DIGITALIZATION OF INDUSTRIAL DOWNSTREAM PROCESSING

MECHANISTIC AND STRUCTURE-BASED MODELING

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

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

DISSERTATION

von
M.Sc. David Saleh
aus Plankstadt

Tag der mündlichen Prüfung: 04.02.2022
Erstgutachter: Prof. Dr. Jürgen Hubbuch
Zweitgutachter: Prof. Dr. Matthias Franzreb

Danksagung

Zahlreiche Kollegen und Freunde haben durch Ihre Unterstützung zum Erfolg dieser Arbeit beigetragen. Mein besonderer Dank gilt folgenden Personen:

Ich danke Prof. Jürgen Hubbuch für die ausgezeichnete Betreuung, die regelmäßigen Gespräche und die wissenschaftliche Freiheit die mir zur Verfügung gestellt wurde. Zudem danke ich Prof. Matthias Franzreb für die Übernahme des Korreferats.

Mein Dank gilt zudem meinen Kollegen bei Boehringer Ingelheim. Jan Visser danke ich für die tatkräftige Unterstützung von Technologieprojekten in seiner Abteilung. Mein besonderer Dank gilt Joey Studts, Julia Spitz, Stefan Oelmeier und Simon Kluters, die frühzeitig das Potential der mechanistische Modellierung erkannten und eine großartige Arbeitsatmosphäre geschaffen haben. Meinen Langzeit-Modellierungs-Kollegen Gang Wang und Federico Rischawy danke ich für die wissenschaftlichen Diskussionen, die bedingungslose Unterstützung und die unterhaltsamen Kaffeepausen. Tausend Dank gehen an das gesamte DSP Modeling-Team: Dominik Hiltmann, Dara da Silva Weirich, Jan-Phillip Sawall, Martin Leipnitz, Adrian Schimek, Till Briskot, Steffen Großhans und Jan Hendrik-Grosch. Ohne die fleißige Unterstützung meiner Studenten wäre diese Arbeit nicht möglich gewesen, vielen Dank Benedict Müller, Michelle Ahlers-Hesse und Rudger Hess. Natürlich gilt mein Dank allen Kollegen in der Downstream Enticklung, G71, 3 OG. Ihr habt mir nicht nur Unmengen an Protein bereitgestellt und mir zahlreiche Methoden beigebracht, ihr habt mich auch in jeder noch so ausweglosen Situation unterstützt - egal ob im Labor oder beim Skifahren. :)

Bei Tobias Hahn, Thiemo Huuk, Nora Geng, Alex Gutzler und dem gesamten GoSilico Team bedanke ich mich für die großartige Zusammenarbeit und die schnelle Unterstützung bei technischen Fragestellungen. Der gesamten MAB Arbeitsgruppe danke ich für die herzliche Aufnahme bei meinen gelegentlichen Besuchen in Karlsruhe und für die zuverlässige Versorgung mit Kaffee.

Meinen Eltern Claudia und Isam, sowie meinen Geschwistern Leila und Nabil, danke ich für die Unterstützung in sämtlichen Lebenssituationen. Zudem danke ich meiner Freundin Saskia für den Rückhalt, die Liebe und das Verständnis gegenüber meiner Entscheidung diese Arbeit in Biberach zu schreiben.

"Biology enables, Culture forbids."

- Yuval Noah Harari

Kurzfassung

Monoklonale Antikörper (mAbs) und andere biologische Therapien kommen Millionen von Patienten zugute, die unter schwerwiegenden Krankheiten leiden. Das Spektrum der therapeutischen Bereiche, in denen Biologika eingesetzt werden, umfasst die Onkologie, die Hämatologie, Entzündungskrankheiten und neuerdings auch Infektionskrankheiten wie die Coronavirus-Disease 2019 (COVID-19). Die Herstellung und Materialbereitstellung für präklinische und klinische Studien ist ein wichtiger Baustein in der Entwicklung eines therapeutischen Antikörpers. MAbs und komplexe Antikörperformate werden in Zellkulturprozessen, dem sogenannten Upstream Processing (USP), hergestellt. Das anschließende Downstream Processing (DSP) zielt darauf ab, das Zielprotein aus der heterogenen Zellkulturflüssigkeit abzutrennen und zu reinigen. Das DSP von mAbs basiert auf dem Plattformkonzept. Aufgrund der strukturellen Ähnlichkeiten der verschiedenen mAb-Produkte erfolgt deren Aufreinigung in einer standardisierten Abfolge von Prozessschritten mit antikörperspezifischer Anpassung von Prozessparametern. Hier wird häufig die Kationenaustauschchromatographie (CEX) als Polishing-Schritt eingesetzt, da sie in der Lage ist, produktbezogene Verunreinigungen, wie Größen- und Ladungsvarianten des mAb-Produkts, zu entfernen. Die Adsorption von Proteinen an chromatographischen Medien hängt von der Zusammensetzung der mobilen Phase, der Ligandenstruktur und der Struktur des Zielproteins ab. Während die präparative Chromatographie eine einzigartige Selektivität bei der Abreicherung von produkt- und prozessbedingten Verunreinigungen bietet, widerspricht die komplexe und zeitaufwändige Prozessentwicklung der ursprünglichen Idee des Plattformkonzepts. Das Streben nach einer standardisierten Aufreinigung verschiedener Antikörperprodukte wird zusätzlich durch bispezifische und multispezifische Antikörperformate erschwert, die die strukturelle Heterogenität der biopharmazeutischen Entwicklungspipelines erhöhen. Aufgrund der unbekannten Beziehungen zwischen Proteinstruktur und Adsorptionsverhalten stützen sich aktuelle Entwicklungsstrategien für die präparative Chromatographie auf Hochdurchsatz-Experimente (HTE) und statistische Versuchsplanung (DoE).

Miniaturisierte HTE-Methoden ermöglichen die Untersuchung eines großen Parameterraums innerhalb eines kurzen Zeitrahmens, aber ihre Vergleichbarkeit mit dem Produktionsmaßstab ist begrenzt. In den frühen Phasen der DSP-Entwicklung schränkt der ständige Mangel an Zeit und Proteinmaterial den Einsatz experimenteller Methoden weiter ein. In vielen Fällen sind DoE-Studien in Verbindung mit empirischer Response-Surface-Modellierung nicht in der Lage, die hochgradig nichtlinearen Beziehungen in der präparativen Chromatographie zu erfassen. Aufgrund der Vielzahl von Parametern, die sich potenziell auf die Produktqualität auswirken können, werden die in DoE-Studien unter-

suchten Prozessparameter häufig auf der Grundlage von Expertenwissen und einer unzureichenden Datenmenge ausgewählt. Eine falsche Auswahl von Prozessparametern kann zu unnötigen Experimenten führen, die die Prozessentwicklung verzögern, oder schlimmer, zu einem schlecht kontrollierten Herstellungsprozess, der nicht in der Lage ist, eine konstante Produktqualität zu gewährleisten.

Mit der Quality by Design (QbD) Initiative fordern die Zulassungsbehörden ein klares Verständnis der Zusammenhänge zwischen Prozessparametern und Produktqualität. Die U.S. Food and Drug Administration (FDA) und andere Aufsichtsbehörden unterstützen ausdrücklich die Verwendung mathematischer Modelle zur Entwicklung gut verstandener Herstellungsprozesse, die eine robuste Produktqualität und eine effiziente Marktversorgung ermöglichen. In den letzten Jahren wurden computergestützte Methoden auf der Grundlage von Homologiemodellierung, quantitativen Struktur-Eigenschafts-Beziehungen (QSPR), maschinellem Lernen und mechanistischer Chromatographiemodellierung entwickelt, um vielseitige Aufgaben in der biopharmazeutischen Forschung und Entwicklung zu unterstützen. Mechanistische Chromatographiemodelle sind in der Lage, nichtlineare Beziehungen zwischen Prozessparametern und kritischen Qualitätsattributen (CQAs) vorherzusagen. Die Proteinstruktur ist jedoch die eigentliche Ursache für die Funktionalität eines biologischen Arzneimittels. Diese Arbeit zielt darauf ab die Zusammenhänge zwischen der Proteinstruktur und dem makroskopischen Prozessverhalten zu verstehen, um die strukturbasierte Vorhersage von CQAs für eine verbesserte Herstellung von biologischen Arzneimitteln zu ermöglichen.

Die vorliegende Arbeit besteht aus fünf Manuskripten, die sich mit der Erstellung von strukturbasierten und mechanistischen Modellen für die rationalisierte DSP Entwicklung von therapeutischen Antikörpern befassen. Dies erfordert ein verbessertes Verständnis der Beziehungen zwischen der Proteinstruktur und den makroskopischen Parametern der Adsorptionsisotherme. Lernalgorithmen sollen mit einem umfassenden Datensatz trainiert und validiert werden, der strukturelle Deskriptoren und Isothermenparameter von therapeutischen Antikörpern enthält, die für biopharmazeutische Entwicklungspipelines repräsentativ sind. Es sollen effiziente Methoden zur Modellkalibrierung, -validierung und *in silico* Prozesscharakterisierung entwickelt werden, die den QbD-Richtlinien gerecht werden. Die Kombination von Homologiemodellierung, QSPR-Modellierung und mechanistischer Chromatographiemodellierung in einem holistischen *in silico*-Werkzeug soll den Weg von der Aminosäuresequenz des Antikörperkandidaten zu einem robusten Produktionsprozess weisen. Das erste Manuskript dieser Arbeit untersuchte den Einfluss von Aminosäuresubstitutionen in der Complementary Determining Region (CDR) eines IgG1 mAb auf sein Elutionsverhalten in der präparativen CEX Chromatographie. Die Aminosäuresubstitutionen wurden eingeführt, um die biophysikalischen Eigenschaften des mAb zu beeinflussen, indem oberflächenexponierte hydrophobe und geladene Bereiche verändert wurden. Zusätzliche positiv geladene Gruppen in den CDR der leichten Kette (L) und der schweren Kette (H) der mAb-Varianten führten zu einem erhöhten Retentionsvolumen bei der linearen Salzgradientenelution im Vergleich zum ursprünglichen Antikörper. Die Substitution von Tryptophan durch Lysin in der H-CDR3 erhöhte die Ladungsheterogenität des Produkts und führte zu einer signifikanten Erhöhung des Elu-

tionspoolvolumens. Eine multiskalige *in silico*-Analyse, bestehend aus Homologiemodellierung, Proteinoberflächenanalyse und mechanistischer Chromatographiemodellierung, entschlüsselte die qualitativen Zusammenhänge zwischen Struktureigenschaften und Parametern der Steric Mass Action Isotherme (SMA). Die gewonnenen Erkenntnisse über die Bindungsorientierung und die Proteinadsorption an starke CEX-Medien bilden das theoretische Fundament für QSPR-Modelle, die Isothermenparameter auf der Grundlage von Antikörperstrukturinformationen vorhersagen.

Im zweiten Manuskript wurde eine QSPR Modellierungsmethode zur Vorhersage von Stoichiometric Displacement Model (SDM) Parametern von therapeutischen mAbs vorgestellt. Das Modell nutzt Proteindeskriptoren, die aus Homologiemodellen abgeleitet wurden und experimentelle Daten mehrerer Antikörperformate, einschließlich IgG1 mAbs, IgG4 mAbs, Fabs sowie bispezifische Antikörper, um Chromatogramme von zwei mAbs vorherzusagen, die aus dem Trainingsdatensatz entfernt wurden. Die Berücksichtigung von zwei diskreten Konformationen bei der Homologiemodellierung von IgG4 mAbs lieferte eine mögliche Erklärung für Split-Peak-Chromatogramme. Mit Hilfe der Gaußprozess-Regression wurde eine quantitative Beziehung zwischen den Proteindeskriptoren und den makroskopischen Parametern der SDM-Isotherme hergestellt. Durch rekursive Feature-Eliminierung wurden Proteindeskriptoren innerhalb der variablen Region von mAbs identifiziert, die für die Vorhersage der thermodynamischen Gleichgewichtskonstante relevant sind. Im Gegensatz dazu, war der charakteristische Ladungsparameter der SDM-Isotherme hauptsächlich von der Gesamtnettoladung der untersuchten Antikörper abhängig. Die ersten beiden Manuskripte zeigten, wie Homologiemodellierung, QSPRs und mechanistische Modellierung die Frühphasen-Entwicklung für ein neues Biopharmazeutikum unterstützen können, auch ohne anfängliches Prozesswissen und Proteinmaterial für Laborversuche. Die Manuskripte drei, vier und fünf bilden eine Publikationsreihe, die darauf abzielt, die Verwendung der mechanistischen Chromatographiemodellierung als QbD-Werkzeug in der Spätphasen-Entwicklung zu fördern. Daher werden in den folgenden Manuskripten optimierte Methoden zur Modellkalibrierung, -validierung und -anwendung vorgestellt.

Im dritten Manuskript wurde eine Methode für die Kalibrierung von multikomponenten SMA Chromatographiemodellen entwickelt. Die mechanistische Modellierung ist eine vielversprechende Technologie für die digitale Bioprozessentwicklung, aber die komplexe und zeitaufwändige Modellkalibrierung hemmt noch immer ihre Anwendung in der biopharmazeutischen Industrie. Für die *in silico*-Prozesscharakterisierung und andere komplexe DSP-Anwendungen müssen Kalibrierungs- und Validierungstechniken zu einer Modellsicherheit führen, die den Anforderungen des QbD-Konzepts gerecht wird. In dieser Studie wurde eine pH-abhängige, multikomponenten SMA-Isotherme verwendet, um einen CEX-Chromatographieprozess zu modellieren, der drei mAb-Ladungsvarianten sowie eine Aggregatspezies beinhaltet. Die Modellkalibrierungsmethode basierte auf der systematischen Reduktion unbekannter Modellparameter durch Anwendung grundlegender Kenntnisse über präparative Chromatographie in Kombination mit der inversen Schätzung von Modellparametern unter Verwendung repräsentativer Experimente. Die Parameter, die den linearen Bereich der SMA-Isotherme definieren, wurden anhand einer Reihe von linearen Gradientenelutionsversuchen ohne Fraktionssammlung bestimmt, was den ana-

lytischen Aufwand für die Quantifizierung der Ladungs- und Größenvarianten drastisch reduzierte. Außerdem konnten mit dieser Methode lokale Minima bei der heuristischen Schätzung der übrigen Modellparameter vermieden werden. Die Anreicherung der Aggregatspezies im Ausgangsmaterial reduzierte die Modellunsicherheit für diese niedrig konzentrierte Verunreinigung. Die Modellvalidierung wurde unter Prozessbedingungen durchgeführt, die außerhalb der vorgesehenen Parameterbereiche des CEX-Prozesses lagen. Mit dieser Arbeit wurde eine standardisierte Methode zur Kalibrierung von mechanistischen Chromatographiemodellen eingeführt, die in einem industriellen Umfeld eingesetzt werden kann.

Als Alternative zu experimentellen Scale-Down Modellen (ScDM) wurde im vierten Manuskript das zuvor vorgestellte mechanistische Chromatographiemodell als digitale Repräsentation des Prozesses im Produktionsmaßstab validiert. Experimentelle ScDMs von Chromatographieprozessen ermöglichen eine wirtschaftliche Prozesscharakterisierung und Ursachenforschung im Labormaßstab. Die Vergleichbarkeit zwischen ScDM Säulen und größeren Maßstäben hängt jedoch von systemspezifischen Dispersionseffekten, der Variabilität der Ligandendichte sowie der Variabilität in der Zusammensetzung des Feed-Materials und der Beladungsdichte ab. Darüber hinaus verlangen die Aufsichtsbehörden, dass mathematische Modelle die Auswirkungen der Prozessvariabilität erfassen, die bei der Herstellung im Großmaßstab zu erwarten sind, wenn das Modell zur Festlegung einer Kontrollstrategie für den kommerziellen Herstellungsprozess verwendet wird. Der Vergleich zwischen simulierten und gemessenen Chromatogrammen und Elutionspooldaten vom Labor- bis zum Produktionsmaßstab ermöglichte die frühzeitige Identifizierung von Unterschieden zwischen den Maßstäben, z. B. Systemdispersionseffekte oder Variabilität der Ionenkapazität. Es wurde eine mehrstufige Modellvalidierungsmethode eingeführt, um die Modellqualität zu messen und die Grenzen des Modells in verschiedenen Maßstäben zu verstehen. Das experimentelle ScDM und das *in silico*-Modell wurden mit Hilfe des identischen statistischen Äquivalenztestverfahrens als repräsentative Darstellung des Produktionsmaßstabs validiert. Das mechanistische Chromatographiemodell umging die Limitierungen des experimentellen ScDM, indem es die Auswirkungen von Betthöhe, Beladungsdichte, Feed-Zusammensetzung und Eigenschaften der mobilen Phase erfasste. Die Ergebnisse zeigen die Anwendbarkeit mechanistischer Chromatographiemodelle als mögliche Alternative zu konventionellen ScDM-Ansätzen und ermöglichen ihre Verwendung für komplexe Aufgaben in der Spätphasen-Entwicklung.

Das fünfte und letzte Manuskript demonstriert die Anwendung des zuvor veröffentlichten mechanistischen Chromatographiemodells auf die Prozesscharakterisierung (PCS) eines Aufreinigungsschritts. Studien zur Prozesscharakterisierung stellen die umfangreichsten und zeitaufwändigsten Arbeitspakete während der DSP-Entwicklung eines mAbs dar. Im Allgemeinen besteht das Ziel der PCS in der Identifizierung von Korrelationen zwischen Prozessparametern und CQAs, was die Etablierung einer robusten Prozesskontrollstrategie ermöglichen soll. Aufgrund der Komplexität der präparativen Chromatographie und einer Vielzahl von potenziell kritischen Prozessparametern erfordert eine traditionelle PCS auf der Grundlage statistischer DoEs Dutzende von Laborexperimenten sowie zeitintensive Offline-Messungen. Die in dieser Arbeit vorgestellte Modellierungsmethode deckt die

Hauptaufgaben traditioneller PCS-Studien nach den QbD-Prinzipien ab, einschließlich der Bewertung der Kritikalität von 11 Prozessparametern und der Festlegung ihrer Kontrollbereiche. Die Analyse der Auswirkungen eines multivariaten Samplings von Prozessparametern auf das Aufreinigungsergebnis ermöglichte die Identifizierung der Edge-of-Failure. Die experimentelle Validierung der *in silico*-Ergebnisse erforderte etwa 75% weniger Experimente im Vergleich zu einer rein auf Laborexperimenten basierenden PCS. Monte-Carlo-Simulationen wurden unter Berücksichtigung der gemessenen Varianzen der Prozessparameter und der Zusammensetzung des Feed-Materials im Produktionsmaßstab eingesetzt, um die Fähigkeit des Prozesses abzuschätzen, die Akzeptanzkriterien für CQAs und Prozessausbeute zu erfüllen. Der hier vorgestellte Arbeitsablauf ermöglicht die Implementierung digitaler Zwillinge als QbD-Werkzeug für eine verbesserte Entwicklung biopharmazeutischer Herstellungsprozesse.

In der vorliegenden Arbeit wurden mehrere Hindernisse auf dem Weg von der Primärstruktur zur Etablierung eines robusten Downstream-Prozesses beseitigt. Die multiskalige Modellierung mehrerer Biologika in der CEX-Chromatographie führte zu einem tiefen Verständnis der zugrundeliegenden Adsorptionsmechanismen. Die vorgestellten QSPR-Modelle zur Vorhersage von SDM-Isothermen Parametern ermöglichten einen frühen Start der Prozessentwicklung, bevor Proteinmaterial für Laborexperimente zur Verfügung steht. Um die Lücke zwischen der Frühphasen- und Spätphasen-Entwicklung zu schließen, können erste Chromatographiemodelle, die auf Proteinstrukturinformationen aufbauen, mit Hilfe experimenteller Daten weiter verfeinert werden. Im Kontext der QbD-Richtlinien tragen standardisierte und wissenschaftlich fundierte Methoden zur Modellkalibrierung, Validierung und *in silico*-Prozesscharakterisierung zu einer effizienteren und wirtschaftlicheren DSP-Entwicklung bei und erhöhen gleichzeitig die Prozessrobustheit und Produktqualität. Die in dieser Arbeit vorgestellten Werkzeuge haben das Potenzial, die Akzeptanz gegenüber der mechanistischen Modellierung in der Industrie und bei den Aufsichtsbehörden zu erhöhen.

Abstract

Monoclonal antibodies (mAbs) and other biological therapies benefit millions of patients fighting severe conditions. The spectrum of therapeutic areas in which biologics can be used includes oncology, hematology, inflammatory diseases, and more recently, infectious diseases such as corona virus disease 2019 (COVID-19). Manufacturing and material supply for pre-clinical and clinical trials forms a major building block in the development of a therapeutic antibody. MAbs and complex antibody formats are produced in cell culture processes, the so-called upstream processing (USP). Subsequently, the downstream processing (DSP) aims to capture and purify the target protein from the highly heterogeneous cell culture fluid. The DSP of mAbs is based on the platform concept. Due to the structural similarities of different mAb products their downstream processing is performed in a standardized sequence of unit operations with antibody-specific adaption of process conditions. Here, cation exchange (CEX) chromatography is frequently employed as a polishing step due to its ability to remove product related impurities, such as size- and charge variants of the mAb product. Protein adsorption to chromatographic resins depends on mobile phase conditions, the ligand structure, and the structure of the target protein. While preparative chromatography offers an unparalleled selectivity towards the removal of product- and process-related impurities, the complex and time-consuming process development contradicts the original idea of the platform concept. The pursuit of a standardized purification of different antibody products is further complicated by bispecific and multispecific antibody formats that increase the structural heterogeneity of biopharmaceutical development pipelines. Due to the unknown relationships between protein structure and adsorption behavior, development strategies for preparative chromatography must rely on high-throughput experimentation (HTE) and statistical design of experiments (DoE).

Miniaturized HTE methods enable the investigation of a large parameter space within a short time frame, but their comparability to manufacturing-scale is limited. During the early phases of DSP development, constant shortage of time and protein material further constrains the use of experimental methods. In many cases, wet-lab DoE studies coupled with empirical response surface modeling are not able to capture the highly nonlinear relationships encountered in preparative chromatography. Due to the multitude of parameters that could potentially affect product quality, the process parameters screened in DoE studies are often selected based on subject matter expertise. Incorrect selection of process parameters could lead to unnecessary experimentation delaying process development, or even worse, to a poorly understood manufacturing process incapable of delivering a constant product quality.

With the Quality by Design (QbD) initiative, regulatory authorities have been demanding a clear understanding of correlations between process parameters and product quality. The U.S. Food and Drug Administration (FDA) and other regulatory agencies specifically encourage the use of mathematical models to develop well-understood manufacturing processes that enable robust product quality and efficient market supply. During the last years, computational tools based on homology modeling, quantitative structure-property relationships (QSPR), machine learning, and mechanistic chromatography modeling were developed to support various tasks in biopharmaceutical research and development. Mechanistic chromatography models are capable of predicting non-linear relationships between process parameters and critical quality attributes (CQAs). However, the protein structure is the ultimate root cause for the functionality of a biological drug. This work intends to close the gap between protein structure and macroscopic process behavior to enable the structure-based prediction of CQAs for an improved manufacturing of biological drugs.

The present work consists of five manuscripts that focus on the development of structure-based and mechanistic models for the rationalized DSP of therapeutic antibodies. This demands an increased understanding of the relationships between protein structure and macroscopic adsorption isotherm parameters. Machine learning models should be trained and tested with a data set that includes structural descriptors and adsorption isotherm parameters of a diverse set of therapeutic antibodies representative for biopharmaceutical development pipelines. Efficient model calibration, validation, and process characterization methods for mechanistic models should be developed that cope with the QbD guidelines. A combination of homology modeling, QSPR modeling, and mechanistic chromatography modeling into a common multiscale *in silico* framework should guide the path from the amino acid sequence of the antibody candidate to a robust manufacturing process. The first manuscript of this thesis elucidated the influence of amino acid substitutions in the complementarity-determining region (CDR) of a full-length IgG1 mAb on its elution behavior in preparative CEX chromatography. The amino acid substitutions were introduced to affect biophysical properties of the mAb by modifying surface-exposed hydrophobic and charged patches. Additional positively charged groups in the light chain (L) and heavy chain (H) CDR of mAb variants resulted in an increased retention volume in linear salt gradient elution compared to the original antibody. Substitution of tryptophan with lysine in the H-CDR3 increased charge heterogeneity of the product leading to a significant increase of elution pool volume. A multiscale *in silico* analysis, consisting of homology modeling, protein surface analysis, and mechanistic chromatography modeling revealed qualitative relationships between structural descriptors and macroscopic parameters of the steric mass-action (SMA) adsorption isotherm. The insights gained into binding orientation and protein adsorption to strong CEX media provide the theoretical basis for QSPR models that predict isotherm model parameters based on antibody structure information.

The second manuscript introduced a QSPR modeling method for prediction of stoichiometric displacement model (SDM) isotherm parameters of therapeutic mAbs. The model leverages protein descriptors derived from homology models and experimental data of multiple antibody formats, including IgG1 mAbs, IgG4 mAbs, Fabs, as well as bispecific

antibodies, to predict chromatograms of two mAbs that were removed from the training data set. Consideration of two discrete conformations during homology modeling of IgG4 mAbs gave a possible explanation for split-peak chromatograms. Gaussian process regression was used to build quantitative relationships between protein descriptors and macroscopic SDM isotherm parameters. Recursive feature elimination identified protein descriptors within the variable region of mAbs relevant for prediction of the thermodynamic equilibrium constant. In contrast, the characteristic charge parameter of the SDM isotherm was mainly depending on the overall net charge of investigated antibodies. The first two manuscripts showed how homology modeling, QSPRs, and mechanistic modeling can support early-stage development of a new biological entity without initial process knowledge and protein material for wet-lab experiments. Manuscripts three, four, and five form a publication series, which aims to foster the adoption of mechanistic chromatography modeling as QbD-tool in late-stage biopharmaceutical development. Therefore, the manuscripts introduce straightforward methods for model calibration, validation, and application.

In the third manuscript, a novel method for the calibration of multicomponent SMA chromatography models was developed. While mechanistic modeling is a promising technology for digital bioprocess development, the complex and time-consuming model calibration still inhibits its application in the biopharmaceutical industry. For *in silico* process characterization and other demanding late-stage DSP applications, calibration and validation techniques must result in a model certainty that meets the requirements of the QbD concept. In this study, a multicomponent, pH-dependent SMA isotherm was used to model a CEX chromatography process including three mAb charge variant, as well a high molecular weight species. The model calibration method was based on the systematic reduction of unknown model parameters by applying fundamental knowledge on preparative chromatography in combination with the inverse estimation of model parameters using a representative set of wet-lab experiments. Parameters defining the linear region of the SMA isotherm were estimated using a set of linear gradient elution experiments without fraction collection, which drastically reduced the analytical efforts for quantification of mAb charge- and size-variants. Further, this methodology avoided local minima during the heuristic estimation of the remaining model parameters. Enrichment of the aggregate species in the loading material reduced model uncertainty for this low-concentrated impurity. Model validation was performed at laboratory-scale at process conditions beyond the intended operating ranges of the purification process. This work introduced a standardized method for calibration of mechanistic chromatography models that can be used economically and efficiently in an industrial setting.

As an alternative to experimental scale down models (ScDM), the fourth manuscript validated the previously presented mechanistic chromatography model as a digital representation of the large-scale manufacturing process. Experimental ScDMs of chromatography processes enable an economic process characterization and root cause investigation at laboratory-scale. However, the comparability between ScDM columns and larger scales depends on system-specific dispersion effects, resin variability, as well as variability in feed composition and loading density. Additionally, regulatory authorities require that

mathematical models capture the effects of process variability anticipated at large-scale manufacturing if the model is used to establish a control strategy for the commercial manufacturing process. Comparison between simulated and measured chromatograms and elution pool data ranging from laboratory- to manufacturing-scale allowed early identification of differences between scales, e.g. system dispersion effects or ionic capacity variability. A multi-stage model validation approach was introduced to measure the model quality and to understand the model's limitations across scales. The experimental ScDM and the *in silico* model were qualified against large-scale data using the identical statistical equivalence testing procedure. The mechanistic chromatography model avoided limitations of the ScDM by capturing effects of bed height, loading density, feed composition, and mobile phase properties. The results demonstrate the applicability of mechanistic chromatography models as a possible alternative to conventional ScDM approaches enabling their application to advanced tasks in late-stage DSP development.

The fifth and final manuscript demonstrates the application of the previously published mechanistic chromatography model to the *in silico* process characterization (PCS) of a monoclonal antibody polishing step. Process characterization studies represent the most comprehensive and time-consuming work-packages during the downstream process development of a mAb. In general, the aim of a PCS is the identification of correlations between process parameters and product quality enabling the definition of a robust process control strategy. Due to the complexity of preparative chromatography and a plethora of potentially critical process parameters, a traditional PCS based on statistical DoEs requires dozens of wet-lab experiments as well as off-line analytical measurements. The modeling workflow presented in this study covered the main tasks of traditional PCS studies following the QbD principles, including criticality assessment of 11 process parameters and establishment of their proven acceptable ranges (PARs) of operation. Analyzing effects of multi-variate sampling of process parameters on the purification outcome allowed identification of the edge-of-failure. Experimental validation of *in silico* results demanded approximately 75% less experiments compared to a purely wet-lab based process characterization study. Monte-Carlo simulation, considering the measured variances of process parameters and loading material composition at manufacturing-scale, was used to estimate the capability of the process to meet the acceptance criteria for critical quality attributes and key performance indicators. The proposed workflow enables the implementation of digital process twins as QbD tool for improved development of biopharmaceutical manufacturing processes.

The present thesis removed several roadblocks on the way from the primary structure to establishing a robust downstream process. Multiscale modeling of a diverse set of biologics in CEX chromatography led to a deep understanding of the underlying adsorption mechanisms. The presented QSPR models for prediction of SMA adsorption isotherm parameters enabled an early start of process development before protein material for wet-lab experiments is available. To close the gap between early- and late-stage development, initial chromatography models built on protein structure information can be further refined using experimental data that is collected during the product life cycle. In the light of the QbD concept, standardized and scientifically sound methods for model calibration,

validation, and *in silico* process characterization contribute to a more efficient and economic DSP development, while increasing process robustness and product quality. The tools introduced in this thesis have the potential to increase the acceptance of mechanistic modeling by industry and regulatory agencies.

Contents

1 Introduction

The biopharmaceutical industry is constantly developing novel treatment options for patients fighting severe conditions. The spectrum of biological therapies includes recombinant proteins, such as insulin, monoclonal antibodies (mAbs), and fusion proteins, as well as advanced therapy medicinal products (ATMP), such as gene therapies or oncolytic viruses. Biologics represent a key growth factor for the pharmaceutical industry. From 1999 to 2009, an average of four biologics per year reached market entry. Only a decade later, in 2019 alone, 20 of 53 drugs approved by the US Food and Drug Administration (FDA) were biologics, including nine monoclonal antibodies (mAbs) or antibody-drug conjugates (ADCs) [1]. Researchers harness the specificity on mAbs to address multiple disease targets, including oncology, hematology, and inflammatory diseases. This broad applicability to multiple therapeutic areas made mAbs a highly profitable class of therapeutics. The global mAb market exceeded US$98 billion in sales in 2017 and is expected to reach US$130-200 billion by 2022 [2]. At the current time, the number of mAbs investigated in clinical studies is still rising [3, 4].

Manufacturing and material supply for pre-clinical and clinical trials form a major building block in biopharmaceutical product development and is often referred to as Chemistry, Manufacturing and Controls (CMC) [5]. MAbs and complex biologics are produced in cell culture processes. After harvesting the cell culture, the subsequent downstream processing (DSP) transforms the heterogeneous cell culture fluid into a concentrated and pure drug product [6]. Due to similar structural characteristics and process behavior of different mAb candidates, large-scale purification is based on the so-called platform process [7, 8]. Here, a standardized sequence of orthogonal separation mechanisms is applied to multiple mAbs and adaptation of process conditions is reduced to a minimum. Common DSP platform processes include 2-3 chromatographic unit operations. While the platform concepts enables fast and efficient process development, the increasing structural diversity of the biopharmaceutical product portfolio challenges the existing development workflows. Especially protein adsorption to chromatography resins is affected by the protein structure, which can complicate manufacturing of novel antibody formats under standardized process conditions. Besides technical challenges, the costs demanded to bring a new biological entity to the market drastically increased during the last years [5]. Adjusted for inflation, costs per market entry increased from US$1.6 billion US dollar in 2007 [9] to US$3.1 billion in 2020 [10]. In parallel, the average Phase I to approval success rate dropped from 30% to 12%. Here, CMC costs account for 13% to 17% of the total R&D budget from pre-clinical development to regulatory approval [5].

The above mentioned challenges led to the development of novel technologies for improved

and more economic bioprocesses. For the optimization and characterization of chromatography processes, industry relies on high-throughput experimentation [11, 12] and design of experiments (DoE) coupled with empirical response surface modeling [13, 14]. Miniaturization and automation *via* robotic liquid-handling stations enables parallel screening of multiple process conditions demanding comparably low quantities of protein material [15, 16]. Multiple companies are currently exploring continuous bioprocessing and process intensification for improved equipment utilization rates, reduced facility footprints, accelerated cycle-times, lower production costs, and lower investment costs [17–20]. Single-use technologies, including disposable filtration and chromatography devices, have the potential to save cleaning and validation costs, as well as buffer consumption, and facility space [21].

Independent of new manufacturing technologies, biopharmaceutical industry is currently adopting model-based tools to address challenges caused by the increased number of mAbs entering CMC development [22, 23]. With the Quality by Design (QbD) initiative [24, 25], regulatory authorities support the application of mathematical modeling to achieve high product quality by understanding the fundamental correlations between process parameters and quality attributes. For the downstream processing of biologics, mechanistic chromatography modeling represents a versatile technology for development tasks ranging from *in silico* process optimization to model-guided process control [26, 27]. In the light of the global ongoing pandemic of coronavirus disease 2019, Cardillo *et al.* highlighted that *in silico* tools have the potential to cut vaccine CMC development timelines in half [28]. While mechanistic modeling is a powerful technology to support the development of single unit operations, novel *in silico* tools must enable transfer of process knowledge from existing mAbs to new products entering the development phase [29]. In this context, modeling techniques based on artificial intelligence (AI) are currently limited by the vast amount of data that is needed to train artificial neural networks or other supervised learning algorithms. During the last years, quantitative structure-property relationship (QSPR) models have been developed to predict product stability or process behavior of mAbs based on protein structure models or the amino acid sequence [30]. QSPR models leverage machine learning algorithms to correlate structural characteristics of the target molecules with a measurable property, e.g. elution behavior in preparative chromatography [30, 31]. The following sections will provide a fundamental overview on therapeutic antibodies and protein structure prediction of mAbs. Furthermore, a theoretical introduction to mechanistic modeling of preparative chromatography and QSPR modeling will be given.

1.1 Monoclonal Antibodies

In order to describe protein adsorption to chromatographic resins, it is helpful to understand the structural characteristics of the target proteins. In the present thesis, all investigated molecules are mAbs or complex antibody formats. Immunoglobulins (Ig) are divided in the classes IgA, IgD, IgE, IgG, and IgM. The ability of IgG to interact

with the neonatal F_c receptor located on endothelial cells inhibits their catabolism, which leads to comparably long biological half-life times of >20 days [32, 33]. Hence, IgG form the framework for the development of therapeutic antibodies [34]. Fig. 1.1 (A) shows a schematic representation of an IgG1 mAb. The roughly Y-shaped mAbs have a molecular weight of approximately 150 kDa. MAbs consist of four peptide chains, two identical heavy chains (HC) and the two light (LC) chains, which are interconnected *via* disulfid bonds within the hinge region. Proteolytic digestion of antibodies results in different fragments, the fragment variable (F_v), the fragment antigen binding (F_{ab}), and the fragment crytallizable (F_c) [35, 36]. The first 110 aminoacids of both chains form the variable regions (V_H and V_L) [34]. Fig. 1.1 (B) depicts the conformational structure of a therapeutic F_{ab} including the F_v region (Idarucizumab, [37]). Each variable chain contains three hypervariable peptide stretches, the complementarity-determining regions (CDRs) responsible for antigen-binding. Approximately 5% of the primary structure of human IgG are responsible for antigen specificity, while the remaining 95% are conserved [38].

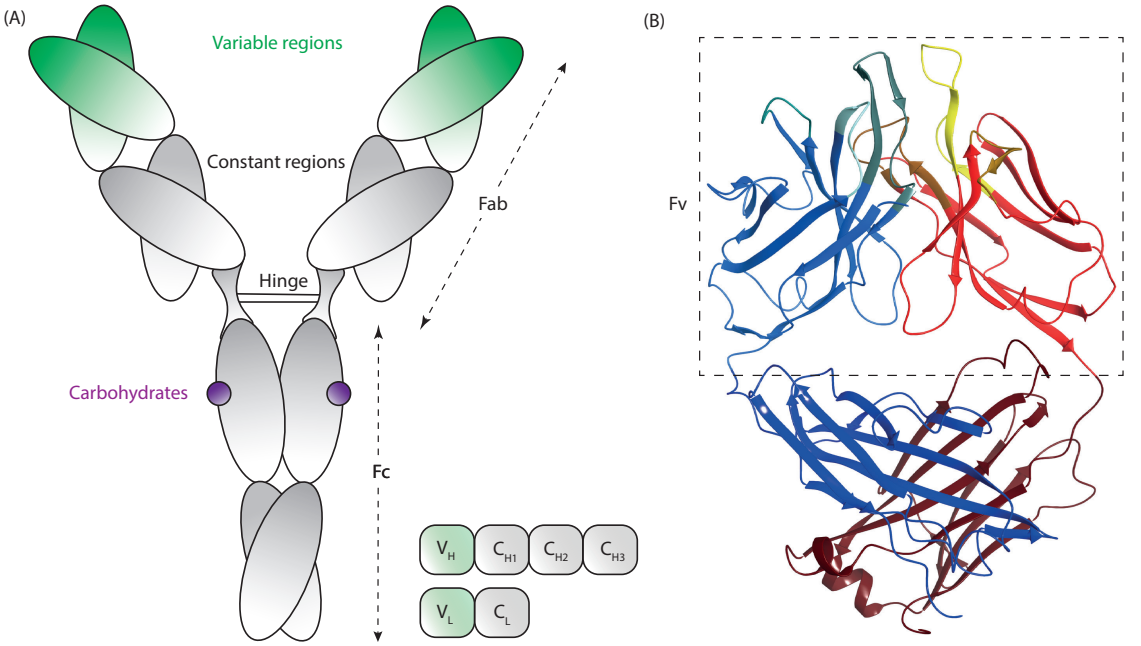

Figure 1.1: (A): Schematic representation of an IgG1 monoclonal antibody structure. (B): Conformational structure of the F_{ab} antibody derivative Idarucizumab, reversal agent for Dabigatran [37], PDB=4YGV. The heavy chain is colored in blue and the light chain in red, respectively. CDRs are depicted in brighter colors and are located in the upper part of (B).

Depending on their abundance in the human serum, IgGs are further divided in four subclasses (IgG1, IgG2, IgG3, and IgG4) with structurally differing F_c regions [39]. Interaction between the constant F_c region and different $F_c\gamma$ receptors elicits antibody-dependent cellular cytotoxicity (ADCC) and complement-dependent cytotoxicity (CDC) [40]. The impact of structural differences between F_c regions of IgG subclasses on effector functions

determines their selection as structural backbone for a potential therapeutic antibody [41, 42]. With 79%, IgG1 mAbs represent the majority of regulatory approved antibody therapeutics [2]. Compared to other subclasses, IgG1 shows the most efficient interaction with the immune system's effector functions. IgG4 mAbs are often selected as structural backbone when immune effector functions are undesired [43], e.g. for receptor inhibition without cell depletion [44].

Figure 1.2 shows schematic representations of therapeutic antibody formats that are relevant for the present thesis. While the majority of approved therapeutic antibodies are IgG1, IgG4, and IgG2 mAbs, novel and more complex antibody derivatives are currently entering the market and clinical pipelines of biopharmaceutical organizations. Multispecific antibody formats represent a promising opportunity for the treatment of cancer [45–47]. Bispecific T-cell engagers that bind a surface target antigen on a cancer cell and a T-cell receptor can redirect cell lysis towards cancer cells. [48]. Figure 1.2 depicts two different bispecific antibody formats. Knob-into-hole formats contain complementary mutations within the the C_{H3} domain that improve generation of heavy-chain hetero-dimers [49]. In contrast, the bispecificity of the IgG-scFv format does not result from heterodimerization of the heavy chain. For IgG-scFv antibodies, two F_v fragments are connected *via* flexible linker peptides to the F_c region of an IgG mAb [50]. While bispecific antibodies show promising results in clinical investigations, their structural complexity can lead to an increased number of product related impuritites that need to be removed during the manufacturing process [51, 52].

The function of proteins, including mAbs, is defined by their three-dimensional structure [53]. Therefore, a full-grown research field has formed around the prediction of protein structures based on the amino acid sequence [54–57]. In recent years, deep learning led to important advances in the field of protein structure prediction. During the CASP14 (Critical Assessment of Structure Prediction) challenge 2020, Deepmind's AI-based AlphaFold2 system [54, 58] outperformed other tools of the structure prediction community, setting new benchmarks for prediction accuracy for the majority of the target structures [59]. Due to the size of mAbs and the abundance of template structures in the Protein Data Bank (PDB), structure prediction of antibodies is mostly conducted *via* homology modeling. Homology modeling describes the construction of a protein structure model based on the amino acid sequence of the target and an experimentally determined template structure that is selected from the PDB based on sequence similarity to the target protein [60]. From a theoretical perspective, homology modeling is based on the following observations [60]:

1. The protein structure is uniquely determined by its amino acid sequence. Therefore, the sequence should enable structure predictions [60, 61].

2. From an evolutionary perspective, the structure changes much slower than the respective amino acid sequence. Hence, distantly related sequences fold into similar structures [60, 62, 63].

With currently over 4000 mAb F_v structures within the PDB and other data bases [64, 65], homology modeling is the ideal method to predict antibody structures supporting discovery, design, and development of therapeutic mAbs [66–68]. The prediction of the hyper-variable CDR loops represents the most challenging part of antibody homology modeling. Recent studies propose loop prediction methods based on a combination of sequence similarity, geometry matching, and the clustering of database structures [66, 69]. The resulting antibody structure models give atomic-level insights into the antibody-antigen interaction enabling model-guided improvement of antibody-affinity [70–72]. Further, structural descriptors derived from antibody homology models can be used to build mathematical models that support the development of a stable product formulation [67, 73] or the prediction of process behavior in preparative chromatography [30, 74].

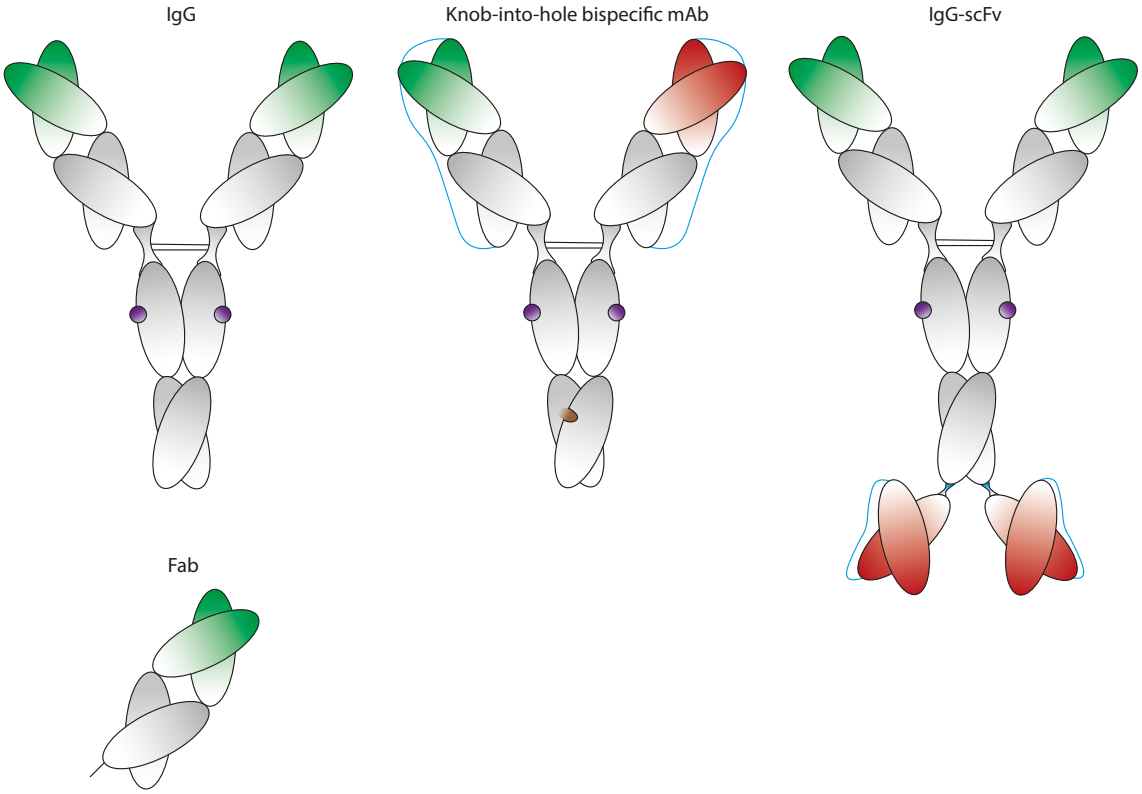

Figure 1.2: Schematic representation of therapeutic antibody formats investigated in this thesis. Variable regions are colored green, or red/green for bispecific antibody formats. Linker sequences within bispecific formats are depicted in blue and the knob-into-hole region within the bispecific F_c is colored brown. Carbohydrates are colored purple.

1.2 Preparative Chromatography for Antibody Purification

The goal of biopharmaceutical downstream processing is the separation of a target protein from a heterogeneous mixture of impurities. Depletion of process- and product related impurities is crucial for the safety and efficacy of biological drugs [75]. Due to its unique selectivity and variety of separation mechanisms, adsorption chromatography is the workhorse for the preparative purification of biologics. In general, liquid-solid chromatography systems consist of a liquid mobile phase that is transported through a solid stationary phase. The components in the mobile phase interact with the stationary phase to different degrees, which leads to the separation of the components [76, 77]. Multiple adsorption mechanisms are exploited for the separation of biologics, including affinity chromatography (AC), ion-exchange chromatography (IEX), and hydrophobic interaction chromatography (HIC). The so-called mixed-mode chromatography (MMC) is based on a combination of at least two different interaction mechanisms between the stationary phase and the components in the mobile phase [78].

In AC, proteins or polypeptides are separated based on a highly specific and reversible interaction with a ligand [79]. For conventional IgG formats, Protein A affinity chromatography is commonly employed as initial capture step [6, 7, 80]. The Protein A ligand is immobilized on the stationary phase. Protein A originates from *Staphylococcus aureus* and displays a high affinity towards the F_c region of IgG at neutral pH [81]. Thus, the harvested cell culture fluid containing the mAb product is loaded on the Protein A column at a mobile phase pH of 6-8. While the antibodies are bound to the resin, process related impurities, such as host cell protein (HCP), DNA, endotoxins, and cell culture media are removed in the flow-through and wash step. Optimized wash steps have the potential to disrupt HCP-mAb interactions, leading to an improved HCP depletion while the target mAb is bound to the Protein A ligand [82, 83]. A low pH-shift of the mobile phase to pH 2.5-4 [7, 84] leads to repulsive effects between the ligand and mAb F_c region, which induces the elution.

HIC and IEX chromatography separate molecules based on electrostatic and hydrophobic interaction with the chromatographic ligands, respectively. IEX, HIC, and MMC are often applied as the so-called polishing chromatography after the initial capture step *via* Protein A chromatography [6, 7]. HIC enables efficient removal of mAb size variants, including high molecular weight species (HMW) and other polar impurities [85]. Adsorption of proteins to HIC ligands is mediated by the reorganization of the water structure on hydrophobic surfaces of the ligand and protein [86, 87] . Therefore, protein binding to HIC resins is performed at high salt concentrations. This often results in comparably high salt concentrations in the elution pool that can complicate the integration of HIC in a process sequence with CEX or AEX chromatography. AEX resins have positively charged ligands bound to their surface [88]. For the purification of mAbs with an isoelectric point (pI) above 7.5, AEX and multi-modal HIC-AEX is frequently performed in flow-through or weak partitioning mode. In flow-through mode, the target protein does not interact with the resin due to the positive protein net charge at a mobile phase pH

below the pI. In contrast to the product, negatively charged impurities, including DNA, HCP, and leached Protein A, bind to the resin and are removed during column regeneration [6, 89]. Weak partitioning chromatography describes an isocratic separation method at mobile phase conditions where the target protein shows a stronger interaction with the AEX resin compared to flow-through mode. Consequently both, acidic and basic impurities can be depleted during weak partitioning chromatography, although at the cost of a reduced step yield [90–92]. The tight operating window demanded for weak partitioning chromatography can be determined *via* automated batch-binding screenings at robotic liquid handling stations [90]. CEX resins have negatively charged groups immobilized on their surface. Selectivity towards product-related impurities such as deamidated, oxidized, N-terminally truncated mAb variants, and HMW species makes CEX chromatography one of the most widely used techniques in the DSP platform. For mAb purification, CEX is predominantly applied in bind-and-elute mode. Here, the positively charged product adsorbs to the CEX media due to the acidic pH of the load solution, while the negatively charged impurities are removed in the load and wash fraction [6, 93]. Subsequently, elution can be induced by an increased counter-ion concentration or an increased pH value, displacing the product from the adsorbent or decreasing the proteins net charge. Impurities with a higher positive charge as the target protein, e.g. mAb aggregates, can be removed during column regeneration. Typical CEX media can reach dynamic binding capacities of approximately 100 g/L_{Resin} in bind-elute mode [94]. For a more efficient and economic DSP, Liu *et al.* [95] proposed the method of overloaded isocratic CEX chromatography at loading densities of up to 1000 g/L_{Resin}. These extreme loading densities intend to exceed the breakthrough of the mAb product. Separation is achieved by the potentially delayed breakthrough of the impurities compared to the target protein [95]. CEX chromatography is considered as one of the most difficult to develop unit operations in the DSP platform. The main reason is the unique relationship between the structure of an antibody and its adsorption behavior to the CEX ligand. Parameters that affect the success of the separation problem are mobile phase conditions during loading, wash, and elution phase, as well as peak collection criteria and the chromatographic media [93]. Bench-scale experimentation and high-throughput methods are still state-of-the-art for the optimization of mobile phase conditions and resin selection [96], as well as for the identification robust operating conditions [14]. Mathematical models are currently gaining interest among biopharmaceutical development organizations due to their potential to complement or even substitute wet-lab experimentation [16, 97].

1.3 Mechanistic Chromatography Modeling

Mechanistic chromatography models are mathematical representations of the macroscopic phenomena within a chromatography column. During the last years, the biopharmaceutical industry worked on the implementation of mechanistic modeling in their development workflows to perform *in silico* experimentation. In multiple case studies, mechanistic models enabled computer-guided process optimization [98, 99], scale-up [27], root-cause investigation [100], or robustness analysis [101]. Compared to wet-lab DoE studies, simulations based on mechanistic models can reduce protein material and time demanded for the early stages of process development [102]. Besides potential economic benefits of mechanistic modeling, regulatory authorities support mathematical models to elucidate the relationships between process parameters and product quality [24, 25, 103].

The multitude of mechanistic chromatography models that can be found in literature differ in their complexity, ranging from simple one-dimensional column models [104, 105] to computational fluid dynamic simulations considering spatial inhomogeneities in the packed-bed due to channeling and wall-effects [106]. The general classification and terminology of mechanistic models described in this thesis follows the books of Schmidt-Traub *et al.* [76] and Guiochon *et al.* [77]. Fig. 1.3 depicts the fundamental phenomena in adsorption chromatography, including fluiddynamic, thermodynamic, and kinetic effects. Here, the ideal equilibrium model [104, 105] is the simplest model, only considering the convective transport applied by the system pump and protein adsorption to the stationary phase. The equilibrium dispersive model further considers peak broadening effects caused by eddy diffusion and mass transfer resistance. The frequently employed transport dispersive model lumps film and pore diffusion effects in a single effective mass transfer resistance. The general rate model considers all effects depicted in Fig. 1.3, including pore diffusion along the radial dimension of the adsorber particle. The introduction of an additional radial dimension makes the general rate model the computationally most demanding chromatography model presented in the following sections.

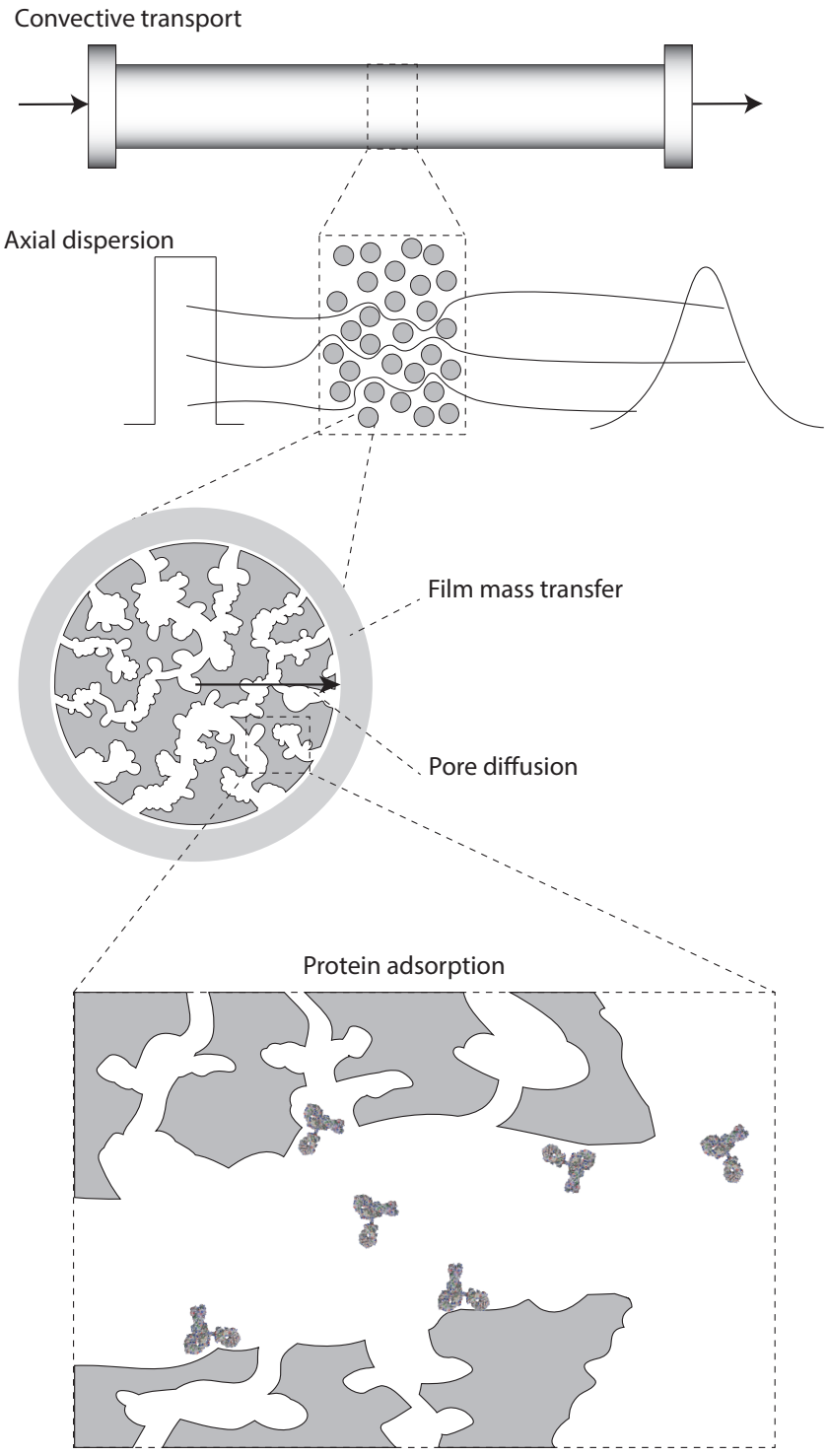

Figure 1.3: Fundamental mechanisms in adsorption chromatography

1.3.1 Differential Mass Balance for Chromatography

The majority of mechanistic chromatography models described in literature are based on the differential mass balance of a compound for an infinitesimally thin slice of the chromatography column [104, 105, 107]. Fig. 1.4 visualizes the fluxes of component i entering and exiting the column slice. Here, the column packing and consequently all column properties are assumed to be radially homogeneous. Further, wall effects are neglected leading to a constant flow velocity within the column cross-section. Hence, the chromatography column is simplified to a single dimension x following the longitudinal axis of the column. A recent study by Benner *et al.* [97] could confirm that wall effects on peak broadening are negligible for column diameters > 5 mm.

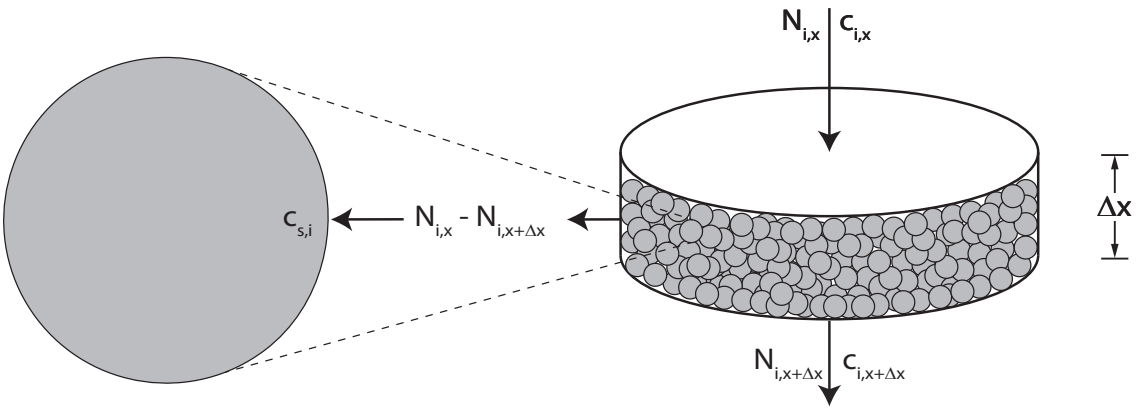

Figure 1.4: Mass balance in a column slice

The mass balance in the column slice depicted in Fig. 1.4 defines that the accumulation of component i in the stationary phase is equal to the difference between the amount of component i entering and the amount of i exiting the slice. The flux $N_{i,x}$ of component i at the inlet of a column slice of thickness Δx is

$$N_{i,x} = \varepsilon S \left(u_{int} c_i - D_{ax} \frac{\partial c_i}{\partial x} \right) \Bigg|_{x,t}, \tag{1.1}$$

where ε is the column porosity of the packed bed and $S = \pi {d_c}^2/4$ the cross-sectional area of the column with the diameter d_c. u_{int} is the mobile phase velocity in the interstitial volume, c_i the mobile phase solute concentration, and D_{ax} is the axial dispersion coefficient. In preparative chromatography, the axial dispersion coefficient D_{ax} comprises deviations of the fluid mechanics from plug flow. Thus, D_{ax} is affected by local non-homogeneities of the packed bed that lead to differences in fluid velocities on a mesoscopic scale, the so-called

eddy diffusion [108]. The flux at the outlet $x + \Delta x$ follows as

$$N_{i,x}|_{x+\Delta x} = \varepsilon S \left(u_{int} c_i - D_{ax} \frac{\partial c_i}{\partial x} \right) \Bigg|_{x+\Delta x,t}. \tag{1.2}$$

The accumulation rate in the slice is defined as

$$N_{i,x}|_{x+\Delta x} - N_{i,x} = S\Delta x \left(\varepsilon \frac{\partial c_i}{\partial t} + (1-\varepsilon) \frac{\partial c_{s,i}}{\partial t} \right) \Bigg|_{\bar{x}}, \tag{1.3}$$

where $c_{s,i}$ is the concentration of component i in the stationary phase and $\bar{x}$ is the average value of x. Insertion of Eqs. 1.1 and 1.2 in Eq. 1.3 results in the differential balance for component i the mobile phase:

$$\varepsilon S \left(u_{int} c_i - D_{ax} \frac{\partial c_i}{\partial x} \right) \Bigg|_{x,t} - \varepsilon S \left(u_{int} c_i - D_{ax} \frac{\partial c_i}{\partial x} \right) \Bigg|_{x+\Delta x,t} \tag{1.4}$$

$$= S\Delta x \left(\varepsilon \frac{\partial c_i}{\partial t} + (1-\varepsilon) \frac{\partial c_{s,i}}{\partial t} \right) \Bigg|_{\bar{x}} \tag{1.5}$$

When making Δx tend towards 0 and assuming constant axial dispersion D_{ax} and velocity u_{int} along the x-axis, Eq. 1.4 can be rewritten as

$$\frac{\partial c_i}{\partial t} = -u_{int} \frac{\partial c_i}{\partial x} + D_{ax} \frac{\partial^2 c_i}{\partial x^2} - \frac{1-\varepsilon}{\varepsilon} \frac{\partial c_{s,i}}{\partial t}. \tag{1.6}$$

The term on the left-hand side of the differential mass balance Eq. 1.6 accounts for accumulation of component i in the mobile phase. On the right-hand side, the first term describes the convective transport, the second term the diffusion, and the last term acoounts for accumulation of i in the stationary phase.

1.3.2 Ideal Model

The ideal model is the simplest chromatography model and assumes permanent equilibrium between stationary phase and mobile phase [76, 77, 104, 105]. Further, axial dispersion, mass transfer resistance and adsorption kinetic effects are neglected. Therefore, the concentration of component i in the interstitial volume equals the concentration in the pore phase. With $D_{ax} = 0$, Eq. 1.6 can be simplified to

$$\frac{\partial c_i}{\partial t} = -u \frac{\partial c_i}{\partial x} - \frac{1-\varepsilon_t}{\varepsilon_t} \frac{\partial q_i}{\partial t}, \tag{1.7}$$

where q_i is the concentration of component i adsorbed to the stationary phase. The total porosity ε_t is the sum of the interstitial and the particle porosity ε_p:

$$\varepsilon_t = \varepsilon + \varepsilon_p (1-\varepsilon), \tag{1.8}$$

and the effective velocity u of components penetrating the pore space is given as

$$u = \frac{\varepsilon}{\varepsilon_t} u_{int}. \tag{1.9}$$

The ideal model found numerous applications in the development of chromatography processes. An analytical solution of the ideal model was derived for simulated moving bed (SMB) chromatography allowing the calculation of concentration profiles within the columns [109]. Further, the ideal model can be used to explain fundamental phenomena in multi-component chromatography, including displacement effects [76, 110].

1.3.3 Equilibrium Dispersive Model

In contrast to the ideal model, the equilibrium dispersive model considers band broadening effects with one additional model parameter, the so-called apparent dispersion coefficient $D_{app,i}$. The apparent dispersion coefficient $D_{app,i}$ lumps peak broadening effects caused by axial dispersion with other mass transfer effects that might occur, including film mass transfer or pore diffusion [76, 111]. The model is based on the assumption that the concentrations of component i in the liquid and the pore phase are identical. Therefore, a differential equation for the pore phase is not demanded for the formulation of the equilibrium dispersive model [76]. The apparent dispersion coefficient depends on the interstitial fluid velocity and for non-linear adsorption isotherms also on the concentration of the solute. With the introduction of the total porosity ε_t and effective velocity u, the apparent dispersion coefficient D_{app} can be rearranged to

$$\tilde{D}_{app,i} = \frac{\varepsilon}{\varepsilon_t} D_{app} = \frac{u}{u_{int}} D_{app}. \tag{1.10}$$

When inserting the apparent dispersion coefficient given in Eq. 1.10 the equilibrium dispersive model follows as

$$\frac{\partial c_i}{\partial t} = -u \frac{\partial c_i}{\partial x} + \tilde{D}_{app,i} \frac{\partial^2 c_i}{\partial x^2} - \frac{1-\varepsilon_t}{\varepsilon_t} \frac{\partial q_i}{\partial t}. \tag{1.11}$$

Previous work showed that the predictive power of the equilibrium dispersive model is comparable to models of higher complexity when applying it to the simulation of highly efficient columns (Plate number N >>100) [112, 113].

1.3.4 Transport Dispersive Model

With the pore phase, the transport dispersive model shown in Eq. 1.12 introduces an additional phase between the interstitial volume and the adsorber surface. Therefore, the transport dispersive model considers two independent band broadening effects. Firstly, the effective mass transfer coefficient $k_{eff,i}$ accounts for film diffusion between interstitial volume and the pore volume, as well as for pore diffusion within the particle phase. Secondly, the axial dispersion coefficient D_{ax} considers peak broadening effects caused by eddy diffusion as a result of the packing quality. After introducing the concentration of the pore volume $c_{p,i}$ and the particle radius r_p in Eq. 1.11, the transport dispersive models

can be derived as

$$\frac{\partial c_i}{\partial t} = -\frac{u}{\varepsilon}\frac{\partial c_i}{\partial x} + D_{ax}\frac{\partial^2 c_i}{\partial x^2} - \frac{1-\varepsilon}{\varepsilon}\left(\frac{3}{r_p}k_{eff,i}\left(c_i - c_{p,i}\right)\right). \tag{1.12}$$

The corresponding pore model in Eq. 1.13 assumes a uniformal concentration in the mobile phase of the particles with the particle porosity ε_p. Thus, the change of concentration in the pore phase $c_{p,i}$ is the difference between the concentration on the particle surface and the concentration of component i adsorbed to the stationary phase q_i:

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

Due to its balance between low computational cost and sufficiently high level of detail, the transport dispersion model is widely used in industrial and academic case studies for simulating ion exchange chromatography processes [114–116].

1.3.5 General Rate Model

The general rate models represent the most detailed class of continuous chromatography models [76]. Here, mass transport effects are considered by at least two separate model parameters. $k_{film,i}$ incorporates film diffusion effects in the particle boundary layer. $D_{pore,i}$ lumps mass transfer inside the macro pores following Fick's law of diffusion, with surface or micro-pore diffusion in the solid phase [76, 77]. Modeling of pore diffusion demands the introduction of an additional dimension $r \in [0, r_p]$ along the particle radius. The transport dispersive model in Eq. 1.12 is used as column model after substituting $k_{eff,i}$ with $k_{film,i}$. Eq. 1.14 describes the mass transfer in the boundary layer of the particle.

$$\frac{\partial c_{p,i}}{\partial t}(x, r_p, t) = \frac{1}{D_{pore,i}}\frac{k_{film,i}}{\varepsilon_p}(c_i(x,t) - c_{p,i}(x, r_p, t)). \tag{1.14}$$

Within the particle pores, the concentration change of component i along the particle radius r can be expressed as

$$\frac{\partial c_{p,i}}{\partial t}(x, r, t) = D_{pore,i}\left(\frac{\partial^2 c_{p,i}}{\partial r^2} + \frac{2}{r}\frac{\partial c_{p,i}}{\partial r}\right) - \frac{1-\varepsilon_p}{\varepsilon_p}\frac{\partial q_i}{\partial t}(x, r, t). \tag{1.15}$$

Both, concentration and loading gradients vanish at the center of the particle ($r = 0$):

$$\left.\frac{\partial c_{p,i}}{\partial r}\right|_{r=0} = \left.\frac{\partial q_i}{\partial r}\right|_{r=0} = 0. \tag{1.16}$$

1.3.6 Adsorption Isotherms for Ion Exchange Chromatography

The present thesis resolves around model-based development tools for CEX chromatography processes in bind-elute mode. Therefore, consideration of protein adsorption is crucial for the simulation of process behavior. In IEX chromatography, charged groups on the proteins interact with ligands that are covalently bonded to the stationary phase. Multiple parameters are known to affect the electrostatic interaction between protein and ligand, including counterion concentration and pH of the mobile phase, ligand structure and ionic capacity of the stationary phase, and protein-specific attributes. The commonly employed steric mass-action (SMA) isotherm developed by Brooks and Cramer [117] accounts for the steric hindrance of salt counterions upon protein binding in multi-component equilibria. The following derivation of the SMA isotherm follows thermodynamic framework of protein adsorption described by Mollerup [118, 119]. For monovalent counterions in ion-exchange chromatography, the equilibrium of a protein P and a counterion S on a ligand L can be described as [117, 120, 121]

$$P + \nu_i LS \rightleftharpoons \nu_i S + PL_{\nu_i}, \tag{1.17}$$

where ν_i is the stoichiometric coefficient of component i and PL_{ν_i} is the protein-ligand complex. In thermodynamic equilibrium it must hold that

$$\Delta G = \sum_i \nu_i \mu_i = 0. \tag{1.18}$$

Here, μ_i is the chemical potential. ΔG is the Gibbs energy change given as

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

where ΔG^0 is the standard Gibbs energy change, K_{eq} denotes the thermodynamic equilibrium constant, R is the universal gas constant, and T the temperature. The chemical potential is given by the sum of ideal contribution μ_i^{id} and excess contribution μ_i^{excess} accounting for non-ideal behavior [119, 122, 123]:

$$\mu_i \equiv \mu_i^0 + RT \ln a_i = \mu_i^{id} + \mu_i^{excess}. \tag{1.20}$$

a_i are the activities and μ_i^0 are the standard state chemical potentials of pure components. Further, the ideal chemical potentials μ_i^{id} are given by

$$\mu_i^{id} = \mu_i^0 + RT \ln c_i - RT \ln c. \tag{1.21}$$

When assuming similar excess potentials in both phases, the activities can be replaced by mole fractions $x_i = c_i/c$. IEX chromatography is performed in aqueous solution, which further leads to the assumption that the solution molarity c is similar in both phases. Consequently, Eqs. 1.17-1.21 result in the stoichiometric displacement-model (SDM) [124, 125]

$$K_{eq} = \frac{q_i}{c_{p,i}} \left(\frac{c_s}{q_s} \right)^{\nu_i}. \tag{1.22}$$

Where c_p and c_s are the protein and counterion concentrations in the fluid phase, respectively. q_s is the concentration of counterion bound to the stationary phase. Due to electroneutrality, the sum of charges on the resin surface must be zero, leading to

$$q_s = \Lambda - \sum_{j=1}^{n} \nu_j q_j. \tag{1.23}$$

Λ is the total ionic capacity of the stationary phase, which can be determined experimentally [126]. Insertion of Eq. 1.23 in Eq. 1.22 results in

$$K_{eq} = \frac{q_i c_s^{\nu_i}}{\left(\Lambda - \sum_{j=1}^{n} \nu_j q_j\right)^{\nu_i} c_{p,i}}. \tag{1.24}$$

The SDM given in Eq. 1.24 is commonly used for simulating IEX chromatography. However, the SDM is only applicable for linear loading conditions, which are of limited relevance for industrial purification tasks. With the introduction of the SMA isotherm, Brookes and Cramer [117] extended the SDM isotherm towards non-linear loading conditions. Fig. 1.5 visualizes protein adsorption in IEX chromatography according to the SMA model. In Eq. 1.25, the steric shielding parameter σ_i considers the ligands on the resin surface that are sterically hindered by protein domains not contributing to the adsorption mechanism.

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

The mobile phase pH affects the protein net charge and consequently its retention behavior in IEX chromatography. Hence, it is crucial to add a pH-dependency to the SMA model given in Eq. 1.25. Hunt *et al.* [115] introduced empirical pH-dependencies for the characteristic charge ν_i and the thermodynamic equilibrium k_{eq} constant:

$$K_{eq,i}(pH) = K_{eq0,i} e^{K_{eq1,i} pH + K_{eq2,i} pH^2}, \tag{1.26}$$

$$\nu_i(pH) = \nu_{0,i} + pH \nu_{1,i}, \tag{1.27}$$

where $K_{eq0,i}$, $K_{eq1,i}$, $K_{eq2,i}$ are regression coefficients that can be determined by fitting the equilibrium constant against the mobile phase pH. $\nu_{0,i}$, $\nu_{1,i}$ are regression coefficients that that can be determined by fitting the characteristic charge against the mobile phase pH. Industrial case studies could successfully apply the empirical pH-dependencies given in Eqs. 1.26-1.27 within a limited pH-range. Nonetheless, their empirical nature prohibits extrapolation beyond the pH-range explored for model calibration. In contrast, the mechanistic pH-dependencies developed by Schmidt *et al.* [127] and Kluters *et al.* [128] consider the protein net-charge calculated based on the primary structure [129]:

$$\nu_i(pH) = \sum_i -\frac{N_{-i}}{1 + 10^{pK_{a,i} - pH}} + \sum_i \frac{N_{+i}}{1 + 10^{pH - pK_{a,i}}}. \tag{1.28}$$

N_{-i} is the number of acidic amino acids and N_{+i} the number of basic amino acids. $pK_{a,i}$ is the acid dissociation constant of respective amino acid i. Effects of the ionic strength on pK_a values can be considered using the Davies equation [130]. The pH-dependency of the equilibrium constant results from inserting the protein charge model in Eq. 1.28 into the fundamental Eqs. 1.17-1.21 [118, 120, 127, 128]:

$$\ln K_{eq,i}(pH) = \nu_i(pH)\frac{\Delta G_s^\circ}{RT} - \frac{\Delta G_{p,i}^\circ}{RT}, \tag{1.29}$$

where ΔG_s° is the standard state Gibbs energy change of the counterion upon adsorption and $\Delta G_{p,i}^\circ$ is the protein-specific standard state Gibbs energy change of adsorption. In contrast to empirical models, the mechanistic pH-dependencies in Eqs. 1.28-1.29 could allow extrapolation. However, model calibration is complex due to the high number of model parameters. The applied charge model in Eq. 1.28 neglects the individual position of a charged amino acid in the tertiary structure and potential effects on the respective pK_a values.

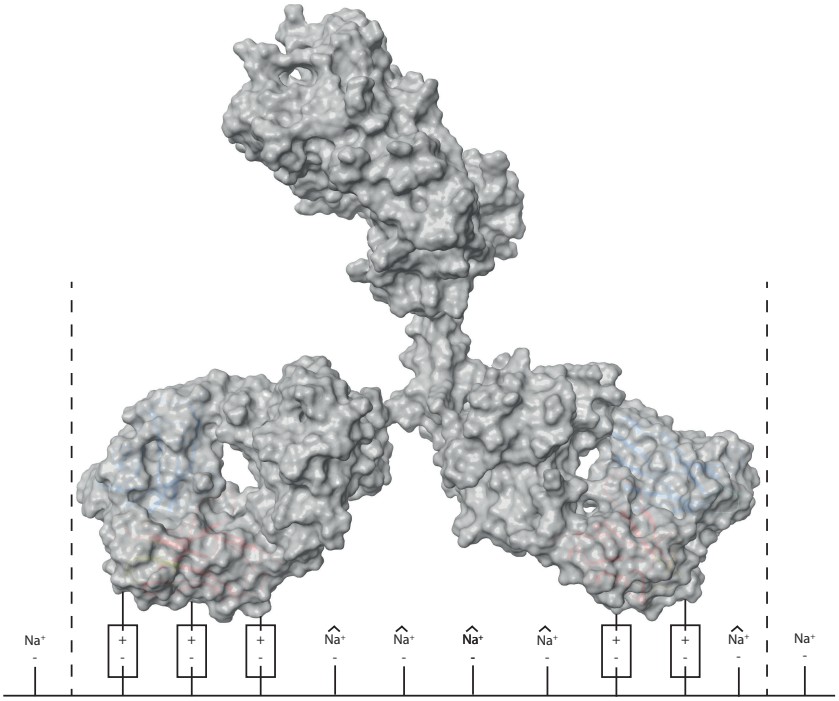

Figure 1.5: Idealized adsorption of a monoclonal antibody on a cation exchange chromatography resin according to the steric-mass action isotherm developed by Brooks & Cramer [117].

1.4 Quantitative Structure-Property Relationships

The development of preparative chromatography processes for the purification of biologics often begins with brute-force experimental methods. Scholars in the field of protein chromatography aim for an *a priori* prediction of retention times or adsorption model parameters based on protein structure information [31, 131–133]. Analyzing the relationships between chromatographic behavior and protein structure can help identify protein domains that mediate adsorption mechanisms. In general, the so-called quantitative structure-property/activity relationships (QSPR/QSAR) are mathematical models that correlate the molecular structure information of a compound with a quantifiable physical property or biological activity [134, 135]. The fundamental assumption of QSPR modeling is that molecules with a similar structure show similar physicochemical properties and functions [136]. Hence, QSPR models are equally relevant to small molecules and biologics, which makes them applicable to a wide spectrum of research areas, including medicinal chemistry [137], antibody drug discovery [138], biopharmaceutical process development [74], and formulation development [139].

In contrast to mechanistic models, QSPRs are empirical models that demand a significant amount of experimental data for training and validation. Due to their empirical nature, QSPRs are often applied to problems where physic-based models are not available. From a practical perspective, Mitchell *et al.* [134] defined QSPR model building as a two-stage process:

1. The molecular structures, typically given as molecular graphs or connection tables, must be converted into a vector of descriptors, usually denoted by the symbol x [134]. Descriptor vectors for the QSPR/QSAR modeling of proteins can be derived from their amino acid sequences or corresponding 3D protein structures [140–143]. When experimentally determined protein structures are not available, homology modeling can be applied to predict the protein structure based on the amino-acid sequence (Details given in Section 1.1).

2. The second part aims to empirically determine a function that maps between the descriptor vectors (x) and the measurable property of the molecules (y) [144]. This mapping function can be based on simple (multi-)linear regression, or more advanced machine learning algorithms, including support vector regression, Gaussian process regression, random forest, k-nearest neighbors, or artificial/deep neural networks [31, 134, 144]. The selection of a machine learning algorithm for QSPR modeling is a non-trivial task. The most relevant factors are size and distribution of experimental data, linearity and complexity of the chemical problem, and internal correlations of the descriptor data set [134].

The two building blocks for QSPR modeling listed above are closely interconnected. Not all descriptors derived from the protein structure must be relevant for the mapping function between feature vector x and property of interest y. Especially for large proteins, including mAbs, the feature vector x can consist of hundreds of columns, including global,

as well as, local, topographical and shape-specific descriptors for each protein domain [142]. Furthermore, the determination of the property of interest y is limited by experimental constraints and an finite number of molecules that is available for model training. To avoid over-determination of the machine learning model, the so-called feature selection aims to reduce the dimensionality of the QSPR problem by selecting descriptors that are relevant for the prediction of the property of interest [134, 145, 146]. For feature selection, an individual feature or a subset of features is compared to labeled training data. The ranking and filtering of features based on simple measures, e.g. regression coefficients (R^2) or mutual information, gives first insights into the relationships between individual descriptors and the property of interest. Individual feature ranking is often used as baseline method due to its simplicity, low computational costs, and high interpretability of the found correlations [147]. Features selected *via* individual ranking can be used as input for the computationally more expensive Wrapper methods. In general, Wrapper methods are based on training and cross-validating a learning machine to evaluate the relevance of a subset of features. The frequently employed recursive feature elimination starts training a machine learning model with all available features and progressively eliminates the least relevant ones [147]. In the final step of QSPR modeling, the predictive power of the trained machine learning model must be validated against an external molecule that was not included in the training or cross-validation data set [148].

1.5 Research Proposal

Monoclonal antibodies (mAbs) and bispecific antibody (bsAb) formats have the ability to address unmet medical needs, resulting in their importance for the pharmaceutical industry. Downstream processing (DSP) of mAbs relies on the so-called platform concept, where different mAb candidates are purified using a standardized sequence of unit operations and with product-specific adaption of process conditions. The increasing structural complexity of novel antibody formats challenges the development of purification processes following the platform paradigm. Cation exchange chromatography has been identified as a development-intensive unit operation, where minimal differences in the primary structure of mAbs have a significant impact on their manufacturability. Due to the unknown relationships between protein structure and process behavior, development strategies for preparative chromatography have to rely on high-throughput experimentation (HTE) and statistical design of experiments (DoE).

While miniaturized HTE methods enable investigation of a large experimental space within a short time frame, their comparability to manufacturing-scale can be limited. With the Quality by Design (QbD) initiative, regulatory authorities demand a clear understanding of the correlations between process parameters and product quality. Process understanding provided by mechanistic chromatography modeling exceeds the empirical correlations given by HTE and DoE studies. In addition, quantitative structure-property relationship modeling (QSPR) opens up the possibility to predict process behavior of biologics based on their protein structure. However, the transfer of digital technologies from academia to industry is slow. Additional research around model building and model application needs to increase the acceptance of mechanistic and QSPR modeling in industry.

The ultimate goal of the present research proposal is the development of multiscale modeling tools bridging the gaps between the antibody structure and a robust downstream process. This includes the investigation of QSPR models for the prediction of adsorption model parameters. Due to experimental limitation, QSPR models for CEX chromatography have been limited to smaller model proteins and the prediction of retention times. In addition, model calibration, validation, and application methods should be established that cope with the quality standards defined by regulatory authorities. Integration of homology modeling, statistical modeling, and mechanistic chromatography modeling into a common multiscale *in silico* framework could guide the path from the primary structure of the antibody candidate to the process model of the final unit operation.

Cation exchange chromatography (CEX) is frequently employed as polishing step to separate a mAb from its process- and product-related impurities. Protein adsorption to CEX media is driven by the structural characteristics of the mAb. Hence, optimal operating conditions have to be identified for each antibody candidate by conducting laborious high-throughput or bench-scale experiments. While mechanistic chromatography modeling may allow to partially replace experiments with computer simulations, the initial hurdle of model calibration demands experimental data for each molecule. Thus, the first

part of this thesis aims to elucidate fundamental relationships between protein structural properties of biologics and their adsorption model parameters. QSPRs could be developed to predict adsorption model parameters based protein descriptors derived from homology models.

The complex and experimentally extensive model calibration complicates the use of mechanistic models in industry. To apply mechanistic chromatography models to real-world tasks in biopharmaceutical process development, it is crucial to ensure adequate model certainty in a limited amount of time. Therefore, the second part of this work is devoted to the development of a straightforward model calibration method based on fundamental knowledge on preparative chromatography and a standardized set of experiments.

In the light of the QbD initiative, regulatory authorities propose the use of mechanistic modeling to ensure robust product quality by understanding the fundamental physical phenomena of the manufacturing process. If models are used to establish a control strategy of the manufacturing process, predictions must capture scaling effects and process variability anticipated during routine manufacturing. Therefore, the last part of this thesis aims to develop methods for model validation and subsequent application for *in silico* process characterization. Besides increased process understanding, novel methods for *in silico* process characterization should reduce the number of wet-lab experiments without taking additional risk, or guide the design of experimental studies required for regulatory approval.

1.6 Outline

Chapter 2 investigates qualitative relations between single amino acid substitutions in the variable region of monoclonal antibodies and their elution behavior in preparative cation exchange chromatography. QSPR models introduced in Chapter 3 enable the prediction of adsorption isotherm parameters based on protein structure models of mAbs. Chapter 4 proposes a method for the calibration of mechanistic cation exchange chromatography models using a standardized set of experiments. In Chapter 5, the previously calibrated mechanistic model is applied to the *in silico* scale up of the unit operation, including clinical-manufacturing of the therapeutic mAb. Ultimately, Chapter 6 allows the *in silico* process characterization following the QbD concept.

Chapter 2: Modeling the impact of amino acid substitution in a monoclonal antibody on cation exchange chromatography

D. Saleh, R. Hess, M. Ahlers-Hesse, N. Beckert, M. Schönberger, F. Rischawy, G. Wang, J. Bauer, M. Blech, S. Kluters, J. Studts, J. Hubbuch

This article investigates the influence of amino acid substitutions in the CDR of a full-length IgG1 mAb on its process behavior in preparative CEX chromatography. Single amino acid substitutions within the investigated mAb resulted in an additional positive charge in the light chain (L) and heavy chain (H) CDR, respectively. The mAb variants showed an increased retention volume in linear gradient elution compared to the wild type antibody. Further, the substitution of tryptophan with lysine in the H-CDR3 increased charge heterogeneity of the product. A multiscale in silico analysis, consisting of homology modeling, protein surface analysis, and mechanistic chromatography modeling increased understanding of the adsorption mechanism. The results elucidate the qualitative relationships between the structure of mAbs and their SMA adsorption isotherm parameters.

Biotechnology and Bioengineering, 2021, 10.1002/bit.27798

Chapter 3: Quantitative structure-property relationships for the prediction of adsorption isotherm parameters of therapeutic antibodies in cation exchange chromatography

D. Saleh, R. Hess, M. Ahlers-Hesse, F. Rischawy, G. Wang, S. Kluters, J. Grosch, J. Studts, J. Hubbuch

In this manuscript, quantitative structure-property relationship (QSPR) models were trained for the prediction of SDM parameters of mAbs. By combining homology modeling, QSPR modeling, and mechanistic modeling, the multiscale modeling method leads from the amino acid sequence of a therapeutic mAb to digital twin of the unit operation. The training and validation data set included SDM parameters and structural descriptors of a

diverse set of IgG1, IgG4, and bispecific antibody formats at varying pH conditions. The relationships between protein structural descriptors and SDM parameters increased the understanding of protein-surface interaction in strong CEX chromatography. Based on the developed multiscale modeling framework, the first *in silico* studies for DSP development can be performed before protein material for wet-lab experiments are available.

Manuscript in preparation

Chapter 4: Straightforward method for calibration of mechanistic cation exchange chromatography models for industrial applications

D. Saleh, G. Wang, B. Müller, F. Rischawy, S. Kluters, J. Studts, J. Hubbuch

To support the application of mechanistic modeling in industry, this article introduces a standardized method for the calibration of multicomponent, pH-dependent SMA isotherm models. In the case investigated, the method was applied to an antibody polishing step including four protein species. The developed strategy combined well-established theories of preparative chromatography and allowed a systematic reduction of unknown model parameters. The model was validated beyond the operating ranges of the final unit-operation, enabling its application to advanced tasks in late-stage downstream process development. Further, the achieved model certainty meets the demands of process development strategies following the QbD concept.

Biotechnology Progress, 2020, 10.1002/btpr.2984

Chapter 5: Cross-scale quality assessment of a mechanistic cation exchange chromatography model

D. Saleh, G. Wang, B. Müller, F. Rischawy, S. Kluters, J. Studts, J. Hubbuch

In this article, the previously calibrated mechanistic model was validated as digital twin of the large-scale unit operation. Based on chromatograms and elution pool data ranging from laboratory- to manufacturing-scale, the proposed modeling workflow enabled early identification of differences between scales, e.g. system dispersion effects or ionic capacity variability. A multi-stage model qualification approach was introduced to measure the model quality and to understand the model's limitations across scales. The mechanistic chromatography model avoided limitations of the SDM by capturing effects of bed height, loading density, feed composition, and mobile phase properties. The results demonstrate the applicability of mechanistic chromatography models as a possible alternative to conventional development approaches based on experimental scale-down models (SDM).

Biotechnology Progress, 2021, 10.1002/btpr.3081

Chapter 6: *In silico* process characterization for biopharmaceutical development following the Quality by Design concept

D. Saleh, G. Wang, F. Rischawy, S. Kluters, J. Studts, J. Hubbuch

In the light of the ObD initiative, this final manuscript explores mechanistic chromatography models for the *in silico* process characterization of a mAb manufacturing processes. The proposed modeling workflow covered the main-tasks of traditional PCS studies following the QbD principles, including criticality assessment of process parameters and establishment of their proven acceptable ranges (PARs) of operation. Analyzing effects of multi-variate sampling of process parameters on the purification outcome allowed identification of the edge-of-failure. Validation of *in silico* results demanded less experimental efforts compared to traditional PCS approaches. Stochastic simulation, considering the measured variances of process parameters and loading material composition, was used to estimate the capability of the process to meet the acceptance criteria for critical quality attributes and key performance indicators. The proposed workflow enables the implementation of digital process twins as QbD tool for improved development of biopharmaceutical manufacturing processes.

Biotechnology Progress, 2021, 10.1002/btpr.3196

2 Modeling the Impact of Amino Acid Substitution in a Monoclonal Antibody on Cation Exchange Chromatography

David Saleh[1, 2], Rudger Hess[1, 2], Michelle Ahlers-Hesse[2], Nicole Beckert[1, 2], Markus Schönberger[2], Federico Rischawy[2], Gang Wang[2], Joschka Bauer[2], Michaela Blech[2], Simon Kluters[2], Joey Studts[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] Boehringer Ingelheim Pharma GmbH & Co. KG, Biberach, Germany
* Corresponding author: Jürgen Hubbuch, +49 721 608 47526, juergen.hubbuch@kit.edu

Abstract

A vital part of biopharmaceutical research is decision making around which lead candidate should be progressed in early-phase development. When multiple antibody candidates show similar biological activity, developability aspects are taken into account to ease the challenges of manufacturing the potential drug candidate. While current strategies for developability assessment mainly focus on drug product stability, only limited information is available on how antibody candidates with minimal differences in their primary structure behave during downstream processing. With increasing time-to-market pressure and an abundance of mAbs in development pipelines, developability assessments should also consider the ability of mAbs to integrate into the downstream platform.

This study investigates the influence of amino acid substitutions in the complementarity-determining region (CDR) of a full-length IgG1 mAb on the elution behavior in preparative cation exchange (CEX) chromatography. Single amino acid substitutions within the investigated mAb resulted in an additional positive charge in the light chain (L) and heavy chain (H) CDR, respectively. The mAb variants showed an increased retention volume in linear gradient elution compared to the wild type antibody. Further, the substitution of tryptophan with lysine in the H-CDR3 increased charge heterogeneity of the product.

A multiscale in silico analysis, consisting of homology modeling, protein surface analysis, and mechanistic chromatography modeling increased understanding of the adsorption mechanism. The results reveal the potential effects of lead optimization during antibody drug discovery on downstream processing.

2.1 Introduction

In recent years, the number of mAbs investigated in clinical studies has steadily increased [3, 4]. Manufacturing and material supply for pre-clinical and clinical trials form a major building block in biopharmaceutical product development and is often referred to as Chemistry, Manufacturing and Controls (CMC). A recent cost evaluation performed by Farid *et al.* (2020) [5] predicted that CMC activities represent 13-17% of the total R&D budget from pre-clinical trials to regulatory approval. Their calculation considered costs caused by 92% of candidates that fail during pre-clinical and clinical development [5]. Biopharmaceutical organizations strive to streamline development by using computer-aided sequence optimization tools in drug discovery that increase the likelihood of successful CMC programs within a strict timeframe [73, 149, 150]. Current strategies for developability assessment rely on in silico techniques to predict aggregation propensities, solubility issues, or long-term stability of the mAb product [73, 151]. Compared to stability aspects in formulation development, how single amino acid substitutions in lead candidates may affect developability of the downstream process (DSP) is poorly understood.

Industrial DSP of biopharmaceuticals relies heavily on chromatographic separation techniques. For the purification of mAbs, preparative cation exchange (CEX) chromatography is frequently employed as a polishing step [6, 7, 152]. CEX chromatography allows the removal of process related impurities, including DNA, host cell proteins, endotoxins, or leached Protein A [6, 95]. Due to its high selectivity towards protein charge, CEX chromatography can also separate an antibody from its product related impurities, e.g. size or charge variants. Despite the favorable properties regarding impurity removal, CEX chromatography remains one of the most development-intensive unit operations in the downstream process. The comparably strong effect of a mAb's structural characteristics on elution behavior in CEX chromatography demands an adaption of process conditions for each product. Therefore, a significant body of research has focused on understanding the relationship between protein structure properties and adsorption behavior. Adsorption in protein chromatography can be investigated using molecular dynamic simulations [131, 153]. Monte Carlo simulations performed by Zhou *et al.* (2004) [154] showed that antibodies tend to have a "head-on" fragment antibody $(Fab)_2$ binding orientation on negatively charged surfaces and an "end-on" fragment crystallizable (Fc) binding orientation on positively charged surfaces, at high surface charge density and low salt concentration. In multimodal (MM) systems, the chromatographic behavior and binding orientation of model proteins [155–157] and mAbs [158] depends on ligand structure and surface properties of the molecule. Further, Robinson *et al.* (2020) [159] showed that the domain contribution of mAbs in MM chromatography is affected by the mobile phase pH value.

In contrast to qualitative analysis, quantitative structure-property relationship (QSPR) models correlate structural descriptors with chromatographic behavior by applying mathematical models. QSPRs based on the crystal structure of non-mAb proteins allowed the prediction of protein retention times [160] and adsorption isotherm parameters [31] in ion exchange chromatography. Robinson *et al.* [30] applied QSPR modeling to the purification of homologous Fab variants on MM resins. Kittelmann *et al.* introduced an orientation sensitive QSPR approach for model proteins [132] and mAbs [133] in IEX chromatography. A comprehensive study performed by Ishihara *et al.* (2005) [74] investigated the elution behavior of 28 mAbs in Protein A affinity and CEX chromatography. For CEX chromatography, the salt concentrations at peak maximum correlated with the surface positive charge distribution of the heavy chain variable region [74]. To the best of our knowledge, results on how substituting residues within the adsorption relevant surface of a full-length mAb would influence CEX chromatography have not been published yet.

In contrast to structure-based modeling techniques, mechanistic models describe the physical effects in chromatography columns on a macroscopic level. During the last years, mechanistic chromatography modeling became a state of the art technology in biopharmaceutical DSP development. Possible applications of mechanistic models are process optimization [26, 161], model-guided scale-up [27, 97], *in silico* robustness analysis [101, 162, 163], or root cause investigation [100]. The process understanding provided by mechanistic chromatography models enables in silico DSP development in line with the Quality by Design (QbD) concept [164]. Chromatography models consist of partial differential equations describing mass transport and protein adsorption phenomena. For IEX chromatography, the adsorption can be modeled using the stoichiometric displacement model (SDM) [124]. The SDM is based on the electrostatic equilibrium theory and formulates the multipoint binding of proteins under consideration of displacement effects. Brooks and Cramer (1992) [117] extended the SDM towards the steric mass-action (SMA) model to cover the shielding effects of bound protein on the resin surface.

Previous work has demonstrated the predictive power of SMA models, even when extrapolating beyond the experimental conditions applied for model calibration [165]. Despite the proven predictive power and the mechanistic nature of the SMA model, it is not clear which structural characteristics of mAbs influence adsorption model parameters. For example, Rüdt *et al.* (2015) [166] built a SDM model for an Fc fusion protein and corresponding aggregates by assuming constant characteristic charge values for both protein species. Other authors [116, 166] reported differing characteristic charge parameters for mAb size and charge variant, due to altering numbers of charged groups interacting with the resin. Furthermore, it is unclear if steric shielding is exclusively a function of the chromatographic ligands blocked by the adsorbed protein or if it further considers non-adsorptive or repulsive effects [31]. Due to the missing correlations between adsorption isotherm parameters and structural characteristics of multi-domain proteins such as mAbs, the preferred practice for model calibration often involves curve fitting to experimental data [114, 165, 167]. Substitution of single charged amino-acid side chains in the adsorption relevant region of full-length mAbs could elucidate the correlations between protein structure and macroscopic adsorption model parameters.

The aim of the present work is to gain a deeper understanding of the binding mechanism of mAbs in preparative CEX chromatography. Therefore, multiple purification experiments were performed for a full-length IgG1 mAb, and two variants differing in a single amino acid within the CDR. An additional positive charge in the L- and H-CDR of the mutated mAbs increased their retention volume during linear salt gradient elution (LGE). Potential effects of amino acid substitutions on the developability of the CEX unit operation were investigated in chromatography runs at low and high loading densities. Homology modeling and protein surface analysis identified exposed mAb regions that mediate the adsorption process. The data set enabled the estimation of SMA model parameters for the three mAbs. The identified effects of amino acid substitutions on adsorption model parameters could support model calibration and QSPR modeling.

2.2 Methods

2.2.1 Process Conditions and mAbs

The mAb polishing step was performed on the strong CEX resin POROS XS (Thermo Fisher Scientific, Waltham, USA). Column specific parameters and the equations used for their calculation are listed in Table 2.1. Tracer injections with blue dextran and 1 M sodium chloride (both Sigma-Aldrich, St. Louis, USA) enabled the calculation of the interstitial volume, V_{int} and total liquid volume, V_t, respectively. The ionic capacity Λ was determined by acid-base-titration [126]. Column characterization experiments were conducted as triplicates. Chemicals used in this study were of pharmaceutical grade. All buffers were prepared with deionized water and filtered with a 0.2 μL sterile filter. Column experiments were performed on the preparative chromatography system ÄKTA Avant 25 controlled via Unicorn 7 (both Cytiva, Uppsala, Sweden). Sodium acetate buffer at pH 5.25 was used for all preparative CEX experiments. The column was equilibrated at pH 5.25 and a counterion concentration of 0.05 M sodium. Buffer exchange of protein samples into the equilibration buffer resulted in defined loading conditions. During gradient elution, counter-ion concentration increased from 0.05 M to 0.50 M sodium. The loading density was 1 g/L_{Resin} for LGEs in the linear region of the adsorption isotherm and 30 g/L_{Resin} for high loading density runs. Samples applied for low and high loading experiments had protein concentration of 1 g/L and 3 g/L, respectively. Loading densities were adjusted via the applied sample volume, considering the 10.68 mL column volume. 1 M and 0.1 M sodium hydroxide were used for column regeneration and storage, respectively.

The three model proteins used in this study are IgG1 mAbs expressed in Chinese hamster ovary (CHO) cells (Boehringer Ingelheim Pharma GmbH & Co. KG, Biberach, Germany). The mAbs were captured via Protein A affinity chromatography. Two mutants (M1 and M2) with additional positively charged groups in the CDRs were derived from a corresponding wild type (WT) antibody. Surface charge of M1 was increased by substitution of serine with lysine in the L-CDR3. For M2, a tryptophan in the H-CDR3 was substituted

Table 2.1: Experimentally determined system- and column-specific model parameters.

Parameter	Symbol	Value	Unit	Equation	Reference
Length	L	136	mm		
Diameter	d	10	mm		
Column Volume	V	10.68	mL	$V = \pi \frac{d^2}{4} L$	
Bead radius	r_p	25	μm		[114]
Interstitial porosity	ε_{col}	0.41	-	$\varepsilon_{col} = \frac{V_{int}}{V}$	[114]
Total porosity	ε_t	0.78	-	$\varepsilon_t = \frac{V_t}{V}$	[114]
Particle porosity	ε_p	0.63	-	$\varepsilon_p = \frac{V_t - V_{int}}{V - V_{int}}$	[114]
Axial dispersion	D_{ax}	0.14	mm^2/s	$D_{ax} = \frac{uLs^2_{NaCl}}{(2V_t)^2}$	[115]
Ionic capacity	Λ	0.49	M	$\Lambda = \frac{c_{NaOH} V_{NaOH}}{V(1-\varepsilon_t)}$	[114, 126]

with lysine. The amino acid substitutions were introduced to affect the biophysical properties of the mAb *via* modification of surface-exposed hydrophobic and charged patches. This methodology can reveal possible effects of mAb lead optimization on CMC properties in upstream, downstream and formulation development.

2.2.2 Homology Modeling and Protein Surface Analysis

Full-length homology models of investigated mAbs were built in Maestro BioLuminate 3.7 (Schrödinger, Munich, Germany) following the method developed by Zhu *et al.* (2014) [69]. Separate templates for light and heavy chains were selected based on sequence identity of framework regions. For WT, M1, M2 the framework region templates were 3SO3 & 3T2N (pdb accession codes) for heavy and light chains, respectively. Comparably low sequence identity (<40%) between available templates and the 14 residues long H-CDR3 demanded an *ab initio* structure prediction using the Prime method [66]. Following the method developed by Zhu *et al.* (2014) [69], all structures were prepared accordingly before starting loop prediction. Structure preparation included the assignment of polar hydrogen positions, protonation states, and energy minimization using the OPLS3e force field [168]. Surface patches and protein descriptors were calculated within BioLuminate, at pH 5.25 [142, 169].

2.2.3 Mechanistic Chromatography Modeling

The simulation software ChromX (GoSilico GmbH, Karlsruhe, Germany) was used for mechanistic chromatography modeling. The transport dispersive model in Eq. 2.1 was

applied as column model [114, 170]. In Eq. 2.1, the change of the concentration $c_i(x,t)$ is a function of the convective mass transport in the interstitial volume of the packed bed with the superficial velocity u. Further, the model considers axial dispersion D_{ax} effects and interfacial mass transfer between the interstitial volume defined by the bed porosity ϵ_{col} and the particle pores. Film diffusion effects in the particle boundary layer and pore diffusion in the particle phase are expressed by the effective mass transfer parameter $k_{eff,i}$. Eq. 2.2 represents the accumulation of mass in the pore volume c_i and the stationary phase q_i. Danckwerts' boundary conditions are given in Eqs. 2.3 and 2.4.

$$\begin{aligned}\frac{\partial c_i(x,t)}{\partial t} = &-\frac{u}{\varepsilon_{col}}\frac{\partial c_i(x,t)}{\partial x} + D_{ax}\frac{\partial^2 c_i(x,t)}{\partial x^2} \\ &- \frac{1-\varepsilon_{col}}{\varepsilon_{col}}\left(\frac{3}{r_p}k_{eff,i}\left(c_i(x,t) - c_{p,i}(x,t)\right)\right)\end{aligned} \tag{2.1}$$

$$\frac{\partial c_{p,i}(x,t)}{\partial t} = \frac{3}{r_p}\frac{k_{eff,i}}{\varepsilon_p}(c_i(x,t) - c_{p,i}(x,t)) - \frac{1-\varepsilon_p}{\varepsilon_p}\frac{\partial q_i(x,t)}{\partial t} \tag{2.2}$$

$$\frac{\partial c_i}{\partial x}(0,t) = \frac{u(t)}{\varepsilon_{col} D_{ax}}(c_i(0,t) - c_{in,i}(t)) \tag{2.3}$$

$$\frac{\partial c_i}{\partial x}(L,t) = 0 \tag{2.4}$$

Protein adsorption is modeled using the SMA isotherm [117]. Eq. 2.5 shows the kinetic form of the SMA isotherm modified by Hahn *et al.* (2016) [167], where q_i and $c_{p,i}$ denote the protein concentration in the solid and liquid phase of the particle, respectively. The SMA model formulates the equilibrium binding behavior of the protein considering the salt concentration in the pore phase c_s, the ionic capacity of the resin Λ and protein specific model parameters. The protein characteristic charge ν_i accounts for the number of charges interacting with the resin, while steric shielding σ_i considers the number of functional groups on the resin blocked by the protein due to steric hindrance. Additionally, the constants $k_{eq,i} = k_{ads,i}/k_{des,i}$ and $k_{kin,i} = 1/k_{des,i}$ comprise adsorption and desorption rates of the modeled proteins.

$$k_{kin,i}\frac{\partial q_i}{\partial t} = k_{eq,i}\left(\Lambda - \sum_{j=1}^{k}\left(\nu_j + \sigma_j\right)q_j\right)^{\nu_i} c_{p,i} - q_i c_s^{\nu_i} \tag{2.5}$$

$$q_{salt} = \Lambda - \sum_{j=1}^{k} \nu_j q_j \tag{2.6}$$

Estimation of protein-specific model parameters is the first step, before a mechanistic model can be applied to real-world tasks in DSP development. The Yamamoto method enabled the analytical solution of ν_i and $k_{eq,i}$ using a set of LGEs at differing salt gradient slopes. Eq. 2.7 describes the linear relationship between the normalized gradient slope GH and the elution salt concentration $c_{s,i}$ of component i at diluted loading conditions [74, 166, 171–173]. Eqs. 2.8 and 2.9 lead to the calculation of the normalized gradient slope GH, where $c_{s,initial}$ is the salt concentration at the gradient begin, $c_{s,final}$ is the salt concentration at the gradient end, and V_G is the gradient length in mL. The isotherm parameters defining the non-linear region of the SMA isotherm, $k_{kin,i}$ and σ_i, were estimated using the inverse method developed by Hahn *et al.* (2016a; 2016b) [114, 167].

$$log(GH) = (\nu_i + 1) log(c_{s,i}) - log\left(k_{eq,i} \Lambda^{\nu_i} (\nu_i + 1)\right) \tag{2.7}$$

$$g = \frac{c_{s,final} - c_{s,initial}}{V_G} \tag{2.8}$$

$$GH = g(V_{col} - \varepsilon_t V_{col}) \tag{2.9}$$

2.3 Results

2.3.1 Process Behavior and Protein Surface Analysis

The aim of the present study was to analyze effects of single amino acid substitutions in the CDR of a mAb on process behavior in CEX chromatography. The two mutants M1 and M2 had an additional positively charged group in the CDR compared to the WT mAb. Details on amino acid substitutions are given in Section 2.2.1. An identical set of preparative CEX experiments was performed for three mAb variants. In this section, results of LGE experiments under low loading conditions (1 g/L_{Resin}) are compared to the *in silico* analysis of corresponding homology models. Figure 2.1 (D)-(F) show LGE experiments at a gradient length of 20 column volumes (CV) and mobile phase pH 5.25. The wild type antibody eluted first, followed by the mutants M1 and M2 with additional positively charged amino acid side chains in the CDR. WT, M1 and M2 eluted at a sodium counter-ion concentration of 0.336 M, 0.416 M and 0.433 M during salt gradient elution, respectively. Beside the delayed retention volume, similar peak shapes could be observed for the WT and M1. In contrast, M2 showed an increased elution pool volume compared to WT and M1, caused by a distinct pre-peak with shoulder. Results of analytical chromatography (HP-CEX) indicated that the pre-peak of M2 was caused by an increased charge heterogeneity of the loading material (data not shown).

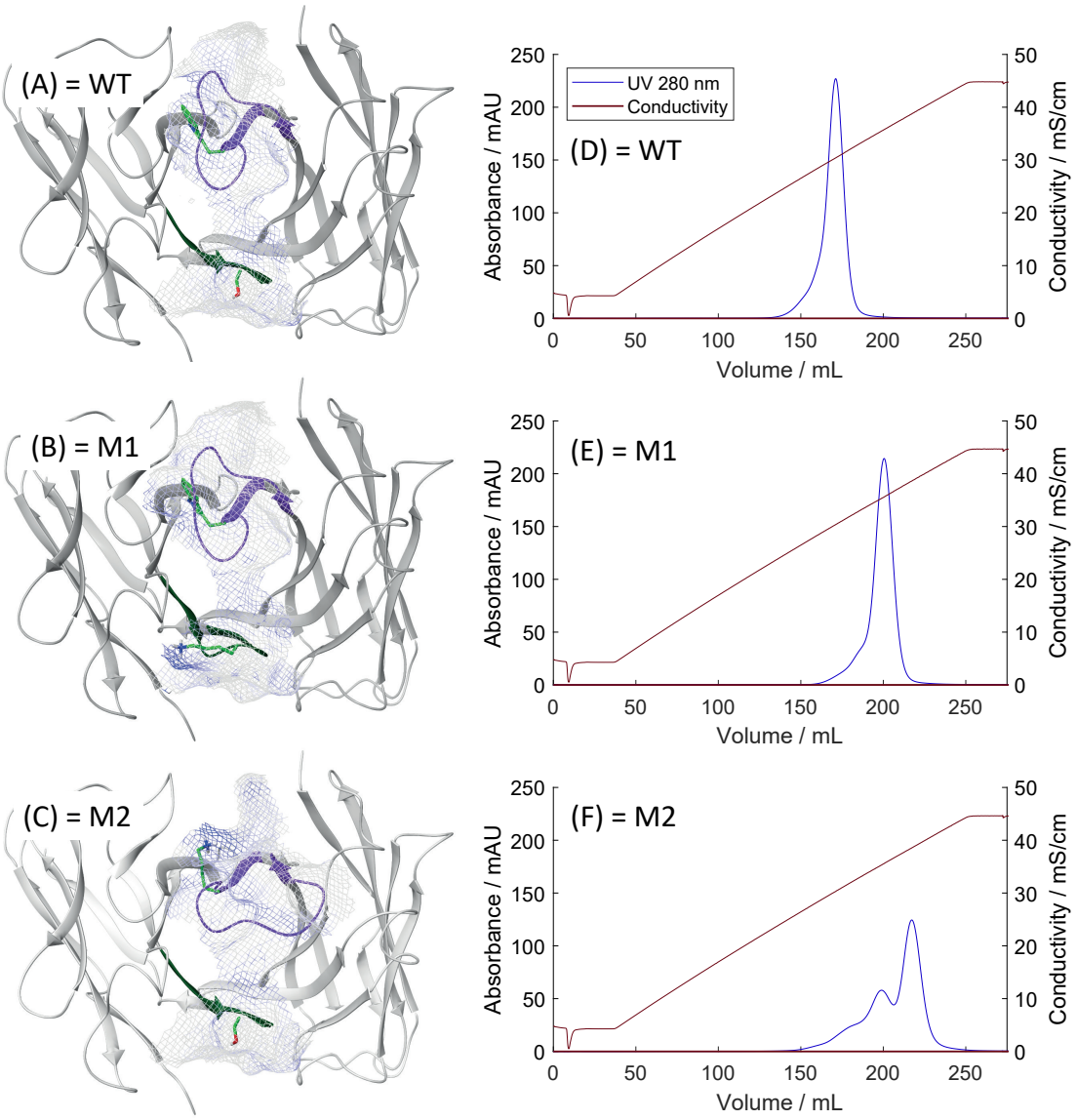

Figure 2.1: Protein surface analysis and elution behavior in CEX chromatography at 1 g/L_{Resin} loading density. Predicted Fv homology models: (A) WT with tryptophan in HC and serine in LC; (B) M1 with tryptophan in HC and lysine in LC; (C) M2 with lysine in HC and serine in LC. The solvent-exposed surface is shown as mesh, and positively charged patches close to H-CDR3 (purple) and L-CDR3 (dark green) are marked in blue. (D), (E) and (F) show the respective CEX chromatograms of LGE experiments conducted on POROS XS at pH 5.25, with 20 CV gradient length, and a flow velocity of 200 cm/h.

The visible effect of point mutations in the CDRs on the elution behavior during CEX chromatography demanded a thorough investigation of structural characteristics causing these phenomena. Homology models were calculated for the three IgG1 mAbs using the workflow described in Section 2.2. Antibody structure predictions in Figure 2.1 (A)-(C) confirmed that the substituted side chain residues were solvent-exposed. Due to identical Fc regions, protein surface analysis was performed for the Fab only, while the pI and net charge calculations were based on full-length homology models. Table 2.2 compares pI values obtained *via* capillary isoelectric focusing (cIEF) with *in silico* predicted pI values and protein descriptors. Both, *in silico* and wet-lab experiments, could not measure significant differences between pI values of the three mutants. Model based net charge of the full-length IgG1 mutants at pH 5.25 increased to 50 from 48. The number of positive charges per CDR increased to 7 from 6 for the mutated mAbs at pH 5.25. Additionally, compared to WT, point mutations increased the sum energy of positive surface patches larger than 30 Å2 in the CDRs by 20% and 28% for M1 and M2, respectively. Protein surface analysis in Figure 2.1 visualizes the additional positively charged patches in the CDR of mutated mAbs. Due to the high degree of solvent-exposure, all depicted surface patches could potentially be involved in the adsorption process. For M1, serine was substituted with lysine, adding a positive patch in a neutral region of the wild type mAb. Beside the differences in positive surface charges, *ab initio* H-CDR3 loop prediction *via* the Prime method [66] resulted in similar conformations for WT and M1. In contrast, the substitution of tryptophan with lysine in M2 influenced the predicted conformation of the H-CDR3, increasing solvent-exposure compared to WT and M1.

Table 2.2: Molecular descriptors and pI values obtained *via in silico* prediction and cIEF measurements. *:*In silico* descriptors were calculated based on homology models at pH 5.25.

	pI* [-]	pI [-]	Net charge*	Formal charge CDR*	Positive patch energy CDR* [kcal/mol]
Method	*In silico*	cIEF	*In silico*	*In silico*	*In silico*
WT	9.2	8.9	48	6	499
M1	9.2	8.9	50	7	599
M2	9.2	8.9	50	7	638

2.3.2 Chromatography Modeling

Homology modeling and protein surface analysis provided insights into the structural properties of the mAbs leading to differences in their chromatographic behavior. Mechanistic chromatography modeling aims to increase process understanding on a macroscopic level by describing the physical effects in the chromatographic system. This section investigates the effects of changes in the protein structure on macroscopic adsorption model parameters.

System and column characterization was crucial for the following estimation of protein-specific model parameters. All results and respective methods of column characterization and are listed in Table 2.1. POROS XS is a perfusion resin, with comparably low mass transfer resistance. Therefore, $k_{eff,salt}$ was approximated with $r_P/3 = 0.0083$ mm/s.[174] For protein species, the penetration correlation [175] enabled calculation of effective mass transfer parameters $k_{eff,i}$ depending on the hydro-dynamic radii of the mAbs and the linear flow rate. Identical hydrodynamic radii of 73 Å were computed based on full-length homology models resulting in an $k_{eff,i}$ of 0.0013 mm/s for WT, M1 and M2. Predictions of the height and width of elution peaks at low loading conditions validated assumptions regarding the mass transfer of protein.

Multiple lab-scale experiments for each mAb allowed the estimation of SMA model parameters. Table 2.3 summarizes resulting isotherm parameters. Five LGE experiments were conducted at 1 g/L_{Resin} loading density and altering gradient lengths between 10 CV and 30 CV. At each gradient slope, the WT eluted first followed by M1 and M2. Figure 2.2 visualizes the Yamamoto correlation (Eq. 2.7), which enabled the analytical solution of the characteristic charge ν_i and equilibrium constant $k_{eq,i}$ based on the five LGE experiments per mAb. The parallel regression lines in Figure 2.2 resulted in comparable characteristic charge ν_i. In addition, the predicted and measured net charges and pI values of the homology models in Table 2.2 did not show a considerable difference between the molecules. In contrast to ν_i, the equilibrium constants $k_{eq,i}$ of the mutants M1 and M2 in Table 2.3 increased to approximately 8- and 23-fold compared to the WT, respectively. The characteristic charge ν_i and the equilibrium constant $k_{eq,i}$ have a similar effect on the retention volume during LGE elution at low loading densities. Thus, the equilibrium constant $k_{eq,i}$ was the parameter affected by the addition of positively charged groups in the CDR and caused the shifts in retention volume.

Steric shielding σ_i and kinetic $k_{kin,i}$, the remaining model parameters defining the non-linear region of the SMA isotherm, were estimated using the inverse method introduced by Hahn *et al.* [167]. Here, the model output was fitted to the UV signal at 280 nm wavelength of the high load LGE runs in Figure 2.3. Similar to $k_{eq,i}$, the kinetic parameter $k_{kin,i}$ increased for the mutated mAbs. The shielding parameters in Table 2.3 show that the substitution of tryptophan with lysine in M2 reduced σ_i by 28% compared to the WT. The single protein species defined for WT and M1 was able to describe peak shapes and retention volumes at 30 g/L_{Resin} loading density. For M2, the increased concentration of acidic charge variants demanded the consideration of additional protein species. Two acidic peak groups (APG1 and APG2) of M2 were defined based on the peak-to-peak ratios measured at the low loading density LGE experiments. The relative input composition of M2 Main, M2 APG1, and M2 APG2 was 67%, 16%, and 17%, respectively. Single amino acid substitutions could not affect the characteristic charge parameters of WT, M1, and M2 in a magnitude that would describe the differences in retention volume. Thus, the characteristic charge ν_i of the M2 charge variants APG1 and APG2 was assumed to be equal to the characteristic charge of the M2 main species.

For all three molecules, the transport-dispersive SMA model could describe the UV sig-

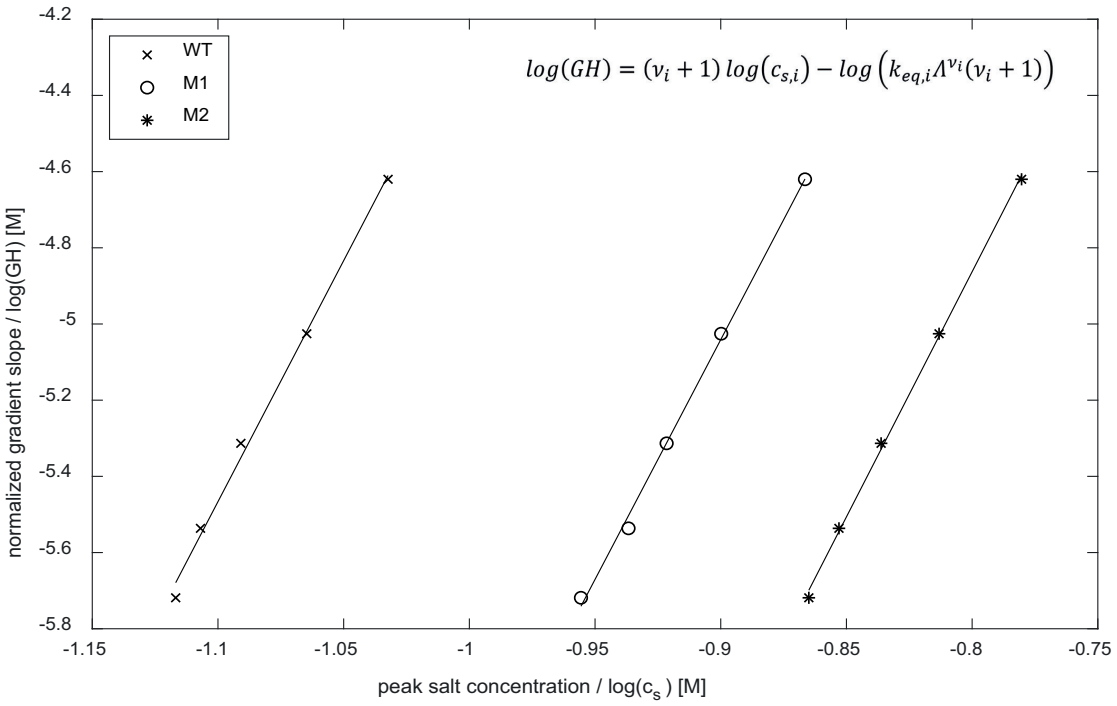

Figure 2.2: Model calibration for the linear region of the SMA isotherm. Yamamoto method for LGE experiments with 10, 15, 20, 25, 30 CV gradient length at pH 5.25 on the strong CEX media POROS XS. Slope and intercept of linear regressions enabled the calculation of characteristic ν_i and $k_{eq,i}$ equilibrium constant, respectively. $R^2 > 0.98$ for all linear regressions.

Table 2.3: SMA Isotherm parameters of WT, M1 and M2 at pH 5.25 on POROS XS. Acidic charge variants of M2 were lumped in two acidic peak groups (APG) based on their retention time in preparative chromatography.

	ν_i [-]	$k_{eq,i}$ [-]	$k_{kin,i}$ [sM$^\nu$]	σ_i [-]
WT	11.7	0.08	4.58E-05	53
M1	11.6	0.65	3.41E-04	51
M2 Main	11.9	1.82	4.09E-04	38
M2 APG1	11.3	0.65	1.05E-04	38
M2 APG2	11.3	0.15	2.74E-04	38

nal. This indicates that the selected model captured relevant physical effects within the chromatographic system. The comparison of isotherm parameters in Table 2.3 with structural descriptors in Table 2.2 shows that macroscopic adsorption model parameters are directly affected by single amino acid substitutions and the resulting protein structure. Besides the identified effects on mechanistic model parameters, amino acid substitutions influenced process performance at both, low- and high loading densities. Retention times, charge heterogeneity, and elution pool volumes in preparative column experiments dif-

fered between the three mAb variants. Thus, lead optimization can potentially affect preparative purification using CEX chromatography.

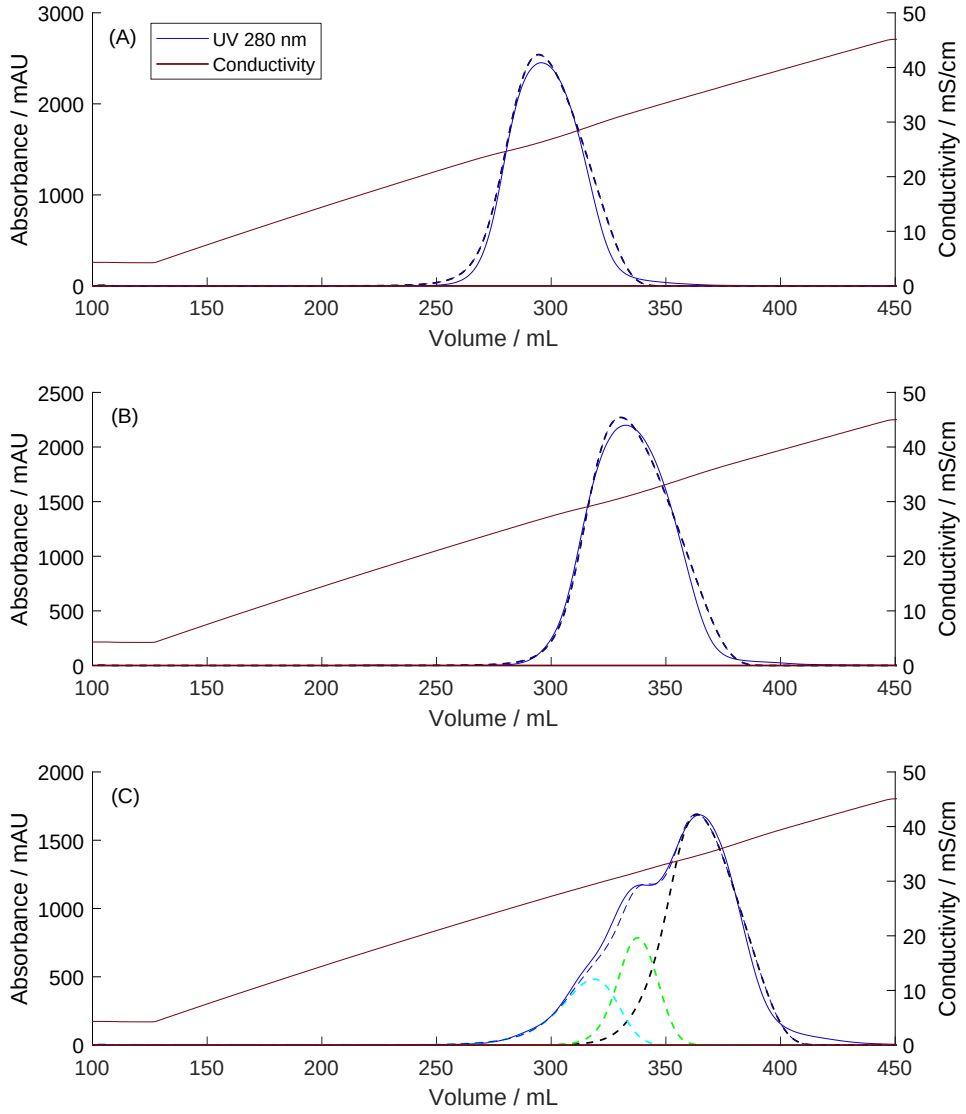

Figure 2.3: Comparison of simulated and measured chromatograms of the CEX process under non-linear loading conditions. LGE experiments on POROS XS at 30 g/L_{Resin} loading density, with 30 CV gradient length, and a flow velocity of 200 cm/h. The blue dashed lines indicate simulated sum signals. Solid lines represent experimental data. (A) = WT. (B) = M1. For WT and M1 a single protein species was simulated. (C) = M2. Three protein species were considered in the simulation of the M2 process. Green and cyan dashed lines are simulations of acidic charge variants of M2.

2.4 Discussion

During the last years, a significant body of research has focused on the development of *in silico* tools for the reduction of aggregation propensity, improvement of solution properties, and optimization of biological activity of therapeutic mAbs [73, 149, 176]. Optimization of protein stability has been shown to be associated with additional favorable CMC properties, e.g. increased expression titers during upstream processing [73, 149]. The present study aimed to investigate the effects of single amino acid substitutions in full-length mAbs on the process behavior during preparative CEX chromatography.

Previous work indicated that mAbs bind with their Fab first to strong CEX media [74, 154]. In this work, modification of CDR residues that contribute to positive patches on the mAb surface proved the vital role of the Fab and CDR in antibody adsorption on strong CEX media. An identical set of preparative CEX experiments with three mAb variants differing in a single amino acid showed a correlation between the sum positive patch energy of the CDR and retention time during gradient elution. Independent of CDR loop confirmations, the substitution of tryptophan with lysine in the H-CDR3 of M2 increased the elution salt concentration by 32%. This resulted in comparably high salt concentrations in the product pool. The final salt concentration in the product pool needs to be considered in designing the process sequence, when combining CEX with other chromatographic modalities for mAb polishing. Thus, exchange of a single charged amino acid during lead optimization has the potential to affect resin selection and process design during downstream processing. Despite differences in patch energies and CEX retention times, pI values and net charges of the homology model did not differ significantly between investigated mAbs. While this may seem trivial due to the minimal variation in primary structure, it is important to notice that the experimental pI value is often considered as initial binding strength indicator for the development of ion exchange chromatography processes. Previous studies could show that solvent-exposed charges on the mAb surface are relevant for protein adsorption in ion exchange and mixed mode chromatography [74, 158, 177]. In this study, targeted amino-acid substitutions within the protein-resin interaction surface emphasized that inhomogeneous charge distribution affects chromatographic behavior of proteins with almost identical net charges. Therefore, local and global molecule properties should be considered equally in the design of IEX chromatography processes, especially for large multi-domain molecules such as mAbs. Nonetheless, the unraveled correlations and binding orientation are only valid for the investigated strong CEX media. In contrast to our data, Müller-Späth *et al.* (2008) [178] separated mAbs with differing numbers of C-terminal lysine using an analytical weak CEX media at pH 6.3. Their data suggests that the Fc part of mAbs contributes to protein adsorption on weak CEX media. Consequently, protein adsorption is equally dependent on the molecule structure and the chromatographic ligand. Future studies are required to shed light on the correlations between antibody structure and process behavior in other chromatographic unit operations.

Beside the shift in retention volume, the substitution of tryptophan with lysine in the

H-CDR3 of M2 had a significant effect on the percentage of acidic charge variants and the resulting elution profiles in Figure 2.1 and Figure 2.3. Interestingly, ab initio H-CDR3 structure prediction using Prime method [66] resulted in an increased solvent-exposure of the M2 H-CDR3 compared to the WT and M1 mAbs (Figure 2.1). Recently, Lan *et al.* (2020) [179] reported an anomalous charge variant profile of an IgG1 mAb in analytical CEX chromatography caused by two discrete conformational states of the H-CDR3. Their analytical and molecular modeling investigation revealed a pH-dependent equilibrium between "open" and "closed" conformational states of the H-CDR3. Further, their work identified a neighboring tryptophan residue in the LC showing reduced solvent-exposure at lower pH values. Previous work underlined that aromatic and hydrophobic groups have a stabilizing effect on CDR loop conformation [68]. Therefore, it can be assumed that the substitution of tryptophan with lysine in M2 destabilized the energetically favored loop conformation and increased solvent-exposure of the H-CDR3 (Figure 2.1). Charge heterogeneity is a common feature of therapeutic mAbs, which does not necessarily influence their efficacy [179, 180]. However, when comparing the results to previous work [181], the distinct peak shoulder in the chromatograms of M2 is uncommon for IgG1 mAbs in preparative CEX chromatography. When applying identical UV based collection criteria to the three mAb variants, the elution pool volume of M2 increases by 33% compared to WT and M1. Due to limited tank capacities in manufacturing facilities and processing time related to the subsequent product concentration *via* ultrafiltration/diafiltration, the high elution pool volume represents an undesired CMC property. Hence, the substitution of potentially stabilizing, aromatic groups within the H-CDR3 should be avoided during lead optimization. Based on these findings, mAb candidates with favorable CMC properties can be selected for a streamlined process development. Alternatively, a qualitative prediction of protein adsorption based on the sequence may allow an early estimation of experimental efforts necessary for developing polishing chromatography. In case a mAb shows a high likelihood for unfavorable CMC properties in CEX chromatography, e.g. increased elution pool volumes or high salt concentration in the elution pool, a broad resin screening with differing chromatographic modalities can be planned early during process development. For a holistic DSP developability assessment, additional unit operations should be added to the multiscale modeling workflow. However, our results indicate that CEX chromatography is sensitive to minimal changes in the primary structure of mAbs. For other unit operations in common DSP platforms, e.g. filtration steps, Protein A affinity chromatography, or anion exchange chromatography in flow-through mode, amino acid substitutions during lead optimization are less likely to have significant effects on the process performance. Thus, the developed in silico workflow supports one of the most development intensive unit operations in our DSP platform. Future work should investigate how the identified correlations integrate into a holistic CMC developability assessment combining upstream, downstream and formulation parameters.

The results in section 2.3.2 confirmed a relationship between mAb structure and adsorption model parameters. Due to the semi-mechanistic nature of the commonly applied SMA adsorption model, an interpretation of model parameters is challenging, especially for multi-domain molecules such as mAbs. The multi-component SMA adsorption model [117] is an extension of the stoichiometric-displacement concept developed by [124]. Here,

the characteristic charge describes multipoint binding of proteins on charged surfaces. In Section 2.3.2, introduction of two additional charges per mAb could not increase the characteristic charge parameters of M1 and M2 in a magnitude that describes the shift in retention time. When comparing the constant ν_i parameters to constant pI values and protein net charges in Table 2.2, the characteristic charge can be assumed to describe the average number of positive patches on the protein surface interacting with the resin. In contrast to the characteristic charge, the equilibrium constant $k_{eq,i}$ was significantly increased by the introduction of positively charged groups within the adsorption-mediating region on the protein surface. Therefore, $k_{eq,i}$ may define the strength of the anisotropic adsorption reaction, depending on positive patches in mAb CDR. Substitution of the tryptophan residue in M2 reduced the steric shielding parameter σ_i. Reduction of σ_i could have been caused by a combination of two effects. Firstly, the conformational change of the H-CDR3 could have reduced the number of ligands blocked by the protein. Additionally, it has been observed that the steric shielding parameter can also describe repulsive effects on the resin surface [31]. Thus, substitution of the aromatic tryptophan residue could have influenced hydrophobic interactions with the resin backbone leading to a reduced σ_{M2} compared to σ_{WT}.

The knowledge gained on structural dependencies of model parameters could support the selection of plausible boundary conditions for model parameter estimation, which is crucial for avoiding local minima [163]. Despite the influence of amino acid substitutions on adsorption model parameters, the dataset is still limited to three variants of a monoclonal antibody. When increasing size and structural heterogeneity of the mAb data set, the found correlations could enable QSPR modeling for the prediction of SMA parameters. Conventional QSPR approaches for developability assessment correlate structural properties with specific performance indicators or quality attributes of the biopharmaceutical [138, 182]. A structure-based prediction of adsorption model parameters could lead to a full digital representation of the unit operation enabling the simulation of an unlimited number of process conditions before material for wet-lab CMC activities is available.

2.5 Conclusion

Our results demonstrate that sequence optimization of mAb candidates can influence downstream processing. Single amino acid substitutions in the CDR had a significant impact on retention volumes and elution profiles during preparative CEX chromatography. The findings enable a relative classification of mAb candidates in weak, medium, and strong adsorption to CEX media based on the number of positively charged amino acid side chains in the CDR. Substitution of tryptophan with lysine destabilized the H-CDR3 loop conformation, leading to an increased charge heterogeneity and broadened elution profiles in CEX chromatography. The identified relationship between mAb primary structure and CMC properties may support the selection of mAb candidates that integrate into common downstream platforms. Further, a structure-based estimation of mAb elution behavior in CEX chromatography could be used to plan initial experiments during early-phase DSP

development. Effects of amino acid substitutions on semi-mechanistic adsorption model parameters underline the possibility of building QSPR models that support the calibration of mechanistic chromatography models. Our results could promote a paradigm shift in DSP development from a strictly generic platform process to a more flexible process design driven by the structural characteristics of the mAb.

2.6 Acknowledgment

The authors would like to acknowledge experimental support of their colleagues in the analytical development department, including Patrick Arendt, Larissa Panhans and Lisa Hospach.

The authors have no conflict of interest to declare.

3 Quantitative Structure-Property Relationships for the Prediction of Adsorption Isotherm Parameters of Therapeutic Antibodies in Cation Exchange Chromatography

David Saleh[1, 2], Rudger Hess[1, 2], Michelle Ahlers-Hesse[2], Federico Rischawy[1, 2], Gang Wang[2], Simon Kluters[2], Jan-Hendrik Grosch[2], Joey Studts[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] Downstream Process Development, Boehringer Ingelheim Pharma GmbH & Co. KG, Biberach, Germany

*Corresponding author: Jürgen Hubbuch, +49 721 608 47526, juergen.hubbuch@kit.edu

Abstract

The development of preparative chromatography processes for purification of monoclonal antibodies (mAbs) is challenged by an unprecedented speed-to-market pressure and the increasing structural complexity of novel antibody formats. This study presents a multiscale *in silico* model consisting of homology modeling, quantitative structure-property relationships (QSPR), and mechanistic chromatography modeling leading from the amino acid sequence of a mAb to the digital twin of a cation exchange chromatography (CEX) process. The model leverages the mAbs' structural characteristics and experimental data of a diverse set of therapeutic antibodies to predict elution profiles of two mAbs that were removed from the training data set. The QSPR modeling procedure identified protein descriptors within the variable region of mAbs that are relevant for adsorption to strong CEX media. Consideration of two discrete conformational states of IgG4 mAbs enabled prediction of split-peak chromatograms. The presented multiscale model has the potential to support process optimization during early stage development before protein material for experimental studies is available.

3.1 Introduction

Monoclonal antibodies (mAbs) and bispecific antibody formats (bsAbs) are intriguing treatment options for a wide spectrum of therapeutic areas, including oncology, hematology, inflammatory diseases [1, 45], and more recently for passive immunization or treatment of infectious diseases, including Coronavirus Disease 2019 (COVID-19) [183]. Biopharmaceutical companies strive to accelerate process development of biologics to bring potentially life-saving medicines to patients as quickly as possible. In the light of the ongoing Coronavirus pandemic, Kelley [184] proposed a CMC development plan that shortened the time from lead identification to start of phase 1 clinical investigation from 12 to 6 months. This strategy combines novel technologies with a drastically templated platform process and a strict parallelization of CMC work packages [184]. While enabling an early start of phase 1 clinical studies, the development plan designed by Kelley results in a higher business risk caused by missing experimental studies during early-stage development. The biopharmaceutical industry demands novel methods supporting process development at pandemic-pace, while achieving highest product quality and robust material supply for clinical investigations.

Due to structural similarities of different mAb products, their large-scale purification is based on the so-called platform process [7, 8]. This platform process consists of a standardized sequence of orthogonal separation mechanisms and the adaptation of process conditions is reduced to a minimum. While enabling an efficient and rapid process development for most molecules, the increasing structural complexity of antibody formats as well as the poorly understood adsorption mechanisms in preparative chromatography challenge a downstream processing under standardized process conditions. For cation exchange (CEX) chromatography, a previous study of our group revealed that single substitutions in the amino acid sequence of full-length mAbs affect elution behavior and the resulting optimal operating conditions that must be identified during early-stage development [185]. Luo *et al.* reported split-peak elution profiles of IgG2 and IgG4 mAbs in CEX chromatography caused by reversible self-association [186] and histidine-protonation-based charge variants [187]. Multiple authors propose to increase process understanding by using mechanistic models as digital twins of the manufacturing process [23, 28, 188]. Mechanistic chromatography models consist of a set of partial differential equations that describe mass transport and thermodynamic adsorption phenomena within the chromatography column on a macroscopic level. *In silico* screening of process conditions *via* mechanistic modeling enables efficient process optimization for the increasing number of mAbs entering the development phase. For ion exchange chromatography (IEX), the semi-mechanistic steric mass-action (SMA) isotherm [117, 118] is frequently employed for the design [27], optimization [98], and characterization [116] of antibody polishing steps. As a simplification of the SMA model, the stoichiometric displacement model (SDM) allows simulation of protein adsorption and desorption under diluted loading conditions [124, 125]. In the light of the Quality by Design (QbD) initiative [24], regulatory authorities support the use of mechanistic models to increase product quality by understanding the fundamental relationships between process parameters and quality attributes [103, 164]. Despite successful

case studies and regulatory initiatives, the cumbersome model calibration can negate the potential benefits of mechanistic modeling for industrial application. Mechanistic model parameters are not directly measurable and are often determined *via* curve fitting to experimental data [114]. Especially during the early phases of DSP development, protein material and time for model calibration are limited. Further, the experiments demanded for the determination of SMA and SDM parameters, e.g. experiments under diluted loading conditions and fraction collection with a high number of analytical samples, are not performed during traditional development workflows. Understanding the relationships between the protein structure of therapeutic antibodies and their macroscopic adsorption isotherm parameters could support a model calibration demanding less experiments.

The inability of current model-based development approaches to transfer process knowledge from existing mAb products to a new biological entity (NBE) led to an increased interest in machine learning and AI-based methods for bioprocessing [23, 29, 189, 190]. Quantitative structure-property relationship (QSPR) models are powerful tools for the *a priori* prediction of process behavior based on empirical relationships between protein structure descriptors and a target property. When process understanding and experimental data is limited, QSPR models give initial insights into the developability of mAb candidates during early-stage development [191]. For the DSP of proteins, QSPRs allowed the prediction of protein retention times in IEX [160] and mixed mode chromatography [30]. Further, Kittelmann *et al.* [132, 133] introduced QSPR models capable of predicting binding orientations of mAbs and model proteins in IEX chromatography. For mAbs in CEX chromatography, Ishihara *et al.* found a correlation between the elution salt concentration and positively charged patches in the heavy chain variable region [74]. Multiscale models use QSPRs to connect protein structure properties with macroscopic model parameters. In contrast to the prediction of single quality attributes and elution salt concentrations, QSPR models for the prediction of isotherm parameters could lead to a digital twin of the unit operation that enables *in silico* process development before protein material for experimental studies is available. Ladiwala *et al.* [31] built a QSPR model for the prediction of SMA isotherm parameters of small model proteins. Noteworthy, their multiscale model enabled the prediction of entire IEX chromatograms based on physicochemical descriptors derived form experimentally determined protein structures [31]. However, experimentally determined protein structures of full-length mAbs are not available during early-stage process development. Further, the binding orientation and adsorption mechanism of large multi-domain proteins, such as mAbs, are not fully understood.

The goal of the present study is a multiscale *in silico* model that leads from the amino acid sequence of a therapeutic antibody to the mechanistic chromatography model of the corresponding CEX unit operation. The multiscale model combines mAb homology modeling, QSPRs, and mechanistic chromatography modeling to predict macroscopic adsorption isotherm parameters. Antibody-specific descriptors are derived from homology models built for a diverse set of full-length antibodies. The experimental data set for training and validation of the QSPR model consists of 21 mAbs and corresponding Fabs, including IgG1, IgG4, and two different bispecific antibody formats. Protein descriptors of IgG4 mAbs are calculated based on two discrete protein conformations, the typical Y-

shape and a closed, self-associated λ-shape where one Fab-arm is oriented towards the Fc portion [192]. SDM parameters are determined at three mobile phase pH conditions for all IgG1 and IgG4 mAbs, which aims to increase the understanding of adsorption mechanisms of different antibody formats in preparative CEX chromatography. QSPR modeling is than used to build an empirical relationship between the molecular descriptors and the macroscopic SDM parameters. Here, the initial feature selection aims to identify regions on the protein surface that affect SDM parameters, and thus also affect process behavior in CEX chromatography. External validation of the multiscale model is achieved by predicting chromatograms of two molecules not included in the training data. QSPRs coupled with mechanistic models can predict process behavior based on the amino acid sequence without performing additional experiments during early-stage DSP development.

3.2 Methods

3.2.1 Antibodies and Homology Modeling

A set of 21 therapeutic Fab (n=1), IgG1 mAbs (n=10), IgG4 mAbs (n=5), and bsAbs (n=5) was used to train and test a QSPR model for the prediction of SDM isotherm parameters. Additionally, enzymatic fragmentation of IgG1 and IgG4 mAbs was performed following the protocol by Andrew *et al.* [193] to further increase size and structural diversity of the data set. All proteins were expressed in Chinese hamster ovary cells (Boehringer Ingelheim GmbH & Co. KG, Biberach, Germany). Full-length homology models of investigated mAbs and complex antibody formats were built in Maestro BioLuminate 3.7 (Schrödinger, Munich, Germany) following the method developed by Zhu *et al.* (2014) [69]. Template structures for full-length homology models are given in Fig. 3.1. The mAb crystal structure with the PDB entry 1HZH [194] was used as Fc-template for IgG1 mAbs and bAbs. Due to the recently discovered conformational diversity of IgG4, two different Fc-templates were used for the homology modeling of each IgG4 mAb [192]. The PDB entry 5DK3 [195] was used as Y-shaped template, while 6GFE [192] was used as template for the λ-shaped IgG4 conformation. Fig 3.1 shows exemplary homology models for each antibody format, including the two conformations of IgG4 mAbs. Further, two different bispecific antibody formats were included in the data set, one knob-in-hole format (n=2) with two different Fabs, and one appended IgG(H)-scFv format (n=3). For the variable regions, separate templates for light and heavy chains were selected based on sequence identity of framework regions. All structures were prepared following the method developed by Zhu *et al.* [66] before starting descriptor calculation. Structure preparation included the assignment of polar hydrogen positions, protonation states, and energy minimization using the OPLS3e force field [168]. Movement of heavy atoms during energy minimization was limited to a maximum root-mean-square-deviation of 0.3 Å between the initial homology model and the energy minimized structure.

Surface patches and protein descriptors were calculated within BioLuminate, at pH con-

ditions of the respective chromatography experiments [142, 169]. The descriptor set consisted of 166 parameters, including 135 global and 31 local descriptors per region. Protein descriptors calculated for the entire antibody structure were defined as global descriptors. Additionally, local descriptors were defined for 30 individual regions of the antibody. These regional mAb descriptors were organized in a topological hierarchy with six levels of detail, ranging from the overall mAb structure to local descriptors for each CDR loop. The descriptors were calculated *via* physical models or parameterized empirical models, including patch potential energy analysis with size- and value-dependent binning. Further, structure specific characteristics, which were describing the shape or scalar values of the different regions, were calculated within BioLuminate. The 135 global descriptors plus 31 local descriptors for 30 individual mAb regions resulted in a vector containing 1065 features per full-length mAb [142].

3.2.2 Process Conditions

All molecules were captured *via* Protein A affinity chromatography and further polished using the strong CEX resin POROS XS (Thermo Fisher Scientific, Waltham, USA). Column characterization experiments were conducted as triplicates. Column dimensions and column-specific model parameters are listed in Tab. 3.1. Chemicals used in this study were of pharmaceutical grade. All buffers were prepared with deionized water and filtered with a 0.2 μm sterile filter. Preparative CEX experiments were performed on an ÄKTA Avant 25 controlled using Unicorn 7 (both Cytiva, Uppsala, Sweden). Tracer injections with blue dextran and 1 M sodium chloride (both Sigma-Aldrich, St. Louis, USA) enabled the calculation of the interstitial volume, V_{int} and total liquid volume, V_t, respectively. The ionic capacity Λ was determined by acid-base-titration [126].

An identical set of preparative CEX experiments in bind-elute mode was performed for all 21 molecules and corresponding Fabs of IgG1/IgG4 mAbs. Linear salt gradient elution (LGE) runs were conducted at gradient lengths ranging from 10 CV to 30 CV, at three different mobile phase pH values, pH 5.00, pH 5.25, and pH 5.50. The counterion concentration increased from 50 mM to 500 mM sodium during gradient elution. The LGEs were conducted in the linear region of the adsorption isotherm, at a loading density of 1 g/L_{Resin}, allowing for the calculation of the characteristic charge ν_i and the equilibrium constant $k_{eq,i}$ at varying pH conditions using the Yamamoto method described in the following section. 1 M and 0.1 M sodium hydroxide were used for column regeneration and storage, respectively.

3.2.3 Mechanistic Chromatography Modeling

The simulation software ChromX was used for mechanistic chromatography modeling (GoSilico GmbH, Karlsruhe, Germany). The transport dispersive model in Eq. 3.1 was selected as column model [114, 170]. Here, the change of the concentration $c_i(x,t)$ in

the mobile phase depends on convective mass transport in the interstitial volume of the packed bed with the superficial velocity u. Peak broadening effects are described by axial dispersion D_{ax} and interfacial mass transfer between the interstitial volume and the particle phase. The effective mass transfer parameter $k_{eff,i}$ lumps film diffusion effects in the particle boundary layer and pore diffusion in the particle phase. Eq. 3.2 describes accumulation of component i in the pore volume $c_{p,i}$. Danckwerts' boundary conditions are given in Eqs. 3.3 and 3.4.

$$\begin{aligned}\frac{\partial c_i(x,t)}{\partial t} = &-\frac{u}{\varepsilon_{col}}\frac{\partial c_i(x,t)}{\partial x} + D_{ax}\frac{\partial^2 c_i(x,t)}{\partial x^2} \\ &- \frac{1-\varepsilon_{col}}{\varepsilon_{col}}\left(\frac{3}{r_p}k_{eff,i}\left(c_i(x,t) - c_{p,i}(x,t)\right)\right)\end{aligned} \tag{3.1}$$

$$\frac{\partial c_{p,i}(x,t)}{\partial t} = \frac{3}{r_p}\frac{k_{eff,i}}{\varepsilon_p}(c_i(x,t) - c_{p,i}(x,t)) - \frac{1-\varepsilon_p}{\varepsilon_p}\frac{\partial q_i(x,t)}{\partial t} \tag{3.2}$$

$$\frac{\partial c_i}{\partial x}(0,t) = \frac{u(t)}{\varepsilon_{col}D_{ax}}(c_i(0,t) - c_{in,i}(t)) \tag{3.3}$$

$$\frac{\partial c_i}{\partial x}(L,t) = 0 \tag{3.4}$$

Protein adsorption is modeled using the stoichiometric displacement model [124, 125]. Eq. 3.5 shows the kinetic form of the SDM isotherm, where q_i and $c_{p,i}$ denote the protein concentration in the solid and liquid phase of the particle, respectively. The SDM describes the equilibrium binding behavior of the protein considering the salt concentration in the pore phase c_s, the ionic capacity of the resin Λ, and protein specific parameters. The protein characteristic charge ν_i accounts for the number of charges interacting with the resin, while the constants $k_{eq,i} = k_{ads,i}/k_{des,i}$ and $k_{kin,i} = 1/k_{des,i}$ comprise adsorption and desorption rates of the modeled proteins.

$$k_{kin,i}\frac{\partial q_i}{\partial t} = k_{eq,i}q_{salt}^{\nu_i}c_{p,i} - q_i c_s^{\nu_i} \tag{3.5}$$

$$q_{salt} = \Lambda - \sum_{j=1}^{k}\nu_j q_j \tag{3.6}$$

The aim of the multiscale modeling workflow was the prediction of a mAb's SDM parameters, characteristic charge ν_i and the equilibrium constant $k_{eq,i}$, which are the model parameters defining the linear region of the SMA isotherm. To generate training and testing data, the Yamamoto method was used for the analytical solution of ν_i and $k_{eq,i}$ using a set of LGEs at differing salt gradient slopes for all therapeutic antibodies at varying mobile phase pH values. Eq. 3.7 describes the linear correlation between the normalized gradient slope GH and the elution salt concentration $c_{s,i}$ of component i at diluted loading conditions [74, 166, 172, 173, 196]. Eqs. 3.8 and 3.9 allow to the calculation of the normalized gradient slope GH, where c_s,initial is the salt concentration at the gradient begin, $c_{s,final}$ is the salt concentration at the gradient end, and V_G is the gradient length in mL.

$$log(GH) = (\nu_i + 1)log(c_{s,i}) - log\left(k_{eq,i}\Lambda^{\nu_i}(\nu_i + 1)\right) \tag{3.7}$$

$$g = \frac{c_{s,final} - c_{s,initial}}{V_G} \tag{3.8}$$

$$GH = g(V_{col} - \varepsilon_t V_{col}) \tag{3.9}$$

3.2.4 Quantitative Structure-Property Relationship Modeling

The QSPR is the central building block of the multiscale model and connects molecular-level protein descriptors x with macroscopic adsorption isotherm parameters y. The QSPR modeling workflow consisted of feature selection, machine learning *via* Gaussian process regression (GPR), and model validation. Machine learning and visualization was performed with Python 3.8.10 including scikit-learn [197]. Before feature selection, the data set was split into training and test data using two different sampling techniques. In the first section of this manuscript, 20% of the overall data was randomly removed from the data set for model validation. For external validation in the second part of the manuscript, all isotherm parameters and descriptors associated with a single mAb at different process conditions, fragmentation states, and conformations were removed from the training data set. To avoid over-determination of the GPR, a two-staged feature selection was used to identify protein descriptors x relevant for prediction of isotherm parameters y. Firstly, a filter method was employed to reduce the computational costs of the subsequent recursive feature elimination (RFE). The filter method removed descriptors based on low variance and low mutual information [198, 199] between individual descriptors x and isotherm parameters y of the training data set. The dimensionality of the remaining descriptor set was further reduced using RFE as supervised feature selection method [146]. During each

iteration of the RFE, 30 GPR models were trained using a 5-fold cross validation with 6 repetitions using a sub-set of 80% of the training data. After each iteration, the feature with the weakest contribution to the GPR model was removed from the descriptor vector x [200]. The optimal number of features used to train the final GPR model was determined *via* a scoring function considering the mean absolute error (MAE) calculated during cross-validation and the log-marginal-likelihood (LML) [201] obtained from corresponding GPR models.

Gaussian process regression was used for supervised learning due to its probabilistic nature and wide variety of kernel functions [202]. The general aim of the machine learning method is the prediction of an isotherm parameter y based on a high-dimensional vector x of protein descriptors. A Gaussian process can be defined as generalization of a Gaussian distribution over a vector space to a function space of infinite dimensions [203]. The GPR relies on Bayesian inference, assuming a *prior* probability distribution for the values of the function $y(x)$ and updates the probability distribution in the presence of observed data to yield a *posterior* probability distribution. Following Obrezanova *et al.* [204], the Bayesian update rule is

$$P(y(x)|\mathscr{D}) \propto P(Y|y(x), X)P(y(x)) \tag{3.10}$$

where $\mathscr{D} = \{X, Y\}$ is the training data, $P(y(x)|\mathscr{D})$ describes the *posterior* distribution, $P(y(x))$ is the *prior*, and $P(Y|y(x), X)$ is the likelihood. The covariance of the *prior* was specified by addition of three sub-kernels: a linear term; a non-linear Matérn class kernel; and a white noise kernel [197, 202]. The GPR was fitted by optimizing the hyperparameters Θ of the kernel functions to maximize the LML with the L-BFGS-B algorithm [205].

3.3 Results & Discussion

In this manuscript, a QSPR model is used for *a priori* prediction of elution behavior of mAbs in preparative CEX chromatography. SDM isotherm parameters were determined based on experimental data for a diverse set of therapeutic antibodies, including IgG1 mAbs, IgG4 mAbs, Fabs and bsAbs. Protein descriptors were derived from the corresponding homology models calculated based on the amino acid sequence. For external validation, data of an IgG1 and an IgG4 mAb was removed separately from the QSPR training set. The predicted SDM isotherm parameters were used to simulate entire chromatograms and validated with experimental chromatograms. The developed multiscale modeling pipeline bridged the gap between the protein structure of mAbs and their macroscopic elution behavior.

3.3.1 QSPR Modeling for the Prediction of SDM Isotherm Parameters

For the training and testing data set, SDM model parameters were determined based on CEX experiments at bench-scale. All purification experiments were performed on the identical chromatography column with the strong CEX media POROS XS and an inner diameter of 1 cm, at a linear flow rate of 200 cm/h. Tab. 3.1 shows experimentally determined column-specific parameters necessary for mechanistic modeling. The measured column porosities, ionic capacity, and axial dispersion coefficient were in accordance with values found in literature [114, 163, 174].

Table 3.1: Experimentally determined column-specific model parameters of the POROS XS CEX chromatography column.

Parameter	Symbol	Value	Unit	Equation	Reference
Length	L	136	mM		
Diameter	d	10	mM		
Column Volume	V	10.68	mL	$V = \pi \frac{d^2}{4} L$	
Bead radius	r_p	25	μm		[114]
Interstitial porosity	ε_{col}	0.39	-	$\varepsilon_{col} = \frac{V_{int}}{V}$	[114]
Total porosity	ε_t	0.77	-	$\varepsilon_t = \frac{V_t}{V}$	[114]
Particle porosity	ε_p	0.62	-	$\varepsilon_p = \frac{V_t - V_{int}}{V - V_{int}}$	[114]
Axial dispersion	D_{ax}	0.14	mm^2/s	$D_{ax} = \frac{uLs^2_{NaCl}}{(2V_t)^2}$	[115]
Ionic capacity	Λ	0.41	M	$\Lambda = \frac{c_{NaOH} V_{NaOH}}{V(1-\varepsilon_t)}$	[114, 126]

The thermodynamic equilibrium constant $k_{eq,i}$ and characteristic charge ν_i of 21 mAbs were determined experimentally using the Yamamoto method as described in Eqs. 3.7-3.9. IgG1 and IgG4 mAbs were further digested with Papain to increase the structural diversity of the training data set. Thus, for the 10 IgG1 and the 5 IgG4 mAbs, SDM isotherm parameters were determined for full-length molecules and corresponding FAbs at pH 5.00, pH 5.25, and pH 5.50 respectively. For 5 bsAb and the single Fab, SDM isotherm parameters were only determined for full-length molecules at pH 5.25. The resulting distributions of measured parameters in the training and testing data are shown in Fig. 3.1. For reasons of clarity, values of the equilibrium constant are shown as their natural logarithm $ln(k_{eq,i})$. Equilibrium constants $k_{eq,i}$ in the overall data set ranged from -6.34 to 3.23. Characteristic charge parameters ν_i ranged from 3.8 for a single Fab at pH 5.5 to 13.9 for a bispecific mAb of the IgG(H)-scFv format at pH 5.25. In comprison, Hunt *et al.* [115] reported characteristic charge parameters are typically ranging between 5 and 20 for mAbs on strong (SO_3) and weak (COO) CEX resins. For full-length IgG1 and IgG4 mAbs investigated in this work, characteristic charge parameters had values between 6.2 and 11.79, while ν_i parameters of corresponding Fabs were reduced by approximately 40%. Thus, the bimodal distribution observable in Fig. 3.1 (A) results from fragmentation of IgG1 and IgG4 mAbs. All mAbs had theoretical pI values above the investigated pH conditions of pH 5.00 to pH 5.50. Hence, the data in Fig. 3.1 suggests a relationship of the characteristic charge ν_i with the molecular weight and net charge of investigated antibodies. In general, $k_{eq,i}$ and ν_i parameters increased with decreasing mobile phase pH.

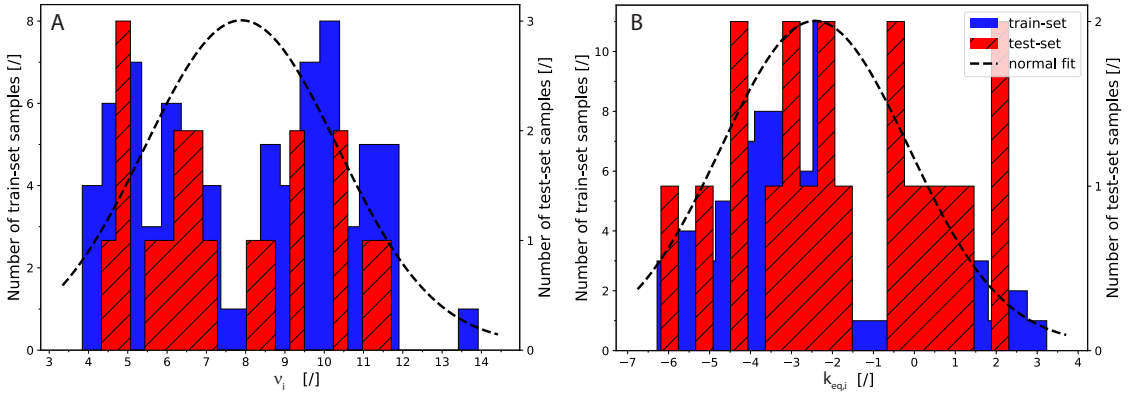

Figure 3.1: Distributions of SDM parameters in the test and training set. Characteristic charge ν_i (A) and equilibrium constant $k_{eq,i}$ (B) were determined *via* the Yamamoto method for 21 IgG1 mAbs, IgG4 mAbs, Fabs, and bsAbs at mobile phase pH conditions of pH 5.00, pH 5.25, and pH 5.5.

LGE experiments used for parameter estimation were performed under diluted loading conditions of 1 g/L_{Resin}. Noteworthy, all IgG4 molecules showed a distinct split-peak elution behavior similar to the results reported by Luo *et al.* [187]. An exemplary split-peak chromatogram of one IgG4 investigated in this study is shown in Fig. 3.5 (B). When collecting one of the peaks and repeating the identical experiment, the split-peak elution behavior was observed again. Double elution peaks where only observed for full-length

IgG4 mAbs and not visible for corresponding Fabs. In order to understand underlying adsorption phenomena, the Yamamoto method was applied separately to both elution peaks of full-length mAbs. This resulted in two pairs of equilibrium constant $k_{eq,i}$ and characteristic charge ν_i for each IgG4 molecule and mobile phase pH. Both parameters, $k_{eq,i}$ and ν_i, affect retention times in gradient elution experiments. Interestingly, $k_{eq,i}$ values were comparable for both peaks of each IgG4 molecule. In contrast, characteristic charge parameters ν_i of early eluting species were 8% to 23% lower compared to the second peak, and thus caused the shift in retention volume.

A recent study performed by Blech *et al.* [192] revealed that IgG4 mAbs can adopt multiple conformational states coexisting in a dynamic equilibrium, including the typical Y-shaped conformation and a self-associated λ-conformation where one Fab is directed towards the Fc-portion of the mAb. Further, Blech *et al.* reported the x-ray crystal structure off the intact, full-length IgG4 mAb in its λ-conformation (PDB=6GFE) [192]. Based on the two published structure templates 5DK3 and 6GFE, it was possible to consider the conformational diversity of IgG4 mAbs during homology modeling and the subsequent descriptor calculation. Exemplary homology models of a IgG4 in Y- and λ-conformations are depicted in Fig. 3.2 (D) and (E), respectively. Global and local descriptors were derived from homology models of the 21 full-length antibodies and used as independent variables x in the subsequent GPR QSPR modeling workflow. Global descriptors, such as the surface net charge, of identical IgG4 mAbs in different conformations changed significantly due to the reduced solvent accessible surface area (SASA) of the self-associated λ-conformation compared to the open Y-conformation. In contrast, local descriptors associated with the variable region of individual IgG4 mAbs were not affected by conformational diversity. Due to the lower protein net charge of λ-conformations and lower measured ν_i parameters of the early eluting peak, we hypothesize that the split-peak phenomenon is caused by the reversible, conformational diversity of IgG4 mAbs as reported by Blech *et al.* [192]. Thus, the early eluting peak represents the IgG4 λ-conformation and the second peak the Y-conformation. To build a predictive QSPR model, SDM parameters y of the first peak were combined with protein descriptors x of the λ-conformation and SDM isotherm parameters of the second peak were combined with descriptors of the Y-conformation. The hypothesis was tested in the second result section, by an external validation including the prediction of an unknown IgG4 split-peak chromatogram.

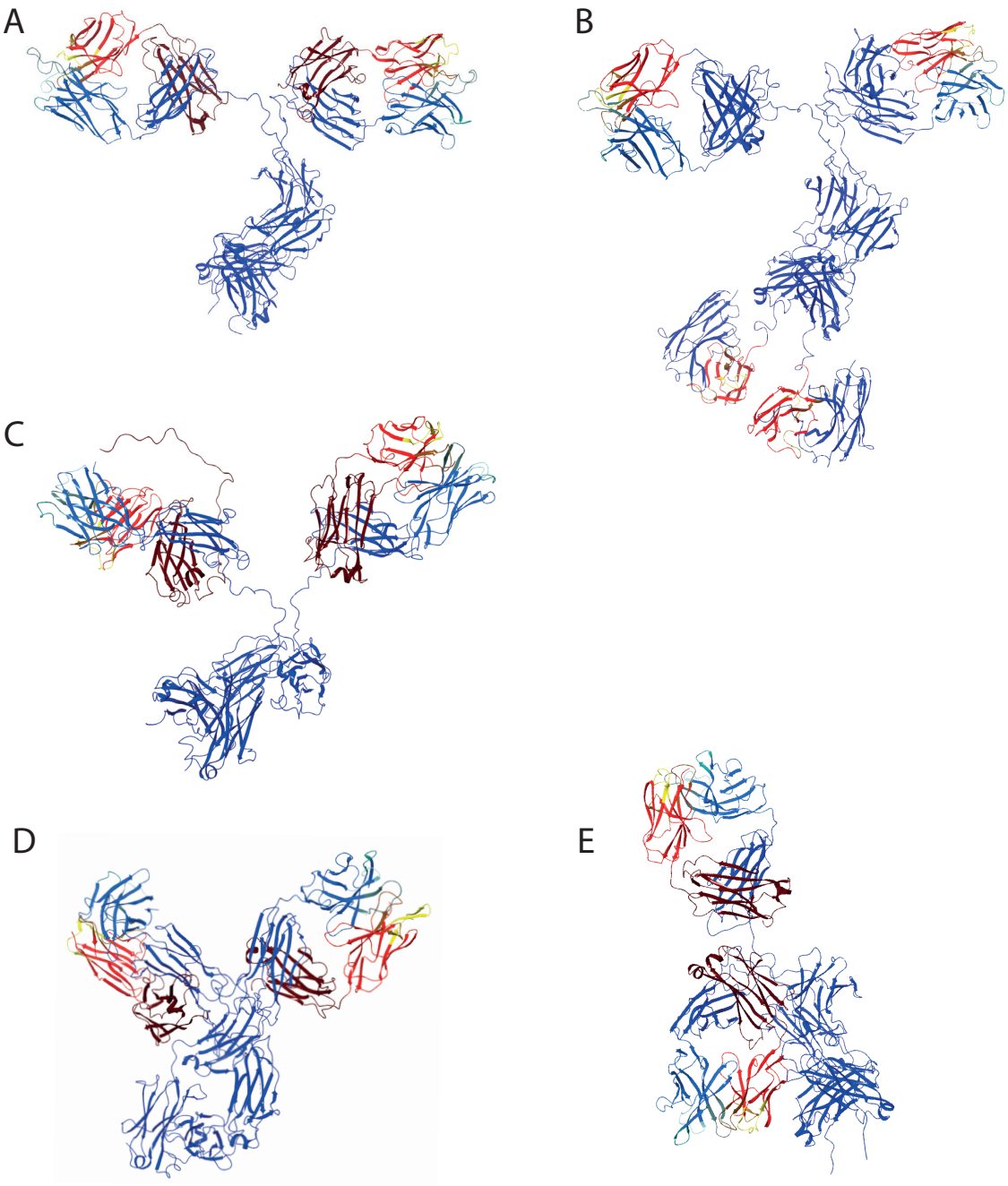

Figure 3.2: Template structures for homology modeling of full-length IgG1, IgG4 mAbs and bsAbs. (A) displays the IgG1 mAb template, a modified version of the PDB entry 1HZH [194]. 1HZH was also used as source structure for both bsAbs formats, the appended IgG(H)-scFv (B) and the knob-in-hole format (C). Two discrete conformations were considered for the homology modeling of IgG4 mAbs. (D) = IgG4 Y-conformation (PDB=5DK3) [195]. (E) = IgG4 λ-conformation (PDB=6GFE) [192]

The QSPR models depicted in Fig. 3.3 represent the central building block of the multiscale modeling pipeline. Gaussian process regression was selected as machine learning algorithm that maps between the descriptor vector x and the macroscopic adsorption isotherm parameters y. The two-stage feature selection aimed to avoid over-determination of the GPR model. Results of the feature elimination can give deeper insights into binding orientation and dependencies of SDM isotherm parameters on protein structure descriptors. The initial filtering method removed descriptors from the normalized feature vector x showing < 0.01 variance across the training data set. Subsequently, assessment of the mutual information between descriptors x and isotherm parameters y in the training set resulted in the elimination of 97% of the 1065 descriptors. Global descriptors, such as the net charge, the hydrodynamic radius, and the total aromatic SASA, shared the highest mutual information with the characteristic charge parameter ν_i. The thermodynamic equilibrium constant $k_{eq,i}$ shared the highest mutual information with global as well as local, charge and aggregation propensity estimating descriptors associated with the framework region (FR) and the complementarity-determining regions (CDR). 15-20 descriptors remained for the second level of feature selection based on RFE. Fig. 3.3 (A) and (B) depicts the protein structural descriptors selected during supervised feature elimination. During each iteration of RFE, a multivariate GPR was trained, cross-validated, and the feature with the lowest permutation importance was removed until the optimal number of features was reached. Based on the MAE of the GPR cross-validation, n=5 descriptors was determined as optimal number of features to train the final GPR models for characteristic charge and equilibrium constant. When comparing the changes of cross-validation MAEs of both SDM isotherm parameters in Fig. 3.3 (A) and (B), the model prediction for the equilibrium constant is 50% less accurate compared to the characteristic charge. Further, Fig. 3.3 (B) indicates that the predictive power of the GPR model for $k_{eq,i}$ significantly decreases when removing descriptors estimating the sum positive surface patch energy in the heavy-chain CDRs and the FR atomic contact energy descriptor. Based on the selected descriptors, it can be assumed that the variable region of mAbs is involved in protein adsorption and "Fab-first" is the preferred binding orientation.

The final predictions of GPR models with a 20% randomly selected test set are plotted in Fig. 3.3 (C) and (D). Comparably high correlation coefficients of $R^2 > 0.99$ as well as the uniformly dispersed residuals of training and testing data, indicate an adequate accuracy of the model predictions for both SDM isotherm parameters. The comparably wide 95% confidence intervals in the testing set of $k_{eq,i}$ suggest a higher prediction uncertainty of the equilibrium constant compared to the characteristic charge parameter. Further, the $k_{eq,i}$ test predictions with the furthest distance to the ideal prediction line also had the widest 95% CIs. This indicates that the GPR model could successfully estimate the *posterior* probability distribution of $y(x)$ based on the underlying data that was used for model training. The higher model uncertainty of $k_{eq,i}$ compared to ν_i could be a consequence of inaccurate predictions of CDR loop conformations during homology modeling and subsequent calculation of local descriptors. The QSPR model for the characteristic charge parameter was exclusively trained with global protein descriptors, which are less dependent on the challenging prediction of hyper-flexible CDR loop structures. Recent advances in protein structure prediction based on deep neural networks [54] could possibly improve

homology models for mAb CDRs and resulting multiscale model predictions. However, the predictive power of the QSPR models in Fig. 3.3 is sufficient for simulation of elution profiles, as shown in the following section.

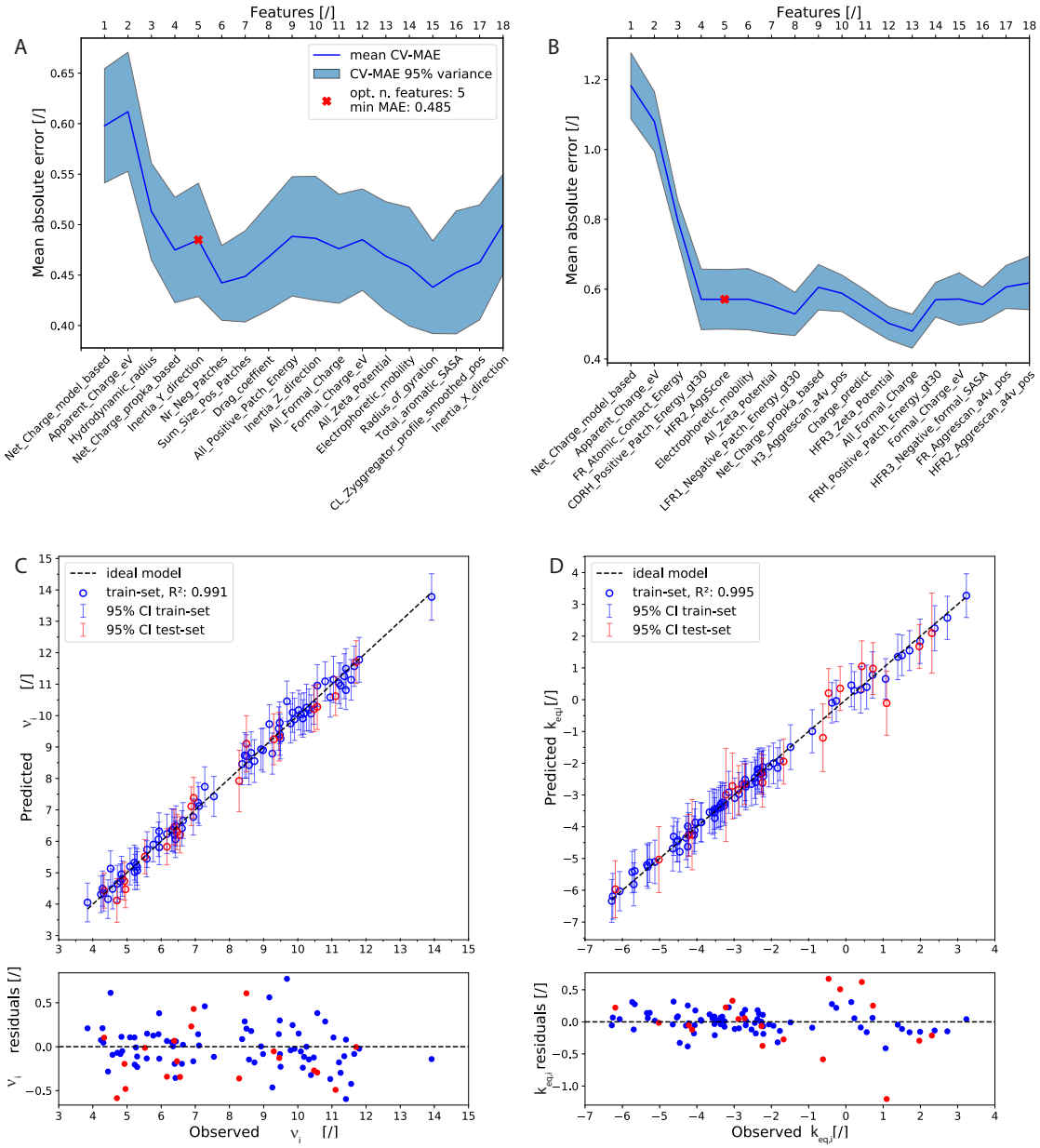

Figure 3.3: Recursive feature elimination and QSPR model for prediction of SDM parameters of mAbs on POROS XS at pH 5.00, 5.25, and 5.50. (A) and (B) show the results of feature selection for characteristic charge ν_i and equilibrium constant $k_{eq,i}$ respectively. The corresponding QSPR models for the prediction of ν_i (C) and $k_{eq,i}$ (D) of a randomly selected test set representing 20% of the overall data.

The ability to calculate probability distributions on model predictions is a main benefit of Gaussian process regression as machine learning method for QSPR modeling. This is especially relevant for multiscale modeling tasks as presented in this work, where the propagation of model uncertainty from homology to mechanistic modeling could affect prediction accuracy. Feature selection and QSPR modeling revealed quantitative relationships between the structure of therapeutic antibodies and their macroscopic adsoprtion model parameters. These relationships were also suggested by a previous work of our group, where single amino acid substitution in the CDR of mAbs had an significant impact on $k_{eq,i}$ parameters [185]. The three IgG1 mAbs with single amino acid substitution were also included in the present data set of 21 mAbs, increasing the predictive power of the QSPR model for $k_{eq,i}$. It is important to consider that the random sampling method applied in Fig. 3.3 potentially included identical molecules in training and testing data sets, but at differing process conditions or fragmentation states. Therefore, the trained GPR model in Fig. 3.3 could be used to simulate process behavior of a mAb at an unknown mobile phase pH in the range of pH 5.00 to pH 5.50, if an initial experiment at a different pH value was already performed. The following chapter explores the possibility to predict entire chromatograms of a new antibody candidate without conducting initial experiment during early-stage process development.

3.3.2 External Validation and Prediction of Chromatograms

For external validation, all data points associated with a specific mAb product must be removed from the training and cross-validation data set. Consequently, for one IgG1 and one IgG4 mAb, all model parameters at varying pH values, fragmentation states, and conformations were removed from the training data set. Fig. 3.4 shows the QSPR models for the prediction of ν_i and $k_{eq,i}$ with individually selected testing data for the IgG1 (Fig. 3.4 A and B) and the IgG4 mAb (Fig. 3.4 C and D). Similar to the randomly sampled QSPR model, the QSPR models in Fig. 3.4 enabled an acceptable prediction accuracy for both isotherm parameters. However, the wider CIs of equilibrium constants in Fig. 3.4 compared to the ones in Fig. 3.3 indicate that predictions for unknown mAbs are more challenging compared to isotherm parameters of mAbs that were already investigated at differing pH conditions, conformations, or fragmentation states. The ultimate goal of the multiscale modeling workflow is the simulation of elution profiles. Fig. 3.5 compares measured CEX chromatgrams with simulations based on the chromatography model built with SDM parameters predicted for the IgG1 and IgG4 mAbs shown in Fig. 3.4. For both, IgG1 and IgG4 split-peak, accurate predictions of UV curves were achieved. Two protein species were considered for simulation of the IgG4 elution profile in Fig. 3.5 (B). The first peak represents the self-associated IgG4 λ-conformation that showed a reduced theoretical surface charge compared to the Y-conformation that elutes in the second peak. The dynamic equilibrium between λ- and Y-conformation discovered by Blech *et al.* [192] could explain why the split-peak phenomenon is reproducible when collecting the peaks separately and repeating the experiment [186, 187]. Due to the semi-mechanistic nature of the SDM isotherm, model predictions are not limited to mobile phase conditions applied

in wet-lab experiments shown in Fig. 3.5. Hence, the model could be used to test varying gradient slopes, pH values, or step elution salt concentrations to define an initial process design space before protein material is available. In contrast to the mechanistic chromatography model, the GPR-based QSPR model is based on empirical relationships between protein descriptors x and macroscopic isotherm parameters y. Therefore, extrapolations towards molecule formats not included in the training set are not feasible. Continuous improvement of QSPR predictions can be achieved by retraining the GPR model when a new biological entity enters the developmental phase.

Coupling of homology modeling, QSPRs, and mechanistic modeling led from the amino acid sequence to an initial mechanistic model allowing prediction of elution profiles at varying mobile phase conditions in gradient and step elution mode. In 2005, Ladiwala *et al.* [31] introduced the first multiscale QSPR model enabling the prediction of chromatograms based on descriptors derived from published crystal structures. To the best of our knowledge, the present work is the first to predict macroscopic isotherm parameters based on protein descriptors derived from mAb homology models. This is crucial for applications in biopharmaceutical industry, since experimentally determined x-ray crystal structures of full-length molecules are not available during product development. Currently the QSPR model is limited to the linear region of the SMA isotherm. An extension towards the non-linear region would demand additional wet-lab experiments at higher loading density. Nonetheless, ν_i and $k_{eq,i}$ are typically estimated *via* multiple time-consuming LGE experiments at low loading densities, which are not performed during traditional process development workflows. A non-linear SMA isotherm model could be achieved by combining the presented QSPR model with inverse model calibration using a single experiment at higher loading densities [167]. Additionally, fraction analysis with off-line analytical measurements could be used to include charge- and size-variants of the mAb.

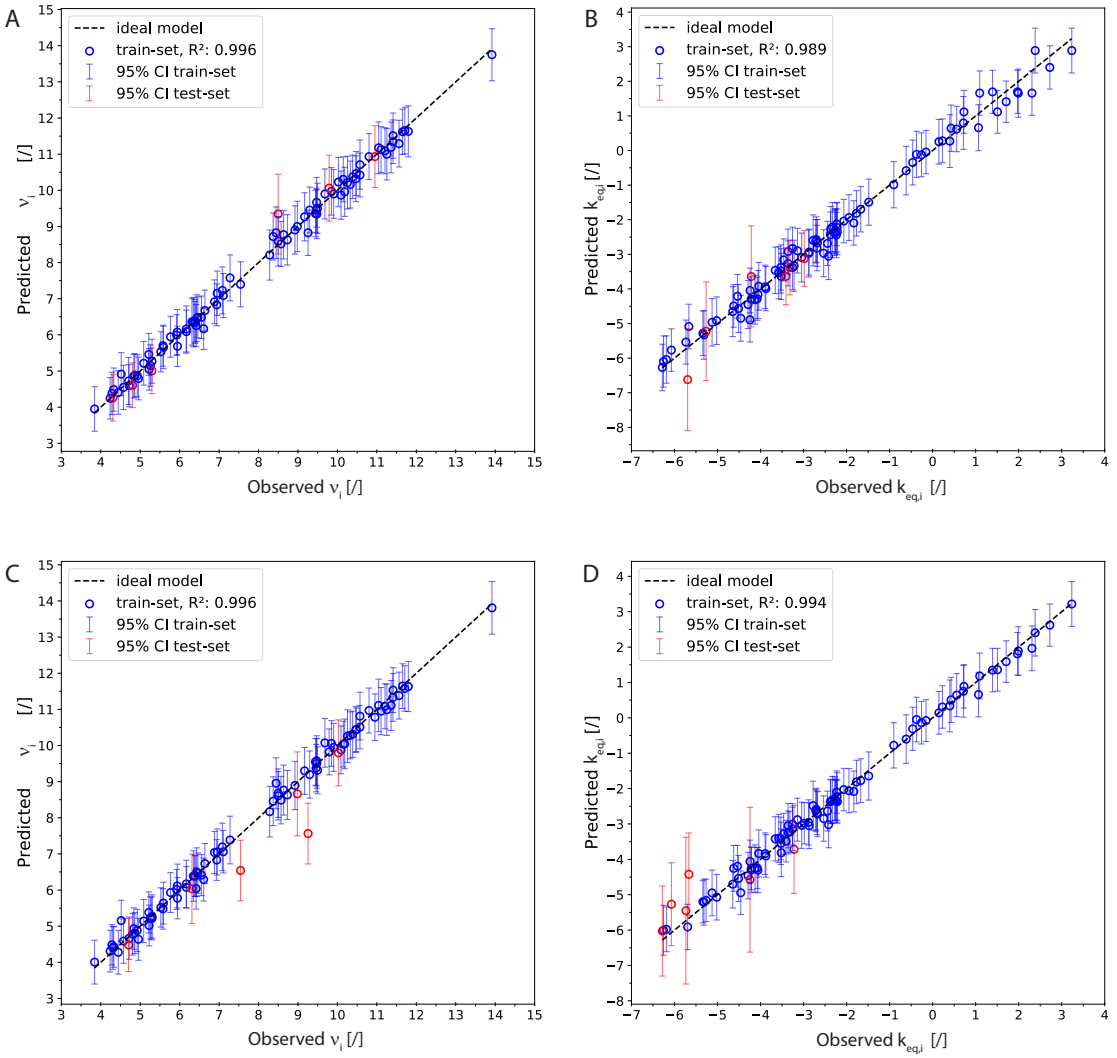

Figure 3.4: External validation of the QSPR model for one IgG1 and one Ig4 mAb. The upper plots show the QSPR models for the prediction of characteristic charge ν_i (A) and equilibrium constant $k_{eq,i}$ (B) of a full-length IgG1 mAb and corresponding Fab at pH 5.00, 5.25, and 5.50. The identical method was applied in (C) and (D) showing QSPR models for prediction of SDM parameters of a full-length IgG4 mAb in Y- and λ-conformation and corresponding Fab.

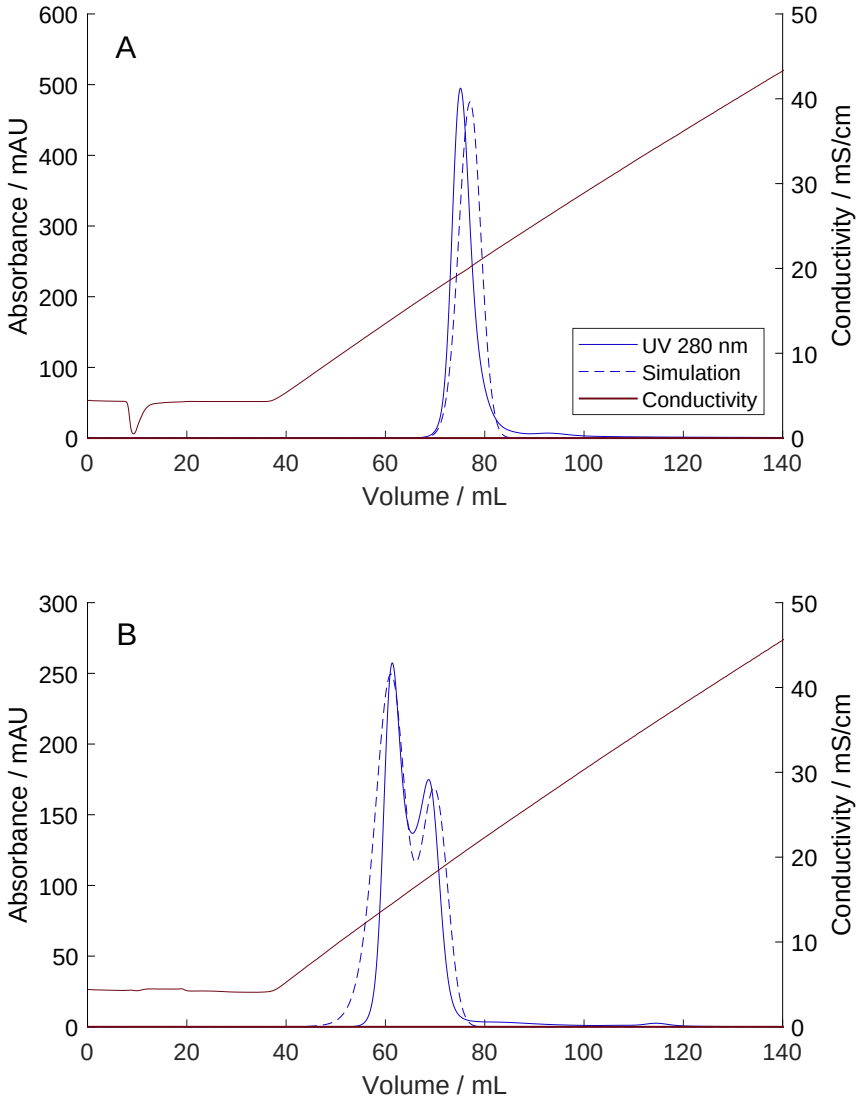

Figure 3.5: Measured and predicted chromatograms of IgG1 (A) and IgG4 (B) mAb on POROS XS at pH 5.25, 10 CV gradient slope, 200 cm/h linear flow rate. The corresponding SDM isotherm parameters for the model prediction are shown in Fig 3.4. Two protein species representing the Y-conformation and the λ-conformation were used to simulate the IgG4 split peak phenomenon.

3.4 Conclusion

The present study introduced a multiscale model leading from the amino acid sequence of a therapeutic antibody to the mechanistic model of its preparative CEX chromatography process. The data set included IgG1 mAbs, IgG4 mAbs, FAbs, and bispecific antibody formats at different pH conditions. Two molecules, one IgG1 and one IgG4, were individually removed from the training data set for external validation. The GPR-based QSPR model predicted the SDM isotherm parameters of the test set molecules with an accuracy that enabled the simulation of CEX chromatograms. Good agreement between simulated and measured chromatograms was observed for both test molecules. Noteworthy, the multiscale model could predict complex split-peak elution curves of the IgG4 mAb by considering two discrete conformations coexisting in a dynamic equilibrium. The closed λ-conformation of the IgG4 mAb had a reduced positive surface charge and characteristic charge ν_i compared to the Y-conformation leading to an early elution during gradient elution experiments. The two-staged feature selection method *via* regression-based filtering and recursive feature elimination effectively avoided over-fitting. Further, the final protein descriptors selected for GPR modeling gave insights into relationships between antibody structure and macroscopic adsorption isotherm parameters. QSPR models for the prediction of the thermodynamic equilibrium constant were highly dependent on local charge-specific descriptors in the variable region of the mAb.

The gained knowledge on antibody adsorption to CEX resins could be used to select mAb candidates that integrate into the DSP platform. During early-stage process development, the multiscale model allows simulation of different process conditions and elution modes before protein material is available. Thus, early *in silico* optimization based on mechanistic modeling coupled with QSPR modeling could reduce the experimental burden and shorten the time from antibody drug discovery to start of phase 1 clinical investigations. GPR was identified as suitable machine learning method for prediction of macroscopic model parameters. In general, the multiscale modeling method could be extended to other target properties relevant for CMC development. Adding parameters defining the non-linear region of the SMA isotherm, e.g. steric shielding, could further increase the relevance of the applied multiscale modeling workflow for chromatography modeling. If sufficient data is available, the model could further be used for prediction of mAb charge heterogeneity, aggregation propensity, or process behavior in other unit operations of the DSP platform. If the fundamental relationships between antibody structure and process behavior are uncovered across the entire DSP platform, wet-lab experiment for CMC development could be reduced to a minimum.

3.5 Acknowledgment

The authors would like to acknowledge experimental and scientific support of their colleagues within the DSP and PDB departments at Boehringer Ingelheim.

4 Straightforward Method for Calibration of Mechanistic Cation Exchange Chromatography Models for Industrial Applications

David Saleh[1, 2], Gang Wang[2], Benedict Müller[2], Federico Rischawy[2], Simon Kluters[2], Joey Studts[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] Boehringer Ingelheim Pharma GmbH & Co. KG, Biberach, Germany
* Corresponding author: Jürgen Hubbuch, +49 721 608 47526, juergen.hubbuch@kit.edu

Abstract

Mechanistic modeling of chromatography processes is one of the most promising techniques for the digitalization of biopharmaceutical process development. Possible applications of chromatography models range from *in silico* process optimization in early phase development to *in silico* root cause investigation during manufacturing. Nonetheless, the cumbersome and complex model calibration still decelerates the implementation of mechanistic modeling in industry. Therefore, the industry demands model calibration strategies that ensure adequate model certainty in a limited amount of time. This study introduces a directed and straightforward approach for the calibration of pH-dependent, multicomponent steric mass-action (SMA) isotherm models for industrial applications. In the case investigated, the method was applied to a monoclonal antibody (mAb) polishing step including four protein species. The developed strategy combined well-established theories of preparative chromatography (e.g. Yamamoto method) and allowed a systematic reduction of unknown model parameters to 7 from initially 32. Model uncertainty was reduced by designing two representative calibration experiments for the inverse estimation of remaining model parameters. Dedicated experiments with aggregate-enriched load material lead to a significant reduction of model uncertainty for the estimates of this low-concentrated

product related impurity. The model was validated beyond the operating ranges of the final unit-operation, enabling its application to late-stage downstream process development. With the proposed model calibration strategy, a systematic experimental design is provided, calibration effort is strongly reduced and local minima are avoided.

4.1 Introduction

Preparative chromatography is essential for the downstream processing (DSP) of biopharmaceuticals. Conventional purification platforms for monoclonal antibodies (mAb) contain two to three chromatographic unit operations [6, 7]. In the light of the Quality by Design (QbD) guidelines [24], high product quality has to be ensured by understanding the complex elution behavior observed in nonlinear chromatography [164]. The physical understanding provided by mechanistic models meets this demand for an in-depth process knowledge. During the last years, academia and industry showed great effort to realize an *in silico* DSP development based on mechanistic chromatography modeling. Numerous publications presented successful applications of mechanistic models for process optimization,[16, 98, 114, 163] process characterization [101, 162, 206], root-cause investigation [100], and for the development of process control strategies [207]. Nonetheless, the industry is still cautious with the implementation of mechanistic modeling in DSP development. Technically, the bottleneck is the time-consuming and complex model calibration.

Model calibration can be achieved using miniaturized high-throughput[208–210] or lab-scale experiments. For the stoichiometric displacement model (SDM), linear gradient elution (LGE) experiments under diluted load conditions enable the estimation of the proteins characteristic charge and the equilibrium constant [74, 117, 173, 211, 212]. Rüdt *et al.* [166] introduced the so-called combined Yamamoto method for the simultaneous estimation of isotherm and mass transfer parameters. Wang *et al.* [190] proposed to estimate the SDM parameters using artificial neural networks. Sequence and structural information allow the calculation of SDM parameters for model proteins [31] and mAb [74]. Regarding nonlinear binding conditions, the shielding parameter of the steric mass-action (SMA) isotherm can be determined by frontal analysis using the correlation developed by Osberghaus *et al.* [171]. A successful application of the techniques mentioned above for multi-component systems in the nonlinear region of the adsorption isotherm has –to the best of our knowledge- not yet been published.

In recent publications and industrial applications, the estimation of model parameters using the inverse method is the preferred practice [114, 167, 171], which is based on fitting the model output to experimental data. This iterative procedure enables the simultaneous estimation of multiple parameters without prior knowledge on the parameter value and is not limited to linear load conditions. Additionally, the computational effort and the success of the inverse model calibration depends on the number of model parameters to be estimated, the experiments chosen, and the optimization algorithm. Industry-relevant chromatography models, being able to describe pH-dependent and multicomponent sys-

tems, can lead to ill-posed estimation problems [213, 214] with dozens of model parameters unknown. Their simultaneous estimation by curve fitting can result in local minima and unreasonable parameter estimates. Briskot and coworkers were aware of this problem and proposed to use principles of Bayesian statistics for model calibration based on a small set of experiments with limited variation [165, 214]. Recently, Rischawy *et al.* [163] introduced principles of good modeling practice (GMoP) for industrial ion-exchange chromatography. The GMoP approach mitigated the risks related to the inverse model calibration by parameter subset selection, evaluation of confidence intervals, visual sensitivity analysis, and model validation across multiple scales.

In the presented work, a novel straightforward model calibration strategy was proposed for the calibration of multicomponent SMA models within industrial operating ranges. A case study showed this strategy circumventing the pitfalls of an entirely curve fitting based model calibration method. A representative lab-scale data set consisting of multiple linear gradient elution experiments (LGE) in the linear region and two experiments in the nonlinear region of the adsorption isotherm was used to calibrate a mechanistic chromatography model consisting of a lumped rate model and a pH-dependent SMA isotherm. The application of the inverse method was kept to a minimum and enabled the parameter estimation for the mAb subspecies and the parameters in the nonlinear region of the SMA isotherm. Dedicated experiments for low-concentrated product related impurities further reduced model uncertainty and ensured the model's predictive power. Comprehensive validation of the calibrated model was achieved using twelve experiments beyond calibration space. The present study aims to streamline the existing calibration techniques towards a standardized method for building CEX chromatography models with a predictive power fulfilling the requirements of advanced applications in biopharmaceutical process development.

4.2 Model Design

4.2.1 Mathematical Model

The transport dispersive model is a commonly applied column model for the simulation of preparative ion exchange chromatography [99, 116, 167, 190]. The system is of convection diffusion reaction type and describes the macroscopic transport of component i through the column. The change of the concentration $c_i(x,t)$ described in Eq. 4.1 depends on the convective mass transport in the interstitial volume of the packed bed with the superficial velocity u. D_{ax} denotes the axial dispersion coefficient and describes the peak broadening effect caused by molecular diffusion in the interstitial volume. The interfacial mass transfer between the interstitial volume and the particle pores depends on the column porosity ϵ_{col}, the component specific effective mass transfer coefficient $k_{eff,i}$ and the particle radius r_p. In the lumped rate model, film diffusion effects in the particle boundary layer and the pore diffusion in the particle phase are combined to $k_{eff,i}$. Eq. 4.2 describes the accumulation

of mass in the pore volume c_i and the stationary phase q_i. The model is complemented with the Danckwerts boundary conditions in Eqs. 4.3 and 4.4.

$$\begin{aligned}\frac{\partial c_i(x,t)}{\partial t} = &-\frac{u}{\varepsilon_{col}}\frac{\partial c_i(x,t)}{\partial x} + D_{ax}\frac{\partial^2 c_i(x,t)}{\partial x^2} \\ &- \frac{1-\varepsilon_{col}}{\varepsilon_{col}}\left(\frac{3}{r_p}k_{eff,i}\left(c_i(x,t) - c_{p,i}(x,t)\right)\right)\end{aligned} \tag{4.1}$$

$$\frac{\partial c_{p,i}(x,t)}{\partial t} = \frac{3}{r_p}\frac{k_{eff,i}}{\varepsilon_p}(c_i(x,t) - c_{p,i}(x,t)) - \frac{1-\varepsilon_p}{\varepsilon_p}\frac{\partial q_i(x,t)}{\partial t} \tag{4.2}$$

$$\frac{\partial c_i}{\partial x}(0,t) = \frac{u(t)}{\varepsilon_{col} D_{ax}}(c_i(0,t) - c_{in,i}(t)) \tag{4.3}$$

$$\frac{\partial c_i}{\partial x}(L,t) = 0 \tag{4.4}$$

The protein adsorption to the resin is modeled with the SMA isotherm [117]. Eq. 4.5 shows the kinetic form of the SMA isotherm modified by Hahn *et al.* [167], where q_i and $c_{p,i}$ denote the concentration in the solid and liquid phase of the particle, respectively. The SMA isotherm formulates the equilibrium binding behavior of the protein in consideration of the counter-ion concentration in the pore phase c_s and the ionic capacity of the resin λ. Adsorption behavior depends on protein specific parameters. ν_i accounts for the proteins characteristic charge interacting with the resin. The steric shielding σ_i considers the number of functional groups on the resin blocked by the protein due to steric hindrance. Adsorption and desorption rates are considered by the constants $k_{eq,i} = k_{ads,i}/k_{des,i}$ and $k_{kin,i} = 1/k_{des,i}$, respectively. In this formulation, $k_{eq,i}$ has a strong influence on the retention time of the elution peak, while $k_{kin,i}$ affects the peak width [167].

$$k_{kin,i}\frac{\partial q_i}{\partial t} = k_{eq,i}(pH)\left(\Lambda - \sum_{j=1}^{k}\left(\nu_j(pH) + \sigma_j\right)q_j\right)^{\nu_i(pH)} c_{p,i} - q_i c_s^{\nu_i(pH)} \tag{4.5}$$

$$q_{salt} = \Lambda - \sum_{j=1}^{k}\nu_j(pH)q_j \tag{4.6}$$

For industrial application, the introduction of pH-dependent isotherm parameters is essential. pH-dependencies within the SMA isotherm framework can be approximated empirically [115] or mechanistically [127–129]. In this study, the pH-dependencies of the characteristic charge ν_i and the equilibrium constant $k_{eq,i}$ were approximated with the regression models developed by Hunt *et al.* [115], Eqs. 4.7 and 4.8. The regression models are assumed to be sufficiently accurate for the process relevant pH range of pH 5.8 ± 0.3 used in this study.

$$k_{eq,i}(pH) = k_{eq0,i} e^{k_{eq1,i} pH + k_{eq2,i} pH^2}, \tag{4.7}$$

$$\nu_i(pH) = \nu_{0,i} + pH\nu_{1,i}, \tag{4.8}$$

4.2.2 Estimation of Model Parameters

Figure 4.1 shows the model calibration strategy introduced in this work leading to a multicomponent, pH-dependent SMA chromatography model. Following the GMoP concept [163], the model calibration experiments were designed according to the requirements of future applications. Process parameters, such as mobile phase pH, salt concentration, and loading density were selected based on the final unit operation enabling the model to be applied to *in silico* process characterization studies in late-stage DSP development. In the first step, fundamental knowledge on preparative chromatography was used for the *a priori* estimation of the effective mass transfer parameter $k_{eff,i}$, and the linear SMA isotherm parameters ν_i and $k_{eq,i}$. For calibration experiments used for inverse parameter estimation, fractions were collected with a volume of 0.5 CV. It is important to notice, that non-ideal elution behavior [215] could be encountered during high load density experiments depending on the protein and the selected chromatography resin. The presented calibration approach leads to an early identification of unusual peak shapes enabling model discrimination. However, model complexity has to be selected based on the requirements of the following applications. The model's predictive power for yield, elution volume and purity are the key requirements for performing *in silico* process characterization. Therefore, peak shapes above the 3000 mAU measurement limit were neglected during inverse parameter estimation. Estimation of HMW isotherm parameters was facilitated using an aggregate enriched load material combined with the Yamamoto method. This method circumvented model uncertainty arising from the inverse estimation of very low concentrated product related impurities (<0.4%). For antibody products with a higher HMW content in the load material, or a more sensitive HMW quantification method, it could be sufficient to estimate HMW parameters using the inverse method. In the last step, the pH-dependency of ν_i and $k_{eq,i}$ of the monomer and HMW species was derived from LGEs under load conditions in the linear range of the adsorption isotherm. In the investigated case study, the pH-dependency of all monomeric charge variants was calculated based on Yamamoto correlations of the single monomer species at two pH values around the set-point pH 5.8.

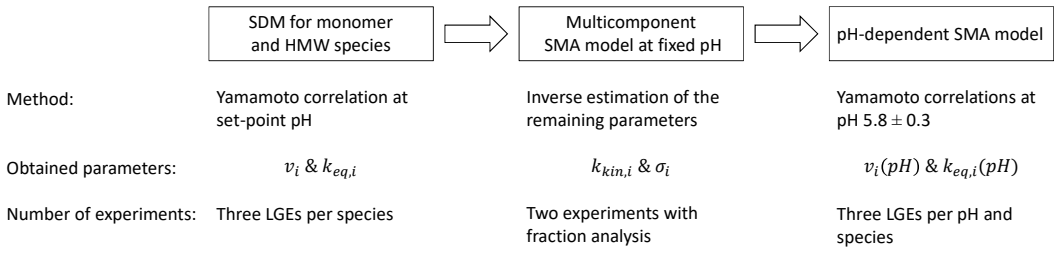

Figure 4.1: Calibration strategy for multicomponent, pH-dependent SMA chromatography models.

The Yamamoto correlation allows the estimation of ν_i and $k_{eq,i}$ using a set of LGEs at different salt gradient slopes. Eq. 4.9 shows the linear relationship between the normalized gradient slope GH and the elution salt concentration $c_{s,i}$ at the peak maximum of component i in the linear range of the adsorption isotherm [74, 166, 171–173]. The UV signal at 280 nm wavelength was approximated to an exponentially modified Gauss function for the determination of the peak maxima and the corresponding elution salt concentration. Eqs. 4.10 and 4.11 lead to the calculation of the normalized gradient slope GH where $c_{s,initial}$ is the salt concentration at the gradient begin, $c_{s,final}$ is the salt concentration at the gradient end and V_G is the gradient length in mL.

$$log(GH) = (\nu_i + 1)log(c_{s,i}) - log\left(k_{eq,i}\Lambda^{\nu_i}(\nu_i + 1)\right) \tag{4.9}$$

$$g = \frac{c_{s,final} - c_{s,initial}}{V_G} \tag{4.10}$$

$$GH = g(V_{col} - \varepsilon_t V_{col}) \tag{4.11}$$

The nonlinear isotherm parameters $k_{kin,i}$, σ_i were estimated using the inverse method implemented in the ChromX software environment by Hahn *et al.* [114, 167]. Linear isotherm parameters of the charge isoforms were estimated inversely with boundaries derived from the Yamamoto correlation. The estimation of an unknown parameter set $\bar{p}$ solves the least square optimization problem

$$\min_{\bar{p}} \sum_j \left(m(t_j) - \sum_{i \geq 1} c_i\left(L, t_j; \bar{p}\right) a_i \right)^2 \tag{4.12}$$

where $m(t_j)$ is the UV measurement of the chromatogram at point t_j, given in mAU. The molar or mass concentration $c_i(L, t_j)$ at the outlet of the column with the length L is transformed into absorbance units with the scaling factor a_i. The scaling factor was calculated according to Beer's law under consideration of the extinction factor and UV cell path length. Parameter uncertainty and confidence ellipsoids were calculated based on the covariance matrix according to Rischawy *et al.* [163].

4.3 Material and Methods

4.3.1 Resin, Buffers and Protein

The mAb polishing step was performed on the strong cation exchanger POROS 50 HS (Thermo Fisher Scientific, Waltham, USA). Tracer injections with blue dextran and 1 M sodium chloride (both Sigma-Aldrich, St. Louis, USA) enabled the calculation of the interstitial volume and total liquid volume, respectively. The ionic capacity λ was determined by acid-base-titration [114, 126]. Tracer injections and acid-base-titration were conducted as triplicates. All column specific parameters and the equations used for their calculation are listed in Table 4.2. Chemicals used in this study were of pharmaceutical grade. All buffers were prepared with deionized water and filtered with a 0.2 μm sterile filter. Table 4.1 shows pH values and counter-ion concentrations of the buffers and load materials used for the calibration (C1-C8) and validation experiments (V1-V12). For column regeneration and storage, 1 M and 0.1 M sodium hydroxide were used, respectively. The model protein (Boehringer Ingelheim GmbH & Co. KG, Biberach, Germany) is an IgG1 monoclonal antibody expressed in stably transfected Chinese hamster ovary (CHO). mAb1 was captured via Protein A affinity chromatography. Titration of the acidic Protein A eluate to pH 5.5, pH 5.8 or pH 6.1 using 1 M acidic acid resulted in the CEX load material. For the calibration experiments C6-C8, a load material with enriched aggregate content was used.

Table 4.1: Summary of all calibration (C) and validation (V) experiments used in this study. Counter-ion concentration in mM represent the sum of sodium ions and deprotonated Tris at the given pH value. *: Experiments were conducted with HMW enriched load material containing 39.2% HMW.

No.	Elution mode	Purpose	pH	Salt concentration cs [mM]			LGE length [CV]	Load density [g/L]
				Equi./ Wash	Load	Elution		
C1	LGE	Inverse estimation	5.8	50	82	50-500	30	45
C2	Step	Inverse estimation	5.8	87	90	217	-	10
C3	3 LGEs	Yamamoto monomer	5.5	50	82	50-500	10, 20, 30	1
C4	3 LGEs	Yamamoto monomer	5.8	50	82	50-500	10, 20, 30	1
C5	3 LGEs	Yamamoto monomer	6.1	50	82	50-500	10, 20, 30	1
C6*	3 LGEs	Yamamoto HMW	5.5	50	50	50-500	10, 20, 30	0.05
C7*	3 LGEs	Yamamoto HMW	5.8	50	50	50-500	10, 20, 30	0.05
C8*	3 LGEs	Yamamoto HMW	6.1	50	50	50-500	10, 20, 30	0.05
V1	LGE	Validation	5.8	50	82	50-500	30	10
V2	LGE	Validation	5.8	50	82	50-500	50	10
V3	Step	Validation	5.8	87	90	247	-	45
V4	Step	Validation	5.8	87	90	277	-	10
V5	LGE	Validation	5.8	50	82	50-500	15	1
V6	LGE	Validation	5.8	50	82	50-500	25	1
V7	LGE	Validation	5.5	50	82	50-500	30	45
V8	Step	Validation	5.5	87	82	247	-	10
V9	Step	Validation	5.5	87	82	217	-	10
V10	LGE	Validation	6.1	50	82	50-500	30	45
V11	Step	Validation	6.1	87	82	247	-	10
V12	Step	Validation	6.1	87	82	217	-	10

4.3.2 Instruments and Software

Calibration and validation experiments were performed on the preparative chromatography system ÄKTA Avant 25. The system was controlled *via* Unicorn 7 (both GE Healthcare, Uppsala, Sweden). The NanoDropTM 2000c (Thermo Fisher Scientific, Waltham, USA) spectrophotometer was used for offline UV measurements. Analytical high performance size exclusion chromatography (HPSEC) and strong cation exchange chromatography (HPSCX) were carried out at the high performance liquid chromatography system Waters Alliance 2695. Analytical chromatograms were integrated in the Empower Software (all Waters, Eschborn, Germany). All computational work was performed on a custom workstation (Lenovo, Stuttgart, Germany) with an Intel® Xeon® Platinum 8160T CPU (48 logical threads) and 128 GB of installed memory (RAM). The chromatography simulation software ChromX (GoSilico GmbH, Karlsruhe, Germany) was used for inverse parameter estimation and process simulation.

4.3.3 Analytical Methods

HPSEC measurements of the collected fractions allowed the quantification of HMW and monomer species. Protein samples were injected on the 30 cm TSKgel® G3000SWXL column (Tosoh Bioscience, Griesheim, Germany). The mobile phase of the isocratic separation was a 200 mM L-Arginin,100 mM sodium phosphate buffer at pH 6.8. Charge variants were quantified via HPSCX measurements. Fractions were analyzed on a 25 cm ProPacTM SCX-10 column (Waters, Eschborn, Germany). The charge isoforms were separated in a linear salt gradient from 0 mM to 100 mM KCl. The heterogenic charge pattern was divided in one acidic peak group (APG), one basic peak group (BPG) and the main peak (Main). In order to stay consistent with the mass balance for all simulated protein species, the relative percentage of the charge isoforms was calculated based on the monomer concentration obtained from HPSEC analysis.

4.4 Results

4.4.1 System and Column Characterization

Results of the system and column characterization experiments are listed in Table 4.2. System dead volumes were determined by tracer injections with an empty column with 0 mm bed height. Consequently, the dead volume of the column-inlet and –outlet tubing was considered in the calculation of the column porosities and the ionic capacity. The NaCl tracer experiments allowed the calculation of the axial dispersion coefficient using the equation shown in Table 4.2 [77, 115, 175], where s denotes for the variance of the conductivity peak. This correlation is based on the assumption that pore diffusion is not limiting and the peak broadening is primarily caused by axial dispersion [115, 163].

Table 4.2: System and column specific parameters.

Parameter	Symbol	Value	Unit	Equation	Reference
Length	L	157	mm		
Diameter	d	10	mm		
Column Volume	V	12.33	mL		
Bead radius	r_p	25	μm		
Interstitial porosity	ϵ_{col}	0.41		$\varepsilon_{col} = \frac{V_{int}}{V}$	[114]
Total porosity	ϵ_t	0.73		$\varepsilon_t = \frac{V_t}{V}$	[114]
Particle porosity	ϵ_p	0.53		$\varepsilon_p = \frac{V_t - V_{int}}{V - V_{int}}$	[114]
Axial dispersion	D_{ax}	0.17	mm^2/s	$D_{ax} = \frac{uLs^2_{NaCl}}{(2V_t)^2}$	[115]
Ionic capacity	Λ	0.292	M	$\Lambda = \frac{c_{NaOH}V_{NaOH}}{V(1-\varepsilon_t)}$	[114, 161]

4.4.2 Model Calibration at Set Point pH

A LGE with a process relevant load density (C1), one step elution experiment (C2), and LGEs under load conditions in the linear range of the adsorption isotherm (C4, C7) were selected for the model calibration of the transport-dispersive SMA model at pH 5.8. All isotherm parameters at pH 5.8 are listed in Table 4.3. The effective mass transfer parameters $k_{eff,i}$ were calculated based on the penetration correlation, under consideration of the molecular weight [77, 216]. Identical hydrodynamic radii of the charge isoforms lead to identical mass transfer rates of $k_{eff,APG,Main,BPG} = 0.00147\ mm/s$. A molecular weight of 300 kDa for a dimeric HMW species was assumed resulting in $k_{eff,HMW} = 0.00117\ mm/s$. The resulting simulations allowed the accurate prediction of peak widths and heights for the LGEs C4 and C7 in the linear range of the adsorption isotherm (Chromatograms not shown). Mass transfer for the small salt ion was assumed to be not limiting. Therefore, $k_{eff,salt}$ was approximated with $r_p/3$ resulting in $0.00833\ mm/s$.

The estimation of ν_i and $k_{eq,i}$ started with the lumped monomer and the HMW species using the LGE experiments at pH 5.8 (C4, C7) shown in Figure 4.4. The load material applied in C4 was representative for the process feed material, containing 0.4% HMW. Thus, C4 allowed the estimation of ν_{mono} and $k_{eq,mono}$ while the HMW species remained undetectable in the UV signal of the chromatograms. The parameters ν_i and $k_{eq,i}$ of the charge isoforms were estimated using the inverse method. ν_{mono} and $k_{eq,mono}$ allowed the definition of physical plausible boundary conditions for the inverse estimation problem, $\pm 50\%$ of $\nu mono$ and $log_{10}(k_{eq,mono})$. Parameters defining the nonlinear region of the SMA isotherm, $k_{kin,i}$ and σ_i of the three charge isoforms, were estimated without initial guess. The inverse estimation was performed using the LGE at 45 g/L (C1) and step elution at 10 g/L load density (C1) shown in in Figure 4.2. The simulated sum signal in Figure 4.2

matched with the UV measurement curve. Further, the model could describe the distinct displacement effects of the charge isoforms observed in the LGE at 45 g/L load density. The simulation of the step elution experiment described the elution order of weak to strong interacting species (APG<Main<BPG) resulting in the tailing shape of the elution peak.

Table 4.3: SMA isotherm parameters at pH 5.8 and confidence intervals in %.

Parameter	APG	Main	BPG	HMW
$\nu_i[-]$	7.38 ± 0.3%	7.50 ± 0.1%	7.70 ± 4.9%	11.08 ± 0.7%
$k_{eq,i}[-]$	1.45 ± 0.7%	1.41 ± 0.3%	1.69 ± 10.7%	1.86 ± 0.3%
$k_{kin,i}[sM^{\nu}]$	8.08E-06 ± 17.7%	1.00E-04 ± 4.3%	5.00E-04 ± 74.1%	3.4E-05 ± 3.5%
$\sigma_i[-]$	128.6 ± 1.8%	56.3 ± 2.6%	107.1 ± 6.0%	0

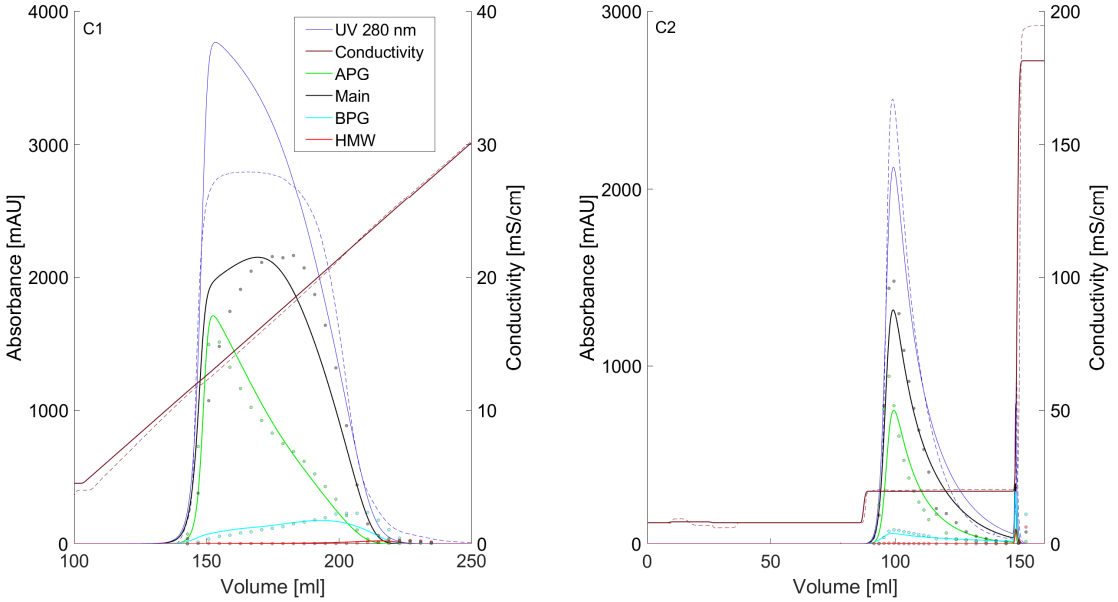

Figure 4.2: Model calibration of the SMA model at pH 5.8. Dashed lines show measurement data. C1: LGE at 45 g/L load density. The simulation exceeds the saturated UV measurement signal. C2: step elution at 10 g/L load density. The absolute concentration of the charge variants (scatter plot) is shown in mAU. Both experiments were used for the inverse estimation of model parameters. Process conditions are listed in Table 4.1.

Additional LGE experiments were performed for the model calibration for the low concentrated HMW impurities. Initial HMW parameters were estimated by fitting the model output to the fraction data of experiments C1 and C2. The initial parameters were able to describe the calibration experiments. However, large confidence intervals of the HMW estimates resulted in unreliable predictions of the product purity for validation experiments. Consequently, additional calibration experiments with HMW enriched load material were performed (C7). In order to obtain representative HMW load material, containing aggregates with physicochemical properties comparable to the normal process feed, a HMW enrichment experiment was performed on the CEX column also used for model calibra-

tion. The final experiment consisted of two consecutive salt steps for the elution of a first monomer fraction and a HMW enriched fraction. The obtained load material was applied for the LGEs C6-C8. Due to the elevated HMW content of 39.2%, a separate HMW peak could be detected in the UV trace of the LGE chromatograms, allowing the calculation of ν_{HMW} and $k_{eq,HMW}$ using the Yamamoto correlation. $k_{kin,HMW}$ was estimated inversely. Shielding effects were neglected for the HMW species ($\sigma_{HMW} = 0$), due to the comparably low load density.

Figure 4.3 shows the confidence ellipsoids of the HMW estimates before and after performing LGE experiments with HMW enriched material (C7). All dimensions of the confidence space could be reduced significantly by conducting the dedicated experiments for HMW parameter estimation. Figure 4.3 also shows a reduction of the negative correlation between ν_{HMW} and $k_{eq,HMW}$ after determining these parameters using the Yamamoto method.

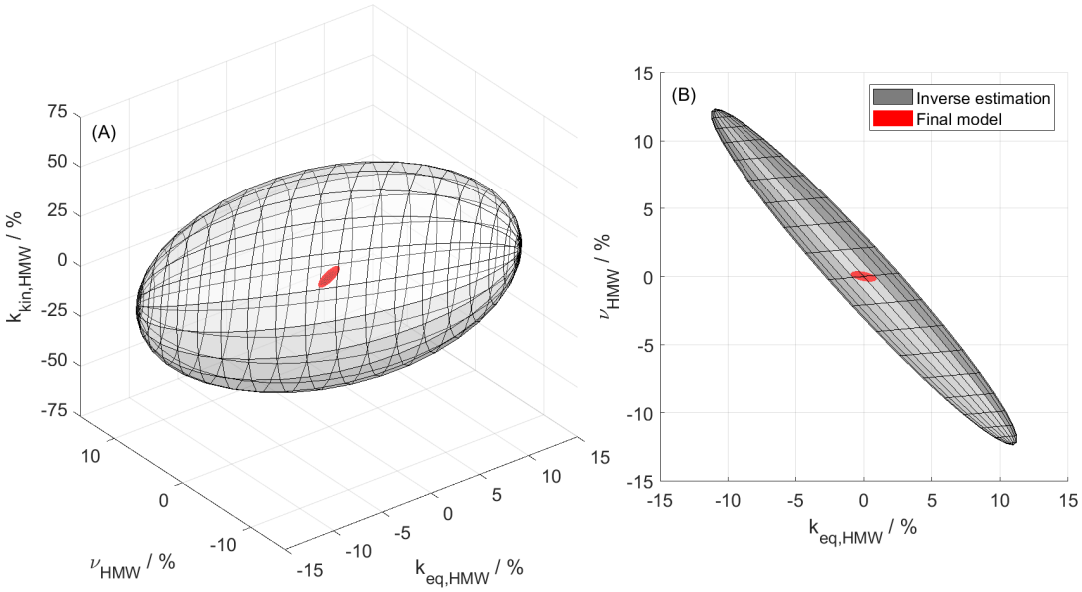

Figure 4.3: Ellipsoids visualizing the 95% confidence space for HMW isotherm parameters. Confidence space before (grey) and after (red) conducting calibration experiments with HMW enriched load material. (A) Aerial view, (B) x-y view.

4.4.3 Estimation of pH-dependent Model Parameters

For the introduction of the pH-dependency to the SMA isotherm, three additional protein specific parameters had to be estimated for each protein species. The Yamamoto method allowed the estimation of the pH-dependent model parameters for the monomer and HMW species (Table 4.4). The pH range of pH 5.8 ±0.3 covers the parameter space investigated in future robustness analysis. LGEs with altering slopes and load material were conducted

at pH 5.5, 5.8 and 6.1 (C3-C8). Figure 4.4 (A) shows the resulting Yamamoto correlations for monomer and HMW with R^2 values > 0.99. The characteristic charge ν_i and the equilibrium constant $k_{eq,i}$ of the lumped monomer and HMW species were derived from the slope and intercept of the Yamamoto correlation using the equations given in chapter 2.2. ν_i and $k_{eq,i}$ are plotted in Figure 4.4 (B) and (C). Within the investigated pH range, the assumed linear pH-dependency for the characteristic charge could be verified. The slopes of the linear approximations differed between the monomer and HMW species. For the equilibrium constant, the empirical pH model was able to describe the $k_{eq,i}$ values obtained from the Yamamoto method. Due to the high structural similarity between charge variants, the slope parameters $\nu_{1,monomer}$, $k_{eq,1,monomer}$ and $k_{eq,2,monomer}$ were assumed to be equal for APG, Main and BPG species. Intercepts $\nu_{1,i}$ and $k_{eq,0,i}$ of the charge variants were taken from the model at set point pH 5.8. This method was developed by Rischawy *et al.* [163] and ensures the model to be unbiased at set point pH by forcing the pH-dependent equation to cross the initial parameter values for charge and equilibrium.

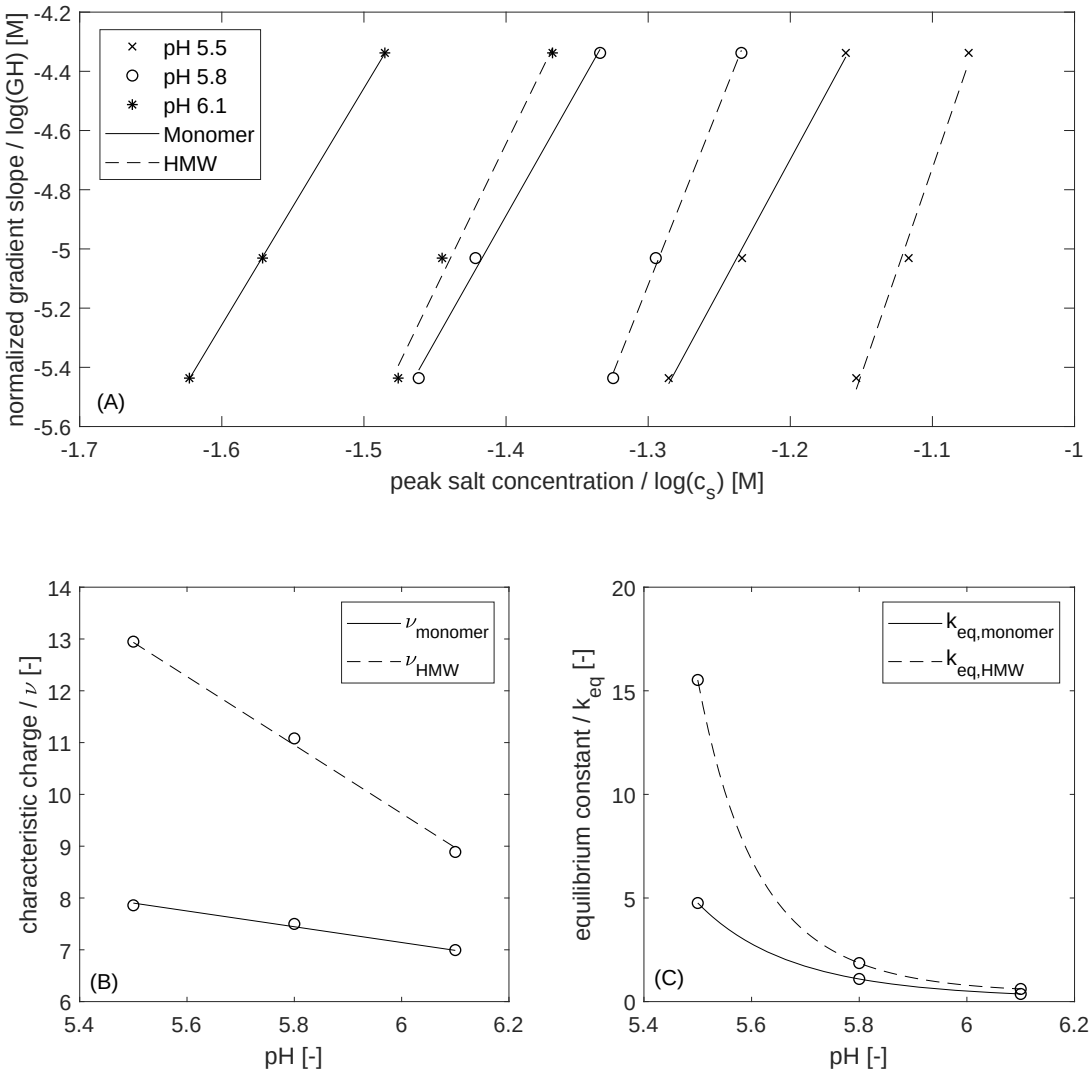

Figure 4.4: Calibration of the pH-dependent CEX model. Yamamoto correlations (A) for LGE experiments with 10 CV, 20 CV and 30 CV gradient length at pH 5.5, 5.8 and 6.1, for monomer (C3-C5) and HMW (C6-C8). Process conditions are listed in Table 4.1. pH-dependencies of the characteristic charge (B) and equilibrium constant (C) for the monomer and HMW species were obtained from the Yamamoto correlation

Table 4.4: Protein specific model parameters for the pH-dependent SMA model. pH-dependent parameters were determined from Figure 4.4 and Eqs. 4.7, 4.8, and 4.9. For a clear representation of model parameters at pH 5.8, the pH was normalized to zero. pH 5.5 = -0.3, pH 5.8 = 0, pH 6.1 = 0.3.

Parameter	APG	Main	BPG	HMW
$\nu_{pH5.8,i}[-]$	7.38	7.50	7.70	10.97
$\nu_{1,i}[-]$	-1.44	-1.44	-1.44	-6.77
$k_{eq,pH5.8,i}[-]$	1.45	1.41	1.69	1.86
$k_{eq,1,i}[-]$	-4.26	-4.26	-4.26	-5.39
$k_{eq,2,i}[-]$	2.19	2.19	2.19	5.59

4.4.4 Model Validation

In the light of the GMoP concept introduced by Rischawy *et al.* [163], a thorough model validation was performed using twelve additional experiments (V1-V12). Process conditions for model validation were selected beyond the range of future process characterization studies. Model validation results are shown in Figure 4.5. The validation set includes LGEs and step elution experiments at pH 5.5, 5.8, and 6.1 under low (1 g/L), moderate (10 g/L) and high load (45 g/L) conditions. Simulated elution and regeneration peaks during the step elution and high load LGE experiments exceed the measurement signal, due to the saturated UV sensor at 3000 mAU. The measurement signal below 3000 mAU revealed that all experiments at pH 5.8 (V1-V6) were predicted with a satisfactory accuracy. Further, step elution experiments at moderate load conditions and LGEs at high load conditions at pH 5.5 (V7-V9) and pH 6.1 (V10-V12) were performed for the validation of the pH-dependent model. Visual comparison of simulation and measurement data shows that validation experiments were well reproduced by the model. Peak shapes in V4 and V12 indicate an elution behavior, which cannot be described with the SMA isotherm. Further, the delayed desorption of protein during the isocratic elution in V9 was overestimated by the simulation. Nonetheless, simulations of validation experiments met the demanded accuracy for future application of the model during *in silico* process characterization studies.

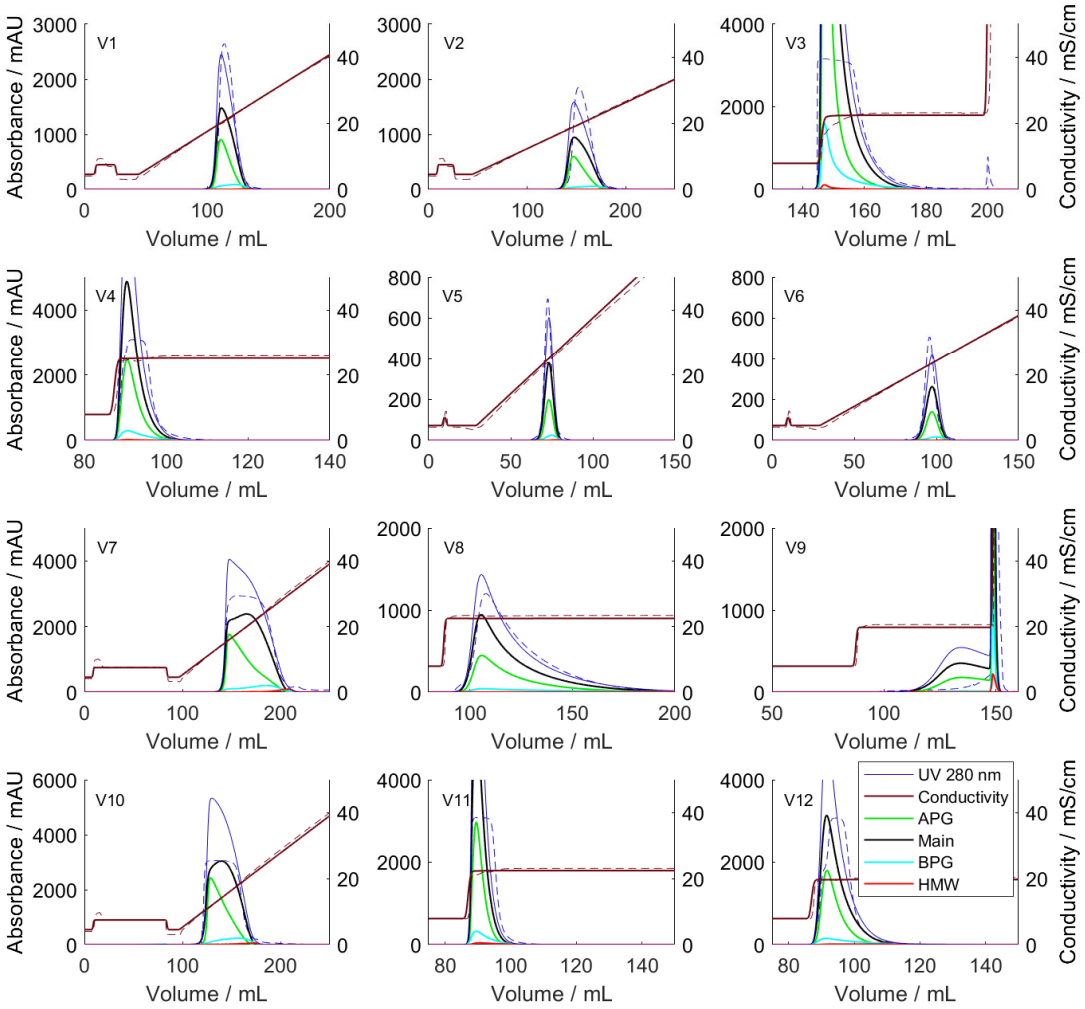

Figure 4.5: Model validation. Predictions of V1-V12 using the pH-dependent mechanistic model. Dashed lines show measurement data. Process conditions are listed in Table 4.1. Validation experiments V1-V6 were conducted at set point pH 5.8, V7-V9 at pH 5.5, and V10-V12 at pH 6.1.

4.5 Discussion

4.5.1 Model Calibration

The presented model calibration workflow includes fundamental parameter estimation techniques for the determination of the effective mass transfer coefficient and the linear SMA isotherm parameters ν_i and $k_{eq,i}$. Nonlinear isotherm parameters were estimated by inverse parameter estimation. The Yamamoto method enabled the estimation of all pH-dependent isotherm parameters using solely LGEs under linear load conditions. The effective mass transport parameters of the protein species could be validated by visual comparison of simulation and measurement data of the 1 g/L LGEs (V5 and V6). Under linear load conditions, the width and height of the elution peak is mainly depending on the mass transfer effects, while the retention time depends on characteristic charge ν_i and the equilibrium constant $k_{eq,i}$. The similar effect of ν_i and $k_{eq,i}$ on the retention time complicates the estimation problem. Similar correlations can be found in the data of Osberghaus *et al.* [171], Briskot *et al.* [165], and Rischawy *et al.* [163]. In the present study, multiple combinations of ν_i and $k_{eq,i}$ were able to describe the Chromatograms of the LGEs C3-C8 (data not shown). Multiple solutions indicate an ill-posed estimation problem [213]. Possible reasons for ill-posed estimation problems are over-parameterization, parameter correlation, inappropriate model selection, limitations in experimental information, parameters with little or no influence on observable parameters, and inappropriate initial parameter guesses [214]. The initial and direct calculation of ν_i and $k_{eq,i}$ using the Yamamoto method addressed these problems by ensuring reasonable estimation boundaries and initial guesses located close to the global optimum of the estimation problem. This procedure did also reduce the computational effort for the inverse estimation of the remaining isotherm parameters. Experiments C1 and C2 were designed from a frequentists point of view, increasing the probability of precise estimates for the kinetic parameter $k_{kin,i}$ and the shielding σ_i. $k_{kin,i}$ and σ_i define the non-linear region of the SMA isotherm. Visual inspection of the fraction data in C1 showed that the comparably high load density of 45 g/L provoked distinct displacement or competitive binding effects of the charge isoforms. These effects are mainly caused by differing adsorption ($k_{eq,i} = k_{ads,i}/k_{des,i}$) and desorption rates ($k_{kin,i} = 1/k_{des,i}$) of the charge isoforms and the amount of ligands occupied by the respective charge isoform. Further, the retention time of the UV elution peak in C1 depends on the occupied ligands on the resin surface, causing a high sensitivity for the estimation of the shielding parameter σ_i. The step elution experiment C2 at moderate load density (10 g/L) was conducted at a salt elution concentration of 30 mM less compared to the set-point condition. As a result, the elution peak showed an increased tailing behavior compared to runs at higher salt concentration. Thus, C2 contains additional information on the desorption behavior of the protein facilitating the estimation of $k_{kin,i}$. All obtained parameters were located in physically reasonable ranges and were consistent with values found in literature [114, 115]. In contrast to Hahn *et al.* [114], the present calibration technique resulted in differing shielding values σ_i for the monomeric charge isoforms. A possible mechanistic explanation for this phenomenon are altering binding orientations of the APG, Main, and BPG species caused by their changing charge

distributions on the protein surface. Additionally, it has been shown that the shielding parameter also describes repulsive effects [31]. Therefore, shielding parameters could differ between charge isoforms of a mAb.

The subsequent introduction of the pH-dependency was based on the LGEs C3, C5, C6, and C8. The estimates were divided in two subsets by splitting the calibration workflow in: 1. SMA model parameters, 2. pH-dependent model parameters. The Yamamoto based estimation of pH-dependent parameters was performed without inverse curve fitting and therefore reduced the computational effort. This approach mitigated the risk of over-fitting and ensured physical plausible estimates. Another benefit of this technique is the early identification of effects not covered by the selected model. In the present study, the selected pH-model [115] could describe the pH-dependency of the equilibrium constant and the characteristic charge within the investigated pH range. However, the applied model does not allow extrapolation beyond this pH range due to its empirical nature. Mechanistic pH-dependencies as proposed by Schmidt *et al.* [127] are designed to cover a broader pH range. The increased number of model parameters in this type of model also requires an increased number of experiments for model calibration.

The pH-dependency of the HMW's characteristic charge showed a differing slope compared to the monomeric species (Figure 4.4). A possible reason could be the complex structure and differing binding orientation of the antibody aggregates. The unique pH-dependency of the HMW species was identified using the Yamamoto method with enriched HMW load material. HMW enrichment *via* preparative CEX under mild salt elution conditions at pH 5.8 resulted in representative mAb aggregates. Stress induced antibody aggregation can lead to HMW species with differing physicochemical properties depending on the aggregation inducing treatment [217, 218] and was therefore not performed. Compared to the inverse calibration of the HMW species, this procedure significantly reduced the model uncertainty for the prediction of the product purity.

The presented model calibration procedure itself can be completed in a relatively short time frame, typically within one day, and is consistent with the fundamental principles of preparative chromatography. Two experiments with fraction analysis and LGEs under linear load conditions were needed for achieving a high model certainty. This model lends itself to advanced applications in DSP development, including model-based process characterization and scale-up. Due to the straightforward protocol, this model calibration workflow has the potential to debottleneck the daily work of chromatography modelers and making chromatography modeling accessible for applications in the biopharmaceutical industry.

4.5.2 Model Validation

The model was validated by visual comparison of simulation and measurement data of twelve additional experiments under varying process conditions. The representative calibration experiments C1 and C2 at load conditions in the non-linear range of the adsorption

isotherm enabled the prediction of validation experiments at pH 5.8. The LGEs at 45 g/L load density and differing pH values (C1, V7 ,V10) revealed the strong influence of the pH on retention volume, elution volume, peak shape and the displacement effects of the charge variants. The consideration of charge variants was found to be essential for accurate modeling of the CEX elution profile, which is in good agreement with recent publications [163, 165]. The pH-dependency used in the presented model was found to be suitable to describe the elution, competitive binding and displacement effects within the investigated pH range. Although the estimation of the pH-dependent parameters was solely based on LGEs at 1 g/L load density, the model is able to predict all LGEs and step elution experiments at pH 5.5 and pH 6.1 with high load density conditions. Assuming identical pH-dependencies for different charge isoforms was valid within the investigated pH range. The overestimated desorption of protein during the elution step of V9 is most likely related to the constant kinetic $k_{kin,i}$ and shielding parameters σ_i of the subspecies. In contrast to ν_i and $k_{eq,i} = k_{ads,i}/k_{des,i}$ the non-linear isotherm parameters σ_i and $k_{kin,i} = 1/k_{des,i}$ are assumed to be pH-independent. Therefore, the selected model considers pH-dependent changes in the adsorption rate while the protein desorption rates are constant for all pH values. Further, possible changes in the binding orientation are not described due to the constant σ_i parameter. An introduction of pH- dependencies for $k_{kin,i}$ and σ_i could further improve the models ability to describe step elution experiments at different pH values. Additionally, the selected SMA isotherm was not able to describe parts of the peak shapes in V4 and V12 (Figure 4.5). Using Mollerup's generalized ion-exchange isotherm [215], it might be possible to describe the complex elution behavior in V4 and V12. However, all possible extensions to the selected model would increase the number of parameters to estimate and the complexity of the model calibration process, without gaining additional benefit for future applications of this model.

4.6 Conclusion

The present study introduced a straightforward and rapid workflow for the calibration of pH-dependent and multicomponent SMA chromatography models. The selection of representative process conditions allowed model calibration using only two experiments with offline fraction analysis. Fundamental knowledge on preparative chromatography enabled the systematic reduction of unknown model parameters and avoided pitfalls, such as ill-posed estimation problems, which can be related to a model calibration completely based on curve fitting. Multiple LGEs and the Yamamoto correlation led to the early elimination of the estimation problem caused by the parameter correlation of the characteristic charge and equilibrium constant. Thus, the estimation of the remaining isotherm parameter using the inverse method could be simplified. Comparably small confidence intervals, resulting in a small confidence region of the model prediction, indicate that the chosen experiments were representative and well suited for solving the estimation problem. The model uncertainty of the HMW species was reduced by performing dedicated calibration experiments with HMW enriched load material. This procedure may be essential for model-based process characterization when considering the importance of modeling

low-concentrated product related impurities for the prediction of the product purity. The subsequent introduction of the pH-dependency was entirely based on LGEs under load conditions in the linear range of the adsorption isotherm. Nonetheless, the final model enabled predictions of step elution and gradient elution experiments at process relevant, increased load densities, and different mobile phase conditions. The obtained accuracy of the model prediction fulfills the requirements of advanced applications in DSP development, such as model-guided scale-up and *in silico* process characterization studies. Having the ability to calibrate complex chromatography models within hours not only debottlenecks *in silico* process development but has also the potential to increase the acceptance of mechanistic modeling in the biopharmaceutical industry.

5 Cross-scale Quality Assessment of a Mechanistic Cation Exchange Chromatography Model

David Saleh[1, 2], Gang Wang[2], Benedict Müller[2], Federico Rischawy[2], Simon Kluters[2], Joey Studts[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] Boehringer Ingelheim Pharma GmbH & Co. KG, Biberach, Germany
* Corresponding author: Jürgen Hubbuch, +49 721 608 47526, juergen.hubbuch@kit.edu

Abstract

Cation exchange chromatography (CEX) is an essential part of most monoclonal antibody (mAb) purification platforms. Process characterization and root cause investigation of chromatographic unit operations are performed using scale down models (ScDM). ScDM chromatography columns typically have the identical bed height as the respective manufacturing-scale, but a significantly reduced inner diameter. While ScDMs enable process development demanding less material and time, their comparability to manufacturing-scale can be affected by variability in feed composition, mobile phase and resin properties, or dispersion effects depending on the chromatography system at hand. Mechanistic models can help to close gaps between scales and reduce experimental efforts compared to experimental ScDM applications.

In this study, a multicomponent steric mass-action (SMA) adsorption model was applied to the scale-up of a CEX polishing step. Based on chromatograms and elution pool data ranging from laboratory- to manufacturing-scale, the proposed modeling workflow enabled early identification of differences between scales, e.g. system dispersion effects or ionic capacity variability. A multi-stage model qualification approach was introduced to measure the model quality and to understand the model's limitations across scales. The experimental ScDM and the *in silico* model were qualified against large-scale data using the

identical state of the art equivalence testing procedure. The mechanistic chromatography model avoided limitations of the ScDM by capturing effects of bed height, loading density, feed composition, and mobile phase properties. The results demonstrate the applicability of mechanistic chromatography models as a possible alternative to conventional ScDM approaches.

5.1 Introduction

Over the last decade, ever-rising numbers of monoclonal antibodies (mAbs) in development pipelines increased the demand for novel technologies accelerating mAb process development [3, 4, 184]. Multiple publications highlighted the potentialities of *in silico* process models for rapid and rationalized bioprocess development [16, 188]. However, there is so far no consensus within the biopharmaceutical industry on how to apply digital process models to real-world tasks.

Cation exchange chromatography (CEX) is a frequently employed polishing step for the downstream processing (DSP) of mAbs. Its selectivity towards protein charge allows the depletion of high molecular weight species (HMW) and other product and process related impurities [6, 95]. HMW removal is of high importance because antibody aggregates may cause an immune response towards the monomeric drug [75]. An in-depth process understanding is crucial for robust process performance and consistent product quality. Development of CEX processes is typically based on a high number of small-scale experiments. Resin selection and process optimization can be performed using automated batch-binding screenings [11, 219] and miniaturized column processes [12, 208, 220, 221]. Automation, parallelization, and miniaturization are of high value for early stage process development, where a large number of process conditions are screened for numerous mAb candidates. Late-stage work packages, such as process characterization, rely on bench-scale experiments using a scale-down model (ScDM) column representative for the respective manufacturing-scale unit operation. The ScDM, as a physical representation of the manufacturing process, enables effect analysis of process parameters on critical quality attributes (CQA) and key performance indicators (KPI). Hakemeyer *et al.* [13] described key elements to be considered in ScDM design, ranging from impurity levels in load material to the use of sound engineering principles for scaling. In order to keep key process parameters such as residence time and separation distance constant, ScDM columns typically have identical bed heights as the respective manufacturing-scale unit operation, but a reduced inner diameter. Furthermore, miniaturized columns for robotic liquid handling stations are currently explored as ScDM for chromatographic unit operations [97]. Reduction of column diameter enables a fast and resource-saving development. However, the effectiveness of the purification is not exclusively affected by column dimensions. Variability of input material composition and impurity levels, may impact the comparability between a ScDM and its respective large-scale unit operation. Consequently, statistical ScDM qualification demands numerous bench-scale experiments with varying input material. Additional effects caused by dispersion in different chromatography systems, resin

lot-to-lot variability [222, 223] and manual column packing procedures [224, 225] may further lead to systematic differences between scales.

In order to follow the Quality by Design (QbD) concept [24], biopharmaceutical companies are working on process models to generate an in-depth process understanding [97, 164, 226–228]. Mechanistic chromatography models are mathematical representations of the physical effects occurring in the chromatographic system [118, 229]. They consist of partial differential equations, describing macroscopic transport through the column, mass transport within the stationary phase, and adsorption of protein to the resin. For mechanistic modeling of mAbs and other proteins in ion exchange chromatography, the SMA adsorption isotherm is frequently used in academic and industrial case studies [97, 117, 128, 163, 165, 171, 230]. The SMA isotherm describes the multipoint binding of proteins to the resin under consideration of a protein's characteristic charge, the thermodynamic equilibrium of the adsorption process, and steric shielding effects. Multiple studies have demonstrated successful application of mechanistic models for the scale-up of chromatography processes [27, 221, 231]. Benner *et al.* [97] used mechanistic modeling to explain systematic offsets between large-scale processes and an experimental ScDM based on miniaturized columns for robotic liquid handling stations. The scientific explanation for scaling effects enabled them to utilize the miniaturized system for a parallelized and material saving process characterization study [97]. Ladwig *et al.* [226] published a mechanistic model describing pH and excipient concentrations for an ultrafiltration and diafiltration (UF/DF) unit operation of a mAb purification process. Similar to the experimental ScDM approach, the mechanistic UF/DF model was qualified against large-scale data validating the model's capability to reduce experimental efforts during process development [226]. Beside the capability of explaining scaling effects, the physical principles of mechanistic chromatography models allow the reduction of experimental effort by *in silico* experimentation at manufacturing-scale.

Successful application of mechanistic models in bioprocess development and manufacturing requires clear guidelines for model development, qualification, and application. As a first building block, we recently introduced a standardized workflow for model calibration to build the quality into the model by applying both, engineering and statistic principles [181]. The subsequent model validation at calibration-scale included twelve experiments with operating conditions beyond the calibration space of the final unit operation. The present study aims to propose the other part of the model quality system with respect to qualifying the mechanistic chromatography model from laboratory- to manufacturing-scale. A multi-stage evaluation using statistical criteria and engineering knowledge was introduced and applied to measure the quality of model prediction and to understand the model's limitations. Six CQAs and KPIs were derived from the predicted chromatograms and corresponding cutting criteria and compared to the wet-lab purification outcomes. Comparable to the experimental ScDM, qualification of the mechanistic model against manufacturing-scale enabled rational evaluation of model predictions for CQAs and KPIs. Direct benchmarking of the mechanistic model against the experimental ScDM showed the benefits and perils of both techniques.

5.2 Modeling

This section gives an overview on the mechanistic model and complementation necessary for model-guided scale-up. Additional details about model discrimination, model parameters and the model calibration strategy can be found in our previous publication [181]. Protein-specific model parameters were kept constant for all simulations and are listed in Table 5.1. Simulation and inverse parameter estimation was performed using the ChromX software (GoSilico, Karlsruhe, Germany). Depending on the large-scale chromatography system, pre-column dispersion was approximated by simulating a continuous stirred-tank reactor (CSTR) at the column inlet. Assuming ideal mixing within the CSTR, the change of concentration $c_i(t)$ of component i in Eq. 5.1 is a function of residence time defined by the reactor length L_{CSTR} and superficial velocity u.

$$\frac{\partial c_i(x,t)}{\partial t} = -\frac{u}{L_{CSTR}} \left(c_{in,i}(t) - c_i(t)\right) \tag{5.1}$$

The transport dispersive model was selected as column model, due to multiple successful case studies for the simulation of ion exchange chromatography systems [99, 114, 116, 167, 190]. Eq. 5.2 describes the macroscopic transport of component i through the chromatography column. The change of the concentration c_i at position x in time t is a function of convective mass transport in the interstitial volume, peak broadening caused by axial dispersion D_{ax}, and mass transfer from the interstitial volume into the pore phase of the particle with the radius r_p. Further, mass transfer between the interstitial volume and the particle pores is affected by the interstitial porosity ε_{col} and the effective mass transfer coefficient $k_{eff,i}$. The accumulation of mass in the pore phase c_i and the stationary phase q_i is described in Eq. 5.3. The Danckwerts' boundary conditions are given in Eqs. 5.4 and 5.5.

$$\begin{aligned}\frac{\partial c_i(x,t)}{\partial t} = &-\frac{u}{\varepsilon_{col}}\frac{\partial c_i(x,t)}{\partial x} + D_{ax}\frac{\partial^2 c_i(x,t)}{\partial x^2} \\ &- \frac{1-\varepsilon_{col}}{\varepsilon_{col}}\left(\frac{3}{r_p}k_{eff,i}\left(c_i(x,t) - c_{p,i}(x,t)\right)\right)\end{aligned} \tag{5.2}$$

$$\frac{\partial c_{p,i}(x,t)}{\partial t} = \frac{3}{r_p}\frac{k_{eff,i}}{\varepsilon_p}(c_i(x,t) - c_{p,i}(x,t)) - \frac{1-\varepsilon_p}{\varepsilon_p}\frac{\partial q_i(x,t)}{\partial t} \tag{5.3}$$

$$\frac{\partial c_i}{\partial x}(0,t) = \frac{u(t)}{\varepsilon_{col} D_{ax}}(c_i(0,t) - c_{in,i}(t)) \tag{5.4}$$

$$\frac{\partial c_i}{\partial x}(L,t) = 0 \tag{5.5}$$

Linear flow rates ranged from 155 cm/h to 360 cm/h between investigated scales, demanding the introduction of flow dependencies for the axial dispersion coefficient D_{ax} [232] and effective mass transfer parameter $k_{eff,i}$. The penetration correlation allowed the direct calculation $k_{eff,i}$ for monomer and HMW species at relevant flow rates under consideration of their hydrodynamic radii [77, 216]. Within the investigated range, flow dependencies for D_{ax} and $k_{eff,i}$ could be approximated using linear regression, Eqs. 5.6 and 5.7. Experimental validation of D_{ax} and k_{eff} parameters was performed *via* pulse injections at low loading conditions with dextran and protein, respectively [181].

$$D_{ax}(u) = D_{ax0} + uD_{ax1} \tag{5.6}$$

$$k_{eff}(u) = k_{eff0,i} + uk_{eff1,i} \tag{5.7}$$

Protein adsorption is simulated using the semi-mechanistic SMA adsorption model [117]. The SMA model formulates the equilibrium binding behavior of the protein in consideration of the salt concentration in the pore phase c_s, the ionic capacity of the resin Λ and the proteins characteristic charge ν_i. Eq. 5.8 shows the kinetic form of the SMA isotherm modified by Hahn *et al.* [167], where $k_{eq,i} = k_{ads,i}/k_{des,i}$ and $k_{kin,i} = 1/k_{des,i}$ describe adsorption and desorption rates of component i, respectively. In addition, the steric shielding parameter σ_i denotes the number of functional groups on the resin surface blocked by the protein. ν_i and $k_{eq,i}$, the SMA parameters defining the linear region of the adsorption isotherm, were estimated using the Yamamoto method at differing pH values [166, 211]. Inverse estimation [167] of the remaining isotherm parameters $k_{kin,i}$ and σ_i was facilitated by designing two experiments representative for the final unit operation. Firstly, a linear gradient elution experiment at 45 g/L loading density contained distinct information on steric shielding and competitive binding effects of the four protein species. Secondly, a step elution experiment at 10 g/L loading density and a counter ion concentration below the set point condition increased the sensitivity for estimating the desorption rate defining $k_{kin,i}$ parameter.

$$k_{kin,i}\frac{\partial q_i}{\partial t} = k_{eq,i}(pH)\left(\Lambda - \sum_{j=1}^{k}\left(\nu_j(pH) + \sigma_j\right)q_j\right)^{\nu_i(pH)} c_{p,i} - q_i c_s^{\nu_i(pH)} \tag{5.8}$$

$$q_{salt} = \Lambda - \sum_{j=1}^{k} \nu_j(pH) q_j \tag{5.9}$$

The introduction of pH-dependent isotherm parameters is crucial for industrial applications. For the identical mAb polishing step, our previous work showed significant effects on the purification outcome when varying the mobile phase pH between pH 5.5 and pH 6.1 [181]. Eqs. 5.10 and 5.11 show the empirical pH dependencies of the characteristic charge ν_i and the equilibrium constant $k_{eq,i}$ developed by Hunt *et al.* [115]. This model was found to be sufficient for the process relevant pH range of pH 5.8 ±0.3 used in this study [181].

$$k_{eq,i}(pH) = k_{eq0,i} e^{k_{eq1,i} pH + k_{eq2,i} pH^2}, \tag{5.10}$$

$$\nu_i(pH) = \nu_{0,i} + pH\nu_{1,i}, \tag{5.11}$$

Table 5.1: Protein specific model parameters for the pH-dependent SMA model. Details regarding the model calibration procedure are described in our previous publication [181]. For a clear representation of model parameters at pH 5.8, the pH was normalized to zero. pH 5.5 = -0.3, pH 5.8 = 0, pH 6.1 =0.3.

Parameter	APG	Main	BPG	HMW
$k_{eff0,i}$ [mm/s]	1.4E-3	1.4E-3	1.4E-3	1.2E-3
$k_{eff1,i}$ [-]	4.7E-05	4.7E-05	4.7E-05	3.3E-05
$\nu_{pH5.8,i}$ [-]	7.38	7.50	7.70	10.97
$\nu_{1,i}$ [-]	-1.44	-1.44	-1.44	-6.77
$k_{eq,pH5.8,i}$ [-]	1.45	1.41	1.69	1.86
$k_{eq,1,i}$ [-]	-4.26	-4.26	-4.26	-5.39
$k_{eq,2,i}$ [-]	2.19	2.19	2.19	5.59
$k_{kin,i}[sM^{\nu}]$	8.08E-06	1.00E-04	5.00E-04	3.4E-05
σ_i [-]	128.6	56.3	107.1	0

5.3 Material and Methods

5.3.1 CEX Unit Operation

The mAb used in this study is an IgG1 monoclonal antibody expressed in stably transfected Chinese hamster ovary (CHO) cells (Boehringer Ingelheim GmbH & Co. KG, Biberach, Germany). The mAb was captured *via* Protein A affinity chromatography and

further polished using anion exchange chromatography in flow-through mode. The presented mechanistic model describes the subsequent CEX unit operation using the strong CEX resin POROS 50 HS (Thermo Fisher Scientific, Waltham, USA). The process was performed at constant pH 5.8 in bind-elute mode and at a maximal loading density of 45 g/L. The column was equilibrated at a counter-ion concentration of 87 mM Na^+, with the same buffer applied to the wash phase after column loading. Subsequently, elution was induced at a counter-ion concentration of 247 mM Na^+. For column regeneration and storage, 1 M and 0.1 M NaOH were applied.

Charge variant and HMW concentrations in the elution pool were quantified using analytical CEX chromatography and analytical size exclusion chromatography, respectively. In order to stay consistent with the mass balance for all simulated protein species, the relative percentage of the charge isoforms was calculated based on the monomer concentration obtained from HPSEC analysis. Acidic (APG), neutral (Main) and basic charge variants (BPG), as well as HMW species were considered as CQAs. Process step yield and elution volume were defined as KPIs and quantified using protein concentration determined via absorbance at 280 nm and gravimetric volume measurement. Details about the model calibration strategy and model validation, as well as analytical chromatography methods, are presented in the previous publication of our group [181].

5.3.2 *In silico* Scale-up and Model Qualification

Figure 5.1 summarizes scales investigated in this study, ranging from the 12.3 mL column used for model calibration to the 441 L manufacturing-scale column. Additional information about system- and column-specific properties is listed in Table 5.2. Model-guided scale-up started with the technical investigation and model development for large-scale chromatography systems. If necessary, system dispersion was simulated by adding a CSTR at column-inlet. Details about system-specific effects considered for simulations of each scale are given in Section 5.4.1. For model qualification, the predictive power of the mechanistic model was evaluated across scales. The model was applied to multiple chromatography runs at 200 L 2000 L and 12000 L scale. Simulations considered relevant input parameters, such as bed height, flow rate, load material composition, loading density, buffer and resin variability. In the present work, simulations captured the effects of real variances during large-scale experimentation. In contrast, model validation in our previous publication included experiments at challenging operating conditions far beyond the intended set-point condition [181].

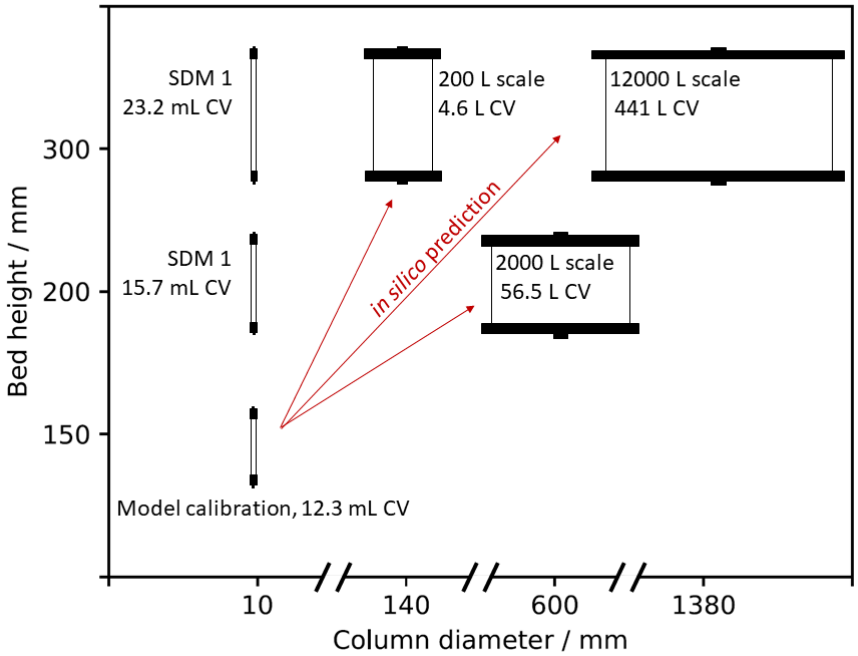

Figure 5.1: Summary of column scales for the CEX unit operation. The mechanistic model was calibrated at bench-scale and applied to 200 L, 2000 L and 12000 L scales. For a clear representation of results, the shown purification scales are named according to the bioreactor volume of preceding cell culture processes. SDM = Scale down model, CV = Column volume

Model qualification consisted of three consecutive stages:

- Investigation of the correlation between predicted and measured CQAs and KPIs across scales: Calculate linear correlation coefficients (R^2) and normalized root-mean-square error of predictions (NRMSEP)
- Testing the statistical significance of the linear correlation: Perform t-test on the slope and intercept of linear regression
- Qualification against manufacturing-scale data using a two one-sided t-tests (TOST)

For scale-independent qualification of the mechanistic model, correlations between predicted and measured values for CQAs and KPIs were evaluated using linear regression and statistical hypothesis testing. t-tests were performed on both, slope and intercept of the linear regression. The tested hypothesis for slope and intercept were one and null, respectively. The NRMSEP given in Eqs. 5.12 and 5.13 was calculated for quantification of the models' predictive power across scales.

$$RMSEP = \sqrt{\frac{\sum_{t=1}^{T}(\hat{y}_t - y_t)^2}{T}} \tag{5.12}$$

$$NRMSEP = \frac{RMSEP}{y_{max} - y_{min}} 100\% \tag{5.13}$$

In the final step, the predictive capability of the *in silico* model was compared to the experimental ScDM. Following state of the art practices in DSP development, the experimental ScDM column had the identical bed height as the respective manufacturing process and an inner diameter of 1 cm. To evaluate the comparability between ScDM and large-scale, ScDM experiments were performed in triplicate with multiple load materials obtained from different large-scale runs. In contrast, mechanistic model simulations considered the variability in the input material, loading density and mobile phase properties for each of the chromatographic cycles. Experimental ScDM and mechanistic model were qualified against large-scale data using the identical equivalence testing procedure [13, 226].

- Step 1- Calculation of the equivalence acceptance criteria (EAC) based on the sample mean ± 3 standard deviations (SD) of historical large-scale runs.
- Step 2- Equivalence test: A TOST was performed according to Schuirmann [233] using Python 3.8.2. For both, ScDM and *in silico* model, the 90% confidence interval (CI) for the difference in means to large-scale data was compared to the EAC limits. The model was defined as equivalent to the large-scale unit-operation if p-values were below the significance level of $\alpha = 0.05$.
- Step 3- Visualization and qualification: The 90% CIs of ScDM and *in silico* model difference in means were visualized and compared to the EAC. Model and large-scale unit operation were considered "equivalent", if the 90% CI on the difference in means fell entirely within the EAC. The model was "equivalent in sample mean only", if the 90% CI overlapped with one or both EAC. The model "failed to be equivalent", if the difference in means was located outside the EAC and the 90% CI on overlapped with EAC. When the 90% CI on the difference fell entirely outside the EAC, the model was ranked "not equivalent" [13].

5.4 Results and Discussion

In this work, a mechanistic cation exchange chromatography model calibrated at bench-scale, was applied to chromatography runs of multiple large-scales, including 2000 L pilot and 12000 L manufacturing-scale. The following chapters describe the approach for the simulation of large-scale systems. Consideration of system dispersion and variations in

ionic capacity between scales enabled the successful qualification of the mechanistic model across scales. For 12000 L manufacturing-scale, mechanistic model predictions and experimental ScDM results were compared considering the most relevant CQAs and KPIs.

5.4.1 Simulation of System-specific Effects

Table 5.2 shows system and column specific parameters and considerations for simulation of the investigated large-scale chromatography processes. The mechanistic model presented in this study was calibrated at a column bed height of 157 mm. Protein specific mass transfer parameters and SMA adsorption model parameters estimated in our previous publication were applied to simulations of large-scale experiments (Table 5.1) [181]. 200 L, 2000 L, and 12000 L scale experiments were conducted at differing linear flow rates compared to calibration-scale, demanding the introduction of flow dependent mass transfer parameters. Details about flow dependencies and parameters are given in Section 5.2, Table 5.1.

Table 5.2: System and column specific parameters applied for the simulation of 200 L, 2000 L and 12000 L scales. *Porosities were determined at calibration-scale and kept constant across scales.

Parameter	Symbol	Calibration	200 L	2000 L	12000 L	Unit
Bed height	L	157	300	200	295	mm
Diameter	d	10	140	600	1380	mm
Column Volume	V	0.0123	4.62	56.6	441	L
Bead radius	r_p	25	25	25	25	μm
Interstitial porosity*	ε_{col}	0.41	0.41	0.41	0.41	-
Total porosity*	ε_t	0.73	0.73	0.73	0.73	-
Particle porosity*	ε_p	0.53	0.53	0.53	0.53	-
Ionic capacity	Λ	0.292	0.310	0.310	0.292	M
Flow rate	u	188	360	240	206	cm/h
Flow rate during elution	$u_{elution}$	188	360	240	155	cm/h
Axial dispersion y-intercept	D_{ax0}	0.0501	0.0501	0.0501	0.0501	mm^2/s
Axial dispersion slope	D_{ax1}	0.2499	0.2499	0.2499	0.2499	mm
CSTR length	L_{CSTR}	-	79	-	-	mm

Visual inspection of the conductivity signal at column outlet was the first step of *in silico* scale-up. This procedure allowed identification and correction of dead volumes and system dispersion effects before simulating protein elution at a new scale. When the salt simulation followed the trend of the conductivity signal, the simulated elution peak was compared to UV measurement data. If further systematic offsets in peak width and retention time were observed, additional evaluation of the elution peak discrepancy was necessary. Variability of resin lot and manual column packing were identified as the most probable root causes for the observed differences between scales causing small variations that were lumped into the

ionic capacity. Alternatively, an increased mass transfer resistance resulting in a decreased value for k_{eff} could also describe the broadened peak shape. However, a scale-dependent change of k_{eff} is considered less likely compared to variations in ionic capacity, which is a well-known phenomenon in industrial protein chromatography. The direct measurement of ionic capacity using acid-base titration is frequently used for model calibration [114]. However, this potentially harmful procedure is inapplicable to large-scale chromatography columns that are used in compliance with good manufacturing practice (GMP). Further, ionic capacity is normalized to resin backbone volume. Hence, ionic capacity is correlated to column packing density, which complicates the reproduction of acid-base titration of different resin lots in small-scale. Therefore, ionic capacity had to be estimated by applying the inverse method developed by Hahn *et al.* using a single chromatogram of one scale [167]. The inverse method enabled prediction of all further chromatography runs and all other process outputs at the respective scales. The estimated increase in ionic capacity for 200 L and 2000 L scale was in the range of 6% compared to the calibration-scale. The observed variance in ionic capacity is considered plausible, as the manufacturer specifies the dynamic binding capacity of lysozyme on POROS 50 HS in a range between 57.0 g/L and 75.3 g/L [94].

Figure 5.2 depicts simulations of an exemplary 200 L scale chromatogram with and without the final corrections necessary for prediction of conductivity and UV signal. The 200 L scale conductivity signal at column outlet showed a distinctive curvature caused by an increased system dispersion. Compared to other scales, the 200 L scale system was not flushed with high salt buffer before starting step elution. Mixing of wash and elution buffer within the bubble trap led to a comparably slow increase of the conductivity signal. Back-mixing within the bubble trap of the 200 L scale system was approximated by simulating a CSTR at the column inlet. As a result, the simulation of the conductivity signal followed the trend of the measured conductivity signal at column-outlet. Further, the slowly increasing salt concentration during step elution and the estimated ligand density corrected retention time and width of the simulated protein peak. System dispersion in 2000 L and 12000 L scale chromatography systems could be neglected, because the large-scale systems were pre-flushed with elution buffer. The pre-flush led to a steep increase of the conductivity signal and the corresponding salt simulation (Figure 5.3). Consequently, simulated and measured elution volumes in 2000 L and 12000 L scale were smaller compared to 200 L scale, shown in Figure 5.4.

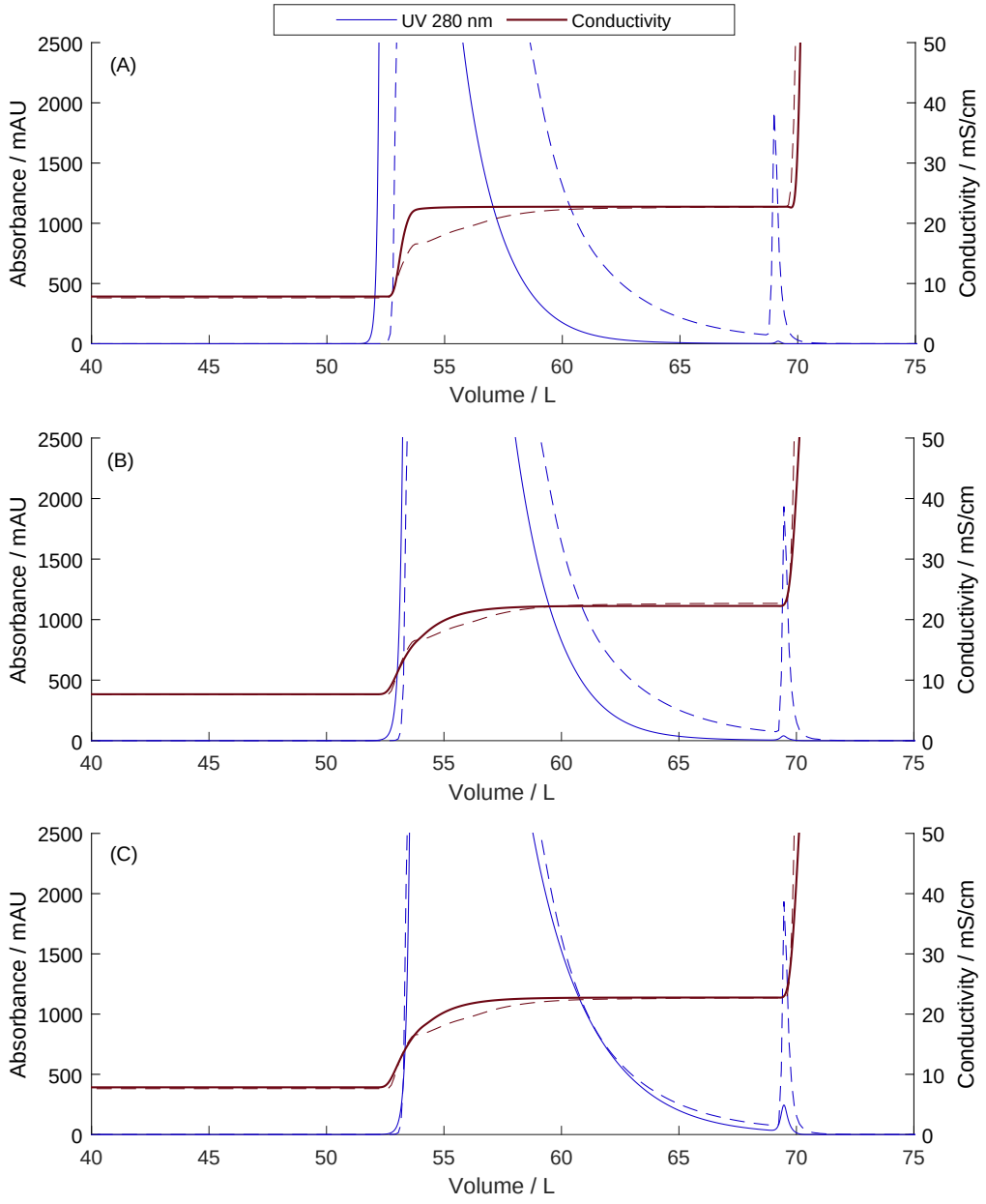

Figure 5.2: Step elution run on the 200 L scale column. Dashed lines show measurement data and solid lines are mechanistic model predictions. A: Dispersive effects outside the column were neglected during the step elution and ligand density was equal to calibration scale, $\Lambda = 0.292$ M. B: CSTR in front of the column simulates dispersion caused by the bubble trap, $\Lambda = 0.292$ M. C: Pre-column CSTR and estimated ligand density, $\Lambda = 0.310$ M.

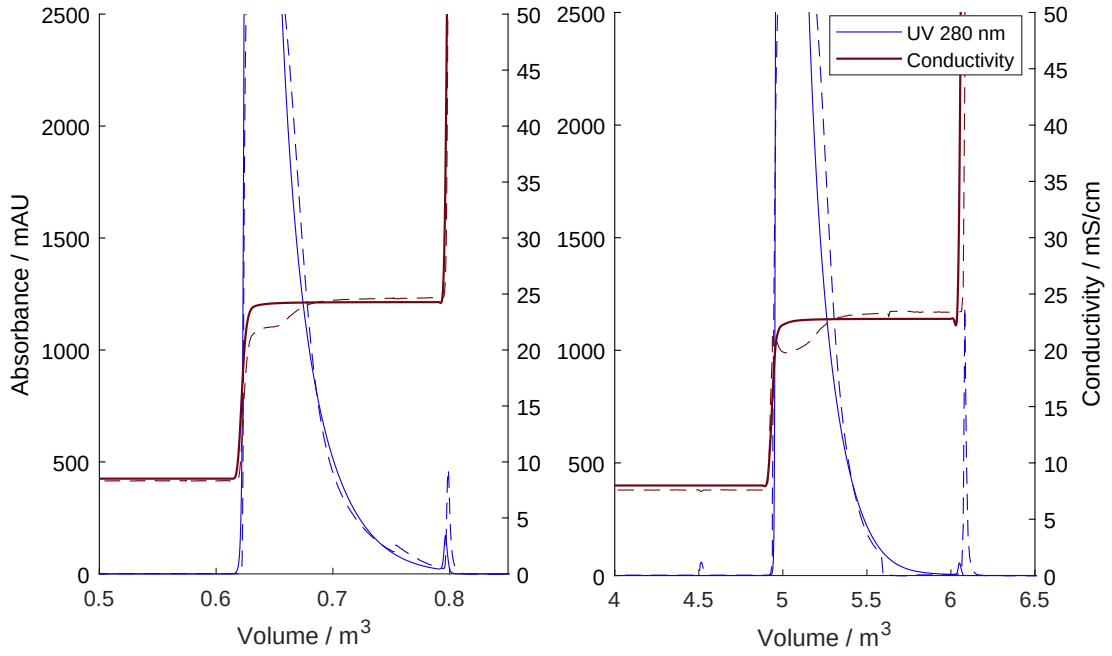

Figure 5.3: Prediction of elution profiles of 2000 L pilot and 12000 L manufacturing-scale. Dashed lines show measurement data and solid lines are mechanistic model predictions. A: 2000 L pilot-scale, B: 12000 L manufacturing-scale.

5.4.2 Cross-scale Qualification of the Mechanistic Model

This chapter evaluates the predictive power of the mechanistic model across scales. Main differences between scales were column volume, bed height, and flow rate (Table 5.2). It is important to notice, that all shown chromatography runs were conducted on the set point. Consequently, only small variation in process parameters occurred and their effect on CQAs and KPIs was small compared to the previous model validation at calibration-scale [181]. The small variation of model inputs and outputs further challenged the predictive capabilities and accuracy of the model. Multiple chromatographic cycles were simulated for each scale. Simulations accounted for all variations in load composition, loading density, and mobile phase pH and salt concentrations. In Figure 5.4, model predictions for CQAs and KPIs of the CEX purification are correlated to the respective measurement data. Investigation of correlation plots are the state of the art procedure in industrial chromatography modeling. However, a decision purely based on these correlation plots is considered too subjective. Hence, linear regression and statistical criteria were introduced for further evaluation of model quality (Table 5.3).

Table 5.3: Regression and *t*-test results for cross-scale model qualification.

	R^2	NRMSEP [%]	slope	intercept	$p_{s,1}$	$p_{i,0}$
APG	0.95	7	1.04	-1.21	0.54	0.58
Main	0.95	8	0.98	0.86	0.77	0.81
BPG	0.97	12	0.93	0.15	0.15	0.56
HMW	0.65	49	0.97	0.07	0.89	0.09
Yield	0.05	37	0.13	86.12	0.00	0.00
Elution volume	0.97	8	0.94	0.19	0.07	0.01

The very first stage of model qualification is to evaluate R^2 and NRMSEP, since a high R^2 value means the given variances are well covered by the linear regression and a low NRMSEP means the regression line is close to the optimal expectation. In this stage, with $R^2 > 0.95$ and NRSMEP $< 12\%$, the model's predictive power for the charge variants (APG, Main, and BPG) and the elution volume were confirmed and the qualification is considered as completed successfully. Qualification *via* R^2 and NRSEMP is considered the case one scenario. Elution volume measurements ranged from 1.4 CV to 2.7 CV. Compared to other CQAs and step yield, the elution volume mainly depended on scale effects, as shown in Figure 5.4. For simulations of 200 L scale experiments, a pre-column CSTR was added resulting in a larger elution volume compared to other scales. Elution volume of 200 L and 2000 L scale was also increased by a 6% higher ionic capacity compared to 12000 L manufacturing-scale. Further, the model accounted for the effect of differing bed heights on the elution volume. 2000 L scale was performed on a column with 200 mm bed height, leading to a higher elution volume compared to 12000 L scale with 295 mm bed height. It should be noted that model parameters were estimated on a small-scale column with a 157 mm bed height. Also the model validation presented in our previous publication was performed at a bed height of 157 mm [181]. Thus, the prediction of large-scale runs with increased bed heights further confirmed the plausibility of the estimated

model parameters.

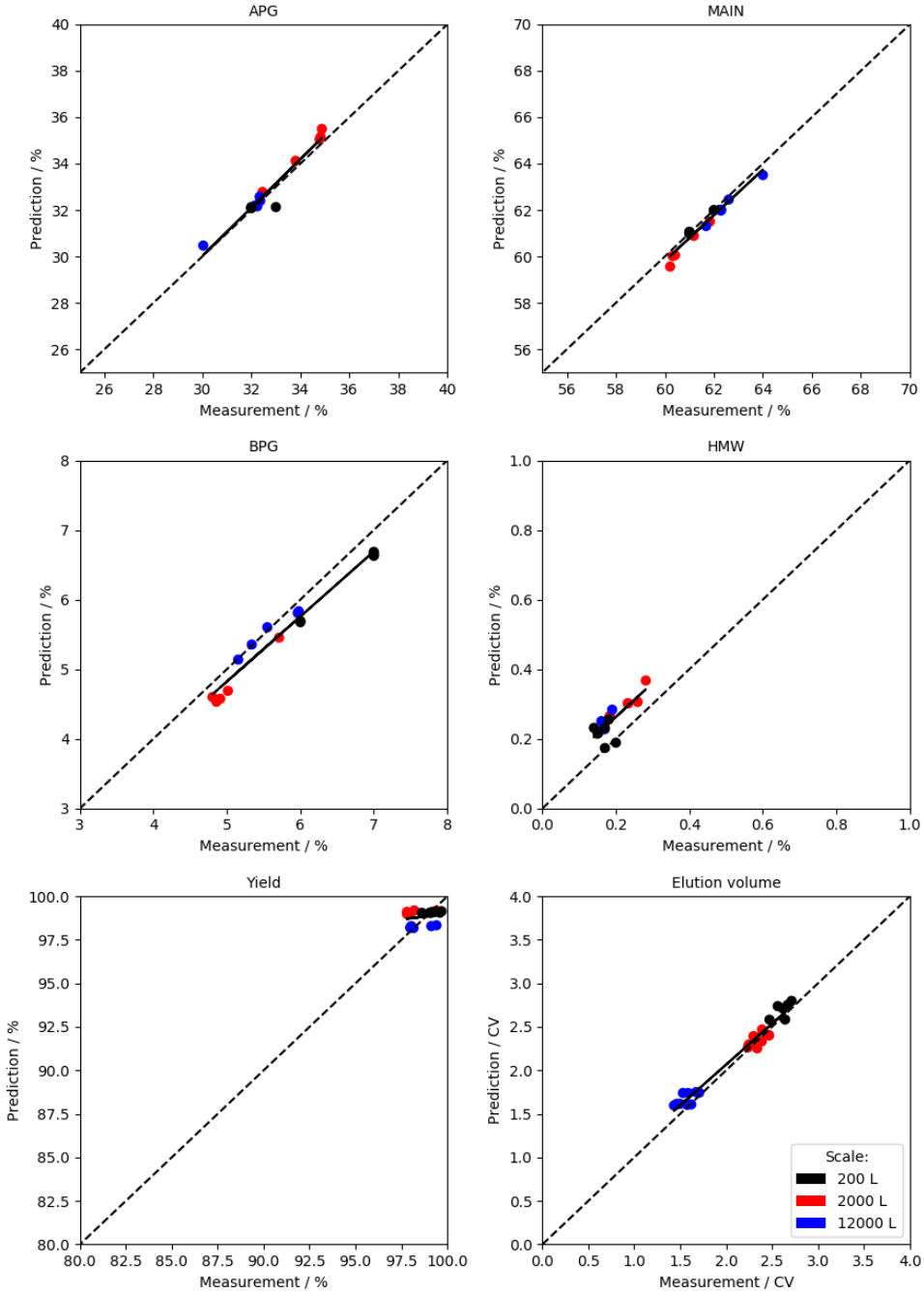

Figure 5.4: Cross-scale analysis of mechanistic model predictions for CQAs and KPI of the CEX unit operation at 200 L, 2000 L and 12000 L full manufacturing-scale. Solid lines show linear regression for the respective CQA/KPI across scales.

For HMW species with both mediocre R^2=0.65 and the NRMSEP=49%, a second case scenario is described. The distribution of data around the regression line is explained by the very low HMW level in the input material of approximately 0.4% and the elution pool ranging from 0.14% to 0.28%. Furthermore, total errors of 0.07% in HMW concentrations can originate from analytical method variability and/or different sample handling schemes between scales. In the second scenario, the slope and intercept of the regression line were taken into account being both close to the expectation of being 1 and 0, respectively. The results of a *t*-test with an error-probability of 5% confirmed the slope and intercept being not significantly different from the expectation. The trust in the model prediction is strengthened by these results from second stage model qualification. Additionally, a systematic shift of HMW concentrations along the y-axis in Figure 5.4 (D) indicates that the mechanistic model overestimated HMW concentrations for all investigated scales. This over estimation is most likely caused by differences in material and sample handling between manufacturing and process development laboratories. Specifically, the impact of freeze and thaw on mAb aggregate formation is a well know phenomenon and is the most probable root cause for the observed model offset in the <0.1% range.

The process step yield is considered a third case scenario as neither R^2 and NRMSEP, nor slope and intercept of linear regression support a belief in model prediction. The *t*-test for slope and intercept of the regression (p <0.05) rejects the hypothesis, that correlations in slope and intercept are significant. Measured yields across scales ranged from 97.8% to 99.7%. The observed variance is located within method variability of protein concentration and gravimetric volume measurements of the elution pools. For the investigated experiments, process parameters most likely had no measurable effect on the step yield. Therefore, the model's predictive power of yield is seemingly low and required a direct comparison between the model prediction and the manufacturing-scale outcomes in the final stage of model qualification as presented in the subsequent section.

5.4.3 Comparison of Mechanistic Model with Experimental ScDM

Equivalence between a process model and its respective full-scale unit operation is the fundamental requirement for applying the model to process characterization studies in late stage DSP development. The TOST is currently the state of the art equivalence testing approach for ScDM qualification [13, 226]. In Figure 5.5, the *in silico* model was qualified against manufacturing-scale data using the TOST qualification scheme described in section 5.3.2. ScDM qualification results are presented as a benchmark for mechanistic model performance. The data included 10 clinical manufacturing runs with two CEX cycles per harvest. The ScDM was limited to six load materials of clinical manufacturing runs, with three lab-scale experiments per load material. Further, ScDM experiments were conducted with all process parameters at set point conditions. In contrast, the mechanistic model can be seen as a digital twin of the manufacturing-scale process, considering variations in feed composition, column length, as well as mobile phase pH and salt concentration.

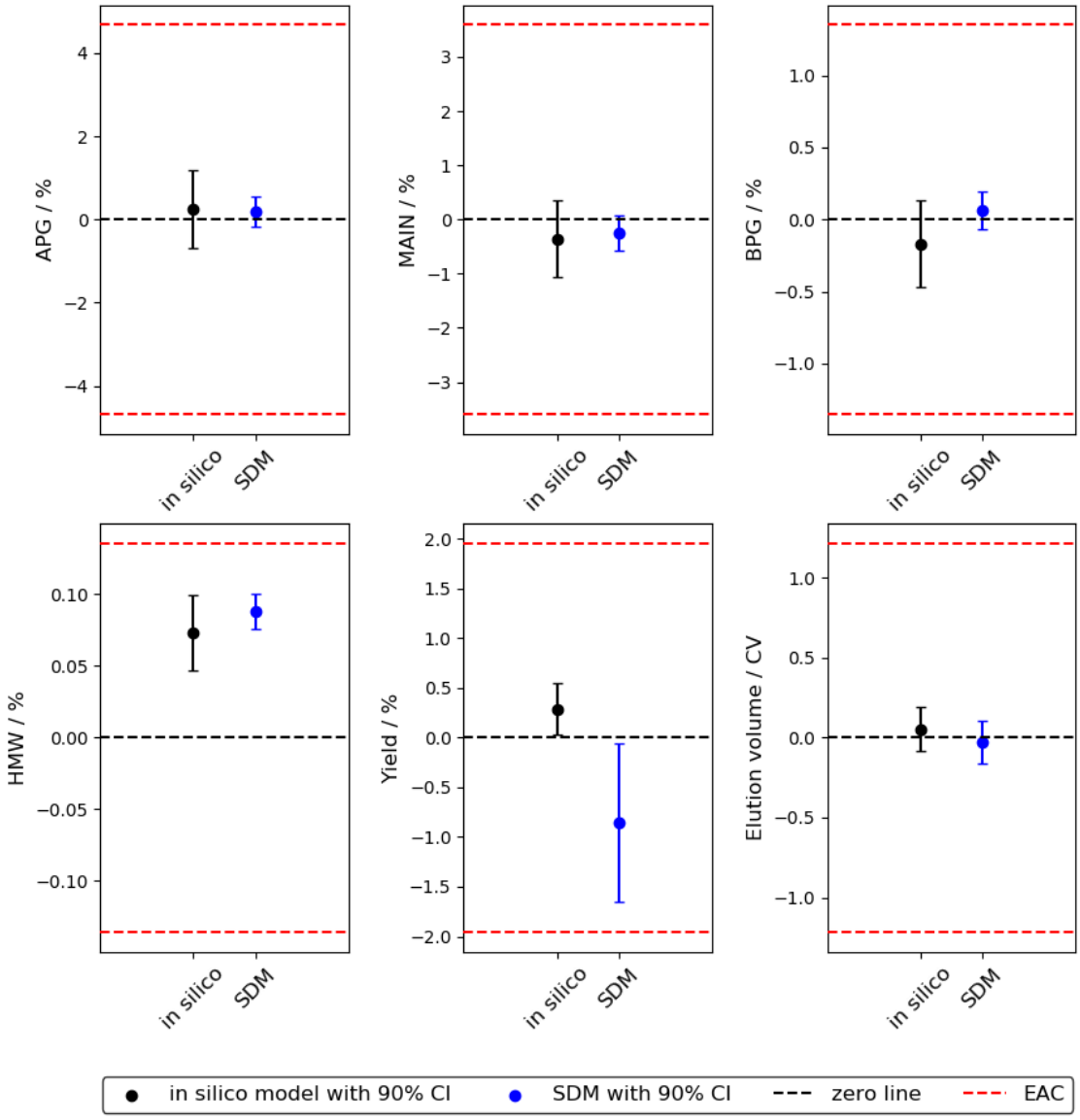

Figure 5.5: Equivalence test comparing large-scale results to mechanistic model prediction and experimental ScDM data. Visual representation of TOST analysis. The zero line represents the mean of 20 chromatography cycles for clinical manufacturing. p-values were < 0.05 for all shown CQAs and KPIs. EAC were defined as large-scale mean ± 3 SD of the respective CQA or KPI.

The visualization of TOST analysis in Figure 5.5 qualifies *in silico* model and experimental ScDM for all investigated CQAs and KPIs. The 90% CIs of model predictions were located within the acceptance criteria. Therefore, *in silico* model and experimental ScDM were equivalent to the large-scale unit operation. ScDM results and mechanistic model predictions for APG, Main, BPG concentrations and elution volume were located close to

the mean value of large-scale runs. In contrast, both models consistently overestimated HMW concentrations. The overestimation of HMW concentration was most likely caused by differing sample handling schemes as discussed in the previous chapter. Comparable sample handling procedures for ScDM and model calibration experiments in small-scale resulted in similar HMW values for ScDM and mechanistic model predictions. *In silico* predictions for CQAs led to broader CIs compared to the experimental ScDM. This observation was caused by the consideration of load material composition in all simulations, while the experimental ScDM was limited to six load materials. For example, the percentage of APG species in clinical manufacturing load materials varied between 30.4% and 35.5%, while the ScDM only included load materials with 32.1% to 34.1% APG content. The ScDM sample mean of yield was located within the EAC, but the error bars indicate a higher variance compared to *in silico* prediction. This variance was caused by analytical variability of protein concentration and gravimetric volume measurements of the elution pool during ScDM experimentation.

The TOST equivalence test is an objective method for model qualification. However, it is difficult to evaluate the true predictive capabilities of the mechanistic model by comparing mean values. Therefore, control charts in Figure 5.6 compare measured and predicted results of five 12000 L manufacturing runs with two CEX cycles per harvest. Input data from batch records was used for modeling every chromatographic cycle. Consideration of these input parameters allowed the mechanistic model to predict the trend of CQAs as a function of the run number. Additionally, the mechanistic model enabled quantitative predictions for yield and elution volume within the EAC.

The data shown in this section qualifies the mechanistic model as an *in silico* representation of the 12000 L scale unit operation for charge variants, HMW species, step yield and elution volume. Thus, the mechanistic model could be applied to process characterization studies for late stage DSP development. Further, the mechanistic model avoided limitations of the experimental ScDM by capturing the minimal effect of process parameter variation on the purification outcome. With a loading density of 45 g/L and a column bed height of 300 mm, a single ScDM experiment consumed more protein than the entire model calibration process on the 157 mm column. It is important to notice, that the presented model is limited to the six CQAs and KPIs. Additional CQAs, such as fragments, host cell protein or leached Protein A concentrations need to be added to the *in silico* model of the CEX unit operation if required.

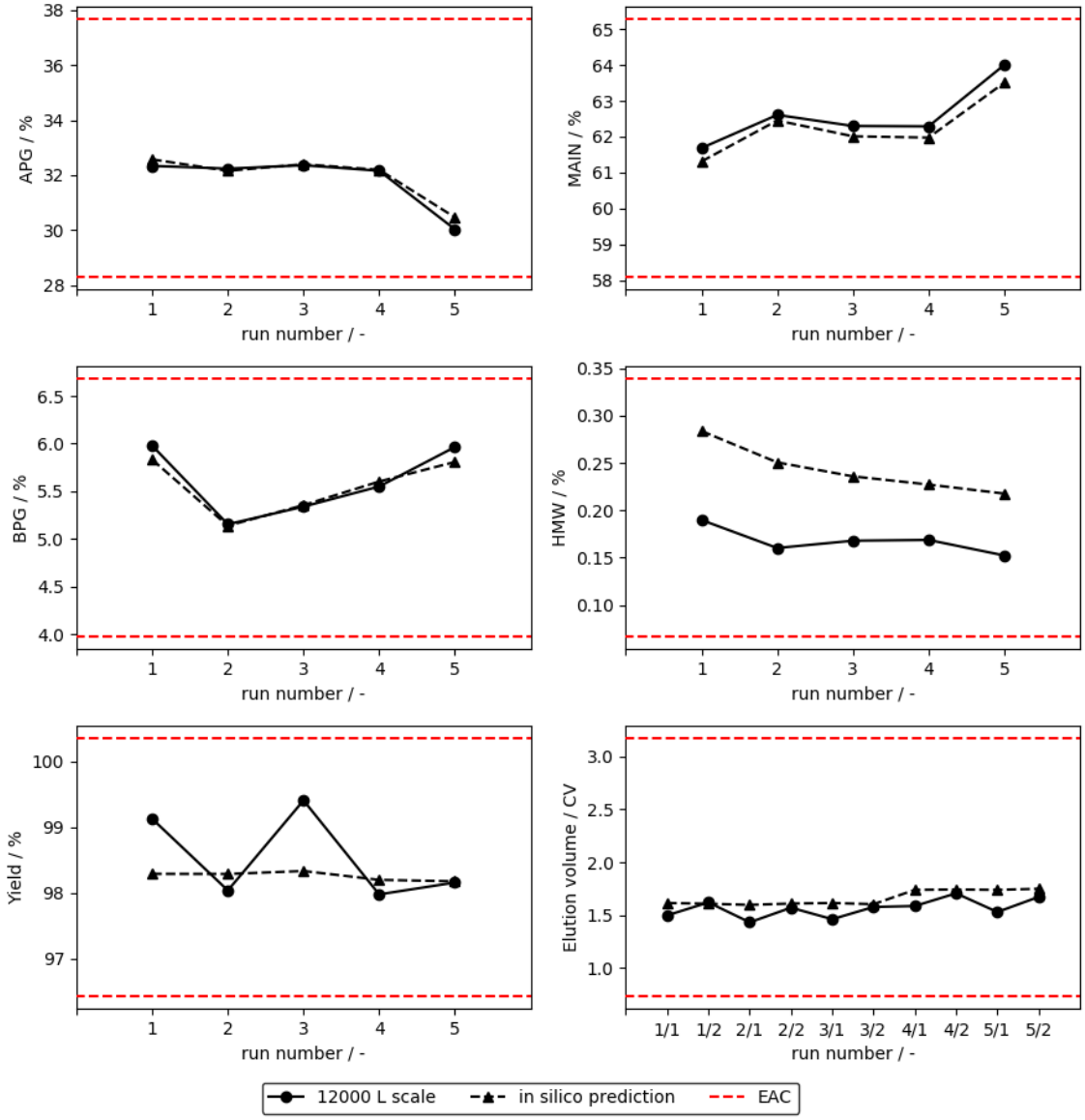

Figure 5.6: Control charts for CQAs and KPIs of 12000 L scale CEX purification runs. Load composition and mobile phase properties of each simulation were adapted according to the inputs of the respective chromatography cycle.

5.5 Conclusion

In the presented case study, a mechanistic model calibrated at bench-scale enabled the prediction of chromatography runs in multiple larger scales. *In silico* experimentation increased process understanding and allowed explanation of offsets between investigated scales. Consideration of scale dependent effects, such as pre-column dispersion and varying ionic capacities was found to be essential for accurate prediction of large-scale CEX processes. All investigated large-scale runs were performed at set-point conditions with only minimal variance of input parameters. The relatively small effect of loading density, input material composition, and mobile phase pH and salt concentration variability on the purification outcome challenged model accuracy. Nonetheless, the consideration of these relatively small input variations together with scale-specific features allowed accurate predictions of CQAs and KPIs across all scales from laboratory to manufacturing-scale. The predictive power across different column dimensions confirmed the physical relevance of the previously estimated model parameters [181]. For manufacturing-scale, model performance was directly benchmarked against the experimental ScDM. The main purpose of qualifying an experimental ScDM is to use it in subsequent process characterization studies. Both, ScDM and mechanistic model were successfully qualified against manufacturing-scale using well-established equivalence testing procedures. Additionally, the mechanistic model could describe the run-to-run trend of CQAs and KPIs. Therefore, the application of a thoroughly calibrated and validated mechanistic model for process characterization purposes can be considered as a scientifically sound and suitable complementation to experimental approaches.

This work presents a systematic framework for qualification of mechanistic chromatography models prior to their applications to late stage biopharmaceutical process development. Rules provided allow a more objective and gradual decision-making. However, the definition of model quality criteria is a complex task involving deep technical understanding, statistics, and understanding of the pharmaceutical quality system.

5.6 Acknowledgement

The authors would like to acknowledge the support of many colleagues across process development, statistics and manufacturing teams, including Roland Pfänder, Stephanie Combe, Jessica Stolzenberger, Marco Kunzelmann, Carolin Ehrhart and Markus Stützle.

6 *In silico* Process Characterization for Biopharmaceutical Development following the Quality by Design Concept

David Saleh[1, 2], Gang Wang[2], Federico Rischawy[2], Simon Kluters[2], Joey Studts[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] Boehringer Ingelheim Pharma GmbH & Co. KG, Biberach, Germany
* Corresponding author: Jürgen Hubbuch, +49 721 608 47526, juergen.hubbuch@kit.edu

Abstract

With the Quality by Design (QbD) initiative, regulatory authorities demand a consistent drug quality originating from a well-understood manufacturing process. This study demonstrates the application of a previously published mechanistic chromatography model to the *in silico* process characterization (PCS) of a monoclonal antibody polishing step. The proposed modeling workflow covered the main tasks of traditional PCS studies following the QbD principles, including criticality assessment of 11 process parameters and establishment of their proven acceptable ranges (PARs) of operation. Analyzing effects of multi-variate sampling of process parameters on the purification outcome allowed identification of the edge-of-failure. Experimental validation of *in silico* results demanded approximately 75% less experiments compared to a purely wet-lab based process characterization study. Stochastic simulation, considering the measured variances of process parameters and loading material composition, was used to estimate the capability of the process to meet the acceptance criteria for critical quality attributes and key performance indicators. The proposed workflow enables the implementation of digital process twins as QbD tool for improved development of biopharmaceutical manufacturing processes.

6.1 Introduction

The biopharmaceutical industry is under an unprecedented pressure to implement technologies for rapid process development. Main reasons are rising numbers of monoclonal antibodies (mAbs) in development [3, 4] and strongly accelerated development timelines [184]. While achieving a short time-to-market timeline, mAb manufacturers have to ensure high product quality by following the Quality by Design (QbD) concept. The QbD concept demands a consistent product quality originating from an intrinsic quality built into the design and the control of the manufacturing process. In recent years, regulatory authorities and biopharmaceutical organizations formulated clear concepts for the implementation of QbD in pharmaceutical development [25, 103, 234].Yu *et al.* [25] listed the key elements for a development strategy that complies with the QbD concept:

- A quality target product profile (QTPP) for the identification of critical quality attributes (CQAs) of the drug product
- Identification of critical material attributes (CMAs) and critical process parameters (CPPs) potentially effecting CQAs
- Measuring the effect of CPPs and CMAs on CQAs
- Development of a control strategy
- Process capability and continual improvement.

Development workflows comprising the above listed QbD elements make use of general process knowledge and statistical design of experiments (DoE) for the characterization of a unit operation [13, 14, 235]. For many process steps, including preparative chromatography, it is not feasible to include all controllable process parameters in a DoE study. Even on small-scale systems [97], it is challenging to screen hundreds of process conditions when considering the subsequent analytical bottleneck. Therefore, process parameters have to undergo a risk-based criticality assessment considering their potential impact on CQAs before designing an experimental process characterization study (PCS) [13, 14]. Risk-based decision trees for process parameter classification are able to reduce the dimensionality of DoE studies, and thus reduce the experimental burden for process development. However, parameter criticality assessment can be influenced by subjective decision-making caused by the lack of experimental data at this development stage. As a result, incorrectly classified process parameters could lead to avoidable experimental effort, or worse, to a poorly understood control strategy. Further, PCS approaches based on DoE are limited to regression models correlating CPPs to CQAs with a limited amount of data points per CPP.

In the ICH Q8/Q9/Q10 (R2) documents [24], regulatory authorities propose the use of mathematical models to support bioprocess development and manufacturing. These models include mechanistic models describing the physical phenomena within a unit operation,

which can be used to predict process outcomes under varying conditions [24]. Digitalization initiatives in biopharma industry and academia identified mechanistic chromatography modeling as a promising tool for *in silico* development of downstream processes (DSP) [16, 22, 23, 164]. After overcoming the initial hurdle of model calibration [114, 165, 171, 181, 190], mechanistic chromatography models show a broad applicability to bioprocess development, including process optimization [161], robustness analysis [101, 162, 206], or scale-up [27, 97]. Recently, Andris *et al.* [236] developed a mechanistic model for the separation of antibody-drug conjugates. Their work allowed the characterization of a design space, revealing the relevance of digital process twins in the light of QbD. For ion exchange chromatography, Jakobsson *et al.* [162] used mechanistic modeling to design a robust pooling strategy under consideration of model uncertainty. A mechanistic modeling study performed by Close *et al.* [101] identified robust operating conditions for a hydrophobic interaction chromatography process, where resin and loading material had a considerable impact on process performance. Following the QbD concept, Rischawy *et al.* [163] used mechanistic modeling for the identification of CPPs for a cation exchange chromatography step applied to the polishing of a bispecific mAb. Shekhawat *et al.* [227] developed a model that improved understanding around resin fouling in Protein A chromatography. The here mentioned mechanistic modeling studies increased process understanding or solved specific problems regarding process robustness. However, as described earlier, regulatory authorities defined clear perspectives on the implementation of QbD in process development and the related tasks. To the best of our knowledge, it is still to be shown how mechanistic models could be applied to a process characterization study addressing the essential QbD elements.

Our previous publications introduced a quality system for mechanistic chromatography modeling in biopharmaceutical process development. The selection of representative experiments for model calibration ensured adequate model certainty with minimal resources [181]. This mechanistic model was validated against data of multiple scales, including clinical manufacturing-scale [237]. As a sequel of this publication series, the mechanistic model is applied to the process characterization of a cation-exchange chromatography (CEX) step. Simulations are performed at manufacturing-scale avoiding limitation of experimental scale-down model studies. The *in silico* strategy aims to fulfill the fundamental tasks of a PCS following the QbD concept. This includes criticality assessment of process parameters and measuring their effect on CQAs and key performance indicators (KPIs). Further, simulations provide the database to identify proven acceptable ranges (PARs) for process parameters as part of the control strategy. An experimental design is derived from mechanistic model predictions to reduce the experimental effort compared to wet-lab driven DoE approaches. As a last element, Monte-Carlo simulation allows the calculation of process capability under consideration of CPP, KPP, KMA, and CMA variances measured during clinical manufacturing. The presented methodology generates in-depth process understanding following the QbD concept, while debottlenecking experimental limitation of DoE approaches. Mechanistic modeling for *in silico* process characterization can improve decision-making in DSP development, assuring product quality throughout the entire value chain.

6.2 Modeling

Details about model discrimination, model parameters, the model calibration strategy, and scale-dependent considerations can be found in our previous publications [181, 237]. Protein-specific model parameters are listed in Table 6.1. This section gives an overview on the mechanistic model and complementations necessary for model-guided scale-up. The one dimensional (1D) transport dispersive model was selected as column model, due to multiple successful case studies for the simulation of ion exchange chromatography systems [99, 114, 116, 167, 190]. Eq. 6.1 describes the macroscopic transport of component i through the chromatography column. The change of the concentration c_i at position x in time t is a function of convective mass transport in the interstitial volume, peak broadening caused by axial dispersion D_{ax}, and mass transfer from the interstitial volume into the pore phase of the particle with the radius r_P. Further, mass transfer between the interstitial volume and the particle pores is affected by the interstitial porosity ε_{col} and the effective mass transfer coefficient $k_{eff,i}$. The accumulation of mass in the pore phase $c_{p,i}$ with the particle porosity ε_p and the stationary phase q_i is described in Eq. 6.2. The Danckwerts' boundary conditions are given in Eqs. 6.3 and 6.4.

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

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

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

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

Linear flow rates ranged from 155 cm/h to 360 cm/h between investigated scales, demanding the introduction of flow dependencies for the axial dispersion coefficient D_{ax} and effective mass transfer parameter $k_{eff,i}$. The penetration correlation allowed the direct calculation $k_{eff,i}$ for monomer and HMW species at relevant flow rates, respectively [175, 216]. Within the investigated range, flow dependencies for D_{ax} and k_{eff} could be approximated using linear regression according to Eq. 6.5 and 6.6.

$$D_{ax}(u) = D_{ax0} + uD_{ax1} \tag{6.5}$$

$$k_{eff}(u) = k_{eff0,i} + uk_{eff1,i} \tag{6.6}$$

Protein adsorption is simulated using the semi-mechanistic SMA adsorption model [117]. The multicomponent SMA model formulates the equilibrium binding behavior of the protein in consideration of the salt concentration in the pore phase c_s, the ionic capacity of the resin Λ, and the proteins characteristic charge ν_i. Eq. 6.7 shows the kinetic form of the SMA isotherm modified by Hahn *et al.* [167], where $k_{eq,i} = k_{ads,i}/k_{des,i}$ and $k_{kin,i} = 1/k_{des,i}$ describe adsorption and desorption rates of component i, respectively. In addition, the steric shielding parameter σ_i denotes the number of functional groups on the resin surface blocked by the protein.

$$k_{kin,i}\frac{\partial q_i}{\partial t} = k_{eq,i}(pH)\left(\Lambda - \sum_{j=1}^{k}\left(\nu_j(pH) + \sigma_j\right)q_j\right)^{\nu_i(pH)} c_{p,i} - q_i c_s^{\nu_i(pH)} \tag{6.7}$$

$$q_{salt} = \Lambda - \sum_{j=1}^{k}\nu_j(pH)q_j \tag{6.8}$$

The introduction of pH-dependent isotherm parameters is crucial for industrial applications. Eqs. 6.9 and 6.10 show the empirical pH dependencies of the characteristic charge ν_i and the equilibrium constant $k_{eq,i}$ developed by Hunt *et al.* [115] This model was found to be sufficient for the process relevant pH range of pH 5.8 $\pm$0.3 used in this study [181].

$$k_{eq,i}(pH) = k_{eq0,i}e^{k_{eq1,i}pH + k_{eq2,i}pH^2}, \tag{6.9}$$

$$\nu_i(pH) = \nu_{0,i} + pH\nu_{1,i}, \tag{6.10}$$

6.3 Material and Methods

6.3.1 CEX unit operation

The investigated protein is an IgG1 mAb expressed in stably transfected Chinese hamster ovary cells (Boehringer Ingelheim GmbH & Co. KG, Biberach, Germany). The mAb was

Table 6.1: Protein specific model parameters for the pH-dependent SMA model. Details regarding the model calibration procedure are described in our previous publication.36 For a clear representation of model parameters at pH 5.8, the pH was normalized to zero. pH 5.5 = -0.3, pH 5.8 = 0, pH 6.1 =0.3.

Parameter	APG	Main	BPG	HMW
$k_{eff0,i}$ [mm/s]	1.4E-3	1.4E-3	1.4E-3	1.2E-3
$k_{eff1,i}$ [-]	4.7E-05	4.7E-05	4.7E-05	3.3E-05
$\nu_{pH5.8,i}$ [-]	7.38	7.50	7.70	10.97
$\nu_{1,i}$ [-]	-1.44	-1.44	-1.44	-6.77
$k_{eq,pH5.8,i}$ [-]	1.45	1.41	1.69	1.86
$k_{eq,1,i}$ [-]	-4.26	-4.26	-4.26	-5.39
$k_{eq,2,i}$ [-]	2.19	2.19	2.19	5.59
$k_{kin,i}[sM^{\nu}]$	8.08E-06	1.00E-04	5.00E-04	3.4E-05
σ_i [-]	128.6	56.3	107.1	0

captured *via* Protein A affinity chromatography and further polished using anion exchange chromatography in flow-through mode. The presented mechanistic model describes the subsequent CEX unit operation using the strong CEX resin POROS 50 HS (Thermo Fisher Scientific, Waltham, USA). The process was performed at constant pH 5.8 in bind-elute mode and at a maximal load density of 45 g/L_{Resin}. The column was equilibrated at a counter-ion concentration of 87 mM Na^+, with the same buffer applied to the wash phase after column loading. Subsequently, elution was induced at a counter-ion concentration of 247 mM Na^+. For column regeneration and storage, 1 M and 0.1 M NaOH were applied respectively. Selected experiments from wet-lab PCS studies were used to validate the most critical relationships between process parameters and CQAs/KPIs. Bench-scale experiments were performed on an Äkta Avant 25 (Cytiva, Uppsala, Sweden) using an experimental ScDM column with a bed height of 300 mm and an inner diameter of 10 mm.

Charge variant and HMW concentrations in the elution pool were quantified using analytical CEX chromatography and analytical size exclusion chromatography, respectively. Acidic (APG), neutral (Main) and basic peak groups (BPG), as well as HMW species were considered as CQAs. Process step yield and elution volume were defined as KPIs and quantified using protein concentration determined *via* absorbance at 280 nm and gravimetric volume measurement. Details of the model calibration and validation, as well as analytical chromatography methods, are presented in one of our previous publications [181].

6.3.2 *In silico* PCS Workflow

This chapter describes the methodology of an *in silico* process characterization following the QbD concept. The PCS workflow consisted of three major building blocks: (1) process parameter criticality assessment and establishment of PARs; (2) identification

and validation of the edge-of-failure; (3) calculation of process capabilities. Protein- and system-specific mechanistic model parameters were kept constant and were obtained from our previous publications [181, 237]. Only process parameters were varied during *in silico* sampling. All simulations were performed at manufacturing-scale, under consideration of system and column dimensions. Before starting *in silico* experimentation, the applied mechanistic model was validated as digital representation of the real-world process. Model validation must consider the intended purpose of the model and its potential impact on the control strategy at manufacturing-scale. Small-scale experiments validated that the model captures the impact of process parameter variation on the purification outcome [181]. Model validation across scales showed that the model captures relevant system effects and proofed equivalence between the mechanistic model and manufacturing-scale [237]. Based on this previously published validation strategy, it is reasonable to use the model for *in silico* process characterization.

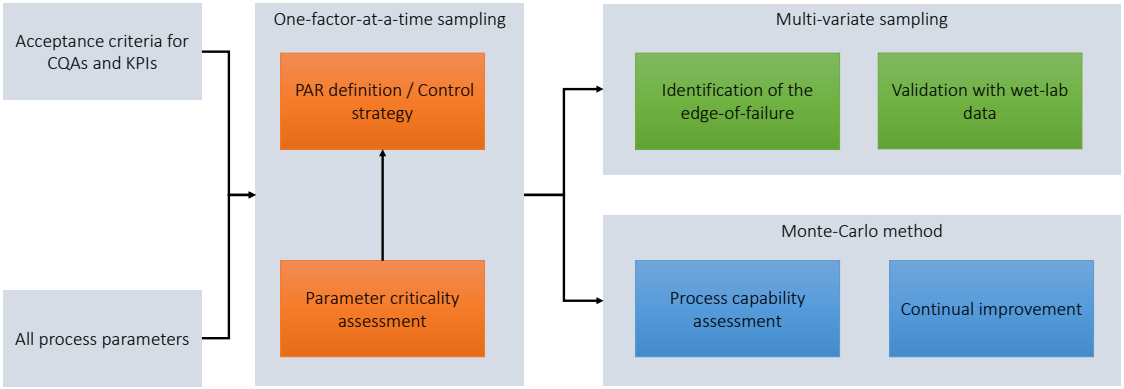

Figure 6.1: *In silico* process characterization of a unit operation for monoclonal antibody purification.

Figure 6.1 depicts the three different parameter-sampling methods. Initially, a one-factor-at-a-time sampling (OFAT) scheme enabled criticality assessment of process parameters and definition of proven acceptable ranges (PAR). During OFAT sampling, one parameter was sampled in a wide range around its intended set point, while the other process parameters were kept constant. The loading density was sampled below its upper limit of 45 g/L_{Resin}. Following the decision tree in Figure 6.2, process parameters were ranked as non-KPP, CPP, KPP, CMA or KMA based on their effect on CQAs and KPIs.

- non-KPP: Process parameter does not affect a CQA or KPP
- CPP: Critical process parameter affects at least one CQA
- KPP: Key process parameter affects at least one KPI and not affects CQAs
- CMA: Critical material attribute affects at least one CQA
- KMA: Key material attribute affects at least one KPI and not affects CQAs

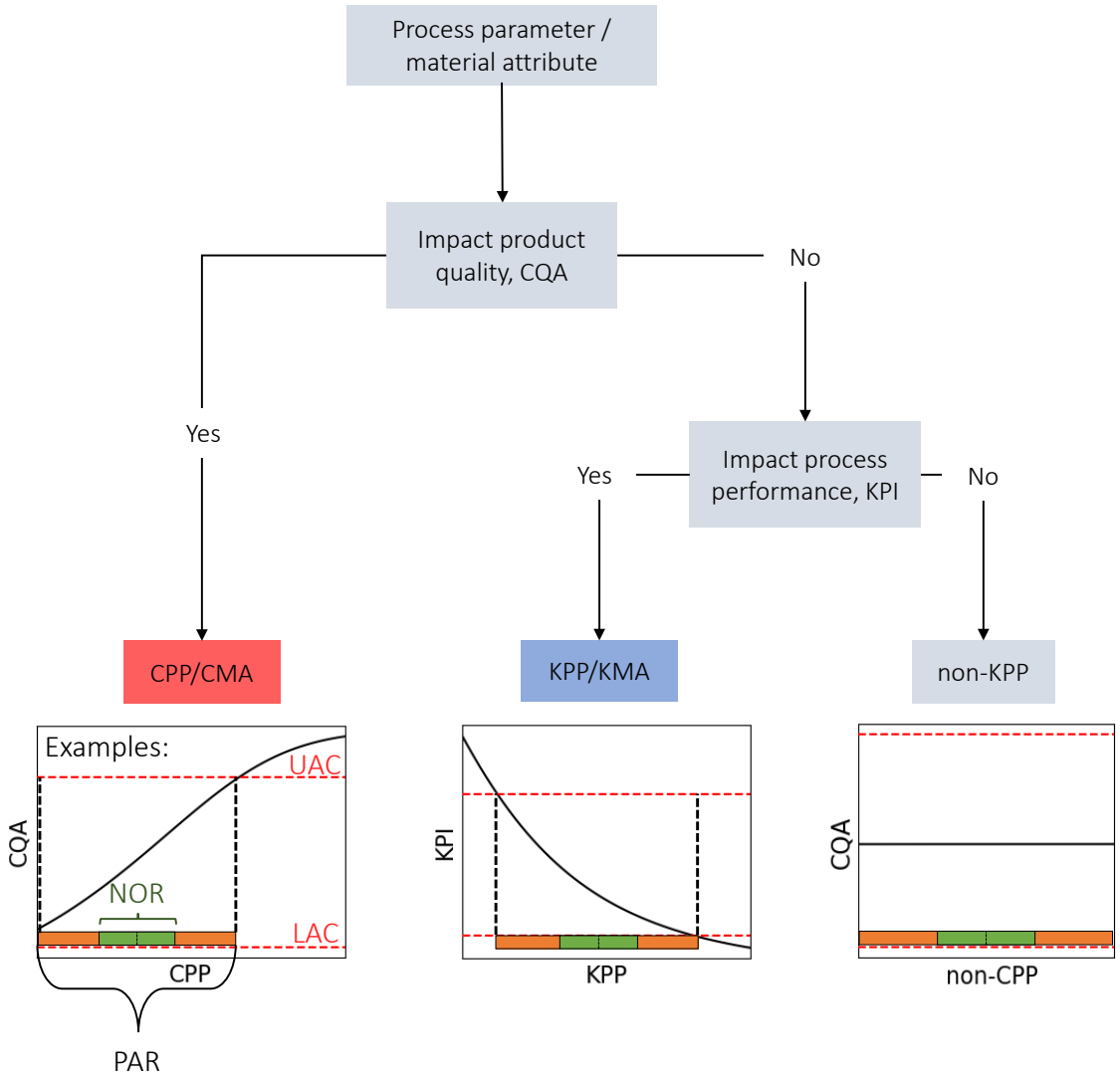

Figure 6.2: Decision tree for model-guided criticality assessment of process parameters and establishment of their PARs. UAC = Upper acceptance criterion, LAC = Lower acceptance criterion

Subsequently, the same data obtained from OFAT sampling allowed definition of PARs for all investigated process parameters. The establishment of PARs is a fundamental part of the control strategy and represents the main goal of a PCS. According to the European Medicines Agency (EMA) and ICH Q8 R2 guideline [24, 238], the PAR is defined as the operating range of a process parameter for which the unit operation will produce a drug substance meeting the relevant quality criteria. When all process parameters are kept constant, but one parameter varies within its PAR, all CQAs and KPIs measured in the elution pool must be located within their predefined acceptance criteria (AC). Thus, OFAT sampling of input parameters is a suitable method for the establishment of PARs.

As presented in Figure 6.2, the intersection of ACs and the curve obtained *via in silico* sampling defined the lower and upper boundary of the PAR. If a process parameter did not cause CQAs or KPIs to violate the AC, the entire *in silico* screened parameter range of this process parameter was defined as PAR.

Process parameters ranked as CPPs and KPPs were analyzed in subsequent multi-parametric sampling studies. The multi-parametric sampling study represented the second building block of the *in silico* PCS. Here, CPPs and KPPs were varied jointly to study the worst-case operating scenarios. This procedure enabled the identification of the edge-of-failure under consideration of the AC. Historical wet-lab experiments at process conditions around the edge-of-failure were used to validate the *in silico* findings.

In a last step, the process capability of the unit operation was calculated based on stochastic simulation (Monte-Carlo simulation), as described in Figure 6.3. Therefore, probability functions of process parameters and loading material composition were calculated based on 20 chromatographic cycles at clinical manufacturing-scale. Subsequently, 1000 simulations were performed using random samples of the previously determined probability function as model input. The resulting CQA and KPI distributions were then plotted and compared to the AC. The standard deviations $\hat{\sigma}$ obtained from *in silico* generated CQA and KPI distributions enabled calculation of the corresponding process capabilities c_{pl} and c_{pu} for the lower and upper AC (LAC and UAC), respectively,

$$c_{pl} = \frac{Mean - LAC}{3\hat{\sigma}}, \tag{6.11}$$

$$c_{pu} = \frac{UAC - Mean}{3\hat{\sigma}}, \tag{6.12}$$

$$c_{pk} = min(c_{pl}; c_{pu}). \tag{6.13}$$

For each CQA or KPI, the overall process capability c_{pk} was defined as the minimum of c_{pl} and c_{pu}. When only an LAC or an UAC was defined, the overall process capability could be simplified to $c_{pk} = c_{pl}$ or $c_{pk} = c_{pu}$.

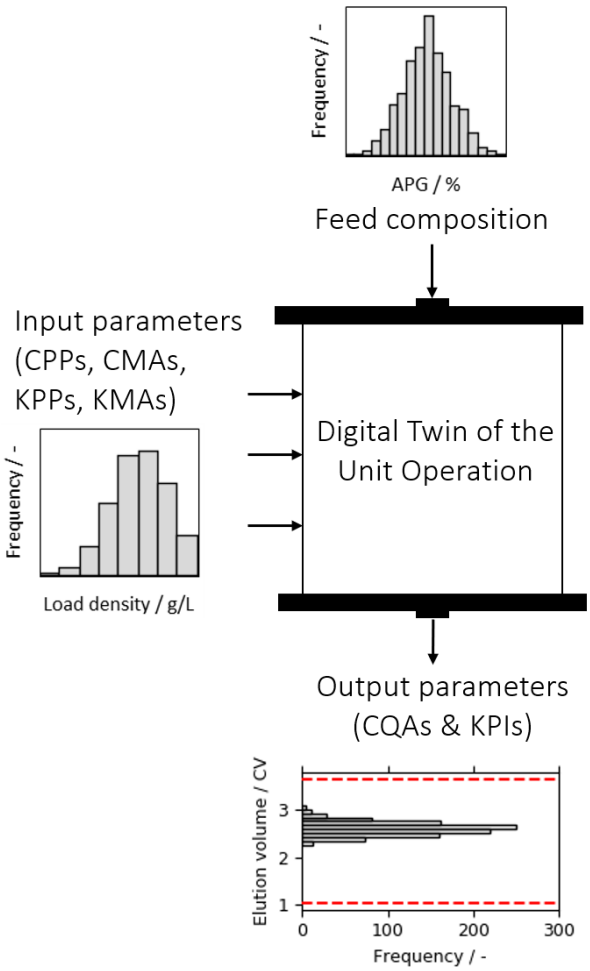

Figure 6.3: Monte-Carlo simulation for the calculation of process capability. The stochastic simulation procedure considered loading material compositions and input parameter distributions resulting in the calculation of process capabilities for six CQAs and KPIs. Exemplary input and output distributions are shown.

6.4 Results and Discussion

In the following chapters, a previously published mechanistic chromatography model was applied to the *in silico* PCS of a CEX unit operation [181, 237]. The multi-stage modeling workflow aimed to fulfill essential tasks of process characterization following the QbD concept. This includes CPP identification and PAR definition. Multi-parametric effects on the purification outcome were identified and validated with wet-lab experiments. Monte-Carlo simulation allowed the determination of process capability under consideration of real CPP, KPP, CMA, and loading material composition variability.

6.4.1 Parameter Criticality Assessment and Control Strategy

Before starting characterization of a unit operation, process parameters must be classified according to their impact on CQAs and KPIs. The mechanistic chromatography model enabled effect analysis of process parameters following an OFAT sampling scheme. Table 6.2 lists the results of the parameter criticality assessment following the decision tree in Figure 6.2. *in silico* investigation of one process parameter consisted of 50 simulations, with equidistant steps in a wide range around the set-point condition. While one parameter was varied, all other process parameters were kept on the set point. Five process parameters were ranked as CPPs or CMAs, showing effects on at least one CQA. Salt concentration in the equilibration/wash buffer was ranked as KMA, since it only affected the KPI step yield. All remaining process parameters were ranked as non-KPPs and did not affect CQAs or KPIs within the screened parameter ranges.

Mobile phase pH and salt concentrations were amongst the process parameters showing the strongest impact on CQAs and KPIs. Thus, Figure 6.4 highlights the effects of mobile phase conditions during equilibration/wash and elution on the purification result. The non-linear correlation between elution pH and HMW concentration was identified as the most considerable effect. The mechanistic model predicted that an elution buffer with pH above pH 5.9 results in HMW levels violating the upper AC. Typically, the initial criticality-assessment of process parameters is based on failure mode and effect analysis (FMEA). The FMEA allows a risk-ranking depending on initial experiments and available data from process development, historical knowledge from different mAbs at comparable process steps, and process understanding of subject matter experts. A validated mechanistic model could be used to support a knowledge-based FMEA. The effects of potential CPPs and KPPs identified *via* FMEA on CQAs and KPIs are then screened in a DoE approach. The *in silico* OFAT screening allowed a rationalized identification of critical input parameters without experimental limitations. Process understanding leveraged from 550 simulations was used to generate the parameter classifications given in Table 6.2. Such a number of experiments containing similar amount of information cannot be screened economically in wet-lab.

Table 6.2: Criticality assessment of *in silico* screened process parameters of the CEX unit operation. Process parameters were classified according to the decision tree depicted in Figure 6.2.

Process parameter	Unit	Tested range	Effect on CQA	Effect on KPI	Classification
pH elution buffer	pH	5.5 - 6.1	Yes	Yes	CMA
Salt elution buffer	mM Na^+	230 - 265	Yes	Yes	CMA
Flow rate elution	cm/h	100 - 350	Yes	Yes	CPP
pH equilibration/ wash buffer	pH	5.5 - 6.1	Yes	Yes	CMA
Loading density	g/L_{Resin}	22.5 - 45	Yes	Yes	CPP
Salt equilibration/ wash buffer	mM Na^+	74 - 99	No	Yes	KMA
Flow rate loading	cm/h	100 - 350	No	No	non-KPP
pH load	pH	5.5 - 6.1	No	No	non-KPP
Salt load	mM Na^+	62 - 85	No	No	non-KPP
Flow rate wash	cm/h	100 - 350	No	No	non-KPP
Column length	mm	270 - 330	No	No	non-KPP

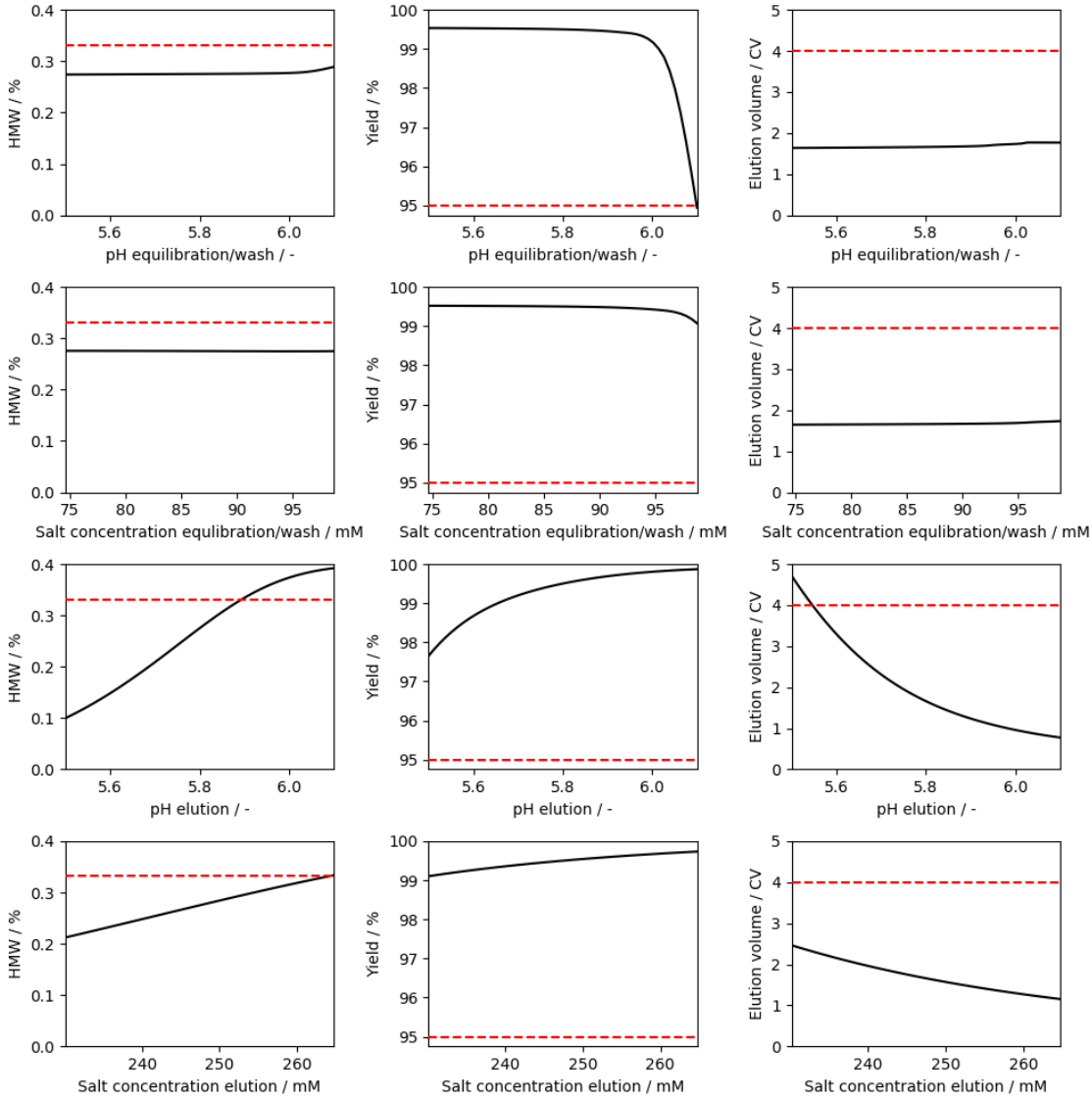

Figure 6.4: Criticality assessment of process parameters *via in silico* OFAT sampling. Each sub-figure contains the information of 50 simulations at varying process conditions. The figure shows effects of mobile phase conditions during elution and wash phase on CQAs and KPIs.

Following the methodology described in section 6.3.2, upper and lower limits for PARs (not presented in numbers) were directly derived from the intersection of simulated data and predefined AC in Figure 6.4. PARs can be established using wet-lab data obtained from OFAT or DoE studies. Due to experimental limitations, data evaluation is often limited to first- or second-degree regression modeling. The data in Figure 6.4 reveals how the non-linear correlations between CQAs and CPPs affect the establishment of PARs. Simple regression modeling based on a small number of experiments would result in different PARs. The *in silico* established PARs could be used as part of the control strategy during commercial manufacturing. From a regulatory perspective, the process understanding obtained *via* OFAT sampling based on mechanistic modeling represents a Level 3 control strategy [25]. The process control assures product quality meeting the specifications when a single process parameter deviates within its PAR. Further, the effect of deviating process parameters on CQAs and KPIs is well-understood enabling the possibility to adapt controls upstream in the process chain. Potentially, the mechanistic model could be applied to a Level 1 control strategy, substituting traditional testing of the intermediate product. In this case, continual generation of *in silico* data would enable automated adjustment of process parameters assuring a consistent product quality within the AC [25]. Application of mechanistic models as soft-sensors in a Level 1 control strategy could be useful for continuous manufacturing [17, 239], when the adoption of process analytical technology (PAT) is not feasible.

The applied mechanistic model considered large-scale column dimensions and properties. Further, the model was validated against manufacturing-scale data. Consequently, the mechanistic model enabling *in silico* PAR definition was representative to the final manufacturing-scale. Traditional DoE approaches rely on scale-down experimentation. The ICH guidelines support the establishment of PARs using small-scale experimentation. However, all simplifications and assumptions made during ScDM experimentation must be justified during approval process. A pure *in silico* PCS is currently not recommended if not all CQAs are fully covered by the mechanistic model. Therefore, the following section focuses on a minimal amount of wet-lab experiments for validating the relevant correlations between process parameters and CQAs and KPIs.

6.4.2 Identification and Validation of the Edge-of-failure

In the previous section, *in silico* OFAT screening enabled classification of process parameters and establishment of PARs. This chapter aims to validate the identified effects of process parameters on CQAs and KPIs using experimental data obtained from previous wet-lab PCS experiments. Here, the minimal number of experiments demanded for validation of *in silico* results was compared to the experimental effort of an entirely wet-lab based PCS. The multi-variate sampling investigated effects on step yield and aggregate concentration.

Amongst the investigated parameters, mobile phase properties showed the strongest impact on HMW removal and step yield. Mechanistic model predictions showed that an

increased counter ion concentration and mobile phase pH during the wash phase caused an early desorption of protein, negatively affecting step yield. Both process parameters, wash salt concentration and wash pH, were simultaneously varied *in silico* using a parametric sweep study consisting of 400 simulations. As a result, Figure 6.5 (A) shows step yield as a function of mobile phase conditions during the wash phase. The edge-of-failure was defined as the cutting curve of the surface function calculated based on *in silico* results and the AC for step yield. Set-point conditions for wash salt and pH conditions are located in the center of x- and y-axis, respectively. Therefore, the contour plot reveals that step yield cannot fall below the AC when varying only one factor at a time. When increasing both, salt and pH during wash above set-point conditions the step yield drops from >98% to a minimum of 77% within the investigated parameter space. Elution of protein during the wash phase resulted in non-linear correlations between process parameters and step yield, which would be difficult to cover using an experimentally limited DoE approach coupled with empirical response surface modeling. The selection of wet-lab experiments at conditions close to the edge-of-failure (scatter plot in Figure 6.5) validated that a simultaneous increase of salt concentration and pH during the wash phase would result in a violation of the AC for step yield. Instead of conducting wet-lab experiments in the entire parameter space, *in silico* identification of the edge-of-failure enabled a reduction of the experimental design to process conditions relevant for proofing process robustness.

The identical methodology was applied to mobile phase conditions during the elution phase and their effect on HMW concentration in the elution pool. Figure 6.5 (B) depicts HMW concentration as a function of elution salt concentration and elution pH. Compared to elution salt concentration, the elution pH had a strong impact on the HMW levels in the product. Again, wet-lab experiments around the edge-of-failure could validate the correlations obtained using *in silico* data. With targeted experiments close to the *in silico* determined edge-of-failure, the total number of wet-lab experiments was reduced from 35 to 9 compared to the traditional DoE-based PCS. The contour plot in Figure 6.5 (B) supports the finding of the previous OFAT analysis, that elution pH 5.7 could be a more robust set point, showing an increased distance to the edge-of-failure compared to pH 5.8. The true capability of the process to deplete HMW species in the desired quantity demands further *in silico* analysis considering material and process parameter variability.

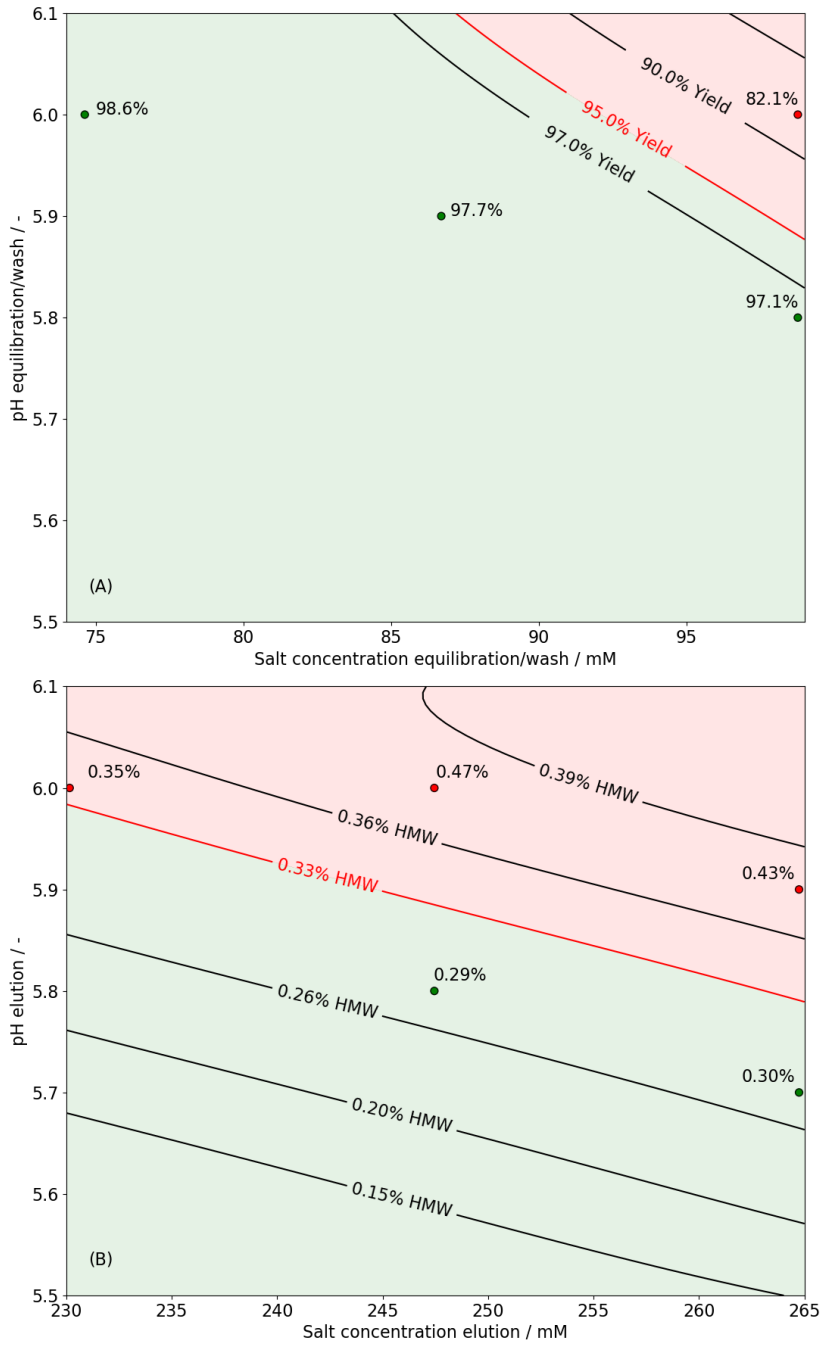

Figure 6.5: Effect of mobile phase conditions on step yield (A) and HMW removal (B). Scatter plots show the results of wet-lab experiments performed at process conditions close to the edge-of-failure (green = within AC, red = outside AC). Red contours represent the edge-of-failure, as the cutting line of model prediction and AC. Each contour plot is calculated based on 400 simulations at varying process conditions.

6.4.3 Process Capability and Continual Improvement

Following the QbD elements described by Yu et al. [25], process capability and continual improvement represents the final building block of the *in silico* PCS. Process capability c_{pk} describes the ability of the purification process to achieve CQAs and KPIs located within the AC under consideration of the intrinsic process variability. As depicted in Figure 6.3, Monte-Carlo simulation enabled calculation of process capabilities. Feed stream and process parameter variability was used as model input. The input distributions were obtained from 20 CEX chromatography cycles at clinical manufacturing-scale. Variance in load composition and mobile phase properties were approximated with Gaussian probability functions. The input variance of the loading density was described by an asymmetric Gaussian function, with a maximum of 45 g/L_{Resin}. 1000 samples were taken from the distributions calculated based on manufacturing-scale data. The intended mobile phase pH value of the unit-operation was pH 5.8 for all chromatographic phases. Although a pH range of pH 5.8±0.1 is well controllable, simulations in Figure 6.4 suggest that pH 5.7 is a more robust set point for HMW removal. Therefore, both elution pH scenarios were evaluated using the Monte-Carlo method. Figure 6.6 and Figure 6.7 show the resulting distribution of CQAs and KPIs for pH 5.8 and 5.7, respectively.

The comparison between Figure 6.6 and Figure 6.7 reveals that a reduction of the elution pH from pH 5.8 to pH 5.7 increases process capability for HMW removal when considering the intrinsic variance of the CEX unit operation. The capability of the process to achieve an HMW concentration below the AC increased from 0.43 to 1.54. Assuming normal distribution of model outputs, the probability for an HMW concentration be located outside the AC reduced to 0.0004% from 19.4%. The adaption of the elution set point pH had no negative effect on process capabilities of other CQAs and KPIs. Consequently, Monte-Carlo simulation could support the decision to shift the set point pH from pH 5.8 to pH 5.7.

Despite the simplification of assuming normal distribution for the majority of CPPs, KPIs, CMAs, and KMAs as model input, step yield and elution volume showed an asymmetric distribution at pH 5.7. These trends underline the importance of considering non-linear correlations in preparative chromatography. Similar to the loading material compositions, charge variant concentrations in the elution pool were found to be normally distributed. Process capabilities for charge variants ranged between 0.92 and 1.23. A $c_{pk} = 1$ corresponds to a distance of 3 sigma between the mean output value and the AC, resulting in a 0.27% probability for a CQA or KPI to be located outside the AC. Probabilistic simulation using mechanistic modeling is a simple and effective way to estimate process capabilities before a sufficient amount of real data from commercial manufacturing campaigns is available. Here, adaption of the set point pH based on Monte-Carlo simulation improved process robustness with regards to aggregate removal and reduced the risk of an out of specification (OOS) event. During the product lifecycle, input distributions for CMAs, KMAs, CPPs. and KPPs can be continuously updated and fed-back into the mechanistic model. This procedure would allow an early identification of root-causes for process variability enabling an adjustment of the control strategy if needed.

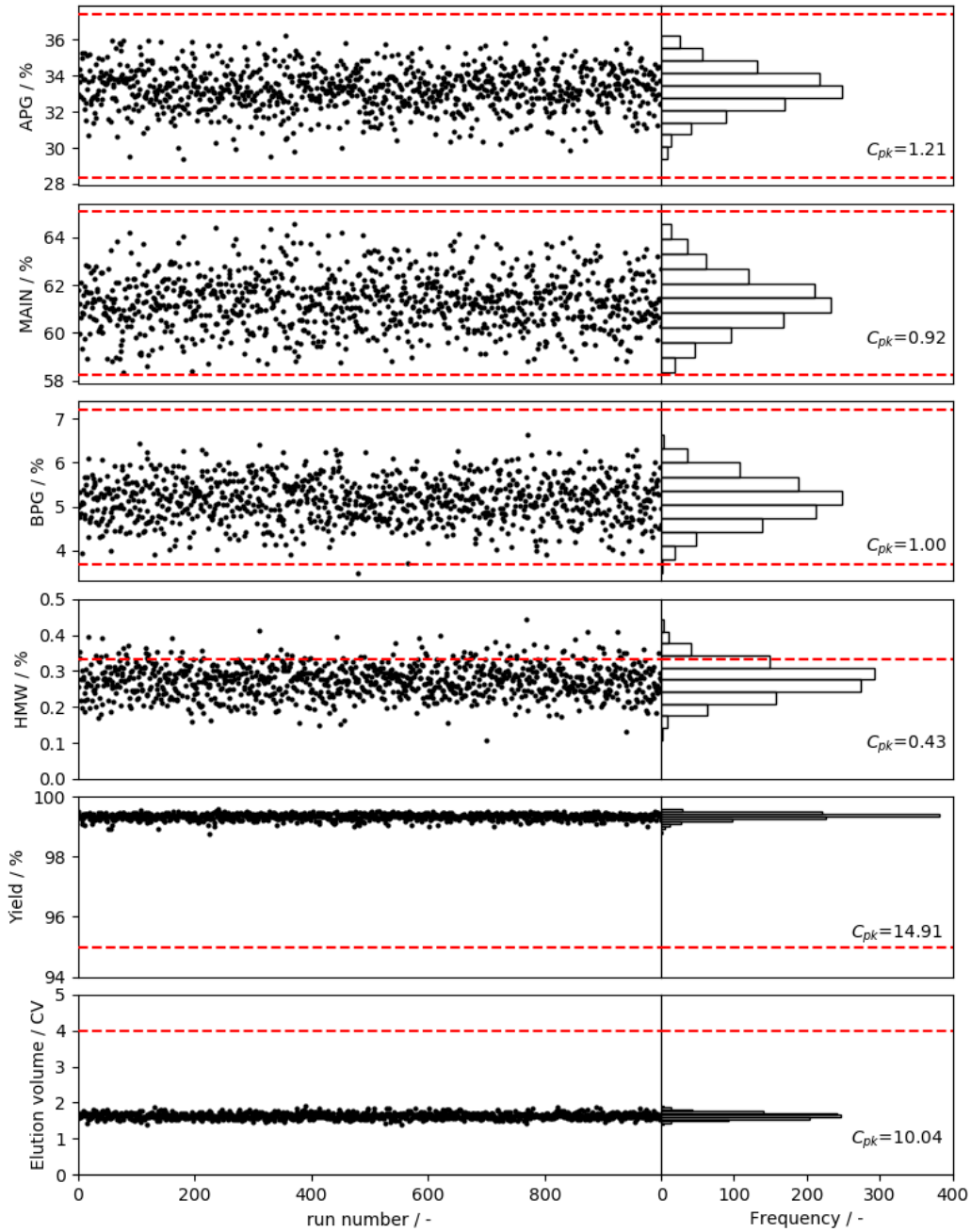

Figure 6.6: Monte-Carlo simulation of the CEX unit operation at pH 5.8 during elution phase. Dashed red lines indicate acceptance criteria. Each data point represents a simulation at 12000 L manufacturing-scale. Measurement data of 20 clinical manufacturing runs was used to simulate the variance of load material composition, loading density, pH, and salt concentrations of the different chromatographic phases

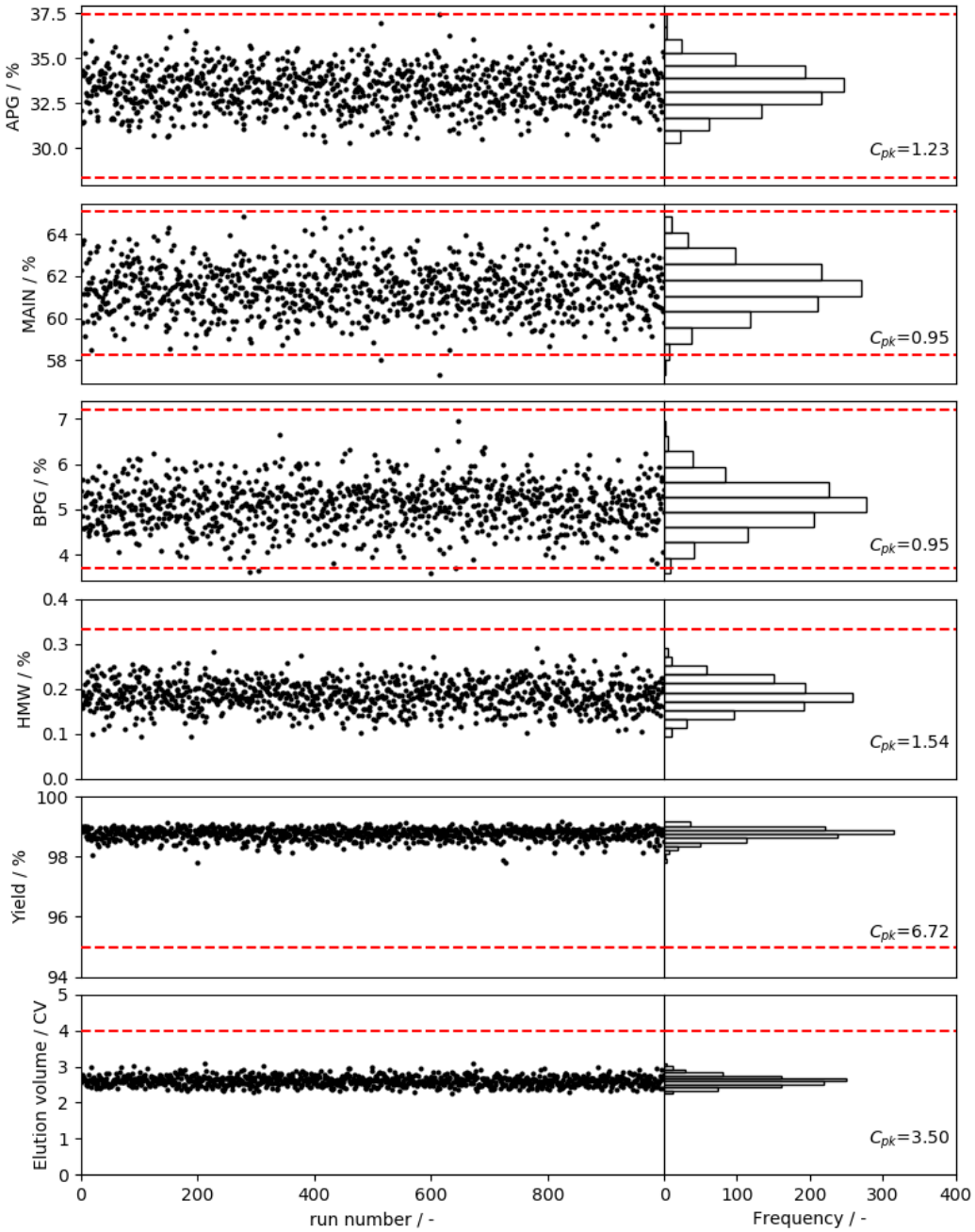

Figure 6.7: Monte-Carlo simulation of the CEX unit operation at pH 5.7 during elution phase. Dashed red lines indicate acceptance criteria. Each data point represents one simulation at 12000 L manufacturing-scale. Measurement data of 20 clinical manufacturing runs was used to simulate the variance of load material composition, loading density, pH, and salt concentrations of the different chromatographic phases

6.5 Conclusion

In the present study, a mechanistic chromatography model was applied to the process characterization of mAb polishing step. The *in silico* methodology fulfilled the essential elements of the QbD concept. OFAT sampling allowed classification of process parameters and establishment of PARs. Wet-lab studies derived from *in silico* screening can lead to a significantly reduced experimental effort compared to purely DoE driven PCS studies. Calculation of process capability considering *a posteriori* variabilities of feed stream materials and process parameters at manufacturing-scale enabled the identification of a robust set point condition. In this study, *in silico* PCS results were complemented with experimental data, which reduced the overall impact of mechanistic modeling on the control strategy. When relying exclusively on *in silico* predictions, considering the effects of parameter uncertainty on model predictions will further increase the trust in the final control strategy. Considering the complexity of polishing chromatography steps compared to other unit operations in mAb purification processes and the related experimental efforts that must be invested for their characterization, the here presented *in silico* techniques have the potential to debottleneck process development timelines. While accelerating development and disrupting experimental constraints, mechanistic modeling generated a deep process understanding ensuring consistent product quality in the light of QbD. This work represents a possible concept for the application of digital process twins to QbD related tasks in biopharmaceutical process development, with the focus on preparative chromatography. The proposed methods could further enable the *in silico* process characterization of other unit operations when validated mechanistic models are available.

Acknowledgement

The authors would like to acknowledge the experimental support and thorough review of their colleagues, including Roland Pfänder, Stephanie Combe, Jessica Stolzenberger, Heike Krischollek, and Jens Smiatek.

7 Conclusion & Outlook

The present thesis removed several roadblocks on the path from the primary structure of a therapeutic antibody to a robust downstream process. Protein surface analysis and QSPR modeling were identified as powerful tools to support candidate selection and process optimization with a minimal amount of experimental information. The gained knowledge on relationships between protein structure and process behavior could promote a paradigm shift in DSP development from a strictly generic platform process to a more flexible process design driven by the structural characteristics of antibody candidates. To close the gap between early- and late-stage DSP development, initial chromatography models built on protein structure information can be further refined using experimental data that is collected during the product life cycle. Here, the standardized and straightforward methods for model calibration, validation, and application could increase the acceptance of mechanistic modeling by industry and regulatory agencies.

The first part of this thesis demonstrated that sequence optimization of mAb candidates can influence their downstream processing. Single amino acid substitutions in the CDR had a significant impact on retention volumes and charge heterogeneity during preparative CEX chromatography. The identified relationship between mAb structure and CMC properties may support selection of mAb candidates that integrate into common downstream platforms. Further, initial experiments during early-phase DSP development could be planed based on a structure-based prediction of the binding strength of a mAb in CEX chromatography. Effects of amino acid substitutions on semi-mechanistic adsorption model parameters underlined the potential of building QSPR models that support the calibration of mechanistic chromatography models.

Based on this knowledge and data collected for a diverse set of over twenty biologics, QSPR models were built for the prediction of SDM parameters using structural descriptors derived from homology models. Consideration of two conformational states of IgG4 mAbs allowed for the prediction of split-peak elution profiles CEX chromatography. Coupling of homology modeling, machine learning, and mechanistic modeling led from the amino acid sequence to an initial process model that could be used for *in silico* process optimization. The addition of parameters defining the non-linear region of the SMA isotherm could further increase the impact of the applied multiscale modeling workflow for early-stage process development

To support technology transfer from academia to industry, this thesis introduced a straightforward and rapid method for the pH-dependent, multicomponent SMA chromatography models. The selection of representative process conditions allowed model calibration using

only two experiments with offline fraction analysis. Fundamental knowledge on preparative chromatography systematically reduced the number of unknown model parameters and avoided local minima during parameter estimation. The comparably small confidence intervals did not only indicate that the chosen experiments were well suited for solving the estimation problem, but they also fulfill the requirements of advanced applications in DSP development. Thus, the final chapters explored possible application scenarios in late-stage DSP that could support the regulatory approval of a commercial manufacturing process.

In silico experimentation at manufacturing-scale has been identified as a suitable alternative to experimental scale-down models traditionally used for development of chromatography processes. The mechanistic model could describe the run-to-run trend of CQAs and KPIs, due to the consideration of scale dependent effects, as well as minimal variances of input parameters, such as loading density, material composition, and mobile phase pH and salt concentration. Thus, the model was not only valid for process conditions beyond the intended operating ranges, but it could also capture the impact of variances anticipated during the routine production of the biologic. Simulating larger scales further improved the understanding of system- and-column specific properties that potentially affect elution profiles and product quality.

The workflow for *in silico* process characterization presented in the final chapter of this work demonstrated how mechanistic chromatography modeling can reduce experimental efforts during late-stage DSP development. Following the QbD principles, OFAT and multivariate sampling studies with thousands of simulations were able to reveal the nonlinear correlations between process parameters and CQAs. Calculation of process capabilities based on Monte-Carlo simulations was identified as a useful method for the *a priori* estimation of process robustness at manufacturing-scale. While accelerating development, mechanistic modeling increased process understanding ensuring consistent product quality and economic production.

The theoretical understanding and computational tools provided by this thesis are already in use to streamline the development of safe and effective biological drugs. Correlating protein structural data with mechanistic model parameters and purification outcomes may be the most promising, if not the only feasible, approach to build models that efficiently use existing information to give an *a priori* prediction on a mAbs' process behavior. The presented methods focused on CEX chromatography, due to the high experimental efforts that need to be invested for process optimization and characterization of this unit operation. In addition, CEX chromatography is a critical factor for product quality and is mechanistically well understood. Future work should extend the multiscale modeling techniques to other purification steps in the DSP platform, including AC, HIC, and MMC. If the fundamental relationships between antibody structure and process behavior are revealed across the entire DSP platform, the vision of a holistic digital transformation of DSP development could become a reality.

Bibliography

1. Morrison, C. Fresh from the biotech pipeline-2019. *Nat. Biotechnol* **38,** 126–131 (2020).
2. Grilo, A. L. & Mantalaris, A. The increasingly human and profitable monoclonal antibody market. *Trends in biotechnology* **37,** 9–16 (2019).
3. Kaplon, H., Muralidharan, M., Schneider, Z. & Reichert, J. M. Antibodies to watch in 2020. *mAbs* **12,** 1703531. ISSN: 1942-0862 (2020).
4. Kaplon, H. & Reichert, J. M. Antibodies to watch in 2019. *mAbs* **11,** 219–238. ISSN: 1942-0862 (2019).
5. Farid, S. S., Baron, M., Stamatis, C., Nie, W. & Coffman, J. *Benchmarking biopharmaceutical process development and manufacturing cost contributions to R&D* in *Mabs* **12** (2020), 1754999.
6. Liu, H., Ma, J., Winter, C. & Bayer, R. Recovery and purification process development for monoclonal antibody production. *mAbs* **2,** 480–499. ISSN: 1942-0862 (2010).
7. Shukla, A. A., Hubbard, B., Tressel, T., Guhan, S. & Low, D. Downstream processing of monoclonal antibodies-application of platform approaches. *J Chromatogr B Analyt Technol Biomed Life Sci* **848,** 28–39. ISSN: 1570-0232 (Print) 1570-0232 (Linking) (2007).
8. Shukla, A. A. & Thömmes, J. Recent advances in large-scale production of monoclonal antibodies and related proteins. *Trends in biotechnology* **28,** 253–261 (2010).
9. DiMasi, J. A. & Grabowski, H. G. The cost of biopharmaceutical R&D: is biotech different? *Managerial and decision Economics* **28,** 469–479 (2007).
10. DiMasi, J. A., Grabowski, H. G. & Hansen, R. W. Innovation in the pharmaceutical industry: new estimates of R&D costs. *Journal of health economics* **47,** 20–33 (2016).
11. Kelley, B. D., Switzer, M., Bastek, P., Kramarczyk, J. F., Molnar, K., Yu, T. & Coffman, J. High-throughput screening of chromatographic separations: IV. Ion-exchange. *Biotechnology and Bioengineering* **100,** 950–963. ISSN: 1097-0290 (2008).
12. Lacki, K. M. High-throughput process development of chromatography steps: advantages and limitations of different formats used. *Biotechnol J* **7,** 1192–202. ISSN: 1860-7314 (Electronic) 1860-6768 (Linking) (2012).
13. Hakemeyer, C., McKnight, N., John, R. S., Meier, S., Trexler-Schmidt, M., Kelley, B., Zettl, F., Puskeiler, R., Kleinjans, A., Lim, F., *et al.* Process characterization and design space definition. *Biologicals* **44,** 306–318 (2016).

14. Kepert, J. F., Cromwell, M., Engler, N., Finkler, C., Gellermann, G., Gennaro, L., Harris, R., Iverson, R., Kelley, B. & Krummen, L. Establishing a control system using QbD principles. *Biologicals* **44,** 319–331. ISSN: 1045-1056 (2016).

15. Bhambure, R., Kumar, K. & Rathore, A. S. High-throughput process development for biopharmaceutical drug substances. *Trends in biotechnology* **29,** 127–135 (2011).

16. Baumann, P. & Hubbuch, J. Downstream process development strategies for effective bioprocesses: trends, progress, and combinatorial approaches. *Engineering in life sciences* **17,** 1142–1158 (2017).

17. Pollock, J., Coffman, J., Ho, S. V. & Farid, S. S. Integrated continuous bioprocessing: Economic, operational, and environmental feasibility for clinical and commercial antibody manufacture. *Biotechnology progress* **33,** 854–866 (2017).

18. Mahajan, E., George, A. & Wolk, B. Improving affinity chromatography resin efficiency using semi-continuous chromatography. *Journal of chromatography A* **1227,** 154–162 (2012).

19. Warikoo, V., Godawat, R., Brower, K., Jain, S., Cummings, D., Simons, E., Johnson, T., Walther, J., Yu, M., Wright, B., *et al.* Integrated continuous production of recombinant therapeutic proteins. *Biotechnology and bioengineering* **109,** 3018–3029 (2012).

20. Nadar, S., Shooter, G., Somasundaram, B., Shave, E., Baker, K. & Lua, L. H. Intensified downstream processing of monoclonal antibodies using membrane technology. *Biotechnology Journal* **16,** 2000309 (2021).

21. Gottschalk, U. in *Disposable Bioreactors* 171–183 (Springer, 2009).

22. Roush, D., Asthagiri, D., Babi, D. K., Benner, S., Bilodeau, C., Carta, G., Ernst, P., Fedesco, M., Fitzgibbon, S., Flamm, M., *et al.* Toward in silico CMC: An industrial collaborative approach to model-based process development. *Biotechnology and Bioengineering* **117,** 3986–4000 (2020).

23. Smiatek, J., Jung, A. & Bluhmki, E. Towards a digital bioprocess replica: computational approaches in biopharmaceutical development and manufacturing. *Trends in Biotechnology* **38,** 1141–1153 (2020).

24. On Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use, I. C. ICH-Endorsed Guide for ICHQ8/Q9/Q10 Implementation.

25. Lawrence, X. Y., Amidon, G., Khan, M. A., Hoag, S. W., Polli, J., Raju, G. & Woodcock, J. Understanding pharmaceutical quality by design. *The AAPS journal* **16,** 771–783 (2014).

26. Hahn, T., Sommer, A., Osberghaus, A., Heuveline, V. & Hubbuch, J. Adjoint-based estimation and optimization for column liquid chromatography models. *Computers & Chemical Engineering* **64,** 41–54. ISSN: 0098-1354 (2014).

27. Mollerup, J. M., Hansen, T. B., Kidal, S., Sejergaard, L. & Staby, A. Development, modelling, optimisation and scale-up of chromatographic purification of a therapeutic protein. *Fluid Phase Equilibria* **261,** 133–139. ISSN: 0378-3812 (2007).

28. Cardillo, A. G., Castellanos, M. M., Desailly, B., Dessoy, S., Mariti, M., Portela, R. M., Scutella, B., von Stosch, M., Tomba, E. & Varsakelis, C. Towards in silico process modeling for vaccines. *Trends in biotechnology* (2021).

29. Von Stosch, M., Portela, R. M. & Varsakelis, C. A roadmap to AI-driven in silico process development: bioprocessing 4.0 in practice. *Current Opinion in Chemical Engineering* **33,** 100692 (2021).

30. Robinson, J. R., Karkov, H. S., Woo, J. A., Krogh, B. O. & Cramer, S. M. QSAR models for prediction of chromatographic behavior of homologous Fab variants. *Biotechnology and bioengineering* **114,** 1231–1240. ISSN: 0006-3592 (2017).

31. Ladiwala, A., Rege, K., Breneman, C. M. & Cramer, S. M. A priori prediction of adsorption isotherm parameters and chromatographic behavior in ion-exchange systems. *Proceedings of the National Academy of Sciences* **102,** 11710–11715. ISSN: 0027-8424 (2005).

32. Kuo, T. T., Baker, K., Yoshida, M., Qiao, S.-W., Aveson, V. G., Lencer, W. I. & Blumberg, R. S. Neonatal Fc receptor: from immunity to therapeutics. *Journal of clinical immunology* **30,** 777–789. ISSN: 0271-9142 (2010).

33. Roopenian, D. C. & Akilesh, S. FcRn: the neonatal Fc receptor comes of age. *Nature reviews. Immunology* **7,** 715. ISSN: 1474-1733 (2007).

34. Davis, J. D., Deng, R., Boswell, C. A., Zhang, Y., Li, J., Fielder, P., Joshi, A. & Kenkare-Mitra, S. in *Pharmaceutical Biotechnology* 143–178 (Springer, 2013).

35. Wang, W., Singh, S., Zeng, D. L., King, K. & Nema, S. Antibody structure, instability, and formulation. *Journal of pharmaceutical sciences* **96,** 1–26 (2007).

36. Crommelin, D. J., Sindelar, R. D. & Meibohm, B. *Pharmaceutical biotechnology: fundamentals and applications* (Springer Science & Business Media, 2013).

37. Schiele, F., van Ryn, J., Litzenburger, T., Ritter, M., Seeliger, D. & Nar, H. *Structure-guided residence time optimization of a dabigatran reversal agent* in *MAbs* **7** (2015), 871–880.

38. Harris, R. J., Shire, S. J. & Winter, C. Commercial manufacturing scale formulation and analytical characterization of therapeutic recombinant antibodies. *Drug development research* **61,** 137–154 (2004).

39. Vidarsson, G., Dekkers, G. & Rispens, T. IgG subclasses and allotypes: from structure to effector functions. *Frontiers in immunology* **5,** 520 (2014).

40. Daeron, M. Fc receptor biology. *Annual review of immunology* **15,** 203–234 (1997).

41. Labrijn, A. F., Buijsse, A. O., Van den Bremer, E. T., Verwilligen, A. Y., Bleeker, W. K., Thorpe, S. J., Killestein, J., Polman, C. H., Aalberse, R. C., Schuurman, J., *et al.* Therapeutic IgG4 antibodies engage in Fab-arm exchange with endogenous human IgG4 in vivo. *Nature biotechnology* **27,** 767–771 (2009).

42. Kretschmer, A., Schwanbeck, R., Valerius, T. & Rösner, T. Antibody isotypes for tumor immunotherapy. *Transfusion Medicine and Hemotherapy* **44,** 320–326 (2017).

43. Labrijn, A. F., Aalberse, R. C. & Schuurman, J. When binding is enough: nonactivating antibody formats. *Current opinion in immunology* **20,** 479–485 (2008).

44. Rizvi, N. A., Mazières, J., Planchard, D., Stinchcombe, T. E., Dy, G. K., Antonia, S. J., Horn, L., Lena, H., Minenza, E., Mennecier, B., *et al.* Activity and safety of nivolumab, an anti-PD-1 immune checkpoint inhibitor, for patients with advanced, refractory squamous non-small-cell lung cancer (CheckMate 063): a phase 2, single-arm trial. *The lancet oncology* **16,** 257–265 (2015).

45. Weidle, U. H., Tiefenthaler, G., Weiss, E. H., Georges, G. & Brinkmann, U. The intriguing options of multispecific antibody formats for treatment of cancer. *Cancer genomics & proteomics* **10,** 1–18 (2013).

46. Arlotta, K. J. & Owen, S. C. Antibody and antibody derivatives as cancer therapeutics. *Wiley Interdisciplinary Reviews: Nanomedicine and Nanobiotechnology* **11,** e1556 (2019).

47. Seckinger, A., Delgado, J. A., Moser, S., Moreno, L., Neuber, B., Grab, A., Lipp, S., Merino, J., Prosper, F., Emde, M., *et al.* Target expression, generation, preclinical activity, and pharmacokinetics of the BCMA-T cell bispecific antibody EM801 for multiple myeloma treatment. *Cancer cell* **31,** 396–410 (2017).

48. Baeuerle, P. A. & Reinhardt, C. Bispecific T-cell engaging antibodies for cancer therapy. *Cancer research* **69,** 4941–4944 (2009).

49. Ridgway, J. B., Presta, L. G. & Carter, P. 'Knobs-into-holes' engineering of antibody CH3 domains for heavy chain heterodimerization. *Protein Engineering, Design and Selection* **9,** 617–621 (1996).

50. Spiess, C., Zhai, Q. & Carter, P. J. Alternative molecular formats and therapeutic applications for bispecific antibodies. *Molecular immunology* **67,** 95–106 (2015).

51. Kuglstatter, A, Stihle, M, Neumann, C, Müller, C, Schaefer, W, Klein, C, Benz, J, Research, R. P. & Development, E. Structural differences between glycosylated, disulfide-linked heterodimeric Knob-into-Hole Fc fragment and its homodimeric Knob–Knob and Hole–Hole side products. *Protein Engineering, Design and Selection* **30,** 649–656 (2017).

52. Williams, A. J., Giese, G. & Persson, J. Improved assembly of bispecific antibodies from knob and hole half-antibodies. *Biotechnology progress* **31,** 1315–1322 (2015).

53. Dill, K. A. & MacCallum, J. L. The protein-folding problem, 50 years on. *science* **338,** 1042–1046 (2012).

54. Senior, A. W., Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T., Qin, C., Žídek, A., Nelson, A. W., Bridgland, A., *et al.* Improved protein structure prediction using potentials from deep learning. *Nature* **577,** 706–710 (2020).

55. McGuffin, L. J., Bryson, K. & Jones, D. T. The PSIPRED protein structure prediction server. *Bioinformatics* **16,** 404–405 (2000).

56. Schwede, T., Kopp, J., Guex, N. & Peitsch, M. C. SWISS-MODEL: an automated protein homology-modeling server. *Nucleic acids research* **31,** 3381–3385 (2003).

57. Xu, J., Mcpartlon, M. & Li, J. Improved protein structure prediction by deep learning irrespective of co-evolution information. *Nature Machine Intelligence,* 1–9 (2021).

58. AlQuraishi, M. AlphaFold at CASP13. *Bioinformatics* **35,** 4862–4865 (2019).

59. Kinch, L. N., Pei, J., Kryshtafovych, A., Schaeffer, R. D. & Grishin, N. V. Topology Evaluation of Models for Difficult Targets in the 14th Round of the Critical Assessment of Protein Structure Prediction (CASP14). *Proteins: Structure, Function, and Bioinformatics* (2021).

60. Krieger, E., Nabuurs, S. B. & Vriend, G. Homology modeling. *Methods of biochemical analysis* **44,** 509–524 (2003).

61. Epstain, C., Goldberger, R. & Anfinsen, C. *The genetic control of tertiary protein structure. Model systems* in *Cold Spring Harbor Symposia on Quantitative Biology* **28** (1963), 439–449.

62. Chothia, C. & Lesk, A. M. The relation between the divergence of sequence and structure in proteins. *The EMBO journal* **5,** 823–826 (1986).

63. Sander, C. & Schneider, R. Database of homology-derived protein structures and the structural meaning of sequence alignment. *Proteins: Structure, Function, and Bioinformatics* **9,** 56–68 (1991).

64. Dunbar, J., Krawczyk, K., Leem, J., Baker, T., Fuchs, A., Georges, G., Shi, J. & Deane, C. M. SAbDab: the structural antibody database. *Nucleic acids research* **42,** D1140–D1146 (2014).

65. Sircar, A., Kim, E. T. & Gray, J. J. RosettaAntibody: antibody variable region homology modeling server. *Nucleic acids research* **37,** W474–W479 (2009).

66. Zhu, K. & Day, T. Ab initio structure prediction of the antibody hypervariable H3 loop. *Proteins: Structure, Function, and Bioinformatics* **81,** 1081–1089 (2013).

67. Norman, R. A., Ambrosetti, F., Bonvin, A. M., Colwell, L. J., Kelm, S., Kumar, S. & Krawczyk, K. Computational approaches to therapeutic antibody design: established methods and emerging trends. *Briefings in bioinformatics* **21,** 1549–1567 (2020).

68. Van der Kant, R., Bauer, J., Karow-Zwick, A. R., Kube, S., Garidel, P., Blech, M., Rousseau, F. & Schymkowitz, J. Adaption of human antibody λ and κ light chain architectures to CDR repertoires. *Protein Engineering, Design and Selection* **32,** 109–127 (2019).

69. Zhu, K., Day, T., Warshaviak, D., Murrett, C., Friesner, R. & Pearlman, D. Antibody structure determination using a combination of homology modeling, energy-based refinement, and loop prediction. *Proteins: Structure, Function, and Bioinformatics* **82,** 1646–1655 (2014).

70. Lippow, S. M., Wittrup, K. D. & Tidor, B. Computational design of antibody-affinity improvement beyond in vivo maturation. *Nature biotechnology* **25,** 1171–1176 (2007).

71. Mariuzza, R., Phillips, S. & Poljak, R. The structural basis of antigen-antibody recognition. *Annual review of biophysics and biophysical chemistry* **16,** 139–159 (1987).

72. Sivasubramanian, A., Chao, G., Pressler, H. M., Wittrup, K. D. & Gray, J. J. Structural model of the mAb 806-EGFR complex using computational docking followed by computational and experimental mutagenesis. *Structure* **14,** 401–414 (2006).

73. Bauer, J., Mathias, S., Kube, S., Otte, K., Garidel, P., Gamer, M., Blech, M., Fischer, S. & Karow-Zwick, A. R. Rational optimization of a monoclonal antibody improves the aggregation propensity and enhances the CMC properties along the entire pharmaceutical process chain. *mAbs* **12,** 1787121. ISSN: 1942-0862 (2020).

74. Ishihara, T., Kadoya, T., Yoshida, H., Tamada, T. & Yamamoto, S. Rational methods for predicting human monoclonal antibodies retention in protein A affinity chromatography and cation exchange chromatography. *Journal of Chromatography A* **1093,** 126–138. ISSN: 0021-9673 (2005).

75. Rosenberg, A. S. Effects of protein aggregates: an immunologic perspective. *The AAPS journal* **8,** E501–E507 (2006).

76. Schmidt-Traub, H. *Preparative Chromatography* ISBN: 9783527605903 (Wiley, Hoboken, GERMANY, 2006).

77. Guiochon, G., Felinger, A. & Shirazi, D. G. *Fundamentals of preparative and nonlinear chromatography* ISBN: 0080457223 (Elsevier, 2006).

78. Zhao, G., Dong, X.-Y. & Sun, Y. Ligands for mixed-mode protein chromatography: principles, characteristics and design. *Journal of biotechnology* **144,** 3–11 (2009).

79. Cuatrecasas, P. & Anfinsen, C. B. in *Methods in enzymology* 345–378 (Elsevier, 1971).

80. Hober, S., Nord, K. & Linhult, M. Protein A chromatography for antibody purification. *Journal of Chromatography B* **848,** 40–47 (2007).

81. Hjelm, H., Hjelm, K. & Sjöquist, J. Protein a from Staphylococcus aureus. Its isolation by affinity chromatography and its use as an immunosorbent for isolation of immunoglobulins. *FEBS Letters* **28,** 73–76. ISSN: 0014-5793 (1972).

82. Shukla, A. A. & Hinckley, P. Host cell protein clearance during protein A chromatography: development of an improved column wash step. *Biotechnology progress* **24,** 1115–1121 (2008).

83. Li, F., Hashimura, Y., Pendleton, R., Harms, J., Collins, E. & Lee, B. A systematic approach for scale-down model development and characterization of commercial cell culture processes. *Biotechnology progress* **22,** 696–703. ISSN: 8756-7938 (2006).

84. Shukla, A. A., Gupta, P. & Han, X. Protein aggregation kinetics during Protein A chromatography: case study for an Fc fusion protein. *Journal of Chromatography A* **1171,** 22–28 (2007).

85. Chen, J., Tetrault, J. & Ley, A. Comparison of standard and new generation hydrophobic interaction chromatography resins in the monoclonal antibody purification process. *Journal of Chromatography A* **1177,** 272–281 (2008).

86. Tanford, C. *The hydrophobic effect: formation of micelles and biological membranes 2d ed* (J. Wiley., 1980).

87. Wang, G., Hahn, T. & Hubbuch, J. Water on hydrophobic surfaces: mechanistic modeling of hydrophobic interaction chromatography. *Journal of Chromatography A* **1465,** 71–78 (2016).

88. Jungbauer, A. Chromatographic media for bioseparation. *Journal of Chromatography A* **1065,** 3–12 (2005).

89. Kelley, B. Very Large Scale Monoclonal Antibody Purification: The Case for Conventional Unit Operations. *Biotechnology Progress* **23,** 995–1008. ISSN: 1520-6033 (2007).

90. Kelley, B. D., Tobler, S. A., Brown, P., Coffman, J. L., Godavarti, R., Iskra, T., Switzer, M. & Vunnum, S. Weak partitioning chromatography for anion exchange purification of monoclonal antibodies. *Biotechnol Bioeng* **101,** 553–66. ISSN: 1097-0290 (Electronic) 0006-3592 (Linking) (2008).

91. Shamashkin, M., Godavarti, R., Iskra, T. & Coffman, J. A tandem laboratory scale protein purification process using Protein A affinity and anion exchange chromatography operated in a weak partitioning mode. *Biotechnology and bioengineering* **110,** 2655–2663 (2013).

92. Zhang, S., Iskra, T., Daniels, W., Salm, J., Gallo, C., Godavarti, R. & Carta, G. Structural and performance characteristics of representative anion exchange resins used for weak partitioning chromatography. *Biotechnology progress* **33,** 425–434 (2017).

93. Yigzaw, Y, Hinckley, P, Hewig, A & Vedantham, G. Ion exchange chromatography of proteins and clearance of aggregates. *Current pharmaceutical biotechnology* **10,** 421–426 (2009).

94. Scientific, T. F. *PRODUCT INFORMATION SHEET: POROS^{TM} Strong Cation Exchange Resins: XS and 50 HS* Report (2018).

95. Liu, H. F., McCooey, B., Duarte, T., Myers, D. E., Hudson, T., Amanullah, A., van Reis, R. & Kelley, B. D. Exploration of overloaded cation exchange chromatography for monoclonal antibody purification. *Journal of Chromatography A* **1218,** 6943–6952. ISSN: 0021-9673 (2011).

96. McDonald, P., Tran, B., Williams, C. R., Wong, M., Zhao, T., Kelley, B. D. & Lester, P. The rapid identification of elution conditions for therapeutic antibodies from cation-exchange chromatography resins using high-throughput screening. *Journal of Chromatography A* **1433,** 66–74 (2016).

97. Benner, S. W., Welsh, J. P., Rauscher, M. A. & Pollard, J. M. Prediction of lab and manufacturing scale chromatography performance using mini-columns and mechanistic modeling. *Journal of Chromatography A* **1593,** 54–62. ISSN: 0021-9673 (2019).

98. Pirrung, S. M., van der Wielen, L. A., Van Beckhoven, R. F., van de Sandt, E. J., Eppink, M. H. & Ottens, M. Optimization of biopharmaceutical downstream processes supported by mechanistic models and artificial neural networks. *Biotechnology progress.* ISSN: 1520-6033 (2017).

99. Hahn, T., Huuk, T., Heuveline, V. & Hubbuch, J. Simulating and Optimizing Preparative Protein Chromatography with ChromX. *Journal of Chemical Education* **92,** 1497–1502. ISSN: 0021-9584 (2015).

100. Wang, G., Briskot, T., Hahn, T., Baumann, P. & Hubbuch, J. Root cause investigation of deviations in protein chromatography based on mechanistic models and artificial neural networks. *Journal of Chromatography A* **1515,** 146–153 (2017).

101. Close, E. J., Salm, J. R., Bracewell, D. G. & Sorensen, E. A model based approach for identifying robust operating conditions for industrial chromatography with process variability. *Chemical Engineering Science* **116,** 284–295. ISSN: 0009-2509 (2014).

102. Osberghaus, A., Hepbildikler, S., Nath, S., Haindl, M., Von Lieres, E. & Hubbuch, J. Optimizing a chromatographic three component separation: A comparison of mechanistic and empiric modeling approaches. *Journal of chromatography A* **1237,** 86–95 (2012).

103. Rathore, A. S. & Winkle, H. Quality by design for biopharmaceuticals. *Nature biotechnology* **27,** 26–34 (2009).

104. Wicke, E. Empirische und theoretische Untersuchungen der Sorptionsgeschwindigkeit von Gasen an porösen Stoffen II. *Kolloid-Zeitschrift* **86,** 295–313 (1939).

105. DeVault, D. The theory of chromatography. *Journal of the American Chemical Society* **65,** 532–540 (1943).

106. Püttmann, A., Nicolai, M., Behr, M. & von Lieres, E. Stabilized space–time finite elements for high-definition simulation of packed bed chromatography. *Finite elements in analysis and design* **86,** 1–11 (2014).

107. Wilson, J. N. A theory of chromatography. *Journal of the American Chemical Society* **62,** 1583–1591 (1940).

108. Tsotsas, E. *Über die Wärme-und Stoffübertragung in durchströmten Festbetten VDI Fortschrittsberichte* 1987.

109. Zhong, G. & Guiochon, G. Analytical solution for the linear ideal model of simulated moving bed chromatography. *Chemical Engineering Science* **51,** 4307–4319 (1996).

110. Katti, A. & Guiochon, G. Prediction of band profiles in displacement chromatography by numerical integration of a semi-ideal model. *Journal of Chromatography A* **449,** 25–40 (1988).

111. Kaczmarski, K., Antos, D., Sajonz, H., Sajonz, P. & Guiochon, G. Comparative modeling of breakthrough curves of bovine serum albumin in anion-exchange chromatography. *Journal of Chromatography A* **925,** 1–17 (2001).

112. Golshan-Shirazi, S. & Guiochon, G. Comparison of the various kinetic models of non-linear chromatography. *Journal of Chromatography A* **603,** 1–11 (1992).

113. Golshan-Shirazi, S & Guiochon, G. in *Theoretical Advancement in Chromatography and Related Separation Techniques* 35–59 (Springer, 1992).

114. Hahn, T., Huuk, T., Osberghaus, A., Doninger, K., Nath, S., Hepbildikler, S., Heuveline, V. & Hubbuch, J. Calibration-free inverse modeling of ion-exchange chromatography in industrial antibody purification. *Engineering in Life Sciences* **16,** 107–113. ISSN: 1618-2863 (2016).

115. Hunt, S., Larsen, T. & Todd, R. J. in *Preparative Chromatography for Separation of Proteins* (eds Staby, A. & Ahuja, S.) (John Wiley & Sons, Incorporated, Somerset, UNITED STATES, 2017). ISBN: 9781119031154.

116. Borg, N., Brodsky, Y., Moscariello, J., Vunnum, S., Vedantham, G., Westerberg, K. & Nilsson, B. Modeling and robust pooling design of a preparative cation-exchange chromatography step for purification of monoclonal antibody monomer from aggregates. *Journal of Chromatography A* **1359,** 170–181. ISSN: 0021-9673 (2014).

117. Brooks, C. A. & Cramer, S. M. Steric mass-action ion exchange: Displacement profiles and induced salt gradients. *AIChE Journal* **38,** 1969–1978. ISSN: 1547-5905 (1992).

118. Mollerup, J. M. The thermodynamic principles of ligand binding in chromatography and biology. *Journal of Biotechnology* **132,** 187–195. ISSN: 0168-1656 (2007).

119. Mollerup, J. M. Applied thermodynamics: A new frontier for biotechnology. *Fluid phase equilibria* **241,** 205–215 (2006).

120. Mollerup, J. M., Hansen, T. B., Frederiksen, S. S. & Staby, A. Thermodynamic modeling of chromatographic separation. *Adv. Chromatogr.* **48,** 57–97 (2010).

121. Pedersen, L., Mollerup, J., Hansen, E. & Jungbauer, A. Whey proteins as a model system for chromatographic separation of proteins. *Journal of chromatography B* **790,** 161–173 (2003).

122. Bett, K. E., Rowlinson, J. S. & Saville, G. *Thermodynamics for chemical engineers* (Mit Press, 1975).

123. Michelsen, M. L. & Mollerup, J. *Thermodynamic modelling: fundamentals and computational aspects* (Tie-Line Publications, 2004).

124. Boardman, N. & Partridge, S. Separation of neutral proteins on ion-exchange resins. *Biochemical Journal* **59,** 543–552. ISSN: 0006-2936 (1955).

125. Velayudhan, A. & Horváth, C. Preparative chromatography of proteins: analysis of the multivalent ion-exchange formalism. *Journal of Chromatography A* **443,** 13–29. ISSN: 0021-9673 (1988).

126. Huuk, T. C., Briskot, T., Hahn, T. & Hubbuch, J. A versatile noninvasive method for adsorber quantification in batch and column chromatography based on the ionic capacity. *Biotechnology Progress* **32,** 666–677. ISSN: 1520-6033 (2016).

127. Schmidt, M., Hafner, M. & Frech, C. Modeling of salt and pH gradient elution in ion-exchange chromatography. *Journal of separation science* **37,** 5–13 (2014).

128. Kluters, S., Wittkopp, F., Johnck, M. & Frech, C. Application of linear pH gradients for the modeling of ion exchange chromatography: Separation of monoclonal antibody monomer from aggregates. *J Sep Sci* **39,** 663–75. ISSN: 1615-9314 (Electronic) 1615-9306 (Linking) (2016).

129. Ribeiro, J. M. & Sillero, A. A program to calculate the isoelectric point of macromolecules. *Computers in biology and medicine* **21,** 131–141 (1991).

130. Davies, C. W. *Ion Association* (Butterworths, London, 1962).

131. Dismer, F. & Hubbuch, J. 3D structure-based protein retention prediction for ion-exchange chromatography. *Journal of Chromatography A* **1217,** 1343–1353. ISSN: 0021-9673 (2010).

132. Kittelmann, J., Lang, K. M., Ottens, M. & Hubbuch, J. An orientation sensitive approach in biomolecule interaction quantitative structure–activity relationship modeling and its application in ion-exchange chromatography. *Journal of Chromatography A* **1482,** 48–56. ISSN: 0021-9673 (2017).

133. Kittelmann, J., Lang, K. M., Ottens, M. & Hubbuch, J. Orientation of monoclonal antibodies in ion-exchange chromatography: A predictive quantitative structure–activity relationship modeling approach. *Journal of Chromatography A* **1510,** 33–39. ISSN: 0021-9673 (2017).

134. Mitchell, J. B. Machine learning methods in chemoinformatics. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **4,** 468–481 (2014).

135. Roy, K. Advances in QSAR modeling. *Applications in pharmaceutical, chemical, food, agricultural and environmental sciences. Springer, Cham* **555** (2017).

136. Emmert-Streib, F. *Statistical modelling of molecular descriptors in QSAR/QSPR* (John Wiley & Sons, 2012).

137. Rodríguez-Pérez, R. & Bajorath, J. Interpretation of compound activity predictions from complex machine learning models using local approximations and shapley values. *Journal of Medicinal Chemistry* **63,** 8761–8777 (2019).

138. Jetha, A., Thorsteinson, N., Jmeian, Y., Jeganathan, A., Giblin, P. & Fransson, J. Homology modeling and structure-based design improve hydrophobic interaction chromatography behavior of integrin binding antibodies. **10,** 890–900 (2018).

139. Bauer, K. C., Haemmerling, F., Kittelmann, J., Duerr, C., Goerlich, F. & Hubbuch, J. Influence of structure properties on protein–protein interactions—QSAR modeling of changes in diffusion coefficients. *Biotechnology and Bioengineering* **114,** 821–831 (2017).

140. Vormittag, P., Klamp, T. & Hubbuch, J. Ensembles of Hydrophobicity Scales as Potent Classifiers for Chimeric Virus-Like Particle Solubility–An Amino Acid Sequence-Based Machine Learning Approach. *Frontiers in Bioengineering and Biotechnology* **8,** 395 (2020).

141. Shukla, A. A., Etzel, M. R. & Gadam, S. *Process scale bioseparations for the biopharmaceutical industry* (CRC press, 2006).

142. Sankar, K., Krystek Jr, S. R., Carl, S. M., Day, T. & Maier, J. K. AggScore: Prediction of aggregation-prone regions in proteins based on the distribution of surface patches. *Proteins: Structure, Function, and Bioinformatics* **86,** 1147–1156 (2018).

143. Conchillo-Solé, O., de Groot, N. S., Avilés, F. X., Vendrell, J., Daura, X. & Ventura, S. AGGRESCAN: a server for the prediction and evaluation of" hot spots" of aggregation in polypeptides. *BMC bioinformatics* **8,** 1–17 (2007).

144. Lusci, A., Pollastri, G. & Baldi, P. Deep architectures and deep learning in chemoinformatics: the prediction of aqueous solubility for drug-like molecules. *Journal of chemical information and modeling* **53,** 1563–1575 (2013).

145. Kira, K. & Rendell, L. A. in *Machine learning proceedings 1992* 249–256 (Elsevier, 1992).

146. Kumar, V. & Minz, S. Feature selection: a literature review. *SmartCR* **4,** 211–229 (2014).

147. Guyon, I. & Elisseeff, A. An introduction to variable and feature selection. *Journal of machine learning research* **3,** 1157–1182 (2003).

148. Gramatica, P. Principles of QSAR models validation: internal and external. *QSAR & combinatorial science* **26,** 694–701 (2007).

149. Van der Kant, R., Karow-Zwick, A. R., Van Durme, J., Blech, M., Gallardo, R., Seeliger, D., Aßfalg, K., Baatsen, P., Compernolle, G. & Gils, A. Prediction and reduction of the aggregation of monoclonal antibodies. *Journal of Molecular Biology* **429,** 1244–1261. ISSN: 0022-2836 (2017).

150. Bailly, M., Mieczkowski, C., Juan, V., Metwally, E., Tomazela, D., Baker, J., Uchida, M., Kofman, E., Raoufi, F. & Motlagh, S. *Predicting Antibody Developability Profiles Through Early Stage Discovery Screening* in *Mabs* **12** (Taylor & Francis), 1743053.

151. Seeliger, D., Schulz, P., Litzenburger, T., Spitz, J., Hoerer, S., Blech, M., Enenkel, B., Studts, J. M., Garidel, P. & Karow, A. R. *Boosting antibody developability through rational sequence optimization* in *MAbs* **7** (2015), 505–515.

152. Glynn, J., Hagerty, T., Pabst, T., Annathur, G., Thomas, K., Johnson, P., Ramasubramanyan, N. & Mensah, P. The Development and Application of a Monoclonal Antibody Purification Platform. *Biopharm International,* 16–20 (2009).

153. Lang, K. M., Kittelmann, J., Dürr, C., Osberghaus, A. & Hubbuch, J. A comprehensive molecular dynamics approach to protein retention modeling in ion exchange chromatography. *Journal of Chromatography A* **1381,** 184–193. ISSN: 0021-9673 (2015).

154. Zhou, J., Tsao, H.-K., Sheng, Y.-J. & Jiang, S. Monte Carlo simulations of antibody adsorption and orientation on charged surfaces. *The Journal of chemical physics* **121,** 1050–1057. ISSN: 0021-9606 (2004).

155. Banerjee, S., Parimal, S. & Cramer, S. M. A molecular modeling based method to predict elution behavior and binding patches of proteins in multimodal chromatography. *Journal of Chromatography A* **1511,** 45–58. ISSN: 0021-9673 (2017).

156. Freed, A. S., Garde, S. & Cramer, S. M. Molecular simulations of multimodal ligand–protein binding: elucidation of binding sites and correlation with experiments. *The Journal of Physical Chemistry B* **115,** 13320–13327. ISSN: 1520-6106 (2011).

157. Woo, J., Parimal, S., Brown, M. R., Heden, R. & Cramer, S. M. The effect of geometrical presentation of multimodal cation-exchange ligands on selective recognition of hydrophobic regions on protein surfaces. *Journal of Chromatography a* **1412,** 33–42. ISSN: 0021-9673 (2015).

158. Robinson, J., Roush, D. & Cramer, S. Domain contributions to antibody retention in multimodal chromatography systems. *Journal of Chromatography A* **1563,** 89–98. ISSN: 0021-9673 (2018).

159. Robinson, J., Roush, D. & Cramer, S. M. The effect of pH on antibody retention in multimodal cation exchange chromatographic systems. *Journal of Chromatography A,* 460838. ISSN: 0021-9673 (2020).

160. Mazza, C., Sukumar, N, Breneman, C. & Cramer, S. Prediction of protein retention in ion-exchange systems using molecular descriptors obtained from crystal structure. *Analytical chemistry* **73,** 5457–5461. ISSN: 0003-2700 (2001).

161. Huuk, T. C., Hahn, T., Osberghaus, A. & Hubbuch, J. Model-based integrated optimization and evaluation of a multi-step ion exchange chromatography. *Separation and Purification Technology* **136,** 207–222. ISSN: 1383-5866 (2014).

162. Jakobsson, N., Degerman, M., Stenborg, E. & Nilsson, B. Model based robustness analysis of an ion-exchange chromatography step. *Journal of Chromatography A* **1138,** 109–119. ISSN: 0021-9673 (2007).

163. Rischawy, F., Saleh, D., Hahn, T., Oelmeier, S., Spitz, J. & Kluters, S. Good modeling practice for industrial chromatography: Mechanistic modeling of ion exchange chromatography of a bispecific antibody. *Computers & Chemical Engineering* **130,** 106532. ISSN: 0098-1354 (2019).

164. Mollerup, J. M., Hansen, T. B., Kidal, S. & Staby, A. Quality by design—thermodynamic modelling of chromatographic separation of proteins. *Journal of Chromatography A* **1177,** 200–206 (2008).

165. Briskot, T., Stückler, F., Wittkopp, F., Williams, C., Yang, J., Konrad, S., Doninger, K., Griesbach, J., Bennecke, M., Hepbildikler, S., *et al.* Prediction uncertainty assessment of chromatography models using Bayesian inference. *Journal of Chromatography A* **1587,** 101–110 (2019).

166. Rüdt, M., Gillet, F., Heege, S., Hitzler, J., Kalbfuss, B. & Guélat, B. Combined Yamamoto approach for simultaneous estimation of adsorption isotherm and kinetic parameters in ion-exchange chromatography. *Journal of Chromatography A* **1413,** 68–76 (2015).

167. Hahn, T., Baumann, P., Huuk, T., Heuveline, V. & Hubbuch, J. UV absorption-based inverse modeling of protein chromatography. *Engineering in Life Sciences* **16,** 99–106. ISSN: 1618-2863 (2016).

168. Roos, K., Wu, C., Damm, W., Reboul, M., Stevenson, J. M., Lu, C., Dahlgren, M. K., Mondal, S., Chen, W., Wang, L., *et al.* OPLS3e: Extending force field coverage for drug-like small molecules. *Journal of chemical theory and computation* **15,** 1863–1874 (2019).

169. Olsson, M. H., Søndergaard, C. R., Rostkowski, M. & Jensen, J. H. PROPKA3: consistent treatment of internal and surface residues in empirical p K a predictions. *Journal of chemical theory and computation* **7,** 525–537 (2011).

170. Baumann, P., Hahn, T. & Hubbuch, J. High-throughput micro-scale cultivations and chromatography modeling: Powerful tools for integrated process development. *Biotechnology and bioengineering* **112,** 2123–2133 (2015).

171. Osberghaus, A, Hepbildikler, S, Nath, S, Haindl, M, von Lieres, E & Hubbuch, J. Determination of parameters for the steric mass action model—A comparison between two approaches. *Journal of Chromatography A* **1233,** 54–65. ISSN: 0021-9673 (2012).

172. Yamamoto, S., Nakanishi, K. & Matsuno, R. *Ion-exchange chromatography of proteins* ISBN: 0824779037 (CRC Press, 1988).

173. Yamamoto, S., Nomura, M. & Sano, Y. Resolution of proteins in linear gradient elution ion-exchange and hydrophobic interaction chromatography. *Journal of Chromatography A* **409,** 101–110. ISSN: 0021-9673 (1987).

174. Rodrigues, A. E. Permeable packings and perfusion chromatography in protein separation. *Journal of Chromatography B: Biomedical Sciences and Applications* **699,** 47–61. ISSN: 0378-4347 (1997).

175. Guiochon, G., Felinger, A. & Shirazi, D. G. in *Fundamentals of preparative and nonlinear chromatography* 221–279 (Academic Press, San Diego, USA, 2006). ISBN: 0080457223.

176. Kumar, S., Roffi, K., Tomar, D. S., Cirelli, D., Luksha, N., Meyer, D., Mitchell, J., Allen, M. J. & Li, L. Rational optimization of a monoclonal antibody for simultaneous improvements in its solution properties and biological activity. *Protein Engineering, Design and Selection* **31,** 313–325 (2018).

177. Gudhka, R. B., Bilodeau, C. L., McCallum, S. A., McCoy, M. A., Roush, D. J., Snyder, M. A. & Cramer, S. M. Identification of preferred multimodal ligand-binding regions on IgG1 FC using nuclear magnetic resonance and molecular dynamics simulations. *Biotechnology and Bioengineering* **118,** 809–822 (2021).

178. Müller-Späth, T., Aumann, L., Melter, L., Ströhlein, G. & Morbidelli, M. Chromatographic separation of three monoclonal antibody variants using multicolumn countercurrent solvent gradient purification (MCSGP). *Biotechnology and bioengineering* **100,** 1166–1177 (2008).

179. Lan, W., Valente, J. J., Ilott, A., Chennamsetty, N., Liu, Z., Rizzo, J. M., Yamniuk, A. P., Qiu, D., Shackman, H. M. & Bolgar, M. S. Investigation of anomalous charge variant profile reveals discrete pH-dependent conformations and conformation-dependent charge states within the CDR3 loop of a therapeutic mAb. **12,** 1763138 (2020).

180. Liu, H., Gaza-Bulseco, G., Faldu, D., Chumsae, C. & Sun, J. Heterogeneity of Monoclonal Antibodies. *Journal of Pharmaceutical Sciences* **97,** 2426–2447. ISSN: 0022-3549 (2008).

181. Saleh, D., Wang, G., Müller, B., Rischawy, F., Kluters, S., Studts, J. & Hubbuch, J. Straightforward method for calibration of mechanistic cation exchange chromatography models for industrial applications. *Biotechnology progress* **36,** e2984 (2020).

182. Sankar, K., Hoi, K. H., Yin, Y., Ramachandran, P., Andersen, N., Hilderbrand, A., McDonald, P., Spiess, C. & Zhang, Q. *Prediction of methionine oxidation risk in monoclonal antibodies using a machine learning method* in *MAbs* **10** (2018), 1281–1290.

183. Wang, C., Li, W., Drabek, D., Okba, N. M., van Haperen, R., Osterhaus, A. D., van Kuppeveld, F. J., Haagmans, B. L., Grosveld, F. & Bosch, B.-J. A human monoclonal antibody blocking SARS-CoV-2 infection. *Nature communications* **11,** 1–6 (2020).

184. Kelley, B. Developing therapeutic monoclonal antibodies at pandemic pace. *Nature Biotechnology,* 1–6. ISSN: 1546-1696 (2020).

185. Saleh, D., Hess, R., Ahlers-Hesse, M., Beckert, N., Schönberger, M., Rischawy, F., Wang, G., Bauer, J., Blech, M., Kluters, S., *et al.* Modeling the impact of amino acid substitution in a monoclonal antibody on cation exchange chromatography. *Biotechnology and Bioengineering* (2021).

186. Luo, H., Macapagal, N., Newell, K., Man, A., Parupudi, A., Li, Y. & Li, Y. Effects of salt-induced reversible self-association on the elution behavior of a monoclonal antibody in cation exchange chromatography. *Journal of Chromatography A* **1362,** 186–193 (2014).

187. Luo, H., Cao, M., Newell, K., Afdahl, C., Wang, J., Wang, W. K. & Li, Y. Double-peak elution profile of a monoclonal antibody in cation exchange chromatography is caused by histidine-protonation-based charge variants. *Journal of Chromatography A* **1424,** 92–101 (2015).

188. Narayanan, H., Luna, M. F., von Stosch, M., Cruz Bournazou, M. N., Polotti, G., Morbidelli, M., Butté, A. & Sokolov, M. Bioprocessing in the digital age: the role of process models. *Biotechnology journal* **15,** 1900172 (2020).

189. Hutter, C., von Stosch, M., Bournazou, M. N. C. & Butté, A. Knowledge transfer across cell lines using Hybrid Gaussian Process models with entity embedding vectors. *arXiv preprint arXiv:2011.13863* (2020).

190. Wang, G., Briskot, T., Hahn, T., Baumann, P. & Hubbuch, J. Estimation of adsorption isotherm and mass transfer parameters in protein chromatography using artificial neural networks. *Journal of Chromatography A* **1487,** 211–217 (2017).

191. Karlberg, M., von Stosch, M. & Glassey, J. Exploiting mAb structure characteristics for a directed QbD implementation in early process development. *Critical reviews in biotechnology* **38,** 957–970 (2018).

192. Blech, M., Hörer, S., Kuhn, A. B., Kube, S., Göddeke, H., Kiefer, H., Zang, Y., Alber, Y., Kast, S. M., Westermann, M., *et al.* Structure of a therapeutic full-length anti-NPRA IgG4 antibody: dissecting conformational diversity. *Biophysical journal* **116,** 1637–1649 (2019).

193. Andrew, S. M. & Titus, J. A. Fragmentation of immunoglobulin G. *Current protocols in immunology* **21,** 2–8 (1997).

194. Saphire, E. O., Parren, P. W., Pantophlet, R., Zwick, M. B., Morris, G. M., Rudd, P. M., Dwek, R. A., Stanfield, R. L., Burton, D. R. & Wilson, I. A. Crystal structure of a neutralizing human IGG against HIV-1: a template for vaccine design. *science* **293,** 1155–1159 (2001).

195. Scapin, G., Yang, X., Prosise, W. W., McCoy, M., Reichert, P., Johnston, J. M., Kashi, R. S. & Strickland, C. Structure of full-length human anti-PD1 therapeutic IgG4 antibody pembrolizumab. *Nature structural & molecular biology* **22,** 953–958 (2015).

196. Shukla, A. A., Bae, S. S., Moore, J., Barnthouse, K. A. & Cramer, S. M. Synthesis and characterization of high-affinity, low molecular weight displacers for cation-exchange chromatography. *Industrial & engineering chemistry research* **37,** 4090–4098 (1998).

197. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. & Duchesnay, E. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **12,** 2825–2830 (2011).

198. Kraskov, A., Stögbauer, H. & Grassberger, P. Estimating mutual information. *Physical review E* **69,** 066138 (2004).

199. Ross, B. C. Mutual information between discrete and continuous data sets. *PloS one* **9,** e87357 (2014).

200. Breiman, L. Random forests. *Machine learning* **45,** 5–32 (2001).

201. Akaike, H. A new look at the statistical model identification. *IEEE transactions on automatic control* **19,** 716–723 (1974).

202. Rasmussen, C. *CKI Williams Gaussian processes for machine learning* 2006.

203. MacKay, D. J. & Mac Kay, D. J. *Information theory, inference and learning algorithms* (Cambridge university press, 2003).

204. Obrezanova, O., Csányi, G., Gola, J. M. & Segall, M. D. Gaussian processes: a method for automatic QSAR modeling of ADME properties. *Journal of chemical information and modeling* **47,** 1847–1857 (2007).

205. Zhu, C., Byrd, R. H., Lu, P. & Nocedal, J. Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization. *ACM Transactions on mathematical software (TOMS)* **23,** 550–560 (1997).

206. Jakobsson, N., Karlsson, D., Axelsson, J. P., Zacchi, G. & Nilsson, B. Using computer simulation to assist in the robustness analysis of an ion-exchange chromatography step. *Journal of Chromatography A* **1063,** 99–109. ISSN: 0021-9673 (2005).

207. Steinebach, F., Angarita, M., Karst, D. J., Müller-Späth, T. & Morbidelli, M. Model based adaptive control of a continuous capture process for monoclonal antibodies production. *Journal of Chromatography A* **1444,** 50–56 (2016).

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

209. Pirrung, S. M., Parruca da Cruz, D., Hanke, A. T., Berends, C., Van Beckhoven, R. F., Eppink, M. H. & Ottens, M. Chromatographic parameter determination for complex biological feedstocks. *Biotechnology progress* **34,** 1006–1018 (2018).

210. Susanto, A., Knieps-Grünhagen, E., von Lieres, E. & Hubbuch, J. High throughput screening for the design and optimization of chromatographic processes: assessment of model parameter determination from high throughput compatible data. *Chemical Engineering & Technology: Industrial Chemistry-Plant Equipment-Process Engineering-Biotechnology* **31,** 1846–1855 (2008).

211. Yamamoto, S., Nakanishi, K., Matsuno, R. & Kamikubo, T. Ion exchange chromatography of proteins—prediction of elution curves and operating conditions. I. Theoretical considerations. *Biotechnology and bioengineering* **25,** 1465–1483. ISSN: 0006-3592 (1983).

212. Yoshimoto, N. Simplified Methods Based on Mechanistic Models for Understanding and Designing Chromatography Processes for Proteins and Other Biological Products—Yamamoto Models and Yamamoto Approach—. *Preparative chromatography for separation of proteins,* 111–157 (2017).

213. Bertero, M., Poggio, T. A. & Torre, V. Ill-posed problems in early vision. *Proceedings of the IEEE* **76,** 869–889 (1988).

214. Barz, T., Körkel, S., Wozny, G., *et al.* Nonlinear ill-posed problem analysis in model-based parameter estimation and experimental design. *Computers & Chemical Engineering* **77,** 24–42 (2015).

215. Huuk, T. C., Hahn, T., Doninger, K., Griesbach, J., Hepbildikler, S. & Hubbuch, J. Modeling of complex antibody elution behavior under high protein load densities in ion exchange chromatography using an asymmetric activity coefficient. *Biotechnology journal* **12,** 1600336 (2017).

216. Bird, R. B., Stewart, W. E. & Lightfoot, E. N. *Transport phenomena* ISBN: 0470115394 (John Wiley & Sons, 2007).

217. Skamris, T., Tian, X., Thorolfsson, M., Karkov, H. S., Rasmussen, H. B., Langkilde, A. E. & Vestergaard, B. Monoclonal antibodies follow distinct aggregation pathways during production-relevant acidic incubation and neutralization. *Pharmaceutical research* **33,** 716–728 (2016).

218. Potty, A. & Xenopoulos, A. Stress-induced antibody aggregates. *BioProcess Int* **11,** 44–52 (2013).

219. Coffman, J. L., Kramarczyk, J. F. & Kelley, B. D. High-throughput screening of chromatographic separations: I. Method development and column modeling. *Biotechnology and bioengineering* **100,** 605–618 (2008).

220. Wiendahl, M., Schulze Wierling, P, Nielsen, J., Fomsgaard Christensen, D, Krarup, J., Staby, A. & Hubbuch, J. High throughput screening for the design and optimization of chromatographic processes–miniaturization, automation and parallelization of breakthrough and elution studies. *Chemical Engineering & Technology: Industrial Chemistry-Plant Equipment-Process Engineering-Biotechnology* **31,** 893–903 (2008).

221. Keller, W. R., Evans, S. T., Ferreira, G., Robbins, D. & Cramer, S. M. Use of MiniColumns for linear isotherm parameter estimation and prediction of benchtop column performance. *J Chromatogr A* **1418,** 94–102. ISSN: 1873-3778 (Electronic) 0021-9673 (Linking) (2015).

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

223. Hardin, A. M., Harinarayan, C., Malmquist, G., Axén, A. & van Reis, R. Ion exchange chromatography of monoclonal antibodies: Effect of resin ligand density on dynamic binding capacity. *Journal of Chromatography A* **1216,** 4366–4371 (2009).

224. Keener, R. N., Maneval, J. E. & Fernandez, E. J. Toward a robust model of packing and scale-up for chromatographic beds. 2. Flow packing. *Biotechnology progress* **20,** 1159–1168 (2004).

225. Keener, R. N., Maneval, J. E. & Fernandez, E. J. Toward a robust model of packing and scale-up for chromatographic beds. 1. Mechanical compression. *Biotechnology progress* **20,** 1146–1158 (2004).

226. Ladwig, J. E., Zhu, X., Rolandi, P., Hart, R., Robinson, J. & Rydholm, A. Mechanistic Model of pH and Excipient Concentration during Ultrafiltration and Diafiltration Processes of Therapeutic Antibodies. *Biotechnology Progress.* ISSN: 1520-6033 (2020).

227. Shekhawat, L. K., Manvar, A. P. & Rathore, A. S. Enablers for QbD implementation: Mechanistic modeling for ion-exchange membrane chromatography. *Journal of Membrane Science* **500,** 86–98 (2016).

228. Shekhawat, L. K., Pathak, M., Sakar, J. & Rathore, A. S. Process development in the Quality by Design paradigm: Modeling of Protein A chromatography resin fouling. *Journal of Chromatography A* **1570,** 56–66 (2018).

229. Nicoud, R.-M. *Chromatographic Processes* ISBN: 1107082366 (Cambridge University Press, 2015).

230. Baumann, P., Huuk, T., Hahn, T., Osberghaus, A. & Hubbuch, J. Deconvolution of high-throughput multicomponent isotherms using multivariate data analysis of protein spectra. *Engineering in Life Sciences* **16,** 194–201. ISSN: 1618-2863 (2016).

231. Kolodziej, M., Sauer, D. G., Beck, J., Marek, W. K., Hahn, R., Jungbauer, A., Dürauer, A., Piątkowski, W. & Antos, D. Scale up of a chromatographic capture step for a clarified bacterial homogenate–influence of mass transport limitation and competitive adsorption of impurities. *Journal of Chromatography A,* 460856. ISSN: 0021-9673 (2020).

232. Chung, S. & Wen, C. Y. Longitudinal dispersion of liquid flowing through fixed and fluidized beds. *AIChE Journal* **14,** 857–866 (1968).

233. Schuirmann, D. J. A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *Journal of pharmacokinetics and biopharmaceutics* **15,** 657–680 (1987).

234. Luciani, F., Galluzzo, S., Gaggioli, A., Kruse, N. A., Venneugues, P., Schneider, C. K., Pini, C. & Melchiorri, D. *Implementing quality by design for biotech products: Are regulators on track?* in *MAbs* **7** (2015), 451–455.

235. Jiang, C., Flansburg, L., Ghose, S., Jorjorian, P. & Shukla, A. A. Defining process design space for a hydrophobic interaction chromatography (HIC) purification step: application of quality by design (QbD) principles. *Biotechnology and bioengineering* **107,** 985–997. ISSN: 0006-3592 (2010).

236. Andris, S. & Hubbuch, J. Modeling of hydrophobic interaction chromatography for the separation of antibody-drug conjugates and its application towards quality by design. *Journal of biotechnology* **317,** 48–58 (2020).

237. Saleh, D., Wang, G., Mueller, B., Rischawy, F., Kluters, S., Studts, J. & Hubbuch, J. Cross-scale quality assessment of a mechanistic cation exchange chromatography model. *Biotechnology progress* **37,** e3081 (2021).

238. European Medicines Agency, Questions and answers: Improving the understanding of NORs, PARs, DSp and normal variability of process parameters. **EMA/ CHMP/ CVMP/ QWP/ 354895/ 2017** (2017).

239. Vogg, S., Müller-Späth, T. & Morbidelli, M. Current status and future challenges in continuous biochromatography. *Current opinion in chemical engineering* **22,** 138–144 (2018).

List of Tables

List of Figures

Abbreviations and Symbols

Abbreviations

AC	Affinity Chromatography
ADC	Antibody-Drug Conjugate
AI	Artificial Intelligence
APG	Acidic Peak Groups
bsAb	bispecific Antibody
BPG	Basic Peak Groups
CDR	Complementarity-Determining Region
CEX	Cation Exchange Chromatography
CHO	Chinese Hamster Ovary
cIEF	Capillary Isoelectric Focusing
CMA	Critical Material Attribute
CMC	Chemistry, Manufacturing and Controls
COVID-19	Coronavirus-Disease 2019
CPP	Critical Process Parameter
CSTR	Continuous Stirred-Tank Reactor
CV	Column Volume
CQA	Critical Quality Attributes
DoE	Design of Experiments
DSP	Downstream Processing
EAC	Equivalence Acceptance Criteria
EMA	European Medicines Agency
Fab	Fragment antigen binding
Fc	Fragment crystallizable region
FDA	Food and Drug Administration
FMEA	Failure Mode and Effect Analysis
FPLC	Fast Protein Liquid Chromatography
Fv	Fragment variable region
FR	Framework Region

GMoP	Good Modeling Practice
GMP	Good Manufacturing Practice
GPR	Gaussian Process Regression
GRM	General Rate Model
HC	Heavy Chain
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
HPSCX	High Performance Size Exclusion Chromatography
HPSEC	High Performance Strong Cation Exchange Chromatography
HTE	High-Throughput Experimentation
HTS	High-Throughput Screening
IEX	Ion-Exchange Chromatography
IgG	Immunoglobulin G
KMA	Key Material Attribute
KPI	Key Performance Indicator
KPP	Key Process Parameter
LAC	Lower Acceptance Criteria
LC	Light Chain
LGE	Linear Gradient Elution
LHS	Liquid Handling Station
LML	Log-Marginal-Likelihood
LMW	Low Molecular Weight Impurities
mAb	Monoclonal Antibody
MAE	Mean Absolute Error
MSE	Mean Square Error
NBE	New Biological Entity
OFAT	One-Factor-at-a-Time Sampling
RMSEP	Root Mean Square Error
NRMSEP	Normalized Root Mean Square Error of Prediction
PAR	Proven Acceptable Range
PAT	Process Analytical Technologies
PCS	Process Characterization Study
PDB	Protein Data Bank
pI	Isoelectric Point

QbD	Quality by Design
QSAR	Quantitative Structure-Activity Relationship
QSPR	Quantitative Structure-Property Relationship
RFE	Recursive Feature Elimination
RMSEP	Root Mean Square Error of Prediction
SASA	Solvent Accessible Surface Area
SD	Standard Deviation
SDM	Stoichiometric Displacement Model
ScDM	Scale-Down Model
SMA	Steric Mass-Action
TDM	Transport-Dispersive Model
TOST	Two One-Sided t-Test
UF/DF	Ultrafiltration and Diafiltration
UAC	Upper Acceptance Criteria
UPLC	Ultra High Performance Liquid Chromatography
USP	Upstream Processing

Symbols

a_i	Activity	M
c_i	Bulk phase concentration of species i	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
d	Column diameter	cm
D_{app}	Apparent dispersion coefficient	$\frac{mm^2}{s}$
D_{ax}	Axial dispersion coefficient	$\frac{mm^2}{s}$
D_{pore}	Pore diffusion coefficient	$\frac{mm^2}{s}$
G	Gibbs free energy	-
G^0	Standard Gibbs free energy	-
h	Ionic hydration number	-
k_{ads}	Adsorption coefficient	-
k_{des}	Desorption coefficient	-
k_{eff}	Mass transfer coefficient	$\frac{mm}{s}$
K_{eq} or k_{eq}	Equilibrium constant of isotherm	-
k_{film}	Film diffusion coefficient	$\frac{mm}{s}$
k_{kin}	Kinetic coefficient	-
L	Column length	cm
L_{CSTR}	Reactor Length of CSTR	cm
N	Plate number	-
N_i	Molar flux of species i	$\frac{mol}{s}$
$pK_{a,i}$	Acid Dissociation Constant	
q_i	Stationary phase concentration of species i	M
r	Radius	mm
r_p	Particle radius of adsorbent	mm
R	Ideal gas constant	$\frac{J}{kg \cdot K}$
S	Cross Sectional Area	mm^2
t	Time dimension	s
T	Temperature	K
u	Effective velocity	$\frac{mm}{s}$
u_{int}	Interstitial velocity	$\frac{mm}{s}$
ε	Column porosity	-
ε_t	Total porosity	-
ε_p	Particle porosity	-

Λ	Ionic capacity	M
ν	Characteristic charge	-
μ	Chemical potential	-
μ^0	Standard chemical potential	-
σ	Shielding parameter	-