zeitaufwändig sind und für im experimentellen System neuauftretende Komponenten wiederholt durchgeführt werden müssen. Kalibrierte mechanistische Modelle enthalten eine große Menge an Information, welche jedoch bisher kaum auf andere Szenarien als die Prozessentwicklung und Robustheitsanalyse angewendet wurden. Neue Strategien werden benötigt, um mechanistische Modelle im vollen Umfang nutzen zu können.

Das Ziel der vorliegenden Arbeit war es, Probleme aus diesen drei Bereichen zu lö-sen. Im Detail sollte die Menge der mechanistischen Modelle um diejenigen für HIC-Prozesse und PEG-induzierte Präzipitation ergänzt werden; Es sollte eine Quasi-Echtzeit-Modellkalibrierungsmethode entwickelt werden, mit deren Hilfe Modellparameter für im experimentellen System neuauftretende Komponenten ohne zusätzlichen Aufwand bestimmt werden können. Neue Anwendungsszenarien sollten erforscht werden, indem Informationen zugänglich gemacht werden, die in mechanistischen Modellen verborgen vorliegen. Diese Arbeit umfasst sechs Publikationen bzw. Manuskripte, die sich mit verschiedenen Aspek-ten der Modellbildung, Modellkalibrierung und Modellanwendung in der mechanistischen Modellierung befassen.

In der ersten Veröffentlichung wurde ein mechanistisches Modell für HIC unter Berück-sichtigung des Gleichgewichts zwischen strukturierten und weniger geordneten Wasser-molekülen auf hydrophoben Oberflächen von Proteinen und Liganden in der stationären Phase hergeleitet. Während der Herleitung der Gleichungen wurden Annahmen getrof-

fen und Vereinfachungen angewendet, um die Verwendbarkeit mit der Genauigkeit und Zuverlässigkeit der Modellantwort in Einklang zu bringen. Die Fähigkeit des Modells, Prozesse in Chromatographiesäulen zu beschreiben, wurde unter Verwendung von Proteinen mit unterschiedlichen Merkmalen bestätigt. Nach bestem Wissen des Autors wurden erstmals kinetische Experimente mit linearen und stufenweisen Salzgradienten und Batch-Chromatographie-Experimente in HIC mit weitgehend identischen Modellparametern beschrieben und exakt vorhergesagt. Diese neue Adsorptionsisotherme ist nicht nur von akademischem Interesse, sondern bietet auch neue Möglichkeiten für die Entwicklung von HIC-Prozessen in der pharmazeutischen Industrie, da sie den derzeit gängigen Ansatz basierend auf Faustregeln und HTE ersetzen kann.

Im zweiten Manuskript wurde ein mechanistisches Modell für die PEG-induzierte Proteinpräzipitation vorgeschlagen, wobei die Ähnlichkeiten zwischen Proteinpräzipitation und HIC berücksichtigt wurden. Es wurde dabei die Vorstellung angenommen, dass die Bildung von Protein-Protein-Grenzflächen hauptsächlich durch den hydrophoben Effekt wie in HIC, also die Reorientierung der strukturierten und weniger geordneten Wasserstruktur auf hydrophoben Bereichen der Proteinoberflächen getrieben ist. Es wurden Hochdurchsatz-Präzipitationsexperimente durchgeführt, um Präzipitationsdaten von Proteinen unterschiedlicher Größen zu erzeugen. Nach der Kalibrierung konnte das Modell neue Prozesse präzise vorhersagen. Der Hauptunterschied zwischen dem vorgeschlagenen Modell und weit verbreiteten Modellen, wie der Cohn-Gleichung und den daraus hergeleiteten Modellen, ist die Möglichkeit, den Einfluss der PEG- und Proteinkonzentration auf die PEG-induzierte Proteinpräzipitation zu beschreiben und das Verhalten kleiner Proteine vorherzusagen.

In der dritten Publikation wurde die einfachste Form der künstlichen Intelligenz (KI) – das künstliche neuronale Netzwerk (KNN) – verwendet, um eine Methode zur Kalibrierung von mechanistischen Chromatographiemodellen zu entwickeln. Die systematisch durchgeführten *in silico* Experimente durch Variation der Zielmodellparameter machten die in einem unkalibrierten mechanistischen Modell verborgene Information zugänglich. Die resultierenden Datensätze bestehend aus simulierten Chromatogrammen und Modellparametern wurden verwendet, um ein KNN-Modell zu konstruieren, das zur simultanen Bestimmung der unbekannten Parameter aus dem Transport-Dispersiven Modell (TDM) und dem stöchiometrischen Verdrägungsmodell für die Trennung einer tertiären Proteinlösung in Kationenaustauschchromatographie (CEX) verwendet wurde. Das trainierte neuronale Netzwerk wurde durch Kreuzvalidierung verifiziert und das kalibrierte mechanistische Modell wurde durch drei Salzgradienten-Experimente erfolgreich validiert. Diese neuartige Methode bietet die Möglichkeit, Modellparameter für eine im Chromatographiesystem neuauftretende Komponente zu bestimmen, wo zeitaufwändige traditionelle Ansätze wiederholt durchgeführt werden müssen. Da der Rechenaufwand des KNN extrem gering ist (mit einer Berechnungszeit von wenigen Millisekunden im untersuchten Fall), kann es für eine automatisierte In-Prozesskalibrierung während chromatographischer Experimente implementiert werden. Darüber hinaus hat diese Methode das Potenzial, für die mechanistische Modellierung in anderen Forschungsbereichen zu Kalibrierungszwecken eingesetzt zu werden.

Während sich die bisherigen Publikationen und Manuskripte auf die notwendigen Voraussetzungen der Modellierung (Modellbildung und Modellkalibrierung) konzentrieren, untersuchen die Folgenden mögliche Anwendungsszenarien der mechanistischen Modellierung. Basierend auf mechanistischen Chromatographiemodellen wurde in der vierten Veröffentlichung eine gründliche Untersuchung der Stofftransport- und Adsorptionsmechanismen von PEGylierten Proteinen in CEX durchgeführt. Diese Studie verwendete eine Kombination aus dem umfassenden *General Rate* Modell (GRM) und der SMA Isothermen, um Informationen über Proteindiffusion durch die Filmschicht um die Adsorberpartikeln, Diffusion innerhalb der Partikelporen sowie Adsorptions-/Desorptionsverhalten der PEGylierten Spezies im linearen und nichtlinearen Bereich der Isothermen zu erhalten. Anomalien wie Peakhöhen und -breiten, die trotz eines zunehmenden hydrodynamischen Radius der PEGylierten Spezies konstant bleiben, und die nichtlineare Korrelation zwischen der Bindekapazität und der Molekülmasse der PEGylierten Proteine konnten jeweils durch den vorgeschlagenen Diffusions-Desorptions-Kompensationseffekt und die berechnete exponentielle Abhängigkeit des Abschirmparameters von der Molekülmasse erklärt werden.

In der fünften Veröffentlichung wurde die mechanistische Chromatographiemodellierung mit HTE und DoE kombiniert, um einen integrierten Prozess zu entwickeln, der aus PEG-induzierter Präzipitation, selektiver Rücklösung und CEX für die Aufreinigung eines industriellen mAb aus geernteter Zellkulturflüssigkeit besteht. Die Prozessentwicklungsstrategie kombinierte Vorteile von HTE, DoE und mechanistischer Modellierung, was zu einem reduzierten Probenverbrauch und minimalem experimentellen Aufwand führte. In Bezug auf die Ausbeute, Reinheit und Produktionsrate von mAb-Monomer sowie die Reduzierung von Hauptkontaminanten, wie DNA, Wirtszellproteinen und mAb-Aggregaten, stellt der integrierte Prozess eine geeignete Alternative zum konstenintensiven Industriestandard der Protein A Chromatographie dar. Darüber hinaus wurde das kalibrierte und validierte mechanistische Modell auf eine periodische Gegenstromchromatographie mit drei CEX-Säulen übertragen, um die Möglichkeiten und das Potenzial der kontinuierlichen Multi-Säulenchromatographie zu untersuchen.

In der sechsten und letzten Veröffentlichung wurde eine neue Methode zur Fehler-Ursachen-Analyse von Abweichungen in der Proteinchromatographie entwickelt, die auf mechanistischer Chromatographiemodellierung und KNN basierte. Prozessvariationen in der Proteinchromatographie, insbesondere die Abweichung der Bindekapazität aufgrund von Säulenalterung, täglichem Betrieb oder Säulenmaterialaustausch, können enorme Auswirkung auf die kritischen Qualitätsmerkmale des pharmazeutischen Produkts haben. Simulationen potenzieller Prozessabweichungen auf Grundlage eines zuvor kalibrierten mechanistischen Chromatographiemodells wurden als Daten für das Training, die Validierung und das Testen des KNN-Modells erstellt. Experimentelle Fehler, die zu einem sehr ähnlichen Erscheinungsbild führten, hier Fehler in der ionischen Kapazität und Salzgradientenlänge, wurden absichtlich eingeführt, um abweichende Chromatogramme eines Tertiärtrennproblems zu erzeugen. Das trainierte KNN konnte die Abweichungen erkennen, ihre Ursachen identifizieren und innerhalb von Millisekunden ihre Größen ausgeben. Die ionische Kapazität, welche einen bisher nicht während des Betriebs messbaren Parameter darstellt, konnte mit hoher Genauigkeit bestimmt werden. Die grundlegende Idee, ein

mechanistisches Modell und ein KNN zu kombinieren, um von ihren jeweiligen Vorteilen zu profitieren, konnte erfolgreich umgesetzt werden. Folglich wurden nur wenige Experimente für die Kalibrierung des mechanistischen Modells benötigt, wodurch der hohe experimentelle Aufwand, der sonst für das Training des KNN-Modells benötigt wäre, vermindert wurde.

Insgesamt trägt diese Arbeit zu den aktuellen Trends in DSP bei, die von empirisch getriebenen Methoden (HTE und DoE) zu Anwendungen auf Grundlage von mechanistischem Verständnis und künstlicher Intelligenz führen. Die Arbeit befasst sich nicht nur mit Fragen der Modellverfügbarkeit in wichtigen Reinigungstechniken wie HIC und PEG-induzierter Proteinpräzipitation und ermöglicht dort eine modellbasierte Prozessentwicklung, sondern schlägt auch eine KNN-basierte Kalibrierungsmethode vor, um die Schwächen herkömmlicher Methoden zu überwinden. Darüber hinaus werden verschiedene Anwendungsszenarien mechanistischer Modelle gezeigt, wie z.B. das Erlangen des grundlegenden Verständnisses für die Entwicklung von Aufreinigungsprozessen einer nicht-typischen Molekülklasse, die Entwicklung eines integrierten Prozesses als kosteneffizientere Alternative zum standardmäßigen Protein A Prozess und die Verwirklichung einer Methode zur Fehler-Ursachen-Analyse von Abweichungen in der Proteinchromatographie für Anwendungen in Quasi-Echtzeit. Langfristig könnte die Kombination aus mechanistischer Modellierung und KNN ein fundamentaler Baustein eines künstlichen Prozessexperten für DSP sein, der eine vollautomatisierte Strategie in Quasi-Echtzeit die Entwicklung, Überwachung und Rückkopplungskontrolle von Aufreinigungsprozessen darstellt.

Abstract

The current technological advancements encounter the population growth, increasing living standard and life expectancy around the globe, opening a rapidly expanding biopharmaceutical market with high demand. As one of the fastest growing industries, the biotechnology industry focuses on development and manufacturing of novel biological molecules – the biologics – and strives literally to save lives. One of the key steps in the manufacturing of biologics is the so-called downstream processing (DSP), which aims to purify an active ingredient – here a target molecule of biological origin. DSP represents the biggest share of the total expenditure and is faced with price pressure resulting from numerous events in the recent past and near future. Among them, the rise of biosimilars due to patent expiration of blockbuster biopharmaceuticals, the increasing titer in upstream processing – USP (strain construction, strain screening and cultivation), the invention of novel, structurally diverse molecules, and cuts in public healthcare budget demand disruptive, more cost-efficient technologies. Furthermore, the development of a new drug requires a lead time of up to a decade, and is constantly threatened by an overwhelming likelihood of failure (approximately 90 %) until its final approval. It is in interest of all parties involved, *i.e.* the patients, the health authorities, and the manufacturers, to shorten the drug development timeline.

Currently, the development of purification processes is mainly performed based on platform processes and high-throughput experimentation (HTE). The so-called platform processes are processes consisting of fixed order of unit operations, once developed in DSP for the purification of standard monoclonal antibody (mAb). It provides a trade-off between the reduction of development efforts and limitation of degrees of freedom to be optimized for a specific molecule. To purify novel biologics, such as non-standard antibodies, conjugated proteins, or virus-like particles, the fixed sequence of unit operations in platform processes is not sufficient. In the current industrial work-flow, high-throughput technologies combined with Design-of-Experiments (DoE) have been introduced to cover a larger design space by miniaturizing, parallelizing, and automatizing the experiments to a certain extent. Due to its empirical nature, high-throughput experimentation (HTE) is a brute-force method, *i.e.* a problem-solving technique that requires a systematical screening of possible candidate solutions, and needs a comparably high amount of sample, that is often not available in the early stage of drug development.

The approach of mechanistic modeling takes advantage of knowledge about the physical fundamentals of the processes and offers advantage of both material and time saving in DSP. Mechanistic models commonly used are, in most cases, differential equations derived from consideration of mass balances and/or laws of thermodynamics, and describe the fluid dynamics and other phenomena observed. Thus, mechanistic modeling is in line with the demand of the Quality by Design (QbD) initiative suggested by the US Food and Drug Administration (FDA), that process development and manufacturing of drugs should be backed up by scientific understanding. As a consequence, commercialized simulation tools are finding more and more applications in the biopharmaceutical industry.

The main tasks of mechanistic modeling can be assigned to the fields of model building, model calibration, and/or model application. In all three fields, existing challenges impede the use of model-based approaches on the one hand, and hinder the development and the implementation of new applications on the other. With respect to model building, an example for a successful model is the steric mass action (SMA) isotherm, which has led to a general consensus about the ability of mechanistic modeling to describe ion-exchange chromatography (IEX) processes in the academic and industrial community. To cover other important protein purification processes, such as hydrophobic interaction chromatography (HIC) or polyethylene glycol (PEG) induced precipitation, however, a widely accepted model with sufficient accuracy is lacking. With respect to model calibration, the common practice to determine unknown model parameters is to perform analytical calculation or numerical methods such as curve fitting. Major disadvantages of these traditional methods are that they are time-consuming and have to be carried out repeatedly for new components in the experimental system. Calibrated mechanistic models contain a large amount of information, but they have been barely applied to other scenarios than the process development and robustness analysis. New strategies are needed to enable the use of mechanistic models to their full extent.

The objective of this thesis was to solve some issues in these three fields. In detail, the set of mechanistic models should be complemented by the ones for HIC processes and PEG-induced precipitation; A near real-time model calibration method should be proposed, which can be used to determine model parameters for new components in the experimental system without additional effort; New application scenarios should be explored by making the information accessible, which is hidden in mechanistic models.

This thesis contains six publications/manuscripts, that focus on different aspects of model building, model calibration, and model application in mechanistic modeling. In the first publication, a mechanistic isotherm model for HIC was derived by considering the equilibrium between well-ordered and bulk-like ordered water molecules on the hydrophobic surfaces of proteins and ligands in the stationary phase. During isotherm derivation, assumptions and simplifications were made to balance the ease of use with accuracy and reliability of model response. The model's capability of describing column chromatography processes was backed up using proteins of different features. To the author's best knowledge, kinetic linear/step-wise salt gradient and batch chromatography experiments in HIC were described and predicted with largely identical model parameters for the very first time. This new adsorption isotherm is not only of academic interest, but also opens up new

possibilities for HIC process development in industry, *i.e.* superseding the current common practice based on rules of thumb and HTE.

In the second manuscript, a new mechanistic model for PEG-induced protein precipitation was proposed by taking the similarities between protein precipitation and HIC into account. The formation of protein-protein interfaces was thought to be mainly driven by the hydrophobic effect as in HIC – the reorganization of the well-ordered and bulk-like water structure on the hydrophobic areas of the protein surfaces. High-throughput precipitation experiments were conducted to generate precipitation data of differently sized proteins. After calibration, the presented model could predict uninvolved data precisely. The main difference between the models widely used, such as the Cohn equation and derivatives, and the proposed model is its capability to describe the influence of both PEG and protein concentration on PEG-induced protein precipitation and to predict the behavior of small-sized proteins.

In the third publication, the most simple form of artificial intelligence (AI) – the artificial neural network (ANN) was employed to develop a method for calibration of mechanistic chromatography models. Performing *in silico* experiments by varying the model parameters of interest systematically made the information hidden in an uncalibrated mechanistic model accessible. The resulting data sets consisting of simulated chromatograms and model parameters were used to construct an ANN model, which was applied for the simultaneous determination of the transport dispersive model (TDM) and stoichiometric displacement model (SDM) parameters for the separation process of a tertiary protein mixture in cation-exchange chromatography (CEX). The trained ANN model was verified by cross validation and the calibrated mechanistic model was successfully validated by three binde-and-elute salt gradient chromatography experiments. This novel method offers the possibility to determine model parameters for a new component in the chromatography system, where time-consuming traditional approaches require to be carried out repeatedly. Because the ANN's computational expense is extremely low, as indicated by the calculation time of only milliseconds in the case investigated, it has the potential to be implemented for automated real-time model calibration during chromatography experiments. Furthermore, this method has the potential to be applied to calibration purposes in mechanistic modeling in other research fields.

While the publications/manuscripts so far concentrate on the necessary preconditions of modeling, *i.e.* model building and model calibration, the following ones explore possible application scenarios of mechanistic modeling. Based on mechanistic chromatography models, a full investigation on the mass transfer and adsorption mechanisms of PEGylated proteins in CEX was carried out in the fourth publication. This study used a combination of the most detailed general rate model (GRM) and SMA to gain information on protein diffusion through the film layer around the adsorbent particles, diffusion inside the particle pore, as well as adsorption/desorption behavior of the PEGylated species in the linear and non-linear region of the isotherm. Anomalies observed, such as peak heights and widths remaining constant in spite of an increasing hydrodynamic radius of PEGylated species

and the non-linear correlation between the column capacity and the MW of PEGylated proteins, could be explained by the proposed diffusion-desorption-compensation effect and the calculated exponential dependency of the shielding parameters on the MW increase, respectively.

In the fifth publication, mechanistic chromatography modeling was combined with HTE and DoE to develop an integrated process consisting of PEG-induced precipitation, selective resolubilization, and CEX chromatography for the isolation of an industrial mAb from harvested cell culture fluid (HCCF). The process development strategy combined advantages of HTE, DoE, and mechanistic modeling, resulting in reduced sample consumption and minimal experimental effort. In terms of yield, purity, and production rate of mAb monomer, as well as the reduction of main contaminants, such as DNA, host cell proteins (HCP), and mAb aggregates, the integrated process represents a feasible alternative to the expensive industrial standard Protein A chromatography. Furthermore, the calibrated and validated mechanistic model was transferred to a three-column CEX periodic counter-current chromatography (3C-PCCC) mode to explore the opportunities and potential offered by multi-column continuous chromatography.

In the sixth and final publication, a novel method for root cause investigation of deviations in protein chromatography was developed based on both mechanistic chromatography modeling and ANN. Process variations in protein chromatography, especially the deviation of capacity due to column aging, day-to-day operations, or column exchange, can have enormous impact on the critical quality attributes (CQA) of the pharmaceutical product. Simulations of potential process deviations based on a beforehand calibrated mechanistic chromatography model were generated as learning material for training, validation, and testing of the ANN model. Experimental errors leading to highly similar observable deviation, here errors in ionic capacity and salt gradient length, were deliberately introduced to generate deviating chromatograms of a tertiary separation problem. The trained ANN could recognize the deviations, identify their root causes, and respond with their magnitudes within milliseconds. The ionic capacity – a hitherto non-observable parameter during operation – could be determined with high accuracy. The fundamental idea to combine a mechanistic model and an ANN to benefit from their respective advantages could be successfully implemented. Consequently, only few experiments were required for the calibration of the mechanistic model, avoiding the high wet-lab experimental effort needed otherwise for the training of the ANN model.

In summary, this thesis contributes to the current trends in DSP moving from empirically driven, state-of-the-art methods (HTE and DoE) towards applications based on mechanistic understanding and artificial intelligence. It not only addresses the issues of model availability in key purification techniques, such as HIC and PEG-induced protein precipitation, to enable model-based process development, but also proposes an ANN-based calibration tool to overcome the weaknesses of traditional methods. Furthermore, diverse application scenarios of mechanistic models are shown, such as providing a thorough *a priori* understanding of a non-typical molecule class for its purification process development, developing an integrated process as a cost-efficient alternative to the standard Protein A process, and

suggesting a combination of mechanistic modeling and ANN to develop a near real-time method for root cause investigation of deviations in protein chromatography. In long term, the combination of mechanistic modeling and ANN could be a fundamental building block of an AI process expert in DSP, representing a fully automated, near real-time strategy for development, monitoring, and feedback control of purification processes.

Contents

Appendix   **163**

1. Introduction

Being in the intersection of biology and technology, the biopharmaceutical industry has changed rapidly in the last decades, providing treatment opportunities for conditions like dementia and cancer. The so-called biologics, such as insulin, human growth hormone, monoclonal antibodies (mAbs) and their derivatives, or virus-like particles, represent one of the currently fastest expanding economic sectors across many different industries [1–3]. From early 1990s to 2015, 47 mAbs have been approved by the US Food and Drug Administration (FDA), or the European Medicines Agency (EMA) – giving an average of two approved mAbs per year [4]. After that period, the average annual approvals of mAbs have been increased by fivefold, as reported in 2017 [5]. At the current time, 52 and 230 mAbs have been reported to be in the phase 3 and phase 2 clinical studies, respectively. By 2020, the global mAb market is expected to reach 125 billion USD [6].

Compared to small-molecule drugs, biologics are highly promising due to their molecular specificity towards cells and receptors. However, due to the accompanied size and complex structure of biopharmaceutics, they are often challenging to be produced synthetically and require to be manufactured in and extracted from biological systems. After expression and folding, the target molecule is accompanied by a huge amount of impurities, such as media components, host-cell proteins, product related molecules, as well as charge variants and glycovariants. These compounds have to be reduced and well-defined in the final drug product to guarantee patient safety and drug efficiency, hence the so-called downstream processing – DSP (cell disruption, clarification, and purification) is of great importance. Since the product attributes are sensitive to slight changes in the manufacturing processes, high standards ensured by sophisticated process design, especially in DSP, are required [7].

From many perspectives, the biopharmaceutical industry is at a turning point. There is the so-called patent cliff of blockbuster therapeutic proteins taking place between 2012 and 2019. By definition, a blockbuster is a drug with annual sales exceeding 1 billion USD [8]. Within a time period of 7 years, the patents of 10 blockbusters capturing an annual market of up to 12.5 billion USD each are about to expire. A new market for generics – the biosimilars – is expected to arise, resulting in drastical falling of the revenue obtained through the sales of the originators [9]. The first biosimilars approved since 2006 were hormones for human growth, red blood cell production, and bone marrow stimulation, followed by the first biosimilar monoclonal antibody in 2013 [10, 11]. Additionally, there is a trend from large quantity blockbuster production of therapeutic drugs towards a more personalized medicine [12]. Thus, process developments are shifted from straightforward routines to individual customization. Recent advances in upstream processing – USP (strain construction, strain screening and cultivation) indicate a titer increase at commercial scale

by approximately 20 % from 2014 to 2019. At the same time, however, adoption of improvements in DSP has been predicted to be much slower [13]. Consequently, the rising competition, the achievements in USP, as well as the demand for fast and reliable tools result in pressure on cutting manufacturing costs and development of innovative technologies in DSP.

As a reaction of research and development in both academia and industry, novel technologies have been tested, developed, and introduced. Great progress can be seen in the preparative liquid-solid column chromatography, which is the workhorse for present purification purposes because of its high resolution and the know-how accumulated in the biopharmaceutical industry over last decades [14, 15]. The standard process development strategy for chromatographic unit operations is mainly based on time-consuming and labor-intensive wet laboratory experiments. High-throughput screening in combination with the approach of design-of-experiment (DoE) has been adopted in recent years to improve such development setups, but still is not sufficient under the given circumstances [16]. Especially in line with the Quality by Design (QbD) initiative suggested by the US FDA [17], mechanistic modeling presents one of the most promising method in process development [18–25]. The artificial neural network (ANN) modeling, which has already been applied in numerous areas and proven itself as an extremely powerful tool, has been explored in combination with different liquid chromatography modes and formats [26–30]. Moreover, the switch to disposable and single-use technology has been implemented to provide a reduction in sanitization expense, more compact, as well as flexible and multi-product facilities [31]; Alternatives to liquid chromatography, such as precipitation has been explored [32]; Non-standard chromatography stationary phases such as membrane and fiber have been designed and tested [33]; A shift from batch to continuous processing could be observed in both chromatographic and non-chromatographic techniques. In the following sections, theoretical foundation of liquid-solid chromatography and precipitation with focus on protein purification as well as a brief introduction to the most simple form of artificial intelligence – the artificial neural network – will be presented.

1.1. Fundamentals of Liquid-solid Chromatography

In liquid-solid chromatography, a solid stationary phase and a mobile liquid phase can be found. The stationary phase consists of a solid bed packed with porous particles, fibers, or a monolith in a cylinder made of plastics, glass, or metal. During the transport in the chromatographic system, the mobile phase – a heterogeneous solution containing different species – can be separated due to differences in their retention behavior. In size exclusion chromatography (SEC) for instance, the stationary phase has a defined pore size distribution. In contrast to larger molecules, smaller ones are able to penetrate into smaller pores, leaving the column later. In bioseparation, the most used chromatography class is the adsorption chromatography, which is also the focus of this thesis. Here, the effects leading to the retention differences between different species include fluid dynamics within the

chromatographic column and their strength of adsorption on the surface of the solid phase. Differentiated by the adsorption mechanism, the most common chromatography techniques in the current field are affinity chromatography (AC), ion-exchange chromatography (IEX), and hydrophobic interaction chromatography (HIC). AC has its basis on the specific binding of target molecules to ligands on adsorber surface, such as immobilized enzymes or metal ions. In IEX and HIC, molecules and ligands undergo electrostatic and hydrophobic interactions, respectively.

In preparative chromatography, it is aimed to collect target molecules as highly concentrated as possible, which cannot be achieved ideally due to the broadening of residence time distribution of the solutes during travel along the column. The effects are summarized in the axial dispersion and mass transfer resistance. The former has its roots in non-ideal fluid distribution as shown in Fig. 1.1. Fluid dynamic adhesion inside the microscopic channels of the packed bed results in a non-uniform flow velocity profile with its maximum in the middle of the channel [34], which is schematically shown in Fig. 1.1 (a). Local inhomogeneous voidages in the packed bed results in the mesoscopic fluid distribution effect responsible for differences in fluid velocities and path-length of molecular traveling (Eddy-diffusion) [34] as shown in Fig. 1.1 (b). From the macroscopic view presented in Fig. 1.1 (c)., local non-uniformities of the voidage the column wall and the adsorbent material lead to the so-called wall effect [34]. As column chromatography is usually conducted at chromatographic plant consisting of pumps, valves, detectors, and tubings as its main components, the contribution of the extra column effects cannot be neglected. All the effects contribute to a symmetrical Gaussian distributed peak width increase – adding up to the axial dispersion. It should be mentioned, that in large-diameter columns, the radial fluid distribution at column inlet and column outlet is the most critical point due to the pressure drop of the packed bed.

(a)   (b)   (c)
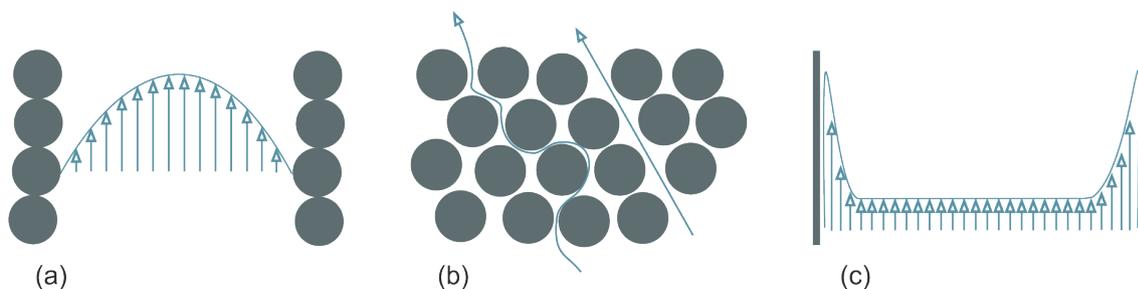
Figure 1.1.: Non-ideal fluid distribution responsible for the broadening of residence time distribution according to Tsotsas [34].

Besides the axial dispersion, mass transfer resistance as shown in Fig. 1.2 contributing to the total band broadening include convective and diffusive transport towards the particle, film diffusion through the film layer between the interstitial and the pore volume, diffusion into the particle pores and at the surface, as well as the adsorption/desorption on the inner adsorbent surface. Are these mass transfer phenomena slower than the microscopic, mesoscopic, and macroscopic fluid distribution effects mentioned above, a non-symmetrical band broadening can be observed. Here, some molecules retain longer in the column due to

their deeper penetration into the pore system and/or stronger interaction with adsorbent, resulting in a late-eluting part called the peak tailing. Unlike the symmetrical peaks, which can be described with the Gaussian distribution, the Exponential-Modified-Gauss (EMG) function is recommended to cover the unsymmetrical ones [35–38].
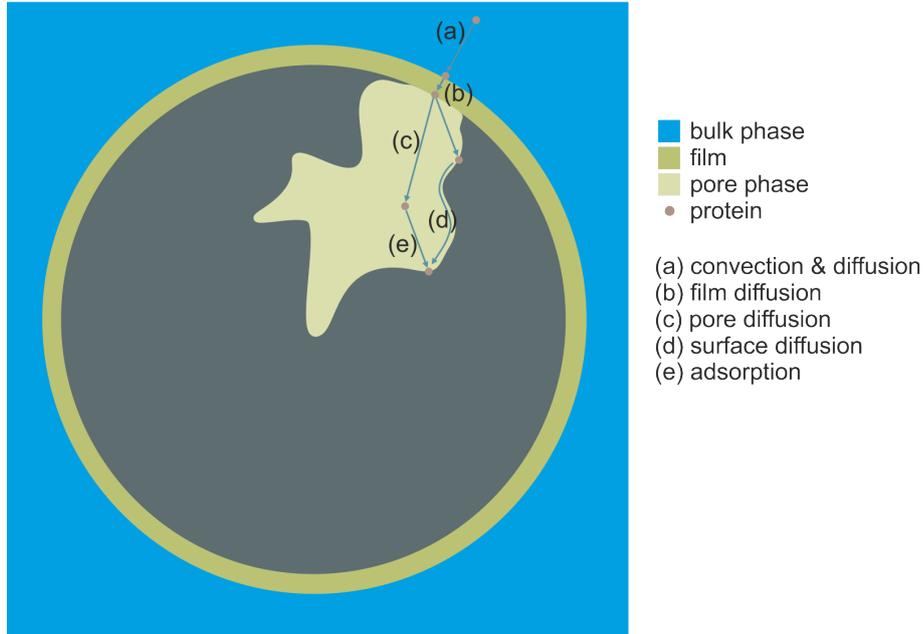
Figure 1.2.: Mass transfer phenomena during protein adsorption in a porous bead.

As common in other fields of chemical engineering, the plate number N as well as the corresponding height of an equivalent theoretical plate $HETP$ are employed to characterize the efficiency of packed beds. In chromatography, $HETP$ accounts for all contributions to the total band broadening (fluid distribution non-idealities and mass transfer resistance) and is defined as the column length L_c divided by N as shown in Eq.1.1.

$$HETP = \frac{N}{L_c} \tag{1.1}$$

In practice, a non-penetrating and non-interacting molecule such as dextran is used as tracer to approximate the $HETP$ without the contribution of the mass transfer resistance. Following the correlation in Eq.1.2, the axial dispersion coefficient of a column D_{ax} can be calculated with the average interstitial velocity of the flowing fluid u_{int}, which depends on pump speed and bed porosity.

$$D_{ax} = \frac{HETP \cdot u_{int}}{2} \tag{1.2}$$

The resistance caused by the boundary layer between the interstitial volume and the pore volume is described with a film diffusion coefficient. The intra-particle diffusion is modeled with a pore diffusion coefficient. In less complex models, a lumped film transfer coefficient

instead of both diffusion coefficients can be found. The occurrence of the surface diffusion is often neglected in adsorption chromatography, as it is exactly then physically plausible, if the attractive forces between the adsorbent and the adsorbate is very weak [39]. The adsorption of adsorbate onto the surface is described by an isotherm equation.

1.1.1. Overview of Chromatography Models

Since the invention of column chromatography, there always has been desire to understand it mathematically and great attempts to describe the chromatographic effects in different scales [40]. Assumptions have been made in the derivation of the differential mass balances. When modeling liquid-solid chromatography, these assumptions and the resulting simplifications have to be kept in mind:

- Homogeneous bed packing with spherical particles of constant diameter
- Homogeneous radial distributions
- Incompressible fluid with a viscosity independent of the fluid flow state
- Isothermal and isobaric process
- Constant axial dispersion coefficient
- Constant partial molar volume of components in both mobile and stationary phases
- Constant operating conditions, such as temperature, pressure, and mobile phase flow rate
- Inert effluent
- No convection and movement inside the pore volume caused by the interstitial flow
- No size-exclusion effects

These assumptions simplify *i.a.* the mass balances for liquid chromatography to be one-dimensional as depicted in Fig. 1.3.
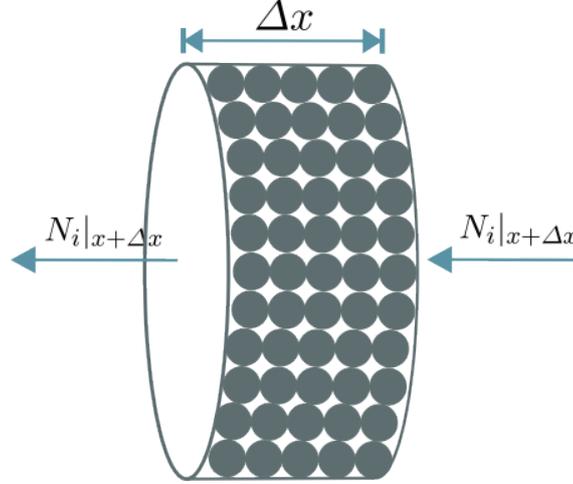
Figure 1.3.: The mass fluxes at the inlet and outlet of a column slice.

Considering the mass fluxes of the component i at the inlet and outlet of a column slice with the thickness Δx, the flux entering the slice is

$$N_i|_x = \epsilon S(u_{int}c_i - D_{ax}\frac{\partial c_i}{\partial x})|_{x,t}, \qquad (1.3)$$

the flux leaving the slice

$$N_i|_{x+\Delta x} = \epsilon S(u_{int}c_i - D_{ax}\frac{\partial c_i}{\partial x})|_{x+\Delta x,t}, \qquad (1.4)$$

and the accumulation rate in the slice

$$N_i|_{x+\Delta x} - N_i|_x = S\Delta x(\epsilon\frac{\partial c_i}{\partial t} + (1-\epsilon)\frac{\partial c_{s,i}}{\partial t})|_{\bar{x}}. \qquad (1.5)$$

Here, S equals $\pi d_c^2/4$ is the circular cross-sectional area of the column with a diameter of d_c, ϵ the column porosity, D_{ax} the axial dispersion coefficient, c_i the concentration of a component in the mobile phase at a certain time t and position x at the inlet, $x + \Delta x$ at the outlet, or the average value in the slice $\bar{x}$, and $c_{s,i}$ the concentration of a component in the stationary phase (pore volume and on the inner pore surface). In Eqs. 1.3 and 1.4, the axial dispersion is defined in analogy to Fick's first law of diffusion. Insert the Eqs. 1.3 and 1.4 into Eq. 1.5, the mass balance equation results:

$$\epsilon S(u_{int}c_i - D_{ax}\frac{\partial c_i}{\partial x})|_{x,t} - \epsilon S(u_{int}c_i - D_{ax}\frac{\partial c_i}{\partial x})|_{x+\Delta x,t} = S\Delta x(\epsilon\frac{\partial c_i}{\partial t} + (1-\epsilon)\frac{\partial c_{s,i}}{\partial t})|_{\bar{x}}$$
$$(1.6)$$

After rewriting and assuming an infinitesimally small slice, *i.e.* approaching Δx towards 0,

Eq. 1.7 follows.

$$D_{ax}\frac{\partial^2 c_i}{\partial^2 x} = u_{int}\frac{\partial c_i}{\partial x} + \frac{\partial c_i}{\partial t} + \frac{1-\epsilon}{\epsilon}\frac{\partial c_{s,i}}{\partial t}$$
$$\Leftrightarrow \frac{\partial c_i}{\partial t} = D_{ax}\frac{\partial^2 c_i}{\partial^2 x} - u_{int}\frac{\partial c_i}{\partial x} - \frac{1-\epsilon}{\epsilon}\frac{\partial c_{s,i}}{\partial t} \tag{1.7}$$

The left-hand side is the accumulation in the mobile phase. The first term on the right-hand side accounts for the diffusion, the second the convection, and the third the accumulation in the stationary phase.

Based on the differential mass balance of the solute in an infinitesimally small slice of a column, various models describing chromatographic processes have been developed. The classification of different model approaches presented here follows Guiochon and Schmidt-Traub *et al.* [40, 41]. According to the endeavor to reflect the reality, the models differ in their complexity. The general rate model (GRM) covers convective transport, axial dispersion, mass transfer resistance, and adsorption equilibrium/kinetic. In comparison to the GRM, the lumped rate models (LRM) neglect either axial dispersion, adsorption equilibrium/kinetic, or mass transfer resistance. The most simple ideal equilibrium model takes only convective transport and adsorption in equilibrium into consideration.

1.1.2. Ideal Equilibrium Model

The ideal equilibrium model, also known as the ideal or basic model, was first describe by Wicke [42] and later derived in the current form by De Vault [43]. The generous assumption of local equilibrium between mobile and stationary phase neglects all effects caused by axial dispersion, mass transfer resistance and adsorption kinetic. In mathematical terms, the axial dispersion coefficient is set to zero and the coefficients which account for mass transfer resistance to infinity. The concentration in the mobile phase is identical to that in the pore volume. Hence the general mass balance equation is reduced to:

$$\frac{\partial c_i}{\partial t} = -u_{int}\frac{\partial c_i}{\partial x} - \frac{1-\epsilon}{\epsilon}\left(\epsilon_p\frac{\partial c_i}{\partial t} + (1-\epsilon_p)\frac{\partial q_i}{\partial t}\right). \tag{1.8}$$

The diffusion term disappears and the accumulation term in the stationary phase is divided into pore volume and adsorbent surface with the particle porosity ϵ_p and q_i the concentration of absorbed component i. The rearranged form can commonly be found in literature:

$$\frac{\partial c_i}{\partial t} = -u_m\frac{\partial c_i}{\partial x} - \frac{1-\epsilon_t}{\epsilon_t}\frac{\partial q_i}{\partial t}. \tag{1.9}$$

Here, the column porosity is replaced by the total porosity

$$\epsilon_t = \epsilon + \epsilon_p(1-\epsilon)$$

and the interstitial velocity is replaced by the effective velocity

$$u_m = \frac{\epsilon}{\epsilon_t} u_{int}.$$

The ideal model delivers the foundation for a variety of pioneering work on thermodynamic behavior of a chromatographic column in linear [44, 45] and nonlinear region [46–48]. To date, this model is still meaningful, as it can be used (after some rearrangements) to deliver rough estimates for propagation velocity of a component inside the column, the observable retention time of a Dirac pulse, or the position of the shocks – the sharp part of a peak/breakthrough curve.

1.1.3. Equilibrium Dispersive Model

The equilibrium dispersive model as described by Eq. 1.10 is an extension of the ideal equilibrium model. Additionally, peak broadening effects caused by axial dispersion and mass transfer resistances are lumped into the apparent dispersion coefficient D_{app}, which was first introduced by van Deemter and coworkers [45]. Thermodynamic equilibrium between the interstitial phase, pore volume, and adsorbent surface is assumed. Formally, all coefficients regarding mass transfer resistance are set to infinity and the film resistance between the interstitial phase and pore volume is neglected.

$$\frac{\partial c_i}{\partial t} = D_{app,i} \frac{\partial^2 c}{\partial^2 x} - u_{int} \frac{\partial c_i}{\partial x} - \frac{1-\epsilon}{\epsilon} \left(\epsilon_p \frac{\partial c_i}{\partial t} + (1-\epsilon_p) \frac{\partial q_i}{\partial t} \right) \tag{1.10}$$

After introducing total porosity ϵ_t and the effective velocity u_m as stated above, the commonly used form is as described as in Eq. 1.11.

$$\frac{\partial c_i}{\partial t} = \frac{\varepsilon}{\varepsilon_t} D_{app,i} \frac{\partial^2 c_i}{\partial^2 x} - u_m \frac{\partial c_i}{\partial x} - \frac{1-\epsilon_t}{\epsilon_t} \frac{\partial q_i}{\partial t}$$

$$\Leftrightarrow \frac{\partial c_i}{\partial t} = \tilde{D}_{app,i} \frac{\partial^2 c_i}{\partial^2 x} - u_m \frac{\partial c_i}{\partial x} - \frac{1-\epsilon_t}{\epsilon_t} \frac{\partial q_i}{\partial t} \tag{1.11}$$

In nonlinear isotherms, $\tilde{D}_{app,i}$ includes concentration effects and therefore varies with concentration changes. This model possesses the capability of describing asymmetrical elution profiles. The observed tailing can be explained by the fact, that the late eluting part is affected longer by the broadening effect.

1.1.4. Lumped Rate Models

In the lumped rate models, a second parameter describing rate limitation has been introduced besides the axial dispersion coefficient. A commonly used lumped rate model is the transport dispersive model, that summarizes the internal and external mass transfer resistance in the effective film transfer coefficient $k_{eff,i}$. The effective film transfer is

formally thought to be independent of the axial dispersion, but dependent on the effects caused by mass transfer resistance, such as film diffusion, pore diffusion, surface diffusion, and concentration of the solutes. The general mass balance equation

$$\frac{\partial c_i}{\partial t} = D_{ax,i}\frac{\partial^2 c_i}{\partial^2 x} - u_{int}\frac{\partial c_i}{\partial x} - \frac{1-\epsilon}{\epsilon}\left(\epsilon_p\frac{\partial c_i}{\partial t} + (1-\epsilon_p)\frac{\partial q_i}{\partial t}\right) \qquad (1.12)$$

has been extended by the mass balance in the stationary phase

$$k_{eff,i}4\pi r_p^2(c_i - c_{p,i}) = \frac{4}{3}\pi r_p^3\left(\epsilon_p\frac{\partial c_{p,i}}{\partial t} + (1-\epsilon_p)\frac{\partial q_i}{\partial t}\right)$$

$$\Leftrightarrow k_{eff,i}\frac{3}{r_p}(c_i - c_{p,i}) = \epsilon_p\frac{\partial c_{p,i}}{\partial t} + (1-\epsilon_p)\frac{\partial q_i}{\partial t}. \qquad (1.13)$$

The film resistance within the liquid boundary layer surrounding the spherical particles with radius r_p between the interstitial phase and pore volume is stated on the left-hand side, whereas the accumulation in the pore volume and the adsorbent surface is declared on the right-hand side. The solution of the transport dispersive models is always an asymmetrical elution profile. The degree of asymmetry increases with increasing axial dispersion coefficient D_{ax} and decreasing effective mass transfer coefficient k_{eff}.

1.1.5. General Rate Models

The general rate models are the most complex models in chromatography. Besides axial dispersion and film diffusion, the impacts of other mass transfer phenomena are considered. The widely used model proposed by Gu *et al.* [49, 50] complements the general mass balance as shown in Eq. 1.12 by the description of diffusion inside the particle pores, which is therefore also known as the pore diffusion model. Analogue to the transport dispersive model, the external mass transfer – the mass transfer through the film surrounding the particles for $r = r_p$ – is described as in Eq. 1.13. The internal mass transfer is the mass transfer inside the particle pores for $r \in [0, r_p]$. It is modeled with Fick's diffusion:

$$\epsilon_p\frac{\partial c_{p,i}}{\partial t} + (1-\epsilon_p)\frac{\partial q_i}{\partial t} = \epsilon_p\frac{1}{r^2}\frac{\partial}{\partial r}(r^2 D_{p,i}\frac{\partial c_{p,i}}{\partial r}). \qquad (1.14)$$

Neglecting the surface diffusion is justified, if a pronounced micropore system is not given and the adsorbent affinity not high. In most cases of adsorption chromatography, the pore diffusion coefficient exceeds the surface diffusion coefficient by orders of magnitude [51–53].

To solve the partial differential equations, initial and boundary conditions are required. At the column inlet the boundary conditions derived by Danckwerts [54] is frequently employed. Here, at the column outlet a zero gradient for the mobile phase concentration is

assumed:

$$\frac{\partial c}{\partial x}(L_c,t) = 0 \tag{1.15}$$

$$\frac{\partial c}{\partial x}(0,t) = \frac{u_{int}}{D_a x}\left(c(0,t) - c_{in}(t)\right) \tag{1.16}$$

1.1.6. Adsorption Isotherms

Besides the models covering the fluid dynamics in the chromatographic column in the interstitial phase and pore volume, an adsorption isotherm has to be added for the description of the adsorption processes on the inner adsorbent surface. The derivation of the commonly used adsorption isotherms is based on the law of mass action. The rate of adsorption is thought to be proportional to the product of activities of the participating components. A comprehensive review on the current thermodynamic framework of protein chromatography adsorption isotherms has been delivered by Mollerup [55]. An equilibrium between all components is assumed under isothermal and isobaric condition, leading to

$$\Delta G = \sum_i \nu_i \mu_i = 0. \tag{1.17}$$

Here, ΔG is the the Gibbs free energy change, ν_i the stoichiometric coefficients, and μ_i the chemical potentials defined as:

$$\mu_i \equiv \mu_i^0 + RT \ln a_i = \mu_i^0 + RT \ln \tilde{N}_i \gamma_i. \tag{1.18}$$

μ_i^0 are the standard chemical potentials, R the ideal gas constant, T the thermodynamic temperature, a_i the activities, γ_i the activity coefficients, and $\tilde{N}_i$ the mole fractions. The Gibbs free energy is defined as

$$\Delta G = \Delta G^0 + RT \ln K, \tag{1.19}$$

with the standard Gibbs free energy G^0 and the thermodynamic equilibrium constant K. In ion-exchange chromatography, a protein P and ν ligands L occupied by counter-ions S are thought to be in equilibrium with ν replaced counter-ions and a protein-ligand complex PL_ν:

$$P + \nu LS \Leftrightarrow \nu S + PL_\nu. \tag{1.20}$$
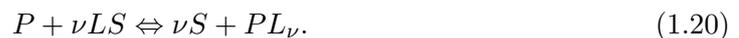
Based on the this assumption and the consideration of Gibbs free energy change, the stoichiometric displacement model (SDM) has been introduced as in Eq. 1.21 [56].

$$K_{eq} = \frac{q_i c_{salt}^{\nu_i}}{q_{salt}^{\nu_i} c_{p,i}} \tag{1.21}$$

Here, K_{eq} is the equilibrium constant, c_{salt} the concentration of counter-ion in the mobile phase, q_{salt} the concentration of counter-ion bound to the ligands, $c_{p,i}$ the protein concentration in the pore volume. q_{salt} can be replaced by applying the electroneutrality on the adsorbent surface with a total ligand density of Λ

$$q_{salt} = \Lambda - \sum_{j=1}^{n} \nu_j q_j, \tag{1.22}$$

resulting in the commonly used form in the literature as in Eq.1.23.

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

Brooks and Cramer [57] extended the SDM isotherm by the steric hindrance factor σ_i, which represents the number of binding sites on the adsorbent surface sterically hindered by the protein i upon binding. Inserting the extended electroneutrality on the adsorbent surface

$$q_{salt} = \Lambda - \sum_{j=1}^{n} \left(\nu_j + \sigma_j\right) q_j \tag{1.24}$$

into Eq. 1.21, the widely accepted steric mass action (SMA) isotherm follows:

$$K_{eq} = \frac{q_i c_{salt}^{\nu_i}}{\left(\Lambda - \sum_{j=1}^{n} \left(\nu_j + \sigma_j\right) q_j\right)^{\nu_i} c_{p,i}}. \tag{1.25}$$

Practically, the SMA isotherm accounts for the reduction of ionic capacity due to protein binding, hence it delivers almost identical results as the SDM isotherm in linear region, but is more precise in nonlinear region of adsorption.

In hydrophobic interaction chromatography (HIC), a protein P and n ligands L are thought to be in equilibrium with a protein-ligand complex PL_n:

$$P + nL \Leftrightarrow PL_n. \tag{1.26}$$

Based on the this assumption and the consideration of Gibbs free energy change, Mollerup introduced an isotherm model for protein adsorption in HIC analogously to the SMA isotherm as shown in Eq. 1.27 [58].

$$K_{eq} = \frac{q_i}{c_{p,i} \gamma_{p,i} c_L^n} \tag{1.27}$$

Here, c_L is the concentration of free ligands and $\gamma_{p,i}$ the activity coefficient of the protein i. It is assumed that the chemical potential of the immobilized hydrophobic ligands and the ligand bound protein are less depends on the ionic strength, hence the neglect of the

activity coefficients of q_i and c_L. The total concentration of ligands is defined as

$$\Lambda = c_L + (n_i + s_i)q_i. \tag{1.28}$$

As the steric hindrance factor in IEX, s_i accounts for the number of ligands sterically hindered upon protein binding. γ_i is calculated according to van der Waals equation of state [58]:

$$\ln \gamma_{p,i} = k_{p,i}c_{p,i} + k_s c_{salt}. \tag{1.29}$$

Here, k_p and k_s are protein and salt specific constants. Consequently, the final HIC isotherm results

$$K_{eq} = \frac{q_i}{c_{p,i}\left(k_{p,i}c_{p,i} + k_s c_{salt}\right)\left(\Lambda - (n_i + s_i)q_i\right)^n}. \tag{1.30}$$

Details on the derivation of the models can be found in the aforementioned references.

1.2. Principles of Precipitation

For protein purification, precipitation is seen as an alternative technique to preparative chromatography. The solubility of proteins in water, diluted aqueous solutions, and concentrated salt solution was extensively studied in the last century. The protein solubility is a thermodynamic quantity defined as a certain protein concentration in saturated solution that is in equilibrium with a solid phase under a given set of conditions. Protein solubility depends thereby on both intrinsic factors, which are primarily defined by the amino acids on the protein surface, and extrinsic factors including pH, ionic strength, temperature, and solvent additives [59]. In the beginning, proteins relatively insoluble in water were investigated at their isoelectric points, forcing them to precipitate. The protein solubility was found to be specific with respect to protein characteristics and the extrinsic experimental conditions. The solubility was thought to be generally independent of the amount of the precipitate, resulting in the suggestion of a simple heterogeneous equilibrium between the solution and the precipitate, until the observation was made, that the solubility of serum globulin is proportional to the amount of precipitate [60]. Furthermore, it was found that some proteins relatively insoluble in water are very soluble in diluted acid, base, or salt solutions. Later on, the observation of protein precipitation induced by adding neutral salts was made [60]. At critical amount of salt added to solution containing differently soluble proteins, the less soluble ones are precipitated, whereas the others remain in solution. It has been observed that not only the salt concentration, but also the salt type is of great importance for the protein solubility. Based on these effects, methods for protein separation using salt-induced precipitation (*i.a.* by shielding electrostatic interactions of proteins) have been developed. Chick and Martin [61] investigated the effect of protein concentration upon solubility at a constant salt concentration with the model

system consisting of egg albumin and ammonium sulfate. Based on the observation, that an increasing protein concentration led to a greater amount of precipitated protein, they concluded that protein solubility would be a function of the protein concentration itself. An explanation was delivered later by Sørensen and Høyrup [62], who developed a "method of proportionality" and demonstrated that due to the hydration of the egg albumin, the apparent ammonium sulfate concentration was increased proportionally to the increased egg albumin concentration. Later, the effect of the salt concentration upon solubility was studied by Sørensen and Høyrup [63]. The relation between the salt concentration and the protein solubility has been postulated by a logarithm equation:

$$\log \tilde{s} = \tilde{\alpha}\tilde{m} + \tilde{\beta}. \tag{1.31}$$

Here, the correlation is purely empirical with $\tilde{s}$ as the solubility of the protein, $\tilde{m}$ the molecular concentration of the salt, as well as $\tilde{\alpha}$ and $\tilde{\beta}$ the constants dependent on protein and salt.

This equation has been later applied to precipitation induced by polyethylene glycol (PEG), with $\tilde{m}$ as the molecular concentration of PEG, which has shown promising results for the separation of complex mAb solutions [64]. Although precipitation is well investigated and shows various applicable scenarios in protein purification, the fundamental mechanisms are not completely understood [65–67]. Two theories have been proposed to describe the mechanisms of PEG-induced precipitation of proteins.
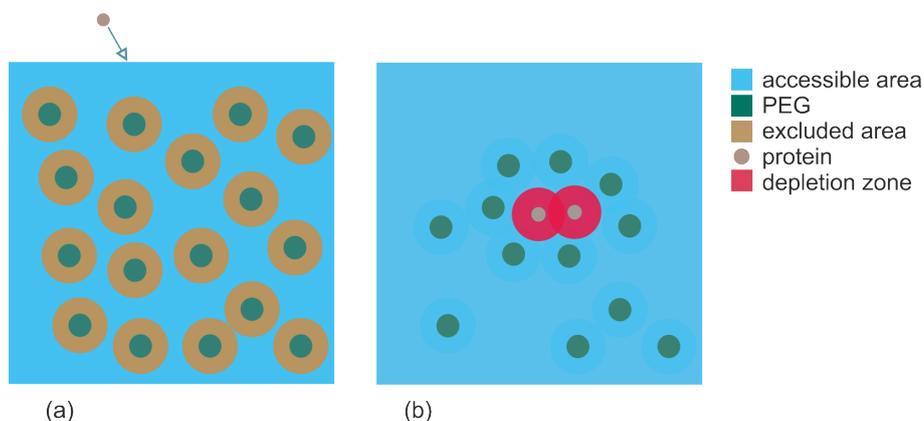
Figure 1.4.: Theories explaining the mechanisms of protein precipitation using PEG.

The theory of excluded volume as depicted in Fig. 1.4 (a) assumes that PEG traps the solvent in its coiled structure and occupies solvent regions, leading to a steric exclusion of the proteins. The result is the increase of protein concentration in the non-occupied regions until the protein solubility is exceeded and precipitation occurs [68]. The theory of attractive depletion as depicted in Fig. 1.4 (b) extends the former theory by postulating a formation of concentration gradient due to the overlapping of two depletion zones of proteins. This gradient results in an osmotic pressure and finally in precipitation of proteins.

1.3. Artificial Neural Networks

At first inspired by the biological neural networks in an animal brain, the artificial neural networks (ANNs) are computing systems which can receive information from the surrounding, store it, process it, and generate output information. They learn without any *a priori* knowledge and evolve their own "understanding" based on learning material. Mathematically, an ANN is a collection of interconnected artificial neurons which can perform simple operations, e.g. multiplication and addition. Since the ANNs' first introduction by McCulloch and Pitts in 1943 [69], they have shown successes in various application scenarios and finally thrives due to improved computing power in the beginning of the 21th century in different fields of artificial intelligence. For example, the Blue Brain project aims to understand the fundamental principles of mammalian brain structure and function based on a digital reconstruction [70]; AlphaGO defeated the humankind in the classic board game of Go [71], which is highly abstract due to its huge number of possible game outcomes of approximately 10^{170}; In self-driving cars, ANNs are trained to mimic a human driver's acceleration, braking, and steering as responses to visual inputs from sensor detections [72].

1.3.1. Architecture

An ANN consists of a number of processing units — the so-called artificial neurons or nodes — in different arrangements. Besides the completely free and self-optimizing architectures, the nodes are usually prearranged in layers as shown in Fig. 1.5.
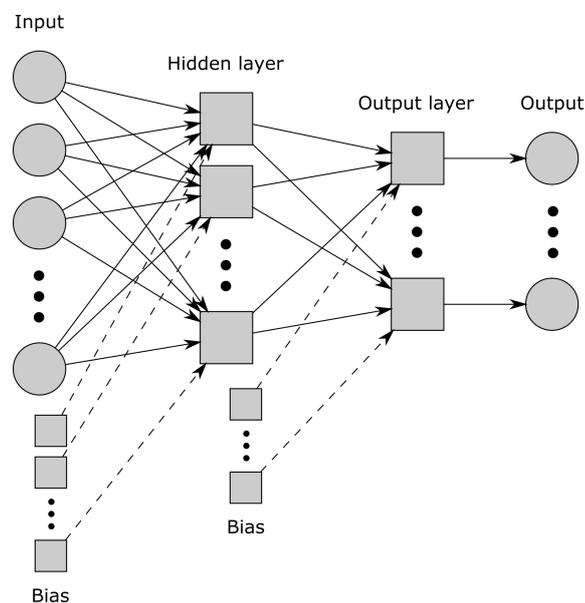
Figure 1.5.: An ANN model of simple architecture.

The input layer is the interface between the input data and the subsequent layers in the ANN. The nodes in the input layer receive information from surrounding and forward it to

nodes in the hidden layer. The nodes in the hidden layers and output layer are responsible for the calculation of the weighted sum of received signals and the biases, as well as its processing with transfer functions. Eventually, the output signal is sent to all receiving nodes in the next layer.

The operation in the ith node in the jth layer can be described mathematically as

$$\omega_{i,j} = \phi_{i,j} \left(\sum_{k=1}^{m} \xi_k \omega_{k,j-1} + \zeta_{i,j} \right).$$ (1.32)

A node responses with the output signal $\omega_{i,j}$ on m received input signals $\omega_{k,j-1}$ by multiplying them with their specific weights ξ_k under consideration of the bias $\zeta_{i,j}$. $\phi_{i,j}$ is a linear, nonlinear, or step transfer function.

The linear transfer function as shown in Eq. 1.33 transforms the input signals linearly by adding a bias term to their weighted sum and is mostly used for function fitting problems.

$$y(\tilde{u}) = a \cdot \tilde{u} + b$$ (1.33)

Here, $\tilde{u}$ is the input signal, and a and b the constants in the activation function. The most used nonlinear transfer function is the sigmoid transfer function as shown in Eq. 1.34 which accepts input signals from negative to positive infinity, and generates outputs between 0 and 1. The derivative of sigmoid neurons can be calculated easily, making it attractive for calculating the weight updates during backpropagation.

$$y(\tilde{u}) = \frac{a}{1 + \exp(-\tilde{u})} + b$$ (1.34)

The step transfer function as in Eq. 1.35 offers a binary output and is therefore useful in the output layer of an ANN to conduct binary classification of the inputs.

$$y(\tilde{u}) = \begin{cases} 1 & \tilde{u} \geq a \\ 0 & \tilde{u} < a \end{cases}$$ (1.35)

Here, y and $\tilde{u}$ are the generated output and received input of the transfer function, respectively, while a and b are any real numbers.

While the size of the input and output layers are defined by the dimension of the processed data, the number of hidden layers, their size, and the overall connectivity are case specific. Special attention should be paid to the complexity of the ANN architecture, since there is always a trade-off between the ANN's capability of solving complex problems and its generalization performance [73, 74].

1.3.2. Supervised Learning

The paradigm of supervised learning are the important tasks mastered by the artificial intelligence: Classification and regression. The former is also known as pattern recognition,

the latter as function approximation. In supervised learning, input data from a labeled training data set are mapped to the corresponding target data by applying learning algorithms to modify the network [75, 76]. Based on the processed learning material, the neural networks should be able to evolve their own sets of relevant characteristics. Commonly, the weights and biases of the artificial neurons are adjusted. More advanced meta learning algorithms, however, are capable of updating the entire network, including the learning algorithms, the activation functions, or even the network architecture [77].

The learning success of an ANN is measured by the cost function, which is – like in other mathematical optimization problems – a function that requires to be minimized. It is often defined as the discrepancy between the target data and the network's output, e.g. the softmax cross-entropy, Kullback-Leibler divergence, or the most common mean-squared error (MSE) as defined in Eq. 1.36.

$$C = \frac{1}{2\hat{n}} \sum_{i=1}^{\hat{n}} \left(\hat{y}_i - \hat{y}'_i \right)^2 \tag{1.36}$$

C represents the cost, $\hat{n}$ the total number of training data sets, $\hat{y}$ the desired output, and $\hat{y}'$ the ANN response. To update the weights and biases after every iteration, the gradient of the cost function is commonly calculated by employing the method of backpropagation [78].

In backpropagation, changing the weights and biases based on the calculation of the partial derivatives of the cost function with respect to weights and biases is the core task. Usually, the algorithm of backpropagation is computed in the propagation phase and the weight update phase. In the propagation phase, the input is forwarded through the ANN to generate the initial outputs of each node in each layer according to Eq. 1.32. According to the cost calculated based on the cost function chosen, the error $\vec{\delta}_J$ in the output layer J is defined as

$$\vec{\delta}_J = \nabla_{\hat{y}'} C \odot \phi'(\vec{z}_J), \tag{1.37}$$

with ϕ' as the changing rate of the activation function ϕ at the weighted sum of the received signal and the bias $\vec{z}_J$ in the output layer, which can be defined as

$$\vec{z}_j = \left(\vec{\xi}_j \vec{\omega}_{j-1} + \vec{\zeta}_j \right) \tag{1.38}$$

for any hidden and output layer. The error $\vec{\delta}_j$ in terms of the error in the next layer is defined as

$$\vec{\delta}_j = \left((\vec{\xi}_{j+1})^T \vec{\delta}_{j+1} \right) \odot \phi'(\vec{z}_j), \tag{1.39}$$

where ϕ' represents the changing rate of the activation function ϕ at the weighted sum of the received signal and the bias $\vec{z}_j$, $\vec{\delta}_{j+1}$ the error and $\vec{\xi}_{j+1}$ the weight matrix in the next layer. In practice, Eqs. 1.37 and 1.39 can be used to calculate the errors for any layer in the ANN. After computing of the error of output layer, the error of the hidden layer

prior to output player can be calculated. In this manner, errors are propagated backwards through the entire ANN.

In the weight update phase, the gradients of the cost function with respect to weights and biases are calculated based on the already computed errors $\delta_{i,j}$ as

$$\frac{\partial C}{\partial \xi_{i,j}} = \omega_{i,j-1} \delta_{i,j} \tag{1.40}$$

and

$$\frac{\partial C}{\partial \zeta_{i,j}} = \delta_{i,j}, \tag{1.41}$$

respectively. For starting of the next iteration, a ratio of the gradient is subtracted from the corresponding initial weight and biases.

1.4. Research Proposal

Biologics represent a growing share of the current pharmaceutical market. The diverse biomolecules e.g., monoclonal antibodies (mAbs), conjugated proteins, or virus-like particles, challenge the bioseparation processes because of their different molecular properties. This has resulted in the advancement of DSP technologies such as, chromatography and alternative unit operations, continuous processing, disposables, process analytical tools, and process development strategies. Fixed orders of unit operations – the so-called platform processes – have been introduced in industry to reduce process development effort for mAbs. With rise of other biopharmaceuticals, the standard platform processes are no longer applicable and need to be cumstomized. Hence, high-throughput experimentation (HTE) has been developed and adapted in academia and industry to screen a large region of the process design space, accepting the drawbacks of its requirement for high amount of sample often unavailable in the early development stages. Later, design-of-experiments (DoE) has been coupled with HTE to balance the accuracy and reliability of the results and the experimental effort needed.

Besides the advantages of saving time and material, the approach of mechanistic modeling copes the demand of the Quality by Design (QbD) initiative suggested by the US Food and Drug Administration (FDA), that the process development and manufacturing should be backed up by thorough scientific understanding. In recent years, great development attempts have been placed upon transfer model-based approaches from academia into industrial laboratory work-flows. Successes have been shown in the usual unit operations, especially in model-based process development and scale-up in ion-exchange chromatography for mAb purification. However, the model-based approach is not straightforward. There are still problems unsolved requiring attention of academic research and development, which can be assigned to different fields of mechanistic modeling in an elementary way: Model building, model calibration, and model application.

The ultimate objective of the present research proposal arises in these fields, to solve urgent issues that prevent the use of model-based approaches on the one hand, and to suggest new applications on the other. In doing so special attention was paid to artificial intelligence based on artificial neural networks (ANNs) which was one of the fastest growing research topics in the last decades and showed successes in diverse utilization scenarios in various fields of technology. In chromatography processes, however, the large amount of data required for the training of the ANNs are often not given and the usage of ANNs has been limited to modeling of retention times. Combining mechanistic modeling and ANNs could open up new possibilities.

Hydrophobic interaction chromatography (HIC) is frequently used in DSP for intermediate and polishing purposes, to remove mAb aggregates and other non-polar contaminants, and to separate antibody-drug conjugate species. In industry, HIC process development relies entirely on rules of thumb or HTE, since an accurate description of the complex salt-dependent protein-ligand interaction is not given. In the selective protein precipitation using PEG – an extraordinary alternative to preparative chromatography because of its

cost efficiency and ease of handling – a similar problem is shared. The development of precipitation processes for biomolecules relies heavily on wet laboratory experiments and empirical correlations. Hence, the first part of this thesis is devoted to build mechanistic HIC and protein precipitation models, which cannot only be used to describe and develop processes, but also as theoretical frameworks for further model extensions.

In modeling of chromatography, some of the model parameters cannot be measured directly and are determined by employing time-consuming methods such as gradient experiments coupled with frontal analysis, combined Yamamoto approach, or curve fitting. These traditional calibration approaches have certain particularities in common *i.e.*, requirement for repeated calibration upon new compounds' entering the system of observation and for profound model understanding. To circumvent these disadvantages, the second part sets the goal to develop a novel model calibration method based on artificial neural networks which represent the most simple form of artificial intelligence.

Mechanistic chromatography modeling is a powerful tool in every aspect, nonetheless its application scenarios reported in the literature are surprisingly limited to model-based process development and robustness analysis. Thus, in the final part of this thesis studies on model application are carried out to investigate the mass transfer and adsorption mechanisms of protein conjugates, to explore an integration of precipitation and continuous chromatography, and eventually to develop a tool for real-time root cause investigation in deviating chromatography processes.

1.5. Outline

Chapter 2 proposes an adsorption isotherm for the description of salt-dependent protein-
ligand interaction in HIC. Chapter 3 introduces a mechanistic model for protein precipitation
using PEG by considering variations in both protein and PEG concentration. Chapter 4
presents the development of a novel calibration method to overcome the disadvantages of
the traditional calibration methods. Chapter 5 addresses the issue of abnormal behavior
of PEGylated proteins in IEX by conducting a full investigation based on mechanistic
chromatography models. In chapter 6, the combination of various process development
strategies and the integration of precipitation and IEX chromatography as the alternative
to expensive Protein A chromatography are discussed. Eventually, chapter 7 explores a
new application of chromatography modeling upon process deviation.

Chapter 2: Water on hydrophobic surfaces: Mechanistic modeling of hydrophobic interaction chromatography

G. Wang, T. Hahn, J. Hubbuch

Journal of Chromatography A (1465), 2016, 71-78

This article presents the derivation of a mechanistic model for hydrophobic interaction
chromatography (HIC). By taking the new insights into the water structure on hyd-
rophobic surfaces into account, assumptions and simplifications were made to cope
the requirement for practical applicability in model-based HIC process development.
The identifiability and reliability of model parameters were demonstrated by column
experiments, application of the inverse method, and calculation of the confidence
intervals for parameter estimates. High-throughput batch chromatography experi-
ments and column experiments beyond calibration space underline the suitability of
the isotherm model.

Chapter 3: Water on hydrophobic surfaces: Mechanistic modeling of protein precipitation using polyethylene glycol

S. Großhans*, G. Wang*, J. Hubbuch (*contributed equally)

Manuscript submitted to Journal of Chromatography A

In this manuscript, the mechanistic similarities between precipitation and HIC
reported in the literature were used to develop a model to describe protein preci-
pitation using PEG quantitatively. After calibration, it was capable of predicting
the precipitation behavior of proteins with very different features. By providing an
opportunity to model the entire precipitation curve in dependence of protein and
PEG concentration, it supplements the widely accepted Cohn equation.

Chapter 4: Estimation of adsorption isotherm and mass transfer parameters in protein chromatography using artificial neural networks

G. Wang, T. Briskot, T. Hahn, P. Baumann, J. Hubbuch

Journal of Chromatography A (1487), 2017, 211-217

This paper presents a calibration method for protein chromatography models based on a combination of mechanistic modeling and ANN. In contrast to traditional calibration methods, it was capable of identifying the model parameters immediately when trained and presented with chromatograms. This method circumvents the requirement for repeated model calibration of traditional methods avoiding the associated high computational expense.

Chapter 5: Model-based investigation on the mass transfer and adsorption mechanisms of mono-PEGylated lysozyme in ion-exchange chromatography

J. Morgenstern*, G. Wang*, P. Baumann, J. Hubbuch (*contributed equally)

Biotechnology Journal (12), 2017, 1700255

This article delivers a thorough investigation on PEGylated lysozyme species in IEX carried out using mechanistic chromatography modeling. The physical meaning of the determined model parameters was utilized to improve the understanding of unusual behavior of PEGylated proteins in preparative chromatography. The knowledge gained is not only of academic interest, but also could help in industrial purification process design and scale-up.

Chapter 6: An integrated precipitation and ion-exchange chromatography process for antibody manufacturing: Process development strategy and continuous chromatography exploration

S. Großhans*, G. Wang*, C. Fischer, J. Hubbuch (*contributed equally)

Journal of Chromatography A (1533), 2018, 66-76

This paper studies the utilization of a development toolbox consisting of high-throughput experimentation (HTE), design of experiments (DoE), and mechanstic modeling for the design and optimization of mAb purification unit operations. The resulting process match the performance of an industrial standard Protein A chromatography step providing an inexpensive alternative. Eventually, the capability of a continuous multicolumn chromatography process was explored based on a calibrated single column chromatography model.

Chapter 7: Root cause investigation of deviations in protein chromatography based on mechanistic models and artificial neural networks

G. Wang, T. Briskot, T. Hahn, P. Baumann, J. Hubbuch

Journal of Chromatography A (1515), 2017, 146-153

This study explores the applicability of mechanistic chromatography models beyond the common scenarios of process development and robustness analysis. With the presented tool consisting of mechanstic models and artificial neural networks, a root cause investigation of deviation during day-to-day operation could be realized without stopping the manufacturing process. As an example, root cause investigation

was carried out on deviation in the ionic capacity of ion-exchange chromatography column, which so far had to be measured by acid-base titration. This shows one of many hitherto unexplored possibilities to apply and to benefit from mechanistic chromatography models.

2. Water on Hydrophobic Surfaces: Mechanistic Modeling of Hydrophobic Interaction Chromatography

Gang Wang[1], Tobias Hahn[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] GoSilico GmbH, Karlsruhe, Germany
* Corresponding author. E-mail address: juergen.hubbuch@kit.edu

Abstract

Mechanistic models are successfully used for protein purification process development as shown for ion-exchange column chromatography (IEX). Modeling and simulation of hydrophobic interaction chromatography (HIC) in the column mode has been seldom reported. As a combination of these two techniques is often encountered in biopharmaceutical purification steps, accurate modeling of protein adsorption in HIC is a core issue for applying holistic model-based process development, especially in the light of the Quality by Design (QbD) approach.

In this work, a new mechanistic isotherm model for HIC is derived by consideration of an equilibrium between well-ordered water molecules and bulk-like ordered water molecules on the hydrophobic surfaces of protein and ligand.

The model's capability of describing column chromatography experiments is demonstrated with glucose oxidase, bovine serum albumin (BSA), and lysozyme on Capto^{TM} Phenyl (high sub) as model system. After model calibration from chromatograms of bind-and-elute experiments, results were validated with batch isotherms and prediction of further gradient elution chromatograms.

2.1. Introduction

Biologics, especially therapeutic proteins, count among the fastest-growing market segments in the current pharmaceutical industry. To cope with the growing number of drug candidates, downstream process (DSP) development faces the challenge of set-up processes for new molecules with limited sample volume within a short time frame. For purification of therapeutic proteins, especially monoclonal antibodies (mAbs), preparative chromatography is the core technology to deliver a highly pure product [79].

Hydrophobic interaction chromatography (HIC) is a frequently used method for intermediate and polishing purposes in DSP [80–82]. HIC is well-known for its remarkable capability to remove aggregates and other non-polar contaminants [82–84]. To date, HIC process development commonly relies on rules of thumb [85] or high-throughput experimentation [86]. Alternative methodologies based on the three-dimensional structure of the protein and its surface hydrophobicity were examined by Mahn and Asenjo [87].

To meet the demands of the Quality by Design approach (QbD) [17–19], a high degree of process understanding is required that can be demonstrated by model building and simulations. At the same time, the sample volume needed for model calibration is expected to be much less than the amount needed for screening with the Design-of-Experiments (DoE) approach. To create a chromatography model, the fluid dynamical principle of mobile phase mass transfer through the chromatography column as well as the thermodynamical principles of protein-ligand interaction [88] must be described. The mass transfer within a chromatography column is well-studied. Several variants of equilibrium and transport-dispersive models are widely accepted [41]. Model-based process development was demonstrated several times for ion-exchange chromatography (IEX) [21, 89] with the steric mass action (SMA) isotherm [90]. The SMA isotherm describes the salt-protein relationship based on the electrostatic equilibrium theory and displacement effect, such that the salt-dependent protein-ligand interaction is directly incorporated. In the case of HIC, this derivation cannot be applied. The challenge of HIC modeling lies in accurately describing the more complex salt-dependent protein-ligand interaction.

The very first theoretical framework for modeling salt effects in hydrophobic interaction chromatography was derived by Melander and Horvath [91], facilitating experimental design and result interpretation. Staby and Mollerup investigated the thermodynamical nature of HIC and presented a mechanistic isotherm [58, 92]. The salt dependence of protein adsorption was described with an exponential term containing protein activity as well as salt activity coefficients. Perkins *et al.* applied the preferential interaction model to examine the salt dependence of the HIC capacity factors [93]. Chen *et al.* also used the preferential interaction model to determine the water molecules released upon salt-related protein binding [94]. Haimer *et al.*

applied an extended Langmuir kinetic derived by Lundstrom to model the spreading phenomenon resulting from injection of a pure protein sample [95, 96]. In addition, Jungbauer *et al.* introduced a modified salt-dependent Langmuir isotherm for linear and nonlinear conditions [97]. Water molecules were first included as participants in the adsorption process by Deitcher *et al.* [98]. Their isotherm described the salt-dependent release of water molecules from the contact area of protein and ligand. Chen and Sun proposed to consider a dehydrated state of the protein, caused by the hydration effect of salt ions [99]. Mirani and Rahimpour extended this two-state isotherm with the consideration of activities instead of concentrations [100].

The formation of a hydrophobic protein-ligand interface is thought to be driven by the hydrophobic effect, involving a reorganization of the water structure on the hydrophobic surfaces of protein and ligand [101]. Frank and Evans assumed a well-ordered ice-like formation of water around hydrophobic groups [102]. Later, a clathrate-like structure of water on hydrophobic surfaces was proposed [103]. In 2014, Shiraga *et al.* carried out terahertz spectroscopy to study the water structure on hydrophobic surfaces of biomolecules and found water molecules around the hydrophobic groups to be more ordered [104].

The present manuscript was enabled by the aforementioned contributions. The new insights into the water structure on hydrophobic surfaces were taken into account and a new mechanistic isotherm was derived for HIC. Keeping the requirements of model-based HIC process development in mind, several assumptions and simplifications were made to facilitate practical applicability. Glucose oxidase, bovine serum albumin (BSA), and lysozyme were chosen as model proteins. This set of proteins covers very different hydrophobic properties and a wide range of sizes, being 160 kDa, 66 kDa, and 14.6 kDa. Gradient elution as well as high-throughput batch experimentation was carried out to prove the predictability of the new isotherm beyond the calibration space.

2.2. Theory

2.2.1. Derivation of a HIC Isotherm Considering the Water Structure

The adsorption mechanism in HIC has been proposed to be an equilibrium between a protein molecule P with stoichiometric number n of ligands L and the protein-ligand complex P_{Ln} [58, 93]:

$$P + nL \longleftrightarrow P_{Ln} \tag{2.1}$$
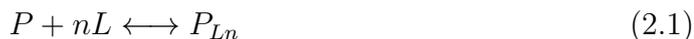
This proposition does not explicitly describe the salt dependence in the case of HIC. Later, Deitcher *et al.* proposed water to be included as a product W with stoichiometric number ξ by assuming a release of water molecules at the protein-ligand contact area [98]:

$$P + nL \longleftrightarrow P_{Ln} + n\xi W \tag{2.2}$$

A water displacement process originally proposed by Geng *et al.* [105] was used as theoretical foundation. The salt was included as a participant by considering the thermodynamic activity of water in electrolyte solutions. They applied a linear empirical model, which describes the natural logarithm of water activity to be equal to the product of salt concentration and a constant factor [98].

Meanwhile, thermodynamical studies have shown that hydrogen bonds between water molecules on hydrophobic surfaces are disrupted to create a cavity. As a counterreaction, water molecules form well-ordered structures, such that a thermodynamically favored state is achieved [106–108]. Upon adsorption, these well-ordered water molecules are thought to be reorienting in a bulk-like structure. This so-called hydrophobic hydration phenomenon is added to the general consideration of the adsorption mechanism according to Eqs. (2.1) and (2.2) to formulate the new adsorption isotherm.

The hydrophobic surfaces of protein P and ligand L are thought to be stabilized by well-ordered water molecules. The protein-ligand complex P_{Ln} is assumed to be bound with n binding sites, which are stabilized by β bulk-like water molecules W_B each:

$$P + nL \longleftrightarrow P_{Ln} + n\beta W_B \tag{2.3}$$

Under consideration of Gibbs free energy and following the approach presented by Mollerup [58], the equilibrium contant K is derived:

$$K \cong a_{W_B}^{n\beta} \frac{q}{c_p c_L^n} \tag{2.4}$$

Here, q and c_p are the concentrations of bound protein and protein in solution, respectively. c_L depicts the concentration of ligands available for binding and a_{W_B} the activity of the bulk-like water molecules. In the next step, parameterizations for a_{W_B}, c_L and β have to be found.

According to stoichiometric considerations, the number of water molecules involved is linearly correlated to the protein concentration in the stationary phase q. Thus, a linear correlation is proposed to substitute a_{W_B}:

$$a_{W_B}^{n\beta} \cong \nu q^{n\beta} \tag{2.5}$$

Here, the stoichiometric constant ν is assumed to be independent of the salt concentration.

In Eq. (2.6), the free ligand concentration c_L is defined as a function of total capacity Λ, hydrophobic binding sites n, and steric hindrance factor s similarly to the SMA model and formulation by Mollerup for HIC [58, 90].

$$c_L^n = (\Lambda - (n+s)\, q)^n \tag{2.6}$$

Inserting Eqs. (2.5) and (2.6) into Eq. (2.4) and collecting all constants on the left-hand side, we obtain the following isotherm equation:

$$K = \frac{\nu q^{1+n\beta}}{c_p\, (\Lambda - (n+s)q)^n} \tag{2.7}$$

$$\xrightarrow{\cdot \frac{\Lambda^n}{\nu}}: k_{eq} = \frac{q^{1+n\beta}}{c_p \left(1 - \frac{q}{q_{max}}\right)^n} \tag{2.8}$$

with the saturation capacity $q_{max} = \Lambda/(n+s)$.

In the final step, the salt dependency of the bulk-like ordered water molecules β is modeled. To this, β and the hydration number of the salt ions h are assumed to be reciprocal. As the salt ions attract water molecules to form a hydration shell, the ionic hydration number h is described with high reliability by an exponential relation as [109]:

$$h = h_0 \exp(-k c_{salt}) \tag{2.9}$$

with h_0 being the ionic hydration number at infinite dilution and k the constant that describes the change of hydration number with increasing ionic concentration c_s. Thus, the salt-dependence of the model parameter β can be described by the exponential term:

$$\beta = \beta_0 \exp(\beta_1 c_{salt}) \tag{2.10}$$

This completes the derivation of the equilibrium formulation of the isotherm model.

Assuming the interaction to be close to the thermodynamic equilibrium but time-dependent, the kinetic formulation of adsorption isotherm results in:

$$k_{kin}\frac{dq}{dt} = k_{eq}\left(1 - \frac{q}{q_{max}}\right)^n c_p - q^{1+n\beta} \tag{2.11}$$

In accordance with Hahn *et al.*, the equilibrium coefficient $k_{eq} = k_{ads}/k_{des}$ and kinetic coefficient $k_{kin} = 1/k_{des}$ are used instead of adsorption and desorption coefficients k_{ads} and k_{des} [110]. The benefit of this formulation is that a change in k_{eq} strongly affects the retention time, while the peak height is mainly affected by a change in k_{kin}.

2.2.2. Transport-dispersive Model

To model the macroscopic mass transport inside the chromatography column with length L and adsorber beads with radius r_p, the widely accepted general rate model (GRM) is employed [41]:

$$\frac{\partial c_i(x,t)}{\partial t} = -u(t)\frac{\partial c(x,t)}{\partial x} + D_{ax}\frac{\partial^2 c(x,t)}{\partial x^2}$$
$$- \frac{1-\epsilon_b}{\epsilon_b}k_{film,i}\frac{3}{r_p}\left(c_i(x,t) - c_{p,i}(x,t)\right) \tag{2.12}$$

$$\frac{\partial c_i(0,t)}{\partial x} = \frac{u(t)}{D_{ax}}(c_i(0,t) - c_{in,i}(t)) \tag{2.13}$$

$$\frac{\partial c_i(L,t)}{\partial x} = 0 \tag{2.14}$$

$$\frac{\partial c_{p,i}(x,t)}{\partial t} = \begin{cases} \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 D_{p,i}\frac{\partial c_{p,i}}{\partial r}\right) - \frac{1-\epsilon_p}{\epsilon_p}\frac{\partial q_i}{\partial t} & \text{for } r \in (0, r_p), \\ \frac{k_{film,i}}{\epsilon_p D_{p,i}}(c_i - c_{p,i}) & \text{for } r = r_p, \\ 0 & \text{for } r = 0. \end{cases} \tag{2.15}$$

Eq. (2.12) describes the mass transfer between the interstitial volume of the mobile phase $c_i(x,t)$ and the pore volume $c_{p,i}(x,t)$. It depends on the peak-broadening effects in the interstitial volume that are lumped in the axial dispersive coefficient D_{ax}, the film transfer coefficient $k_{film,i}$, and the voidage of the bed ϵ_b. Eqs. (2.13) and (2.14) are the Danckwerts boundary conditions. Eq. (2.15) models the exchange between the pore volume concentration $c_{p,i}(x,t)$ and stationary phase q_i depending on the particle voidage ϵ_p and the component-specific pore diffusion coefficient D_p. Combined with the kinetic formulation of the adsorption isotherm Eq. (2.11), a chromatography process can be modeled and simulated.

2.2.3. Numerical Solution

The discretization in space on a grid with equidistant nodes is performed using the finite element method. The discretization in time is carried out with the fractional step θ-scheme [111]. The non-linear equation system is treated with Picard iteration [112]. For parameter estimation, adaptive simulated annealing [113] and the Levenberg-Marquardt algorithm [114] are employed. For initial simulations, the former heuristic algorithm is chosen. It searches a larger space with random jumps to prevent local minima. Following common approaches, the results are refined by applying the latter deterministic algorithm.

2.3. Materials and Methods

2.3.1. Instruments

Gradient elution experiments were carried out using an ÄKTA purifier 10 fast protein liquid chromatography system equipped with pump unit P-903, UV cell (10 mm path length), conductivity cell, and autosampler A-900 (all GE Healthcare, Little Chalfont, Buckinghamshire, UK). The control software UNICORN 5.31 (GE Healthcare, Little Chalfont, Buckinghamshire, UK) was used. High-throughput batch chromatography experiments were conducted on a Freedom EVO 200 robotic platform equipped with a liquid handler, a robotic moving arm for transportation of plates, six fixed pipette tips, an orbital shaker, a vacuum separation module, and a photometer infinite M200 pro (all Tecan, Männedorf, Switzerland). The instrument was controlled by Freedom EVOware 2.5 software (Tecan, Männedorf, Switzerland).

2.3.2. Software

The kinetic adsorption isotherm was calibrated and simulated using the chromatography simulation software ChromX (GoSilico, Karlsruhe, Germany). ChromX provides numerical tools for discretization, solving of partial differential equation systems, as well as optimization and statistical analysis [115]. The equilibrium adsorption isotherm was calibrated and simulated with Matlab® R2015a (MathWorks, USA). fsolve was used to solve the nonlinear equation, and simulated annealing was used to estimate equilibrium isotherm parameters.

2.3.3. Adsorber, Proteins, and Chemicals

As a HIC adsorber medium, Capto$^{\text{TM}}$ Phenyl (high sub) supplied by GE Healthcare (Carlsbad, CA, USA) was used. For high-throughput batch chromatography experiments, adsorber medium was dispensed using MediaScout® ResiQuot provided by ATOLL (Weingarten, Germany), such that $20.8\,\mu\text{L}$ medium was packed in each well of MultiScreen HTS-DV supplied by Merck Millipore (Billerica, MA, USA). For linear-gradient elution experiments, a pre-packed $1\,\text{mL}$ MediaScout® miniChrom column with dimension $50\,\text{mm} \times 5\,\text{mm}$ provided by ATOLL (Weingarten, Germany) was used. Between the runs, the adsorber media were stored in $20\,\%$ ethanol. Before experimentation, the storage solution was removed by prolonged equilibration with water, low-salt and high-salt buffer.

Three proteins of different sizes ($160\,\text{kDa}$, $66\,\text{kDa}$ and $14.6\,\text{kDa}$) and different isoelectric points (acidic and alkaline) were used as model proteins to investigate the isotherm predictability. Glucose oxidase (*Aspergillus niger*, no. G7141) and BSA (bovine serum, no. A4612) were purchased from Sigma-Aldrich (St. Louis, MO, USA). Lysozyme (hen egg, no. HR7-110) was purchased from Hampton Research (Aliso Viejo, CA, USA).

To perform experiments near the isoelectric point of glucose oxidase and BSA, a $40\,\text{mM}$ 1-methylpiperazine buffer (Sigma-Aldrich, St. Louis, MO, USA) was used at pH 4.5 with additional 0 and $4\,\text{M}$ NaCl. For lysozyme, a $40\,\text{mM}$ CAPSO buffer (Sigma-Aldrich, St. Louis, MO, USA) was used at pH 10 with additional 0 and $4\,\text{M}$ NaCl. All solutions were prepared with ultra pure water (UPW) (arium pro UV, Sartorius, Göttingen, Germany). All buffers and storage solutions were filtrated with a membrane cutoff of $0.22\,\mu\text{m}$ and degassed by sonification. All proteins were prepared equal to the loading conditions and were filtrated with a membrane cutoff of $0.22\,\mu\text{m}$ prior to usage.

2.3.4. Extra Column Effects

The ÄKTA system and chromatography column were characterized via tracer injection at $60\,\text{cm/h}$. For determination of the system and total voidage of the column, $25\,\mu\text{L}$ $1\,\%(\text{v/v})$ acetone (Merck, Darmstadt, Germany) was used as a pore-penetrating, non-interacting tracer. Filtrated $10\,\text{g/L}$ dextran $2000\,\text{kDa}$ (Sigma-Aldrich, St. Louis, MO, USA) solution was used as non-pore-penetrating, non-interacting tracer to determine the bed voidage of the column. The UV signal at $300\,\text{nm}$ and the conductivity signal were recorded to determine these parameters and correct all measurements regarding dead volumes.

2.3.5. Bind-and-elute Experiments

Protein solutions of varying concentrations dissolved in corresponding binding buffer were loaded via a $500\,\mu$L or $1000\,\mu$L loop connected to the autosampler. Different gradients and steps were mixed from high-salt and low-salt buffer. A subset of the chromatograms was used to perform the inverse method for parameter estimation. The remaining runs were used to validate the calibrated model. The operating conditions were chosen manually from the ranges given in Table 2.1. All experiments were conducted at a linear flow of $60\,$cm/h to ensure a residence time of five minutes.

2.3.6. Parameter Determination

The bed voidage ϵ_b and particle voidage ϵ_p were calculated from the aforementioned tracer injections following the methodology presented by Hahn *et al.* [116]. The axial dispersion coefficient D_{ax} was calculated from the dextran peak broadening evaluated by UNICORN 5.31. The film transfer coefficient $k_{film,s}$ and the pore diffusion coefficient $D_{p,s}$ of salt were estimated based on the recorded conductivity signal. In the case of proteins, the film transfer coefficients $k_{film,i}$ and the pore diffusion coefficients $D_{p,i}$ were estimated, where the upper bound was set to $r_p/3$

Isotherm parameters of the kinetic formulation in Eq. (2.11) were also determined using the inverse method in combination with simulated annealing and the Levenberg-Marquardt algorithm. Since the adsorber medium and salt type did not change, β_0 and β_1 were estimated once for glucose oxidase and then directly adopted for BSA and lysozyme.

2.3.7. Model Validation

The model quality was assessed by comparison of simulation and experimental data. The 95 % confidence intervals of estimated parameters were calculated from the parameter covariance matrices, such that not only the accuracy, but also the reliability of the calibrated model could be evaluated. As β_0 and β_1 were kept constant

Table 2.1.: Ranges of operating conditions for column experiments used for model calibration and validation.

	Lower bound	Upper bound
Gradient start concentration [mM]	440	4020
Gradient end concentration [mM]	0	600
Gradient length [CV]	0	20

when modeling all three protein species, this reduction of the degrees of freedom underlines the legitimacy of the model assumptions. Furthermore, the calibrated isotherm parameters of β_0, β_1 and n were used to describe batch adsorption behavior as additional validity check.

2.3.8. High-throughput Batch Chromatography Experiments

High-throughput batch chromatography experiments were performed at 25 °C on the robotic platform as follows. The preparation of aliquots from Capto™ Phenyl (high sub) suspension was conducted by using the vacuum manifold device ResiQuot. After removal of storage solution from the filter plate by applying a vacuum of 300 mbar, the adsorber resin was washed five times with 250 μL UPW per well. The equilibration was carried out three times with 250 μL buffer of corresponding composition. Protein of varying concentrations dissolved in binding buffer was loaded into each well of the filter plate. Thereafter, a wash step, elution, and regeneration were performed. For all steps, the incubation was carried out on a shaking incubator at 300 rpm for 1 h. Vacuum at 300 mbar was applied to collect the supernatant in a UV 96-well MTP. Salt concentrations from 1 M to 0 M NaCl were applied in 1/9 M steps.

Isotherm fitting determined only the packing-dependent parameters k_{eq} and q_{max}. All other isotherm parameters, β_0, β_1 and n, were adopted directly from the parameter set obtained from chromatogram fitting.

2.4. Results

2.4.1. System Characterization

The system dead volume of 330 μL of the FPLC was determined by tracer injection without a chromatography column attached. This volume was subtracted from all other data obtained from the FPLC. Based on the results of the tracer injections with connected chromatography column, bed and particle voidage, and axial dispersion were calculated. The results are given in Tab 2.2.

Table 2.2.: Voidages and axial dispersion are calculated from retention volume and peak broadening of tracer injections.

		Capto$^{\text{TM}}$ Phenyl (high sub)
Bed voidage	ϵ_b	0.422
Particle voidage	ϵ_p	0.986
Total voidage	ϵ_t	0.992
Axial dispersion $[cm^2/s]$	D_{ax}	$5.6 \cdot 10^{-4}$

2.4.2. Parameter Estimation

The determined parameters and confidence intervals for all mass transfer and isotherm parameters are given in Tab. 2.3. First, adaptive simulated annealing was used for parameter estimation. Thereafter, a Levenberg-Marquardt algorithm was employed to find the final optimum.

The film transfer and pore diffusion coefficients for salt, $k_{film,s}$ and $D_{p,s}$, were estimated from the experimentally obtained conductivity data. Each protein-specific parameter set was then estimated from three bind-and-elute experiments.

Fig. 2.1 shows the simulated and experimental chromatograms used for model calibration for glucose oxidase (Figs. 2.1 a – 2.1 c), BSA (Figs. 2.1 d – 2.1 f) and lysozyme (Figs. 2.1 g – 2.1 i). Measurements have not been post-processed, and the simulated sum signal takes into account a linear baseline drift with a slope of $3.5\,\text{mAU/M}$ caused by salt gradients. To assess the predictive power of the calibrated models, two additional experiments have been conducted per protein as test set. Figs. 2.2 a – 2.2 f show the comparison of model prediction and experimental data. To compare the performance of the presented model with previously proposed salt-dependent isotherms, the three experiments with glucose oxidase were used to calibrate the models by Mollerup [58] and Deitcher *et al.* [98]. The quality of model fit was evaluated with the normalized root-mean-square error (NRMSE). The obtained NRMSE of the proposed new isotherm was 0.032, compared to 0.039 for Deitcher *et al.*, and 0.058 for Mollerup. An exemplary simulation result for all three isotherms is shown in Figs. 2.3 a – 2.3 c.

2.4.3. Equilibrium Adsorption Isotherm

The result of the batch chromatography experiments for glucose oxidase, BSA, lysozyme, and ten different salt concentrations between $1\,\text{M}$ and $0\,\text{M}$ in $1/9\,\text{M}$ steps, is shown in Fig. 2.4. The (c,q,c_s)-triplets were fitted with the equilibrium isotherm formulation shown in Eq. (2.8). As the general binding behavior is identical, the

isotherm parameters β_0, β_1, and n from Tab. 2.3 were kept constant for each protein species. Only k_{eq} and q_{max} were adjusted to account for the difference in skeleton-based and bed-volume-based capacity. The lines in Fig. 2.4 show the fitted adsorption isotherms. The resulting $k_{eq,batch}$ value is 0.129 and $q_{max,batch}$ equals $4.604 \cdot 10^{-4}$ M for glucose oxidase, $8.557 \cdot 10^{-2}$ and $1.609 \cdot 10^{-3}$ M for BSA, 0.502 and $1.705 \cdot 10^{-2}$ M for lysozyme.
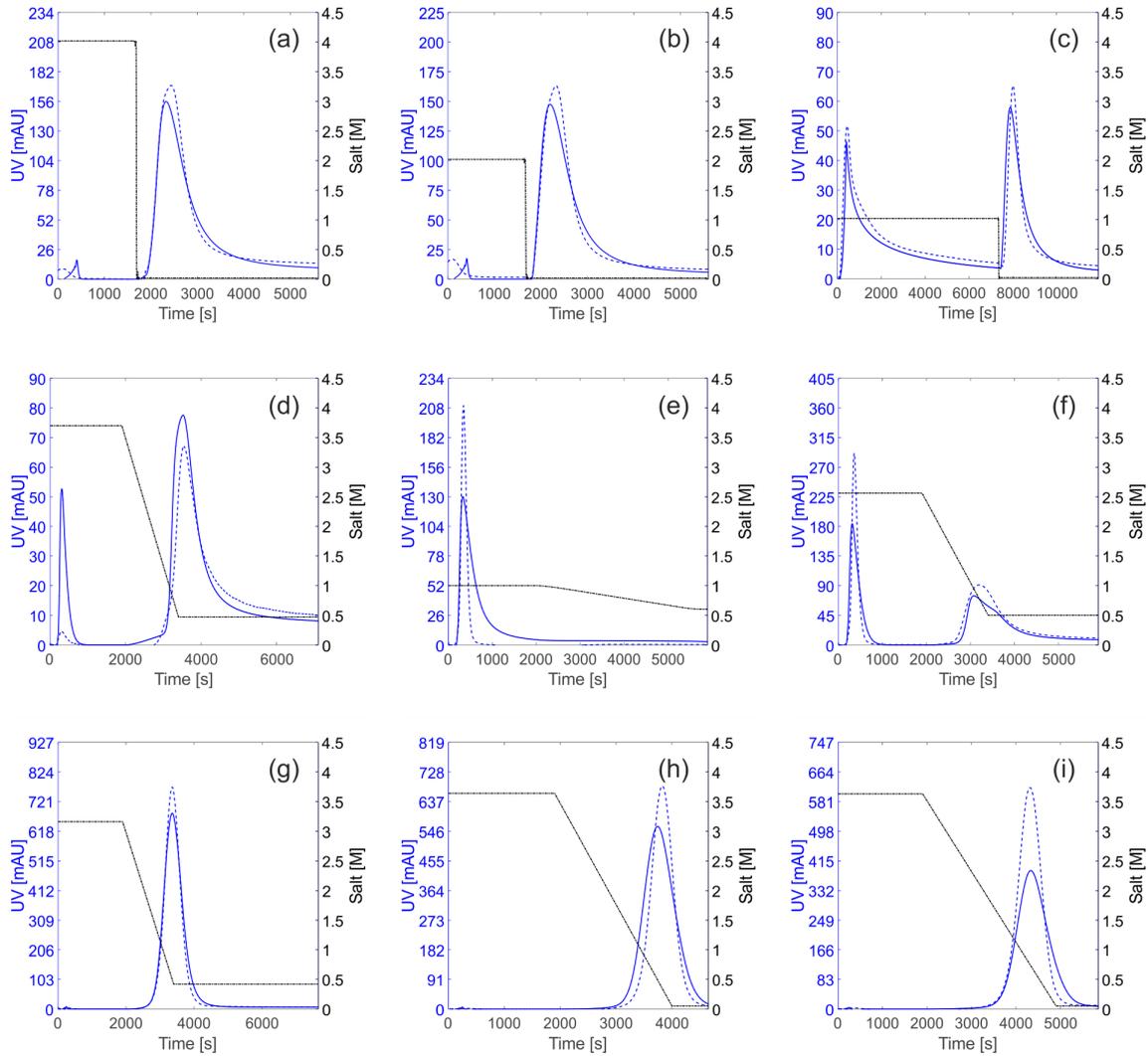
Figure 2.1.: Plots of UV signals over process run time for bind-and-elute experiments. Blue dashed lines display the UV signal measured at column outlet. Black dashed lines show the adjusted salt gradients. Solid lines represent the simulated chromatograms. Plots (a) – (c) show glucose oxidase within a salt concentration between $4.02\,\mathrm{M}$ and $0.02\,\mathrm{M}$, plots (d) – (f) BSA within a salt concentration between $3.7\,\mathrm{M}$ and $0.47\,\mathrm{M}$, plots (g) – (i) lysozyme within a salt concentration between $3.63\,\mathrm{M}$ and $0.051\,\mathrm{M}$. The normalized root-mean-square errors (NRMSE) are 0.032 for glucose oxidase, 0.045 for BSA, and 0.061 for lysozyme.
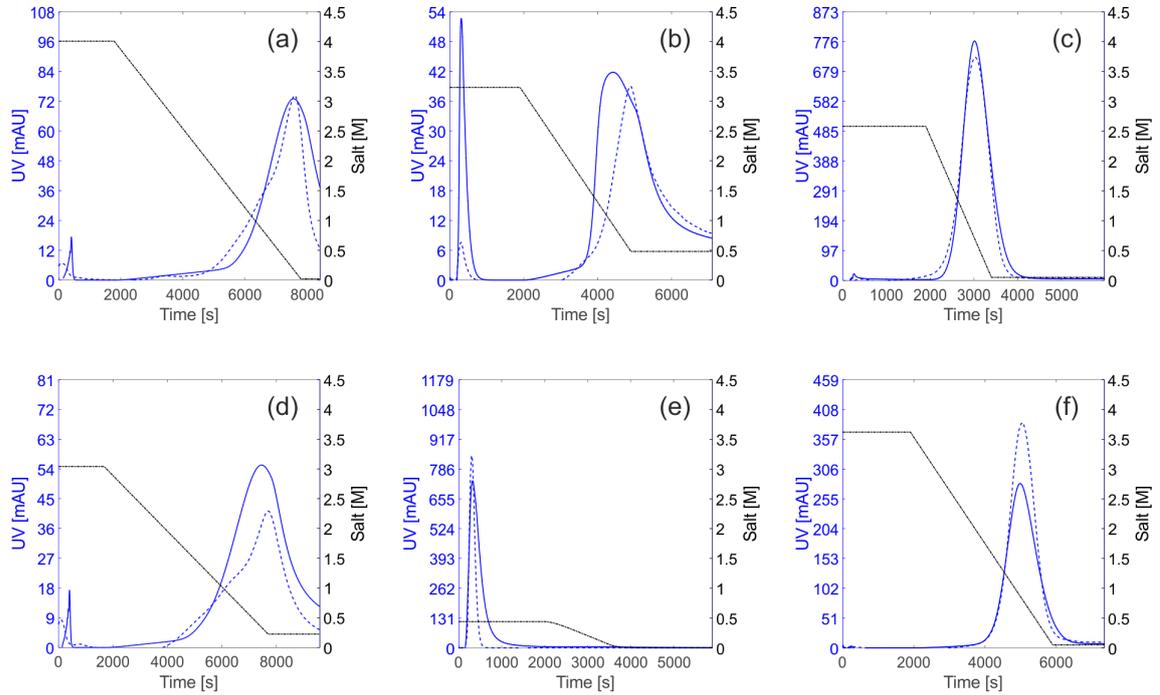
Figure 2.2.: Comparison of simulated UV signal and measurements. (a) and (d) represent glucose oxidase within a salt concentration between $4.02\,\mathrm{M}$ and $0.02\,\mathrm{M}$, (b) and (e) represent BSA within a salt concentration between $3.23\,\mathrm{M}$ and $0\,\mathrm{M}$, (c) and (f) the ones for lysozyme within a salt concentration between $4.02\,\mathrm{M}$ and $0.05\,\mathrm{M}$. The NRMSE of prediction are 0.150 for glucose oxidase, 0.080 for BSA, 0.089 for lysozyme.
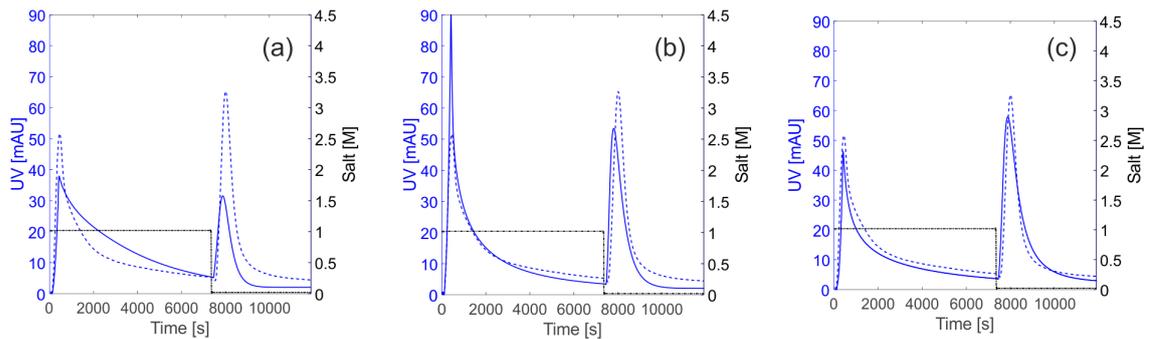
Figure 2.3.: Comparison of three salt-dependent isotherm models. Blue dashed lines display the UV signal measured at column outlet. Black dashed lines show the adjusted salt gradients. Solid lines represent the simulated chromatograms. Plot (a) shows the simulation result of glucose oxidase within a salt concentration between $1.02\,\mathrm{M}$ and $0.02\,\mathrm{M}$ using the isotherm model proposed by Mollerup[58], (b) the simulation result using the one proposed by Deitcher *et al.*[98], and (c) the result using the presented isotherm model. Fig. 2.3 (c) is identical with Fig. 2.1 (c) and shown here for comparability.
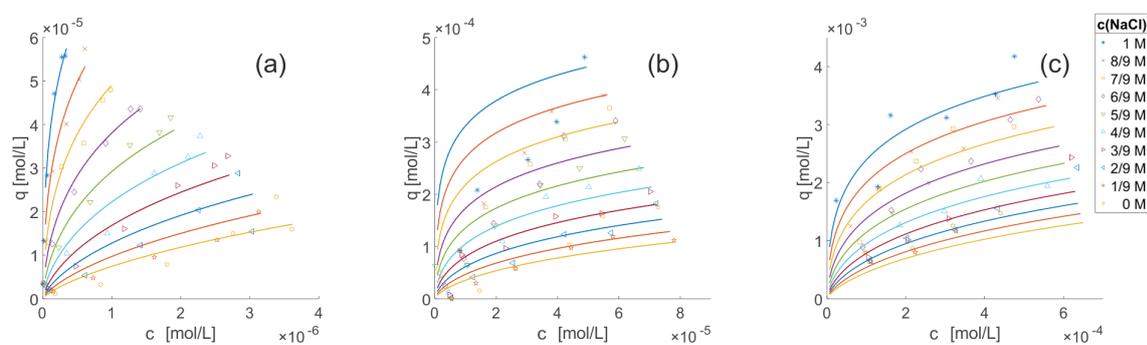
Figure 2.4.: Equilibrium adsorption data points from high-throughput experimentation and resulting equilibrium isotherms. Plot (a) shows glucose oxidase, plot (b) BSA, plot (c) lysozyme. The uppermost isotherms represent the adsorption behavior for 1 M NaCl. The salt concentration decreases in 1/9 M steps from top to bottom. The NRMSE are 0.266 for glucose oxidase, 0.093 for BSA, and 0.010 for lysozyme.

2.5. Discussion

As the results of the model calibration for glucose oxidase, BSA, and lysozyme show in Fig. 2.1, the overall model fit is very good. The glucose oxidase runs were used for comparing the model's performance with other salt-dependent isotherms. With a comparably small NRMSE the proposed new isotherm performed best, followed by Deitcher *et al.*[98], and Mollerup[58]. Despite of very different protein characteristics, peak area, shape and height of the elution peaks are described accurately. For BSA, the calibration was performed primarily by varying bind-and-elute salt concentration. For lysozyme, the gradient length was primarily varied. In the case of glucose oxidase, both bind-and-elute salt concentration and gradient length were varied strongly. Regardless of the differences in experimental design, a good identifiability of isotherm parameters is indicated by the remarkably small confidence intervals given in Tab. 2.3.

Once the model parameters given in Tab. 2.2 and 2.2, Eqs. 2.10 - 2.15 are determined, the mechanistic approach can be used to predict the protein adsorption/desorption behavior for any given salt concentration and gradient length. The model was validated by comparison between simulations and experimental test sets, that were not involved in parameter estimation. In accordance with the small confidence intervals, the chromatograms of the test sets could be predicted as depicted in Fig. 2.2 for all three protein species. In a second validation, batch chromatography data of each protein species was used to fit an equilibrium isotherm while keeping most parameters constant. Here, only parameters $k_{eq,batch}$ and $q_{max,batch}$ are expected to differ from the column ones, because of differences in packing as well as static and dynamic binding capacity. The two estimated isotherm parameters lead to

Table 2.3.: Parameters of the mass transfer model and kinetic isotherm formulation estimated from three bind-and-elute experiments each. β_0 and β_1 were determined based on glucose oxidase and adopted directly for BSA and lysozyme. Film transfer and pore diffusion coefficients of salt were estimated from the conductivity signals.

Parameter	Salt	Glucose oxidase	BSA	Lysozyme
$k_{film}\,[cm/s]\cdot 10^{-5}$	1.435	$5.378 \pm 3.458\cdot10^{-1}$	$9.9 \pm 6.410\cdot10^{-1}$	6.878 ± 3.684
$D_p\,[cm^2/s]\cdot 10^{-8}$	4.005	$2.799 \pm 1.797\cdot10^{-1}$	$3 \pm 7.512\cdot10^{-2}$	$3.675 \pm 7.095\cdot10^{-1}$
$k_{eq}\,[-]$	-	34.138 ± 12.940	264.457 ± 451.483	3.555 ± 4.174
$k_{kin}\,[s]\cdot 10^{-2}$	-	15.700 ± 8.500	4.389 ± 8.576	1.721 ± 12.5
$\beta_0\,[-]\cdot 10^{-2}$	-	$3.574 \pm 3.461\cdot10^{-1}$	3.574 ± 1.810	3.574 ± 2.354
$\beta_1\,[M^{-1}]$	-	1.001 ± 0.043	1.001 ± 0.215	$1.001 \pm 9.106\cdot10^{-3}$
$n\,[-]$	-	9.576 ± 1.839	10.845 ± 1.440	8.863 ± 6.958
$q_{max}\,[M]\cdot 10^{-2}$	-	$1.280 \pm 2.926\cdot10^{-1}$	$1.493 \pm 8.350\cdot10^{-2}$	48.100 ± 39.700

a very good agreement with the measurements. As to be expected, the $q_{max,batch}$ values of $4.604 \cdot 10^{-4}$ M, $1.609 \cdot 10^{-3}$ M and $1.705 \cdot 10^{-2}$ M are slightly larger than the corresponding column parameter values taken with respect to the whole column volume $q_{max,column}$ $1.024 \cdot 10^{-4}$ M, $1.194 \cdot 10^{-4}$ M and $3.848 \cdot 10^{-3}$ M. The maximum binding capacity of the whole column is given as $q_{max,column} = q_{max} \cdot (1 - \epsilon_t)$. Compared with the batch isotherm data for BSA and lysozyme obtained at neutral pH [98, 99], the results confirm the observation that an increase in HIC binding capacity can be achieved by processing near the protein's pI [86]. The related reduced solubility leads to scattered excessive deviations at higher concentrations of salt and protein.

The parameters β_0 and β_1, describing bulk-like water molecules involved in the adsorption/desorption process, are independent of the protein species. This indicates that the effect of the water structure on the protein surface in investigated cases is not relevant compared to the contribution of water molecules on the adsorber surface. Once they are determined for a particular experimental set-up consisting of a certain adsorber medium, buffer system and salt type, a direct adoption to other proteins can be legitimate. The analogy to the widely employed SMA isotherm for IEX explains this aspect. In SMA the characteristic charge depicts the number of ligands involved in the electrostatic interaction. Here, the "strength" of every single ligand is assumed to be constant and not dependent on the binding site. Similarly, β_0 can be assumed to be dominated by the ligand characteristics and independent from the protein structure. When modeling multi-component systems with m protein-species involved, the benefit is obvious. Since only 4 isotherm parameters are protein specific, the number of protein-specific parameters is reduced by $2m$ compared to Mollerup [58] and Deitcher *et al.* [98].

The magnitude of maximum binding capacity q_{max} shows good correlation with the size of the corresponding protein. Due to the relatively high equilibrium coefficient k_{eq} of BSA, the hydrophobic binding affinity seems to be very strong. This observation fits to findings reported in the literature [117]. Since this isotherm model does not take the irreversible binding of BSA described by Haimer *et al.* into account [95], the prediction shows slightly more deviation here.

Although assumptions and simplifications were made during the isotherm derivation to balance ease of use with accuracy and reliability, validity and predictability are backed up by experimental data performed under very different operating conditions. To the best of our knowledge, this is the first work to describe and predict kinetic bind-and-elute and batch chromatography experiments in HIC with largely identical parameter sets.

2.6. Conclusion

In this work, the insights into the water structure on a hydrophobic surface are used to develop a mechanistic model for the adsorption behavior of proteins in HIC.

The present approach proposes an equilibrium between well-ordered and bulk-like ordered water molecules on the hydrophobic surface. This equilibrium is described for a constant buffer composition, in particular without change of ionic concentration over time. The challenge of incorporating salt effects in HIC modeling is solved by considering the hydration number of ions as a function of the salt concentration. This affects the equilibrium of water structures reciprocally. The parameter identifiability and reliability are demonstrated by bind-and-elute experiments in column format, application of the inverse method and calculation of the confidence intervals for the estimated parameters. High-throughput batch chromatography experimentation in combination with a largely identical isotherm parameter set underlines the suitability of the isotherm model.

For further studies, it is possible to incorporate other changes in buffer composition to describe protein-ligand adsorption behavior in HIC more generally. Mechanistic extensions of the presented new thermodynamic framework are conceivable, for example to model pH dependence, irreversible binding, protein-spreading, and multilayer adsorption.

2.7. Acknowledgment

3. Water on Hydrophobic Surfaces: Mechanistic Modeling of Protein Precipitation Using Polyethylene Glycol

Steffen Großhans*, Gang Wang*, and Jürgen Hubbuch**

Institute of Engineering in Life Sciences, Section IV: Biomolecular Separation Science, Karlsruhe Institute of Technology (KIT), 76131 Karlsruhe, Germany
* These authors contributed equally to this work.
** Corresponding author.

Abstract

For the purification of biopharmaceutical proteins, liquid chromatography is still the golden standard. Especially with rising product titers, drawbacks like slow volumetric throughput and high resin costs lead to the requirement of alternative technologies. Selective preparative protein precipitation is one promising alternative technique. Although the capability was proved, there is no precipitation process for large scale monoclonal antibody (mAb) production yet. One reason might be, that the mechanism behind protein phase behavior is not completely understood and the precipitation process development is still empirical.

Mechanistic modeling can provide faster, material saving process development and a better process understanding at the same time. In preparative chromatography, mechanistic modeling was successfully shown for a variety of applications. Latest, a new isotherm for hydrophobic interaction chromatography (HIC) under consideration of water molecules as participants was proposed, enabling an accurate description of HIC.

In this work, based on similarities between protein precipitation and HIC, a new precipitation model was derived. In the proposed model, the formation of protein-protein

interfaces is thought to be driven by hydrophobic effects, involving a reorganization of the well-ordered water structure on the hydrophobic surfaces of the protein-protein complex. To demonstrate model capability, high-throughput precipitation experiments with the model proteins lysozyme, myoglobin, bovine serum albumin (BSA), and one mAb were conducted at various pH values. Polyethylene glycol (PEG) 6000 was used as precipitant. The precipitant concentration as well as the initial protein concentrations were varied systematically. The calibrated models were finally validated with experimental data.

3.1. Introduction

Biologics represent a growing share of the pharmaceutical market, reached global sales of USD 228 billion in 2016 [118]. Among them, monoclonal antibodies (mAb) are the most important family of products [119]. The fact, that many mAbs have a relatively low potency requiring high doses, makes mAbs one of the most expensive drugs [120]. Since 2015 the patent of first generations mAbs began to expire, resulting in a number of biosimilars approved in the US and Europe [121]. Together with increasing pressure on healthcare budgets, cost savings are desired [10]. Therefore improvements in downstream processes such as alternative methods or novel development strategies are necessary [122].

Selective protein precipitation is known as a cost efficient alternative purification step for a long time [123]. Phase separation is carried out by adding precipitation agents, like inorganic salts, organic solvents, or nonionic polymers to the protein solution [124–127]. For some biopharmaceutical products precipitation is already well established. For example, ethanol or polyethylene glycol (PEG) precipitation is the basis of the extraction of immunoglobulin G from human plasma [128]; Viral vaccines and virus-like particles (VLP) are purified or concentrated through PEG or salt precipitation [129, 130]. Although there are a lot of studies on precipitation of recombinant mAbs as well, it has not been applied for large scale mAb production yet [131–133].

For downstream process (DSP) development mechanistic understanding is needed to meet the demands of the Quality by Design (QbD) approach suggested by the US Food and Drug Administration (FDA) [18]. For precipitation the understanding of protein phase behavior is mandatory [134]. Although protein phase behavior is experimental well investigated, the mechanism behind precipitation has not been completely understood. This leads to a high degree of freedom in process development and makes it challenging.

Models and simulations can reduce the number of experiments and lead to a more thorough process understanding [135]. Cohn *et al.* derived an equation to describe

protein precipitation [65]. This equation is a useful empirical expression, but less a mechanistic description [136]. Its parameters were specified for salt precipitation and its application to precipitation using polymers was shown [68, 137]. Sim *et al.* generalized the model on the basis of hydrodynamic radii [66]. Quantitative structure activity relationship (QSAR) was used to estimate the parameters of the Cohn equation [67]. Anyhow, all these models have struggle with dealing small molecular weight proteins and are not capable to cover variations in the initial protein concentration.

PEG induced precipitation can be described by the theory of excluded volume [138–140]. According to this theory, adding a certain amount of PEG to a protein solutions results in phase transition of the proteins, although there is no direct interaction between polymer and protein. The polymers are reported to trap the solvent and therefore, sterically excludes proteins from solvent regions occupied by the polymers. In other words, the polymers and proteins compete for the solvent which they are solved in. Electrostatic interaction is known to have mainly repulsive influence on protein-protein interaction, so that attractive forces can be reduced to hydrophobic effects [141]. On molecular level, water molecules next to hydrophobic surfaces are thought to have a well-ordered structure, different to the bulk-like ordered water in free solution [142–144]. During precipitation a rearrangement of the protein occurs and the hydrophobic area is reduced. Simultaneously, the water structure has to be reorganized. A new equilibrium between well-ordered and bulk-like water is given, resulting in an increased entropy [97, 101]. The above described hydrophobic effect is thought to occur between two proteins in precipitation, as well as between protein and ligand in hydrophobic interaction chromatography (HIC). Similarities between HIC and precipitation were already shown. Melander and Horváth investigated the salt effect on the hydrophobic effect in precipitation and HIC [145]; Nfor and coworkers studied interrelation between the number of released water molecules in protein precipitation and HIC [146]; and Baumgartner and coworkers investigated the retention behavior during HIC and its correlation to protein-protein interaction [147]. The mechanistic effects of HIC are well investigated [58, 148, 149].

In this study a mechanistic protein precipitation isotherm model was derived, inspired by recent results in HIC modeling [150]. Based on existing precipitation theories, water was introduced as an additional component for the model building. With the help of high-throughput experimental technique precipitation, data for lysozyme, myoglobin, bovine serum albumin (BSA), and a mAb were generated. After randomization, half of the date were used to calibrate the model, the other was used as validation set (Fig. 3.1). The model predictability for a wide range of properties such as size, hydrophobicity, and the isoelectric point (pI) was shown.
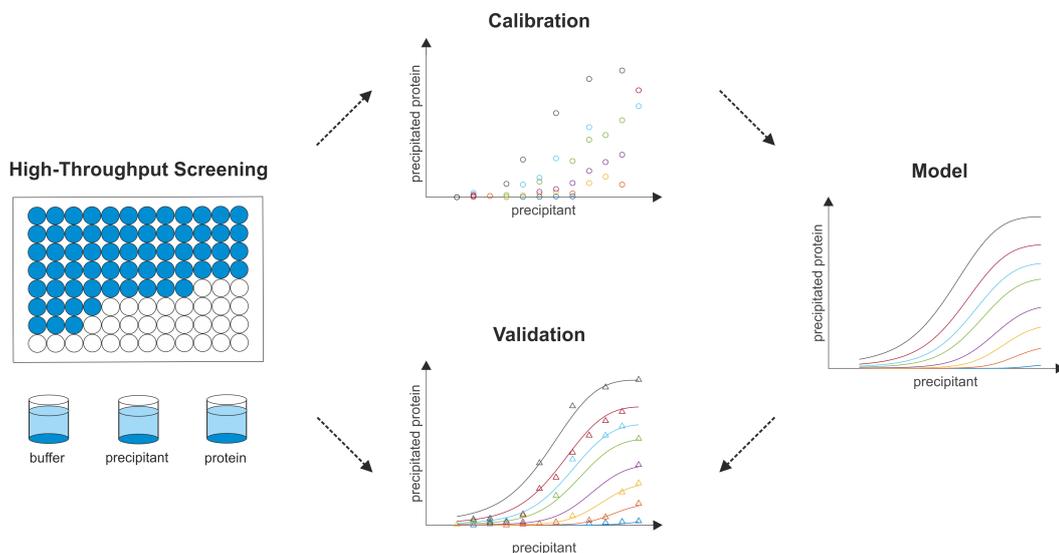
Figure 3.1.: Mechanistic protein precipitation modeling. By varying the amount of buffer, protein and precipitant stock solutions the precipitant and protein concentration were varied in high-throughput experiments. After phase separation the protein concentration was detected using UV 280 measurement. Half of the data were used as calibration set. With this data the parameters of the model were estimated. The so generated model was validated with the other half of the experimental data.

3.2. Theory

In 1925 Cohn and coworkers introduced a later widely excepted semi-logarithmic equation for modeling protein precipitation:

$$log(\tilde{s}) = \tilde{\alpha}\tilde{m} + \tilde{\beta}. \tag{3.1}$$

The useful empirical equation describes the protein solubility $\tilde{s}$ in mg/mL in the presence of $\tilde{m}$ PEG in $\% \, w/w$. Here, the phase behavior of a constant protein concentration depends on the precipitation efficiency $\tilde{\alpha}$ and the protein solubility in the absence of PEG $\tilde{\beta}$. This equation represents a summary of a macroscopic observation on the precipitation behavior.

In the present work, the focus is placed on a mechanistic level, especially on the behavior of water molecules during protein precipitation. In the model assumption, the precipitation mechanism of n protein molecules P by a PEG molecule PEG forming the precipitate P_{nPEG} is considered. The hydrophobic surfaces of proteins are thought to be stabilized by well-ordered water molecules. The precipitate P_{nPEG}

is assumed to be stabilized by β bulk-like ordered water molecules W_B:

$$nP + PEG \longleftrightarrow P_{nPEG} + n\beta W_B. \tag{3.2}$$

Precipitation using PEG is known to be a fast process [151]. Thus, following equilibrium was considered:

$$\Delta\mu = \mu_{P_{nPEG}} + n\beta\mu_{W_B} - n\mu_P - \mu_{nPEG}. \tag{3.3}$$

The chemical potentials are deduced to apply this constraint. It was assumed, that the protein surface charges can be considered negligible in the present hydrophobically driven mechanism.

Considering the equilibrium $\Delta\mu = 0$ at constant temperature and pressure, it is

$$
\begin{aligned}
RT \ln x_{P_{nPEG}} \quad &+ \quad n\beta RT \ln x_{W_B}\gamma_{W_B} - nRT \ln x_P\gamma_P - RT \ln x_{PEG}\gamma_{PEG} \tag{3.4}\\
&= \quad -\mu^0_{P_{nPEG}} - n\beta\mu_{W_B} + n\mu_P + \mu_{PEG} \tag{3.5}\\
&= \quad -\Delta G^0 = RT \ln K. \tag{3.6}
\end{aligned}
$$

To simplify the model equation, it is assumed the activity coefficient of protein, PEG, and precipitate to be constants. The equilibrium constant K is derived

$$K = \frac{x_{P_{nPEG}}x_{W_B}^{n\beta}}{x_P^n x_{PEG}} \Rightarrow K = \frac{q_{P_{nPEG}}a_{W_B}^{n\beta}}{c_P^n c_{PEG}} \tag{3.7}$$

with $q_{P_{nPEG}}$ and c_P are the precipitated protein and protein in solution, respectively. c_{PEG} depicts the concentration of PEG in solution and a_{W_B} the activity of the bulk-like ordered water molecules. In the following step, parameterizations for a_{W_B} and β have to be found. According to stoichiometric considerations, the number of water molecules involved is linearly correlated to the precipitated protein $q_{P_{nPEG}}$. Thus, a linear correlation is proposed to substitute a_{W_B}:

$$a_{W_B}^{n\beta} \quad \cong \quad \nu q^{n\beta}. \tag{3.8}$$

The stoichiometric constant ν is assumed to be independent of the PEG concentration. Inserting Eq. 3.8 into Eq. 3.7, and collecting all constants on the left-hand site, the following isotherm equation is obtained:

$$K \quad = \quad \frac{\nu q^{1+n\beta}}{c_p^n c_{PEG}} \tag{3.9}$$

$$\xrightarrow{\cdot\frac{1}{\nu}}: k_{eq} \quad = \quad \frac{q^{1+n\beta}}{c_p^n c_{PEG}}. \tag{3.10}$$

Finally, the PEG and protein dependency of the bulk-like ordered water molecules β

is modeled. Since hydration of PEG, protein, and salt ions have high similarities e.g., attracting water molecules to form a hydration shell, the model originally describing the hydration number of the salt ions h is employed [152]:

$$h = h_0 \exp(-kc_s) \tag{3.11}$$

h_0 is the ionic hydration number at infinite dilution and k the constant that accounts the dependency of hydration number on ionic concentration c_s. β and h are assumed to be reciprocal, so that the model parameter β can be approximated by the exponential term:

$$\beta = \beta_0 \exp(\beta_1 c_{PEG} + \beta_2 c_p) \tag{3.12}$$

β_0 is the hydration number at infinite dilution of PEG and protein, whereas β_1 and β_2 the constants that account for the dependency of hydration number on PEG and protein concentration, respectively. This completes the derivation of the equilibrium formulation of the precipitation isotherm model.

3.3. Materials and Methods

3.3.1. Disposables

All precipitation experiments were carried out in $350\,\mu\text{L}$ polypropylene flat bottom 96-well micro plates (Greiner Bio-One, Kremsmünster, Austria). For spectroscopic measurements samples were diluted into Greiner UV-Star® micro plates (Greiner Bio-One, Kremsmünster, Austria).

3.3.2. Chemicals and Stock Solutions

As buffer substances, sodium hydrogen carbonate and tris (hydroxymethyl)-aminomethane (all Merck KGaA, Darmstadt, Germany) were used. Tris hydrochloride was obtained from PanReac AppliChem (Darmstadt, Germany). Sodium carbonate was obtained from Sigma Aldrich (St. Louis, MO, USA). The PEG with an average molecular mass of 6000 was obtained from Merck KGaA (Darmstadt, Germany). All buffers were prepared with a concentration of $50\,mM$. For this, the appropriate amounts of associated buffer components were weighted and dissolved in ddH_2O. The desired pH was reached by varying the amount of acid and basic component for each buffer. For the $40\,\%\,(\text{w/w})$ PEG 6000 and $50\,\%\,(\text{w/w})$ PEG 6000 stock solution, the buffer components were first dissolved in ddH_2O followed by adding the appropriate amount of PEG 6000.

3.3.3. Preparation of Protein Stock Solutions

Lysozyme from chicken egg white was purchased from Hampton Reseach (Aliso Viejo, CA, USA). Myoglobin and BSA were purchased from Sigma Aldrich (St. Louis, MO, USA). The mAb was provided as purified mAb from LEK d.d. (Ljubljana, Slovenia). Lyzoszym, myoglobin and BSA were provided as lyophilized powder and therefore first solved in the appropriate buffer. Afterward all proteins including the mAb were filtered using $0.2\,\mu$m syringe filters (Satorius, Göttingen, Germany). Following the filtration, proteins were rebuffered and desalted into the associated buffer using PD 10 desalting columns (GE Healthcare, Little Chalfont, UK).

3.3.4. Generation of Precipitation Curves

All precipitation experiments were carried out on a Tecan Freedom Evo 200 System liquid handling station (Tecan, Männedorf, Switzerland). The liquid handling station was equipped with an 8-tips liquid handling arm, a robotic manipulator arm, a Te-Shake orbital shaker, an Infinite® 200 UV-Vis spectrophotometer (all Männedorf, Switzerland), and a Rotanta 46RSC centrifuge (Hettlich GmbH & Co. KG, Tuttlingen, Germany). The system was controlled by Evoware 2.5 (Tecan, Männedorf, Switzerland). Excel 2016 (Microsoft, Redmond, WA, USA) was used as data import format and for data storage. All calculations were done using Matlab® R2016a (The Mathworks, Natick, MA, USA). All experiments were carried out at $20°C$, controlled by air conditioning. Systems with a total volume of $250\,\mu$L containing varying protein and PEG concentration were prepared. The PEG concentration was varied in 12 equidistant steps. The protein concentration was varied from $1.5\,mg/mL$ to $12.0\,mg/mL$ in 12 steps. The position for each system on the 96-well micro plate was randomized. After adding the protein stock solution, the system was incubated for $15\,min$ on the orbital shaker at $1000\,rpm$, and afterwards $15\,min$ without shaking. To analyze the amount of precipitated protein, the microplate was centrifuged for $30\,min$ at $4000\,rpm$. Then, the supernatant was sampled and diluted at a ratio of 1:6 for lysozyme, 1:3 for BSA, and 1:4 for myoglobin and the mAb. Subsequently, UV-Vis absorption at $280\,nm$ was measured. The protein concentration was calculated with based on a linear calibration curve. All experiments were conducted in triplicates.

3.3.5. Numerical Procedures

The equilibrium precipitation isotherm model Eqs. 3.10 and 3.12 proposed in the previous section contain unknown parameters, which cannot be measured experimentally. Model calibration and simulation were carried out in Matlab® R2016a (The Mathworks, Natick, MA, USA). To solve the nonlinear equation, *fsolve* was used. The heuristic algorithm simulated annealing *simulannealbnd* was employed

to deliver the initial parameter estimates. Afterwards, the deterministic algorithm Levenberg-Marquardt *nlinfit* was used to find the final parameter estimates for k_{eq}, n, β_0, β_1, and β_2.

The data generated in high-throughput experimentation were labeled randomly with either 0 or 1 using *randi*. According to the labels, the data were split into equally sized calibration data and validation data sets of 48 samples each. The former ones were used to proof the description capability and accuracy of the suggested model; The latter were compared to the model prediction to back up the model's accuracy and predictability.

3.4. Results and Discussion

3.4.1. Model Calibration

All precipitation data from high-throughput experimentation were randomized and divided into two equally sized data sets of 48 samples each. First, simulated annealing was used for parameter estimation with the calibration data sets. Thereafter, a Levenberg-Marquardt algorithm was employed to find the final optimum. The equilibrium coefficient k_{eq}, the number of proteins affected by one PEG molecule n, the hydration number at infinite dilution β_0, and the constants accounting for the dependency of hydration number on PEG β_1 and protein concentration β_2 for lysozyme, myoglobin, BSA, and mAb at pH 7.5 and pH 8.5 are given in Tab. 3.1. Here, the natural logarithm of the k_{eq} is presented for a better overview.

Fig. 3.2 shows the model simulation as solid curves and the experimental data used for model calibration as circles for lysozyme (Fig. 3.2 a), myoglobin (Fig. 3.2 b), BSA (Fig. 3.2 c), mAb at pH 7.5 (Fig. 3.2 d) and pH 8.5 (Fig. 3.2 e). In all cases, the results cover 8 protein concentrations to $12\,mg/mL$ from $1.5\,mg/mL$ with steps of

Table 3.1.: Parameters of the precipitation model estimated from the calibration high-throughput experimental data. The natural logarithm of the equilibrium coefficient k_{eq} is presented for a better overview.

parameter	lysozyme	myoglobin	BSA	mAb 7.5	mAb 8.5
$\ln k_{eq}\,[-]$	31.61	16.13	12.48	18.46	19.83
$n\,[-]$	5.32	3.73	3.18	3.28	3.53
$\beta_0\,[-]$	$3.33{\cdot}10^{-3}$	$3.79{\cdot}10^{-4}$	$4.64{\cdot}10^{-4}$	$1.12{\cdot}10^{-3}$	$1.03{\cdot}10^{-2}$
$\beta_1\,[L/mol]$	$1.16{\cdot}10^{5}$	$1.24{\cdot}10^{5}$	$1.92{\cdot}10^{5}$	$4.53{\cdot}10^{5}$	$4.38{\cdot}10^{5}$
$\beta_2\,[L/mol]$	$1.40{\cdot}10^{3}$	$1.46{\cdot}10^{3}$	$6.82{\cdot}10^{2}$	$7.56{\cdot}10^{3}$	$1.62{\cdot}10^{3}$

$1.5\,mg/mL$. The investigated range of PEG concentration varies according to the precipitation behavior of every protein. For lysozyme, myoglobin, and BSA, PEG concentrations up to $5.6{\cdot}10^{-5}\,mol/L$ are shown. The mAb precipitated at lower PEG concentration, so that PEG concentrations up to $2{\cdot}10^{-5}\,mol/L$ are presented. Despite
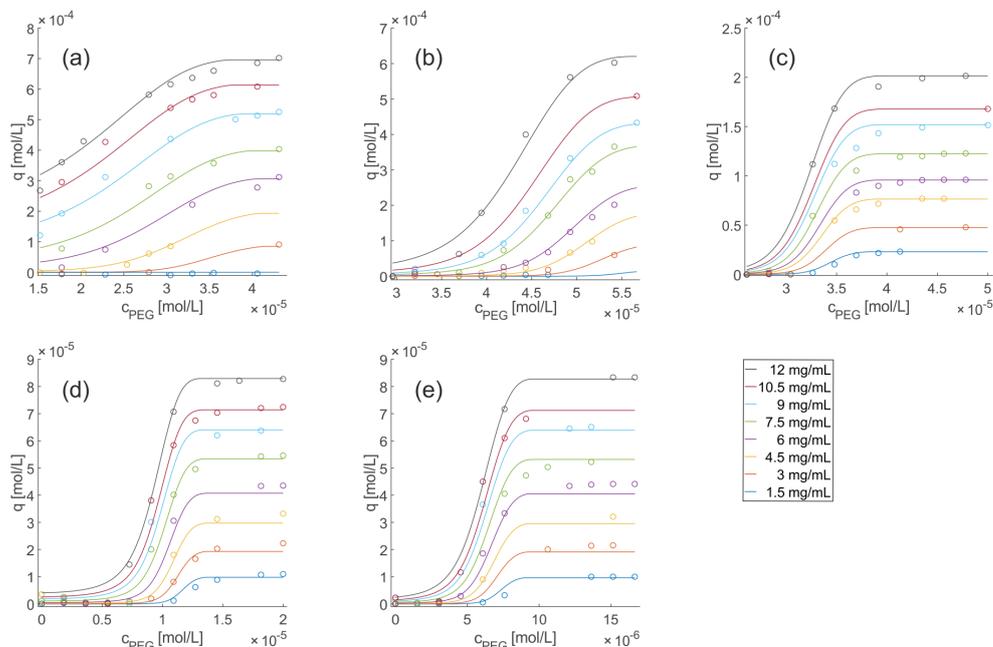
Figure 3.2.: Protein precipitation data points from high-throughput experimentation used for model calibration (circles) and resulting precipitation model (solid lines). Plot (a) - (e) shows lysozyme, myoglobin, BSA, and mAb at pH 7.5 and pH 8.5, respectively.

of very different protein characteristics such as size, hydrophobicity, and surface charge distribution, the calibration data sets are described by the precipitation model accurately in all cases. Regardless of the differences in experimental conditions such as PEG concentration and pH value, a good identifiability of precipitation model parameters is observed. As the smallest protein investigated, lysozyme ($14.6\,kDa$) shows the highest number of proteins affected by one PEG molecule with $n = 5.32$. The slightly larger protein myoglobin ($17\,kDa$) shows the next highest n with 3.73. For all other proteins the n values were similar, compared to those of myoglobin.

The hydration number constants β_0, β_1, and β_2 of each protein, influence the β function in the same order of magnitude, compared to each other. β_1 accounts for the dependency of hydration of PEG. Consequently, the behavior can be reconciled to the theory of excluded volume. By adding PEG, the accessible water for the protein is reduced. As this exclusion is caused by a steric phenomena, the influence can be attributed to protein size. A linear correlation between β_1 and the molecular weight is observed, in accordance with the linear correlation of precipitation behavior and the hydrodynamic radius of the protein for PEG as precipitant reported by Sim *et al.*. Furthermore, Hämmerling *et al.* confirmed this assumption, but pointed out the

influence of additional factors, like protein shape and other surface characteristics. The similarity of the β_1 values determined for the mAb at pH 7.5 and pH 8.5, supports this assumption.

In case of mAb, a pH shift toward its pI (pH 8.3-8.5) to pH 8.5 from pH 7.5 leads to increased n, k_{eq}, and β_0, resulting in earlier precipitation of mAb. At the same time a decrease of β_2 was observed. While β_1 could be assigned to the protein size, the nature of β_0 and β_2 appears to be more complex. The pH dependence of these parameters suggests a correlation to the surface characteristics of protein. The pH dependency of the hydration number β_0 is consistent with the results delivered by Xia and coworkers [153] that the water release increases as the buffer pH approaches the protein's pI in HIC. The hydration number parameter β_2 is referred to the influence of the protein concentration. Closer to the pI, this influence appears to be less important. In the absence of electrostatic interactions the well-ordered water conformation is reported to be less stable and therefore it is favored to set free well-ordered water molecules [154].

3.4.2. Model Validation

To assess the predictive power of the calibrated precipitation models, the validation data sets were used. Fig. 3.3 shows the model prediction as solid curves and the experimental data excluded from model calibration as triangles for lysozyme (Fig. 3.3 a), myoglobin (Fig. 3.3 b), BSA (Fig. 3.3 c), mAb at pH 7.5 (Fig. 3.3 d) and pH 8.5 (Fig. 3.3 e). A very good prediction can be found especially for myoglobin, BSA, mAb at both pH values, and lysozyme at higher PEG concentration. For PEG concentrations below $2.5 \cdot 10^{-5} \, mol/L$ however, the model prediction tends to underestimate the precipitated amount of lysozyme. Although simplifications and assumptions were made during the precipitation model derivation to balance accuracy with ease of use, predictability and validity are backed up by high-throughput experimental data of very different proteins under diverse operating conditions. The main difference between the widely used models, such as Cohn equation and models related to it, and the proposed precipitation model is its mechanistic nature and capability to describe the protein precipitation process dependent on both PEG and protein concentration.
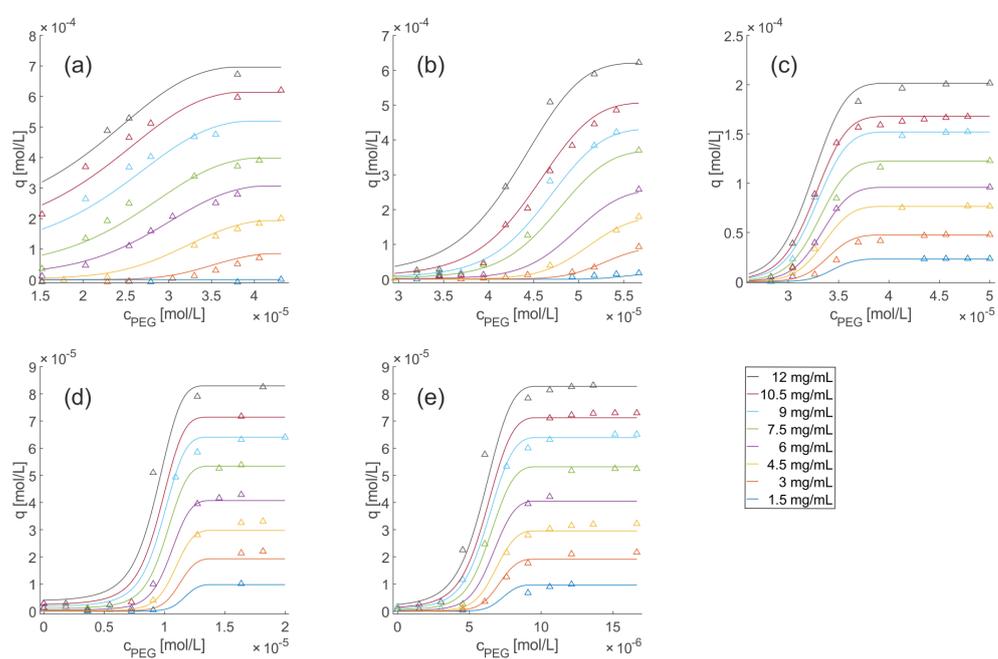
Figure 3.3.: Comparison of model prediction (solid lines) and high-throughput experimental data (triangles). (a) represents lysozyme, (b) myoglobin, (c) BSA, (d) mAb at pH 7.5, and (e) mAb at pH 8.5. The solid lines are identical with Fig. 3.2 and shown here for comparability.

3.5. Conclusion

In the presented work, a mechanistic model for protein precipitation behavior with PEG was introduced by considering the insights into the water structure on a hydrophobic surface. The present approach proposes the equilibrium between well-ordered and bulk-like ordered water molecules on the hydrophobic surfaces of protein as the driving force for the precipitation process. This equilibrium is described for a constant buffer composition, in particular without change of pH and PEG type.

In further studies, the dependency of model parameters on changes of pH and PEG type should be investigated to enhance the mechanistic understanding of protein precipitation with PEG. The applicability of the suggested model to protein precipitation with salt should be tested in a systematic manner. Multi-component systems such as harvested cell culture fluid should be described to enable model-based optimization of selective precipitation processes.

3.6. Acknowledgment

4. Estimation of Adsorption Isotherm and Mass Transfer Parameters in Protein Chromatography Using Artificial Neural Networks

Gang Wang[1], Till Briskot[1], Tobias Hahn[2], Pascal Baumann[1], Jürgen Hubbuch[1,*]

[1] Karlsruhe Institute of Technology (KIT), Institute of Process Engineering in Life Sciences, Section IV: Biomolecular Separation Engineering, Karlsruhe, Germany
[2] GoSilico GmbH, Karlsruhe, Germany
* Corresponding author. E-mail address: juergen.hubbuch@kit.edu

Abstract

Mechanistic modeling has been repeatedly successfully applied in process development and control of protein chromatography. For each combination of adsorbate and adsorbent, the mechanistic models have to be calibrated. Some of the model parameters, such as system characteristics, can be determined reliably by applying well-established experimental methods, whereas others cannot be measured directly. In common practice of protein chromatography modeling, these parameters are identified by applying time-consuming methods such as frontal analysis combined with gradient experiments, curve-fitting, or combined Yamamoto approach. For new components in the chromatographic system, these traditional calibration approaches require to be conducted repeatedly.

In the presented work, a novel method for the calibration of mechanistic models based on artificial neural network (ANN) modeling was applied. An *in silico* screening of possible model parameter combinations was performed to generate learning material

for the ANN model. Once the ANN model was trained to recognize chromatograms and to respond with the corresponding model parameter set, it was used to calibrate the mechanistic model from measured chromatograms. The ANN model's capability of parameter estimation was tested by predicting gradient elution chromatograms. The time-consuming model parameter estimation process itself could be reduced down to milliseconds. The functionality of the method was successfully demonstrated in a study with the calibration of the transport-dispersive model (TDM) and the stoichiometric displacement model (SDM) for a protein mixture.

4.1. Introduction

Preparative liquid chromatography is the workhorse for the purification purposes of today's pharmaceutical and biopharmaceutical industry [14, 15]. Several methods for chromatography process development are proposed in the literature, including purely experimental approaches which are time-consuming and labor-intensive. High-throughput approaches have improved such experimental development setups, but still these methods are often highly ineffective [16].

Especially in the light of the Quality by Design (QbD) initiative [17], mechanistic modeling is currently the most promising method in protein chromatography process development as it allows to gain a high degree of process understanding [18, 19]. Successful studies have been conducted for conventional single-column batch chromatography [20, 21], and continuous multi-column chromatography processes [22]. Furthermore, model-based process control strategies are enabled [23–25].

The chromatographic models consist of mathematical equations describing the mass transfer phenomena within the column on macroscopic level and the thermodynamics on the adsorber surfaces. For the mass transfer part, the transport-dispersive model (TDM) is often chosen because of its simplicity and wide acceptance in the literature [21, 41]. The thermodynamic binding term can be described by different isotherm models. Among them, the stoichiometric displacement model (SDM) is known to be well suited for ion-exchange chromatography (IEX) [56]. It describes the multi-point protein binding in IEX based on the electrostatic equilibrium theory and displacement effect, such that the salt-dependent protein-ligand interaction is included. Brooks and Cramer introduced the popular steric mass-action (SMA) model as an extension to cover the steric effects on the adsorber surface [90].

While modeling has several advantages over the experimental approach, it also possesses several limitations. This includes mainly the need for profound mathematical and thermodynamic understanding. Already in the early stage of model calibration, the hurdle becomes noticeable: Some parameters, such as system characteristics, can be determined reliably by applying well-established experimental methods. Others, like the mass transfer and isotherm parameters, cannot be measured directly

but are estimated by applying time-consuming methods such as frontal analysis combined with gradient experiments [90, 155], curve-fitting [116, 150], or combined Yamamoto approach [156]. For the linear region of adsorption isotherm, Parente and Wetlaufer proposed a correlation between retention parameters and gradient elution retention data [157]. These traditional calibration approaches, however, require to be conducted repeatedly if new components are introduced in the chromatographic setup.

In other research fields, the artificial neural network (ANN) modeling has been applied to obtain an initial guess for inverse problems in heat and radiative transfer [158, 159], and in gas-liquid adsorption [160]. In all cases, a combination with stochastic methods was reported to be necessary to achieve satisfactory estimates. To our best knowledge, ANN modeling approach has not been applied to the protein chromatography model calibration.

Inspired by the aforementioned contributions, the presented work introduces a novel approach for the estimation of the mass transfer parameter and the isotherm parameters for a chromatography model consisting of the TDM and the SDM isotherm. Here, the ANN modeling is not used for initial guess, but directly for the final estimation. The fundamental idea is to consider all possible parameter combinations as the superset of the true parameter set. The information contained in an uncalibrated model is approximated with *in silico* experiments by varying the model parameters of interest. An ANN model is trained by mapping the results of the *in silico* experiments to the corresponding model parameters. This ANN model was used to recognize particular chromatograms generated experimentally in the lab, and deliver the model parameters in demand. The ANN model's capability of estimating the film transfer coefficient, the equilibrium coefficient, and the characteristic charge is tested with a spiked protein mixture consisting of an industrial monoclonal antibody (mAb), cytochrome c, and lysozyme. The accuracy and the reliability of the parameter estimates is proven by predicting further gradient elution chromatograms and calculating the confidence intervals at 95 % level.

4.2. Theory

4.2.1. Artificial Neural Networks

ANN modeling offers numerous advantages over first principles modeling such as the ability to detect unconsidered nonlinear relationships and possible interactions between variables, and the requirement for less formal statistical training [161]. Hence, it is not surprising that extensive utilization has been found in many complex real-world problems [162]. Silva Neto et al. carried out ANN as a solution for inverse problems [158–160]. In the field of chromatography, the most popular ANN type is the back-propagation ANN (BP-ANN). It is considered a useful tool to model retention times in different chromatographic modes and formats [26–30]. As shown in Fig. 4.1, a BP-ANN of simple architecture typically consists of an input layer with nodes representing input data, a hidden layer, and an output layer returning the output data. Each neuron is connected to every neuron in the next layer. In the nodes of the hidden layer, the sum of products of inputs and weights, and the products of biases and weights n is transferred with a hyperbolic tangent sigmoid activation function to an output signal between -1 and $+1$ as shown in Eq. (4.1). Here, the nonlinearity in the data is captured.

$$f(\tilde{u}) \quad = \quad \frac{2}{1 + e^{-2\tilde{u}}} \tag{4.1}$$
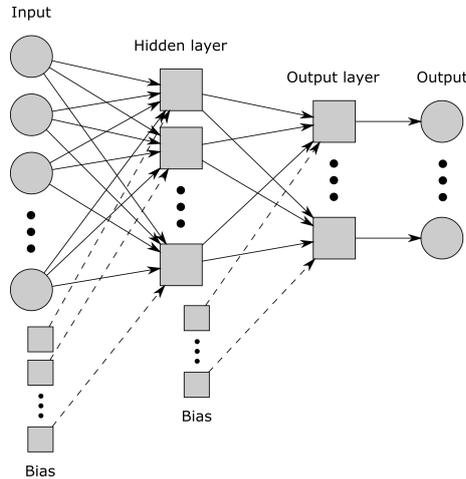
Figure 4.1.: The topological structure of a simple multi-layer BP-ANN.

In the nodes of the output layer, a linear activation function transfers the sum of weighted outputs from the former layer and the weighted biases to the final outputs. By mapping the input data to the correct output data using supervised learning with

the delta rule, the difference between the model response and the desired output is iteratively minimized by updating the weights in the network [163].

4.2.2. Transport-dispersive Model

To model the macroscopic mass transport inside a chromatography column with length L and adsorber beads with radius r_p, the transport-dispersive model is employed [41]:

$$
\begin{aligned}
\frac{\partial c_i(x,t)}{\partial t} &= -u(t)\frac{\partial c_i(x,t)}{\partial x} + D_{ax}\frac{\partial^2 c_i(x,t)}{\partial x^2} \\
&\quad - \frac{1-\epsilon_b}{\epsilon_b}k_{eff,i}\frac{3}{r_p}\left(c_i(x,t) - c_{p,i}(x,t)\right)
\end{aligned}
\tag{4.2}
$$

$$
\frac{\partial c_{p,i}(x,t)}{\partial t} = -\frac{1-\epsilon_p}{\epsilon_p}\frac{\partial q_i(x,t)}{\partial t} + k_{eff,i}\frac{3}{\epsilon_p r_p}\left(c_i(x,t) - c_{p,i}(x,t)\right)
\tag{4.3}
$$

$$
\frac{\partial c_i(0,t)}{\partial x} = \frac{u(t)}{D_{ax}}(c_i(0,t) - c_{in,i}(t))
\tag{4.4}
$$

$$
\frac{\partial c_i(L,t)}{\partial x} = 0
\tag{4.5}
$$

Eq. (4.2) describes the convectional flow, the axial dispersion, and the mass transfer between the concentration in interstitial volume of the mobile phase $c_i(x,t)$ and the pore volume $c_{p,i}(x,t)$ of component i. It depends on the peak-broadening effects in the interstitial volume that are lumped in the component-specific film transfer coefficient $k_{eff,i}$, the axial dispersive coefficient D_{ax}, and the voidage of the bed ϵ_b. Eq. (4.3) models the exchange between stationary phase concentration q_i and the pore volume concentration $c_{p,i}(x,t)$ depending on the particle voidage ϵ_p and the overall mass transport coefficient $k_{eff,i}$. Eqs. (4.4) and (4.5) are the Danckwerts boundary conditions.

4.2.3. Stoichiometric Displacement Isotherm Model

To describe the adsorption and desorption mechanisms on the IEX adsorber surface, the kinetic formulation of the stoichiometric displacement isotherm model (SDM) is employed [56, 164]:

$$
k_{kin,i}\frac{\partial q_i(x,t)}{\partial t} = k_{eq,i}q_{salt}(x,t)^{\nu_i}c_{p,i}(x,t) - c_{salt}^{\nu_i}q_i(x,t)
\tag{4.6}
$$

As proposed in [116], the equilibrium coefficient $k_{eq,i} = k_{ads,i}/k_{des,i}$ and the kinetic coefficient $k_{kin,i} = 1/k_{des,i}$ are used instead of the adsorption coefficient $k_{ads,i}$ and the desorption coefficient $k_{des,i}$. The advantage of this formulation is that $k_{eq,i}$ strongly affects the retention time, whereas $k_{kin,i}$ has a slight influence on the peak height in most cases. ν_i is the characteristic charge of the component i. The counter-ion concentration in the pore phase is depicted as c_{salt}, and the salt concentration in the stationary phase q_{salt} is given as:

$$q_{salt}(x,t) \;=\; \Lambda - \sum_{i=1}^{m} \nu_i q_i(x,t) \qquad (4.7)$$

with the ionic capacity of the stationary phase being Λ. Since the derivation of the SDM isotherm does not consider the adsorber binding sites which are sterically hindered by adsorbed molecules, the shielding effect is neglected.

4.2.4. Numerical Solution

For mechanistic modeling, the discretization in space is performed using the finite element method on a grid with equidistant nodes. The discretization in time is carried out with the fractional step θ-scheme [111]. The nonlinear equation system is treated with Picard iteration [112]. For parameter estimation, the Levenberg-Marquardt algorithm [114] is employed.

4.3. Materials and Methods

4.3.1. Instruments

Gradient elution experiments were carried out using an Ettan liquid chromatography (LC) system equipped with pump unit P-905, dynamic single chamber mixer M-925 (90 μl mixer volumne), UV monitor UV-900 (3 mm optical path length), conductivity cell pH/C-900, and autosampler A-905 (all GE Healthcare, Little Chalfont, Buckinghamshire, UK).

4.3.2. Software

The control software UNICORN 5.31 (GE Healthcare, Little Chalfont, Buckinghamshire, UK) was used in combination with the LC system. The model was simulated using the chromatography simulation software ChromX (GoSilico, Karlsruhe, Germany). ChromX was employed for numerical simulation of the system of partial differential equations, as well as optimization and statistical analysis [115]. The ANN modeling was carried out in Matlab$^{\ddot{o}}$ R2016a (MathWorks, USA).

4.3.3. Adsorber, Proteins, and Chemicals

The strong cation-exchange chromatography (CEX) adsorber medium, TOYOPEARL®️
GigaCap S-650M supplied by Tosoh Bioscience (Griesheim, Germany) was used as
pre-packed 0.965 mL Toyoscreen®️ column with dimension 30 mm × 6.4 mm. Between
the runs, the adsorber media were stored in 20 % ethanol. Before experimentation,
the storage solution was removed by prolonged equilibration with water, low-salt
and high-salt buffer.

Three proteins of different sizes and different isoelectric points were used as model
proteins to investigate the ANN's capability of model parameter identification. A
monoclonal antibody (mAb, 144-147 kDa, pI 8.3-8.5) of the IgG isotype was kindly
provided by Lek (Ljubljana, Slovenia). Cytochrome c (bovine heart, no. A4612,
12.3 kDa, pI 10.4-10.8) was purchased from Sigma-Aldrich (St. Louis, MO, USA).
Lysozyme (hen egg, no. HR7-110, 14.6 kDa, pI 11.0) was purchased from Hampton
Research (Aliso Viejo, CA, USA).

To perform experiments below the isoelectric points of all three proteins, 50 mM
sodium phosphate buffer was used at pH 6.5 with additional 0 and 1 M NaCl.
For cleaning-in-place, 0.5 M NaOH was used. All solutions were prepared with
ultra pure water (arium pro UV, Sartorius, Göttingen, Germany). All buffers and
storage solutions were filtrated with a membrane cutoff of 0.22 μm and degassed by
sonication. All proteins were prepared in accordance with the loading conditions and
were filtrated with a membrane cutoff of 0.22 μm prior to usage.

4.3.4. System Characterization

The ÄKTA system and chromatography column were characterized via tracer pulse
injection at a constant flow rate of 0.2 ml/min. For determination of the system and
total voidage of the column, 25 μL 1 %(v/v) acetone (Merck, Darmstadt, Germany)
was used as a pore-penetrating, non-interacting tracer. Filtrated 10 g/L blue dextran
2000 kDa (Sigma-Aldrich, St. Louis, MO, USA) solution was used as non-pore-
penetrating, non-interacting tracer to determine the interstitial volume of the column.
The UV signal at 260 nm was recorded to determine these parameters and correct
all measurements regarding dead volumes. The ionic capacity Λ of GigaCap S-
650M was determined by acid-base titration following Huuk et al. [165]. The UV
signal at 280 nm was recorded during pulse injection of pure protein samples. The
measurement factors of each protein were calculated from the protein concentration
and adsorption units according to the Lambert-Beer law [116].

4.3.5. *In Silico* Screening

Since the model parameter $k_{kin,i}$ has just a minor impact on the peak height, it was eliminated from the screening process and set to $1 \cdot 10^{-9}$ with the aim of reducing the design space. 780 model parameter combinations of $k_{eff,i}$, $k_{eq,i}$ and ν_i within boundaries as shown in Tab. 4.1 were used to carry out *in silico* experimentation using ChromX. Wide ranges were chosen to cover very different protein adsorption behaviors on CEX. $k_{eq,i}$ was logarithmized to be within reasonable range. Finally, all model parameters and the resulting 780 chromatograms were normalized to the minimum and maximum values.
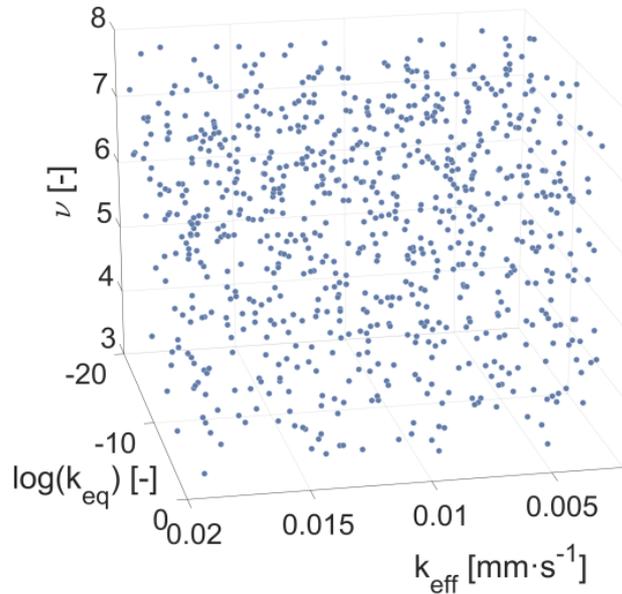
Figure 4.2.: Design space of the *in silico* screening:$k_{eff} \in [0.002,0.02]\,mm/s$, $k_{eq} \in [1 \cdot 10^{-9},0.1]$, and $\nu \in [3,8]$.

Table 4.1.: Design space of model parameters for *in silico* screening.

	Lower boundary	Upper boundary
k_{eff} [mm·s^{-1}]	0.002	0.02
k_{eq} [-]	$1.0 \cdot 10^{-10}$	0.1
ν [-]	3	8

4.3.6. ANN Model Calibration and Validation

A BP-ANN model with a hidden layer with 18 neurons and an output layer with three neurons was created in MATLAB®. In the nodes of the hidden layer, the hyperbolic tangent sigmoid activation function *tansig* was implemented. In the ones of the output layer, the linear activation function *purelin* was chosen. The function *divideint* was applied to divide the *in silico* data into training, validation, and test data sets by a ratio of 7:2:1 using interleaved indices. The network training function *traingdx* was used to update weight values for inputs and biases according to gradient descent momentum and an adaptive learning rate. The momentum constant and the learning rate were set to 0.9 and 0.05. A root-mean-square error (RMSE) of $1 \cdot 10^{-5}$ and a maximal iteration number of $1 \cdot 10^3$ were set as stopping criteria.

4.3.7. Bind-and-elute Experiments

$1 \, \text{mg/ml}$ of each protein were dissolved in corresponding binding buffer and loaded via a $100 \, \mu\text{L}$ loop. Different gradients and steps were mixed within the LC system from low-salt and high-salt buffer. A subset of the chromatograms was used as input to perform the parameter estimation with the previously calibrated ANN model. The remaining runs were used for model validation.

The linear gradient experiments were conducted by applying the varying gradient lengths 5 CV, 15 CV, 20 CV, and 25 CV. After a post-loading wash step over 1 CV equilibration buffer, elution was started with 50 mM sodium phosphate and an increasing salt gradient ranging from 0 mM to 350 mM sodium chloride. After protein elution, the column was stripped over 3 CV at a sodium chloride concentration of 1 M and re-equilibrated with 1 CV equilibration buffer.

The step elution comprised of three elution steps. First, the column was equilibrated with 2 CV equilibration buffer. After a wash step over 2 CV equilibration buffer, the salt concentration was raised to 50 mM sodium chloride. After 8 CV the concentration was further increased to 140 mM sodium chloride and was kept constant over 8 CV. The column was finally stripped over 3 CV at a sodium chloride concentration of 1 M and re-equilibrated with equilibration buffer.

All experiments were conducted at a flow rate of 0.2 ml/min to ensure a constant residence time.

4.3.8. Mechanistic Model Calibration and Validation

Two bind-and-elute chromatograms using linear gradient elution of 15 CV and 25 CV were fitted with exponentially modified Gaussian distribution [166] with the aim of signal smoothing and peak separation. After normalization, the separated peaks were

fed into the calibrated BP-ANN model which returned the estimates of $k_{eff,i}$, $k_{eq,i}$ and ν_i for each protein. As fine adjustment of the peak height, Levenberg-Marquardt algorithm [114] was employed. Three uninvolved bind-and-elute chromatograms were predicted and compared to wet-lab experiments. To prove the estimation reliability, confidence intervals at 95 % level were calculated subsequently.

4.4. Results and Discussion

4.4.1. System Characterization

The system dead volume of $280\,\mu l$ of the LC was determined by tracer injection without a column attached. This volume was subtracted from all other data obtained from the LC. Bed and particle voidage, and axial dispersion were calculated based on the results of the tracer injections. The ionic capacity was calculated based on the acid-basic titration. The results are listed in Tab. 4.2. The transformation from protein concentration to adsorption units was performed with the measurement factors $6.651{\cdot}10^7\ mAU \cdot L/mol$ for IgG, $6.762{\cdot}10^6\ mAU \cdot L/mol$ for cytochrome c, and $1.054{\cdot}10^7\ mAU \cdot L/mol$ for lysozyme, according to the Lambert-Beer law.

4.4.2. ANN Model Calibration and Validation

For ANN model calibration, the results of *in silico* screening were correlated with the corresponding model parameter sets using back-propagation. 546 training data sets were used for the ANN's adjustment according to the respective errors. 156 validation data sets were used to measure the capability of generalization, and to stop training when generalization stopped improving. Test data consisting of 78 chromatograms and parameter sets were excluded during model calibration for cross validation. The best correlations achieved for the mass transfer coefficient k_{eff} are

Table 4.2.: The voidages and axial dispersion are calculated from retention volume and peak broadening of tracer injections.

		GigaCap S-650M
Bed voidage	ϵ_b	0.404
Particle voidage	ϵ_p	0.765
Total voidage	ϵ_t	0.860
Axial dispersion $[mm^2 \cdot s^{-1}]$	D_{ax}	0.048
Ionic Capacity $[M]$	Λ	1.284

given in Figs. 4.3 (a) - (c) which represent the training data set, validation data set, and test data set, respectively. Figs. 4.3 (d) - (f) show the correlation for the equilibrium coefficient k_{eq} and Figs. 4.3 (g) - (i) the characteristic charge ν.
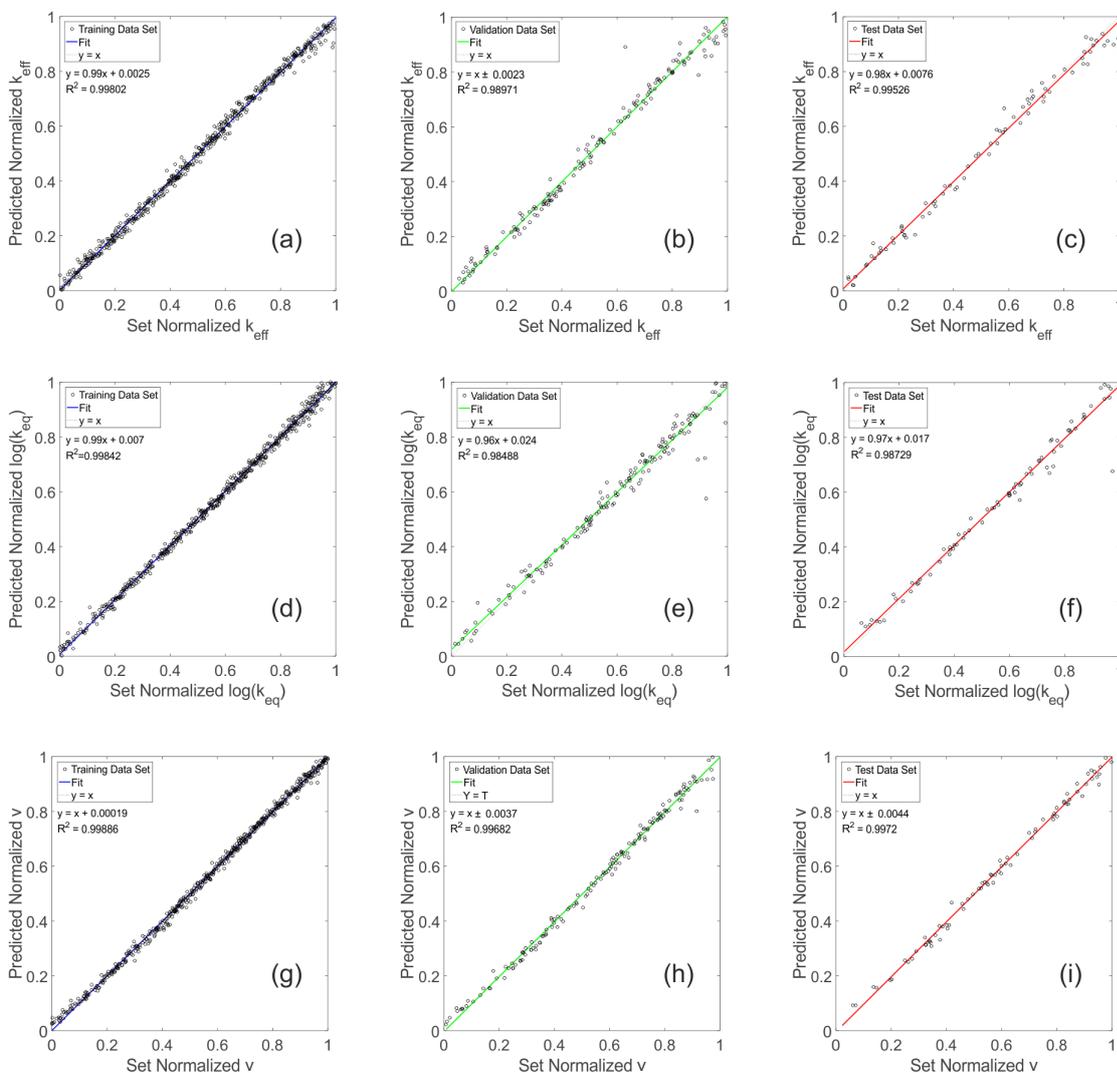

Figure 4.3.: Plots of screened parameters versus simulated parameters using ANN model. (a) - (c) present the training data set, validation data set, and test data set of the parameter k_{eff}, respectively. (d) - (e) present k_{eq}. (f) - (h) present ν.

A good agreement between the predictions by the ANN model and the parameters set for the *in silico* screening was observed. The deviations found were in a statistically acceptable range ($R^2 = 0.98971$ for the validation data set of k_{eff}, $R^2 = 0.98488$ for the validation data set of k_{eq}, and $R^2 = 0.98729$ for the validation data set of ν). According to correlation results of test data sets ($R^2 = 0.99526$ for k_{eff}, $R^2 = 0.98729$ for k_{eq}, and $R^2 = 0.9972$ for ν), the generalization capability of the ANN model was

verified, making it suitable for the estimation of model parameters.

4.4.3. Mechanistic Model Calibration and Validation

To estimate the model parameters k_{eff}, k_{eq}, and ν, two linear gradient chromatography experiments were performed with gradient lengths of 15 CV and 25 CV. After data preparation, these chromatograms were presented to the beforehand calibrated and verified ANN model. The returned parameters are given in Tab. 4.3. The confidence intervals at 95 % level confirm the reliability of the parameter estimates. For a slight adjustment of the peak height, the kinetic coefficients k_{kin} were estimated subsequently by applying the Levenberg-Marquardt algorithm. Figs. 4.4 (a)-(b) show the simulated and experimental chromatograms used for the parameter estimation. The weakly binding component mAb, moderately binding component cytochrome c, and the strongly binding component lysozyme are eluted due to the salt gradient in successive order. To assess the predictive power of the calibrated mechanistic model, three additional experiments have been conducted as validation data. Figs. 4.5 (a)-(c) show good agreement of model prediction and experimental data. Thus, the established method for model calibration can reliably identify SDM parameters in the linear or weakly nonlinear isotherm region based on chromatograms presented to ANN.
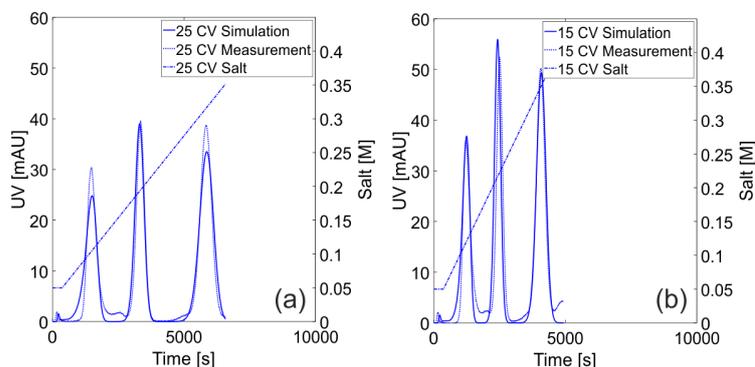
Figure 4.4.: Plots of UV signals over process run time for bind-and-elute experiments. Dashed lines display the UV signal measured at column outlet and the adjusted salt gradients. Solid lines represent the simulated chromatograms. The elution peaks of the first eluting component mAb, the intermediate eluting component cytochrome c, and the last eluting component lysozyme by applying linear salt gradients from 0.05 M to 1.05 M over 25 CV and 15 CV are shown in (a) and (b), respectively.

Table 4.3.: Parameters of the mass transfer model and kinetic isotherm formulation estimated from two bind-and-elute experiments. k_{eff}, k_{eq}, and ν were determined using the ANN model. k_{kin} was estimated using the inverse method.

Parameter	mAb	Cytochrome c	Lysozyme
k_{eff} [mm/s]	$6.593 \cdot 10^{-3} \pm 3.685 \cdot 10^{-3}$	$6.300 \cdot 10^{-3} \pm 1.447 \cdot 10^{-3}$	$6.644 \cdot 10^{-3} \pm 3.333 \cdot 10^{-4}$
k_{eq} [-]	$8.363 \cdot 10^{-6} \pm 2.594 \cdot 10^{-7}$	$6.062 \cdot 10^{-6} \pm 8.707 \cdot 10^{-6}$	$1.445 \cdot 10^{-2} \pm 1.489 \cdot 10^{-5}$
ν [-]	5.220 ± 0.015	$5.167 \pm 9.900 \cdot 10^{-3}$	$5.195 \pm 8.079 \cdot 10^{-3}$
k_{kin} [s^{-1}]	$2.239 \cdot 10^{-4} \pm 3.973 \cdot 10^{-5}$	$3.440 \cdot 10^{-9} \pm 7.133 \cdot 10^{-4}$	$1.225 \cdot 10^{-8} \pm 2.540 \cdot 10^{-3}$
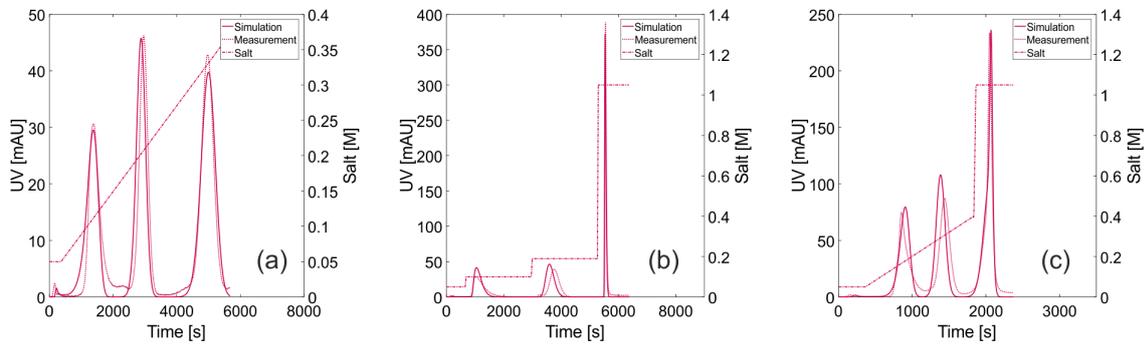
Figure 4.5.: Plots of UV signals over process run time for bind-and-elute experiments. Dashed lines display the UV signal measured at column outlet and the adjusted salt gradients. Solid lines represent the simulated chromatograms. The elution peaks of the first eluting component mAb, the intermediate eluting component cytochrome c, and the last eluting component lysozyme by applying linear salt gradients from 0.05 M to 1.05 M over 5 CV, 20 CV, and a step-wise salt gradient from 0.05 M to 0.20 M and up to 1.05 M are shown in (a), (b), and (c).

4.5. Conclusion

In the presented work, artificial neural network modeling was applied for the simultaneous estimation of TDM and SDM parameters mass transfer coefficient k_{eff}, equilibrium coefficient k_{eq}, and characteristic charge ν for a tertiary protein mixture in CEX. After data generation by conducting *in silico* experiments, data preparation, and calibration of the ANN model by mapping the resulting simulated chromatograms to the corresponding model parameter sets, parallel estimations of nine model parameters were carried out in milliseconds using the ANN model. The ANN model's capability of generalization was verified by coefficients of determination of the cross validation set between 0.98729 and 0.99526. The correctness of the estimated model parameters was proved by good agreement between simulation and measurement of two linear gradient chromatography experiments. The reliability of the parameter estimates was underlined by small confidence intervals at 95 % level. Furthermore, three uninvolved bind-and-elute gradient chromatography experiments were predicted

with high accuracy using the estimated parameters. Because of the ability of ANN to cover nonlinearities, there is no necessity to carry out the calibration experiments in the linear region of the adsorption isotherm.

The presented novel method has several unique characteristics compared to established methods. Once the ANN model is calibrated, it possesses the ability to process new chromatograms and determine the underlying model parameter sets without overhead. Also, it can assign suitable model parameters to new components in the chromatographic system, where traditional model calibration approaches require to be conducted again and are therefore computationally more expensive. Potentially, this method can be implemented for automated real-time model parameter estimation during the chromatographic experiments, since the parameter estimation process itself only takes milliseconds. As a next step, the presented method will be applied to different modes of chromatography and isotherm equations in order to check universal applicability to mechanistic model calibration.

4.6. Acknowledgment

5. Model-based Investigation on the Mass Transfer and Adsorption Mechanisms of Mono-PEGylated Lysozyme in Ion-exchange Chromatography

Josefine Morgenstern*, Gang Wang*, Pascal Baumann and Jürgen Hubbuch**

Institute of Engineering in Life Sciences, Section IV: Biomolecular Separation Science, Karlsruhe Institute of Technology (KIT), 76131 Karlsruhe, Germany
* These authors contributed equally to this work.
** Corresponding author.

Abstract

Recent studies highlighted the potential of PEGylated proteins to improve stabilities and pharmacokinetics of protein drugs. Ion-exchange chromatography (IEX) is among the most frequently used purification methods for PEGylated proteins. However, the underlying physical mechanisms allowing for a separation of different PEGamers (proteins with a varying number of attached PEG molecules) are not yet fully understood.

In this work, mechanistic chromatography modeling was applied to gain a deeper understanding of the mass transfer and adsorption/desorption mechanisms of mono-PEGylated proteins in IEX. Using a combination of the general rate model (GRM) and the steric mass action (SMA) isotherm, simulation results in good agreement with the experimental data were achieved. During linear gradient elution of proteins attached with PEG of different molecular weight, similar peak heights and peak shapes at constant gradient length were observed. A superimposed effect of increased desorption rate and reduced diffusion rate as a function of the hydrodynamic radius

of PEGylated proteins was identified to be the reason of this anomaly. That is why the concept of the diffusion-desorption-compensation effect is proposed. In addition to the altered elution orders, PEGylation resulted in a considerable decrease of maximum binding capacity. By using the SMA model in a kinetic formulation, the adsorption behavior of PEGylated proteins in the highly concentrated state was described mechanistically. An exponential increase in the steric hindrance effect with increasing PEG molecular weight was observed. This suggests the formation of multiple PEG layers in the interstitial space between bound proteins and an associated shielding of ligands on the adsorber surface to be the cause of the reduced maximum binding capacity. The presented *in silico* approach thus complements the hitherto proposed theories on the binding mechanisms of PEGylated proteins in IEX.

5.1. Introduction

It is estimated that in 2020 about 46% of the sales volume of the 100 highest-selling pharmaceutical products will be achieved by biopharmaceutical products [167]. Biopharmaceuticals contain active substances based on biological molecules, such as recombinant proteins. Compared to conventional small molecular pharmaceuticals, proteins have a complex three-dimensional structure allowing for a more efficient and specific intervention in cellular metabolic pathways. The efficacy of systemically administered protein drugs however, may be hampered by a low bioavailability due to a poor solubility under physiological conditions, a short in vivo half-life due to a rapid elimination by the body and proteolysis. A promising approach to overcoming these drawbacks is the covalent attachment of polyethylene glycol (PEG) to protein drugs [168]. As early as in 1977, the group of Abuchowski and Davis found an increased blood circulation half-life and a reduced immunogenicity of PEGylated proteins compared to the native form [169, 170]. Additional positive effects of PEGylation are an increased thermal stability as well as a higher solubility allowing for higher concentrated protein formulations [171]. Two successfully approved PEGylated protein drugs are interferon α-2a (Pegasys®, Hoffman-LaRoche) for the treatment of hepatitis C and granulocyte-colony stimulating factor (Neulasta®, Amgen) for the treatment of leukemia.

The emergence of conjugates with varying number (PEGamers) and site of attachment (positional isoforms) upon PEGylation reactions creates a need for a thorough purification in order to gain regulatory approval [172, 173]. Ion-exchange chromatography (IEX) is among the most frequently used purification methods for PEGylated proteins [172, 174]. Understanding the underlying physical mechanisms is an important prerequisite to optimize, control, predict, and scale-up the separation of PEGamers to pilot and production level. In this context, mechanistic modeling provides an

excellent opportunity to generate various information about mass transport and adsorption isotherm parameters *in silico*.

The physico-chemical properties and thus the behavior of a protein in chromatographic separation processes are significantly influenced by its PEGylation [172, 173]. Due to the high hydration of the hydrophilic PEG, PEGylated proteins have a distinctly higher hydrodynamic radius than unmodified proteins with the same molecular weight. A non-linear correlation introduced by Fee and Van Alstine allows a reliably mathematical prediction of the hydrodynamic radius $h_{R,PEGprot}$ based on the molecular weight of the protein and the attached PEG [172, 175]. In case of chromatographic separation, the PEG 'cloud' around the protein results in an increased distance between protein binding site and adsorber surface [173]. Seely and Richey [176] observed that the elution order of different PEGamers was the same in both cation-exchange and anion-exchange chromatography. They proposed the 'charge-shielding effect' which links the weakened electrostatic interactions to the increased distance between protein binding site and adsorber surface. A deeper process understanding was achieved by Yamamoto et al. [177] using mechanistic chromatography modeling. They applied the stoichiometric displacement model (SDM) to verify the 'charge-shielding effect' quantitatively and associated it with the decreased elution volume of PEGamers. Moreover, it was shown that mono-PEGylated proteins are bound to the ion-exchange adsorber with binding sites similar to the unmodified protein. In following studies, this model was applied to PEGylated lysozyme and BSA [178, 179]. The aforementioned contributions demonstrated the successful application of mechanistic modeling to understand the adsorption behavior of PEGylated proteins in the linear region of the adsorption isotherm.

This work presents a full investigation of the behavior of mono-PEGylated proteins in IEX based on mechanistic chromatography modeling. In contrast to previous studies, information on the adsorption and desorption behavior in the non-linear region of the isotherm, i.e. the overloaded state, is included by using the steric mass action (SMA) model [57] in kinetic formulation. Compared to the equilibrium isotherm used hitherto, the kinetic formulation is suitable for the description of protein behavior in higher concentrated state on adsorber surface. To further account for mass transfer effects within the chromatography column the general rate model (GRM) [180] is employed. To best of our knowledge, mechanistic modeling of polymer grafted proteins in IEX using a combination of GRM and SMA isotherm has not been studied. By connecting these two approaches, this study delivers supplements by the quantitative investigation on the film diffusion, pore diffusion, charge and shielding parameters, as well as the adsorption and desorption rate coefficients.

The model protein lysozyme from chicken egg was chosen as PEGylation target and conjugated to activated PEG of three different molecular weights (2 kDa, 5 kDa and 10 kDa). The preparative isolation of the mono-PEGylated species was carried out using a single cation-exchange (CEX) chromatography step. For each purified protein

species, three linear gradient elution (LGE) experiments with different gradient slopes were conducted to confirm the constancy of the characteristic charge. Breakthrough experiments were carried out to gain insight into the binding behavior of PEGylated proteins in the highly non-linear region and to investigate whether the perceivable behavior of PEGylated proteins originates from adsorption/desorption or mass transfer. Confidence intervals at 95 % level were calculated for parameter estimates.

5.2. Materials and Methods

5.2.1. Adsorber, Proteins, and Chemicals

All stock solutions and buffers were prepared with ultra-pure water (PURELAB Ultra water purification system, ELGA Labwater, Germany), filtrated using a cellulose-acetate filter with a membrane cut-off of $0.22\,\mu$m (Satorius, Germany) and degassed by sonication. The used buffer substances were sodium acetate trihydrate (Sigma-Aldrich, USA) for pH 5 and sodium phosphate monobasic dihydrate (Sigma-Aldrich, USA) as well as di-sodium hydrogen phosphate dihydrate (Merck, Germany) for pH 7.0 and pH 7.2, respectively. Hydrochloric acid and sodium hydroxide (NaOH) for pH adjustment were obtained from Merck (Germany). Lysozyme from chicken egg-white (no. HR7-110) was purchased from Hampton Research (USA). Methoxy-PEG-propionaldehyde (mPEG-aldehyde) with an average molecular weight (MW) of 2 kDa (Sunbright®ME-020 AL), 5 kDa (Sunbright®ME-050 AL) and 10 kDa (Sunbright®ME-100 AL) was obtained from NOF Corporation (Japan). Sodium cyanoborohydride (NaCNBH$_3$) and L-lysine were purchased from Sigma-Aldrich (USA). For preparative isolation of PEGamers as well as for modeling purposes, the strong cation-exchange (CEX) chromatography adsorber medium TOYOPEARL® GigaCap S-650M (Tosoh Bioscience, Germany) was used. It is a high capacity polymer grated cation exchange resin based on hydroxylated methacrylic polymer with a 100 nm pore size and a 75 μm particle size. For preparative isolation of PEGamer species, 5 mL pre-packed MiniChrom columns (dimension: 100 mm $\times$ 8 mm) and for modeling purposes, a pre-packed 0.965 mL Toyoscreen®column (dimension: 30 mm $\times$ 6.4 mm) were used. Between the runs, the resin media were stored in 20 % ethanol. The storage solution was removed by prolonged equilibration with ultra-pure water and flushed with binding and elution buffer before experimentation. Sodium chloride (NaCl) used for protein elution was purchased from Merck (Germany). 0.5 M NaOH (Merck, Germany) was used for cleaning-in-place.

5.2.2. Instrumentation and Software

pH adjustment of all buffers was performed using a five-point calibrated pH-meter HI-3220 (Hanna Instruments, USA) equipped with a SenTix®62 pH electrode (Xylem Inc., USA). Protein concentration measurements were conducted using a NanoDrop2000c UV–vis spectrophotometer (Thermo Fisher Scientific, USA). Purity of isolated mono-PEGylated lysozyme was determined by high-throughput capillary gel electrophoresis (HT-CGE) using the Caliper LabChip®GX II device (PerkinElmer, USA). For data processing and purity determination, the LabChip®GX 3.1 software (PerkinElmer, USA) was used.

Preparative isolation of mono-PEGylated lysozyme species was performed on an ÄKTA™ purifier system equipped with a Fraction Collector Frac-950 (GE Healthcare, Sweden). All experiments for chromatography model calibration were carried out using an Ettan liquid chromatography (LC) system with the UV monitor UV-900 (3 mm optical path length), pump unit P-905, dynamic single chamber mixer M-925 (90 μl mixer volumne), and conductivity cell pH/C-900 (all GE Healthcare, Little Chalfont, Buckinghamshire, UK). The UNICORN 5.31 software (GE Healthcare, UK) was used to control both chromatographic systems and to record the signals. The protein chromatography simulation software ChromX (GoSilico, Germany) was used for the numerical simulations of the system of partial differential equations, estimation of model parameters, as well as for statistical analysis [115]. Other data evaluations were conducted in Matlab® R2016a (MathWorks, USA).

5.2.3. PEGylation Reaction

As reaction buffer 25 mM sodium phosphate (pH 7.2) containing 20 mM sodium cyanoborohydride (NaCNBH₃) as reducing agent was used. PEGylation experiments were performed batch-wise in 50 mL Falcon Tubes (BD Biosciences, USA). Lysozyme (5 mg/mL) and mPEG-aldehyde were dissolved in the reaction buffer with a molar polymer to protein ratio of 6.67:1 [181, 182]. The tube was continuously shaken in an overhead shaker LabincoLD79 (Labinco BV, Netherlands) for 3.5 h at 25°C. The PEGylation reaction was stopped by adding 200 mM of L-lysine according to [183].

5.2.4. Preparative Purification of Mono-PEGylated Lysozyme

For preparative isolation of mono-PEGylated lysozyme, the stopped PEGylation batch was diluted to a ratio of 1:12 in 10 mM sodium acetate buffer (pH 5) [181]. For column loading, the system was equilibrated in 10 mM sodium acetate buffer (pH 5). Sample application was performed using a 50 mL super loop (GE Healthcare, Sweden). Elution was initiated by applying an NaCl step gradient with 10 mM sodium acetate buffer (pH 5) containing 1.0 M sodium chloride. The NaCl molarities

used for the step elution of the different PEGamers are displayed in Tab. 5.1 as a function of the molecular weight of the attached PEG molecules. The flow rate for binding and elution was set to 1 mL/min. Fractions of 2 mL were collected into a 96-well deep well plate (VWR, USA). To obtain sufficient sample volume for the linear gradient and the breakthrough experiments, fractions containing mono-PEGylated lysozyme of multiple chromatography runs were pooled. To ensure similar binding conditions for all PEG molecular weights during the calibration runs, the mono-PEGylated samples were concentrated to approximately $3.76 \cdot 10^{-4}$ M. This was accomplished by evaporation using a vacuum concentration unit RVC 2-33CDplus (Martin Christ Gefriertrocknungsanlagen GmbH, Germany) operated at 24 mbar. After concentrating, the protein samples were transferred to 25 mM sodium phosphate buffer (pH 7) using Slide-A-Lyzer™ Dialysis Cassettes (Thermo Fisher Scientific, USA) with a molecular weight cut-off of 2 kDa. All chromatography experiments were carried out at 25°C.

5.2.5. Offline Identification and Quantification of PEGamer Species

Purity of isolated mono-PEGylated lysozyme was determined by high-throughput capillary gel electrophoresis (HT-CGE) as described in [181]. The experiments were performed with an HT Protein Express LabChip® and an HT Protein Express Reagent Kit (Perkin Elmer, Hopkinton, MA, USA). The LabChip® installation, sample preparation and analysis were performed according to the manufacturer's standard protocol [184]. Sample preparation was performed in skirted 96-well polypropylene twin.tec® PCR plates from Eppendorf (Hamburg, Germany). Molecular weight determination was performed according to protein standards from the HT Protein Express Reagent Kit.

For protein quantification, absorption measurements at 280 nm were performed. Since the bound PEG molecules do not absorb at 280 nm, the extinction coefficient of $\varepsilon^{1\%}_{280\,nm,lysozyme} = 22.00$ was used for native as well as for mono-PEGylated lysozyme [181, 185]. Appropriate blanks were subtracted. Molar concentrations were

Table 5.1.: NaCl steps in mM used for the elution of different PEGamer species from Toyopearl GigaCap S-650M at pH 5 as a function of the PEG molecular weight

	Native lysozyme	mono-PEGylated lysozyme	di-PEGylated lysozyme
M_w=2 kDa	1000	460	290
M_w=5 kDa	1000	350	160
M_w=10 kDa	1000	250	75

calculated using a lysozyme molecular mass of 14.6 kDa [186]. The final concentrations of native lysozyme and mono-PEGylated species attached with 2 kDa, 5 kDa, and 10 kDa PEG used for the linear gradient and breakthrough experiments were
$3.87 \cdot 10^{-4} \pm 7.19 \cdot 10^{-7}$ M, $3.63 \cdot 10^{-4} \pm 1.41 \cdot 10^{-7}$ M,
$3.60 \cdot 10^{-4} \pm 1.61 \cdot 10^{-7}$ M, and $3.81 \cdot 10^{-4} \pm 1.57 \cdot 10^{-5}$ M, respectively. The slight deviations in PEGamer concentrations are due to concentrating and buffer exchange. For subsequent modeling the exact concentrations were employed.

5.2.6. Chromatography System Characterization

Tracer pulse injections at constant flow rate of 0.33 mL/min were carried out to characterize the ÄKTA™ system and chromatography column. For determination of the interstitial volume of the column, 25 μL of 10 g/L non-interacting, non-pore-penetrating tracer blue dextran 2000 kDa (Sigma-Aldrich, St. Louis, MO, USA) in ultra-pure water was used. 25 μL of 1 %(v/v) pore-penetrating, non-interacting tracer acetone (Merck, Darmstadt, Germany) in ultra-pure water was used to determine system and total voidage of the column. The UV signals at 260 nm were recorded for that purpose. All measurements were corrected with respect to system dead volumes. The ionic capacity Λ of GigaCap S-650M was determined via acid-base titration following Huuk and coworkers [165].

5.2.7. Linear Gradient Experiments for Model Calibration

Protein solutions with lysozyme and its PEGylated species were prepared in binding buffer (25 mM sodium phosphate buffer, pH 7.0). Before injection, the protein solutions were filtrated with a membrane cut-off of 0.22 μm.

Linear gradient elution (LGE) experiments were used for determining model parameters for native lysozyme, lysozyme attached with PEG 2 kDa, PEG 5 kDa, and PEG 10 kDa. Protein solutions were injected via a 100 μL loop. After a post-loading wash step of 1 CV binding buffer, elution was carried out by increasing the salt gradient from 0 M to 1.0 M NaCl. From low-salt and high-salt buffer, linear gradients with a gradient length of 15 CV, 20 CV, and 25 CV were mixed within the LC system. After that, the column was stripped over 3 CV at an NaCl concentration of 1.0 M and re-equilibrated for 5 CV binding buffer. To ensure a constant residence time, all experiments were carried out at a flow rate of 0.33 mL/min.

5.2.8. Breakthrough Experiments for Model Calibration

Breakthrough experiments were used for modeling of the SMA isotherm model in the non-linear region. Protein solutions with native lysozyme, lysozyme attached

with PEG 2 kDa, PEG 5 kDa, and PEG 10 kDa were prepared in binding buffer and injected via a 50 mL superloop (GE Healthcare, UK). The loading was carried out under strong binding condition at 0 M NaCl until 100 % breakthrough was observed. To ensure a constant residence time, all experiments were carried out at a flow rate of 0.33 mL/min.

5.2.9. General Rate Model

In the presented study, the general rate model (GRM) was employed to cover convection and diffusion within a one-dimensional chromatography column of length L. Here, the concentrations of all components i in the bulk phase c, in the pore phase c_p, and adsorbed to the stationary phase q depend on time t and axial position x. Eq. 7.1 describes the mass transfer between the bulk phase and the pore phase depending on the flow velocity u, axial dispersion D_{ax}, bed porosity ϵ_b, film diffusion coefficient k_{film}, particle radius r_p, and the concentrations c and c_p. The chosen Danckwerts boundary conditions are shown in Eqs. 7.2 and 7.3. In Eq. 7.4 the mass transfer between the pore phase and the stationary phase is described to be dependent on the radial position in the pore r, pore diffusion coefficient D_p, particle porosity ϵ_p, film diffusion coefficient k_{film}, and concentrations in the bulk phase c, pore phase c_p, and stationary phase q.

$$
\begin{aligned}
\frac{\partial c_i(x,t)}{\partial t} \;=\; & -u(t)\frac{\partial c_i(x,t)}{\partial x} + D_{ax}\frac{\partial^2 c(x,t)}{\partial x^2} \\
& -\frac{1-\epsilon_b}{\epsilon_b}k_{film,i}\frac{3}{r_p}\left(c_i(x,t) - c_{p,i}(x,t)\right)
\end{aligned}
\tag{5.1}
$$

$$
\frac{\partial c_i(0,t)}{\partial x} \;=\; \frac{u(t)}{D_{ax}}(c_i(0,t) - c_{in,i}(t))
\tag{5.2}
$$

$$
\frac{\partial c_i(L,t)}{\partial x} \;=\; 0
\tag{5.3}
$$

$$
\frac{\partial c_{p,i}(x,t)}{\partial t} \;=\;
\begin{cases}
\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 D_{p,i}\frac{\partial c_{p,i}(x,t)}{\partial r}\right) - \frac{1-\epsilon_p}{\epsilon_p}\frac{\partial q_i(x,t)}{\partial t} & \text{for } r \in (0, r_p), \\
\frac{k_{film,i}}{\epsilon_p D_{p,i}}\left(c_i(x,t) - c_{p,i}(x,t)\right) & \text{for } r = r_p, \\
0 & \text{for } r = 0.
\end{cases}
\tag{5.4}
$$

5.2.10. Adsorption Isotherm Model

Based on the stoichiometric displacement model (SDM) [187], Brooks and Cramer derived the steric mass action (SMA) isotherm model by introducing the shielding factor σ, which accounts for the sterically hindered binding sites on the adsorber surface due to protein binding [57]. In Eq. 5.5, the kinetic formulation according

to Nilsson and coworkers is shown [188]. It describes the protein concentration in the stationary phase q as a function of q itself, in the pore phase c_p, and salt concentration $c_{p,salt}$ in the pore phase.

$$k_{kin,i}\frac{\partial q_i(x,t)}{\partial t} = k_{eq,i}(\Lambda - \sum_{j=1}^{k}(\nu_i + \sigma i)q_j(x,t))^{\nu_i}c_{p,i}(x,t) \tag{5.5}$$
$$-c_{p,salt}(x,t)^{\nu_i}q_i(x,t), \quad \forall i \neq salt$$

Eq. 5.6 describes the salt concentration in the stationary phase as a function of proteins bound to the adsorber surface.

$$q_{salt}(x,t) = \Lambda - \sum_{j=1}^{k}\nu_j q_j(x,t) \tag{5.6}$$

Instead of the adsorption rate coefficient k_{ads} and the desorption rate coefficient k_{des}, the equilibrium coefficient $k_{eq}=k_{ads}/k_{des}$ and the kinetic coefficient $k_{kin} = 1/k_{des}$ were used. In this way, parameter estimation was simplified, since k_{eq} and k_{kin} correlate mainly with the retention time and peak height, respectively [115]. ν is the characteristic charge, also known as the number of binding sites directly involved in binding. Λ is the column-specific ionic binding capacity equal to the number of potential binding sites. Here, the SMA isotherm has been chosen to cover the overloading state in investigated breakthrough experiments. For the description of low protein loading as usually applied in the step gradient experiments for preparative separation, the SDM isotherm would be sufficient. The kinetic formulation has been chosen out of several reasons. According to Carta and Jungbauer, protein adsorption is often slower than small molecules because of limitations in the binding kinetics. In addtion, a true adsorption equilibrium may not be established since the protein may undergo molecular changes due to unfolding, aggregation, or degradation before reaching equilibrium with the surface [189]. Furthermore, Toyopearl GigaCap S-650M is a hydroxylated methacrylic polymer grafted adsorber providing high ligand density. As result, fast adsorption rates may be favored initially, but with increasing protein binding, steric crowding and electrostatic repulsion may limit the access to binding sites [190, 191].

5.2.11. Numerical Methods

The chromatograms resulting from LGE and breakthrough experiments were used to estimate the parameters with the inverse method [116].The adaptive simulated annealing (ASA) [113] yielding the first guess was followed by the Levenberg-Marquardt (LM) algorithm [114] for the fine adjustment of the parameter estimates. Subsequently, the confidence intervals at 95 % level were calculated to verify estimation

reliability. Discretization in space on a grid with equidistant nodes and θ-scheme discretization in time were carried out by employing the finite element method and the fractional step [111], respectively. Picard iteration was employed to approximate the solution of the non-linear equation system [112].

5.3. Results

5.3.1. PEGylation and Purification

In case of lysozyme, six lysine residues and the N-terminal amino group are available as binding sites for the PEG aldehyde reaction [192]. The large number of binding sites allows for the formation of different PEGamers. Preparative isolation of the mono-PEGylated species was performed by a single cation-exchange step. In Fig. 5.1 the resulting chromatograms are shown for 1:12 diluted PEGylation batches with 2 kDa (a), 5 kDa (b) and 10 kDa (c) PEG. After peak fractionation, HT-CGE analysis was performed according to [181] to verify purity and PEGylation degree. As observed and discussed by [174, 176, 193], a decrease in elution volume with increasing PEGylation degree was observed for all PEG molecular weights. The red areas in Figs. 5.1a-c indicate the respective pooling limits for the mono-PEGylated species based on purity requirements greater than 97 %. Purity was determined by HT-CGE analysis according to the analytical protocol established by [181]. The resulting fluorescence signals of HT-CGE for the native lysozyme and the purified mono-PEGylated species with a concentration of $6.99 \cdot 10^{-5}$ M showed a distinct peak broadening of PEGylated proteins compared to the native species (Supplementary Fig. 5.5). By using the calibration established by [181], this peak broadening was taken into account in the calculation of purities.
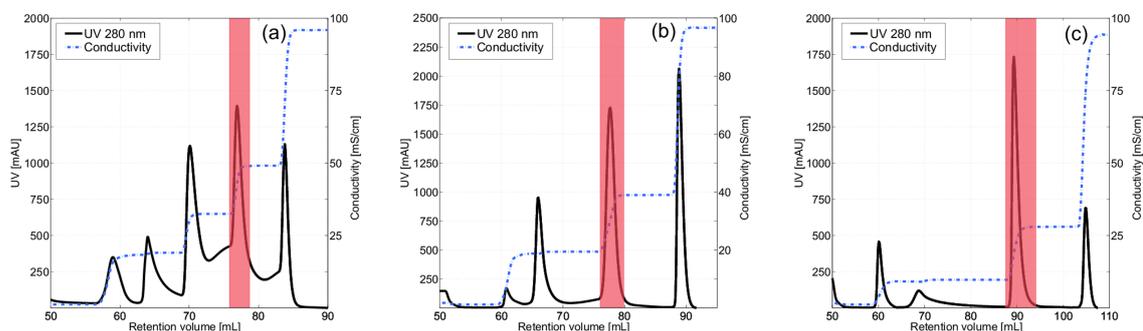
Figure 5.1.: Chromatograms of preparative CEX for 1:12 diluted PEGylation batches (r=6.67, pH 7.2, 3.5 h) loaded with a 50 mL loop for 2 kDa PEG (a), 5 kDa PEG (b) and 10 kDa PEG (c). The red area indicates the respective pooling limits for the mono-PEGylated species.

5.3.2. System Characteristics

Tracer experiments were carried out to determine the system parameters bed voidage, particle voidage, and axial dispersion. The ionic capacity was determined by applying acid-base titration. The results are shown in Tab. 5.2. The axial dispersion was found to be similar to literature data [194].

### 5.3.3. Linear Gradient Elution and Breakthrough Experiments

Linear gradient experiments were carried out to generate information about proteins in the linear region of the adsorption isotherm. The retention time of every protein species over three different salt gradient lengths yielded information about the isotherm parameters characteristic charge ν and equilibrium coefficient k_{eq}. The height, width, and shape of the elution peaks provided partial information about the mass transfer parameters film diffusion coefficient k_{film} and pore diffusion coefficient D_p. Thus, by employing ASA and LM, ν and k_{eq} were estimated with high reliability, for k_{film} and D_p an initial guess was delivered. As can be seen by comparing the dashed lines in Figs. 5.2a, d, g, and j, lysozyme in its native form was the strongest binding species for all investigated gradient lengths. Comparison of the elution peaks of native and PEGylated species at a constant gradient length in Fig. 5.2 reveals that the elution times decreased with increasing PEG chain length. Except for the different elution times of all protein species, their peak heights and widths are highly similar at each salt gradient conditions. A small shoulder peak behind the main peak can be seen in Fig. 5.2d-f, indicating a small amount of a stronger binding protein species. Presumably this species is by unmodified lysozyme, since for the 2 kDa PEGylation no peak baseline seepearation between the different PEGamer species could be achieved in in preparative chromatography (compare Fig. 5.1a).

Table 5.2.: For the Toyoscreen column,the voidages and axial dispersion are calculated from the retention volume and peak broadening of tracer injections. The ionic capacity is determined by acid-base titration.

		GigaCap S-650M
Particle diameter	d_p	$75\,\mu m$
Bed voidage	ϵ_b	0.414
Particle voidage	ϵ_p	0.779
Total voidage	ϵ_t	0.871
Axial dispersion $[mm^2/s]$	D_{ax}	$6.691 \cdot 10^{-2}$
Ionic capacity [M]	Λ	1.389

Additionally, breakthrough experiments were carried out under strong binding condition. The 280 nm signals were highly non-linear above 2000 mAU and reached the detector saturation at approximately 2500 mAU. As shown by the dashed lines in Fig. 5.3, lysozymes with 10 kDa, 5 kDa and 2 kDa PEG attached, and the native lysozyme exhibited their breakthrough in successive order. Based on this information, the shielding parameter σ was estimated and the correlation between k_{kin} and k_{film} that both affect the peak height in the linear part of the adsorption isotherm was dissolved.

The final parameter estimates and the related confidence intervals at 95 % level are summarized in Tab. 5.3. The simulated LGE for the four protein species are displayed in measurements was found for the retention time, peak width, and peak shape. Overall, the conformity was highest for the native lysozyme. The peak heights of PEGylated species were slightly overestimated. The simulated breakthrough curves for all protein species are displayed in Fig. 5.3. Here, the model accurately accounted for the overall slopes and reflected the process relevant times at 10 % and 50 % breakthrough. The relative offsets for the process times at 10 % breakthrough were 1.83 % for lysozyme in the native condition, 3.53 % for lysozyme attached with 2 kDa PEG, 1.93 % for lysozyme attached with 5 kDa, and 4.17 % for lysozyme attached with 10 kDa.
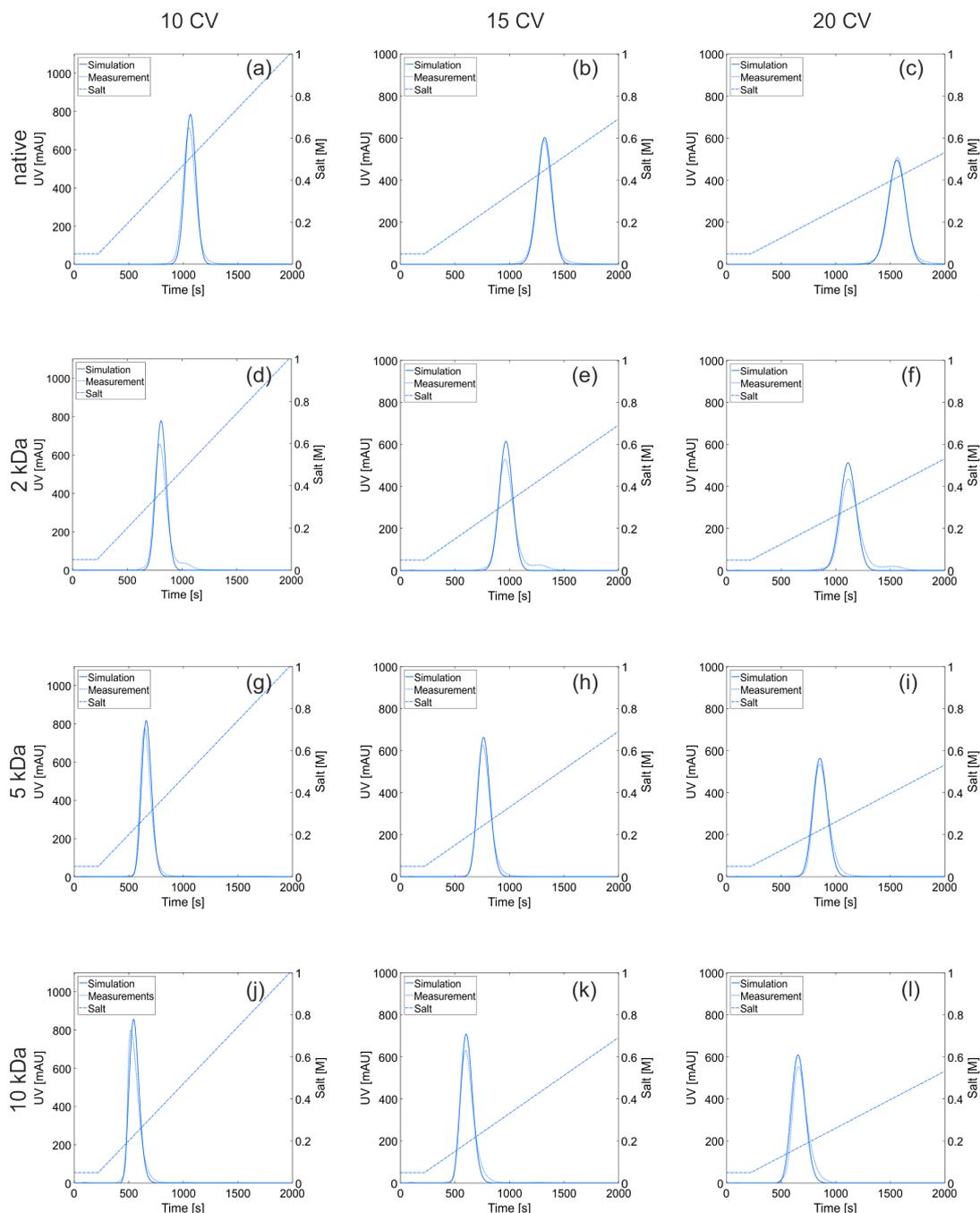
Figure 5.2.: Plots of UV signals over process run-time for bind-and-elute experiments. Dashed lines display the UV signals measured at the column outlet and the adjusted linear salt gradients. Solid lines represent the simulated chromatograms. The elution peaks of native lysozyme, lysozyme attached with 2 kDa PEG, 5 kDa PEG, and 10 kDa PEG by applying linear salt gradients from 0.05 M to 1.0 M over 10 CV, 15 CV, and 20 CV are shown in (a) - (c), (d) - (f), (g) - (i), and (j) - (l). Similar peak heights and widths, but different retention times can be seen for different protein species. Here, the Toyoscreen column was employed.
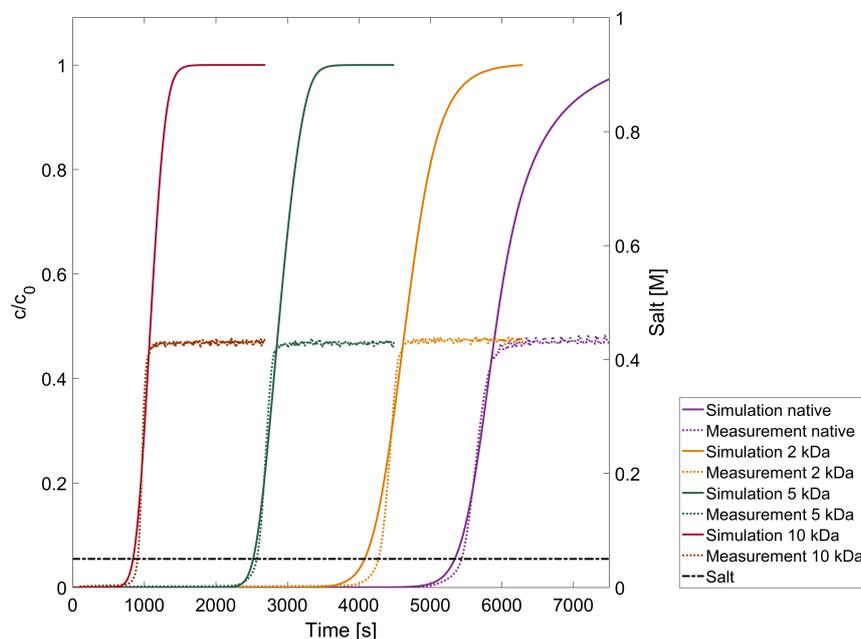
Figure 5.3.: Plots of normalized protein concentration over process run-time for breakthrough experiments. Dashed lines display the normalized protein concentrations calculated from UV Fig. 5.2 as solid lines. In all cases, a good agreement between simulations and signals measured at the column outlet and the constant salt concentration at 0.05 M. Solid lines represent the normalized protein concentrations calculated from the simulated chromatograms. The native lysozyme and lysozyme attached with 2 kDa PEG, 5 kDa PEG, and 10 kDa PEG are shown in purple, yellow, green, and red, respectively. Here, the Toyoscreen column was employed.

5.3.4. Mass Transfer and Kinetic Phenomena

The GRM assumes that the adsorbent particles have a spherical shape and a uniform diameter. The shape of PEGylated proteins is influenced by the surrounding PEG layer which is higlhly dynamic. Due to the high hydration of PEG, PEGylated proteins have a significantly greater hydrodynamic radius than unmodified proteins with a comparable molecular weight. Fee and Van Alstine introduced a nonlinear relationship between the degree of PEGylation in terms of total molecular weight of PEG attached and the hydrodynamic radius of the PEGylated protein [172, 175]. This non-linearity is the reason why the behavior of conjugated proteins in IEX must necessarily be described as a function of the hydrodynamic radius and not in terms of the total molecular weight of attached polymer. According to the correlation introduced by Fee and Van Alstine [172], the hydrodynamic radii $h_{R,PEGprot}$ were calculated to be 2.00 nm, 2.37 nm, 3.08 nm, and 3.84 nm for the four lysozyme species with increasing PEG MW.

Table 5.3.: Parameters of the mass transfer model and kinetic isotherm formulation estimated from bind-and-elute experiments with linear salt gradient and breakthrough curves using the inverse method are shown for native and mono-PEGylated lysozyme species. Confidence intervals at 95 % level reflect the reliability of the parameter estimates.

Parameter	Native	2 kDa PEGylated	5 kDa PEGylated	10 kDa PEGylated
k_{film} $[mm/s]$	$9.95 \cdot 10^{-2}$ $\pm 5.20 \cdot 10^{-2}$	$8.92 \cdot 10^{-2}$ $\pm 5.49 \cdot 10^{-2}$	$6.62 \cdot 10^{-2}$ $\pm 3.12 \cdot 10^{-2}$	$4.07 \cdot 10^{-2}$ $\pm 2.53 \cdot 10^{-2}$
D_p $[mm^2/s]$	$2.85 \cdot 10^{-4}$ $\pm 2.38 \cdot 10^{-5}$	$1.33 \cdot 10^{-4}$ $\pm 7.56 \cdot 10^{-6}$	$8.42 \cdot 10^{-5}$ $\pm 2.08 \cdot 10^{-6}$	$5.75 \cdot 10^{-5}$ $\pm 2.28 \cdot 10^{-6}$
k_{eq} [-]	$4.62 \cdot 10^{-2}$ $\pm 9.16 \cdot 10^{-5}$	$5.94 \cdot 10^{-3}$ $\pm 2.79 \cdot 10^{-5}$	$1.16 \cdot 10^{-3}$ $\pm 5.13 \cdot 10^{-6}$	$1.92 \cdot 10^{-5}$ $\pm 1.48 \cdot 10^{-6}$
k_{kin} $[sM^{-\nu}]$	$3.94 \cdot 10^{-2}$ $\pm 7.56 \cdot 10^{-4}$	$4.58 \cdot 10^{-3}$ $\pm 2.10 \cdot 10^{-4}$	$2.31 \cdot 10^{-4}$ $\pm 6.74 \cdot 10^{-5}$	$6.34 \cdot 10^{-6}$ $\pm 2.82 \cdot 10^{-5}$
ν [-]	4.21 $\pm 1.82 \cdot 10^{-3}$	4.21 $\pm 2.65 \cdot 10^{-3}$	4.20 $\pm 1.10 \cdot 10^{-3}$	4.22 $\pm 1.56 \cdot 10^{-3}$
σ [-]	5.61 $\pm 1.27 \cdot 10^{-2}$	6.81 $\pm 1.35 \cdot 10^{-2}$	9.79 $\pm 1.67 \cdot 10^{-2}$	25.90 $\pm 1.08 \cdot 10^{-1}$

For lysozyme, the mass transfer coefficient k_{eff} calculated according to $1/k_{eff} = 1/k_{pore} + 1/k_{film}$ with the internal mass transfer resistance $k_{pore} = 10 D_p \epsilon_p / d_p$ [41] was found to be consistent with literature data [194]. An approximately linear decrease of the film diffusion coefficient k_{film} with increasing $h_{R,PEGprot}$ was determined as displayed in Fig. 5.4 a. A comparable dependency was reported by Mejía-Manzano et al. [195] for affinity chromatography. As shown in Fig. 5.4 b, the pore diffusion coefficient D_p decreased reciprocally with increasing $h_{R,PEGprot}$ according to the Stokes-Einstein equation qualitatively. The adsorption and desorption rate coefficients $k_{ads} = k_{eq}/k_{kin}$ and $k_{des} = 1/k_{kin}$ were calculated and are displayed in Figs. 5.4 c and d. With increasing PEG MW, both k_{ads} and k_{des} showed an exponential increase. A similar trend has been observed by Mejía-Manzano et al. [195] for affinity chromatography. The increase of k_{des} exceeded the increase of k_{ads} by more than two orders of magnitude. For native and 5 kDa mono-PEGylated lysozyme, k_{eq} was found to be of the same magnitude as reported in the literature [178].

5.3.5. Characteristic Charge and Shielding

The characteristic charges ν of PEGylated lysozyme (4.20-4.22) were found to be equal to the value determined for native lysozyme (4.21) as shown in Fig. 5.4 e. ν was unaffected by PEGylation degree and PEG chain length. This finding was consistent with data delivered by Abe and coworker [178]. A small shielding factor
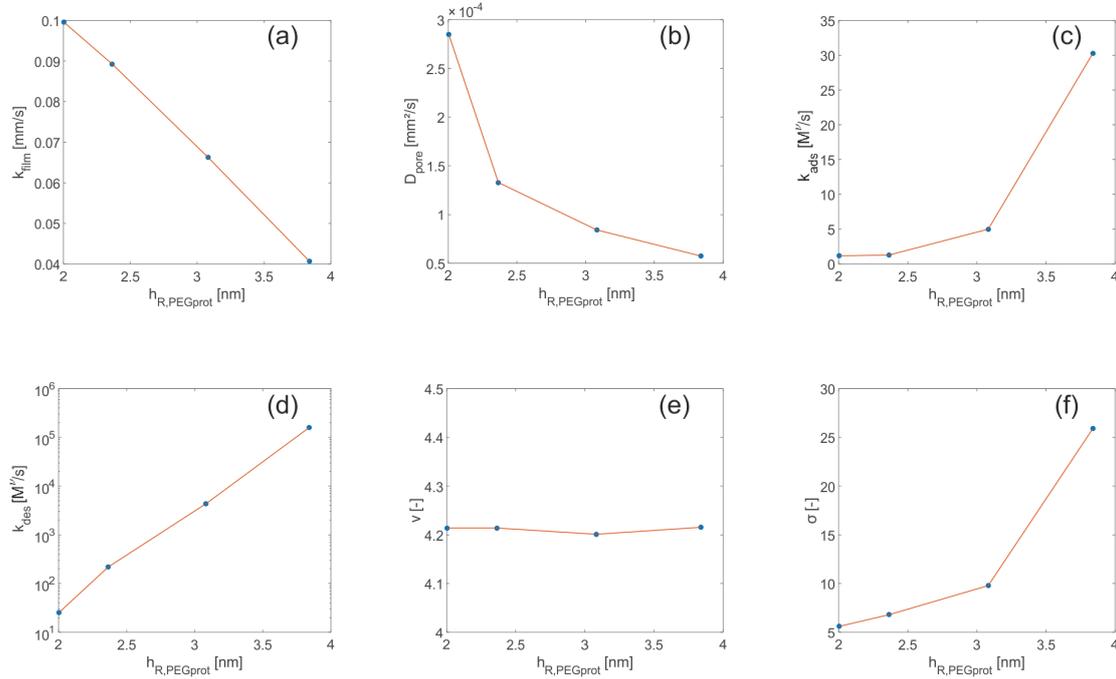
Figure 5.4.: (a) - (f) the show film diffusion coefficient, pore diffusion coefficient, adsorption coefficient, desorption coefficient, characteristic charge, and shielding factor versus the hydrodynamic radius $h_{R,PEGprot}$ of PEGamers. $h_{R,PEGprot}$ takes into account the non-linear relationship between conjugate size and total molecular weight of attached PEG [172, 175]. Blue dots from left to right represent the native lysozyme and lysozyme attached with 2 kDa PEG, 5 kDa PEG, and 10 kDa PEG.

σ of 5.61 was found for the native lysozyme. With increasing PEG chain length, σ increased from 6.81 for 2 kDa to 9.79 for 5 kDa, and up to 25.90 for 10 kDa PEG as displayed in Fig. 5.4 f. The dependency of σ on the hydrodynamic radius was highly non-linear. Based on the definition of q_{max} being $\Lambda/(\sigma + \nu)$, the maximal binding capacity q_{max} for the four protein species was calculated to be $1.41{\cdot}10^{-1}$ M for native species, $1.26{\cdot}10^{-1}$ M for 2 kDa PEGylated species, $9.93{\cdot}10^{-2}$ M for 5 kDa PEGylated species and $4.61{\cdot}10^{-2}$ M for 10 kDa PEGylated species. q_{max} was found to be reduced by 10.6 % when attached with 2 kDa PEG by 29.6 % when attached with 5 kDa PEG, and by 67.3 % when attached with 10 kDa PEG compared to the native lysozyme species.

5.4. Discussion

PEGylation is commonly used in biopharmaceutical industry to improve protein sta-bilities and pharmacokinetics of protein drugs. However, the currently used reaction mechanisms and conditions usually result in a heterogeneous product mixture of unre-acted protein and conjugates with varying number and modification site of attached polymers [172, 173]. For this reason, purification processes of PEGylated proteins are imperative. Chromatographic processes based on electrostatic interactions e.g. ion-exchange chromatography, are among the most effective purification processes for this application [179]. So far, the development of ion-exchange steps for the purification of the individual PEGamers has been driven mainly by expert-based or experimental approaches (high-throughput process development and statistical design of experiments). These approaches are time-consuming and cost-intensive due to the wide variety of proteins and polymers (linear vs. branched, molecular weight etc.). Mechanistic modeling and simulations can help to reduce the number of experiments during process optimization by *in silico* predictions [85]. From the perspective of process development, the parameters estimated by mechanistic modeling can be used for process up scaling, process optimization, and process control, meeting the demands of the Quality by Design approach (QbD) proposed by the US food and drug administration (FDA) [17].

In our work, the SMA isotherm in kinetic formulation coupled with the GRM produced a comprehensive description of the adsorption and desorption behavior on the adsorber surface, steric hindrance, and the mass transfer for native lysozyme and its PEGylated species. The model parameters k_{film}, D_p, k_{eq}, k_{kin}, ν, and σ were determined and k_{ads}, k_{des}, $h_{R,PEGprot}$, and q_{max} were calculated to improve the mechanistic understanding of PEGylated proteins in CEX. It should be mentioned, that the PEGylation reaction usually delivers various PEGamer isoforms. In the presented case, the isoforms behaved highly similar and could not be separated with the used CEX setup. Hence all isoforms of each lysozyme species had to be modeled as lumped components, resulting in slight overestimation of the peak heights.

As reported by many researchers, PEGylated proteins elute earlier than their native analogs [176, 178, 179]. Based on the observation of the elution order of different PEGylated species beeing the same in both cation-exchange and anion-exchange chromatography, Seely and Richey suggested the 'charge-shielding effect' to explain this phenomenon [176]. Later, Abe and coworkers applied the equilibrium stoichiome-tric displacement isotherm model (SDM) to describe the retention time of PEGylated proteins in linear gradient experiments and determined similar numbers of binding sites for lysozyme and BSA attached with PEG of different lengths. Furthermore, they reported the decrease of a lumped parameter consisting of the equilibrium coefficient, the binding site, and the ionic binding capacity with increasing PEG chain length [178]. In this way, the 'charge-shielding effect' hypothesis was verified

and the equilibrium coefficient was identified to be responsible for the weaker binding of PEGylated proteins [178].

k_{film} showed a linear dependency on $h_{R,PEGprot}$ as expected according to the correlation suggested by Jungbauer and Carta [189]. Its decrease with increasing PEG chain length was to be reflected by broader and lower elution peaks. However, the LGE under same operating conditions showed similar peak heights and widths for all PEGylated and native species. Considering the fact that there is a strongly exponential correlation of k_{des} with $h_{R,PEGprot}$, a diffusion-desorption-compensation effect is suggested to be responsible for the uniformity in peak heights and width. The faster desorption of proteins attached with longer PEG chain may be neutralized by the slower film diffusion. This hypothesis is highly consistent with the widely accepted view of the 'charge-shielding effect', since weaker charged proteins increasingly tend to undergo desorption. D_p showed a reciprocal correlation with $h_{R,PEGprot}$, as had been expected according to the Stokes-Einstein equation [196], which is reflected by slight tailing of elution peaks in LGE. D_p was found to exceed the molecular diffusion coefficient for native lysozyme. As intensively studies in literature, there are two opinions to explain this effect. Carta et al. and Rodrigues et al. [197–199] introduced the convection-enhanced effective intra-particle diffusivity. Many more experimental examples of intra-particle convection in protein chromatography could be found in the literature [200–204]. However, convective mass transfer into the bead interior was observed for large pores ($> 5000\,\mathring{A}$)[204, 205]. For small pores up to 700 Å, Nash et al. assumed diffusional mass transport only. For the TOYOPEARL® GigaCap S-650M resin having an average pore size of 1000 Å, the observed molecular diffusion coefficient cannot be explained completely by convective mass transport in the pores. An additional effect observed by Dziennik et al. [206] for porous resins with high charge density applies to TOYOPEARL® GigaCap S-650M. They found indications that non-diffusive mechanisms of electrostatic origin could enhance protein uptake rates in ion exchange particles, resulting in enhanced effective pore diffusivities.

The shielding factor σ showed an exponential increase with increasing PEG chain length. In comparison to native lysozyme, approximately 12 %, 43 %, and 207 % more free binding sites are sterically hindered by the species with 2 kDa, 5 kDa, and 10 kDa PEG attached, respectively. In contrast to this, ν was found to be independent of PEGylation and PEG chain length, indicating the same binding orientation for all species. Thus, the steric hindrance of free binding sites was identified to be the main contributor to the observed exponential decrease of molar binding capacity q_{max} upon PEGylation. It is indicated that the longer PEG chains of an adsorbed protein make many more free binding sites inaccessible than the shorter ones or equally sized unmodified proteins. Fee and Van Alstine proposed a correlation for the average shape of the PEG-layers around a protein over timescale [172, 173]. These layers are expected to have increasing degree of dynamics with increasing PEG molecular weight [207]. This concept could also explain the non-linearity in k_{ads}, k_{des}, and σ shown in Fig. 5.4c, d, f.

Especially in the overloading state under strong binding condition, a high density of proteins bound could result in formation of multiple PEG chain layers covering adjacent free binding sites. The multiple PEG chain layers would not shield the electrostatic interactions, but keep the proteins in the mobile phase distant from the adsorber surface, so that the electrostatic attraction would become weak and binding impossible. This hypothesis is schematically represented by Fig. 5.5. Along the increasing binding density, several transitional states are supposed to exist. First, in the linear part of the adsorption isotherm, the proteins could distribute uniformly on the adsorber surface; secondly, unfavorable binding sites between the proteins covered by thin PEG chain layers could be occupied, though the electrostatic attraction could already be reduced; finally, the multiple PEG chain layers could become dominant, so that the electrostatic attraction could disappear and binding could be suppressed. This concept is consistent with the observation made by Blaschke and coworkers [208]. They found that adsorption was less enthalpy-driven at higher loading states for proteins attached with longer PEG chains. Of course, the PEG-layers around PEGylated proteins are not static, rather of dynamic nature. Thus, the mechanistic chromatography model describes the average behavior of lysozyme species in CEX. As suggested by Fee et al., the dynamicism of PEG-layers tend to increase with increasing PEG chain length. This concept could be an alternative explanation for the nontrivial behavior of PEGylated lysozymes observed in the presented work.

Using mechanistic chromatography modeling and considering insights provided by former pioneer work, the hydrodynamics and thermodynamics of PEGylated lysozymes in CEX were investigated. The diffusion-desorption-compensation effect was introduced to explain the anomaly of peak heights and widths remaining constant in spite of an increasing hydrodynamic radius. Additionally, it reflects the exponential dependency of the shielding factor on the MW of PEGylated proteins and suggests that multiple PEG chain layers formed in the overloading state are responsible for this non-trivial phenomenon. Thus, the model view of PEGylated proteins' behavior in CEX was supplemented by the overloading state.

This study clearly demonstrates that mechanistic chromatography modeling can be applied to describe PEGylated proteins with high accuracy and reliability. Thus it has great potential for the optimization, prediction, and scale-up of purification processes for PEGylated proteins. A future challenge is to show whether the separation of positional isoforms can be predicted by this kind of simulation. In this respect, a combination of mechanistic chromatography modelling combined with molecular modeling could be profitable.
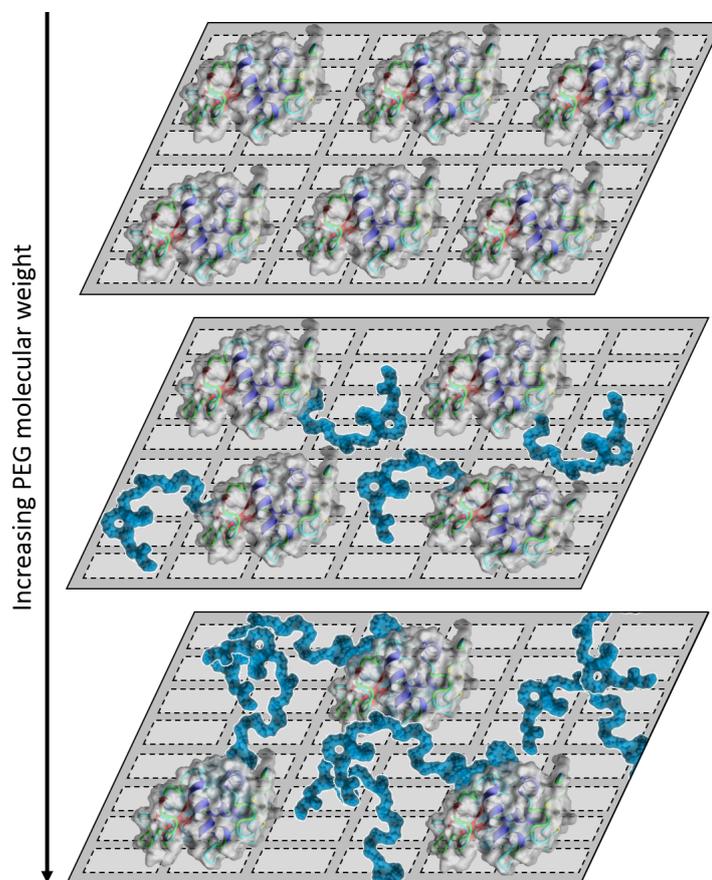
Figure 5.5.: Molecular picture of the adsorption of lysozymes on an adsorber surface. Increasing PEG chain length results in the formation of multiple PEG chain layers hindering the binding of further lysozymes. The reduction of accessible binding sites explains the observed decrease in binding capacity upon protein PEGylation. (Molecular graphic of lysozyme (PDB: 1LYZ) was created with YASARA (www.yasara.org)).

5.5. Acknowledgments

6. An Integrated Precipitation and Ion-exchange Chromatography Process for Antibody Manufacturing: Process Development Strategy and Continuous Chromatography Exploration

Steffen Großhans*, Gang Wang*, Christian Fischer and Jürgen Hubbuch**

Institute of Engineering in Life Sciences, Section IV: Biomolecular Separation Science, Karlsruhe Institute of Technology (KIT), 76131 Karlsruhe, Germany
* These authors contributed equally to this work.
** Corresponding author.

Abstract

In the past decades, research was carried out to find cost-efficient alternatives to Protein A chromatography as a capture step in monoclonal antibody (mAb) purification processes. In this work, polyethylene glycol (PEG) precipitation has shown promising results in the case of mAb yield and purity. Especially with respect to continuous processing, PEG precipitation has many advantages, like low cost of goods, simple setup, easy scalability, and the option to handle perfusion reactors.

Nevertheless, replacing Protein A has the disadvantage of renouncing a platform unit operation as well. Furthermore, PEG precipitation is not capable of reducing high molecular weight impurities (HMW) like aggregates or DNA. To overcome

these challenges, an integrated process strategy combining PEG precipitation with cation-exchange chromatography (CEX) for purification of a mAb is presented.

This work discusses the process strategy as well as the associated fast, easy, and material-saving process development platform. These were implemented through the combination of high-throughput methods with empirical and mechanistic modeling. The strategy allows the development of a common batch process. Additionally, it is feasible to develop a continuous process.

In the presented case study, a mAb provided from cell culture fluid (HCCF) was purified. The precipitation and resolubilisation conditions as well as the chromatography method were optimized, and the mutual influence of all steps was investigated. A mAb yield of over 95.0 % and a host cell protein (HCP) reduction of over 99.0 % could be shown. At the same time, the aggregate level was reduced from 3.12 % to 1.20 % and the DNA level was reduced by five orders of magnitude. Furthermore, the mAb was concentrated three times to a final concentration of $11.9\,mg/mL$.

6.1. Introduction

From 1992 to 2015, 47 monoclonal antibodies (mAbs) have been approved either by the US Food and Drug Administration (FDA), or the European Medicines Agency (EMA) [4]; until January 11, 2017, the number of approved therapeutic mAbs by FDA has increased to 68 [5]. Currently, 230 and 52 mAbs have been reported to undergo the phase 2 and phase 3 clinical studies, respectively. Compared to early 2010 (26 mAbs in late-stage studies), a 100 % increase has been achieved, delivering the evidence for a continuing strong growth of the mAb market [209]. The global mAb market was currently expected to reach almost $ 125 billion by 2020 [6].

Based on recent advances in upstream processing, a titer increase at commercial scale up to approximately 20 % from 2014 to 2019 has been predicted, whereas improvements in downstream processing have been adopted more slowly [13]. In 2016, FDA has approved the first mAb biosimilar infliximab-dyyb for the US market, which will be sold with an estimated discount at 20-30 % to its originator. The game-changing biosimilars, along with the increasing demand for mAbs, and achievements in upstream processing have placed downward pressure on downstream processing.

As Protein A chromatography provides highly specific binding to the Fc-region of a variety of mAb types, resulting in high product yield and purity in a single step, it has become the golden standard for the capture step in the platform process [210–212]. However, its disadvantages such as slow volumetric throughput and high resin costs have become even more pronounced under the circumstances mentioned above. As a result, a growing interest in the development of non-chromatographic mAb purification processes could be observed [131, 213–218]. Especially precipitation

methods based on polyethylene glycol (PEG) showed promising results as alternative capture step [219, 220]. Hammerschmidt *et al.* [221] investigated the economics of precipitation in comparison to the actual platform process. The results showed cost reduction for the precipitation process during clinical as well as in the commercial phase. Especially in the clinical phases and in small scale production replacing Protein A resins showed economical benefit.

Additionally, a trend from conventional single column batch chromatography towards multi-column chromatography has been reported [222–228]. While operating in semi-continuous fashion, the overall operating time can be reduced. Furthermore, higher resin utilization per cycle leads to columns with reduced volume. Thus, an increased productivity may be achieved.

On the process side, Protein A chromatography has been established as an expensive but easy-to-use technique for mAb capturing, which does not require high development effort. In comparison, the process development for both precipitation and multi-column chromatography, however, is much more complicated, which contradicts the desire for shorter development timelines in the biopharmaceutical industry. Thus, the necessity of simplification of process development while maintaining product quality is a major challenge.
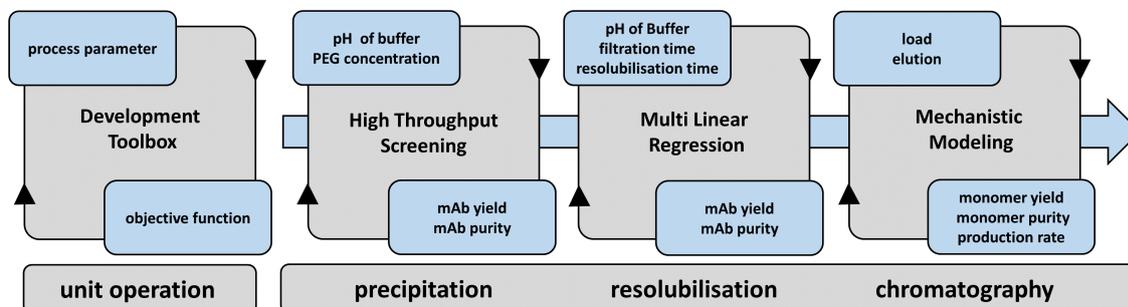
Figure 6.1.: Flowchart of the development toolbox and the integrated process which combines precipitation, resolubilisation and cation-exchange chromatography.

In this work, a case study on industrial mAb purification from a harvested cell culture fluid (HCCF) by integrating PEG precipitation and cation-exchange chromatography (CEX) is presented. The occurring difficulties by renouncing Protein A chromatography are addressed. A combined process development strategy minimizing the

amount of feedstock needed and experimental effort is suggested. In particular, high-throughput experimentation, design of experiments (DoE), and mechanistic chromatography modeling were employed to develop the precipitation, resolubilisation, and CEX chromatography step. The developed toolbox and the process scheme of the case study are illustrated in Fig 6.1. Additionally, based on mechanistic modeling, the optimized single-column batch chromatography process was transferred to a three-column CEX periodic counter-current chromatography (3C-PCCC) mode to investigate the advantages and potential of multi-column chromatography in the considered case.

6.2. Theory

6.2.1. Fundamentals of Precipitation

Protein precipitation has been known for a long time as a separation method for the purification of complex protein solutions [229]. Phase separation is carried out by adding precipitation agents, like inorganic salts, organic solvents, or nonionic polymers to the protein solution [124–127]. For the precipitation of a complex mAb solution, in particular the use of polyethylene glycol (PEG) with an average molecular weight of 6000 DA as precipitation agent has shown promising results [64]. Although this phenomenon is well investigated and there are several models describing the behavior of proteins in PEG solution, the mechanism is not completely understood [65, 66, 230]. The two theories which describe the mechanism of PEG-induced precipitation are, on the one hand, the theory of excluded volume [138–140] and on the other hand the theory of attractive depletion [231, 232]. The theory of excluded volume is based on the assumption that PEG traps the solvent in the coiled structure and the folded filament of the solved molecules. This results in the steric exclusion of proteins from the solvent regions occupied by the polymers. Molecules with a larger hydrodynamic radius are more strongly excluded compared to molecules with a smaller one. According to this view, the sterically excluded proteins are concentrated until their solubility is exceeded and precipitation occurs [68]. In contrast to this assumption, the theory of attractive depletion describes a force of the order of osmotic pressure which leads to protein precipitation. This force is caused by mainly steric exclusion of the PEG molecules from the ambience of the proteins, described as depletion zone. If two depletion zones are overlapping, a PEG-free area is created and, therefore, a concentration gradient is formed. This gradient leads to an osmotic pressure which further leads to precipitation of the proteins. Both theories describe the mechanism of PEG-induced precipitation but additionally other factors which influence the colloidal stability of protein solutions like protein concentration, pH or ionic strength, have to be considered [219].

6.2.2. Ion-exchange Chromatography Modeling

The following theoretical background was used to carry out mechanistic chromatography modeling for the subsequent CEX step. The general rate model (GRM) is employedf to cover the convection and diffusion in a chromatography column of length L and adsorber beads of radius r_p [41]:

$$
\begin{aligned}
\frac{\partial c_i(x,t)}{\partial t} &= -u(t)\frac{\partial c(x,t)}{\partial x} + D_{ax}\frac{\partial^2 c(x,t)}{\partial x^2} \\
&\quad -\frac{1-\epsilon_b}{\epsilon_b}k_{film,i}\frac{3}{r_p}\left(c_i(x,t) - c_{p,i}(x,r_p,t)\right)
\end{aligned}
\tag{6.1}
$$

$$
\frac{\partial c_{p,i}(x,r,t)}{\partial t} =
\begin{cases}
\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 D_{p,i}\frac{\partial c_{p,i}(x,r,t)}{\partial r}\right) - \frac{1-\epsilon_p}{\epsilon_p}\frac{\partial q_i(x,r,t)}{\partial t} & \text{for } r \in (0, r_p), \\
\frac{k_{film,i}}{\epsilon_p D_{p,i}}(c_i(x,t) - c_{p,i}(x,r_p,t)) & \text{for } r = r_p, \\
0 & \text{for } r = 0.
\end{cases}
\tag{6.2}
$$

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

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

Within the mobile phase $c_i(x,t)$, the mass transfer between the pore volume $c_{p,i}(x,r,t)$ and the interstitial volume is described by Eq. (6.1). The voidage of the adsorber bed ϵ_b, the axial dipersive coefficient D_{ax}, and the film transfer coefficient $k_{film,i}$ describe the peak-broadening effects in the interstitial volume. The mass transfer is from the pore volume concentration $c_{p,i}$ to the stationary phase q_i and vise versa. ϵ_p is the adsorber particle voidage and $D_{p,i}$ the component-specific pore diffusion coefficient. Eqs. (6.3) and (6.4) are the Danckwerts boundary conditions.

The widely accepted steric mass action (SMA) isotherm model introduced by Brooks and Cramer describes adsorption and desorption processes on adsorber surfaces [90]. The kinetic formulation suggested by Nilsson *et al.* is shown in Eq. 6.5 [188]. Here, the stationary phase protein concentration q depends on q itself, pore phase protein concentration c_p, and pore phase salt concentration $c_{p,salt}$. ν is the characteristic charge representing the binding sites number involved in binding directly. σ is the number of binding sites on the adsorber surface sterically hindered by protein binding. Λ is the ionic binding capacity of the column related to adsorber skeleton.

$$
\begin{aligned}
k_{kin,i}\frac{\partial q_i(x,t)}{\partial t} &= k_{eq,i}(\Lambda - \sum_{j=1}^{k}(\nu_i + \sigma i)q_j(x,t))^{\nu_i}c_{p,i}(x,t) \\
&\quad -c_{p,salt}(x,t)^{\nu_i}q_i(x,t) \quad \forall i \neq salt
\end{aligned}
\tag{6.5}
$$

The salt concentration in the stationary phase is described in Eq. 6.6.

$$q_{salt}(x,t) = \Lambda - \sum_{j=1}^{k} \nu_j q_j(x,t) \qquad (6.6)$$

The adsorption rate coefficient k_{ads} and the desorption rate coefficient k_{des} were replaced by the equilibrium coefficient $k_{eq}=k_{ads}/k_{des}$ and the kinetic coefficient $k_{kin} = 1/k_{des}$ to simplify parameter estimation, as k_{eq} and k_{kin} correlate mainly with the retention time and peak height, respectively [115]. Out of several reasons, the kinetic formulation has been chosen. Because of limitations in the binding kinetics, protein adsorption is often slow as reported by Carta and Jungbauer. Additionally, true adsorption equilibrium may not be established in column chromatography. The hydroxylated methacrylic polymer-based and polymer-grafted adsorber (Toyopearl GigaCap S-650M) provides high ligand density, resulting in limited access to binding sites with increasing protein binding due to steric crowding and electrostatic repulsion [190, 191].

Combining both GRM and SMA, the presented CEX chromatography process can be modeled using numerical methods.

6.2.3. 3C-PCCC

Periodic counter-current chromatography is a simulated moving-bed process patented by GE Healthcare as a semi-continuous purification process [233]. Usually, it is operated in three- or four-column configurations.

Since PCCC allows the loading of columns much closer to their static binding capacity, it has the potential to reduce resin volume, buffer consumption, and process time in comparison to conventional batch chromatography [222–224, 234]. Economical simulation of a Protein A capture step in 3C-PCCC and 4C-PCCC mode has demonstrated that PCCC can reduce manufacturing costs particularly in early clinical phases [234]. Compared to traditional batch single-column chromatography, where the process steps of a chromatography cycle are conducted consecutively, in PCCC, multiple steps are performed simultaneously but distributed along several columns. In 3C-PCCC, the process consists of two interconnected columns and one short-circuited column. The columns are shifted only periodically after two out of four process steps. In the first stage, column 3 is interconnected with column 2. In steady-state, column 3 was fed with feed material in the previous steps and is now fully saturated. Wash buffer is passed through this column to remove unbound and weakly bound proteins. The wash effluent is directed to the equilibrated column 2 to recover leached product. In the meantime, feed material is loaded onto column 1. The flow-through is discarded until the first product breaks through. As soon as the first product breaks through, column 1 is switched into the interconnected lane and

the outflow is loaded directly onto column 2. Since the load step is now performed with two columns in series, resin capacity utilization can be increased by loading column 1 beyond 1 % product breakthrough [233].

While column loading is continued in the interconnected lane, column 3 is short-circuited and the product is eluted. The short circuit allows the application of arbitrary elution gradients and center-cut separation. After product elution, column 3 is regenerated and equilibrated. Column 1 and column 2 remain in the interconnected lane and continue to receive feed material without product breaking through column 2. The procedure is repeated several times, so that each column is loaded, washed, eluted, and regenerated several times in the process. In addition to column regeneration, it might also be necessary to integrate a CIP step. This step can be carried out after a certain number of process cycles, or, alternatively, as a part of each regeneration step [233].

6.3. Materials and Methods

6.3.1. Feed Stock

The mAb used in this work was provided as HCCF from LEK d.d. (Ljubljana, Slovenia). The HCCF was stored at -80$°C$ for long-term and at -30$°C$ for short-term storage. All samples were filtered through 0.2 μm syringe filters (Satorius, Göttingen, Germany) directly before use.

6.3.2. Disposables

All precipitation experiments were carried out in 350 μl polypropylene flat bottom 96-well micro plates (Grainer Bio-One, Kremsmünster Austria). For all resolubilisation studies, 350 μl AcroPrep filter plates with a 0.2 μm GH polypro membrane (Pall Corporation, Port Washington, NY, USA) were used. For DNA quantification, black 96-well polystyrene microplates (Thermo Fisher Scientific, Waltham, MA, USA) were used. Before loading onto the CEX column, pH of the mAb samples was adjusted using pD 10 desalting columns (GE Healthcare, Little Chalfont, UK).

6.3.3. Chemicals and Stock Solutions

As buffer substance, acetic acid, sodium acetate trihydrate, sodium chloride, sodium hydrogen carbonate, and tris(hydroxymethyl)-aminomethane (all Merck KGaA, Darmstadt, Germany) were used. Tris hydrodochloride was obtained from PanReac AppliChem (Darmstadt, Germany). Sodium carbonate was obtained from Sigma Aldrich (St. Louis, MO, USA). The PEG with an average molecular mass of 6000 was obtained from Merck KGaA (Darmstadt, Germany). All buffers were prepared with a concentration of $100\,mM$. For this, the appropriate amounts of associated buffer components were weighted and dissolved in ddH_2O. The desired pH was reached by varying the amount of acid and basic component for each buffer. For the $40\,\%\,(w/w)$ PEG 6000 stock solution, the buffer components were first dissolved in ddH_2O followed by adding the appropriate amount of PEG 6000.

6.3.4. Liquid Handling Station

All precipitation experiments were carried out on a Tecan Freedom Evo 200 System liquid handling station (Tecan, Männedorf, Switzerland). The liquid handling station was equipped with an 8-tips liquid handling arm, a Te-VacS vacuum separation module, a Te-Shake orbital shaker (all Männedorf, Switzerland), and a Rotanta 46RSC centrifuge (Hettlich GmbH & Co. KG, Tuttlingen, Germany). The system was controlled by Evoware 2.5 (Tecan, Männedorf, Switzerland). Excel 2013 (Microsoft, Redmond, WA, USA) was used as data import format and for data storage. All calculations were done using Matlab R2015a (The Mathworks, Natick, MA, USA).

For reproducible HTS, the 8-tips liquid handling arm was calibrated. A separate liquid class for each buffer, precipitant, and protein solution was created. The calibration procedure was described earlier by Oelmeier *et al.* [235] in detail. By variation of air gaps and the adjustment of the aspiration, dispensing and beakeoff speed accuracy pipetting, even for viscous solutions, could be ensured.

6.3.5. Liquid Chromatography System

The chromatographic experiments for mechanistic model calibration were performed using an Ettan liquid chromatography (LC) system equipped with a pump module P-905, a dynamic single chamber mixer M-925 ($90\,\mu L$ mixer volume), an autosampler A-905, and a fraction collector Frac-950. The LC system was additionally equipped with an UV monitor UV-900 ($3\,mm$ optical path length) and a pH and conductivity monitor pH/C-900 (all GE Healthcare, Little Chalfont, Buckinghamshire, UK). The LC system was controlled with UNICORN™ 5.31.

As chromatography resin, the strong cation exchanger TOYOPEARL™ GigaCap S-650M (TOSOH, Stuttgart, Germany) was applied. The resin beads are comprised

of a polymethacrylate backbone with a mean particle size of 75 μm. The resin was prepacked by TOSOH in three $1\,mL$ ToyoScreen™ columns with dimensions of $6.4\,mm\,x\,3\,cm$.

6.3.6. Analytical Methods

To determine mAb concentration and aggregate content, the UltraHPLC system Ultimate 3000RSLC, controlled with Chromeleon 6.8 (both Thermo Fisher Scientific, Waltham, MA, USA) was used. For mAb concentration, a POROS analytical Protein A column (Thermo Fisher Scientific, Waltham, MA, USA) was connected to a UHPLC system. Analytical size exclusion chromatography (SEC) was performed using a TSKgel SuperSW mAb HTP column (TOSOH, Stuttgart, Germany). Host cell protein (HCP) concentrations of all samples were determined using a microfluidic CD-based ELISA-like assay on the Gyrolab XPlore station controlled by Gyrolab (Gyros AB, Uppsala, Sweden). DNA was quantified using a Quant-iT dsDNA assay kit (Thermo Fisher Scientific, Waltham, MA, USA). The fluorescence intensity was measured using a Tecan Infinity 200 UV-Vis spectrometer (Tecan, Männedorf, Switzerland).

6.3.7. High-throughput Method for Precipitation Screening

The precipitation procedure on the automated liquid handling station was described in detail by Oelmeier *et al.* [32]. All experiments were carried out at $20°C$, controlled by air conditioning. Systems with a total volume of $250\,\mu l$ containing an mAb concentration of $1.5\,mg/mL$ were prepared. For finding the optimal precipitation conditions, the PEG concentration was varied in 24 steps from $0\,\%\,(w/w)$ to $13.8\,\%\,(w/w)\,PEG$ 6000. Additionally, the pH was varied in four steps from pH 7.5 to pH 10.5. After adding the complex protein stock solution, the system was incubated for $15\,min$ on the orbital shaker at $1000\,rpm$, and afterwards $15\,min$ without shaking. To analyze the amount of precipitated protein, the microplate was centrifuged for $30\,min$ at $4000\,rpm$. Then, $100\,\mu l$ of the supernatant was sampled and diluted at a ratio of 1:2. Subsequently, mAb and HCP concentration of the samples were analyzed.

6.3.8. Optimization of the Resolubilisation Using Empirical Modeling

To find optimal resolubilisation process parameters, a DoE was established using the software Modde 10.1.1 (Umetrics, Umeå, Sweden). As design space, a full factorial design with three factors was chosen. In preliminary experiments, the mAb showed instabilities below pH 4.0 during resolublisation (data not shown). Therefore, the pH value of the resolubilisation buffer was varied between pH 4.0 and pH 6.0. The filtration time after precipitation and resolubilisation was varied between $150\,s$ and

$600\,s$, and the incubation time of the resolubilisation was varied between $75\,min$ and $150\,min$. As response, mAb yield and the mAb purity were investigated. The experiments for resolubilisation were carried out on the liquid-handling station with a method similar to the one described before. Instead of mixing the systems in 96-well microplates, 96-well filterplates were used. After incubation, phase separation was carried out applying three times $700\,mbar$ pressure difference for various times. Subsequently, $240\,\mu$l of resolubilisation buffer were added to each system. Afterwards, the filterplates were incubated on the orbital shaker for various times. After the incubation, a second filtration step was carried out using the vacuum separation module three times at $700\,mbar$ pressure difference for various times.

6.3.9. Chromatography System Characterization

The ÄKTA™ system and chromatography column were characterized using tracer pulse injections at constant flow rates of $0.33\,mL/min$. $25\,\mu$L of $10\,g/L$ non-interacting, non-pore-penetrating tracer blue dextran $2000\,kDa$ (Sigma-Aldrich, St. Louis, MO, USA) and $25\,\mu$L of $1\,\%(v/v)$ pore-penetrating, non-interacting tracer acetone (Merck, Darmstadt, Germany) in ultra-pure water were used to determine the system voidage, interstitial volume and total voidage of the column. Here, the UV signals at $260\,nm$ were used and corrected due to system dead volumes. The ionic capacity Λ of GigaCap S-650M was determined by applying acid-base titration suggested by Huuk *et al.* [165].

6.3.10. Bind-and-elute Experiments

After the precipitation and resolubilisation steps, linear gradient elution (LGE) experiments were carried out by means of model calibration. To determine the model parameters in the linear region of the adsorption isotherm, a $0.5\,mL$ loop was used. Post-loading wash steps of $1\,CV$ binding buffer were carried out, followed by an elution step by increasing the salt gradient from $0\,M$ to $1.0\,M$ NaCl with gradient lengths of $10\,CV$, $15\,CV$, and $20\,CV$. For the nonlinear region, $23.16\,mL$ protein solution was injected via $50\,mL$ superloop before a wash step of $5\,CV$ binding buffer. Subsequently, the elution was carried out by increasing the salt gradient from $0\,M$ to $1.0\,M$ NaCl with a gradient length of $25\,CV$. After elution, the column was stripped over $3\,CV$ at an NaCl concentration of $1.0\,M$ and re-equilibrated for $5\,CV$ binding buffer. All experiments were conducted at $0.33\,mL/min$.

6.3.11. Model Calibration

The chromatograms from LGE in both linear and nonlinear region of the adsorption isotherm were used to carry out the parameter estimation via the inverse method [116].

The first guess was delivered using adaptive simulated annealing (ASA) [113]. For fine adjustment, the Levenberg-Marquardt (LM) algorithm [114] was used. Discretization in space on a grid with equidistant nodes was carried out using the finite element method. The θ-scheme discretization in time was conducted using the fractional step [111]. The solution of the non-linear equation system was approximated via the Picard iteration [112].

6.3.12. *In Silico* Optimization and Model Validation

Pareto optimization of the batch single-column chromatography was performed using ChromX. As objective functions, monomer yield, purity, and production rate were chosen. Optimization parameters are the injection volume and sodium chloride concentration in elution step. The salt concentrations of the binding buffer and strip buffer were kept constant to ensure maximal binding and a complete regeneration of the columns. The OpenBeagleMultiObjective algorithm was employed for the *in silico* optimization. One of the Pareto optima that provided monomer yield and purity above 90 % and 99 %, respectively, and the maximal production rate were chosen to be validated in the laboratory. *In silico* screening was carried out to approximate the Pareto optima of the 3C-PCCC process. Here, 1000 simulations of the 3C-PCCC process were generated by varying the protein injection volume and the optimal salt elution concentration. Here, 1000 simulations of the 3C-PCCC process were generated by varying the protein injection volume and the optimal salt elution concentration. The fractionation was chosen to start and end at $300\,mAU$.

6.4. Results

6.4.1. Investigation of Precipitation Conditions

In this work, PEG precipitation of the mAb was performed as alternative capture step for an mAb purification process. In specific, the goal was to find optimal process conditions for selective precipitation of the mAb. On the one hand, the precipitant concentration and, on the other hand, the pH value of the buffer used was varied. As target parameters, mAb yield and HCP reduction were considered.

The mAb as well as the HCP concentration in the supernatant, after phase separation through centrifugation, are shown in Fig 6.2. All values shown are mean values of at least triple determinations and refer to conditions at 0 % PEG. 98.5 % yield was chosen as target parameter for the precipitation, as this was the maximum reached yield in this study.
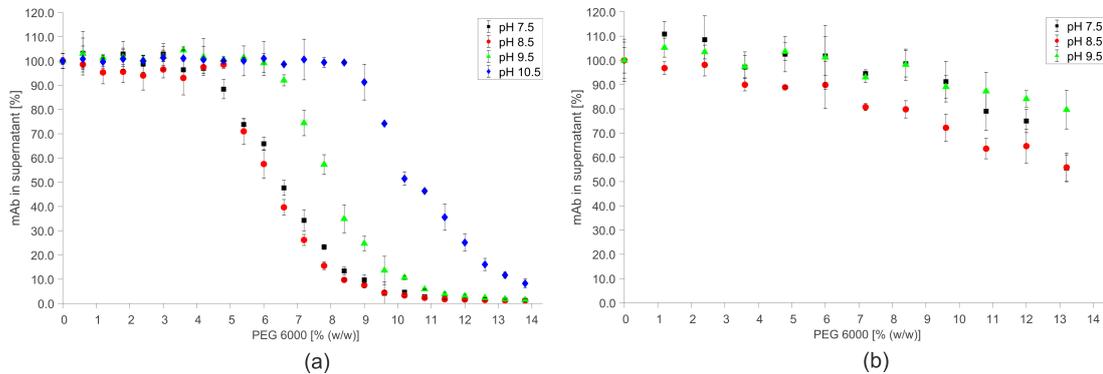
Figure 6.2.: Selected results of the precipitation screening. Data points represent the mean values of at least triplicates. As feedstock HCCF was used. Precipitation was carried out using PEG 6000, varied from $0\,\%\,(w/w)$ to $13.8\,\%\,(w/w)$ PEG, as precipitant. Additionally the pH of the utilized buffer was varied. The relative amount refereed to conditions with $0\,\%\,(w/w)$ PEG of (a) mAb and (b) HCP in the supernatant after centrifugation are displayed.

With increasing PEG concentration, a decrease in the mAb concentration in the supernatant could be observed (Fig. 6.2 (a)). This correlates to an increase in mAb precipitation. For all pH values, the mAb stayed stable in solution for PEG concentrations below $4.2\,\%\,(w/w)$ PEG. With the exception of pH 10.5, all other conditions led to yields above $98.5\,\%$. Therefor, no further investigations were carried out at pH 10.5. The shape of all curves of the different pH values differed in the initial PEG concentration where the precipitation of the mAb began. For conditions with pH 7.5, pH 8.5, and for pH 9.5, PEG concentrations of $12.6\,\%\,(w/w)$ PEG, $13.2\,\%\,(w/w)$ PEG, respectively $13.8\,\%\,(w/w)$ PEG were required to reach the predetermined yield.

All examined systems with pH values except for pH 10.5 were investigated on HCP concentration in the supernatant after phase separation (Fig. 6.2 (b)). For all conditions, a decrease in HCP concentration in the supernatant could be observed with increasing PEG concentration. This correlates to a decrease in the theoretical purity of the precipitated mAb referred to HCP. Precipitation of HCP was not detectable for conditions with PEG concentrations less than $7.2\,\%\,(w/w)$ PEG. For all conditions with pH 8.5, a stronger reduction in the HCP concentration could be observed compared to pH 7.5 and pH 9.5. Furthermore, for PEG concentrations above $11.8\,\%\,(w/w)$ PEG, less HCP reduction could be observed for systems with pH 9.5 compared to pH 7.5. For conditions with the predetermined yield being greater than $98.5\,\%$ and pH 8.5, HCP reduction over $60\,\%$ could be reached. For systems with pH 7.5, $70\,\%$, and for systems with pH 9.5, over $79\,\%$ HCP reduction could be detected.

6.4.2. Investigation of Resolubilisation Conditions

For precipitation, with the exception of pH 10.5, all other investigated pH values achieved satisfying results in the case of mAb yield. One precipitation condition for each investigated pH value with an mAb yield over 98.5 % and with the least amount of PEG was chosen for further process optimization. For these conditions, the resolubilisation was investigated. In detail, for pH 7.5 at least 13.2 % (w/w) PEG, for pH 8.5 12.6 % (w/w) PEG, and for pH 9.5 13.8 % (w/w) PEG were needed to reach the predetermined yield. To optimize the resolubilisation, three factors were taken into account. The pH of the resolubilisation buffer, the filtration time of the phase separation through vacuum filtration, and the incubation time of the resolubilisation. The screenings were carried out as DoE. From the results, a multiple linear regression model (MLR) was built.

For all three precipitation conditions, a similar behavior of the responses, mAb yield and purity, could be observed. No significant influence of the filtration time could be measured and, therefore, this factor was not considered for the model (Fig. 6.3). This results allow the reduction of the parameter set for further investigations. By
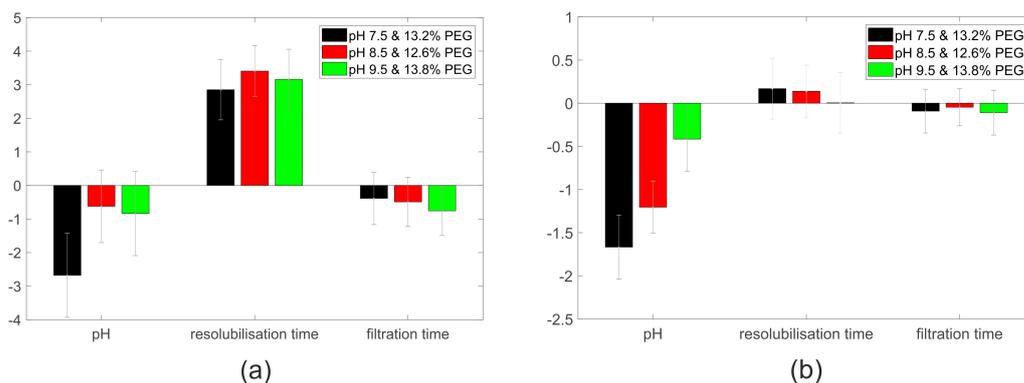
Figure 6.3.: Factors of the MLR model of mAb yield and mAb purity after resolubilisation. HCCF was precipitated at pH 7.5 and 13.2 % (w/w) PEG, pH 8.5 and 12.6 % (w/w) PEG and pH 9.5 and 13.8 % (w/w) PEG. Afterward the precipitated mAb was resolubilisation for incubation times varying from 75 min to 150 min. Additionally the pH of the resolubilisation buffer was varied from pH 4.0 to pH 6.0. Furthermore the filtration time of the phase separation by vacuum filtration was varied from 150 s to 600 s. For further investigation only the resolubilisation time in case of yield and the pH of the resolubilisation buffer in case of the mAb purity were taken into account for the MLR.

reducing the filtration time an overall time saving for the HTS could be reached. No influence of the pH of the resolubilisation could be observed in the case of mAb yield. The only factor which had a significant influence on the mAb yield was the resolubilisation time (Fig. 6.3 (a)). With longer resolubilisation times, higher yields were detected. At 150 min incubation, yields from up to 100 % could be measured (Fig. 6.4 (a)). For the mAb purity, the only factor which had a significant influence

was the pH of the resolubilisation buffer (Fig. 6.3 (b)). With lower pH values, a higher mAb purity could be observed for all precipitation conditions. Purity of mAb up to 97 % could be detected under conditions with pH 4.0 (Fig. 6.4 (b)).
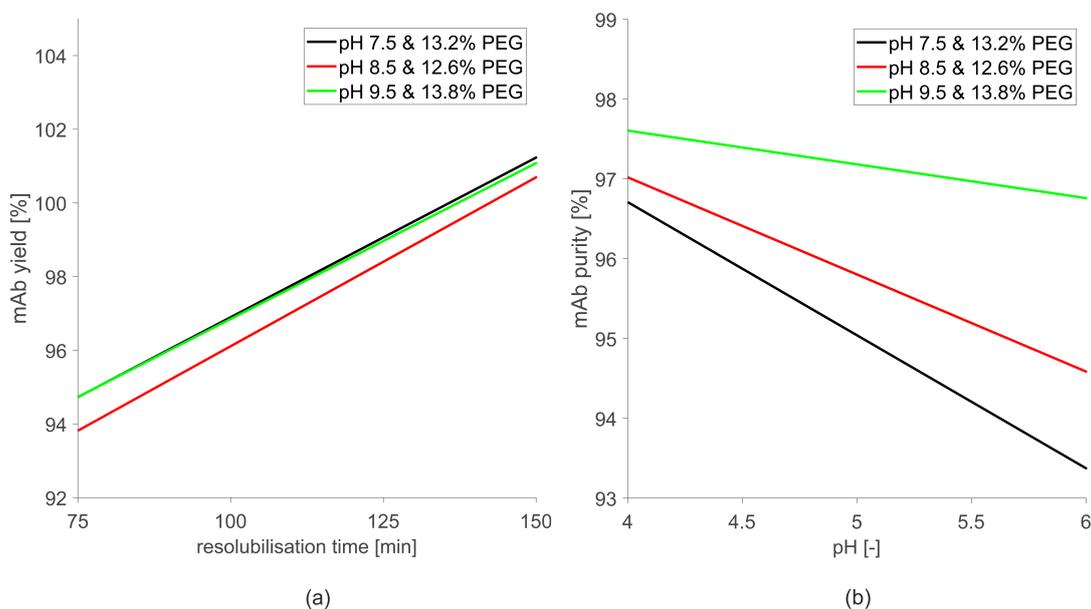
Figure 6.4.: MLR model of mAb yield (a) and mAb purity (b) after resolubilisation. HCCF was precipitated at pH 7.5 and 13.2 % (w/w) PEG, pH 8.5 and 12.6 % (w/w) PEG, and pH 9.5 and 13.8 % (w/w) PEG. Afterwards the precipitated mAb was resolubilised. For the MLR model of the mAb yield the influence of the incubation time was considered. For the MLR model of the purity the pH influence of the resolubilisation buffer was considered.

6.4.3. CEX System Characterization

By conducting tracer experiments without a column attached, the LC system dead volume of $280\,\mu L$ was determined, which was subtracted from all other data from the LC. Based on the results of the tracer injections with a column attached, bed and particle voidage, and axial dispersion were calculated. The acid-basic titration was conducted to determine the ionic capacity. The results are listed in Tab. 6.1.

Table 6.1.: The voidages and axial dispersion are calculated based on tracer experiments. The ionic capacity is determined via acid-base titration.

		GigaCap S-650M
Bed voidage	ϵ_b	0.414
Particle voidage	ϵ_p	0.779
Total voidage	ϵ_t	0.870
Axial dispersion $[mm^2 \cdot s^{-1}]$	D_{ax}	0.067
Ionic capacity $[M]$	Λ	1.383

6.4.4. *In Silico* CEX Process Development

Three linear gradient experiments with gradient lengths of $10\,CV$, $15\,CV$, $20\,CV$, and $25\,CV$ in linear and one with $25\,CV$ in nonlinear region of the adsorption isotherm were used to estimate the chromatography model parameters k_{film}, D_p, k_{eq}, ν, and σ via inverse method by employing ASA and LM algorithm. Non-binding impurities were lumped to one component. The estimated parameter for monomer, aggregate, and lumped impurity are given in Tab. 6.2. k_{kin} was found to be not necessary for the description of the system, since the equilibria of the components considered were reached immediately. Figs. 6.5 (a) - (d) show the simulated and experimental chromatograms used for the model calibration. The lumped impurity broke through during the wash step. The mAb and aggregate show similar behavior and peak overlapping during gradient elution, whereas the latter shows a slightly stronger binding. Figs. 6.5 (e) shows the magnified simulated and measured signals of the same process as in Figs. 6.5 (d).

Knowledge of mass transfer and SMA parameters eventually enabled identification of the optimal step-wise gradient process for the considered separation problem. To find the optimal protein loading volume and elution salt concentration with respect to maximizing monomer yield, purity, and production rate, a multi-objective *in silico* process optimizaton was performed. The resulting Pareto-optimal points can be seen in Fig. 6.6. Constraints for monomer yield and purity were set to be above $90\,\%$ and $99\,\%$, respectively, whilst the maximal production rate was chosen. The optimal protein injection volume and elution NaCl concentration were predicted to be $35.98\,mL$ and 0.161 M. The validation experiment conducted in the laboratory is compared to the model prediction in Figs. 6.8 (a) - (c). It can be seen that the behavior of lumped impurity, mAb monomer, and aggregate are predicted accurately. As magnified details of Fig. 6.8 (a), Figs. 6.8 (b) and (c) show the simulated and measured signal of lumped impurities and the aggregate, respectively. Small deviations, such as the slight underestimation of the aggregate level in the main peak and the tailing of mAb monomer, are shown. The *in silico* predicted and experimentally determined monomer purity, yield, and concentration in the product fraction are compared in

Table 6.2.: The mass transfer model and adsorption isotherm parameters are estimated using the inverse method.

Parameter	mAb	Aggregate	Lumped impurity
$k_{film} \, [mm \cdot s^{-1}]$	0.086	0.049	0.100
$D_p \, [mm^2 \cdot s^{-1}]$	$3.508 \cdot 10^{-5}$	$2.788 \cdot 10^{-5}$	$1.648 \cdot 10^{-5}$
$k_{eq} \, [\text{-}]$	$2.724 \cdot 10^{-7}$	$6.832 \cdot 10^{-9}$	$3.995 \cdot 10^{-12}$
$\nu \, [\text{-}]$	8.899	11.340	-
$\sigma \, [\text{-}]$	101.380	98.262	-

Tab. 6.3. The monomer purity has a deviation of 0.3 %, yields a prediction error of 8.4 %, and the monomer concentration in pooled fraction is underestimated by 8.5 %.

To investigate the potential of a 3C-PCCC process in the considered case, an *in silico* screening was carried out to approximate the Pareto optimal points with respect to maximizing monomer purity and yield simultaneously. Overall, the optimal protein injection volumes were found to be greater and the optimal salt elution concentrations were lower than the ones in the batch single-column process. Improvements of both monomer purity and yield were found as shown in Fig. 6.7.
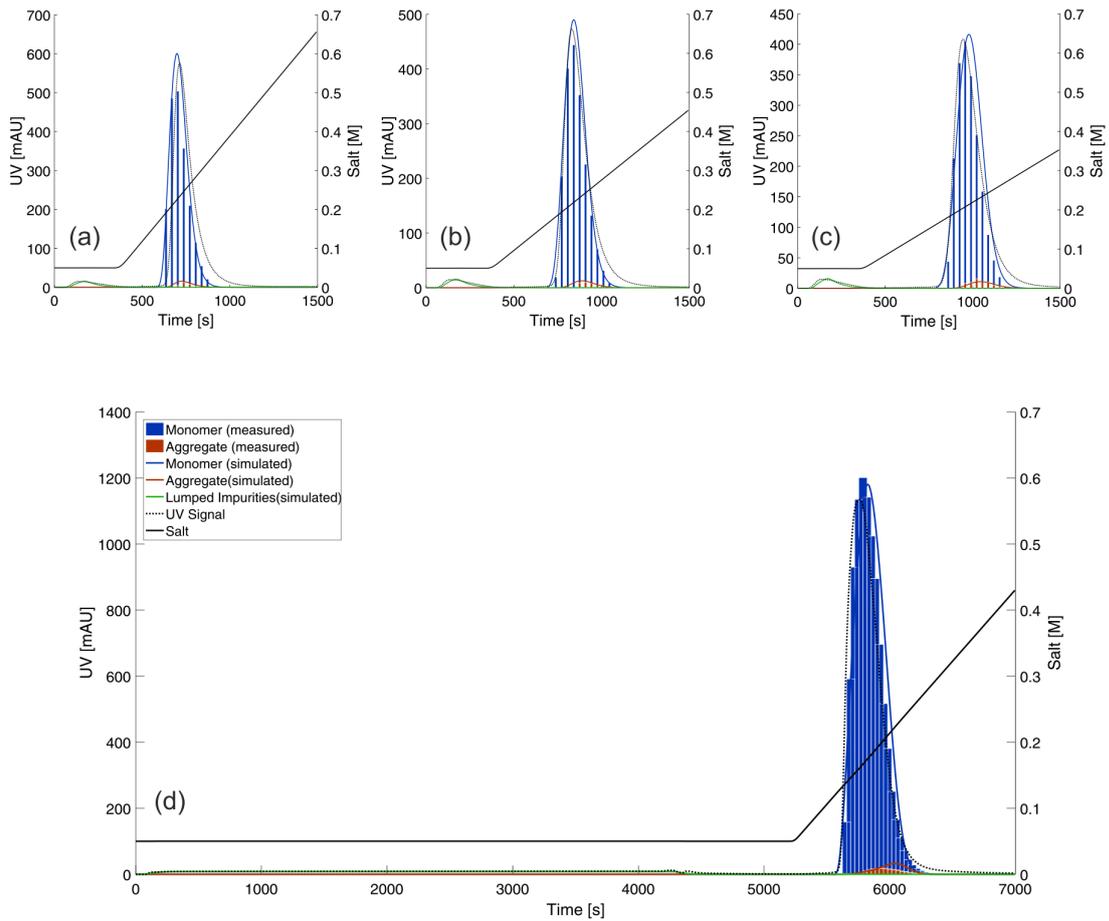
Figure 6.5.: Plots of UV signals over process run time for the bind-and-elute chromatography experiments. Dashed lines display the UV signal measured at column outlet. Black solid lines represent the adjusted salt gradients with elution gradient lengths of $10\,CV$ (a), $15\,CV$ (b), $20\,CV$ (c), and $25\,CV$ (d). Blue solid lines and bars represent the simulation of monomer and offline analytical result, respectively. Red and green are the aggregate and lumped impurity. (d) displays the breakthrough of lumped impurity and the elution of aggregate.
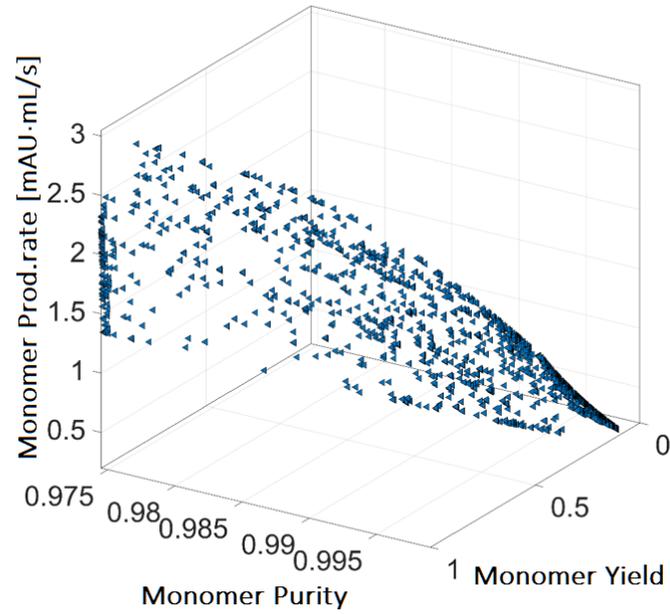
Figure 6.6.: Pareto-optimal operating points for the batch single column chromatography process represent the maximal monomer yield, purity, and production rate by varying protein loading volume and elution salt concentration.
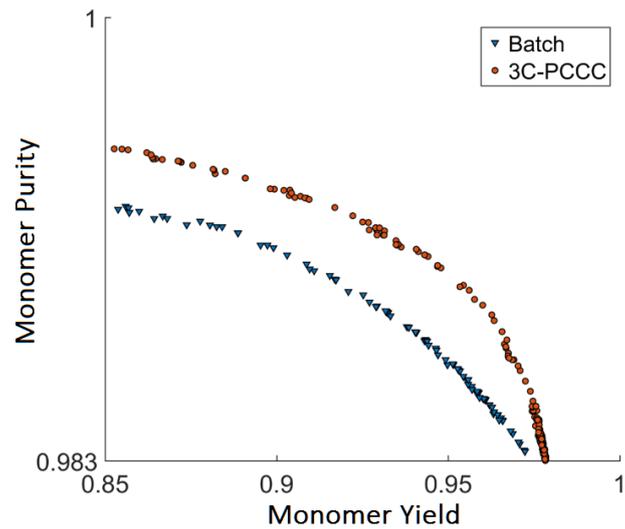
Figure 6.7.: Comparison of the Pareto-optimal operating proteins for the batch (blue triangles) and 3C-PCCC process (orange circles).

6.4.5. Batch Process Validation

After optimization of each process step, a validation run with the optimal process parameters was performed. The precipitation was performed at pH 8.50 with a PEG concentration of 12.6 % (w/w) PEG. Afterwards, the samples were resolubilised at pH 4.00 for 150 min with a filtration time of 150 s. Before the CEX step, the pH of the sample was adjusted to pH 5.00. The subsequent CEX step was conducted at a loading volume of 35.98 mL, and elution NaCl concentration of 0.161 M, as predicted by the mechanistic chromatography model. At the beginning and after each step, the mAb, HCP and DNA concentration as well as the aggregate content were measured. In Tab. 6.4, these results are compared to a common purification run. The values for Protein A are provided from Lek d.d. and were measured directly after Protein A elution. During the integrated process, HCP and DNA reduction were mainly reached during the precpitation and resolubilisation step. Aggregate reduction and concentration of the product were reached during CEX. Compared to Protein A, mAb yield, HCP reduction, and aggregate content reached similar values. The mAb concentration of the integrated process was 63% lower in comparison to Protein A chromatography, the DNA reduction was four magnitudes higher.

Table 6.3.: Comparison between the *in silico* predicted and experimentally determined objectives.

Objective	Model	Experiment	Deviation
Monomer purity [%]	99.00	98.70	0.30
Monomer yield [%]	91.10	98.75	8.40
c(monomer) pooled [mg/mL]	10.94	11.87	8.50

Table 6.4.: Case study validation run after optimisation of the integrated process. Monomer concentration, yield, and purity were detected after each step and compared to a common Protein A process.

Objectives	HCCF	Precipitation	CEX	Protein A
c(monomer) [mg/ml]	3.30	1.69	11.87	32.20
Monomer yield [%]	-	96.47	95.41	> 95.00
HCP reduction [%]	-	98.23	99.28	99.99
Aggregate content [%]	3.12	3.96	1.30	1.20
DNA [ng/ml]	8.09e+07	2.3 e+03	2.48e+02	1.90e+06
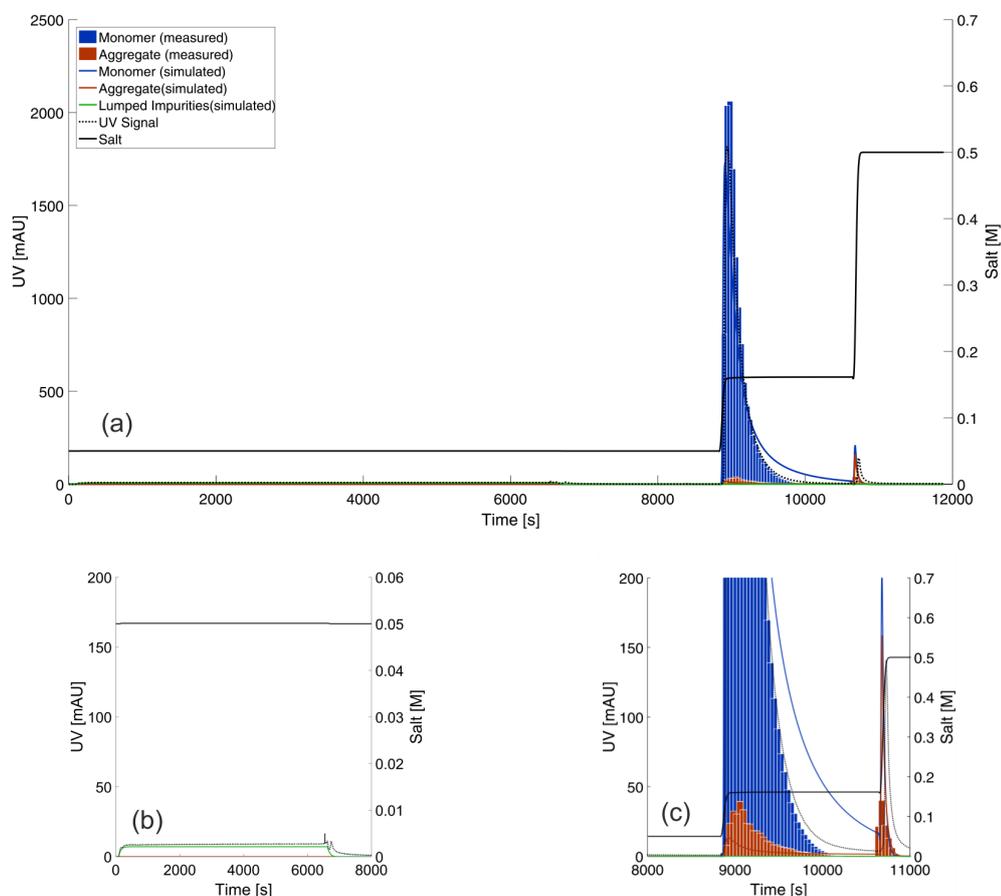
Figure 6.8.: Plot of UV signal over process run time for the validation experiment. Dashed lines display the UV signal measured at column outlet. Black solid lines represent the adjusted step-wise salt gradient. Blue solid lines and bars represent the simulation of monomer and offline analytical result, respectively. Red and green are the aggregate and lumped impurity. (b) and (c) display the magnified details of the breakthrough of lumped impurity and the elution of aggregate in the same process as shown in (a), respectively.

6.5. Discussion

A process strategy for mAb processes was successfully developed. The integration of PEG precipitation and CEX showed the potential of replacing Protein A as capture step in an mAb process.

Through the replacement of Protein A, the necessity of establishing an alternative development platform was given. Furthermore, in comparison to protein A chromatography, PEG precipitation alone is not capable of reducing high molecular weight impurities (HMW) like aggregates or DNA. The integrated process and the development platform presented in this work can solve both problems. The PEG

precipitation showed promising results based on mAb yield and HCP reduction. Resolubilisation at low pH values led to high mAb yields and a high amount of HCP reduction. The results were comparable to common Protein A steps [236]. Furthermore, it allowed a direct loading onto CEX column. Resolubilisation in combination with concentration is reported as being time-consuming and might end up with low yields though incomplete resolubilisation [220]. Using CEX in bind-and-elute mode combined a reduction of HMW with a volume reduction at the same time.

For this integrated process, a process development platform was designed. It combines high-throughput methods with empirical and mechanistic modeling. For the precipitation, a high-throughput method on a liquid handling station was successfully generated. The PEG concentration and the pH of the surrounding buffer showed significant influence on the precipitation behavior. Although PEG is a cheap raw material, recycling is currently not an issue. Therefore, the amount of PEG is one of the driving costs of the process and consequently at least possible PEG should be used. Studies by Hammerschmidt *et al.* have shown that precipitation results, like yield or purity, are independent of the operation mode [237]. This indicates that development strategy presented here could be used for batch and continuous process development. After the precipitation, the resolubilisation was optimized using a DoE and a MLR model, avoiding brute-force methodology. Afterwards, critical process parameters could be detected and optimized in a goal-driven manner. The mechanistic chromatography modeling reduced the number of experiments needed to be carried out in the laboratory. Thus, the material and time required for the development of precipitation, resolubilisation, and chromatography processes could be reduced. Furthermore, based on the mechanistic understanding gained, the potential of a 3C-PCCC process could be explored.

The development strategy was used to purify an mAb, provided as HCCF. The highest observed mAb yield after precipitation was 98.5%. Comparing different pH values, under conditions with pH 8.5, the least amount of PEG was needed to attend this objective. Conditions with pH 7.5 and pH 9.5 reached this yield at slightly higher PEG concentrations as well. Altogether, the lowest colloidal phase stability of the mAb was observed near its isoelectric point (pI). This is equal to least electrostatic repulsion. However, for these conditions, no more than 79.0% HCP reduction were detected. Co-precipitation of HCP following the mAb could be a reason for this partial reduction. Higher purities were detected under conditions with minor colloidal phase stability of the mAb.

As the precipitation alone was not sufficient enough, the resolubilisation was optimized as second purification step. The DoE showed the pH value to be a critical process parameter for the reduction of HCP. The resolubilisation was carried out at pH values near the pI of most of the HCP. This is reported to be around pH 5.0 [238]. A selective resolubilisation was observed, and HCP reductions of over 98.0% were detected. This suggests that most of the HCP were not resolubilised and remained on

the filter. Simultaneously, the investigated pH values were at least 2.5 pH steps below the pI of the mAb. This was probably the reason for a complete resolubilisation of the mAb. Shifting the pH to even lower values could be a possibility to combine virus inactivation with the resolublisation step. The resolubilisation time was detected as the second significant process parameter. Higher yields were detected at longer resolubilisation times. This implies a slow resolubilisation kinetics. The formation of a filter cake is reported as reason for this phenomenon [237]. No reduction of molecules with a hydrodynamic radius bigger than that of the mAb was expected. The aggregate level stayed constant. However, a DNA reduction of over 5 magnitudes could be seen. This might be due to the incubation conditions, as DNA is reported to be sensitive towards low pH values [239].

Because the mAb showed instabilities at high salt concentrations (data not shown), the pH was adjusted from pH 4.0 to pH 5.0 before the CEX step. For the CEX step, the predicted Pareto-optimal process could be validated successfully. Here, the chromatogram predicted showed good agreement with the validation experiment carried out in the laboratory. The elution of aggregate in the first step was slightly underestimated, leading to a monomer purity overestimated by 0.3 %. The tailing of monomer was predicted to be more pronounced by the model, resulting in a yield deviation of 8.4 %. The reason of these deviations might be the simplification of the protein system by neglecting different binding orientations and charge variants. In the case of the 3C-PCCC process, an *in silico* screening predicted higher monomer purity and yield, whereas the optimal protein injection volumes were found to be greater and the optimal salt elution concentrations were lower than the ones in the batch single-column process. This observation can be explained by the nature of the 3C-PCCC mode i.e., by the fact, that a column can be loaded with more protein. For for the monomers, a more nonlinear region of the adsorption isotherm could be reached. This resulted in a lower elution salt concentration required. The aggregates, however, bind in the linear region of the adsorption isotherm due to their low concentration and tend to retain in the column during the elution step. Hence, 3C-PCCC shows a great potential to be combined with a continuous precipitation capture process.

6.6. Conclusion

This work has shown a fast and material saving-process development method for an mAb process. By integrating a PEG precipitation and CEX chromatography, a feasible alternative to Protein A chromatography as capture step was demonstrated. For process development, a strategy combining high-throughput methods with empirical and mechanistic modeling has shown promising results which might overcome the drawback of replacing an affinity step.

Precipitating under conditions with pH 8.5 and 12.6 % (w/w) PEG and resolubilisation at pH 4.0 led to a significant reduction in HCP level in combination with a yield of over 96.4%. Additionally, loading onto a CEX column with only minor pH adjustments was enabled. Using a step gradient, an aggregate reduction in combination with product concentration to $11.9\,mg/mL$ was demonstrated. All in all, a flexible time-and material-saving method was demonstrated which allows screening of even more parameters than shown in this case study. Furthermore, the capability of a continuous chromatography process was explored theoretically based on the calibrated mechanistic chromatography model.

In further studies, the upscale of these results should be demonstrated. The execution of a fully continuous process including continuous PEG precipitation in combination with CEX-PCCC at a larger scale would be desirable.

6.7. Acknowledgment

7. Root Cause Investigation of Deviations in Protein Chromatography Based on Mechanistic Models and Artificial Neural Networks

Gang Wang[1], Till Briskot[1], Tobias Hahn[2], Pascal Baumann[1], Jürgen Hubbuch[1,*]

[1] Karlsruhe Institute of Technology (KIT), Institute of Process Engineering in Life Sciences, Section IV: Biomolecular Separation Engineering, Karlsruhe, Germany
[2] GoSilico GmbH, Karlsruhe, Germany
* Corresponding author. E-mail address: juergen.hubbuch@kit.edu

Abstract

In protein chromatography, process variations, such as aging of column or process errors, can result in deviations of the product and impurity levels. Consequently, the process performance described by purity, yield, or production rate may decrease. Based on visual inspection of the UV signal, it is hard to identify the source of the error and almost unfeasible to determine the quantity of deviation. The problem becomes even more pronounced, if multiple root causes of the deviation are interconnected and lead to an observable deviation. In the presented work, a novel method based on the combination of mechanistic chromatography models and the artificial neural networks is suggested to solve this problem. In a case study using a model protein mixture, the determination of deviations in column capacity and elution gradient length was shown. Maximal errors of 1.5% and 4.90% for the prediction of deviation in column capacity and elution gradient length respectively demonstrated the capability of this method for root cause investigation.

7.1. Introduction

Mechanistic models have found extensive applications in protein chromatography as reported [24, 240–242]. It requires little experimental effort and provides mechanistic process understanding, which satisfies the demands of the Quality by Design (QbD) approach promoted by the US Food and Drug Administration (FDA) [17–19].

Mechanistic models for chromatography consist of partial differential equations describing the convection, diffusion, and adsorption mechanisms in the chromatographic column. For ion-exchange chromatography, a proven approach is given by the combination of the widely accepted general rate model (GRM) covering the mass transfer [41], and the stoichiometric displacement isotherm model (SDM) [56, 164] or its extension, the steric-mass action (SMA) isotherm model [90]. Although these mechanistic models contain a huge amount of information, they has been almost exclusively applied to model-based process development [240, 241, 243–245] and robustness analysis [246–250].

Artificial intelligence and machine learning based on artificial neural networks (ANNs) have been among the fastest growing research topics in the last decades. The multilateral applicability of ANN has been demonstrated in numerous fields for solving many complex real-world problems [162, 251]. In chromatography, ANNs were thought to be an alternative to first-principle modeling. The back-propagation ANN (BP-ANN) of simple architecture was employed to model retention times in different chromatographic modes and formats [26–30]. In this context, however, it lacks mechanistic understanding and requires a large amount of data for the learning process that is often not available.

Combining both mechanistic and non-mechanistic modeling results in the so-called hybrid models or metamodels which enable entirely new opportunities [242, 252–255]. ANNs can make accessible hidden unused information contained in mechanistic models; whereas the mechanistic model supports the learning process of ANNs with overwhelming data amounts generated by *in silico* experimentation. To demonstrate the power of this combined method, root cause investigation is considered. Here, an exact mechanistic model which can immediately identify the root causes of the deviation does not exist. Also the conventional ANN approach would fail, because it is unfeasible to artificially generate process variations of any magnitude in the laboratory to gather enough information for its application.

In this work, the combination of mechanistic models and ANNs is carried out to identify causes for deviation in chromatograms. The presented case study pays special attention to the ionic capacity for two reasons. First, the ionic capacity of a column can be influenced by day-to-day operations, column aging, or even a column exchange - and its deviation can result in deviation of the elution behavior of product and impurities, leading to reduced purity, yield, or production rate. Secondly, a

real-time analytical tool for determination of the ionic capacity is not available, and the current standard method is the time-consuming acid-base titration. The additional parameter salt gradient length was chosen to complicate the case study, since its deviation has a very similar effect on the chromatogram. Of course, the actual values of the salt gradient length can be measured easily using a conductivity sensor. However, the measurement itself does not allow a quantitative explanation of deviating chromatograms.

After the mechanistic model has been calibrated, it is employed for *in silico* experimentation to generate information about the interrelationships between the chromatograms and the two parameters ionic capacity and salt gradient length. Based on this information, the ANN learns to recognize chromatograms and to respond with the corresponding parameter values. To verify the ANN's generalization capability, 20 % of the *in silico* data are used for cross-validation. To demonstrate the practical applicability of this method, experimental errors are imitated by manipulating the parameters in the laboratory.

7.2. Theory

7.2.1. Transport-dispersive Model

As a basis for generating large data sets for ANN calibration, mechanistic models were employed using the following theoretical background. To model the convection and diffusion inside the chromatography column of length L and adsorber beads of radius r_p, the general rate model (GRM) is used [41]:

$$
\begin{aligned}
\frac{\partial c_i(x,t)}{\partial t} &= -u(t)\frac{\partial c(x,t)}{\partial x} + D_{ax}\frac{\partial^2 c(x,t)}{\partial x^2} \\
&\quad - \frac{1-\epsilon_b}{\epsilon_b}k_{film,i}\frac{3}{r_p}\left(c_i(x,t) - c_{p,i}(x,r_p,t)\right)
\end{aligned}
\tag{7.1}
$$

$$
\frac{\partial c_i(0,t)}{\partial x} = -\frac{u(t)}{D_{ax}}(c_i(0,t) - c_{in,i}(t))
\tag{7.2}
$$

$$
\frac{\partial c_i(L,t)}{\partial x} = 0
\tag{7.3}
$$

$$
\frac{\partial c_{p,i}(x,r,t)}{\partial t} = \begin{cases}
\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 D_{p,i}\frac{\partial c_{p,i}(x,r,t)}{\partial r}\right) - \frac{1-\epsilon_p}{\epsilon_p}\frac{\partial q_i(x,r,t)}{\partial t} & \text{for } r \in (0,r_p), \\
\frac{k_{film,i}}{\epsilon_p D_{p,i}}(c_i(x,t) - c_{p,i}(x,r_p,t)) & \text{for } r = r_p, \\
0 & \text{for } r = 0.
\end{cases}
\tag{7.4}
$$

The mass transfer between the pore volume $c_{p,i}(x,r,t)$ and the interstitial volume of the mobile phase $c_i(x,t)$ is described by Eq. (7.1). The peak-broadening effects in the interstitial volume are lumped in the voidage of the bed ϵ_b, the axial dispersive coefficient D_{ax}, and the film transfer coefficient $k_{film,i}$. Eqs. (7.2) and (7.3) are the Danckwerts boundary conditions. Eq. (7.4) describes the exchange between the stationary phase q_i and the pore volume concentration $c_{p,i}$ depending on the adsorber particle voidage ϵ_p and the component-specific pore diffusion coefficient $D_{p,i}$. Here, the less complex linear driving force (LDF) model could also be chosen in spite of large molecules, if the number of transfer units is high and the adsorption isotherm is linear.

7.2.2. Stoichiometric Displacement Isotherm Model

In the linear region of the adsorption isotherm, the steric hindrance effect of proteins can be neglected, resulting in equivalence of the steric mass action (SMA) [90] and stoichiometric displacement isotherm model (SDM). Since the presented work only investigates the linear region, for description of the interaction between protein and ligand on the IEX adsorber surface, SDM in the kinetic formulation is employed [56, 164]:

$$k_{kin,i}\frac{\partial q_i(x,r,t)}{\partial t} \;=\; k_{eq,i}q_{salt}(x,r,t)^{\nu_i}c_{p,i}(x,r,t) - c_{salt}(x,r,t)^{\nu_i}q_i(x,r,t) \qquad (7.5)$$

The adsorption coefficient $k_{ads,i}$ and the desorption coefficient $k_{des,i}$ are replaced by the equilibrium coefficient $k_{eq,i} = k_{ads,i}/k_{des,i}$ and the kinetic coefficient $k_{kin,i} = 1/k_{des,i}$ as proposed in [116]. The advantage of this formulation is that $k_{eq,i}$ strongly affects the retention time, whereas $k_{kin,i}$ has a slight influence on the peak height when the equilibrium is not reached instantly. ν_i is the characteristic charge of the component i. c_{salt} is the counter-ion concentration in the pore phase. The salt concentration in the stationary phase q_{salt} is given as:

$$q_{salt}(x,r,t) \;=\; \Lambda - \sum_{i=1}^{m}\nu_i q_i(x,r,t) \qquad (7.6)$$

with the ionic capacity of the stationary phase being Λ.

Combined with the GRM Eqs. (7.1)-(7.4), an IEX chromatography process can be modeled using numerical methods.

7.2.3. Numerical Solution

The finite element method is employed for the discretization in space on a grid with equidistant nodes. The fractional step θ-scheme is chosen for the discretization in

time [111]. The solution of the non-linear equation system is approximated by Picard iteration [112]. Adaptive simulated annealing [113] and the Levenberg-Marquardt algorithm [114] are employed for parameter estimation. The former is a heuristic algorithm that searches a defined space with random jumps to find the global minimum. Thereafter, the results are refined by applying the latter algorithm which is deterministic.

7.2.4. Artificial Neural Networks

In comparison to first-principles modeling, ANN modeling offers numerous advantages such as the ability to detect possible correlations and unconsidered nonlinear relationships between variables [161]. Its topological structure is shown in Fig. 7.1.Each
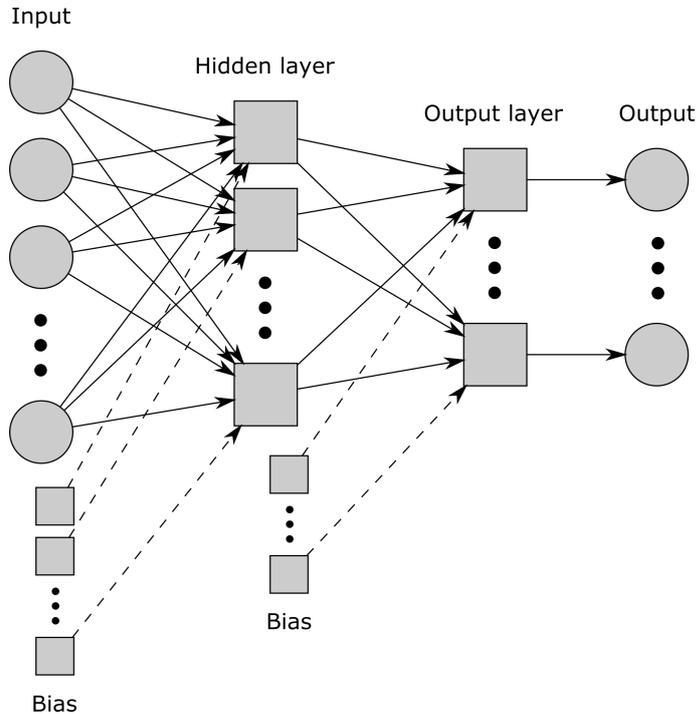
Figure 7.1.: Topological structure of a simple multi-layer BP-ANN.

node is connected to every node in the next layer. The nodes in the input layer represent the input data and the ones in the output layer return the output data. In the nodes of the hidden layer, the sum of products of biases and weights n, and the products of inputs and weights is transferred with a hyperbolic tangent sigmoid activation function to an output signal between -1 and $+1$ as shown in Eq. (7.7).

$$f(\tilde{u}) \quad = \quad \frac{2}{1 + e^{-2\tilde{u}}} \tag{7.7}$$

Here, the nonlinearity in the data can be covered. A linear activation function is employed in the nodes of the output layer to transfer the sum of weighted outputs from the former layer and the weighted biases to the final outputs. By mapping the input data to the correct output data using backpropagation, the difference between the targeted and the actual output values is iteratively minimized by updating the weights in the network [163].

7.3. Materials and Methods

7.3.1. Instruments

All chromatographic experiments were carried out using an Ettan liquid chromatography (LC) system equipped with pump unit P-905, dynamic single chamber mixer M-925 (90 μl mixer volume), UV monitor UV-900 (3 mm optical path length), and conductivity cell pH/C-900 (all GE Healthcare, Little Chalfont, Buckinghamshire, UK).

7.3.2. Software

The control software UNICORN 5.31 (GE Healthcare, Little Chalfont, Buckinghamshire, UK) was used in combination with the LC system. The mechanistic model was simulated using the protein chromatography simulation software ChromX (GoSilico, Karlsruhe, Germany). ChromX was used for the numerical simulation of the system of partial differential equations, estimation of model parameters, as well as optimization and statistical analysis [115]. The ANN modeling was conducted in Matlab® R2016a (MathWorks, USA).

7.3.3. Adsorber, Proteins, and Chemicals

The strong cation-exchange chromatography (CEX) adsorber medium, TOYOPEARL® GigaCap S-650M supplied by Tosoh Bioscience (Griesheim, Germany) was used as pre-packed 0.965 mL Toyoscreen® column with dimension 30 mm × 6.4 mm. The adsorber media were stored in 20 % ethanol between the runs. The storage solution was removed by prolonged equilibration with water, and flushed with low-salt and high-salt buffer before experimentation.

Three proteins of different mass transfer and adsorption behavior according to their different sizes and different isoelectric points were used. A monoclonal antibody (mAb, pI 8.3-8.5, 144-147 kDa) of the IgG isotype was provided by Lek (Ljubljana, Slovenia). Cytochrome c (bovine heart, no. A4612, pI 10.4-10.8, 12.3 kDa) was

purchased from Sigma-Aldrich (St. Louis, MO, USA). Lysozyme (hen egg, no. HR7-110, pI 11.0, 14.6 kDa) was purchased from Hampton Research (Aliso Viejo, CA, USA).

The experiments were conducted below the isoelectric points of all three proteins. 50 mM sodium phosphate buffer was used at pH 6.5 with additional 0 and 1 M NaCl. 0.5 M NaOH was used for cleaning-in-place. All solutions were prepared with ultra pure water (arium pro UV, Sartorius, Göttingen, Germany), filtrated with a membrane cutoff of 0.22 μm and degassed by sonication. Protein solution with 1 mg/ml of IgG, cytochrome c, and lysozyme each was prepared in corresponding binding buffer and filtrated with a membrane cutoff of 0.22 μm prior to usage.

7.3.4. System Characterization

The ÄKTA system and chromatography column were characterized via tracer pulse injection at a constant flow rate of 0.5 ml/min just like all other chromatography experiments in this work. For determination of the interstitial volume of the column, filtrated 10 g/L blue dextran 2000 kDa (Sigma-Aldrich, St. Louis, MO, USA) solution was used as non-interacting, non-pore-penetrating tracer. The system and total voidage of the column were determined using 25 μL 1 %(v/v) acetone (Merck, Darmstadt, Germany) as a pore-penetrating, non-interacting tracer. For determination of these parameters, the UV signal at 260 nm was recorded. All measurements were corrected with respect to dead volumes. Acid-base titration following Huuk et al. was applied to determine the ionic capacity Λ of GigaCap S-650M [165] in triplicate. The column was first equilibrated for 15 CV with 0.5 M HCl to exchange the counter-ion against protons. Subsequently, the columns were washed with ultra-pure water for 20 CV to displace the HCl solution. Finally, the adsorber was titrated with a 0.01 M NaOH solution. The total ionic capacity per adsorber skeleton volume was calculated according to

$$n_{Na^+} = V_{NaOH} \cdot c_{NaOH} \tag{7.8}$$

and

$$\Lambda = \frac{n_{Na^+}}{V_c \cdot (1 - \epsilon_t)}, \tag{7.9}$$

with the amount of exchanged sodium ions n_{Na^+}, column volume V_c, total porosity ϵ_t, and concentration of NaOH solution used c_{NaOH}.

7.3.5. Bind-and-elute Experiments

Protein solution with 1 mg/ml of IgG, cytochrome c, and lysozyme each was prepared in corresponding binding buffer and injected via a 100 μL loop. From low-salt and

high-salt buffers, different steps and gradients were mixed within the LC system. A subset of the chromatograms was used to perform the parameter estimation with the inverse method [116]. The remaining runs were used for model validation and the ANN-based root cause investigation.

The linear gradient experiments were conducted by applying the varying gradient lengths 15 CV, 20 CV, and 25 CV. After a post-loading wash step of 1 CV equilibration buffer, elution was started with 50 mM and an increasing salt gradient ranging from 50 mM to 400 mM NaCl. After protein elution, the column was stripped over 3 CV at a NaCl concentration of 1.05 M and re-equilibrated for 1 CV binding buffer.

The step elution consisted of three elution steps. After an equilibration step with 2 CV equilibration buffer and a wash step of 2 CV binding buffer, the salt concentration was raised to 100 mM NaCl and held for 8 CV. Then, the concentration was further increased to 190 mM NaCl and kept constant for 8 CV. Finally, the column was stripped for 3 CV at a concentration of 1.05 M NaCl and re-equilibrated with equilibration buffer.

All experiments were carried out at a flow rate of 0.5 ml/min to ensure a constant residence time.

7.3.6. Mechanistic Model Calibration and Validation

Two bind-and-elute chromatograms using linear gradient elution of 20 CV and 25 CV were used for model calibration. Widely accepted correlations suggested in the literature were employed to determine the upper and lower boundaries for subsequent parameter estimation. The penetration theory [256] was chosen to calculate the upper boundary and the correlation suggested by Kataoka and coworker [257] was employed to calculate the lower boundary of the film diffusion coefficient. The correlation based on Guiochon [40], Tyn and Gusek [258], and Mackie and Meares [259] was used to estimate the lower boundary, and the correlation based on Guiochon [40], Tyn and Gusek [258], and Wackao and Smith [260] to estimate the upper boundary of the pore diffusion coefficient. In parameter estimation, the genetic algorithm was employed at first. For fine tuning of the parameter estimates, the Levenberg-Marquardt algorithm [114] was employed. It is important to note, that a small discrepancy between simulated and experimental chromatograms does not necessarily indicate an accurate parameter estimation. Thus, the application of parameter estimation need to be subjected to rigorous validation. Here, an uninvolved step-wise gradient bind-and-elute chromatogram was predicted and compared to wet-lab experiment for validation. The confidence intervals at 95 % level were calculated based on parameter sensitivities subsequently to assess the reliability of parameter estimates.

7.3.7. *In Silico* Screening

570 model parameter combinations of salt gradient length and ionic capacity were used to perform *in silico* experimentation using ChromX. The ranges from $1.0\,\text{M}$ to $1.3\,\text{M}$ for the ionic capacity and from $12\,\text{CV}$ to $28\,\text{CV}$ for the salt gradient length were chosen. The salt gradient length and ionic capacity both affect the elution behavior in very similar ways. Visually inspected, a deviation of the ionic capacity or salt gradient length influences the retention time and the peak height. A pronounced problem is given, when multiple root causes are potentially responsible for one observed deviation. The ranges were chosen to cover different scenarios of deviated elution behavior on CEX in a process-relevant parameter window. All parameters and the resulting 570 chromatograms were normalized to the minimum and maximum values observed. Since the *in silico* experimentation generates ideal data, white Gaussian noise was added by a signal-to-noise ratio of 60. After that, all normalized UV signals smaller than 0.2 were cut off with the aim to reduce the data amount. This resulted in cheaper computation and was proven to be time-saving for potential real-time application of the ANN.

7.3.8. ANN Model Calibration and Validation

A simply constructed BP-ANN model with a hidden layer of nine neurons and an output layer of two neurons was built in MATLAB®. The hyperbolic tangent sigmoid activation function *tansig* and the linear activation function *purelin* were implemented in the nodes of the hidden layer and in the ones of the output layer, respectively. The function *divideint* using interleaved indices was employed to divide the *in silico* data into training, validation, and test data sets by a ratio of 5:3:2. The network training function *trainlm* was applied to update weight values for inputs and biases according to an adaptive learning rate and a gradient descent momentum. The learning rate and the momentum were set to 0.05 and 0.9. A root-mean-square error (RMSE) of $1 \cdot 10^{-5}$ and a maximal iteration number of 1000 were set as stopping criteria.

7.3.9. ANN-based Root Cause Investigation

Besides the verification of the ANN's generalization capability with cross-validation using $20\,\%$ of the total *in silico* data sets, a case study with three runs with differently pronounced deviations in their chromatograms was designed to show the practical applicability of the presented root cause investigation method. The calibration experiment with a linear gradient elution of $25\,\text{CV}$ and ionic capacity of $1.172\,\text{M}$ was chosen to be the set point. Three further bind-and-elute experiments using parameter combinations of $15\,\text{CV}$ and $1.172\,\text{M}$, $20\,\text{CV}$ and $1.130\,\text{M}$, as well as $25\,\text{CV}$

and $1.130\,\mathrm{M}$ were carried out in the laboratory. The first experiment represented the deviation of the salt gradient length only, the second the simultaneous deviation of the salt gradient length and ionic capacity, and the third the deviation of the ionic capacity only. Furthermore, they represented the different degrees of deviations in the chromatogram. The chromatograms were fitted with exponentially modified Gaussian distribution [166] in the aim of signal smoothing and peak separation. After normalization, all UV signals smaller than 0.2 were cut off according to the data preparation for ANN model calibration and validation. The separated peaks were entered into the calibrated BP-ANN model which returned the predictions of salt gradient length and ionic capacity for each experiment. Subsequently, the predictions were compared to the settings and the percentage errors were calculated to prove the predictive accuracy.

7.4. Results and Discussion

7.4.1. System Characterization

Without a column attached, the dead volume of $280\,\mu l$ of the LC system was determined by tracer injection. This dead volume was subtracted from all other data generated with the LC. Based on the results of the tracer injections, axial dispersion, bed and particle voidage were calculated. The ionic capacities were calculated based on the acid-base titration in triplicate. The ionic capacity was reduced irreversibly from $1.172\,\mathrm{M} \pm 0.003\,\mathrm{M}$ to $1.130\,\mathrm{M} \pm 0.004\,\mathrm{M}$ by flushing the column with $2.0\,\mathrm{M}$ HCl at a flow rate of $0.2\,\mathrm{ml/min}$ for $24\,\mathrm{h}$. The results are listed in Table 7.1.

Table 7.1.: The voidages and axial dispersion are calculated from the retention volume and peak broadening of tracer injections. The ionic capacity 1 is the initial state of the column, whereas the ionic capacity 2 is determined after acid treatment of the same column.

		GigaCap S-650M
Bed voidage	ϵ_b	0.418
Particle voidage	ϵ_p	0.758
Total voidage	ϵ_t	0.829
Axial dispersion $[mm^2 \cdot s^{-1}]$	D_{ax}	0.120
Ionic Capacity 1 [M]	Λ	1.172 ± 0.003
Ionic Capacity 2 [M]	Λ	1.130 ± 0.004

7.4.2. Mechanistic Model Calibration and Validation

To estimate the model parameters k_{film}, D_{pore}, k_{eq}, and ν with the inverse method, two linear gradient chromatography experiments were carried out with gradient lengths of 20 CV and 25 CV. First, the genetic algorithm was employed. Thereafter, a Levenberg-Marquardt algorithm was used to find the final optimum. The estimated parameters are given in Table 7.2. The confidence intervals at 95 % level confirm the reliability of the parameter estimates. Obviously, k_{kin} was not necessary for the description of the system considered, because the equilibria of all components were reached instantly. Figs. 7.2 (a) and (b) show the experimental and simulated chromatograms used for the model calibration. The weakly binding component mAb, moderately binding component cytochrome c, and the strongly binding component lysozyme are eluted in the salt gradient in successive order. To prove the prediction performance of the calibrated mechanistic model, an additional experiment beyond the calibration space has been conducted as validation data. Fig. 7.2 (c) shows very good agreement of experimental data and model prediction. Since the reliability of the parameter estimates and the accuracy of the calibration are given, this mechanistic model is assessed to be suitable for *in silico* data generation.

7.4.3. ANN Model Calibration and Validation

570 *in silico* experiments were conducted to screen the design space as shown in Fig. 7.3 (b). The resulting chromatograms can be seen in Fig. 7.3 (a). The ranges from 1 M to 1.3 M and from 12 CV to 28 CV were chosen to cover the designed deviations of ionic capacity and salt gradient length in the case study. Thereafter, all normalized UV signals smaller than 0.2 were cut off with the aim to reduce the data containing less relevant case-specific information. The resulting chromatograms can be seen in Fig. 7.4. This resulted in cheaper computation and proved to be time-saving for potential real-time application of the ANN. For ANN model calibration, backpropagation was used to correlate the results of *in silico* screening with the corresponding parameter sets. 50 % of the total data were used as training data

Table 7.2.: Parameters of the mass transfer model and kinetic isotherm formulation estimated from two bind-and-elute experiments using the inverse method.

Parameter	mAb	Cytochrome c	Lysozyme
$k_{film} \left[\frac{mm}{s}\right] \cdot 10^{-2}$	$7.516 \cdot 10^{-1} \pm 2.474 \cdot 10^{-1}$	1.899 ± 2.320	2.098 ± 1.871
$D_{pore} \left[\frac{mm^2}{s}\right] \cdot 10^{-5}$	2.325 ± 0.315	6.823 ± 1.482	6.304 ± 0.836
k_{eq} [-] $\cdot 10^{-5}$	$4.388 \cdot 10^{-5} \pm 0.522 \cdot 10^{-5}$	$4.860 \cdot 10^{-3} \pm 0.227 \cdot 10^{-3}$	2.982 ± 0.138
ν [-]	9.720 ± 0.038	10.353 ± 0.018	10.012 ± 0.033
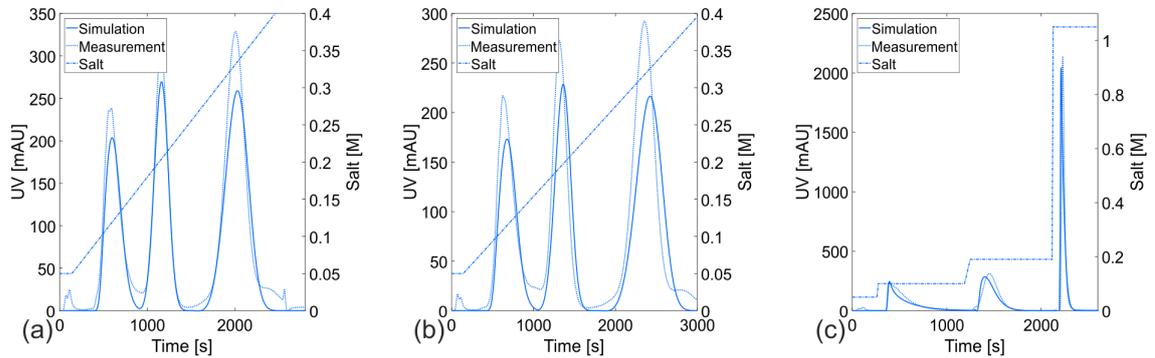
Figure 7.2.: Plots of UV signals over process run time for bind-and-elute experiments. Dashed lines display the UV signal measured at column outlet and the adjusted salt gradients. Solid lines represent the simulated chromatograms. The elution peaks of the early eluting component mAb, the moderate eluting component cytochrome c, and the late eluting component lysozyme by applying linear salt gradients from $0.05\,\mathrm{M}$ to $1.05\,\mathrm{M}$ over $20\,\mathrm{CV}$ and $25\,\mathrm{CV}$ are shown in (a) and (b). (c) represent a step-wise gradient elution from $0.05\,\mathrm{M}$, over $0.1\,\mathrm{M}$ and $0.19\,\mathrm{M}$, to $1.05\,\mathrm{M}$.
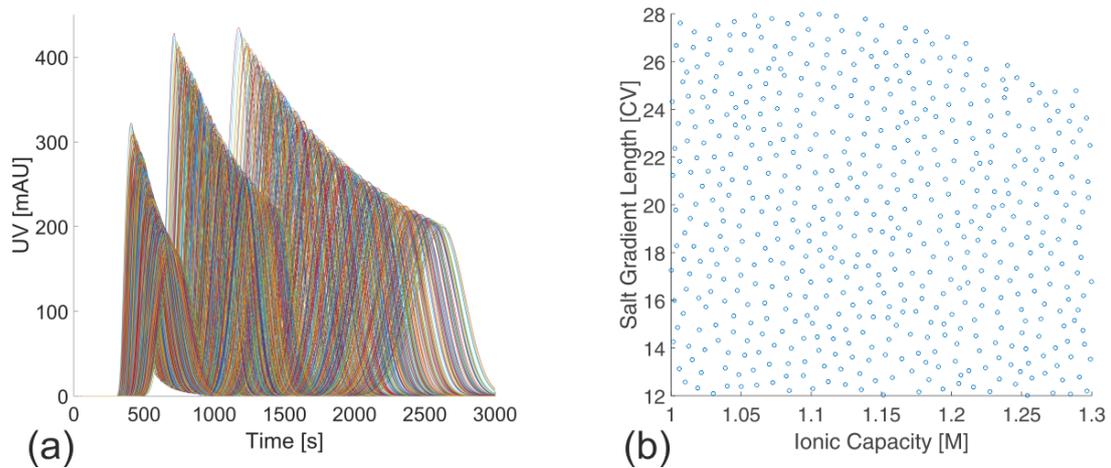
Figure 7.3.: The resulting chromatograms of the *in silico* screening are displayed in (a). The corresponding design space of the *in silico* screening is shown in (b): ionic capacity $\in [1,1.3]\,\mathrm{M}$, and salt gradient length $\in [12,28]\,\mathrm{CV}$.

set for the ANN's adjustment according to the respective errors. $30\,\%$ were used as validation data set to measure the generalization performance, and to stop training when generalization stopped improving. The remaining $20\,\%$ were excluded during model calibration and used subsequently as test data for cross-validation. The final correlation achieved for the salt gradient length is given in Figs. 7.5 (a)-(c) which represent the training data set, validation data set, and test data set, respectively. Figs. 7.5 (d)-(f) show the correlation for the ionic capacity.
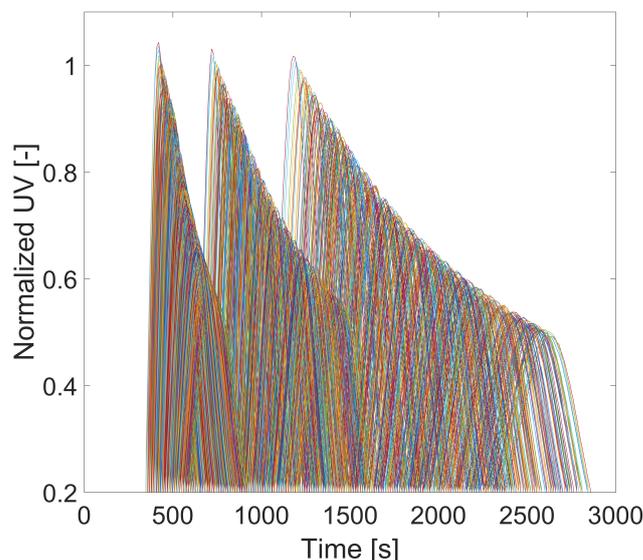
Figure 7.4.: The post-processed results of the *in silico* screening are shown.

A good agreement between the predictions by the ANN model and the parameter sets given in the *in silico* screening was observed. The smallest deviations were found to be in the correlation results of training data set ($R^2 = 0.9997$ for salt gradient length, and $R^2 = 0.9992$ for ionic capacity). The deviations found in the correlation results of validation data set were in a statistically acceptable range ($R^2 = 0.9926$ for salt gradient length, and $R^2 = 0.9822$ for ionic capacity). According to the correlation results of the test data set ($R^2 = 0.9965$ for salt gradient length, and $R^2 = 0.9904$ for ionic capacity), the ANN's generalization was confirmed, verifying its usability for the root cause investigation.

7.4.4. ANN-based Root Cause Investigation

Although the ANN's generalization capability was verified with the cross-validation using 20 % of the total *in silico* data sets, three runs with strong, moderate, and slight deviations in their chromatograms were generated in the laboratory to prove the practical applicability of the presented method for root cause investigation. The salt gradient length and ionic capacity were adjusted separately and simultaneously to challenge the calibrated ANN model with the pronounced problem, when multiple root causes were potentially responsible for one observed deviation. As can be seen in Table 7.3, the casual relations in all three deviated runs were determined properly. Overall, the deviations were found in a statistically acceptable range. Errors of approximately 0 % and 4.90 % were given for the prediction of the deviated run 1. The deviated run 2 with both adjusted salt gradient length and ionic capacity was
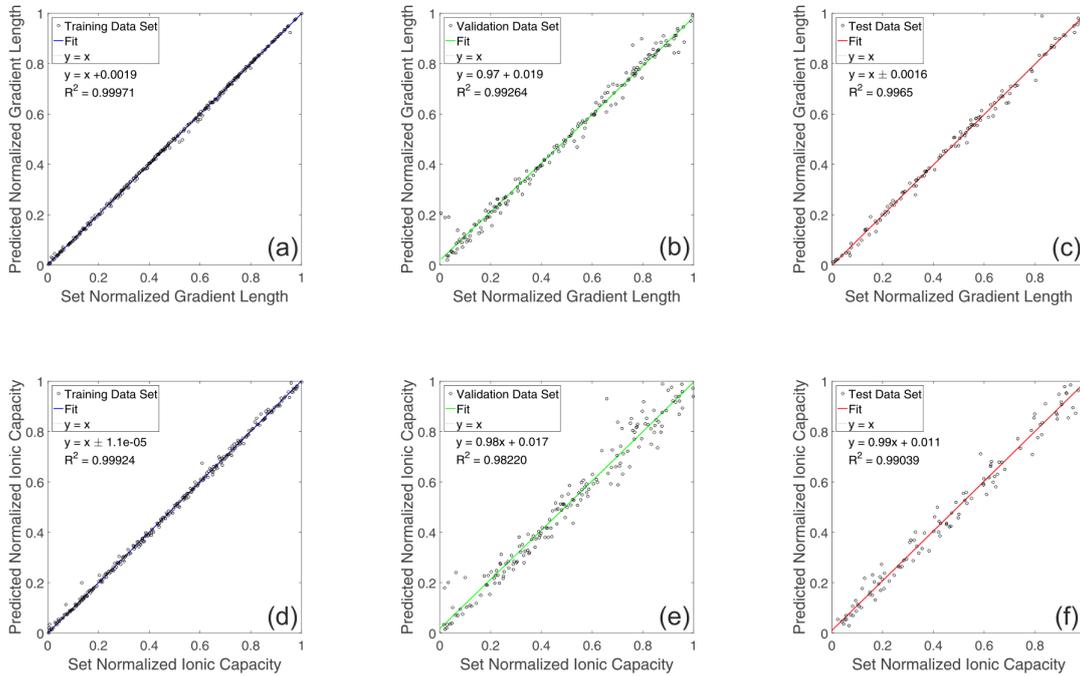
Figure 7.5.: Plots of screened parameters versus simulated parameters using ANN model. (a) - (c) represent the training data set, validation data set, and test data set of the parameter ionic capacity. (d) - (f) represent the salt gradient length.

predicted accurately with the small errors of $0.53\,\%$ and $0.55\,\%$. Errors of $1.5\,\%$ and $3.35\,\%$ were given for the prediction of the deviated run 3. Compared to the inverse modeling approach proposed by Brestrich and coworkers [261], the presented method is advantageous with respect to the computational expense. Unlike the necessity of performing parameter estimation repeatedly, the calibrated ANN model can be carried out with very little computational expense.

A correlation has been derived by Parente and Wetlaufer originally for determination of the adsorption isotherm model parameters in linear region [157]. Its modified formulation delivered by Shukla and coworkers [155] was employed with retention

Table 7.3.: The ionic capacity and salt gradient length adjusted in the laboratory and predicted with the ANN model, and the percentage error of prediction for three deviated runs.

	Ionic capacity [M]			Salt gradient length [CV]		
	Adjusted	Predicted	Error	Adjusted	Predicted	Error
Deviated run 1	1.172	1.172	$0\,\%$	15	15.735	$4.90\,\%$
Deviated run 2	1.130	1.136	$0.53\,\%$	20	20.109	$0.55\,\%$
Deviated run 3	1.130	1.147	$1.50\,\%$	25	24.128	$3.35\,\%$
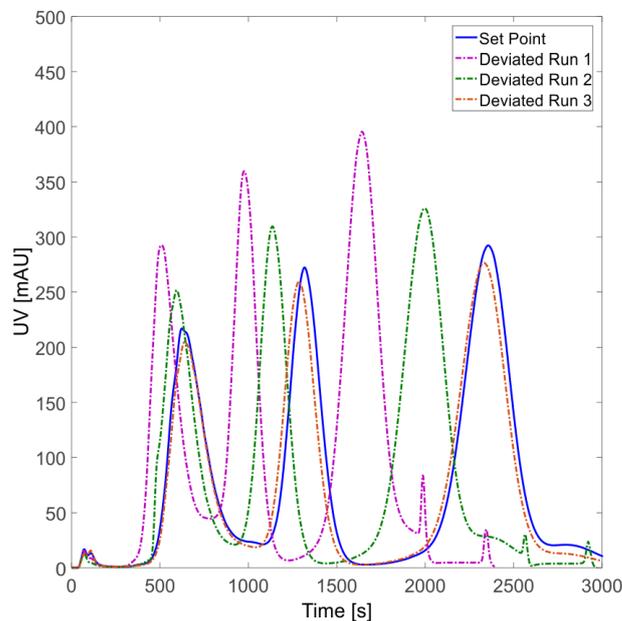
Figure 7.6.: Plots of UV signals over process run time for bind-and-elute experiments. The set point run displayed as solid blue line is identical with Fig. 1 and shown here for comparability. The deviated run 1 as dashed purple line was generated by adjusting the salt gradient length to 15 CV. The deviated run 2 as dashed green line was generated by adjusting both the salt gradient length to 20 CV and the ionic capacity to 1.130 M. The deviated run 3 as dashed orange line was generated by adjusting the ionic capacity to 1.130 M.

volumes of the solutes and solved for ionic capacity and gradient length. To compare the performance of this method and the presented one, the calculated results are shown in Tab. 7.4.

Table 7.4.: The ionic capacity and salt gradient length adjusted in the laboratory and calculated with the correlation delivered by Parente and Wetlaufer [157], and the percentage error of calculation for three deviated runs.

| | Ionic capacity [M] | | | Salt gradient length [CV] | | |
	Adjusted	Calculated	Error	Adjusted	Calculated	Error
Deviated run 1	1.172	1.093	6.76 %	15	15.104	0.65 %
Deviated run 2	1.130	1.309	15.81 %	20	15.003	23.27 %
Deviated run 3	1.130	1.175	3.99 %	25	20.906	15.25 %

7.5. Conclusion

This study presents a novel method for the root cause investigation in protein chromatography when an exact mathematical model which can immediately identify the root causes of the deviation is not given and the amount of data available does not allow the direct application of an ANN approach. The fundamental idea was to combine a mechanistic model and an ANN model to benefit from their respective advantages. The learning material for the ANN model was generated by conducting 570 *in silico* experiments using the calibrated mechanistic model. Based on these data, the ANN model learned the correlations between the appearances of the chromatograms and the parameters of interest. Since only few experiments were required for the calibration of the mechanistic model, the high wet-lab experimental effort needed for the training of the ANN model is circumvented. The generalization performance of the artificial intelligence of a very simple construction was assessed by correlation results of cross-validation. The practical applicability was verified by an experimental case study. Thereby, three experiments were conducted in the laboratory to generate the deviating chromatograms by deliberately causing errors in the parameters salt gradient length, ionic capacity, and both of them. These experiments represented the cases of strong, moderate, and slight deviations in the chromatogram, respectively. This pronounced problem, that multiple root causes could be potentially responsible for one observed deviation, was solved successfully. The ionic capacity that had hitherto been a non-observable parameter during operation, was notably accurate with percentage errors of $0\,\%$, $0.53\,\%$, and $1.50\,\%$.

With this method, one can identify the root causes of the deviation and determine their magnitudes within milliseconds. Thus, it is well-suited for real-time process monitoring and control. It allows an immediate correction of the chromatography process operating condition as a short-term solution. In this way, the economic harm can be reduced drastically, especially when innovative technologies providing high productivity e.g., continuous chromatography, are involved in the manufacturing. For future work, the method should be extended by a decision-making strategy to be able to correct the deviated operating conditions to the closest Pareto optimum.

7.6. Acknowledgment

8. Conclusion and Outlook

The presented thesis addresses issues inhibiting the know-how transfer of mechanistic modeling from academia to industry in DSP. Compared to technologies already transferred to the industrial work-flow such as HTE and DoE, mechanistic modeling provides numerous self-evident advantages, but shows challenges in model building, model calibration, and model application. In these three fields, problems of particular interest were defined *i.e.*, the lack of suitable mechanistic description for HIC and PEG induced protein precipitation, the requirement for repeated model calibration using traditional calibration techniques, the uniform utilization of mechanistic modeling to develop processes and to analyze process robustness. In this context, new mechanistic models were derived, novel model calibration method based on ANN was proposed, and further encouraging applications of mechanistic modeling were demonstrated and suggested in the thesis.

The first part of this thesis pointed out that modeling of protein behavior in HIC columns was seldom used due to the lack of a suitable adsorption isotherm model. Inspired by preceding works and recent insights into water structure on hydrophobic surfaces, an adsorption isotherm describing protein behavior dependent on variation of both salt and protein concentration was derived. Hereby, the hydration number of ions was considered as a function of salt concentration which reciprocally affects the equilibrium of water structure. Coupled with the GRM, convection, diffusion, and adsorption processes in chromatography column could be covered. Model's capability of describing bind-and-elute chromatograms and predictive power beyond calibration space were demonstrated successfully.

As protein precipitation did not only share a similar problem, but also highly similar mechanistic foundation with HIC, the second part of this thesis presented the derivation of a mechanistic model for protein precipitation using PEG. Here, the equilibrium between well-ordered and bulk-like ordered water molecules on hydrophobic surfaces was considered. Compared to the widely used models nowadays, the here introduced model has advantages such as the capability of modeling the entire precipitation curve in dependence of protein and PEG concentration. High-throughput experimental data of different proteins and experimental conditions were generated to prove the validity of the model.

In the general work-flow of mechanistic modeling, the model has to be calibrated properly after successful model selection and prior to the use of it at its full potential.

The third part of this thesis addressed the disadvantages of the commonly employed model calibration methods in protein chromatography, such as the requirement for repeated calibration upon new compounds entering the chromatographic system, the related time consumption, and the resulting computational expense. Solutions were offered with a newly developed model calibration method based on ANN. In the presented case study the trained ANN determined the mechanistic model parameters for uncalibrated compounds immediately and thereby shortened the model calibration procedure itself to milliseconds from hours or an even longer time.

To meet the QbD demand suggested by the US FDA, advanced process understanding provided by mechanistic modeling can be essential in order to transfer chromatographic processes from laboratory to production scale. In the fourth part of this thesis, GRM and SMA adsorption isotherm were employed to investigate the hitherto not completely understood behavior of protein conjugates in IEX. This study demonstrated that an increase in molecular weight of PEG resulted in abnormal behavior of the protein conjugates in IEX, such as unchanged elution peak shape and exponentially decreased binding capacity. Based on the determined mass transfer and adsorption parameters, the anomalies were explained mechanistically.

An other study on utilizing mechanistic chromatography modeling in IEX was presented in the fifth part of this thesis. Here, mechanistic modeling was combined with HTE and DoE resulting in a process development toolbox providing advantages such as, flexibility, time- and material saving. Based on this toolbox, an integrated precipitation and IEX process for the purification of an industrial antibody feed was developed. Its potential for superseding the cost driving Protein A chromatography in mAb manufacturing was demonstrated. Compared to the industrial Protein A chromatography step, the developed integrated process showed highly similar performance with respect to monomer yield, HCP reduction, and aggregate content. It is, however, inferior in terms of monomer concentration after pooling by threefold, but superior regarding DNA reduction by four orders of magnitude. Furthermore, the capability of the single column and three column continuous IEX processes was studied by calculating the Pareto frontiers based on the mechanistic models. The result indicated improvements in both purity and yield of monomer when switched to continuous processing.

In the sixth and last part, a new approach for the root cause investigation of deviations in protein chromatography is introduced. The presented case study focused on the ionic capacity of the chromatography column, which can be influenced by a variety of factors such as, column exchange, column aging, or day-to-day operations. The deviation of the ionic capacity is unavoidable in manufacturing processes and can have severe effects on the elution behavior of product and impurities, resulting in reduced product quality and patient safety. Deviations in salt elution gradient were introduced additionally. To determine the root causes of deviation, simulations of potential process deviations generated employing a mechanistic chromatography model were

used to construct an ANN model. After the ANN learned to recognize deviations in chromatograms and to respond with the corresponding operating parameter values, its practical applicability was verified by predicting deviations in wet laboratory experiments with deliberately introduced errors. The deviations in ionic capacity – a hitherto non-observable parameter during operation, was determined accurately with a maximum error of $1.5\,\%$.

In summary, this thesis not only contributes to the state-of-the-art in HIC modeling, protein precipitation modeling, model calibration technique, understanding for adsorption behavior of protein conjugates in IEX, but also explores further application scenarios of mechanistic modeling in DSP. With the proposed HIC and protein precipitation models, a mechanistic framework is provided for possible model extensions accounting for specific phenomena such as, irreversible binding, spreading, as well as multi-layer adsorption of proteins. Furthermore, the possibility for model-based process development in HIC and selective protein precipitation is given, which would offer deeper process understanding in accordance with the QbD principle, time- and material-saving, and the capability of process development in early stages of product development. The investigation on adsorption and mass transfer behavior of protein conjugates in IEX delivers mechanistic explanation for the anomalies observed, which are of great academic and practical interest. The knowledge gained would simplify the IEX process design and scale up for protein conjugate purification. The suggested tool box for process development combines the advantages of HTE, DoE, and mechanistic modeling. Its potential is demonstrated with the developed and optimized integrated process as a competitive alternative to the widely used Protein A chromatography for purification of industrial mAb feed stock. In future studies, the feasibility of the tool box and integrated process could be tested in larger scale, and both single- and multi-column format.

Compared to the mechanistic approaches, ANN – despite of its extremely versatile applications in other fields of science and engineering, has noticeably less successes in the field of chromatography modeling yet. The general reason is the lack of data in protein chromatography due to its high experimental effort and the limitation of material, which is compensated with the proposed method *i.e.*, to generate a necessary amount of data by conducting mechanistic model simulation for ANN training. On this way, two methods were suggested, which are proven to be efficient to calibrate mechanistic models and to investigate root causes of deviations in protein chromatography, respectively. The properly built and trained ANN models are not only advantageous regarding speed and computational expense, but also easy to use, even without the otherwise prerequisite deep mathematical understanding. In future work, the combined mechanistic and ANN-based approach could be used to realize automated real-time mechanistic model calibration during chromatography experiments, and implemented for real-time process monitoring and control. It could be extended by a process development tool and a decision-making strategy that enables immediate correction of operating parameters upon deviations. In long term,

these mechanistic and ANN-based methods could be very fundamental building blocks for an omniscient AI process expert in DSP.

Bibliography

1. *Mordern Biopharmaceuticals. Design, development and optimization* (ed KNÄBLEIN, J.) (Wiley-VCH, Weinheim, Germany, 2008) (cit. on p. 1).

2. KNEZEVIC, I. & E. GRIFFITHS: Biosimilars–global issues, national solutions. *Biologicals* (2011), vol. **39**: 252–255 (2011) (cit. on p. 1).

3. *From Science to Operations: Questions, Choices and Strategies for Success in Biopharma* (eds OTTO, R., A. SANTAGOSTINO & U. SCHRADER) (McKinsey, 2014) (cit. on p. 1).

4. GJOKA, X., K. ROGLER, R. A. MARTINO, R. GANTIER & M. SCHOFIELD: A straightforward methodology for designing continuous monoclonal antibody capture multi-column chromatography processes. *J. Chromatogr. A* (2015), vol. **1416**: 38–46 (2015) (cit. on pp. 1, 90).

5. CAI, H. H.: Therapeutic monoclonal antibodies approved by FDA in 2016. *MOJ. Immunol.* (2017), vol. **5**: 1–2 (2017) (cit. on pp. 1, 90).

6. ECKER, D. M, S. D. JONES & H. L. LEVINE: The therapeutic monoclonal antibody market. *mAbs* (2015), vol. **7**: 9–14 (2015) (cit. on pp. 1, 90).

7. SCHIESTL, M., T. STRANGLER, C. TORELLA, T. CEPELJNIK, H. TOLL & R. GRAU: Acceptable changes in quality attributes of glycosylated biopharmaceuticals. *Nat. Biotechnol.* (2011), vol. **29**: 310–312 (2011) (cit. on p. 1).

8. DGC, E. C.: Pharmaceutical Sector Inquiry Preliminar Report. *Technical report* (2008), vol. (2008) (cit. on p. 1).

9. FERNÁNDEZ, B. & J. MARTÍNEZ-HURTADO: Biosimilars: Company strategies to capture value from the biologics market. *Pharmaceuticals* (2012), vol. **5**: 1393–1408 (2012) (cit. on p. 1).

10. SEKHON, B. S. & V. SALUJA: Biosimilars: an overview. *Biosimilars* (2011), vol. **1**(1): 1–11 (2011) (cit. on pp. 1, 42).

11. MCKEAGE, K.: A review of ct-p13: An infliximab biosimilar. *BioDrugs* (2014), vol. **28**: 313–321 (2014) (cit. on p. 1).

12. JAIN, K.: *Textbook of Personalized Medicines* (Springer, 2009) (cit. on p. 1).

13. RADER, R. A. & E. S. LANGER: Biopharmaceutical manufacturing: Historical and future trends in titers, yields, and efficiency in commercial-scale bioprocessing. *BioProc. J.* (2015), vol. **13**: 47–54 (2015) (cit. on pp. 2, 90).

14.   Guiochon, G.: Preparative liquid chromatography. *J. Chromatogr. A* (2002), vol. **965**(1–2): 129 –161 (2002) (cit. on pp. 2, 54).

15.   Hubbuch, J.: Editorial: High-throughput process development. *Biotechnol. J.* (2012), vol. **7**(10): 1185–1185 (2012) (cit. on pp. 2, 54).

16.   Treier, K., P. Lester & J. Hubbuch: Application of genetic algorithms and response surface analysis for the optimization of batch chromatographic systems. *Biochem. Eng. J.* (2012), vol. **63**: 66 –75 (2012) (cit. on pp. 2, 54).

17.   *International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use, ICH-Endorsed Guide for ICH Q8/Q9/Q10 Implementation* http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Quality/Q8_9_10_QAs/PtC/Quality_IWG_PtCR2_6dec2011.pdf () (cit. on pp. 2, 24, 54, 83, 114).

18.   Chhatre, S., S. Farid, J. Coffman, P. Bird, A. Newcombe & N. Titchener-Hooker: How implementation of Quality by Design and advances in Biochemical Engineering are enabling efficient bioprocess development and manufacture. *J. Chem. Technol. Biot.* (2011), vol. **86**: 1125–1129 (2011) (cit. on pp. 2, 24, 42, 54, 114).

19.   Hanke, A. & M. Ottens: Purifying biopharmaceuticals: knowledge-based chromatographic process development. *Trends Biotechnol.* (2014), vol. **32**: 210–220 (2014) (cit. on pp. 2, 24, 54, 114).

20.   Nfor, B. K., T. Ahamed, G. W. van Dedem, P. D. Verhaert, L. A. van der Wielen, M. H. Eppink, E. J. van de Sandt & M. Ottens: Model-based rational methodology for protein purification process synthesis. *Chem. Eng. Sci.* (2013), vol. **89**: 185 –195 (2013) (cit. on pp. 2, 54).

21.   Huuk, T. C., T. Hahn, A. Osberghaus & J. Hubbuch: Model-based integrated optimization and evaluation of a multi-step ion exchange chromatography. *Sep. Purif. Technol.* (2014), vol. **136**: 207–222 (2014) (cit. on pp. 2, 24, 54).

22.   Müller-Späth, T., G. Ströhlein, L. Aumann, H. Kornmann, P. Valax, L. Delegrange, E. Charbaut, G. Baer, A. Lamproye, M. Jöhnck, M. Schulte & M. Morbidelli: Model simulation and experimental verification of a cation-exchange IgG capture step in batch and continuous chromatography. *J. Chromatogr. A* (2011), vol. **1218**: 5195 –5204 (2011) (cit. on pp. 2, 54).

23.   Close, E. J., J. R. Salm, D. G. Bracewell & E. Sorensen: A model based approach for identifying robust operating conditions for industrial chromatography with process variability. *Chem. Eng. Sci.* (2014), vol. **116**: 284 –295 (2014) (cit. on pp. 2, 54).

24.   Steinebach, F., M. Angarita, D. J. Karst, Müller-Späth & M. Morbidelli: Model based adaptive control of a continuous capture process for monoclonal antibodies production. *J. Chromatogr. A* (2016), vol. **1444**: 50 –56 (2016) (cit. on pp. 2, 54, 114).

25. Brestrich, N., T. Hahn & J. Hubbuch: Application of spectral deconvolution and inverse mechanistic modelling as a tool for root cause investigation in protein chromatography. *J. Chromatogr. A* (2016), vol. **1437**: 158 –167 (2016) (cit. on pp. 2, 54).

26. D'Archivio, A. A., A. Incani & F. Ruggieri: Cross-column prediction of gas-chromatographic retention of polychlorinated biphenyls by artificial neural networks. *J. Chromatogr. A* (2011), vol. **1218**: 8679 –8690 (2011) (cit. on pp. 2, 56, 114).

27. Madden, J. E., N. Avdalovic, P. R. Haddad & J. Havel: Prediction of retention times for anions in linear gradient elution ion chromatography with hydroxide eluents using artificial neural networks. *J. Chromatogr. A* (2001), vol. **910**(1): 173 –179 (2001) (cit. on pp. 2, 56, 114).

28. Bergés, R., V. Sanz-Nebot & J. Barbosa: Modelling retention in liquid chromatography as a function of solvent composition and pH of the mobile phase. *J. Chromatogr. A* (2000), vol. **869**(1–2): 27 –39 (2000) (cit. on pp. 2, 56, 114).

29. Malenović, A., B. Jančić-Stojanović, N. Kostić, D. Ivanović & M. Medenica: Optimization of Artificial Neural Networks for Modeling of Atorvastatin and Its Impurities Retention in Micellar Liquid Chromatography. *Chromatographia* (2011), vol. **73**(9): 993–998 (2011) (cit. on pp. 2, 56, 114).

30. Gupta, V. K., H. Khani, B. Ahmadi-Roudi, S. Mirakhorli, E. Fereyduni & S. Agarwal: Prediction of capillary gas chromatographic retention times of fatty acid methyl esters in human blood using MLR, PLS and back-propagation artificial neural networks. *Talanta* (2011), vol. **83**(3): 1014 –1022 (2011) (cit. on pp. 2, 56, 114).

31. Kuczeweski, M., E. Schirmer, B. Lain & G. Zarbis-Papastoitsis: A single-use purification process for the production of a monoclonal antibody produced in a PER.C6 human cell line. *Biotechnol. J.* (2011), vol. **1**: 56–65 (2011) (cit. on p. 2).

32. Oelmeier, S. A., C. Ladd-Effio & J. Hubbuch: Alternative separation steps for monoclonal antibody purification: Combination of centrifugal partitioning chromatography and precipitation. *J. Chromatogr. A* (2013), vol. **1319**: 118–126 (2013) (cit. on pp. 2, 97).

33. Winderl, J., T. Hahn & J. Hubbuch: A mechanistic model of ion-exchange chromatography on polymer fiber stationary phases. *J. Chromatogr. A* (2016), vol. **1475**: 18–30 (2016) (cit. on p. 2).

34. Tsotsas, E.: *Über die Wärme- und Stoffübertragung in durchströmten Festbetten, VDI Fortschrittsberichte* (VDI, Düsseldorf, Germany, 1987) (cit. on p. 3).

35. Jeansonne, M. & J. Foley: Review of the exponentially modified Gaussian chromatographic peak model since 1983. *J. Chromatogr. Sci.* (1991), vol. **29**: 258–266 (1991) (cit. on p. 4).

36. Jeansonne, M. & J. Foley: Improved Equations for Calculation of Chromatographic Figures of Merit for Ideal and Skewed Peaks. *J. Chromatogr.* (1992), vol. **594**: 1–8 (1992) (cit. on p. 4).

37. FOLEY, J. & J. DORSEY: Equations for calculation of chromatographic figures of merit for ideal and skewed peaks. *Anal. Chem.* (1983), vol. **55**: 730–737 (1983) (cit. on p. 4).

38. FOLEY, J. & J. DORSEY: A review of the exponentially modified Gaussian (EMG) function: Evaluation and subsequent calculation of universal data. *J. Chromatogr. Sci.* (1984), vol. **22**: 40–46 (1984) (cit. on p. 4).

39. ZHANG, Z., M. MAZZOTTI & M. MORBIDELLI: Power-Feed operation of simulated moving bed units: changing flow-rates during the switching interval. *J. Chromatogr. A* (2003), vol. **1006**: 87–99 (2003) (cit. on p. 5).

40. GUIOCHON, G., A. FELINGER, D. G. SHIRAZI & A. M. KATTI: *Fundamentals of Preparative and Nonlinear Chromatography* 251 (Elservier, 2006) (cit. on pp. 5, 7, 120).

41. *Preparative Chromatography* 2nd (eds SCHMIDT-TRAUB, H., M. SCHULTE & A. SEIDEL-MORGENSTERN) (Wiley, 2012) (cit. on pp. 7, 24, 28, 54, 57, 81, 93, 114, 115).

42. WICKE, E.: Empirische und theoretische Untersuchungen der Sorptionsgeschwindigkeit von Gasen an porösen Stoffen II. *Kollod Z.* (1939), vol. **86**: 295–313 (1939) (cit. on p. 7).

43. DE VAULT, D.: The theory of chromatography. *J. Am. Chem. Soc.* (1943), vol. **65**: 532–540 (1943) (cit. on p. 7).

44. LAPIDUS, L. & N. AMUNDSON: A descriptive theory of leaching: Mathematics of adsorption beds. *J. Phys. Chem.* (1952), vol. **56**: 984–988 (1952) (cit. on p. 8).

45. VAN DEEMTER, J., F. ZUIDERWEG & A. KLINKENBERG: Longitudinal diffusion and resistance to mass transfer as causes of nonideality in chromatography. *Chem. Eng. Sci.* (1956), vol. **5**: 271–289 (1956) (cit. on p. 8).

46. GLUECKAUF, E.: Theory of Chromatography. Part II. Chromatograms of a Single Solute. *J. Chem. Soc.* (1947), vol.: 1302–1308 (1947) (cit. on p. 8).

47. GLUECKAUF, E. & J. COATES: Theory of chromatography, Part IV: The influence of incomplete equilibrium on the front boundaryof chromatogram and on the effectiveness of separation. *J. Chem. Soc.* (1947), vol.: 1315–1321 (1947) (cit. on p. 8).

48. GLUECKAUF, E.: Theory of chromatography. VII. The general theory of two solutes following non-linear isotherms. *Disc. Faraday Soc* (1949), vol. **7**: 12 (1949) (cit. on p. 8).

49. GU, T., G. TSAI & G. TSAO: New approach to a general nonlinear multicomponent chromatography model. *AIChE J.* (1990), vol. **36**: 784–788 (1990) (cit. on p. 9).

50. GU, T., G. TSAI, G. TSAO & M. LADISCH: Displacement effect in multicomponent chromatography. *AIChE J.* (1990), vol. **36**: 1156–1162 (1990) (cit. on p. 9).

51. Furuya, F., Y. Takeuchi & K. Noll: Intraparticle diffusion of phenols within bidispersed macrorecticular resin particles. *J. Chem. Eng. Japan* (1989), vol. **22**: 670 (1989) (cit. on p. 9).

52. Ma, Z., R. Whitley & N. Wang: Pore and Surface Diffusion in Multicomponent Adsorption and Liquid Chromatography Systems. *AIChE J.* (1996), vol. **42**: 1244–1262 (1996) (cit. on p. 9).

53. Suzuki, M.: *Adsorption Engineering* (Elsevier Science Publishers, Amsterdam, 1990) (cit. on p. 9).

54. Danckwerts, P.: Continuous flow systems - Distribution of residence times. *Chem. Eng. Sci.* (1953), vol. **2**: 1–13 (1953) (cit. on p. 9).

55. Mollerup, J. M.: A review of the thermodynamics of protein association to ligands, protein adsorption, and adsorption isotherms. *Chem. Eng. Technol.* (2008), vol. **31**: 864–874 (2008) (cit. on p. 10).

56. Velayudhan, A. & C. Horvath: Preparative chromatography of proteins - analysis of the multivalent ion-exchange formalism. *J. Chromatogr.* (1988), vol. **443**: 13 –29 (1988) (cit. on pp. 10, 54, 57, 114, 116).

57. Brooks, C. A. & S. M. Cramer: Steric mass-action ion exchange: displacement profiles and induced salt gradients. *AIChE Journal* (1992), vol. **38**(12): 1969–1978 (1992) (cit. on pp. 11, 69, 74).

58. Mollerup, J.: Applied thermodynamics: A new frontier for biotechnology. *Fluid Phase Equilibr.* (2006), vol. **241**: 205–215 (2006) (cit. on pp. 11, 12, 24, 26, 27, 33, 36, 38, 39, 43).

59. Kramer, R. M., V. R. Shende, N. Motl & C. N. J. MartinScholtz: Toward a Molecular Understanding of Protein Solubility: Increased Negative Surface Charge Correlates with Increased Solubility. *Biophys. Journal* (2012), vol. **102**: 1907–1915 (2012) (cit. on p. 12).

60. Cohn, E. J.: THE PHYSICAL CHEMISTRY OF THE PROTEINS. *J. Gen. Physiol.* (1922), vol. **4**: 697–722 (1922) (cit. on p. 12).

61. Chick, H. & C. J. Martin: The Precipitation of Egg-Albumin by Ammonium Sulphate. A Contribution to the Theory of the "Salting-out" of Proteins. *Biochem. J.* (1913), vol. **7**: 380–398 (1913) (cit. on p. 12).

62. Sørensen, S. & M. Høyrup. (1917), vol. **7**: 164 (1917) (cit. on p. 13).

63. Sørensen, S. & M. Høyrup. (1917), vol. **7**: 213 (1917) (cit. on p. 13).

64. Polson, A, G. Potgieter, J. Largier, G. Mears & F. Joubert: The fractionation of protein mixtures by linear polymers of high molecular weight. *BBA-Gen. Subjects* (1964), vol. **82**(3): 463–475 (1964) (cit. on pp. 13, 92).

65. Cohn, E. J.: THE PHYSICAL CHEMISTRY OF THE PROTEINS. *Phys. Rev.* (1925), vol. **5**(3): 349–437 (1925) (cit. on pp. 13, 43, 92).

66. SIM, S.-L., T. HE, A. TSCHELIESSNIG, M. MUELLER, R. B. TAN & A. JUNGBAUER:
     Protein precipitation by polyethylene glycol: A generalized model based on hydrody-
     namic radius. *J. Biotechnol.* (2012), vol. **157**(2): 315–319 (2012) (cit. on pp. 13, 43,
     92).

67. HÄMMERLING, F., C. L. EFFIO, S. ANDRIS, J. KITTELMANN & J. HUBBUCH:
     Investigation and prediction of protein precipitation by polyethylene glycol using
     quantitative structure–activity relationship models. *J. Biotechnol.* (2017), vol. **241**:
     87–97 (2017) (cit. on pp. 13, 43).

68. ATHA, D. H. & K. C. INGHAM: Mechanism of precipitation of proteins by polyet-
     hylene glycols. Analysis in terms of excluded volume. *J. Biol. Chem.* (1981), vol.
     256(23): 12108–12117 (1981) (cit. on pp. 13, 43, 92).

69. MCCULLOCH, W. & W. PITTS: A logical calculus of the ideas immanent in nervous
     activity. *Bull. Math. Biophys.* (1943), vol. **5**: 115–133 (1943) (cit. on p. 14).

70. *THE BLUE BRAIN PROJECT - A SWISS BRAIN INITIATIVE* `https://`
     `bluebrain.epfl.ch/`. Accessed: 2018-03-28 (cit. on p. 14).

71. SILVER, D. *et al.*: Mastering the game of Go with deep neural networks and tree
     search. *Nature* (2016), vol. **529**: 484–489 (2016) (cit. on p. 14).

72. *WAYMO* `https://waymo.com/`. Accessed: 2018-03-28 (cit. on p. 14).

73. KERMANI, B., M. WHITE & H. NAGLE: A new method in obtaining a better
     generalization in artificial neural networks. *Conf. Proc. IEEE Eng. Med. Biol. Soc.*
     (1994), vol.: 119–1120 (1994) (cit. on p. 15).

74. *Neural Network Computing for the Electric Power Industry: Proceedings of the 1992
     Inns Summer Workshop* (ed SOBAJIC, D.) (Psychology Press, 1993) (cit. on p. 15).

75. RUSSELL, S. J. & P. NORVIG: *Artificial Intelligence: A Modern Approach* (Prentice
     Hall, 2010) (cit. on p. 16).

76. MOHRI, M., A. ROSTAMIZADEH & A. TALWALKAR: *Foundations of Machine Learning*
     (The MIT Press, 2012) (cit. on p. 16).

77. ABRAHAM, A.: Meta learning evolutionary artificial neural networks. *Neurocomputing*
     (2004), vol. **56**: 1–38 (2004) (cit. on p. 16).

78. GOODFELLOW, I., Y. BENGIO & A. COURVILLE: *Deep learning* (The MIT Press,
     2016) (cit. on p. 16).

79. MOLLERUP, I., S. JENSEN, P. LARSEN, O. SCHOU & L. SNEL: *Encyclopaedia of Bio-
     processTechnology: Fermentation, Biocatalysis, and Bioseparation* (eds FLICKINGER,
     M. & S. DREW) 1491 (John Wiley, 1999) (cit. on p. 24).

80. JAGSCHIES, G., A. GRONBERG, T. BJORKMAN, K. LACKI & H. JOHANSSON:
     Technical and economical evaluation of downstream processing options for monoclonal
     antibody (MAb) production. *Biopharm. Int.* (2006), vol. **19**: 10–19 (2006) (cit. on
     p. 24).

81. SOFER, G. & L. CHIRICA: Improving productivity in downstream processing. *Biopharm. Int.* (2006), vol. **19**: 48–55 (2006) (cit. on p. 24).

82. CHEN, J., J. TETRAULT & A. LEY: Comparison of standard and new generation hydrophobic interaction chromatography resins in the monoclonal antibody purification process. *J. Chromatogr. A* (2008), vol. **1177**: 272–281 (2008) (cit. on p. 24).

83. GUSE, A., A. MILTON, H. SCHULZE-KOOPS, E. R. B. MÜLLER & B. SIMMER: Purification and analytical characterization of an anti-CD4 monoclonal antibody for human therapy. *J. Chromatogr. A* (1994), vol. **661**: 13–23 (1994) (cit. on p. 24).

84. TORNØE, I., I. L. TITLESTAD, K. KEJLING, K. ERB, H. J. DITZEL & J. C. JENSENIUS: Pilot scale purification of human monoclonal IgM (COU-1) for clinical trials. *J. Immunol Methods* (1997), vol. **205**: 11–17 (1997) (cit. on p. 24).

85. NFOR, B., P. VERHAERT, L. van WIELEN, J. HUBBUCH & M. OTTENS: Rational and systematic protein purification process development: the next generation. *Trends Biotechnol.* (2009), vol. **27**: 673–679 (2009) (cit. on pp. 24, 83).

86. BAUMANN, P., K. BAUMGARTNER & J. HUBBUCH: Influence of binding pH and protein solubility on the dynamic binding capacity in hydrophobic interaction chromatography. *J. Chromatogr. A* (2015), vol. **1396**: 77–85 (2015) (cit. on pp. 24, 39).

87. MAHN, A. & J. A. ASENJO: Prediction of protein retention in hydrophobic interaction chromatography. *Biotechnol. Adv.* (2005), vol. **23**: 359–368 (2005) (cit. on p. 24).

88. MOLLERUP, J. M., T. B. HANSEN, S. KIDAL & A. STABY: Quality by design: Thermodynamic modelling of chromatographic separation of proteins. *J. Chromatogr. A* (2008), vol. **1177**: 200–206 (2008) (cit. on p. 24).

89. MOLLERUP, J. M., T. B. HANSEN, S. KIDAL, L. SEJERGAARD & A. STABY: Development, modelling, optimisation and scale-up of chromatographic purification of a therapeutic protein. *Fluid Phase Equilibr.* (2007), vol. **261**: 133–139 (2007) (cit. on p. 24).

90. BROOKS, C. A. & S. M. CRAMER: Steric mass-action ion exchange: Displacement profiles and induced salt gradients. *Fluid Phase Equilibr.* (1992), vol. **38**: 1969–1978 (1992) (cit. on pp. 24, 27, 54, 55, 93, 114, 116).

91. MELANDER, W. & C. HORVÁTH: Salt effect on hydrophobic interactions in precipitation and chromatography of proteins: an interpretation of the lyotropic series. *Archives of biochemistry and biophysics* (1977), vol. **183**: 200–215 (1977) (cit. on p. 24).

92. STABY, A. & J. MOLLERUP: Solute retention of lysozyme in hydrophobic interaction perfusion chromatography. *J. Chromatogr. A* (1996), vol. **734**: 205–212 (1996) (cit. on p. 24).

93. PERKINS, T. W., D. S. MAK, T. W. ROOT & E. N. LIGHTFOOT: Protein retention in hydrophobic interaction chromatography: Modeling variation with buffer ionic strength and column hydrophobicity. *J. Chromatogr. A* (1997), vol. **766**: 1–14 (1997) (cit. on pp. 24, 26).

94. CHEN, J. & S. M. CRAMER: Protein adsorption isotherm behavior in hydrophobic interaction chromatography. *J. Chromatogr. A* (2007), vol. **1165**: 67–77 (2007) (cit. on p. 24).

95. HAIMER, E., A. TSCHELIESSNIG, R. HAHN & A. JUNGBAUER: Hydrophobic interaction chromatography of proteins IV. Kinetics of protein spreading. *J. Chromatogr. A* (2007), vol. **1139**: 84–94 (2007) (cit. on pp. 25, 39).

96. LUNDSTRÖM, I.: Models of Protein adsorption on solid surfaces. *Prog. Coll. Pol. Sci. S.* (1985), vol. **70**: 76–82 (1985) (cit. on p. 25).

97. JUNGBAUER, A., C. MACHOLD & R. H. DEPARTMENT: Hydrophobic interaction chromatography of proteins: III. Unfolding of proteins upon adsorption. *J. Chromatogr. A* (2005), vol. **1079**: 221–228 (2005) (cit. on pp. 25, 43).

98. DEITCHER, R., J. ROME, P. GILDEA, J. O'CONNEL & E. FERNANDEZ: A new thermodynamic model describes the effects of ligand density and type, salt concentration and protein species in hydrophobic interaction chromatography. *J. Chromatogr. A* (2010), vol. **1217**: 199–208 (2010) (cit. on pp. 25, 26, 33, 36, 38, 39).

99. CHEN, J. & Y. SUN: Modeling of the salt effects on hydrophobic adsorption equilibrium of protein. *J. Chromatogr. A* (2003), vol. **992**: 29–40 (2003) (cit. on pp. 25, 39).

100. MIRANI, M. R. & F. RAHIMPOUR: Thermodynamic modelling of hydrophobic interaction chromatography of biomolecules in the presence of salt. *J. Chromatogr. A* (2015), vol. **1422**: 170–177 (2015) (cit. on p. 25).

101. TANFORD, C.: *The Hydrophobic Effect: Formation of Micelles and Biological Membranes* (Wiley-Interscience, New York, 1980) (cit. on pp. 25, 43).

102. FRANK, H. S. & M. W. EVANS: Free volume and entropy in condensed systems. III. *J. Chem. Phys.* (1945), vol. **13**: 507–532 (1945) (cit. on p. 25).

103. HEAD-GORDON, T.: Is water structure around hydrophobic groups clathrate-like. *P. Natl. Acad. Sci. USA* (1995), vol. **92**: 8308–8312 (1995) (cit. on p. 25).

104. SHIRAGA, K., T. SUZUKI, N. KONDO & Y. OGAWA: Hydration and hydrogen bond network of water around hydrophobic surface investigated by terahertz spectroscopy. *J. Chem. Phys.* (2014), vol. **141**: 235103 (2014) (cit. on p. 25).

105. GENG, X., L. GUO & J. CHANG: Study of the retention mechanism of proteins in hydrophobic interaction chromatography. *J. Chromatogr. A* (1990), vol. **507**: 1–23 (1990) (cit. on p. 26).

106. HOWARTH, O. W.: Reassessment of hydrophobic bonding. *J. Chem. Soc., Faraday Trans. 1* (1975), vol. **71**: 2303–2309 (1975) (cit. on p. 26).

107.   BLOKZIJL, W. & J. B.F. N. ENGBERTS: Hydrophobic Effects. Opinions and Facts. *Angew. Chem.-Ger. Edit.* (1993), vol. **32**: 1545–1579 (1993) (cit. on p. 26).

108.   SOUTHALL, N. T., K. A. DILL & A. D. J. HAYMET: A View of the Hydrophobic Effect. *J. Phys. Chem. B* (2002), vol. **106**: 521–533 (2002) (cit. on p. 26).

109.   AFANASIEV, V. N., A. N. USTINOV & I. YU.VASHURINA: Acoustic study of solvent coordination in the hydration shells of potassium iodide. *J. Solution Chem.* (2006), vol. **35**: 1477–1491 (2006) (cit. on p. 27).

110.   HAHN, T., A. SOMMER, A. OSBERGHAUS, V. HEUVELINE & J. HUBBUCH: Adjoint-based estimation and optimization for column liquid chromatography models. *Comput. Chem. Eng.* (2014), vol. **64**: 41–54 (2014) (cit. on p. 28).

111.   GLOWINSKI, R.: *Finite element methods for incompressible viscous flow. In Numerical Methods for Fluids (Part 3)* (eds CIARLET, P. & J. LIONS) 3–1176 (Elsevier, 2003) (cit. on pp. 29, 58, 76, 99, 117).

112.   QUARTERONI, A. & A. VALLII: *Numerical approximation of partial differential equations* 160 (Springer, 1994) (cit. on pp. 29, 58, 76, 99, 117).

113.   KACZMARSKI, K. & D. ANTOS: Use of simulated annealing for optimization of chromatographic separations. *Acta Chromatogr.* (2006), vol. **17**: 20–45 (2006) (cit. on pp. 29, 75, 99, 117).

114.   DEVERNAY, F.: *C/C++ Minpack* http://devernay.free.fr/hacks/cminpack/. 2007 (cit. on pp. 29, 58, 62, 75, 99, 117, 120).

115.   HAHN, T., T. HUUK, V. HEUVELINE & J. HUBBUCH: Simulating and Optimizing Preparative Protein Chromatography with ChromX. *J. Chem. Educ.* (2015), vol. **92**: 1497–1502 (2015) (cit. on pp. 29, 58, 71, 75, 94, 118).

116.   HAHN, T., P. BAUMANN, T. HUUK, V. HEUVELINE & J. HUBBUCH: UV absorption-based inverse modeling of protein chromatography. *Eng. Life Sci.* (2016), vol. **16**: 99–106 (2016) (cit. on pp. 31, 55, 58, 59, 75, 98, 116, 120).

117.   ANTIA, F. D., I. FELLEGVÁRI & C. HORVÁTH: Displacement of Proteins in Hydrophobic Interaction Chromatography. *Ind. Eng. Chem. Res.* (1995), vol. **34**: 2796–2804 (1995) (cit. on p. 39).

118.   MOORKENS, E., N. MEUWISSEN, I. HUYS, P. DECLERCK, A. G. VULTO & S. SIMOENS: The market of biopharmaceutical medicines: a snapshot of a diverse industrial landscape. *Front Pharmacol.* (2017), vol. **8**: 314 (2017) (cit. on p. 42).

119.   MARICHAL-GALLARDO, P. & M. ALVAREZ: State-of-the-art in downstream processing of monoclonal antibodies: Process trends in design and validation. *Biotechnol. Progr.* (2012), vol. **28**(4): 899–916 (2012) (cit. on p. 42).

120.   FARID, S. S.: Process economics of industrial monoclonal antibody manufacture. *J. Chromatogr. B.* (2007), vol. **848**(1): 8–18 (2007) (cit. on p. 42).

121.   MESTRE-FERRANDIZ, J., A. TOWSE & M. BERDUD: Biosimilars: how can payers get long-term savings? *Pharmacoeconomics* (2016), vol. **34**(6): 609–616 (2016) (cit. on p. 42).

122. CHON, J. H. & G. ZARBIS-PAPASTOITSIS: Advances in the production and downstream processing of antibodies. *New Biotechnol.* (2011), vol. **28**(5): 458–463 (2011) (cit. on p. 42).

123. JUNGBAUER, A.: Continuous downstream processing of biopharmaceuticals. *Trends Biotechnol.* (2013), vol. **31**(8): 479–492 (2013) (cit. on p. 42).

124. COHN, E. J., L. E. STRONG, W. HUGHES, D. MULFORD, J. ASHWORTH, M. E. MELIN & H. TAYLOR: Preparation and Properties of Serum and Plasma Proteins. IV. A System for the Separation into Fractions of the Protein and Lipoprotein Components of Biological Tissues and Fluids1a, b, c, d. *J. Am. Chem. Soc.* (1946), vol. **68**(3): 459–475 (1946) (cit. on pp. 42, 92).

125. BROOKS, D., T. BRADFORD & J. HOPWOOD: An improved method for the purification of IgG monoclonal antibodies from culture supernatants. *J. Immunol. Methods* (1992), vol. **155**(1): 129–132 (1992) (cit. on pp. 42, 92).

126. KUEHNER, D. E., H. W. BLANCH & J. M. PRAUSNITZ: Salt-induced protein precipitation: Phase equilibria from an equation of state. *Fluid Phase Equilibr.* (1996), vol. **116**(1): 140–147 (1996) (cit. on pp. 42, 92).

127. GHOSH, R.: Separation of human albumin and IgG by a membrane-based integrated bioseparation technique involving simultaneous precipitation, microfiltration and membrane adsorption. *J. Membrane Sci.* (2004), vol. **237**(1): 109–117 (2004) (cit. on pp. 42, 92).

128. BUCHACHER, A. & G. IBERER: Purification of intravenous immunoglobulin G from human plasma–aspects of yield and virus safety. *Biotechnol. J.* (2006), vol. **1**(2): 148–163 (2006) (cit. on p. 42).

129. TSOKA, S., O. C. CINIAWSKYJ, O. R. THOMAS, N. J. TITCHENER-HOOKER & M. HOARE: Selective Flocculation and Precipitation for the Improvement of Virus-Like Particle Recovery from Yeast Homogenate. *Biotechnol. Progr.* (2000), vol. **16**(4): 661–667 (2000) (cit. on p. 42).

130. ZELTINS, A.: Construction and characterization of virus-like particles: a review. *Mol. Biotechnol.* (2013), vol. **53**(1): 92–107 (2013) (cit. on p. 42).

131. MCDONALD, P., C. VICTA, J. N. CARTER-FRANKLIN & R. FAHRNER: Selective antibody precipitation using polyelectrolytes: A novel approach to the purification of monoclonal antibodies. *Biotechnol. Bioeng.* (2009), vol. **102**: 1141–1151 (2009) (cit. on pp. 42, 90).

132. SOMMER, R., P. SATZER, A. TSCHELIESSNIG, H. SCHULZ, B. HELK & A. JUNGBAUER: Combined polyethylene glycol and CaCl 2 precipitation for the capture and purification of recombinant antibodies. *Process Biochem.* (2014), vol. **49**(11): 2001–2009 (2014) (cit. on p. 42).

133. CRAMER, S. M. & M. A. HOLSTEIN: Downstream bioprocessing: recent advances and future promise. *Curr. Opin. in Chem. Eng.* (2011), vol. **1**(1): 27–37 (2011) (cit. on p. 42).

134. Lewus, R. A., P. A. Darcy, A. M. Lenhoff & S. I. Sandler: Interactions and phase behavior of a monoclonal antibody. *Biotechnol. Progr.* (2011), vol. **27**(1): 280–289 (2011) (cit. on p. 42).

135. Baumann, P. & J. Hubbuch: Downstream process development strategies for effective bioprocesses: Trends, progress, and combinatorial approaches. *Eng. Life. Sci.* (2017), vol. **17**(11): 1142–1158 (2017) (cit. on p. 42).

136. Foster, P., P Dunnill & M. Lilly: The kinetics of protein salting-out: precipitation of yeast enzymes by ammonium sulfate. *Biotechnol. Bioeng.* (1976), vol. **18**(4): 545–580 (1976) (cit. on p. 43).

137. Przybycien, T. M. & J. E. Bailey: Solubility-activity relationships in the inorganic salt-induced precipitation of α-chymotrypsin. *Enzyme Microb. Tech.* (1989), vol. **11**(5): 264–276 (1989) (cit. on p. 43).

138. Laurent, T.: The interaction between polysaccharides and other macromolecules. 5. The solubility of proteins in the presence of dextran. *Biochem. J.* (1963), vol. **89**(2): 253 (1963) (cit. on pp. 43, 92).

139. Iverius, P. & T. Laurent: Precipitation of some plasma proteins by the addition of dextran or polyethylene glycol. *BBA-Protein Struct. M.* (1967), vol. **133**(2): 371–373 (1967) (cit. on pp. 43, 92).

140. Polson, A: A theory for the displacement of proteins and viruses with polyethylene glycol. *Prep. Biochem.* (1977), vol. **7**(2): 129–154 (1977) (cit. on pp. 43, 92).

141. De Young, L. R., A. L. Fink & K. A. Dill: Aggregation of globular proteins. *Accounts Chem. Res.* (1993), vol. **26**(12): 614–620 (1993) (cit. on p. 43).

142. Frank, H. S. & M. W. Evans: Free volume and entropy in condensed systems III. Entropy in binary liquid mixtures; partial molal entropy in dilute solutions; structure and thermodynamics in aqueous electrolytes. *J. Chem. Phys.* (1945), vol. **13**(11): 507–532 (1945) (cit. on p. 43).

143. Head-Gordon, T.: Is water structure around hydrophobic groups clathrate-like? *P. Natl. Acad. Sci. USA* (1995), vol. **92**(18): 8308–8312 (1995) (cit. on p. 43).

144. Shiraga, K, T Suzuki, N Kondo & Y Ogawa: Hydration and hydrogen bond network of water around hydrophobic surface investigated by terahertz spectroscopy. *J. Chem. Phys.* (2014), vol. **141**(23): 12B647_1 (2014) (cit. on p. 43).

145. Melander, W. & C. Horváth: Salt effects on hydrophobic interactions in precipitation and chromatography of proteins: an interpretation of the lyotropic series. *Arch. Biochem. Biophys.* (1977), vol. **183**(1): 200–215 (1977) (cit. on p. 43).

146. Nfor, B. K., N. N. Hylkema, K. R. Wiedhaup, P. D. Verhaert, L. A. Van der Wielen & M. Ottens: High-throughput protein precipitation and hydrophobic interaction chromatography: salt effects and thermodynamic interrelation. *J. Chromatogr. A* (2011), vol. **1218**(49): 8958–8973 (2011) (cit. on p. 43).

147. BAUMGARTNER, K., S. GROSSHANS, J. SCHÜTZ, S. SUHM & J. HUBBUCH: Prediction of salt effects on protein phase behavior by HIC retention and thermal stability. *J. Pharmaceut. biomed.* (2016), vol. **128**: 216–225 (2016) (cit. on p. 43).

148. STABY, A. & J. MOLLERUP: Solute retention of lysozyme in hydrophobic interaction perfusion chromatography. *J. Chromatogr. A* (1996), vol. **734**(1): 205–212 (1996) (cit. on p. 43).

149. CHEN, J. & S. M. CRAMER: Protein adsorption isotherm behavior in hydrophobic interaction chromatography. *J. Chromatogr. A* (2007), vol. **1165**(1): 67–77 (2007) (cit. on p. 43).

150. WANG, G., T. HAHN & J. HUBBUCH: Water on hydrophobic surfaces: Mechanistic modeling of hydrophobic interaction chromatography. *J. Chromatogr. A* (2016), vol. **1465**: 71–78 (2016) (cit. on pp. 43, 55).

151. HAMMERSCHMIDT, N., S. HOBIGER & A. JUNGBAUER: Continuous polyethylene glycol precipitation of recombinant antibodies: Sequential precipitation and resolubilization. *Process Biochem.* (2016), vol. **51**(2): 325–332 (2016) (cit. on p. 45).

152. AFANASIEV, V. N., A. N. USTINOV & I. Y. VASHURINA: Acoustic study of solvent coordination in the hydration shells of potassium iodide. *J. Solution Chem.* (2006), vol. **35**(11): 1477–1491 (2006) (cit. on p. 46).

153. XIA, F., D. NAGRATH & S. M. CRAMER: Effect of pH changes on water release values in hydrophobic interaction chromatographic systems. *Journal of Chromatography A* (2005), vol. **1079**(1): 229–235 (2005) (cit. on p. 50).

154. ZHOU, R., X. HUANG, C. J. MARGULIS & B. J. BERNE: Hydrophobic collapse in multidomain protein folding. *Science* (2004), vol. **305**(5690): 1605–1609 (2004) (cit. on p. 50).

155. SHUKLA, A. A., S. S. BAE, J. A. MOORE, K. A. BARNTHOUSE, & S. M. CRAMER: Synthesis and Characterization of High-Affinity, Low Molecular Weight Displacers for Cation-Exchange Chromatography. *Ind. Eng. Chem. Res.* (1998), vol. **37**(10): 4090–4098 (1998) (cit. on pp. 55, 126).

156. RÜDT, M., F. GILLET, S. HEEGE, J. HITZLER, B. KALBFUSS & B. GUÉLAT: Combined Yamamoto approach for simultaneous estimation of adsorption isotherm and kinetic parameters in ion-exchange chromatography. *J. Chromatogr. A* (2015), vol. **1413**: 68 –76 (2015) (cit. on p. 55).

157. PARENTE, E. S. & D. B. WETLAUFER: Relationship between isocratic and gradient retention times in the high-performance ion-exchange chromatography of proteins: Theory and experiment. *J. Chromatogr. A* (1986), vol. **355**: 29 –40 (1986) (cit. on pp. 55, 126, 127).

158. SOEIRO, F. J.C. P., H. F. C. VELHO, P. O. SOARES & A. J. S. NETO: Using neural networks to obtain initial estimates for the solution of inverse heat transfer problems. *Inverse Problems, Design and Optimization Symposium, Rio de Janeiro, Brazil* (2004), vol. (2004) (cit. on pp. 55, 56).

159. Knupp, D. C. & A. J. S. Neto: Solution of the inverse radiative transfer problem of simultaneous identification of the optical thickness and space-dependent albedo using Bayesian inference. *Comput. Model. Eng. Sci.* (2013), vol. **96**: 339 –360 (2013) (cit. on pp. 55, 56).

160. Lugon Jr, J. & A. J. S. Neto: A hybrid approach with artificial neural networks, Levenberg–Marquardt and simulated annealing methods for the solution of gas–liquid adsorption inverse problems. *Inverse Probl. Sci. Eng.* (2009), vol. **17**: 85 –96 (2009) (cit. on pp. 55, 56).

161. Tu, J. V.: Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *J. Clin. Epidemiol.* (1996), vol. **49**(11): 1225 –1231 (1996) (cit. on pp. 56, 117).

162. Basheer, I. & M Hajmeer: Artificial neural networks: fundamentals, computing, design, and application. *J. Microbiol. Meth.* (2000), vol. **43**(1): 3 –31 (2000) (cit. on pp. 56, 114).

163. Du, X., Q. Yuan, J. Zhao & Y. Li: Comparison of general rate model with a new model – artificial neural network model in describing chromatographic kinetics of solanesol adsorption in packed column by macroporous resins. *J. Chromatogr. A* (2007), vol. **1145**(1–2): 165 –174 (2007) (cit. on pp. 57, 118).

164. Xindu, G. & F. E. Regnier: Retention model for proteins in reversed-phase liquid chromatography. *J. Chromatogr. A* (1984), vol. **296**: 15 –30 (1984) (cit. on pp. 57, 114, 116).

165. Huuk, T. C., T. Briskot, T. Hahn & J. Hubbuch: A versatile noninvasive method for adsorber quantification in batch and column chromatography based on the ionic capacity. *Biotechnol. Progr.* (2016), vol. **32**: 666–677 (2016) (cit. on pp. 59, 73, 98, 119).

166. Grushka, E.: Characterization of exponentially modified Gaussian peaks in chromatography. *Anal. Chem.* (1972), vol. **44**(11): 1733–1738 (1972) (cit. on pp. 61, 122).

167. EvaluatePharma: *World Preview 2015, Outlook to 2020* http://info.evaluategroup.com/rs/607-YGS-364/images/wp15.pdf (cit. on p. 68).

168. Harris, J. M. & R. B. Chess: Effect of PEGylation on pharmaceuticals. *Nat. Rev. Drug Discov.* (2003), vol. **2**(3): 214–221 (2003) (cit. on p. 68).

169. Abuchowski, A, T Van Es, N. Palczuk & F. Davis: Alteration of immunological properties of bovine serum albumin by covalent attachment of polyethylene glycol. *J. Biol. Chem.* (1977), vol. **252**(11): 3578–3581 (1977) (cit. on p. 68).

170. Abuchowski, A., T. V. Es, N. Palczuk & F. Davis: Alteration of immunological properties of bovine serum albumin by covalent attachment of polyethylene glycol. *J. Biol. Chem.* (1977), vol. **252**: 3578–3581 (1977) (cit. on p. 68).

171. Morgenstern, J, P Baumann, C Brunner & J Hubbuch: Effect of PEG molecular weight and PEGylation degree on the physical stability of PEGylated lysozyme. *Int. J. Pharm.* (2017), vol. **519**(1-2): 408–417 (2017) (cit. on p. 68).

172. Fee, C. J. & J. M. V. Alstine: PEG-proteins: Reaction engineering and separation issues. *Chem. Eng. Sci.* (2006), vol. **61**: 934–939 (2006) (cit. on pp. 68, 69, 80, 82–84).

173. Janson, J.-C. Chap. 14 (John Wiley & Sons, 2012) (cit. on pp. 68, 69, 83, 84).

174. Pasut, G & F. M. Veronese: State of the art in PEGylation: the great versatility achieved after forty years of research. *J. Control. Release* (2012), vol. **161**(2): 461–472 (2012) (cit. on pp. 68, 76).

175. Fee, C. J. & J. M. Van Alstine: Prediction of the viscosity radius and the size exclusion chromatography behavior of PEGylated proteins. *Bioconjugate chemistry* (2004), vol. **15**(6): 1304–1313 (2004) (cit. on pp. 69, 80, 82).

176. Seely, J. E. & C. W. Richey: Use of ion-exchange chromatography and hydrophobic interaction chromatography in the preparation and recovery of polyethylene glycol-linked proteins. *J. Chromatogr. A* (2001), vol. **908**(1): 235–241 (2001) (cit. on pp. 69, 76, 83).

177. Yamamoto, S., S. Fujii, N. Yoshimoto & P. Akbarzadehlaleh: Effects of protein conformational changes on separation performance in electrostatic interaction chromatography: Unfolded proteins and PEGylated proteins. *J. Biotechnol.* (2007), vol. **132**(2): 196–201 (2007) (cit. on p. 69).

178. Abe, M, P Akbarzaderaleh, M Hamachi, N Yoshimoto & S Yamamoto: Interaction mechanism of mono-PEGylated proteins in electrostatic interaction chromatography. *Biotechnol. J.* (2010), vol. **5**(5): 477–483 (2010) (cit. on pp. 69, 81, 83, 84).

179. Yoshimoto, N, Y Isakari, D Itoh & S Yamamoto: PEG chain length impacts yield of solid-phase protein PEGylation and efficiency of PEGylated protein separation by ion-exchange chromatography: Insights of mechanistic models. *Biotechnol. J.* (2013), vol. **8**(7): 801–810 (2013) (cit. on pp. 69, 83).

180. Guiochon, G & L. A. Beaver: Separation science is the key to successful biopharmaceuticals. *J. Chromatogr. A* (2011), vol. **1218**(49): 8836–8858 (2011) (cit. on p. 69).

181. Morgenstern, J, M Busch, P Baumann & J Hubbuch: Quantification of PEGylated proteases with varying degree of conjugation in mixtures: An analytical protocol combining protein precipitation and capillary gel electrophoresis. *J. Chromatogr. A* (2016), vol. **1462**: 153–164 (2016) (cit. on pp. 71, 72, 76).

182. Maiser, B, F Dismer & J Hubbuch: Optimization of random PEGylation reactions by means of high throughput screening. *Biotechnol. Bioeng* (2014), vol. **111**(1): 104–114 (2014) (cit. on p. 71).

183. OTTOW, K. E., T LUND-OLESEN, T. L. MAURY, M. F. HANSEN & T. J. HOBLEY: A magnetic adsorbent-based process for semi-continuous PEGylation of proteins. *Biotechnol. J.* (2011), vol. **6**(4): 396–409 (2011) (cit. on p. 71).

184. PERKINELMER: *LabChip GX/GX II - User Manual* http://www.perkinelmer.com/Content/LST_Software_Downloads/LabChip_GX_User_Manual.pdf (cit. on p. 72).

185. BAUMGARTNER, K, L GALM, J NÖTZOLD, H SIGLOCH, J MORGENSTERN, K SCHLEINING, S SUHM, S. A. OELMEIER & J HUBBUCH: Determination of protein phase diagrams by microbatch experiments: Exploring the influence of precipitants and pH. *Int. J. Pharm.* (2015), vol. **479**(1): 28–40 (2015) (cit. on p. 72).

186. RESEARCH, H.: *Lysozyme - User Guide HR7-110* http://hamptonresearch.com/documents/product/hr006812_711020.pdf (cit. on p. 73).

187. VELAYUDHAN, A. & C. HORVATH: Preparative chromatography of proteins - analysis of the multivalent ion-exchange formalism. *J. Chromatogr. A* (1988), vol. **443**: 13 –29 (1988) (cit. on p. 74).

188. JAKOBSSON, N., D. KARLSSON, J. P. AXELSSON, G. ZACCHI & B. NILSSON: Using computer simulation to assist in the robustness analysis of an ion-exchange chromatography step. *J. Chromatogr. A* (2005), vol. **1063**: 99–109 (2005) (cit. on pp. 75, 93).

189. CARTA, G. & A. JUNGBAUER: *Protein Chromatography Process Development and Scale-up* 163–166 (WHILEY-VCH, 2010) (cit. on pp. 75, 84).

190. LENHOFF, A. M.: Protein adsorption and transport in polymer-functionalized ion-exchangers. *J. Chromatogr. A* (2011), vol. **1218**: 8748–8759 (2011) (cit. on pp. 75, 94).

191. WITTEMANN, A., B. HAUPT & M. BALLAUFF: Adsorption of proteins on spherical polyelectrolyte brushes in aqueous solution. *Phys. Chem. Chem. Phys.* (2003), vol. **5**: 1671–1677 (2003) (cit. on pp. 75, 94).

192. MAISER, B, F KRÖNER, F DISMER, G BRENNER-WEISS & J HUBBUCH: Isoform separation and binding site determination of mono-PEGylated lysozyme with pH gradient chromatography. *J. Chromatogr. A* (2012), vol. **1268**: 102–108 (2012) (cit. on p. 76).

193. YOSHIMOTO, N & S YAMAMOTO: PEGylated protein separations: challenges and opportunities. *Biotechnol. J.* (2012), vol. **7**(5): 592–593 (2012) (cit. on p. 76).

194. WANG, G., T. BRISKOT, T. HAHN, P. BAUMANN & J. HUBBUCH: Estimation of adsorption isotherm and mass transfer parameters in protein chromatography using artificial neural networks. *J. Chromatogr. A* (2017), vol. **1487**: 211–217 (2017) (cit. on pp. 77, 81).

195. Mejía-Manzano, L. A., G Sandoval, M. E. Lienqueo, P Moisset, M Rito-Palomares & J. A. Asenjo: Simulation of mono-PEGylated lysozyme separation in heparin affinity chromatography using a general rate model. *Journal of Chemical Technology and Biotechnology* (2017), vol. (2017) (cit. on p. 81).

196. Einstein, A.: On the movement of small particles suspended in stationary liquids required by the molecular-kinetic theory of heat. *Ann. Phys.* (1905), vol. **17**: 549–560 (1905) (cit. on p. 84).

197. Carta, G., M. E. Gregory & D. J. Kirwan: Chromatography with permeable supports: theory and comparison with experiments. *Sep. Technol.* (1992), vol. **2**: 62–72 (1992) (cit. on p. 84).

198. Carta, G. & A. E. Rodrigues: Diffusion and convection in chromatographic processes using permeable particles with a bidisperse pore structure. *Chem. Eng. Sci.* (1993), vol. **48**: 3927–3935 (1993) (cit. on p. 84).

199. Rodrigues, A. E., Z. P. Lu, J. M. Loureiro & G. Carta: Peak resolution in linear chromatography: Effects of intraparticle convection. *J. Chromatogr. A* (1993), vol. **653**: 189–xx198 (1993) (cit. on p. 84).

200. Frey, D. D., E. Schweinheim & C. Horvath: Effect of intraparticle convection on the chromatography of biomacromolecules. *Biotechnol. Progr.* (1993), vol. **9**: 273–284 (1993) (cit. on p. 84).

201. Freitag, R., D. Frey & C. Horvath: Effect of bed compression on high-performance liquid-chromatography columns with gigaporous polymeric packings. *J. Chromatogr. A* (1994), vol. **686**: 165–177 (1994) (cit. on p. 84).

202. Pfeiffer, J. F., J. C. Chen & J. T. Hsu: Permeability of gigaporous particles. *AIChE J.* (1996), vol. **42**: 932–939 (1996) (cit. on p. 84).

203. Gustavsson, P. E., K. Mosbach, K. Nilsson & P. O. Larsson: Superporous agarose as an affinity-chromatography support. *J. Chromatogr. A* (1997), vol. **776**: 197–203 (1997) (cit. on p. 84).

204. Nash, D. C. & H. A. Chase: Comparison of diffusion and diffusion-convection matrices for use in ion-exchange separations of proteins. *J. Chromatogr. A* (1998), vol. **807**: 185–207 (1998) (cit. on p. 84).

205. Lloyd, L. L.: Rigid macroporous copolymers as stationary phases in high-performance liquid chromatography. *J. Chromatogr. A* (1991), vol. **544**: 201–217 (1991) (cit. on p. 84).

206. Dziennik, S. R., E. B. Belcher, G. A. Barker, M. J. DeBergalis, S. E. Fernandez & A. M. Lenhoff: Nondiffusive mechanisms enhance protein uptake rates in ion exchange particles. *PNAS* (2003), vol. **100**: 420–425 (2003) (cit. on p. 84).

207. Van Alstine, J. M. & M Malmsten: Poly (ethylene glycol) Amphiphiles: Surface Behavior of Biotechnical Significance. *Langmuir* (1997), vol. **13**(15): 4044–4053 (1997) (cit. on p. 84).

208. BLASCHKE, T, J VARON, A WERNER & H HASSE: Microcalorimetric study of the adsorption of PEGylated lysozyme on a strong cation exchange resin. *J. Chromatogr. A* (2011), vol. **1218**(29): 4720–4726 (2011) (cit. on p. 85).

209. REICHERT, J. M.: Antibodies to watch in 2017. *mAbs* (2017), vol. **9**: 167–181 (2017) (cit. on p. 90).

210. *Affinity chromatography* (ed FASSINA, G.) (John Wiley & Sons, Ltd, 2001) (cit. on p. 90).

211. HUSE, K., H.-J. BÖHME & G. H. SCHOLZ: Purification of antibodies by affinity chromatography. *J. Biochem. Biophys. Methods* (2002), vol. **51**: 217–231 (2002) (cit. on p. 90).

212. HOBER, S., K. NORD & M. LINHULT: Protein A chromatography for antibody purification. *J. Chromatogr. B* (2007), vol. **848**: 40–47 (2007) (cit. on p. 90).

213. THÖMMES, J. & M. ETZEL: Alternatives to chromatographic separations. *Biotechnol. Prog.* (2007), vol. **23**: 42–45 (2007) (cit. on p. 90).

214. LOW, D., R. O'LEARY & N. S. PUJAR: Future of antibody purification. *J. Chromatogr. B* (2007), vol. **848**: 48–63 (2007) (cit. on p. 90).

215. VENKITESHWARAN, A., P. HEIDER, L. TEYSSEYRE & G. BELFORT: Selective precipitation-assisted recovery of immunoglobulins from bovine serum using controlled fouling crossflow membrane microfiltration. *Biotechnol. Bioeng.* (2008), vol. **101**: 957–966 (2008) (cit. on p. 90).

216. BRODSKY, Y., C. ZHANG, Y. YIGZAW & G. VEDANTHAM: Caprylic acid precipitation method for impurity reduction: An alternative to conventional chromatography for monoclonal antibody purification. *Biotechnol. Bioeng.* (2012), vol. **109**: 2589–2598 (2012) (cit. on p. 90).

217. CAPITO, F., J. BAUER, A. RAPP, C. SCHROTER, H. KOLMAR & B. STANISLAWSKI: Feasibility study of semi-selective protein precipitation with salt-tolerant copolymers for industrial purificiation of therapeutic antibodies. *Biotechnol. Bioeng.* (2013), vol. **110**: 2915–2927 (2013) (cit. on p. 90).

218. SHETH, R. D., M. JIN, B. V. BHUT, Z. LI, W. CHEN & S. M. CRAMER: Affinity precipitation of a monoclonal antibody from an industrial harvest feedstock using an ELP-Z stimuli responsive biopolymer. *Biotechnol. Bioeng.* (2014), vol. **111**: 1595–1603 (2014) (cit. on p. 90).

219. KNEVELMAN, C., J. DAVIES, L. ALLEN & N. J. TITCHENER-HOOKER: High-throughput screening techniques for rapid PEG-based precipitation of IgG4 mAb from clarified cell culture supernatant. *Biotechnol. Prog.* (2010), vol. **26**: 697–705 (2010) (cit. on pp. 91, 92).

220. HAMMERSCHMIDT, N., S. HOBIGER & A. JUNGBAUER: Continuous polyethylene glycol precipitation of recombinant antibodies: Sequential precipitation and resolubilization. *Proc. Biochem.* (2016), vol. **51**: 325–332 (2016) (cit. on pp. 91, 109).

221. Hammerschmidt, N., A. Tscheliessnig, R. Sommer, B. Helk & A. Jungbauer:
      Economics of recombinant antibody production processes at various scales: Industry-
      standard compared to continuous precipitation. *Biotechnol. J.* (2014), vol. **9**(6):
      766–775 (2014) (cit. on p. 91).

222. Warikoo, V. *et al.*: Integrated continuous production of recombinant therapeutic
      proteins. *Biotechnol. Bioeng.* (109), vol. **2012**: 255–296 (109) (cit. on pp. 91, 94).

223. Mahajan, E., A. George & B. Wolk: Improving affinity chromatography resin
      efficiency using semi-continuous chromatography. *J. Chromatogr. A* (2012), vol.
      1227: 154–162 (2012) (cit. on pp. 91, 94).

224. Godawat, R., K. Brower, S. Jain, K. Konstantinov, F. Riske & V. Warikoo:
      Periodic counter-current chromatography - design and operational considerations
      for integrated and continuous purification of proteins. *Biotechnol. J.* (2012), vol. **7**:
      1496–1500 (2012) (cit. on pp. 91, 94).

225. Bisschops, M. & M. Brower: The impact of continuous multicolumn chromato-
      graphy on biomanufacturing efficiency. *Pharm. Bioprocess* (2013), vol. **1**: 361–372
      (2013) (cit. on p. 91).

226. Ng, C. K. S., F. Rousset, E. Valery, D. G. Bracewell & E. Sorensen:
      Design of high productivity sequential multi-column chromatography for antibody
      capture. *Food Bioprod. Process* (2014), vol. **92**: 233–241 (2014) (cit. on p. 91).

227. Angarita, M., T. Müller-Späth, D. Baur, R. Lievrouw, G. Lissens & M.
      Morbidelli: Twin-column capture SMB: A novel cyclic process for protein a affinity
      chromatography. *J. Chromatogr. A* (2015), vol. **1389**: 85–95 (2015) (cit. on p. 91).

228. Brower, M., Y. Hou & D. Pollard: Monoclonal Antibody Continuous Processing
      Enabled by Single Use. *Continuous Processing in Pharmaceutical Manufacturing*
      (2014), vol. (2014) (cit. on p. 91).

229. Green, A. A.: Studies in the physical chemistry of the proteins viii. The solubility
      of hemoglobin in concentrated salt solutions. A study of the salting out of proteins.
      J. Biol. Chem. (1931), vol. **93**(2): 495–516 (1931) (cit. on p. 92).

230. Hämmerling, F., C. L. Effio, S. Andris, J. Kittelmann & J. Hubbuch:
      Investigation and prediction of protein precipitation by polyethylene glycol using
      quantitative structure–activity relationship models. *J. Biotechnol.* (2017), vol. **241**:
      87–97 (2017) (cit. on p. 92).

231. Asakura, S. & F. Oosawa: Interaction between particles suspended in solutions of
      macromolecules. *J. Polym. Sci.* (1958), vol. **33**(126): 183–192 (1958) (cit. on p. 92).

232. Odijk, T.: Depletion theory and the precipitation of protein by polymer. *J. Phys.
      Chem. B* (2009), vol. **113**(12): 3941–3946 (2009) (cit. on p. 92).

233. Bryntesson, M., M. Hall & K. Lacki *Patent* US 7901581 (US) (2011) (cit. on
      pp. 94, 95).

234. Pollock, J., G. Bolton, J. Coffman, S. V. Ho, D. G. Bracewell & S. S. Farid: Optimising the design and operation of semi-continuous affinity chromatography for clinical and commercial manufacture. *J. Chromatogr. A* (2013), vol. **1284**: 17–27 (2013) (cit. on p. 94).

235. Oelmeier, S. A., F. Dismer & J. Hubbuch: Application of an aqueous two-phase systems high-throughput screening method to evaluate mAb HCP separation. *Biotechnol. Bioeng.* (2010), vol. **108**(1): 69–81 (2010) (cit. on p. 96).

236. Shukla, A. A., B. Hubbard, T. Tressel, S. Guhan & D. Low: Downstream processing of monoclonal antibodies application of platform approaches. *J. Chromatogr. B* (2007), vol. **848**(1): 28–39 (2007) (cit. on p. 109).

237. Hammerschmidt, N., B. Hintersteiner, N. Lingg & A. Jungbauer: Continuous precipitation of IgG from CHO cell culture supernatant in a tubular reactor. *Biotechnol. J.* (2015), vol. **10**(8): 1196–1205 (2015) (cit. on pp. 109, 110).

238. Valente, K. N., A. K. Schaefer, H. R. Kempton, A. M. Lenhoff & K. H. Lee: Recovery of Chinese hamster ovary host cell proteins for proteomic analysis. *Biotechnol. J.* (2014), vol. **9**(1): 87–99 (2014) (cit. on p. 109).

239. Sturtevant, J. M., S. A. Rice & E. P. Geiduschek: The stability of the helical DNA molecule in solution. *Discussions of the Faraday Society* (1958), vol. **25**: 138–149 (1958) (cit. on p. 110).

240. Baumann, P., T. Hahn & J. Hubbuch: High-throughput micro-scale cultivations and chromatography modeling: Powerful tools for integrated process development. *Biotechnol. Bioeng.* (2015), vol. **112**: 2123–2133 (2015) (cit. on p. 114).

241. Effio, C. L., T. Hahn, J. Seiler, S. A. Oelmeier, I. Asen, C. Silberer, L. Villain & J. Hubbuch: Modeling and simulation of anion-exchange membrane chromatography for purification of Sf9 insect cell-derived virus-like particles. *J. Chromatogr. A* (2016), vol. **1429**: 142–154 (2016) (cit. on p. 114).

242. Pirrung, S. M., L. A. van der Wielen, R. F. Van Beckhoven, E. J. van de Sandt, M. H. Eppink & M. Ottens: Optimization of biopharmaceutical downstream processes supported by mechanistic models and artificial neural networks. *Biotechnol. Prog.* (2017), vol. (2017) (cit. on p. 114).

243. Iyer, H., S. Tapper, P. Lester, B. Wolk & R. van Reis: Use of the steric mass action model in ion-exchange chromatographic process development. *J. Chromatogr. A* (1999), vol. **832**: 1–9 (1999) (cit. on p. 114).

244. Mollerup, J. M., T. B. Hansen, S. Kidal, L. Sejergaard & A. Staby: Development, modelling, optimisation and scale-up of chromatographic purification of a therapeutic protein. *Fluid Phase Equilibr.* (2007), vol. **261**: 133–139 (2007) (cit. on p. 114).

245. Huuk, T. C., T. Hahn, A. Osberghaus & J. Hubbuch: Model-based integrated optimization and evaluation of a multi-step ion exchange chromatography. *Sep. Purif. Technol.* (2014), vol. **136**: 207–222 (2014) (cit. on p. 114).

246.   Jakobsson, N., M. Degerman & B. Nilsson: Optimisation and robustness analysis of a hydrophobic interaction chromatography step. *J. Chromatogr. A* (2005), vol. **1099**: 157–166 (2005) (cit. on p. 114).

247.   Jakobsson, N., M. Degerman, E. Stenborg & B. Nilsson: Model based robustness analysis of an ion-exchange chromatography step. *J. Chromatogr. A* (2007), vol. **1138**: 109–119 (2007) (cit. on p. 114).

248.   Degerman, M., K. Westerberg & B. Nilsson: Determining critical process parameters and process robustness in preparative chromatography – a model-based approach. *Chem. Eng. Technol.* (2009), vol. **32**: 903–911 (2009) (cit. on p. 114).

249.   Degerman, M., K. Westerberg & B. Nilsson: A model-based approach to determine the design space of preparative chromatography. *Chem. Eng. Technol.* (2009), vol. **32**: 1195–1202 (2009) (cit. on p. 114).

250.   Westerberg, K., E. B. Hansen, M. Degerman, T. B. Hansen & B. Nilsson: Model-based process challenge of an industrial ion-exchange chromatography step. *Chem. Eng. Technol.* (2012), vol. **35**: 183–190 (2012) (cit. on p. 114).

251.   Krogh, A.: What are artificial neural networks? *Nat. Biotechnol.* (2008), vol. **26**: 195–197 (2008) (cit. on p. 114).

252.   Nagrath, D., B. W. Bequette & S. M. Cramer: Evolutionary Operation and Control of Chromatographic Processes. *AlChE J.* (2003), vol. **49**: 82–95 (2003) (cit. on p. 114).

253.   Nagrath, D., A. Messac, B. W. Bequette & S. M. Cramer: A Hybrid Model Framework for the Optimization of Preparative Chromatographic Processes. *Biotechnol. Prog.* (2004), vol. **20**: 162–178 (2004) (cit. on p. 114).

254.   Wang, G. G. & S. Shan: Review of Metamodeling Techniques in Support of Engineering Design Optimization. *J. Mech. Des.* (2006), vol. **129**: 370–380 (2006) (cit. on p. 114).

255.   Caballero, J. A. & I. E. Grossmann: An algorithm for the use of surrogate models in modular flowsheet optimization. *AlChE J.* (2008), vol. **54**: 2633–2650 (2008) (cit. on p. 114).

256.   Bird, R. B., W. E. Stewart & E. N. Lightfoot: *Transport Phenomena* (Wiley, 1962) (cit. on p. 120).

257.   Kataoka, T., H. Yoshida & K. Ueyama: Mass transfer in laminar region between liquid and packing material surface in the packed bed. *J. Chem. Eng. Jpn.* (1972), vol. **5**: 132–136 (1972) (cit. on p. 120).

258.   Tyn, M. T. & T. W. Gusek: Predictions of Diffusion Coefficients of Proteins. *Biotechnol. Bioeng.* (1990), vol. **35**: 327–338 (1990) (cit. on p. 120).

259.   Mackie, J. S. & P. Meares: The Diffusion of Electrolytes in a Cation-Exchange Resin Membrane. I. Theoretical. *Proceedings A* (1955), vol. **232**: 498–509 (1955) (cit. on p. 120).

260. WAKAO, N. & J. M. SMITH: Diffusion in catalyst pellets. *Chem. Eng. Sci.* (1962), vol. **17**: 825–834 (1962) (cit. on p. 120).

261. BRESTRICH, N., T. HAHN & J. HUBBUCH: Application of spectral deconvolution and inverse mechanistic modelling as a tool for root cause investigation in protein chromatography. *J. Chromatogr. A* (2016), vol. **1437**: 158 –167 (2016) (cit. on p. 126).

List of Figures

List of Tables

A. Abbreviations and Symbols

Abbreviations

AC	Affinity Chromatography
AI	Artificial Intelligence
ANN	Artificial Neural Network
ASA	Adaptive Simulated Annealing
BP-ANN	Back-Propagation Artificial Neural Network
BSA	Bovine Serum Albumin
CEX	Cation-Exchange Chromatography
CIP	Cleaning-in-Place
CV	Column Volume
Cyt c	Cytochrome c
DoE	Design-of-Experiments
DSP	Downstream Processing
EMA	European Medicines Agency
EMG	Exponentially-Modified-Gaussian
FDA	Food and Drug Administration
FPLC	Fast Protein Liquid Chromatography
GRM	General Rate Model
HCCF	Harvested Cell Culture Fluid
HCP	Host Cell Protein
HETP	Height of an Equivalent Theoretical Plate
HIC	Hydrophobic Interaction Chromatography
HMW	High Molecular Weight Impurities
HT-CGE	High-Throughput Capillary Gel Electrophoresis
HTE	High-Throughput Experimentation
HTS	High-Throughput Screening
IEX	Ion-Exchange Chromatography

IgG	Immunoglobulin G
LC	Liquid Chromatography
LDF	Linear Driving Force
LGE	Linear Gradient Elution
LHS	Liquid Handling Station
LM	Levenberg-Marquardt
LRM	Lumped Rate Models
Lys	Lysozyme
mAb	Monoclonal Antibody
MCSGP	Multicolumn Counter-Current Solvent Gradient Purification
MLR	Multiple Linear Regression
MSE	Mean Square Error
NRMSE	Normalised Root Mean Square Error
NRMSEP	Normalised Root Mean Square Error of Prediction
PAT	Process Analytical Technologies
PCCC	Periodic Counter-Current Chromatography
PEG	Polyethylene Glycol
pI	Isoelectric Point
RMSE	Root Mean Square Error
RMSEP	Root Mean Square Error of Prediction
QbD	Quality by Design
QSAR	Quantitative Structure Activity Relationship
SEC	Size Exclusion Chromatography
SDM	Stoichiometric Displacement Model
SMA	Steric Mass Action
TDM	Transport-Dispersive Model
UHPLC	Ultra High Performance Liquid Chromatography
USP	Upstream Processing
VLP	Virus-like Particle
3C-PCCC	Three-Column Periodic Counter-Current Chromatography
4C-PCCC	Four-Column Periodic Counter-Current Chromatography

Symbols

a	Constant in activation function	-
a_i	Activity	M
b	Constant in activation function	-
c_i	Bulk phase concentration of species i	M
c_L	Concentration of free ligands	M
$c_{p,i}$	Pore phase concentration of species i	M
$c_{s,i}$	Stationary phase concentration of species i	M
c_{salt}	Salt concentration in mobile phase	M
C	Output of cost function	-
d_c	Column diameter	cm
D_{app}	Apparent dispersion coefficient	$\frac{mm^2}{s}$
$\tilde{D}_{app}$	Apparent dispersion coefficent	$\frac{mm^2}{s}$
D_{ax}	Axial dispersion coefficient	$\frac{mm^2}{s}$
D_p	Pore diffusion coefficient	$\frac{mm^2}{s}$
G	Gibbs free energy	-
G^0	Standard Gibbs free energy	-
h	Ionic hydration number	-
K	Equilibrium constant	-
k_{ads}	Adsorption coefficient	-
k_{des}	Desorption coefficient	-
k_{eff}	Mass transfer coefficient	$\frac{mm}{s}$
K_{eq}	Equilibrium constant of isotherm	-
k_{eq}	Equilibrium coefficient	-
k_{film}	Film diffusion coefficient	$\frac{mm}{s}$
k_{kin}	Kinetic coefficient	-
k_p	Protein specific isotherm constant	$\frac{1}{M}$
k_s	Salt specific isotherm constant	$\frac{1}{M}$
L_c	Column length	cm
$\tilde{m}$	Salt concentration during precipitation	$\% \, w/w$
n	Number of ligands involved in hydrophobic binding	-
$\hat{n}$	Total number of training data sets	-
N	Plate number	-
N_i	Molar flux of species i	$\frac{mol}{s}$

$\tilde{N}_i$	Mole fraction of species i	-
q_i	Stationary phase concentration of species i	M
q_{max}	Saturation capacity of protein binding in chromatography	M
r	Radius	mm
r_p	Particle radius of adsorbent	mm
R	Ideal gas constant	$\frac{J}{kg \cdot K}$
s	Shielding factor in HIC	-
$\tilde{s}$	Solubility of a protein	$\frac{mg}{L}$
S	Area dimension	mm^2
t	Time dimension	s
T	Temperature	K
u_{int}	Interstitial velocity	$\frac{mm}{s}$
u_m	Effective velocity	$\frac{mm}{s}$
$\tilde{u}$	Input signal of activation function	-
x	Space dimension	mm
$\bar{x}$	Mean of x	mm
y	Output signal of activation function	-
$\hat{y}$	Desired output of artificial neural network	-
$\hat{y}'$	Response of artificial neural network	-
$\vec{z}_j$	Bias vector in the layer j	-
$\tilde{\alpha}$	Precipitation constant depending on protein	-
β	Number of bulk-like water molecules involved in adsorption/desorption	-
$\tilde{\beta}$	Precipitation constant depending on salt	-
γ	Activity coefficient	-
δ	Error in an artificial neuron	-
$\vec{\delta}_j$	Error vector in the layer j	-
ε	Column porosity	-
ε_t	Total porosity	-
ε_p	Pariticle porosity	-
Λ	Total ligand density	M
ν	Stoichiometric coefficient	-
μ	Chemical potential	-
μ^0	Standard chemical potential	-
ω	Output signal of an artificial neuron	-

ϕ	Activation function	-
ϕ'	Changing rate of activation function ϕ	-
ξ	Weight	-
$\vec{\xi}_j$	Weight vector in the layer j	-
ζ	Bias	-
σ	Shielding factor in IEX	-