A Priori Approach to Predicting Therapeutic Antibody Behavior in Multimodal Chromatography

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. Rudger Thomas Heß
aus Erlangen

Tag der mündlichen Prüfung: 07.10.2024
Erstgutachter: Prof. Dr. Jürgen Hubbuch
Zweitgutachter: Prof. Dr. Matthias Franzreb

Danksagung

Diese Dissertation entstand im Rahmen einer Direktkooperation zwischen dem Karlsruher Institut für Technologie (KIT) und dem Unternehmen Boehringer Ingelheim. Der Erfolg dieser Arbeit ist auf die langjährige Unterstützung zahlreicher Personen zurückzuführen. Zunächst möchte ich mich bei meinem Doktorvater, Prof. Dr. Jürgen Hubbuch, für seine wissenschaftliche und persönliche Betreuung bedanken. Ebenso gilt mein Dank Prof. Dr. Matthias Franzreb für die Übernahme des Korreferates.

Ich möchte auch meinen Kollegen bei Boehringer Ingelheim Pharma GmbH & Co. KG einen großen Dank aussprechen. Mein Arbeitsumfeld in Biberach an der Riß wurde durch sie jeden Tag bereichert und über die Jahre durch tiefe Freundschaften verschönert. Mein besonderer Dank gilt Dr. Jan Visser, Dr. Markus Wendeler, Dr. Thomas Schwab und Dr. Joey Studts, die die Anfertigung dieser Dissertation in ihrer jeweiligen Abteilung und Gruppe ermöglicht haben. Ich möchte mich auch bei meinen Betreuern Jan-Hendrik Grosch und David Saleh bedanken, die mich stets mit Rat und Tat unterstützt haben. Ein großer Dank gilt zudem Gang Wang, Federico Rischawy und Till Briskot, die mich sowohl fachlich als auch menschlich gefördert und geleitet haben. Des Weiteren möchte ich mich bei dem gesamten Modellierungsteam und meinem Labor für ihre tatkräftige Unterstützung und offene Art bedanken. Ein wesentlicher Bestandteil dieser Dissertation wurde durch meine drei Studenten Doil Yun, Jan Fäßler und Ahmed Mama ermöglicht, und ich möchte ihnen hiermit meinen Dank für ihre herausragende Arbeit aussprechen. Es erfüllt mich mit großem Stolz, dass sich alle drei auf den Weg ihrer eigenen Promotion begeben haben.

Ich will diese Gelegenheit nutzen, um mich bei zahlreichen Kollegen und der gesamten MAB-Arbeitsgruppe zu bedanken. Es ist keine Selbstverständlichkeit, auf so viele zuverlässige, ehrliche und aufgeschlossene Menschen zu treffen.

Zuletzt gilt mein Dank meiner Familie, meiner Partnerin Sabine und engsten Freunden, die mich auf dieser Reise begleitet haben. Ihr habt mich durch eure Geduld und Rücksichtnahme gestützt und seid alle ein Teil dieser Dissertation geworden.

Rudger Heß

"VOLATILITY IS A PHENOMENON IN SEARCH OF A THEORY"

- JAMES TOBIN

Zusammenfassung

Diese Dissertation stellt den ersten wissenschaftlichen Ansatz dar, der eine Verbindung zwischen mechanistischer und strukturbasierter Modellierung für die Prozessentwicklung der multimodalen Chromatographie (MMC) herstellt. MMC ist eine Aufreingungsmethode, welche eine Kombination mehrerer Oberflächenchemien nutzt, um die Qualität und Effizienz der Molekülauftrennung zu verbessern. Die hier vorgestellten Ergebnisse könnten erheblich zu dem routinierten Einsatz von MMC in der biopharmazeutischen Industrie beitragen und die Verwendung von Proteinstrukturwissen in diesem Bereich fördern.

Monoklonale Antikörper (mAbs), gelten als eine der bedeutendsten Produktklassen von Medikamenten biologischen Ursprungs. Sie haben die Krankheitsbehandlung in therapeutischen Bereichen wie Onkologie und Immunologie revolutioniert. Ein erheblicher Kostenfaktor bei der Herstellung von Antikörpermedikamenten ist die Produktaufreinigung. Daher wurde innerhalb der Entwicklung von biopharmazeutischen Aufreinigungsprozessen (engl. “Downstream Processing“, DSP) bereits bedeutsame Arbeit geleistet, um deren Produktivität zu steigern. Dabei wurden fortgeschrittene Reinigungstechniken wie MMC in den DSP-Methodensatz aufgenommen. Die verbesserte Aufreinigungsleistung von MMC ist besonders von Vorteil für neu entwickelte Antikörperformate, welche spezifische Proteineigenschaften und Verunreinigunsprofile aufzeigen und daher etablierte Techniken und Prozeszabläufe überlasten können. Die erhöhte Komplexität von MMC erfordert jedoch eine produktspezifische Entwicklung dieser Aufreinigungsstufe. Oft müssen zahlreiche Prozessbedingungen untersucht werden, um die Anwendbarkeit von MMC für unterschiedlichste mAb Produkte festzustellen. Dies kann zu einer erheblichen Verlängerung der Entwicklungszeiten und Erhöhung des Ressourcenaufwands führen, was die Implementierung dieser vielversprechenden Technik in die mAb-Aufreinigung verzögert.

Üblicherweise sind für die Entwicklung von MMC-Prozessen kostspielige sowie zeitintensive Vorversuche notwendig, die größere Mengen an Probenmaterial erfordern. Jedoch sind vor allem in den frühen Phasen der Prozessentwicklung sowohl das Probenmaterial als auch die verfügbare Entwicklungszeit stark eingeschränkt. Der wirtschaftliche Erfolg zukünftiger Arzneimittel kann durch eine weitere Reduktion dieser Entwicklungszeiten begünstigt werden. Hieraus folgt ein erklärtes Ziel der biopharmazeutischen DSP, arbeitsintensive Laborexperimente durch den Einsatz von vornherein bekannten Strukturwissen zu ersetzen. Diese Methodik ist unter *a priori* Prozessentwicklung bekannt. Da das Prozessverhalten durch Oberflächenwechselwirkungen zwischen Antikörpern und deren Aufreinigungsmethoden beinflusst wird, könnte die gezielte Nutzung von Strukturinformationen, Entwicklungszeiten erheblich verkürzen. Frühere Bemühungen, Strukturinformationen mit Hilfe physikalischer Modelle für die MMC-Entwicklung nutzbar zu machen, trafen

aufgrund des zugrundeliegenden komplexen Interaktionsmechanismus auf Widerstand. Empirische Modellierungsansätze waren wiederum durch die Menge der verfügbaren Kalibrierungsdaten begrenzt. Die Multiskalenmodellierung könnte ein vielversprechender Ansatz sein, um eine *a priori* Prozessentwicklung zu verwirklichen, indem molekulare sowie makroskopische Effekte der Adsorptionschromatographie berücksichtigt werden. Der hier gezeigte Ansatz der Multiskalenmodellierung ist eine Zusammenführung von statistischer und mechanistischer Modellierung. Dabei werden mit Hilfe von statistischen Modellen proteinspezifische Adsorptionsparameter aus molekularen Proteineigenschaften vorhergesagt. Diese Adsorptionsparameter werden dann in mechanistischen Chromatographiemodellen genutzt, um makroskopische Chromatographievorhersagen zu ermöglichen.

Diese kumulative Dissertation umfasst drei für sich selbst stehende Publikationen. Sie beschreiben die aufeinander aufbauende Entwicklung eines Multiskalenmodells zur Vorhersage von MMC-Interaktionen. Die Vorhersagen sollten dabei ausschließlich anhand der Aminosäuresequenzen von Antikörper ermöglicht werden. Diese Form der Multiskalenmodellierung ist ein datengetriebener Ansatz, der eine erhebliche Menge an Kalibrierungsmolekülen benötigt, von denen physikochemische Informationen berechnet sowie mechanistische Chromatographiemodelle erstellt werden können. Aus diesem Grund lag der initiale Fokus auf der effizienten Bereitstellung von molekülspezifischen Adsorptionsparametern durch die Generierung von mechanistischen Chromatographiemodellen.

Die erste Veröffentlichung (Kapitel 3) stellt einen standardisierten Arbeitsablauf vor, der die mechanistische Modellierung von MMC für die industrielle Antikörperaufreinigung ermöglicht. Mechanistische Modellierung dient der Simulation von chromatographischen Trennungen als Funktion variierender Prozessparameter. Aufgrund des komplexen Adsorptionsmechanismus der Protein-Liganden-Interaktion stellt die MMC-Modellierung eine anspruchsvolle Aufgabe dar. Bestehende mechanistische Modelle wurden als unpraktikabel erachtet, ob ihrer großen Anzahl an Modellparametern, die vor der Modellanwendung bestimmt werden mussten. Darüber hinaus erhöhte eine ungenaue Parameterkalibrierung, auf Grundlage von wenig-informativen und nicht standardisierten experimentellen Arbeitsabläufen das Risiko unzuverlässiger Modellvorhersagen. Innerhalb von Kapitel 3 wurde die Komplexität eines bestehenden Adsorptionsmodells erheblich reduziert, während die Genauigkeit der Simulation gängiger Prozessbedingungen in einem linearen Adsorptionsbereich beibehalten werden konnte. Darüber hinaus wurde ein standardisierter labor- und computergestützter Arbeitsablauf etabliert, um eine effiziente Kalibrierung und Validierung der thermodynamischen Modellparameter zu ermöglichen. Diese Methodik stellt einen erheblichen Beitrag im Umfeld des DSP dar, indem ein vereinfachter Zugang zu einer modellgestützten MMC-Entwicklung unter niedrigen Beladungsdichten ermöglicht wurde. Darüber hinaus ist der vorgestellte Arbeitsablauf eine Voraussetzung für die effiziente Erfassung von proteinspezifischen Adsorptionsparametern hinsichtlich des finalen Multiskalenmodells.

In der zweiten Veröffentlichung (Kapitel 4) wird eine Methode zur Vorhersage von Antikörperinteraktionen während linearer pH-Gradienten in MMC vorgestellt. Die Feststellung des pH-Retentionspunktes gilt als informatives Vorexperiment für die MMC-

Prozessentwicklung. Hieraus kann ein optimaler pH-Bereich bei konstanter Salzkonzentration für darauffolgende Aufreinigungsprozesse abgeleitet werden. Die zugrunde liegende Modellklasse wird als quantitative Struktur-Eigenschafts-Beziehung (engl. quantitative structure-property relationship, QSPR) bezeichnet. Dabei handelt es sich um einen mehrstufigen Prozess, bei dem physikochemische Moleküleigenschaften mit experimentellen Beobachtungen mittels multivariater Regressionsalgorithmen korreliert werden. Im Rahmen von Kapitel 4 wurde die Strukturvorhersage von Antikörpern und die Berechnung physikochemischer Eigenschaften automatisiert, ebenso wie ein experimentelles Hochdurchsatzverfahren und ein fortschrittlicher QSPR-Arbeitsablauf etabliert. Dieser Ansatz ermöglichte die Vorhersage der pH-Retention für verschiedene Antikörper und Antikörperformate anhand ihrer Aminosäuresequenz. Der unmittelbare Nutzen dieses Modells besteht in der strukturbasierten Vorhersage von Prozessbedingungen ohne die Notwendigkeit von vorausgehenden Laborexperimenten. Darüber hinaus könnte das kalibrierte QSPR-Modell als computergestützte Methode verwendet werden, um potenzielle Antikörperkandidaten anhand ihrer Prozesssierbarkeit zu identifizieren, was im Englischen als "Manufacturability Assessment" bezeichnet wird. Im Hinblick auf das übergeordnete Ziel, eine *a priori* Prozessentwicklung für MMC zu realisieren, war das Ergebnis dieser Veröffentlichung von zweifachem Nutzen. Erstens wurde eine adaptive Vorlage für die Realisierung von QSPR-Modellen erstellt, die später auf das endgültige Multiskalenmodell übertragen werden konnte. Dies beinhaltet eine automatisierte Strukturvorhersage, Berechnung physikochemischer Eigenschaften, Etablierung von Hochdurchsatzexperimenten sowie eine Modellkalibrierungs- und Validierungsroutine. Zweitens kann der vorhergesagte Prozess-pH Wert als Ausgangspunkt für das endgültige Multiskalenmodell verwendet werden, um eine relevante Prozessvorhersage zu gewährleisten.

Nachdem alle notwendigen Voraussetzungen für die Generierung ausreichender Kalibrierungsdaten erfüllt waren, stellt die dritte und abschließende Veröffentlichung (Kapitel 5) ein Multiskalenmodell zur Vorhersage der Antikörperinteraktion in MMC vor. Im Rahmen dieser Veröffentlichung wurden mechanistische Chromatographie- und statistische QSPR-Modellierung kombiniert. Hierzu wurde ein großer Satz an Antikörperstrukturen und -formaten untersucht. Dieser umfasst 59 volllängen mAbs, fünf unterschiedliche Formate wie IgG1, IgG4 sowie drei voneinander abweichende bi- und tri-spezifische Konstrukte. Von jedem dieser Moleküle wurden antikörperspezifische Eigenschaften berechnet und mechanistische Chromatographiemodelle in verschiedenen pH-Umgebungen kalibriert. Dabei wurden Adsorptionparamter mit Hilfe des neuentwickelte MMC-Modellierungsansatzes aus Kapitel 3 bestimmt. Daraufhin wurden diese Daten verwendet, um ein QSPR-Model, abgeleitet aus Kapitel 4, zu kalibrieren, was die Vorhersage von Adsorptionsparametern aus Antikörperstrukturen ermöglichte. Abschließend wurde die Genauigkeit des Multiskalenansatzes durch die Simulation von MMC-Chromatographie für mehrere Antikörper, Formate und pH-Werte unter niedrigen Beladungsdichten validiert. Dieses Multiskalenmodell ermöglicht die computergestützte Identifizierung von potenziellen MMC-Prozessbedingungen für verschiedenste Antikörperkandidaten. Die gezeigten Ergebnisse veranschaulichen das Potenzial einer *a priori* Prozessentwicklung in der Antikörperaufreinigung durch die Vorhersage chromatographischen Verhaltens auf Grundlage der Proteinstruktur. Darüber

hinaus wurde durch die Untersuchung der strukturellen Antikörpereigenschaften, im Zuge der QSPR-Modellierung, ein Einblick in den multimodalen Adsorptionsprozess gewonnen. Perspektivisch könnten die gewonnen Informationen zur Weiterentwicklung mechanistischer MMC-Modelle führen.

Zusammenfassend leistet diese Dissertation einen erheblichen Beitrag zur beschleunigten Entwicklung von MMC in der biopharmazeutischen Antikörperaufreinigung. Ein standardisierter Arbeitsablauf für die mechanistische MMC-Modellierung wurde eingeführt, um eine modellgestützte Prozessentwicklung zu ermöglichen. Ein QSPR-Modell zur Vorhersage des optimalen Prozess-pH Wertes auf der Grundlage der Aminosäuresequenz wurde etabliert. Diese Methode beschleunigt die Prozessidentifikation und könnte die Auswahl von Antikörperkandidaten durch Berücksichtigung ihrer "Manufacturability" leiten. Die etablierten Arbeitsabläufe wurden in einem Multiskalenmodell zur Vorhersage des MMC-Verhaltens von Antikörpern kombiniert. Dies ermöglicht ein rein strukturbasiertes und computergestütztes Prozessdesign. Während diese Ergebnisse bereits das Potenzial für eine *a priori* Prozessentwicklung aufzeigen, könnten die aufgedeckten strukturellen Erkenntnisse zur Weiterentwicklung von Adsorptionsmodellen in der MMC führen. Angesichts der erhöhten Komplexität der MMC-Interaktion im Vergleich zu anderen DSP-Verfahrensschritten könnte diese Dissertation als Vorlage für weitere *a priori* Prozessmodelle sowie des mechanistischen Verständnisgewinns in verschiedensten Aufreinigungsmethoden dienen.

Abstract

This dissertation depicts the first study to establish a link between mechanistic and structure-based modeling for the development of multimodal chromatography (MMC) processes. MMC combines different surface chemistries to enhance the quality and efficiency of molecule purification. The research presented herein contributes significantly to the adoption of MMC in the biopharmaceutical industry and promotes the use of structural knowledge in this domain.

Monoclonal antibodies (mAbs), a primary product class of biologically derived medicines, have revolutionized disease treatment in therapeutic areas such as oncology and immunology. A significant cost contributor to antibody products is their purification development, known as downstream processing (DSP). Consequently, extensive research has been conducted to enhance the productivity of biopharmaceutical purification. In this context, advanced purification techniques like MMC have been introduced to the DSP toolbox. The improved purification performance of MMC is especially advantageous for novel antibody formats exhibiting specific impurity profiles, which can challenge established purification techniques and processing routines. However, the increased complexity of MMC necessitates a product-specific development for this unit operation. Often, a variety of process conditions need to be explored, to assess the applicability of MMC for a diverse range of mAb products. This can inflate development timelines and resource expenditure, limiting the implementation of this promising technique into mAb purification.

Typically, the development of MMC processes involves costly and time intensive screening experiments requiring substantial amounts of sample material. Particularly in the early stages of process development, sample material as well as development time is scarce. The economic success of launched drug products can be facilitated by further reducing development timelines. Therefore, it has been a long-standing goal of the biopharmaceutical DSP to replace labor-intensive wet lab experimentation by harnessing prior knowledge of molecule structure. This methodology is often referred to as *a priori* process development. As surface-surface interactions between molecules and purification methods largely affect their process behavior, utilizing structural information could significantly accelerate development timelines. Previous attempts in harnessing structural information for MMC development have struggled to capture the complex mechanism using physics-informed models. Conversely, empirical modeling approaches were limited by the amount of available training data. Multiscale modeling could be a promising path towards achieving *a priori* process development by considering molecular as well as macroscopic effects of adsorption chromatography. The multiscale modeling approach employed in this thesis merges statistical and mechanistic modeling. Hereto, protein-specific adsorption parameters are predicted

from the molecular properties of proteins. These adsorption parameters are then incorporated into mechanistic chromatography models to enable macroscopic chromatography predictions.

This publication-based thesis comprises three independent research articles. These articles detail the sequential development of a multiscale model for predicting MMC interaction based solely on antibody sequence information. This kind of multiscale model is a data-driven approach and requires a substantial amount of training molecules from which physicochemical information can be calculated and mechanistic chromatography models be generated. Accordingly, the initial focus of this thesis was the efficient acquisition of molecule specific adsorption parameters by generation of mechanistic chromatography models.

The first publication (Chapter 3) presents a standardized workflow for mechanistic modeling of MMC in antibody purification at industrially relevant process conditions. Mechanistic modeling enables the simulation of chromatographic operations as a function of varying process parameters. MMC modeling is considered challenging due to the complex adsorption mechanism that needs to be simulated during the multimodal protein-ligand interaction. Previous thermodynamic models were deemed impractical due to the large number of model parameters that needed to be determined prior to model application. Moreover, inaccurate parameter calibration, resulting from uninformative and non-standardized experimental routines, increased the risk of unreliable model predictions. Within Chapter 3, the complexity of an existing adsorption model was substantially reduced while maintaining its accuracy for simulating common process conditions in a linear adsorption regime. Furthermore, a standardized wet lab and *in silico* workflow was established to enable efficient calibration and validation of the thermodynamic model parameters. This methodology alone represents a substantial contribution in the field of DSP by facilitating straightforward access to model assisted MMC development at low loading density conditions. Additionally, the presented workflow serves as a prerequisite for the final multiscale model to enable the efficient acquisition of protein-specific adsorption parameters.

In the second publication (Chapter 4), a method for predicting antibody retention during linear pH gradients in MMC is introduced. Determining retention pH is considered an informative screening experiment for MMC process development. It can lead to a reliable indication of an optimal range for process pH at a given salt concentration. The underlying model was based on quantitative structure-property relationship (QSPR), which is a multistage process in which physicochemical molecule characteristics are correlated to experimental properties using multivariate regression algorithms. Within Chapter 4, antibody structure prediction and calculation of physicochemical characteristic were automated, alongside the establishment of an experimental high-throughput setup and an elaborate QSPR workflow. This approach enabled the prediction of pH retention for various antibodies and antibody formats using only their amino acid sequence. The immediate benefit of this model includes the structure-based prediction of process conditions without the need for performing wet lab experimentation. Additionally, the calibrated QSPR model could be utilized as an *in silico* screening tool for identifying antibody candidates suitable for purification at selected process pH. This is often referred to as manufacturability

assessment. Regarding the overarching goal of realizing *a priori* process development in MMC, the outcome of this publication was two-fold. First, an adaptive template for developing QSPR models was established that could later be transferred to the final multiscale model. This includes automated structure prediction, physicochemical characteristic calculation, high-throughput experimentation, as well as a model training and testing routine. Second, the predicted process pH can be utilized as a starting point for the final multiscale model to assure relevant process prediction.

After meeting the necessary prerequisites for gathering sufficient training data, the third and final publication (Chapter 5) introduces a multiscale model for predicting antibody interaction in MMC. Here, mechanistic chromatography and statistical QSPR modeling were combined, and a large quantity of antibody structures and formats investigated. The molecules comprise 59 full-length mAbs, five different formats, including IgG1s, IgG4s, and three different kinds of bi- and trispecific constructs. From each molecule antibody-specific characteristics were calculated and mechanistic chromatography models at various pH environments calibrated. Accordingly, adsorption parameters were determined using the newly developed MMC modeling workflow introduced in Chapter 3. The data was than utilized to train a QSPR model, modified from Chapter 4, to enable the prediction of adsorption parameters from antibody structure. The accuracy of this approach was validated by simulating MMC chromatography for multiple antibodies, formats, and pH values at low loading density conditions. This multiscale model allows *in silico* screening of potential MMC operating conditions targeting the initial process development of antibody candidates. The results demonstrate the potential of *a priori* process development in antibody purification by predicting chromatographic behavior based on protein structure. Furthermore, insights into the multimodal adsorption process through investigation of structural contributions of the QSPR parameter models were gained. These insights could yield further developments of mechanistic MMC models.

In conclusion, this dissertation provides a substantial contribution to the accelerated development of MMC in the biopharmaceutical antibody purification across multiple areas. A standardized workflow for mechanistic MMC modeling was introduced to enable a model-assisted process development. A QSPR model for predicting optimal process pH based on the amino acid sequence was established. This tool accelerates process identification and could guide antibody candidate selection by accounting for manufacturability profiles. The established workflows were combined in a multiscale model for predicting antibody behavior in MMC. The model predictions enable an *in silico* process design requiring only amino acid sequence information. While these results alone demonstrate the potential for *a priori* process development, the structural correlations could lead to further developments of advanced adsorption models for MMC. Given the exceptional complexity of the MMC interaction compared to other DSP operations, this dissertation could serve as a blueprint for enabling *a priori* process development and increasing mechanistic understanding throughout different unit operations.

Contents

1 Introduction

The biopharmaceutical industry has seen remarkable growth in recent years, becoming a vital sector in the field of medicine. This industry focuses on producing recombinant proteins, also known as biologics, which are used to treat a wide range of conditions and diseases. Unlike small molecules like aspirin, biologics are more complex and are produced using living cells rather than chemical synthesis [1]. By 2021, monoclonal antibodies (mAbs) made up more than 80% of the biopharmaceutical market to enable a targeted treatment for various therapeutic needs such as oncology, hematology, and inflammatory diseases [2,3]. The success of mAbs can be attributed to their increased specificity, potency, and half-life compared to other therapeutic modalities, as they lean on the efficacious pathways derived from the natural immune system [4].

While mAb products display clear therapeutic benefits, their structural complexity presents significant manufacturing challenges. These challenges are tackled by the Chemistry, Manufacturing, and Controls (CMC) activities to guarantee product safety while operating under physiological process conditions as required by the regulatory agencies [5]. Instead of using consistent chemical processes as implemented for small molecules, biologics are produced by genetically modified cells in a bioreactor, a process known as upstream processing (USP). Following cell expression, the mixture containing product and impurities is separated from cells or cell debris and purified during downstream processing (DSP) [6]. Purification is essential to isolate the desired product from potentially harmful contaminants, ensuring the safety of intravenous application to patients [7]. Contaminants include DNA and other host cell proteins (HCP) termed process related impurities as well as product related impurities like misfolded mAbs, aggregates, fragments, and post-translational modifications [8].

The structural similarities of mAb products enables a platformed purification process to reduce development and production expenditures, relying on a well-defined sequence of unit operations [9]. Preparative liquid chromatography is particularly important for mAb purification due to its high selectivity towards molecular properties like size, charge, or hydrophobicity, and its mild operating conditions. During chromatographic purification, product-impurity separation is realized by selectively binding target molecules to a solid stationary phase, while non-target molecules remain in a liquid mobile phase, *vice versa*.

In recent years, the success rate for regulatory approval of mAb candidates during clinical trials has decreased due to a saturated market environment, resulting in higher capital requirements for mAb products [10,11]. Accordingly, the exploration of innovative pathways in treating various diseases has been driven forward, which led to substantial

increases in the structural diversity of mAb therapeutics [12,13]. While minor adjustments to platformed purification processes are required for standard formats, non-standard mAbs challenge existing development workflows and can necessitate a resourceful process redesign and characterization. This particularly affects small and medium-sized pharmaceutical companies, leading to increased competition and forced consolidation [14]. To succeed in this competitive market, development timelines and production costs must be significantly reduced through technical innovation [15].

Current research in mAb purification focuses on reducing manufacturing costs and increasing the efficiency of process development, particularly for chromatographic unit operations [16]. To reduce material consumption and improve equipment utilization during mAb production, single-use technologies as well as continuous processing has been implemented by replacing stainless steel with disposable materials or connecting multiple unit operations [17,18]. More importantly, the performance of chromatography purification has progressed by utilizing advanced material and surface chemistries to improve stability, capacity, and impurity depletion [19]. Meanwhile, chromatography development was accelerated by using automated high-throughput systems, harnessing structural information of mAb candidates, as well as using statistical and mechanistic process models [20–23]. Additionally, advanced process analytics benefit both manufacturing and development by improving process monitoring [24]. These developments were also facilitated by the regulatory agencies to further improve product safety through increased process understanding as introduced by the quality by design (QbD) initiative [25,26].

One promising approach for cutting manufacturing costs is the use of multimodal chromatography (MMC), which combines multiple modes of interaction, such as ion exchange, hydrophobic interaction, and affinity, in a single chromatographic step [27]. This technique can provide enhanced selectivity and resolution, leading to improved purification performance and potentially reducing the number of required chromatographic unit operations [28]. However, its process development is complex and lacks fundamental understanding of multimodal protein-ligand interaction. Lacking process understanding is especially disadvantageous as current process development relies on empirical knowledge and has to perform extensive screening experimentation [29,30].

The following sections will establish a theoretical foundation for facilitating the adoption of MMC by utilization of protein structure information. Current approaches in antibody purification are presented, emphasizing the physicochemical characteristics of this product class and MMC. Thereafter, a brief introduction in mechanistic chromatography modeling is given, explaining practices of simulating macroscopic process behavior. Lastly, fundamentals of molecule modeling and physicochemical description as well as statistical analysis of structure-function relationships are addressed.

1.1 Purification of Therapeutic Antibodies

The process of purifying therapeutic antibodies must adhere to the highest regulatory standards to ensure product safety and efficacy. The biological origin and structural complexity of these products necessitate physiological process conditions for their purification unlike chemical compounds. The purification strategy is defined by the physicochemical properties of the product, which must be thoroughly characterized. In the subsequent sections, the structural characteristics of mAbs will be discussed, current downstream processing approaches presented, and the principles of MMC explored.

1.1.1 Monoclonal Antibodies

MAbs have become a powerful tool in modern medicine. They can treat a wide range of diseases due to their unique ability to mimic or augment the body's natural immune response. This has led to a revolution in the field of immunotherapy, opening new avenues for disease treatment and prevention. Antibodies, also known as immunoglobulins (Ig), are Y-shaped proteins naturally found on the B cell membrane. They play a vital role in the adaptive immune system, defending against pathogens. Antibodies are involved in a key process called opsonization. In this process, antibodies bind selectively to multiple surface characteristics of pathogens, known as antigens. This leads to coagulation and marks the pathogens for destruction by immune cells. The term “monoclonal” refers to an antigen-specific antibody that is isolated from a single B cell clone for therapeutic or research applications [31].

Antibodies naturally occur in five main isotypes: IgA, IgD, IgE, IgG, and IgM. Each class has a unique structure and function within the immune response, helping the body fight a variety of pathogens. The IgG class of antibodies is further divided into four subclasses: IgG1, IgG2, IgG3, and IgG4. These subclasses have structural differences that result in different immune responses and catabolic half-lives [32]. The IgG1 subclass, illustrated in Fig. 1.1, is the most prevalent in human serum and is highly effective against viruses and bacteria. Due to its beneficial pharmaceutical properties, IgG1 is the most common format for antibody therapeutics, followed by the IgG4 and IgG2 subclasses [4,12].

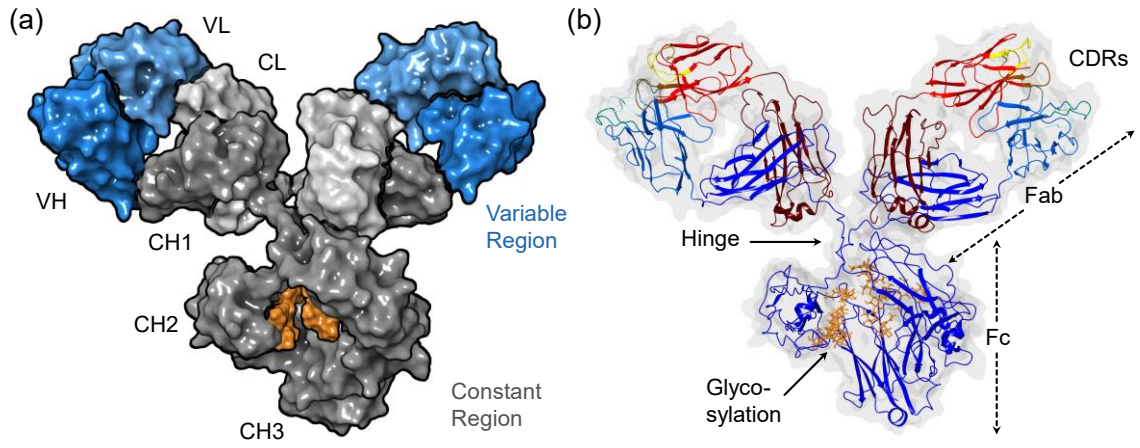

Fig. 1.1: Structure of IgG1 antibody (PDB entry 1IGY).(a) Surface representation of IgG1, colored according to antibody functional domains. (b) Ribbon representation indicating secondary structures, glycosylation and CDRs in brighter colors. Figure adapted from Absolute Antibody [33].

The IgG1 monomer weighs approximately 150 kDa and consists of four polypeptide chains, as depicted in Fig. 1.1. The four amino acid chains include two identical heavy chains (HC) and two identical light chains (LC). They fold into three equal-sized globular portions, stabilized by disulfide bonds and connected by a flexible stretch of polypeptide chain known as hinge region. The number of interchain disulfide bonds and the length of the hinge region differentiate the IgG molecules into their subclasses. The globular regions form a Y-shaped conformation with the two arms termed fragment antigen-binding (Fab) and the trunk known as fragment crystallizable (Fc). The antibody's quaternary structure (read Section 1.3.1) is divided functionally into a variable and a constant region, based on their sequence variability. The constant region is highly conserved within one antibody subclass and includes the Fc domain and the proximal half of each Fab domain, interacting with a limited number of effector molecules and cells [34]. The distal half of the Fab domains, which enables selective binding to varying antigen motives, is classified as the variable region (VR).

The secondary building blocks of these regions are further divided into 110 amino acid long β-barrel domains. Each domain has a molecular weight of approximately 12 kDa and is stabilized by an intrachain disulfide bond. The variable and constant regions are numbered in ascending order in segments of β-barrel domains. Each heavy chain comprises one variable domain (VH) and three constant domains (CH1, CH2, and CH3). Each light chain consists of one variable domain (VL) and one constant domain (CL). An oligosaccharide chain is embedded between the CH2 domains of the Fc region. The composition and precise binding location of the oligosaccharide is cell-line dependent and is referred to as glycosylation [35]. The glycosylation pattern of an antibody is critical for its pharmacokinetic and pharmacodynamic characteristics.

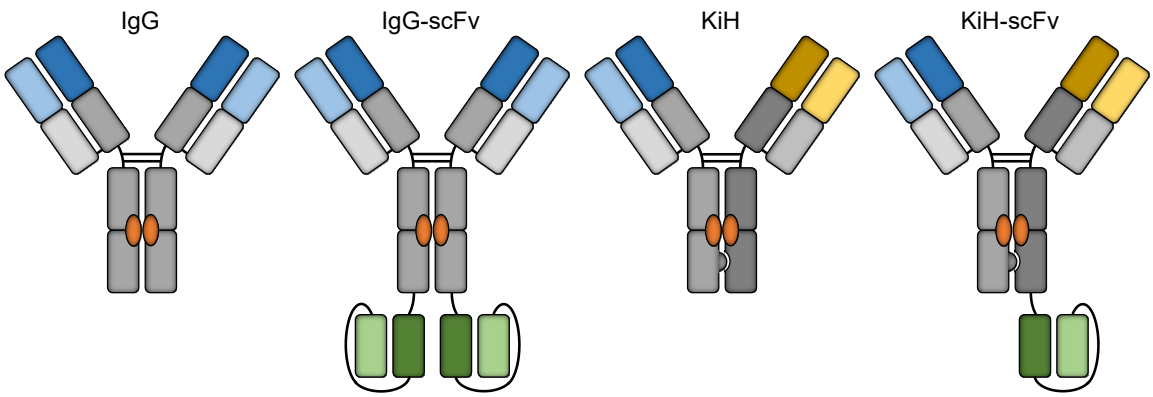

Fig. 1.2: Schematic of therapeutic antibody formats investigated in this thesis.The structure of immunoglobulin isotype G (IgG), IgG appended by a single-chain fragment variables (IgG-scFv), Knob-in-Hole bispecific (KiH), and KiH appended by a scFv (KiH-scFv) are shown from left to right. Variable regions are colored in blue, yellow, and green whereas the constant region is colored in grey. Furthermore, the glycosylation is colored in orange. The antibody hinge region and linkers are indicated by solid black lines.

The variable domains of the heavy and light chains together form the antigen-binding site, known as the paratope. The antigen-binding site comprises six hypervariable loops, which structurally adapt to complement the epitopes of the antigen. These loops are often called the complementarity-determining region (CDR) and include three heavy chain CDRs (H1, H2, H3) and three light chain CDRs (L1, L2, L3), with H3 being the most structurally variable. The framework-regions (FR) account for the intrachain sequences between the CDR loops of the heavy chain (HFR1, HFR2, HFR3, HFR4) and the light chain (LFR1, LFR2, LFR3, LFR4). The binding affinity of an antibody to an antigen reflects the likelihood of paratope-epitope association, which depends on steric restrictions and the release of free energy upon binding [36]. As the IgG subclass has two antigen-binding sites, its interaction is bivalent, and the accumulated binding strength is termed avidity.

There have been significant advancements in the field of therapeutic antibodies. Antibodies are now engineered into various formats to enhance their overall therapeutic potential. This is achieved by increasing specificity, improving efficacy, and reducing adverse effects [37]. Developments include antibody fragments, multispecific antibodies, antibody-drug conjugates, and antibody fusion proteins [38]. Each offers unique advantages in disease treatment. Multispecific antibodies are particularly popular in cancer treatment as they combine multiple paratopes on a single monomer, tackling complex signaling pathways [39]. Fig. 1.2 depicts various antibody therapeutics investigated in this study, including monoclonal IgGs, as well as bi- and tri-specific modalities. Bispecific interaction is achieved either through a Knob-in-Hole (KiH) construct or by appending single-chain fragment variables (scFv) to the Fc portion of the antibody [40,41]. KiH formats contain complementary mutations within the CH3 region to decrease the likelihood of mispairing, while scFv domains are constructed by fusing two Fv fragments using a flexible stretch of

synthetic peptide termed linker. A trispecific modality can be constructed by combining both techniques.

1.1.2 Preparative Chromatography

Preparative chromatography is a crucial component of biopharmaceutical DSP. It stands out for its selectivity, reversibility of association, and physiological processing conditions when compared to other separation methods [42]. Adsorption chromatography is typically used in preparative applications due to its superior loading capacity compared to other modes, such as size-exclusion chromatography (SEC). Impurity depletion is achieved through differential surface energies between the target molecules or impurities and the functionalized chromatographic resin [43,44]. In liquid-solid chromatography, a heterogeneous feed stream containing the target protein, impurities, and other solutes is loaded onto the adsorber matrix. The matrix usually consists of rigid but porous, and spherical adsorber particles packed tightly into a chromatographic bed within a cylindrical housing. This process is known as liquid column chromatography. The liquid feed stream is referred to as the mobile phase, while the solid adsorber particles are known as the stationary phase. The migration of solutes through the column matrix is influenced by fluid and thermodynamic effects, which will be discussed in the subsequent modeling Section 1.2.

Chromatographic purification steps are classified based on their adsorption mechanism and mode of operation. The utilization of different chromatographic modes provides an orthogonal separation utility within DSP [45]. The differential partitioning between the solutes in the mobile phase and the adsorbate in the stationary phase is controlled by the solution conditions of the mobile phase. Chromatographic purification typically follows a six-stage process. The stages are distinguished by their mobile phase conditions, which are regulated using parameters such as pH, conductivity, or buffer excipient concentration. The stages include equilibration of the resin matrix, column loading, washing to remove non-specifically bound product or contaminants, column elution to desorb specifically bound product or impurities, column regeneration to release strongly adsorbed contaminants, and finally, sanitization using a cleaning-in-place (CIP) step. The loading and elution steps differentiate the chromatographic operation modes. In bind-and-elute chromatography, the target molecule is strongly adsorbed to the resin and is desorbed either by a continuous eluent modification (gradient elution) or rapid exchange in eluent composition (step elution). If the column is loaded and eluted at isocratic conditions, meaning a constant buffer environment, the mode is termed flow through (FT) or weak partitioning (WP) operation. In these modes, the target product either flows freely through the column or exhibits a slight adhesion to the resin matrix.

Three to five chromatographic unit operations are typically connected within the antibody purification platform [9]. These are linked via multiple adjustment and filtration steps. Affinity chromatography (AC) serves as the primary capture step of the purification platform, removing the majority of process-related impurities as HCPs, DNA, endotoxins, and

cell culture media through a bind-and-elute operation [6]. AC is a highly selective separation method that interacts with specific motifs on the target protein surface during a multipoint adsorption process. In the case of mAb purification, a targeted adsorption to the Fc region can be achieved by using a Protein A ligand. The mAb desorption is then initiated by a reducing the mobile phase pH, which induces dissociation of the target protein from the Protein A ligand. Following the capture step, the crude drug substance is further purified using two to three chromatographic unit operations, known as polishing chromatography. During polishing chromatography, product-related impurities like size and charge variants of the main product, as well as remaining process-related impurities, are removed. In general ion-exchange and hydrophobic interactions are utilized as primary adsorption mechanism. Anion-exchange chromatography (AEX) operates by adsorbing negatively charged surfaces to a positive ligand group. AEX is often employed in FT mode and excels in the removal of negatively charged impurities like DNA. Conversely, cation-exchange chromatography (CEX) adsorbs positively charged surfaces and is mainly operated in bind-and-elute mode. A gradient or step elution then enables the separation of charge variants of the main product by increasing the ionic strength of the eluent. Hydrophobic interaction chromatography (HIC) adsorbs hydrophobic surfaces to aliphatic or aromatic ligand groups and is invaluable for removing size variants. HIC is mainly operated in bind-and-elute mode by decreasing the concentration of chaotropic modifiers, thereby reducing the strength of hydrophobic attraction. Additionally, multimodal chromatography has been introduced to the mAb purification platform. MMC combines orthogonal forms of physicochemical interaction, often linking electrostatic and hydrophobic groups within one multimodal ligand. This increases the selectivity compared to unimodal polishing methods and has the potential to integrate two or more polishing steps into a single unit operation [46]. MMC will be discussed further in Section 1.1.3.

The development of chromatographic unit operations involves identifying optimal purification conditions to maximize product purity and process economics, while satisfying strict regulatory requirements for product safety and efficacy. These requirements are enforced by government institutes such as the U.S. Food and Drug Administration (FDA) and the European Medicines Agency (EMA). Critical quality attributes (CQAs) of the drug substance are defined and must be met with minimal variability through the identification of critical process parameters (CPPs) [47,48]. Early process development involves selecting chromatographic resins, buffer conditions, and the protein load on the column to improve product purity, process yield, and robustness. Most of the development and regulatory efforts in DSP are dedicated to polishing chromatography. Process development, optimization, and characterization are primarily performed empirically using small-scale and high-throughput methods. Meanwhile, regulatory bodies are advocating for increased process understanding, as outlined in the QbD initiative [49]. QbD represents a shift towards science- and risk-based process development, in contrast to pure empiricism [50]. Mechanistic chromatography and structure-based modeling can support the QbD initiative and have the potential to supplement or replace wet lab experimentation. These topics will be introduced in Sections 1.2 and 1.3, respectively.

1.1.3 Multimodal Chromatography

Multimodal chromatography, also known as mixed-mode chromatography, is a versatile and selective purification technique. It involves multiple types of interactions between the stationary phase and solute, presenting both challenges and opportunities for process design and optimization [28,46]. Multimodal functionality can be achieved by connecting different functional groups or ligand moieties within a single chemical scaffold to form the multimodal ligand. Alternatively, unimodal ligands can be attached to one resin particle, or distinct resins can be mixed in one chromatographic bed [51]. This thesis will focus exclusively on multimodal ligands that are offering differing functionality in close proximity, as shown in Fig. 1.3.

Various forms of multimodal ligands exist, while novel modalities are constantly being developed [27,52]. Affinity chromatography, for instance, can be considered one form of multimodal functionality, often exhibiting a complex and highly specific ligand design. Hydroxyapatite (HT) is a natural adsorbent with mixed-mode functionality, achieved through positively charged calcium ions (C-site) and negatively charged phosphate groups (P-sites). The most common form of multimodal ligands combines hydrophobic and electrostatic functionality. Such a ligand attaches an aliphatic or aromatic moiety to its reactive side, serving the electrostatic functionality, and connects to the resin matrix via a spacer arm, as displayed in Fig. 1.3 (a). Often, an additional hydrogen-bonding group is attached to these ligands to increase binding strength, though it is considered secondary for selectivity purposes [27]. Either anionic or cationic functionality is achieved through strong (quaternary ammonium, sulfonic group) or weak (secondary amines, carboxylic groups) ion exchange moieties. Due to the multimodal binding behavior, the ligands can be operated in a HIC or an IEX predominant mode dependent on buffer pH and conductivity [53]. HIC mode is observed in an electrostatic repulsive regime with ligand and target protein exhibiting similar net charge or elevated conductivity levels, whereas as IEX mode is requiring oppositely charged binding partners and decreased buffer conductivity. A transition zone exists, in which both forms of interaction are strongly pronounced. This behavior is often referred to as "U-shaped binding curve" because of the complex pH and conductivity dependency of hydrophobic-electrostatic ligands [54]. A special case of multimodal functionality is presented by heterocyclic ring structures as pyridine or imidazole, which combine hydrophobic and ionic characteristic in a single group. Furthermore, their weak ion exchange characteristic and pKa values at physiological pH range enables the mechanism of hydrophobic charge induction chromatography (HCIC). In HCIC, binding occurs at a neutral pH through hydrophobic attraction with ion-exchange groups being uncharged. Elution is initiated by shifting pH, inducing similar charge of the ligand and the molecule surface leading to electrostatic repulsion.

MMC was initially introduced to mAb purification as a cost-effective alternative to Protein A capture. Multimodal cation exchange resins are particularly promising for this application, offering similar purification performance at a significantly lower cost than Protein A resins [55,56]. Additionally, they can operate under milder pH conditions, are chemically more stable showing no ligand leakage, display an improved capacity and provide broader

selectivity toward different classes of antibody-like structures [57]. However, MMC development is considered more complex and often necessitates resource intensive high throughput screening (HTS) and design of experiment (DoE) studies for identifying robust set point conditions [51,56,58]. By now, substantial improvements in Protein A ligand chemistry have relativized most of MMCs advantages for replacing this established technique except for their price difference. Nevertheless, the ability of MMC resins in engaging with multiple binding domains on the molecule surfaces still facilitates their utilization for the capture of nonstandard mAb formats [57,59].

There has been increasing interest in using MMC in polishing applications [60–66]. Its benefits include salt-tolerant adsorption capacity over classical ion-exchange chromatography (IEX) resins and improved impurity clearance within a single chromatographic processing step. Salt tolerant adsorption refers to the ability of binding molecules at elevated conductivity values based on complimentary hydrophobic effects, which enables more flexibility in DSP design [67]. Especially, anionic MMC showed synergistic effects to established mAb purification platforms to replace one or more unit operations within a single chromatography step. Classical AEX resins can be enhanced through multimodal functionality to assist the clearance of product related impurities and HCPs.

Arguably, one of the most widely adopted resins, is the multimodal strong anion exchanger Capto™ adhere [68], shown in Fig. 1.3. The N-benzyl-N-methylethanolamine ligand Capto adhere, combines a positively charged quaternary amine with an aromatic phenyl ring and a hydroxyl group.

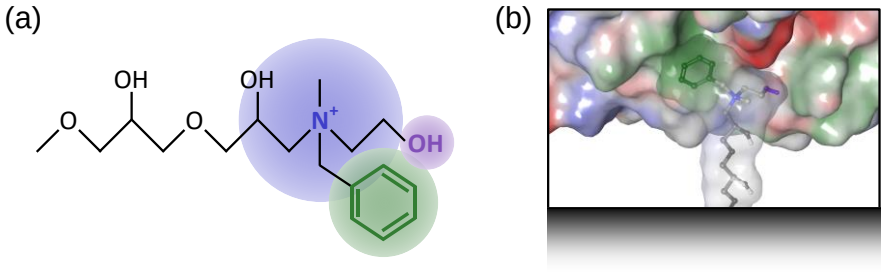

Fig. 1.3: Capto adhere ligand [68].(a) Structural formula of N-benzyl-N-methylethanolamine ligand Capto adhere. The positively charged quaternary amine colored in blue, enables anion exchange functionality, whereas the aromatic phenyl ring colored in green suffices hydrophobic interaction. The additional hydroxyl group, colored in purple assists the solute adsorption through hydrogen bonding. A spacer arm connects the functional groups to the resin matrix. The colored spheres surrounding the different modalities, symbolize the length scale of each interaction. (b) Schematic representation of Capto adhere adsorption to a significant binding domain on a protein surface. Green surface colors indicate hydrophobic interaction while blue and red colors display positive or negative electrostatic surface potentials.

Capto adhere has shown superior purification performance compared to established polishing resins and is predominantly employed in FT or WP operation [60–65]. It allows the simultaneous clearance of aggregates and fragments, HCPs, DNA, viruses, and leached Protein A and will be further investigated throughout this thesis.

Despite the apparent benefits of using MMC in DSP applications, its process development still hinders its widespread adoption. The complex and multimodal functionality can induce significant binding domains or hot spots on the protein surface as visualized in Fig. 1.3 (b). Accordingly, substantial research has been conducted in investigating multimodal interactions by using molecular dynamics (MD) simulations, fragmentation and labeling techniques, or quantitative structure-property relationship (QSPR) methods [69–72]. However, MMC development remains challenging and further research is needed to fully realize it's potential.

1.2 Modeling of Liquid Chromatography

Liquid chromatography modeling plays a crucial role in various applications such as *in silico* process optimization, scale-up, root-cause investigation, and robustness analysis [44,73]. These aspects can be valuable for product safety and process economics. The complexity of modeling chromatographic processes lies in its multiscale nature, encompassing various physical effects at different length scales. These include fluid dynamics, mass transfer phenomena, and adsorption thermodynamics. In the subsequent section, transport phenomena from a macroscopic to microscopic level will be presented. This exploration aims to simulate the mass transfer within the chromatographic column and its resin particles. Following this, the thermodynamics of chromatographic adsorption will be introduced, which describes the interaction between the molecule surface and the adsorber matrix at a molecular level.

1.2.1 Transport Models

Liquid column chromatography involves various transport phenomena, as depicted in Fig. 1.4. The chromatographic system, introduced in Section 1.1.2, is generally divided into a liquid mobile phase and a solid stationary phase.

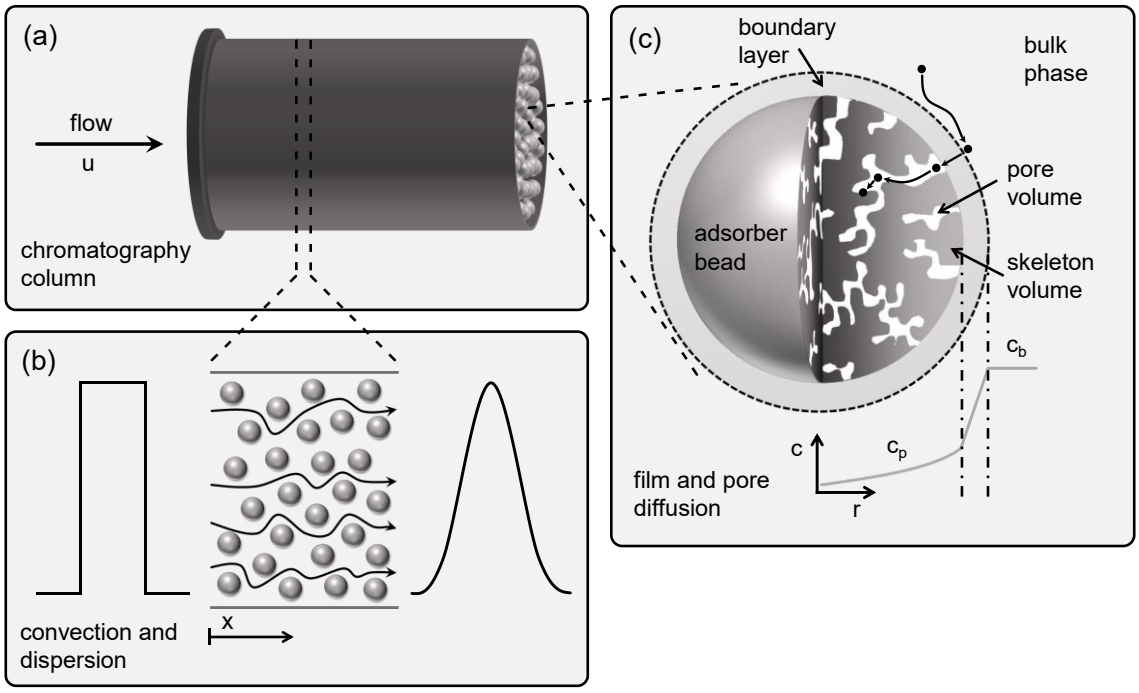

Fig. 1.4: Illustration of mass transfer phenomena in liquid column chromatography.(a) Introduction of load material into the chromatographic column with superficial velocity u. (b) Depiction of convective and dispersive effects within the resin bed's void volume, with axial dispersion causing band broadening effects along the longitudinal x axis. This transforms a rectangular residence time distribution into a Gaussian peak shape. (c) Representation of external and internal mass transfer phenomena in a resin particle, based on the linear driving force (LDF) model. The film mass transfer and pore diffusion establish a distinct concentration profile $c(r)$ for solute molecules, shown beneath the resin particle. The interstitial concentration c_b approaches the pore concentration c_p along the radial r axis. Figure adapted from Schmidt Traub [44].

In adsorption chromatography, a homogeneous mixture containing solute component $i \in [1, n]$ is applied to the column with length L_c. This mixture is pushed through the system in axial direction $x \in [0, L_\mathrm{c}]$, as depicted in Fig. 1.4 (a). The fluid enters the column at a superficial velocity u and is dispersed between the stationary resin particles of radius r_p. This dispersion increases the fluid's flow to an interstitial velocity u_int. Prior adsorption, the fluid penetrates the particles radially $r \in [0, r_\mathrm{p}]$, following a concentration gradient. The interparticle phase is referred to as the void or interstitial volume, whereas the intraparticle phase includes the liquid pore volume and the impermeable skeleton volume. The mobile phase acts as a solvent, dissolving and transporting the components with concentrations $c_{\mathrm{b},i}$ and $c_{\mathrm{p},i}$ in the interstitial and pore volumes, respectively. The phase distribution, distinguished by solute accessibility, is defined by interstitial porosity ε_b, particle porosity ε_p, and total column porosity ε_t. Mass transfer resistances and adsorption to the chromatographic resin selectively separate different components by altering their migration velocity through the column.

Convective transport refers to the average velocity of the bulk fluid through the chromatographic column. Fluid dynamics effects and mass transfer phenomena distort ideal plug-flow conditions, resulting in a Gaussian-like peak shape, as displayed in Fig. 1.4 (b). Band broadening of the idealized rectangular residence time distribution results from axial dispersion and external and internal mass transfer phenomena. Axial dispersion occurs in the void volume and results from turbulent eddy diffusion between the resin particles, molecular diffusion in the longitudinal direction, and uneven flow distribution caused by packing non-idealities and wall effects. These effects are influenced by column geometry, packing characteristics, interstitial velocity, and the viscosity of the mobile phase. External and internal mass transfer phenomena occur in the particle boundary layer and pore system, as shown in Fig. 1.4 (c). Upon entering the particle phase, the phenomena reduce to linear mass transfer and diffusive resistances, as they are excluded from convective transport. These are often described by the linear driving force (LDF) model. According to the LDF model, a stagnant film or laminar boundary layer surrounds the particles. Film mass transfer is induced by a linear concentration gradient from the bulk phase to the entrance of the particle pore system. Pore diffusion then transports the solutes through the tortuous pore system to the particle core in a radial direction. Pore diffusion is often considered the rate-limiting step of adsorption chromatography.

In liquid column chromatography, the system is generally considered axially and radially homogeneous. This implies that the chromatographic bed and its porosities are constant throughout the column. The resin particles are assumed to be of uniform size and experience no convective transport in the pore system. The column is generally considered isothermal, as the volumetric heat capacity of the fluid is large compared to the energy induced by drag effects or adsorption reactions. This assumption has important implications for the adsorption models presented in Section 1.2.2. It also leads to a homogeneous mobile phase of constant viscosity. The mobile phase is considered incompressible, inert, and having a constant linear velocity. All chromatographic mass transfer phenomena can thus be described in relation to time t and location x within the dynamic system.

To quantify the mass transfer phenomena, a differential mass balance equation of an infinitesimal volume segment in the axial direction is defined. For numerical calculation, the column is temporally and spatially discretized along the longitudinal axis and in the radial direction of the resin particle. Various modeling approaches exist, as detailed in the textbooks of Guiochon [43] and Schmidt Traub [44]. These approaches can be broadly categorized by their complexity or the number of band broadening effects they account for. This thesis will focus on a selection of common modeling frameworks, presented in the subsequent sections in order of decreasing model complexity.

1.2.1.1 General Rate Models

The General Rate Models (GRMs), which account for at least three band broadening effects, are the most comprehensive models for describing liquid column chromatography. These models consider not only dispersion but also both external and internal mass transfer

phenomena occurring within the void volume and the particle phase. Consequently, a continuity equation is necessary for each phase. Eq. (1.1) illustrates the continuous mass balance equation specifically for the column's void volume.

$$\frac{\partial c_{\mathrm{b},i}}{\partial t}(x,t) = -u_{\mathrm{int}}\frac{\partial c_{\mathrm{b},i}}{\partial x}(x,t) + D_{ax}\frac{\partial^2 c_{\mathrm{b},i}}{\partial x^2}(x,t) - \frac{1-\varepsilon_{\mathrm{b}}}{\varepsilon_{\mathrm{b}}}\frac{3}{r_{\mathrm{p}}}k_{\mathrm{film},i}\left(c_{\mathrm{b},i}(x,t) - c_{\mathrm{p},i}\left(x,r_{\mathrm{p}},t\right)\right) \tag{1.1}$$

The local temporal change of the component concentration in the void volume is affected by convective transport, axial dispersion, and film mass transfer through the stagnant boundary layer. Dispersive affects are lumped into the axial dispersion coefficient D_{ax}, which is independent from the type or concentration of the solute. The diffusivity through the stagnant film is determined by the film mass transfer coefficient, $k_{film,i}$, as per Fick's law of diffusion. Here, $c_{\mathrm{p},i}(r_{\mathrm{p}})$ represents the concentration at the particle surface. Both the axial dispersion coefficient and the film mass transfer coefficient are influenced by the velocity of the interstitial fluid between the resin particles. To complement the transport equation, Danckwerts' closed boundary condition is used for dispersive systems, as shown in Eq. (1.2) at the column inlet, along with the zero-gradient condition at the column outlet, as shown in Eq. (1.3).

$$\frac{\partial c_{\mathrm{b},i}}{\partial x}(x=0,t) = \frac{u_{\mathrm{int}}(t)}{D_{\mathrm{ax}}}\left(c_{\mathrm{b},i}(x=0,t) - c_{\mathrm{in},i}(t)\right) \tag{1.2}$$

$$\frac{\partial c_{\mathrm{b},i}}{\partial x}(x=L_{\mathrm{c}},t) = 0 \tag{1.3}$$

The component concentration at the column inlet is denoted as $c_{\mathrm{in},i}$. As per the studies conducted by Gu et al. [74,75], the radial mass transport within an idealized spherical particle can be explained with a differential mass balance. This balance is applicable for both the pore system and the stationary phase, as illustrated in Eq. (1.4).

$$\frac{\partial c_{\mathrm{p},i}}{\partial t}(x,r,t) = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 D_{\mathrm{pore},i}\frac{\partial c_{\mathrm{p},i}}{\partial r}(x,r,t)\right) - \frac{1-\varepsilon_{\mathrm{p}}}{\varepsilon_{\mathrm{p}}}\frac{\partial q_{\mathrm{i}}}{\partial t}(x,r,t) \tag{1.4}$$

The mobile phase within the pore system is assumed to be stagnant and the internal mass transport is governed by Fickian diffusion.

The pore diffusion coefficient of the i-th component $D_{\mathrm{pore},i}$, induces a concentration distribution along the particle radius. The local temporal change of the adsorbed component concentration q_i is then defined via adsorption models that will be introduced in Section 1.2.2. Analogous to the void phase, boundary conditions are required for applying the

GRM. Given the symmetry of the particle, it can be assumed that the loading gradient at the particle's center disappears, as expressed in Equation (1.5).

$$\frac{\partial c_{\mathrm{p},i}}{\partial r}(x, r = 0, t) = 0 \tag{1.5}$$

For connecting the continuity equations around the void volume and the pore phase the boundary condition is defined at the particle surface through Eq. (1.6).

$$\frac{\partial c_{\mathrm{p},i}}{\partial r}\left(x, r = r_{\mathrm{p}}, t\right) = \frac{k_{\mathrm{film},i}}{\varepsilon_{\mathrm{p}} D_{\mathrm{pore},i}}\left(c_{\mathrm{b},i}(x,t) - c_{\mathrm{p},i}\left(x, r = r_{\mathrm{p}}, t\right)\right) \tag{1.6}$$

1.2.1.2 Transport Dispersive Model

The Transport Dispersive Model (TDM) is primarily characterized by two rate-limiting parameters. This model simplifies the chromatography process by neglecting the concentration distributions within the particle phase, which is generally adequate for most practical applications. In the context of adsorption chromatography, rate limitation is predominantly due to pore diffusion. This allows for the effective combination of external and internal mass transfer. As a result, the rate limitation is modeled to exist within the liquid boundary layer surrounding the chromatographic particles, assuming infinite pore diffusion $\left(D_{\mathrm{pore},i} \to \infty\right)$. The mass balance equation for the void volume can then be represented by Eq. (1.7).

$$\begin{aligned}\frac{\partial c_{\mathrm{b},i}}{\partial t}(x,t) = &-u_{\mathrm{int}}\frac{\partial c_{\mathrm{b},i}}{\partial x}(x,t) + D_{\mathrm{ax}}\frac{\partial^2 c_{\mathrm{b},i}}{\partial x^2}(x,t) \\ &- \frac{1-\varepsilon_{\mathrm{b}}}{\varepsilon_{\mathrm{b}}}\frac{3}{r_{\mathrm{p}}}k_{\mathrm{eff},i}\left(c_{\mathrm{b},i}(x,t) - c_{\mathrm{p},i}(x,t)\right)\end{aligned} \tag{1.7}$$

The effective mass transfer coefficient of the i-th component $k_{\mathrm{eff},i}$ is introduced, replacing $k_{\mathrm{film},i}$ in Eq. (1.1). This coefficient is influenced by the velocity of the interstitial fluid. The continuity equation around the particle pore system is reduced to Eq. (1.8) according to the LDF model.

$$\frac{\partial c_{\mathrm{p},i}}{\partial t}(x,t) = \frac{3}{r_{\mathrm{p}}}\frac{k_{\mathrm{eff},i}}{\varepsilon_{\mathrm{p}}}\left(c_{\mathrm{b},i}(x,t) - c_{\mathrm{p},i}(x,t)\right) - \frac{1-\varepsilon_{\mathrm{p}}}{\varepsilon_{\mathrm{p}}}\frac{\partial q_i}{\partial t}(x,t) \tag{1.8}$$

1.2.1.3 Ideal Column Model

The ideal or equilibrium column model, introduced by Wicke [73] and DeVault [74], is a simple transport model that primarily accounts for convective transport and adsorption

thermodynamics. This model assumes a constant equilibrium between the mobile and stationary phases. It disregards axial dispersion ($D_{\mathrm{ax}} \to 0$), as well as external ($D_{\mathrm{film},i} \to \infty$) and internal ($D_{\mathrm{pore},i} \to \infty$) mass transfer resistances. As a result, the mass balance equations for the void and pore volumes are combined. This is achieved by introducing the total column porosity $\varepsilon_{\mathrm{t}} = \varepsilon_{\mathrm{b}} + (1 - \varepsilon_{\mathrm{b}})\varepsilon_{\mathrm{p}}$, and the superficial velocity $u = u_{\mathrm{int}}\varepsilon_{\mathrm{b}}$, leading to Eq. (1.9).

$$\frac{\partial c_i}{\partial t}(x,t) = -\frac{u}{\varepsilon_{\mathrm{t}}}\frac{\partial c_i}{\partial x}(x,t) - \frac{1-\varepsilon_{\mathrm{t}}}{\varepsilon_{\mathrm{t}}}\frac{\partial q_i}{\partial t}(x,t) \tag{1.9}$$

The local temporal change of the mobile phase concentration of the i-th component c_i is defined. In this model, deviations of component migration are primarily attributed to variations in the adsorption rate. The straightforward nature of this approach facilitates the exploration of fundamental phenomena in chromatography. For instance, by rearranging Eq. (1.9), an expression for the migration of concentration fronts through chromatographic columns can be derived in Eq. (1.10).

$$w_{\mathrm{p}}(c_i^+) = \frac{u}{\varepsilon_{\mathrm{t}}}\left(1 + \frac{1-\varepsilon_{\mathrm{t}}}{\varepsilon_{\mathrm{t}}}\frac{\partial q_i}{\partial c_i}\bigg|_{c_i = c_i^+}\right)^{-1} \tag{1.10}$$

The velocity of propagation, w_{p}, for any given component concentration, represented as c_i^+, is dependent on its local isotherm slope $\frac{\partial q_i}{\partial c_i}\big|_{c_i = c_i^+}$. When incorporating the retention time $t_i^{\mathrm{R}} = L_{\mathrm{c}}/w_{\mathrm{p}}(c_i^+)$ of component i and assuming a linear adsorption regime ($q_i \to 0$) similar to what is observed in analytical chromatography, Eq. (1.11) is defined.

$$t_i^{\mathrm{R}} = t_0\left(1 + \frac{1-\varepsilon_{\mathrm{t}}}{\varepsilon_{\mathrm{t}}}K_{\mathrm{H},i}\right) \tag{1.11}$$

Eq. (1.11) is termed the basic equation of chromatography and can serve as powerful shortcut for preliminary process design [43,44]. This equation is contingent on the column dead time t_0, which is the retention time under unretained conditions. Additionally, it introduces the Henry coefficient $K_{\mathrm{H},i}$ for the i-th component, which is equivalent to the initial slope of the isotherm in a linear adsorption regime. The application of Eq. (1.11) facilitates an accelerated determination of adsorption parameters. This can be achieved using peak maximum methods during isocratic or gradient elution, as demonstrated by Yamamoto et al. [76].

1.2.2 Adsorption Models

The mass accumulation of component $i \in [1, n]$ onto the adsorber surface is described via adsorption models. Adsorption isotherms depict the functional relationship of the component concentrations in liquid chromatography at equilibrium and constant temperature. They relate the liquid concentration of the i-th solute c_i to adsorbed component concentration per adsorber skeleton volume q_i, as visualized in Fig. 1.5 (a). Adsorption isotherms can be vaguely distinguished into convex or concave functional forms. Convex isotherms are most common in adsorption chromatography and indicate competitive binding of solute components to a finite number of binding partners like the one presented in Fig. 1.5 (a). Conversely, concave isotherms can be observed if adsorbed proteins act as binding partners for solute components. The curvature of adsorption isotherms strongly affects the chromatogram and leads to deviations from an ideal peak shape. Peak-sharpening effects occur and include shock fronts and diffuse peak tails given a convex isotherm shape, *vice versa*. The isotherm operates in a linear and nonlinear adsorption regime, which refers to the local slope of the isotherm function. The linear part of the isotherm ($q_i \ll c_i$) is independent of the component concentration and equals its Henry coefficient $K_{\mathrm{H},i} = \lim_{q_i \to 0} \partial q_i / \partial c_i$. The nonlinear part of the isotherm is most relevant in overloaded conditions, which is often observed in preparative chromatography.

The adsorption models presented in this thesis are based on the thermodynamic framework of Mollerup [78,79]. The models assume a reversible association of solute proteins to immobilized ligands given a constant protein conformation. The phase change of a protein component from solute to adsorbed state introduces a new component, the protein-ligand complex. The association-dissociation process is formulated as a stoichiometric reaction equation introduced by Boardman and Partridge [80] and Regnier et al. [81]. The model assumes a chemical equilibrium at isothermal and isobaric conditions. The equilibrium is defined as the global minimum of Gibbs energy indicating equality of adsorption and desorption rates. The isotherm framework accounts for dissociation reactions of modulators or solvents from chromatographic ligands, as depicted in Fig. 1.5 (b). The reaction equation can be corrected by asymmetric excess potentials, which are calculable from excess Gibbs energy models, or state function-based models. The saturation capacity of the chromatographic resins is modeled by using a material balance for the ligand availability according to the stoichiometric displacement model (SMA) introduced by Brooks and Cramer [77]. Empirical pH extensions often complement the final isotherm equations. In the following section a common framework for describing MMC adsorption is derived. The MMC isotherm is than used as a template for introducing IEX and HIC models through model reduction.

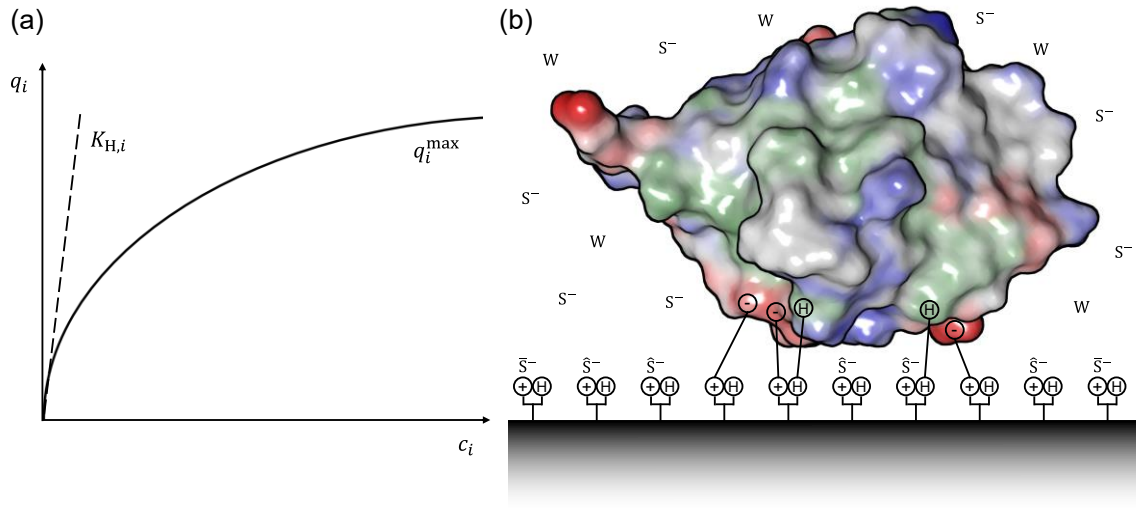

Fig. 1.5: Protein adsorption to multimodal anion exchanger (PDB entry 1UBQ).(a) Convex adsorption isotherm for solute component i. The isotherm defines the relationship between bound and unbound component concentration q_i and c_i. The initial isotherm slope equals the Henry coefficient $K_{H,i}$, whereas the saturation capacity of the chromatography resin is given by q_i^{max}. (b) Illustration of protein-ligand complex after association of positive and hydrophobic ligand groups to negative and hydrophobic ligand binding sites on the protein surface derived from the SMA model [77]. The protein is dissolved in an aqueous solution containing negatively charged counterions S^- and water molecules W. Unbound ligand groups are associated to counterions $\bar{S}^-$. Some ligand groups are sterically hindered due to protein topography and are associated to counterions $\hat{S}^-$ that cannot be displaced by other proteins. Figure adapted from Brooks and Cramer [77].

1.2.2.1 Multimodal Chromatography

To describe the multimodal adsorption process, Nfor et al. have devised a stoichiometric model to account for electrostatic and hydrophobic physicochemical interactions [54]. According to Nfor et al., the multimodal association of reactants is defined as an (informal) chemical equilibrium, shown in Eq. (1.12).

$$\mathrm{P}_i^{z_{\mathrm{p},i}} + \nu_{i,j}\mathrm{S}_j^{z_{\mathrm{s},j}}\mathrm{L}_{z_{\mathrm{s},j}} + n_i\mathrm{L} \rightleftharpoons \mathrm{P}_i^{z_{\mathrm{p},i}}\mathrm{L}_{z_{\mathrm{p},i}n_i} + \nu_{i,j}\mathrm{S}_j^{z_{\mathrm{s},j}} \tag{1.12}$$

Protein P_i reversibly adsorbs to the ligand binding site L, leading to the formation of a protein-ligand complex PL, as depicted in Fig. 1.5 (b). This interaction process involves the stoichiometric displacement of counterion S_j, $j \in [1, m]$, from the ligand's charged functional group. The stoichiometric coefficient of this electrostatic interaction $\nu_{i,j} = z_{\mathrm{p},i}/z_{\mathrm{s},j}$ is termed characteristic charge of the i-th component. Here, $z_{\mathrm{p},i}$ and $z_{\mathrm{s},j}$, represent the effective binding charges of the protein and the counterion, respectively. In addition, the ligand's hydrophobic moiety can reversibly associate with non-polar binding sites on the

protein surface. The stoichiometric coefficient of this hydrophobic interaction is represented by n_i. Eq. (1.12) can be simplified by assuming that the counterion and the ligand binding site are monovalent and uniform. Thus, the solute protein $\mathrm{P}_i^{z_{\mathrm{p},i}}$ and the adsorbed protein ligand complex $\mathrm{P}_i^{z_{\mathrm{p},i}}\mathrm{L}_{z_{\mathrm{p},i}n_i}$ are substituted by $\mathrm{P}_{\mathrm{sol},i}$ and $\mathrm{P}_{\mathrm{ads},i}$, as shown in Eq. (1.13).

$$P_{\mathrm{sol},i} + \nu_i(LS) + n_i L \rightleftharpoons P_{\mathrm{ads},i} + \nu_i S \tag{1.13}$$

Under the principle of the law of mass action, the specific equilibrium constant $K_{\mathrm{eq},i}$ for the protein and buffer system can be derived. This derivation occurs at the thermodynamic equilibrium of the adsorption and desorption reaction. Parameterizing the chemical activities with their respective molar concentrations and activity coefficients γ, leads to Eq. (1.14).

$$K_{\mathrm{eq},i} = \frac{k_{\mathrm{ads},i}}{k_{\mathrm{des},i}} = \left(\frac{q_i \gamma_{\mathrm{P}_{\mathrm{ads},i}}}{c_{\mathrm{p},i}\gamma_{\mathrm{P}_{\mathrm{sol},i}}}\right)\left(\frac{c_{\mathrm{s}}\gamma_{\mathrm{S}}}{\hat{q}_{\mathrm{s}}\gamma_{\mathrm{LS}}}\right)^{\nu_i}\left(\frac{c_v}{\hat{c}_{\mathrm{l}}\gamma_{\mathrm{L}}}\right)^{n_i} \tag{1.14}$$

The equilibrium constant is defined as the ratio of the adsorption rate $k_{\mathrm{ads},i}$ to the desorption rate $k_{\mathrm{des},i}$ of the protein species from the ligand binding site. Here, the subscripts s and l represent the counterion and the ligand species. It is important to note that the ionic strength of the solution may not align with the counterion concentration, especially when dealing with a multivalent modulator. In such cases, the ionic strength should be considered instead. The activity coefficients, which express the excess Gibbs free energy of the thermodynamic association, are used to account for deviations from ideal chemical behavior. Furthermore, the molar density of the liquid phase c_v is influenced by the modulator concentration. Assuming unity of all activity coefficients $\gamma_{\mathrm{P}_{\mathrm{ads},i}} = \gamma_{\mathrm{S}} = \gamma_{\mathrm{LS}} = \gamma_{\mathrm{L}} = 1$ except for the solute protein $\gamma_{\mathrm{P}_{\mathrm{sol},i}}$, leads to Eq. (1.15).

$$K_{\mathrm{eq},i} = \left(\frac{q_i}{c_i}\right)\left(\frac{c_{\mathrm{s}}}{\hat{q}_{\mathrm{s}}}\right)^{\nu_i}\left(\frac{c_v}{\hat{c}_{\mathrm{l}}}\right)^{n_i}\left(\frac{1}{\gamma_{\mathrm{P}_{\mathrm{sol},i}}}\right) \tag{1.15}$$

The concentration of freely available ligands for the electrostatic and the hydrophobic interaction $\hat{q}_{\mathrm{s}}$ and $\hat{c}_{\mathrm{l}}$ can be effectively derived using the SMA formalism [80], as illustrated in Eq. (1.16) and (1.17).

$$\hat{q}_{\mathrm{s}} = \Lambda_{\mathrm{IEX}} - \sum_{i=1}^{m}(\nu_i + \sigma_i)q_i \tag{1.16}$$

$$\hat{c}_{\mathrm{l}} = \Lambda_{\mathrm{HIC}} - \sum_{i=1}^{m}(n_i + s_i)q_i \tag{1.17}$$

The ionic capacity Λ_{IEX} and the hydrophobic ligand density and Λ_{HIC} per adsorber skeleton volume limit the molar concentration of available ligands. During the adsorption process, the concentration of free ligands on the resin surface decreases, corresponding to the amount of the bound component q_i. Each protein adheres to the resin in a stoichiometric manner, while additional free ligand binding sites are obstructed due to steric hindrance. This steric hindrance, affecting both electrostatic and hydrophobic adsorption mechanisms, is quantified by the parameters σ_i and s_i, respectively. For convenience, the activity coefficient of the solute protein is normalized to be unity in pure water at infinite dilution $\gamma_{\mathrm{P_{sol},}i}^{\infty,w} = \lim_{c_{\mathrm{p},i}\to 0} \gamma_{\mathrm{P_{sol},}i}$, introducing the asymmetric activity coefficient $\tilde{\gamma}_{\mathrm{P_{sol},}i} = \gamma_{\mathrm{P_{sol},}i}/\gamma_{\mathrm{P_{sol},}i}^{\infty,w}$. Additionally, the asymmetric activity coefficient can be parameterized according to the van der Waals equation of state, introducing Eq. (1.18).

$$\tilde{\gamma}_{\mathrm{P_{sol},}i} = \exp\left(K_{\mathrm{s},i} c_{\mathrm{s}} + K_{\mathrm{p},i} c_i\right) \tag{1.18}$$

The parameters $K_{\mathrm{s},i}$ and $K_{\mathrm{p},i}$ play a crucial role in the activity coefficient model. They provide a detailed description of the interactions between proteins and modulators, specifically focusing on the salt-protein and protein-protein interactions. The salt-protein interaction parameter, $K_{\mathrm{s},i}$, is defined as $K_{\mathrm{s},i} = 2(a_{12} - a_{32})/RT$, following the van der Waals parameterization method used by Mollerup [82]. Similarly, the protein-protein interaction parameter, $K_{\mathrm{p},i}$, is defined as $K_{\mathrm{p},i} = 2(a_{12} - a_{22})/RT$. Here, the parameter a serves as an interaction constant, providing a measure of the average attraction caused by the intermolecular forces between the components of the system. The subscripts 1, 2, and 3 represent the aqueous solvent, the protein, and the salt, respectively. Therefore, a_{12}, a_{22}, a_{32} represent the interaction constants for water-protein, protein-protein, and salt-protein interactions. The molar gas constant is denoted by R and the absolute temperature by T.

Rearranging and inserting Eq. (1.16), (1.17), and (1.18) into Eq. (1.15), while changing the definition of the equilibrium constant to $\widetilde{K}_{\mathrm{eq},i} = K_{\mathrm{eq},i}\gamma_{\mathrm{P_{sol},}i}^{\infty,w}$, the multicomponent mixed-mode isotherm at equilibrium condition is derived in Eq. (1.19).

$$\begin{aligned} K_i(c_s) &= \frac{q_i}{c_i} \\ &= \widetilde{K}_{\mathrm{eq},i}\left(\frac{\Lambda_{\mathrm{IEX}} - \sum_{i=1}^{m}(\nu_i + \sigma_i)q_i}{c_{\mathrm{s}}}\right)^{\nu_i}\left(\frac{\Lambda_{\mathrm{HIC}} - \sum_{i=1}^{m}(n_i + s_i)q_i}{c_v}\right)^{n_i}\exp\left(K_{\mathrm{s},i}c_{\mathrm{s}}\right. \\ &\left. + K_{\mathrm{p},i}c_i\right) \end{aligned} \tag{1.19}$$

The equilibrium ratio $K_i(c_{\mathrm{s}})$ denotes the ratio between the adsorbed protein species to the solute protein species in the column. At non-equilibrium conditions the mixed-mode reaction can be solved after the rate of change of the adsorbed protein species $\partial q_i/\partial t\,(x,t)$, which is often utilized in chromatographic simulations and termed kinetic formulation [83]. Extensions to this isotherm model exist and include the release of bulk-like water molecules during protein adsorption, alternative HIC mechanisms, or empirical pH dependencies for the parameters $K_{\mathrm{eq},i}$, ν_i, and n_i [84–86].

1.2.2.2 Ion-exchange Chromatography

The SMA model is often used to describe pure ion-exchange mechanisms. This isotherm is based on the principle of stoichiometric displacement and steric hindrance [77]. This model can be derived from the multimodal isotherm, as detailed in Eq. (1.19), by neglecting hydrophobic interactions ($n_i \to 0$) and thermodynamic excess potentials of the salt-protein ($K_{\mathrm{s},i} \to 0$) and protein-protein ($K_{\mathrm{p},i} \to 0$) interaction, leading to Eq. (1.20).

$$\frac{q_i}{c_i} = \widetilde{K}_{\mathrm{eq},i}\left(\frac{\Lambda_{\mathrm{IEX}} - \sum_{i=1}^{m}(\nu_i + \sigma_i)q_i}{c_{\mathrm{s}}}\right)^{\nu_i} \tag{1.20}$$

1.2.2.3 Hydrophobic Interaction Chromatography

Analogous to the IEX mechanism, Mollerup has introduced a stoichiometric model for describing hydrophobic adsorption [82]. As hydrophobic ligand density cannot be derived easily through experimentation, an expression for maximum adsorption capacity $q_i^{\max}$ can be utilized, defined in Eq. (1.21).

$$q_i^{\max} = \frac{\Lambda_{\mathrm{HIC}}}{n_i + s_i} \tag{1.21}$$

Unlike IEX models, thermodynamic excess potentials play a crucial role in HIC modeling. The hydrophobic isotherm for multiple components is also derived from the multimodal isotherm, shown in Eq. (1.19). Accordingly, electrostatic interactions are neglected ($\nu_i \to 0$) while Eq. (1.21) is inserted into Eq. (1.19), resulting in Eq. (1.22).

$$\frac{q_i}{c_i} = \widetilde{K}_{\mathrm{eq},i}\left(\frac{\Lambda_{\mathrm{HIC}}}{c_v}\right)^{n_i}\left(1 - \frac{\sum_{i=1}^{m} q_i}{q_i^{\max}}\right)^{n_i}\exp\left(K_{\mathrm{s},i}c_{\mathrm{s}} + K_{\mathrm{p},i}c_i\right) \tag{1.22}$$

1.3 Quantification of Protein Structure-Function Relation

Proteins can be viewed as molecular building blocks and tools. They exhibit an inherent relationship between their structure and function, each defining the other. The premise of quantitative structure-activity/-property relationship (QSAR/QSPR) states that similar molecular structures exhibit comparable behavior, a concept known as the similarity-property principle (SPP) [87]. The differentiation of QSAR and QSPR is based on the type of property being modeled, either biological or physicochemical [88]. SPP than enables the prediction of molecular property from structural characteristics, or the *de novo* design of molecular structures based on functional requirements [89,90]. These predictions are widely used in fields such as material science and medicinal chemistry, particularly for *in silico*

screening of novel materials or drug discovery aimed at improving stability, efficacy, or toxicity profiles [87,91]. Other applications include predicting developability attributes or process conditions from existing compounds for biotherapeutic development [22,92,93].

QSPR modeling is a strictly correlative analysis used when underlying structure-function relations are not well understood. It is particularly useful for complex interrelations where sufficiently large data sets are available but mechanistic models are not. Central to the QSPR approach is the identification of a suitable mapping function $y(X)$ to predict a property or dependent variable vector y from physicochemical protein characteristics or an independent variable matrix X, which will be discussed in section 1.3.3. The QSPR modeling process involves data acquisition, data selection, model selection and calibration, and model validation, as shown in Fig. 1.6. During data acquisition, property information is gathered and physicochemical characteristics, or descriptors, are extracted from protein structures. The data set is then curated to correct for structural and experimental errors. Data selection involves partitioning the data set into subsets used for model calibration and internal and external model validation. Model selection and calibration is an iterative process where different data treatments and model structures are compared to reduce the dimensionality of the multivariate descriptor space and improve the model's self-consistency. It involves automated calibration of the mapping functions, a process often referred to as machine learning [94].

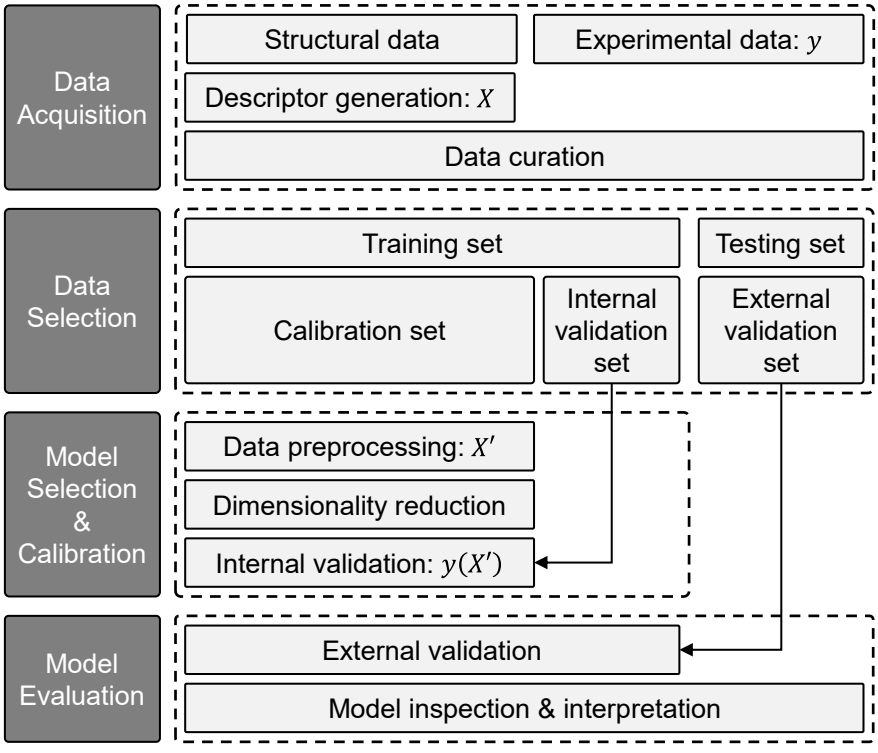

Fig. 1.6: Generalized QSAR/QSPR modeling workflow. The workflow is a multistep process and includes data acquisition, followed by data selection, model selection and calibration, and lastly the model evaluation.

The calibrated empirical model is then evaluated by predicting properties for protein structures not involved during model training. Furthermore, mechanistic interpretation of the correlative model is aspired to increase the understanding of the underlying structure-function relation [95].

Several challenges arise during QSPR modeling, primarily concerning empirical model validation [96,97]. The training and testing process is closely tied to the selection of representative training and testing data. This is particularly challenging for small data set sizes as each structure can significantly contribute to the structural diversity of the entire data set. The goal for validating QSPR models should be the definition of an applicability domain (AD), which is a region of model reliability where interpolation is feasible within a chemical or structural space [98]. Predictions outside the model's AD are considered unreliable as this would require extrapolation of the empirical model.

Defining the AD becomes demanding when working in a multivariate descriptor space and observing nonlinear descriptor interrelations, leading to increased degrees of freedom for calibrating the mapping function. Small ADs, restricted to highly homologous structures, can result in model overfitting or overdetermination. Simply speaking, overfitting is learning a mapping function too well. Consequently, phenomenologically unrelated descriptors or overestimated sensitivity towards related descriptors might be incorporated into model predictions, which would otherwise classify as noise. Therefore, mechanistic interpretation of the mapped structure-property relation is crucial in evaluating QSPR models. Conversely, large ADs, incorporating highly heterogeneous structures, could lead to discontinuities in the model predictions. This phenomenon, often referred to as activity cliff, is a threshold at which small structural variations lead to large deviations in molecule behavior. If picturing a structure-property map, a method for representing structural similarity against property similarity derived from a model response surface, activity cliffs identify as regions where the structure-property relationships break down. This effect can originate from inadequate descriptor representation or model structure and indicates missing information in deriving the true structure-property relation. Activity cliffs are difficult to identify as it is often unclear what classifies as an activity cliff or outlier molecule that introduces erroneous bias to the model. Ideally, the entire data set should be included while the QSPR approach should provide the capacity to correctly assess prediction confidence.

These challenges in empirical model validation illustrate why the QSPR approach should not be mistaken for being mechanistically grounded despite using structural information. QSPR models operate within defined boundaries and are incapable of extrapolation where descriptor information is merely a vector symbol to distinguish molecules within its AD [99]. As correlation should not be confused with causation, QSPR models can only provide evidence for deriving true mechanistic understanding. The subsequent sections will give a brief excurse into important prerequisites for developing reliable QSPR models. Initially, the acquisition of high-quality structural data is presented in Section 1.3.1. Thereafter, the concept of physicochemical descriptors is introduced in Section 1.3.2. Lastly, fundamentals of statistical modeling and machine learning are depicted in Section 1.3.3.

1.3.1 Protein Structure Acquisition

The development of reliable QSPR models depends on the availability of high-quality protein structures. These structures provide the basis for calculating physicochemical characteristics. The aim is to minimize random structural errors and promote more systematic deviations. For instance, post-translational modifications in structural preparation might be disregarded. This can reduce error propagation and enhance signal-to-noise ratio for subsequent statistical modelling. Protein structures can be obtained either through experimental means or predicted using *in silico* models. However, a comprehensive understanding of protein arrangement and the protein folding process is crucial for evaluating their structural characteristics.

Proteins are made up of amino acids linked by peptide bonds, which fold into specific spatial conformations to execute their biological functions. Protein structures can be classified based on their structural similarity, topological class, or shared evolutionary heritage. The three-dimensional arrangement of the polypeptide chain is referred to as a protein fold. Protein structures are divided into four distinct levels, the primary, secondary, tertiary, and quaternary structure [100]. The primary structure is the linear sequence of amino acids in the polypeptide chain, representing the denatured protein state. Secondary structures are highly regular local sub-structures of the polypeptide backbone, including α-helices, β-sheets and β-turns, and loops. The tertiary structure is the three-dimensional structure of a single protein molecule, folding in a way that the hydrophilic sides face the surrounding environment, and the hydrophobic sides face the protein's core. The quaternary structure comprises two or more individual polypeptide chains, forming structures like mAbs, as shown in Fig. 1.1.

Protein folding refers to the process by which a protein chain transforms from an unstable, random coil into a three-dimensional structure. It is guided by hydrophobic interactions, intramolecular hydrogen bond formation, van der Waals forces, and disulfide bond formation, and is counteracted by conformational entropy. This process can be conceptually depicted as an energetic funnel. During folding, a series of meta-stable intermediate structures are formed, eventually reaching an entropically favorable conformation, the protein's native fold. However, the native structure is not static; it populates an ensemble of conformational states, a concept known as conformational polymorphism [101]. The prevalence of any specific conformation follows a Boltzmann distribution and depends on the protein's current environment.

Experimental determination of protein structure is primarily achieved through X-ray crystallography, which resolves the structure by measuring the electron density distribution derived from X-ray diffraction patterns. These patterns can be captured by irradiating the protein in its crystalline state. Other methods include nuclear magnetic resonance (NMR) spectroscopy and cryogenic electron microscopy (cryoEM). The resolved structures are then deposited in protein structure databases, such as the freely accessible Protein Data Bank (PDB). However, experimental structure determination is often expensive, complex, and time-consuming. This is due to exhaustive sample preparation, measurement

sensitivity, and elaborate data evaluation. This leads to a significant discrepancy between the number of proteins with known primary structure and number of proteins with deposited higher order structures.

Computational techniques offer an alternative to experimental structure determination. Various methods have been developed to infer the three-dimensional structure of proteins from their amino acid sequence. These methods can be broadly categorized into ground-up based approaches, which combine physics simulation with structural statistics, or template-based techniques [100]. *De novo* or *ab initio* protein modeling predicts the three-dimensional protein structure based on physical principles using force field equations. Due to the complexity of the protein folding process, which involves assisting proteins like molecular chaperones, these tools are limited to predicting secondary structures, amino acid side chain conformation, and protein packing. For predicting tertiary and quaternary structures, template-based approaches are preferred. Methods like threading or comparative modeling utilize previously solved structures as a template for sequence alignment. This is feasible due to the limited number of tertiary structural motifs but challenging for quaternary structure prediction due to the increased number of possible conformational combinations.

Antibody homology modeling, a special case of quaternary structure prediction using a template-based approach, is practical due to the abundance of well-characterized antibody structures. The antibody structure is evolutionarily conserved to ensure proper functionality, which limits the number of tolerated mutations. Furthermore, antibody-specific libraries exist for categorizing antibody CDR loops into structural classes. Canonical classes distinguish CDR conformations according to their length and key residues. The homology workflow involves template selection, template-target sequence alignment, model building (which includes CDR and side chain modeling), model optimization, and model validation [102]. During the final model validation, structural inconsistencies are inspected, such as steric clashes or the likelihood of the predicted backbone and side-chain conformations compared to known structures.

The progress in predicting protein structures is consistently tracked and evaluated through competitions like the biannual "Critical Assessment of Techniques for Protein Structure Prediction" (CASP) experiment [103]. Recently, AI-based methodologies like AlphaFold have made significant advancements in protein structure prediction [104]. Tools like the AlphaFold Protein Structure Database help bridge the gap between proteins of known sequence and their structure [105]. While current AI-based techniques are limited to tertiary structure prediction, future improvements in the prediction of protein quaternary structure are anticipated.

1.3.2 Physicochemical Property Description

Molecular descriptors are quantifiable features that extract and encode the physicochemical properties of chemical or biological entities. They simplify and classify molecule characteristics into scalar values, vectors, or n-dimensional matrices through several levels of abstraction. The type of structural information needed varies case by case and requires the integration of domain and application knowledge [106].

Descriptors are differentiated based on the level of complexity or information content they account for, as depicted in Fig. 1.7 [90]. For instance, 0D descriptors can be derived by simply adding elements in a list of atoms within a chemical formula. 1D descriptors than account for substructures within this molecule like amino acid residues of the protein primary structure. 2D descriptors consider the connectivity of these elements, such as chemical structural graphs or molecule topology. 3D descriptors are derived from molecule conformation and can be extended to 4D descriptors when considering the time dependency during quantum- or molecular dynamics (QM/DM) simulations or additional forms of information [88,91,94]. This classification can become blurred as the dimensionality of input formats is often modified. Exemplary, MD time-series data can be averaged to a single ensemble structure, or three-dimensional surface characteristics can be projected onto a two-dimensional surface plane.

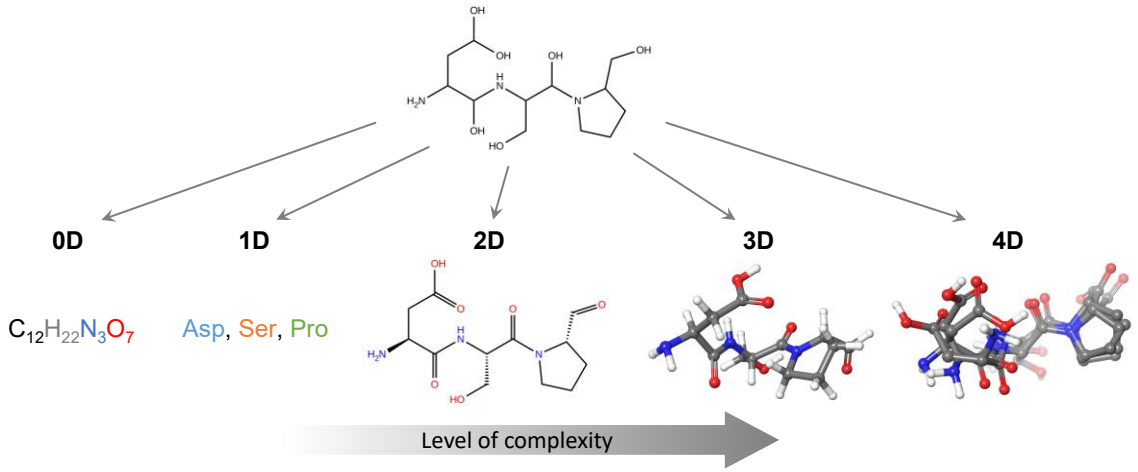

Fig. 1.7: Schematic of descriptor classes for exemplary tripeptide. The chemical graph of an aspartic acid-serine-proline (DSP) tripeptide is depicted at the figure top. An increasing amount of information can be extracted from the molecule, which is classified into 0D to 4D descriptors. Figure adapted from Grisoni et al. [88].

The physicochemical properties used to calculate descriptors can either be derived empirically based on experimental measurements or from first principles models. Empirical descriptors include bioinformatics scales that are assigned to individual residues or are mapped onto a molecular surface to account for surface accessibility. For instance, current residue hydrophobicity scales are derived from experimental studies of amino acid solubility or retention time during analytical HIC or reversed-phase liquid chromatography (RP-LC) experiments. Similarly, physics-based descriptors, which include various field equations, can be calculated for chemical substructures or entire molecules.

A notable example of a physics-based descriptor that is used in this study is the three-dimensional surface charge distribution of biomolecules. It can be calculated by using the Poisson-Boltzmann (PB) model [107]. Fig. 1.8 depicts the charge distribution of proteins that is attributed to surface exposed amino acids and affected by solvent pH and ion composition. Partial surface charges cause an electric field, which leads to the formation of a double layer of counterions to maintain electroneutrality. The PB model than assumes an implicit solvent that is affected by thermal motion to form a diffuse double layer [108]. The three-dimensional distribution of electrostatic potential within this layer can be described according to Poisson, shown in Eq. (1.23).

$$\nabla^2\psi = \frac{\partial^2\psi}{\partial x^2} + \frac{\partial^2\psi}{\partial y^2} + \frac{\partial^2\psi}{\partial z^2} = -\frac{\rho_\mathrm{e}}{\varepsilon_\mathrm{r}\,\varepsilon_0} \tag{1.23}$$

Using the 3D Laplace operator ∇^2, the second order partial-derivative of the electric potential ψ is described in a cartesian coordinate system. It is dependent on the charge density ρ_e as well as the relative permittivity ε_r and vacuum permittivity ε_0, respectively. The spatial distribution of free moving ions within the solution can be accounted for by Boltzmann statistics. An expression for the local charge density is derived when only considering electrostatic interactions and neglecting the finite ion size, as shown in Eq. (1.24).

$$\rho_e = eN_\mathrm{A}\sum z_i c_{\infty,i} e^{-\frac{z_i e\psi}{k_\mathrm{b}T}} \tag{1.24}$$

According to the Boltzmann expression, the charge density is affected by the electric work $e\psi$ required for moving ion i with charge z_i within the diffuse double layer. This leads to an accumulation of counter-ions and a depletion of co-ions close to the dielectric boundary. Furthermore, the expression includes the elementary charge e, the Avogadro constant N_A, the Boltzmann constant k_B, and the absolute temperature T, as well as the bulk concentration of ions $c_{\infty,i}$, whereby definition the electrostatic potential vanishes $(\psi = 0)$. Inserting Eq. (1.24) into Eq. (1.23) leads to the nonlinear Poisson–Boltzmann equation, shown in Eq. (1.25).

$$\nabla^2\psi = -\frac{eN_\mathrm{A}}{\varepsilon_\mathrm{r}\,\varepsilon_0}\sum z_i c_{\infty,i} e^{-\frac{z_i e\psi}{k_\mathrm{b}T}} \tag{1.25}$$

While for simple geometries analytical solution to the PB equation exist, calculation for complex biomolecules requires numerical approximations in a spatially discretized coordinate grid. These can include finite element or finite difference methods as implemented by tools like the Adaptive Poisson-Boltzmann Solver (APBS) [109].

Protein-specific descriptors extend classical molecule descriptors to account for their increased conformational variability, surface anisotropy, and overall size and shape [106,110]. Due to variations in surface charge, hydrophobicity, or topology distribution, visualized in Fig. 1.3 (b), not all protein regions experience the same physicochemical interaction with the surrounding environment. To quantify this heterogeneity, areas on the surface with consistent electrostatic or hydrophobic properties can be defined as surface patches. Descriptor sets often include binning or gating strategies, where individual properties are segregated according to their size, intensity, or frequency (e.g., the 10th largest percentile of positively charged patches). To further refine protein descriptors, a distinction into global and local descriptors can be introduced. Global descriptors are calculated from the entire protein structure, whereas local descriptors decompose the protein into specific subregions [110].

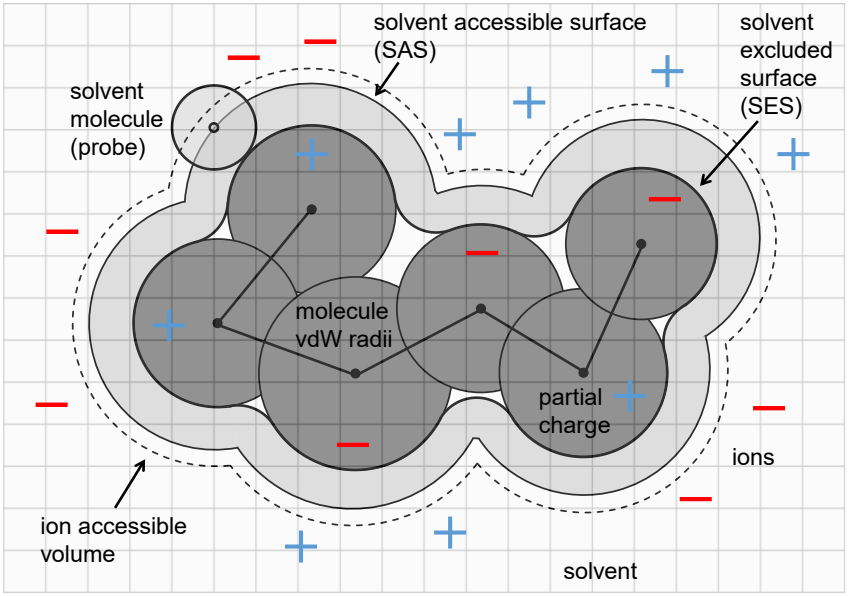

Fig. 1.8: Two-dimensional representation of solvated molecule according to PB model. The PB model describes the electrochemical potential of a macromolecule within a continuous dielectric solvent. The electric potential is affected from partial charges on the molecule surface and its surrounding ions. The protein surface and dielectric boundary is given by overlapping van der Waals (vdW) radii. Alternatively, it can be assessed by rolling a spherical probe over these radii to define a solvent excluded surface (SES), solvent accessible surface (SAS), or in case of an ionic probe, ion accessible volume. The solution to the PB equation is than numerically approximated at discrete locations within a coordinate grid, as indicated in light grey colors. Figure adapted from Grochowski et al. [108].

However, these strategies can introduce biases through arbitrary thresholds in binning or structural domain association and can inflate the number of possible descriptors for subsequent model building. Therefore, the development of protein descriptor sets is a balance between information loss, descriptor bias, and redundancy. Future directions in protein descriptor development aim to quantify spatial patch distribution, proximity, or relative orientation of polarity-induced moments. This could increase their specificity towards predicting aggregation propensity or other forms of multimodal interactions [106].

1.3.3 Statistical Modeling and Machine Learning

QSPR models are used to link the physicochemical properties of molecules to their observable behavior. These models leverage statistical methods to bridge gaps in mechanistic understanding. They utilize probability distributions of calculated descriptors and experimental observations, denoted as independent variable vector x and dependent variable vector y, to infer a mapping function $y(x)$ that describes their relationship. For instance, this mapping function could be a linear model with a slope or regression weight w, as shown in Eq. (1.26).

$$y(x) = wx + \varepsilon \tag{1.26}$$

Statistical models, unlike mechanistic models, include an error term ε, making them non-deterministic [111]. Machine learning than includes automated data preparation techniques as well as training and testing of statistical models [112]. It is utilized in various steps for the development of predictive QSPR models, as depicted in Fig. 1.6.

Machine learning methods can be broadly categorized into supervised, semi-supervised, unsupervised, and reinforcement learning [113]. This categorization is based on the structure of the data and the labeling of independent variables, also known as features in machine learning. A structured data set denoted as $D = \{y, X\}$, includes the labeled feature matrix $X = \{x_i\}_{i=1}^{n}$ with n features per observation. Supervised learning uses this labeled data to address different types of problems such as classification or regression. The type of problem depends on the data type of the feature set and its target variable. Classification applies to categorical or discrete variables, while regression requires numerical or continuous variables. In contrast, an unstructured data set consists of unlabeled information. Unsupervised learning can extract hidden patterns from this data, such as estimating the density of protein surface hydrophobicity by clustering for similarity between neighboring surface points. Semi-supervised learning combines both labeled and unlabeled data to increase information content, while reinforcement learning uses a reward-based approach to train highly flexible algorithms for long term tasks.

Data preprocessing is a crucial step before deriving a mapping function. It connects a target variable with its corresponding features and involves feature discretization (partition continuous features into discrete values), feature encoding (transform categorical values into numerical features), feature extraction (derive features from unlabeled data), or feature imputation (fill sparse feature vectors/matrices). The features can then be transformed and centered using various scaling and normalization techniques, often required for multivariate regression algorithms. Moreover, automated outlier removal can be incorporated during this process. Machine learning can also be utilized to reduce the dimensionality of the feature space through feature selection/elimination or agglomeration. Feature selection involves identifying predictive variables from the feature set, while feature agglomeration reduces the feature space into a latent space by combining variables that are sharing common information.

The machine learning algorithms used for QSPR primarily encompass supervised regression techniques. The choice of these algorithms is influenced by several factors such as data size, linearity, internal correlation, and sparsity, among others [94,112,114]. It is possible to combine multiple algorithms into an ensemble model, which can enhance model predictability, albeit at the expense of interpretability. A key distinction between these techniques lies in their assumptions about the relationship between the target and the feature, which can be either linear or nonlinear. Nonlinear relationships add complexity to the statistical models, affecting both training efficiency and difficulty in model validation. However, a promising method for developing nonlinear regression models, is Gaussian process regression (GPR), which is utilized in this thesis.

Unlike frequentist regression models, GPR is based on Bayesian inference and does not consider the training data D to be randomly distributed, but rather its model parameters [115]. This concept is expressed through Bayes' theorem in Eq. (1.27), which states that a prior hypothesis H can be updated to a posterior hypothesis as more evidence E becomes available through measured data.

$$p(H|E) = \frac{p(E|H)p(H)}{p(E)};\ posterior = \frac{likelihood\ \mathrm{x}\ prior}{marginal\ likelihood} \tag{1.27}$$

The Bayesian approach is an intuitive method for addressing statistical problems as it incorporates prior assumptions about the probability of a certain hypothesis, known as the prior probability distribution or prior $p(H)$. As more evidence becomes available, this space of possibilities is narrowed down, resulting in the posterior probability distribution or posterior $p(H|E)$. The likelihood $p(E|H)$ refers to the conditional probability of the evidence supporting the posterior given the hypothesis. The expression is normalized by the total probability of observing the evidence, often referred to as the marginal likelihood $p(E)$, ensuring a valid probability density of the posterior between one and zero.

By neglecting the marginal likelihood and inserting the linear model introduced in Eq. (1.26) into Eq. (1.27), a simple expression of the Bayesian update rule for linear regression problems is derived in Eq. (1.28), also termed Bayesian linear regression.

$$p(w|y,X) \propto p(w|y,X)p(w|X) \tag{1.28}$$

The main distinction between frequentist and Bayesian inference becomes apparent here as w is defined as a probability distribution over all possible parameter values constrained by measured data, rather than a fixed-point estimate. For N-dimensional multivariate regression, the prior $p(w|X)$ is than defined by its mean μ, typically set to zero, and a joint covariance matrix Σ, with normally distributed regression coefficients $w \sim \mathcal{N}(\mu, \Sigma)$. The symmetric covariance matrix with shape NxN defines the variance on the main diagonal and the covariance on the off diagonal between the regression coefficients.

During model training, the mathematical distance between predicted and observed target values is minimized using error metrics. This minimization is achieved by conditioning model weights or hyperparameters of the underlying covariance matrix. Once the posterior has been estimated, it can be used to make predictions for unobserved data points x'. All regression coefficients of the predictive parameter distribution are averaged and weighted by their posterior probability [116]. Hence, the predictive distribution $f' \triangleq f(x')$ is determined by marginalizing the distribution of $p(f'|x', w)$ over the posterior distribution $p(w|y,X)$, as shown in Eq. (1.29).

$$p(f'|x', y, X) = \int p(f'|x', w)p(w|y,X)dw \tag{1.29}$$

To extend this approach to nonlinear problems, the inputs can be projected into a higher dimensional space or feature space using a set of basis functions, which is referred to as kernelization. These kernels are independent from the regression coefficients and enable the use of a linear model in the transformed feature space. Thus, GPR can be viewed as kernelized Bayesian linear regression. Analogous to Bayesian linear regression, the GPR is defined by its mean function $m(x)$, and its covariance function or kernel $k(x, x')$, as depicted in Eq. (1.30).

$$f(x) \sim GP(m(x), k(x, x')) \tag{1.30}$$

The kernel describes the similarity and smoothness between data points. Its structure allows to introduce prior knowledge of the underlying target-feature relationship to the GPR and is problem specific. Multiple kernels can be combined to improve the accuracy of prediction and quality of model uncertainty, including an experimental noise estimation.

A commonly used kernel in GPR is the radial basis function (RBF) or squared-exponential kernel, shown in Eq. (1.31).

$$k(x, x') = exp\left(-\frac{\|x - x'\|^2}{2l^2}\right) \tag{1.31}$$

The RBF is a smooth, repeated bell curve and is infinitely mean-square differentiable, unlike its generalized form, the Matern kernel that is used in this thesis. It incorporates the Euclidean distance between data points $\|x - x'\|^2$ and the length scale l, characterizing the distance in which y is expected to vary significantly.

GPR is a nonparametric model that isn't confined to a single functional form but comprises an infinite Gaussian distribution of possible functions, termed function space, as depicted in Fig. 1.9. The probability distribution of the function space is limited to kernel functions passing through the measured data points, which reduces the posterior uncertainty around these observations. To illustrate, an initial RBF prior, shown in Fig. 1.9 (a), has a zero-mean function and a broad and uniform model uncertainty. After conditioning the prior, the posterior distribution fits the underlying relationship and reduces the uncertainty of the model respecting the observed data points, as depicted in Fig. 1.9 (b). This example illustrates the main benefits of using GPR for supervised regression tasks, namely, the ability to define a preference bias for the underlying target-feature relationship and directly accessing prediction uncertainty for unseen data points. For further reading, a brief introduction to GPR is provided by Wang [115], while an in-depth derivation can be found in the textbook of Rasmussen [116].

Next to identifying the mapping function, overfitting or overdetermination is a common concern in statistical modeling that needs to be addressed as it can compromise model predictiveness. Overfitting is related to the degree of freedom a mapping function has in describing the target-feature relationship. It depends on the model structure, the dimensionality of the feature set, and the number of observations. In this context, the concept of bias-variance tradeoff describes the balance between model complexity (variance) and error on training data (bias). For instance, assuming linearity between target and feature values shown in Fig. 1.9 (b) by using a linear kernel, would introduce a bias, which reduces the model's variance in fitting the observational data. Therefore, controlling overfitting involves defining the model structure, introducing parameter boundaries, and reducing dimensionality through procedures like feature selection.

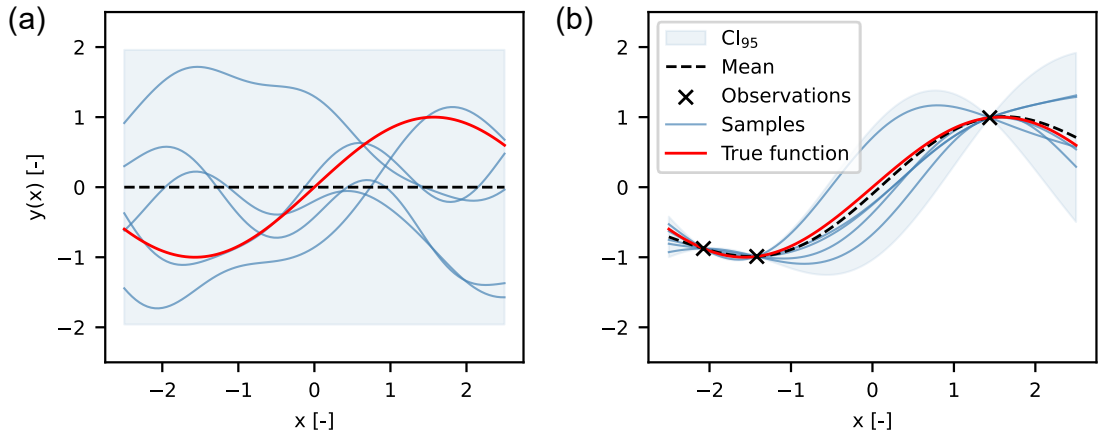

Fig. 1.9: Visualization of GPR using a noise free RBF kernel. (a) Unconditioned prior distribution over possible mapping functions showing a uniform 95% confidence interval and zero mean function. (b) Conditioned posterior distribution over possible mapping functions after incorporating the evidence provided by three observations. Five sample functions are randomly drawn from the kernel's prior and it's posterior to visualize the function space. Figure adapted from Rasmussen [116].

After defining the mapping function, evaluation of QSPR models includes assessing model self-consistency and validating it against withheld observations, also known as testing. During internal model validation, the model's performance and generalizability are quantified. This is done through repeated model building and testing using different combinations of training subsets, a process known as cross-validation [114]. Additionally, y-scrambling can be performed. This involves training and testing a model after randomly assigning target variables to feature values to examine the potential for model overdetermination. A more recent aspect of QSPR model evaluation is model interpretation. This includes inspecting conditioned weight coefficients, conducting sensitivity analysis derived from local feature perturbation, accessing model responses throughout the feature space during single feature variation, or creating decision trees [95].

2 Contribution of this Thesis

2.1 Research Proposal

Monoclonal antibodies represent a major product class in the pharmaceutical industry, primarily targeting the treatment of chronic diseases such as cancer. Increasing development expenditures, regulatory requirements, and industry competition necessitate cost reductions in mAb purification development and production. Technological advancements in downstream processing contribute to reducing these costs through increasing process understanding using computational modeling in line with the implementation of novel separation units. Multimodal chromatography has emerged as a promising technique for mAb purification, improving product quality and enhancing process economics. This is achieved by leveraging orthogonal physicochemical interactions of multimodal chromatographic ligands with the protein surface, which in turn improves mAb-impurity separation and streamlines the entire DSP. However, the development of MMC is often considered less predictable and more complex than unimodal separation units, such as ion exchange chromatography, hindering its widespread industrial adoption.

Particularly, early stages of process development face difficulties in identifying initial purification conditions due to the complex nature of the multimodal protein-ligand interactions and limited process understanding. This leads to a vast experimental space of potential operating conditions that must be explored within a limited timeframe. Additionally, the biopharmaceutical industry is transitioning away from standard antibody formats towards engineered and multispecific entities, where experiential knowledge is scarce. Typically, high-throughput technology and statistical design of experiments are employed to screen for optimal purification conditions due to a lack of process understanding, which inflates development timelines and resource consumption. To overcome the limitations of this empirical process development, it is crucial to understand the relationship between protein structure and its chromatographic behavior, which is not only supported by regulatory agencies through the quality by design initiative but also essential for achieving a competitive advantage in antibody production.

The primary goal of this research is the development of a multiscale model for MMC to increase process understanding and ultimately facilitate its adoption in the purification of mAb therapeutics. This model should connect molecular protein characteristics with macroscopic chromatography behavior to predict chromatographic processes. Hereto, the model should incorporate the amino acid sequence of the target molecules, which is often referred to as *a priori* process development. Thereby, laborious wet lab experimentation for

identification of initial process conditions would be circumvented, assisting the implementation of MMC into DSP development. The model will be derived by inferring chromatography simulation from molecular information within a quantitative structure-property relationship framework, bridging the gap between protein structure and chromatographic behavior. Translating molecular information to chromatographic processes is complex and involves chromatography and molecular modeling connected through statistical models, thus requiring a large quantity of training data.

The first challenge in this research is the efficient acquisition of isotherm parameters and prediction of chromatographic processes by developing a standardized workflow for modeling MMC. Mechanistic chromatography models must be established for a large and diverse set of antibodies and antibody formats, which are then used for statistical modeling. Although highly parameterized isotherm models were available at the beginning of this work, poor parameter identifiability and laborious model calibration prohibited them from the given application. By developing a readily applicable isotherm equation and using an experimental calibration and validation routine that allows robust identification and testing of adsorption parameters, a valuable tool for modeling MMC in academic and industrial applications should be provided.

The second challenge in this research is the development of a QSPR framework for the prediction of initial pH conditions in MMC, which involves the efficient prediction of antibody structures using homology modeling. Typically, chromatography modeling excels in extrapolating chromatography mode and buffer conductivity but is limited to a comparably small range of pH conditions. This is due to the complex pH-dependence of protein adsorption. To overcome this limitation, statistical models can be employed to derive process pH from protein structure that can later be used as a starting point for *in silico* process optimization. Furthermore, such a QSPR model alone would have tremendous value in DSP development. It could either serve as a digital screening tool for accessing antibody manufacturability or accelerate experimental process development through confinement to pH setpoint conditions.

After establishing the necessary models and acquiring sufficient training data, a multiscale model should be developed by combining the mechanistic and statistic approaches. Ultimately, this model should enable an *a priori* prediction of MMC to facilitate the development of mAb purification processes.

2.2 Manuscript Overview

This section presents an overview of all publications that form the foundation of this thesis. The research focuses on the development of a multiscale model for *a priori* process development of MMC in mAb purification. The thesis is organized into three main chapters. Chapter 3 establishes a standardized methodology for mechanistic modeling of MMC, enabling efficient acquisition of isotherm parameters and process simulation. Chapter 4 investigates QSPR methods to predict process pH, a complex parameter that cannot be fully described by first-principal models, serving as a starting point for the multiscale model. Chapter 5 integrates both mechanistic and statistical approaches to predict MMC behavior based solely on amino acid sequence information. All articles have undergone peer review and have been published in reputable international scientific journals. They have not been used in any other publication-based thesis.

Chapter 3: Standardized method for mechanistic modeling of multimodal anion exchange chromatography in flow through operation

Rudger Hess, Doil Yun, David Saleh, Till Briskot, Jan-Hendrik Grosch, Gang Wang, Thomas Schwab, Jürgen Hubbuch

Journal of Chromatography A (1690), 2023, p. 463789

The research article presents a mechanistic modeling approach to accelerate the development of anionic MMC for biopharmaceutical purification. Within the study, an existing multimodal isotherm model was modified to enable model calibration using only three chromatographic experiments. This model was then used to predict the retention of four antibody formats at two pH values. The model reduction improved structural parameter identifiability and enabled an analytical isotherm parameter determination, which was further refined by incorporating size exclusion effects. The models were successfully validated by predicting three MMC operation modes at varying ionic strengths for four different antibody formats. This standardized modeling approach reduces material consumption, experimental and computational burden, and has the potential to accelerate the development of complex MMC purification processes in biopharmaceutical DSP.

Chapter 4: Antibody sequence-based prediction of pH gradient elution in multimodal chromatography

Rudger Hess, Jan Faessler, Doil Yun, David Saleh, Jan-Hendrik Grosch, Thomas Schwab, Jürgen Hubbuch

Journal of Chromatography A (1711), 2023, p. 464437

The research article introduces a methodology for predicting antibody elution behavior in MMC based on amino acid sequences. By analyzing 64 full-length antibodies, the study identified six key structural features correlated with elution behavior. The developed model effectively predicts pH gradient elution for diverse set of therapeutic antibodies. This model can inform process development, reduce trial and error during process optimization, and enable *in silico* manufacturability assessment. The study's QSPR model can replace experimental screening of initial process pH in MMC, advancing early process development. Furthermore, identified feature dependencies could improve mechanistic chromatography models by considering a molecular-level detail.

Chapter 5: Predicting multimodal chromatography of therapeutic antibodies using multiscale modeling

Rudger Hess, Jan Faessler, Doil Yun, Ahmed Mama, David Saleh, Jan-Hendrik Grosch, Gang Wang, Thomas Schwab, Jürgen Hubbuch

Journal of Chromatography A (1718), 2024, p. 464706

The research article describes a multiscale modeling approach for predicting the multimodal chromatographic behavior of therapeutic antibodies using sequence information. QSPR modeling was utilized to correlate isotherm parameters with physicochemical descriptors from 59 full-length antibodies at different pH values. Both the QSPR and the mechanistic chromatography models demonstrated high accuracy and identified the significance of electrostatic interaction, hydrophobicity, and the HFR3 region in antibody binding to multimodal resin. The multiscale modeling approach can lead to more efficient and cost-effective process development in therapeutic antibody purification, benefiting the biopharmaceutical industry, while further improving future isotherm models.

3 Standardized Method for Mechanistic Modeling of Multimodal Anion Exchange Chromatography in Flow Through Operation

Rudger Hess[1, 2], Doil Yun[2], David Saleh[2], Till Briskot[2], Jan-Hendrik Grosch[2], Gang Wang[2], Thomas Schwab[2], Jürgen Hubbuch[1,*]

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

[2] DSP Development, Boehringer Ingelheim Pharma GmbH & Co. KG, Biberach, Germany

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

Journal of Chromatography A (1690), 2023, p. 463789

DOI: 10.1016/j.chroma.2023.463789

Abstract

Multimodal chromatography offers an increased selectivity compared to unimodal chromatographic methods and is often employed for challenging separation tasks in industrial downstream processing (DSP). Unfortunately, the implementation of multimodal polishing into a generic downstream platform can be hampered by non-robust platform conditions leading to a time and cost intensive process development. Mechanistic modeling can assist experimental process development but readily applicable and easy to calibrate multimodal chromatography models are lacking. In this work, we present a mechanistic modeling aided approach that paves the way for an accelerated development of anionic mixed-mode chromatography (MMC) for biopharmaceutical purification. A modified multimodal isotherm model was calibrated using only three chromatographic experiments and was employed in the retention prediction of four antibody formats including a Fab, a bispecific, as well as an IgG1 and IgG4 antibody subtype at pH 5.0 and 6.0. The chromatographic experiments were conducted using the anionic mixed-mode resin Capto adhere at industrial relevant process conditions to enable flow through purification. An existing multimodal isotherm model was reduced to hydrophobic interactions in the linear range of the adsorption isotherm and successfully employed in the simulation of six chromatographic experiments per molecule in concert with the transport dispersive model (TDM). The model reduction to only three parameters did prevent structural parameter non-identifiability and enabled an analytical isotherm parameter determination that was further refined by incorporation of size exclusion effects of the selected multimodal resin. During the model calibration, three linear salt gradient elution experiments were performed for each molecule followed by an isotherm parameter uncertainty assessment. Lastly, each model was validated with a set of step and isocratic elution experiments. This standardized modeling approach facilitates the implementation of multimodal chromatography as a key unit operation for the biopharmaceutical downstream platform, while increasing the mechanistic insight to the multimodal adsorption behavior of complex biologics.

Keywords: *Mixed-mode gradient elution; Therapeutic antibody; Analytical parameter determination; Parameter uncertainty assessment; Model calibration and validation*

3.1 Introduction

The biopharmaceutical downstream purification is heavily reliant on chromatographic unit operations for the production of monoclonal antibodies (mAbs) [16,117]. Due to the highly conserved structure of this product class, the manufacturing process utilizes a generic order of consecutive chromatographic unit operations often referred to as a platform process [9]. Increasing demands of platform productivity and challenging impurity profiles of novel drug candidates led to the introduction of multimodal chromatography to the biopharmaceutical platform [28,60]. The structure of multimodal ligands combines two or more primary physicochemical moieties, which enable a comparably high selectivity compared to singular interaction resins [27]. In particular, the combination of hydrophobic and electrostatic interactions has proven beneficial in the primary capture of harvested cell culture fluid as well as the removal of product- and process-related impurities during subsequent polishing steps [51,58]. Especially, salt-tolerant anionic mixed-mode polishing is deemed to be superior to established platform operations as it can replace multiple orthogonal polishing steps while reducing the DNA and viral burden [61,62]. Nevertheless, the widespread implementation of MMC processes is lacking behind its potential [118]. The templated platform approach struggles to integrate robust MMC unit operations owed to a lack of mechanistic understanding [46,66].

The simulation of complex multimodal interactions by mechanistic chromatography modeling is considered challenging and requires sophisticated isotherm models coupling hydrophobic and electrostatic effects [54,119]. In recent years, the complexity of published MMC isotherms has increased further by incorporating a stoichiometric water displacement as well as a pH dependency into existing isotherm models [84,120]. On the other hand, oversimplified models based on the Langmuir isotherm have been employed for describing the MMC interaction [121,122]. While the increased parameterization of mechanistic models could yield an improved description of physicochemical phenomena there is a demand for readily applicable and easy to calibrate models [29,123]. A way of reducing the effort for multimodal mechanistic modeling is to focus on the governing form of physicochemical interaction under relevant process conditions. A purely electrostatic treatment of the multimodal interaction has been implemented by using the stoichiometric steric mass action model (SMA) [65,124–127] for the simulation of bind-and-elute protein adsorption. To date, no standardized methodology for the simulation of multimodal processes in an electrostatically repulsive environment, which are common for anionic flow through polishing, has been shown.

In this study, we introduce a standardized calibration and validation workflow for the mechanistic MMC modeling under electrostatic repulsive conditions. The model reduction of an existing MMC isotherm [54] is shown, which enabled an analytical isotherm parameter determination. Hereafter, the analytical parameter set was refined numerically to consider resin specific size exclusion effects. The modified isotherm was calibrated for four different antibody formats using linear salt gradients. Additionally, an investigation of the isotherm parameter uncertainty, correlation, and the influence of size exclusion effects to

the parameter determination was performed. Lastly, the predictive power of the model was validated by a comprehensive set of chromatographic experiments.

3.2 Theory and Model Design

3.2.1 Transport Dispersive Model

The mass transfer inside the chromatographic column was described by a lumped mass transfer model referred to as the transport dispersive model [43,44]. According to the TDM the mass balance is of convection diffusion reaction type and describes the macroscopic transport and accumulation of component $i \in [1, n]$ through the chromatographic column, shown in Eq. (3.1).

$$\begin{aligned}\frac{\partial c_{\mathrm{b},i}}{\partial t}(x,t) = &-u_{\mathrm{int}}\frac{\partial c_{\mathrm{b},i}}{\partial x}(x,t) + D_{\mathrm{ax}}\frac{\partial^2 c_{\mathrm{b},i}}{\partial x^2}(x,t) \\ &-\frac{1-\varepsilon_{\mathrm{b}}}{\varepsilon_{\mathrm{b}}}\frac{3}{r_{\mathrm{p}}}k_{\mathrm{eff},i}\left(c_{\mathrm{b},i}(x,t) - c_{\mathrm{p},i}(x,t)\right)\end{aligned} \tag{3.1}$$

The system depicts a chromatographic column with length L_{c} [m] that is uniformly packed with a resin of particle radius r_{p} [m]. The column volume (CV) is divided into the stationary and the mobile phase whereas the mobile phase is further subdivided into the interparticle and the intraparticle volume. The concentration of the solute within the interparticle bulk volume of the resin bed $c_{\mathrm{b},i}$ [molm^{-3}] is referred to as the interstitial concentration and is dependent of the time t [s] and the location $x \in [0, L_{\mathrm{c}}]$ [m]. The convective transport in between the resin beads depends on the interstitial velocity u_{int} [ms^{-1}]. The kinetic peak broadening throughout the column is modeled by the axial dispersion coefficient D_{ax} [m^{2}s^{-1}] and contains contributions of the Eddy and axial diffusion. The interstitial porosity of the column bed is given as ε_{b} [−]. Further, the accumulation term denotes for a solid film linear driving force (LDF) model and includes the solute concentration on the adsorber surface in the resin pores $c_{\mathrm{p},i}$ [molm^{-3}]. Within the model, $c_{\mathrm{p},i}$ is assumed to be equal to the constant component concentration in the particle pore system. The interfacial mass transfer resistance between the mobile and the stationary phase is lumped into the effective mass transfer coefficient $k_{\mathrm{eff},i}$ [ms^{-1}]. The chromatographic model is complemented by Danckwerts' boundary conditions of dispersive systems at the column inlet and outlet [128], defined in Eq. (3.2) and (3.3).

$$\frac{\partial c_{\mathrm{p},i}}{\partial t}(x=0,t) = \frac{u_{\mathrm{int}}(t)}{D_{\mathrm{ax}}}\left(c_{\mathrm{b},i}(x=0,t) - c_{\mathrm{in},i}(t)\right) \tag{3.2}$$

$$\frac{\partial c_{\mathrm{b},i}}{\partial x}(x = L_c, t) = 0 \tag{3.3}$$

The continuity equation of the TDM inside the adsorber particle is derived in accordance with the LDF model as shown in Eq. (3.4).

$$\frac{\partial c_{\mathrm{p},i}}{\partial t}(x,t) = \frac{3}{r_{\mathrm{p}}}\frac{k_{\mathrm{eff},i}}{\varepsilon_{\mathrm{p},i}}\Big(c_{\mathrm{b},i}(x,t) - c_{\mathrm{p},i}(x,t)\Big) - \frac{1-\varepsilon_{\mathrm{p},i}}{\varepsilon_{\mathrm{p},i}}\frac{\partial q_{\mathrm{p},i}}{\partial t}(x,t) \tag{3.4}$$

The volume-wise partitioning of the resin bed is defined by the particle porosity that is occupied by the protein species $\varepsilon_{\mathrm{p},i}\,[-]$. Lastly, the reactive part of the TDM is expressed within the adsorbed component concentration per unit of solid volume $q_{\mathrm{p},i}[\mathrm{molm}^{-3}]$.

3.2.2 Linear Mixed-Mode Isotherm Model

The isotherm equation employed in this study is based on the mixed-mode adsorption and desorption model of Nfor et al. [54] and grounded on the thermodynamic framework of Mollerup [79,82]. The initial model does consider electrostatic and hydrophobic physicochemical interactions at high load density conditions. In course of this study, the initial MMC isotherm model was reduced to a linear representation of the adsorption and desorption behavior. Different model assumptions were made regarding the multimodal interaction mechanism encountered in an electrostatic repulsive environment. At a process pH far below the isoelectric point (pI) of the given molecule species $(\mathrm{pH} \ll \mathrm{pI})$, the electrostatic net-attraction of the molecule-ligand interaction approaches zero, as observed by Nfor et al. [54]. Accordingly, the association of the reactants during the mixed-mode interaction was presumed to be governed by hydrophobic attraction. By assuming the counterion and the ligand binding site of the chemical equilibrium to be uniform and monovalent, the reaction equation is introduced as depicted in Eq. (3.5).

$$P_{\mathrm{sol},i} + n_i L \rightleftharpoons P_{\mathrm{ads},i} \tag{3.5}$$

Within the model, the solute protein species $\mathrm{P}_{\mathrm{sol},i}$ adsorbs to the ligand binding site L and forms the protein-ligand complex $\mathrm{P}_{\mathrm{ads},i}$. During the adsorption process, the hydrophobic moiety of the ligand entails a reversible association with the non-polar binding sites of the protein. The stoichiometric coefficient of the hydrophobic interaction is denoted by $n_i\,[-]$. According to the law of mass action, the protein and buffer system specific equilibrium constant $\widetilde{K}_{\mathrm{eq},i} = k_{\mathrm{ads},i}/k_{\mathrm{des},i}\,[-]$ can be derived at the thermodynamic equilibrium of the mixed-mode adsorption and desorption reaction. The equilibrium constant equals the ratio of the adsorption rate $k_{\mathrm{ads},i}$ to the desorption rate $k_{\mathrm{des},i}$ of the protein species from the ligand binding site. The concentration of free ligands available for hydrophobic interaction can be treated analogous to the SMA model of Brooks and Cramer [77]. Moreover, the

steric hinderance of the free ligand binding sites during the adsorption process can be neglected at low load density conditions $q_i \ll c_{\mathrm{p},i}$ [79,82]. Lastly, the protein-protein interaction was recognized to be of minor importance compared to other model parameters by multiple authors [54,82,129–131]. The aforementioned theoretical considerations led to the linear mixed-mode adsorption model at equilibrium conditions, presented in Eq. (3.6).

$$K_i(c_{\mathrm{p,s}}) = \widetilde{K}_{\mathrm{eq},i}\left(\frac{\Lambda_{\mathrm{HIC}}}{c_{\mathrm{p},v}}\right)^{n_i} \exp\left(K_{\mathrm{s},i}c_{\mathrm{p,s}}\right) \tag{3.6}$$

The isotherm equals the equilibrium ratio $K_i(c_{\mathrm{p,s}})$ as the ratio between the adsorbed protein species to the solute protein species in the column with the subscript s denoting for the counterion species. The molar concentration of the available ligands of the adsorption process is determined by the hydrophobic ligand density Λ_{HIC} $[\mathrm{molm^{-3}}]$ per adsorber skeleton volume. Additionally, the molar density of the solution in the pore volume $c_{\mathrm{p},v}$ $[\mathrm{molm^{-3}}]$ is introduced that is dependent on modulator concentration. Lastly, the salt-protein interaction parameter $K_{\mathrm{s},i}$ $[\mathrm{mol^{-1}m^3}]$ describes the protein and modulator dependent interaction. Assuming $c_{\mathrm{p},v}$ to be constant [132] allowed to modify the definition of the equilibrium constant to $\widetilde{K}'_{\mathrm{eq},i} \approx \widetilde{K}_{\mathrm{eq},i}\left(\Lambda_{\mathrm{HIC}}/c_{\mathrm{p},v}\right)^{n_i}$ $[-]$ as has previously been demonstrated by Andris et al. [131]. The parameter modification led to the final equilibrium equation of the linear mixed-mode isotherm model, given in Eq. (3.7).

$$K_i(c_{\mathrm{p,s}}) = \widetilde{K}'_{\mathrm{eq},i} \exp\left(K_{\mathrm{s},i}c_{\mathrm{p,s}}\right) \tag{3.7}$$

For a dynamic location and time dependent numerical solution of the adsorption and desorption process, the kinetic material balance [83] of Eq. (3.7) is introduced in Eq. (3.8).

$$k_{\mathrm{kin},i}\frac{\partial q_i}{\partial t}(x,t) = \widetilde{K}'_{\mathrm{eq},i} \exp\left(K_{\mathrm{s},i}c_{\mathrm{p,s}}\right) c_{\mathrm{p},i}(x,t) - q_i(x,t) \tag{3.8}$$

Consistent with Hahn et al. [128], the kinetic coefficient is defined as $k_{\mathrm{kin},i} = k_{\mathrm{des},i}^{-1}$ $[\mathrm{s(molm^{-3})^{n_i}}]$.

3.2.3 Isotherm Parameter Determination

The parameter estimation of the modified MMC isotherm model is based on the theoretical foundation of Yamamoto et al. [76,133] and often referred to as the Yamamoto method [134,135]. The estimation method relies on linear modulator gradients and has been mainly used in cation exchange chromatography (CEX) modeling. In this study, negative salt linear gradient elution (LGE) experiments have been incorporated to estimate the linear MMC isotherm parameters shown in Eq. (3.7). In course of the calibration methodology the gradient slope $g = c_{\mathrm{s}}^{\mathrm{F}} - c_{\mathrm{s}}^{\mathrm{I}}/V_{\mathrm{G}}$ $[\mathrm{molm^{-6}}]$ is introduced and depicts the change of the modulator concentration per linear gradient volume V_{G} $[\mathrm{m^3}]$, as defined in Eq. (3.9).

$$G = g\varepsilon_\mathrm{b}V_\mathrm{col} = \frac{c_\mathrm{s}^\mathrm{F} - c_\mathrm{s}^\mathrm{I}}{V_\mathrm{G}}\varepsilon_\mathrm{b}V_\mathrm{col} \tag{3.9}$$

The initial and the final salt concentration at the start and the end of the gradient are given as c_s^I and c_s^F. Further, the method uses a classical phase ratio H $[-]$ definition, shown in Eq. (3.10), in which the adsorbed phase refers to the solid column volume according to Guiochon et al. [43].

$$H = \frac{1 - \varepsilon_\mathrm{b}}{\varepsilon_\mathrm{b}} \tag{3.10}$$

When multiplying Eq. (3.9) with Eq. (3.10) the normalized gradient slope GH $[\mathrm{molm^{-3}}]$ is introduced in Eq. (3.11).

$$GH = g(1 - \varepsilon_\mathrm{b})V_\mathrm{col} \tag{3.11}$$

Finally, the salt LGEs are combined with an equilibrium model, which considers the zone movement of the components through the eluting column. When solving the equilibrium model after the rate of change of the components, Eq. (3.12) can be derived [133].

$$GH = \int_{c_\mathrm{s}^\mathrm{I}}^{c_{\mathrm{s},i}^\mathrm{R}} \frac{dc_\mathrm{s}}{K_i(c_\mathrm{s}) + K_{\mathrm{sec},i} - K_{\mathrm{sec},s}} \tag{3.12}$$

The denominator of the integral contains the salt dependent isotherm of the protein species $K_i(c_\mathrm{s})$. The upper limit of the integral $c_{\mathrm{s},i}^\mathrm{R}$ depicts the salt concentration at the peak maximum of the eluting molecule species whereas the lower limit is defined as the initial gradient concentration c_s^I. Further, the distribution coefficients $K_{\mathrm{sec},i} = \varepsilon_{\mathrm{p,i}}$ $[-]$ and $K_{\mathrm{sec},s} = \varepsilon_{\mathrm{p},s}$ $[-]$ denote for the fraction of the pore volume accessible to the protein and the salt component at non-interacting conditions, comparable to the environment found during size exclusion chromatography (SEC). The SEC distribution coefficients equal the protein species pore porosity $\varepsilon_{\mathrm{p},i}$ $[-]$ and the pore porosity of the pore penetrating salt tracer component $\varepsilon_{\mathrm{p},s}$ $[-]$. Inserting Eq. (3.7) into Eq. (3.12) yields Eq. (3.13).

$$GH = \int_{c_\mathrm{s}^\mathrm{I}}^{c_{\mathrm{s},i}^\mathrm{R}} \frac{dc_\mathrm{s}}{\widetilde{K}'_{\mathrm{eq},i}\exp\left(K_{\mathrm{s},i}c_\mathrm{s}\right) + K_{\mathrm{sec},i} - K_{\mathrm{sec,s}}} \tag{3.13}$$

If the distribution coefficient of the protein species is assumed to be equal to the distribution coefficient of the salt component $K_{\mathrm{sec},i} \approx K_{\mathrm{sec,s}}$ and the protein is strongly retained at the start of the gradient, hence neglecting the lower limit of the integral, an analytical solution can be derived, shown in Eq. (3.14) [133,136]:

$$\log(-GH) = \log\left(-\frac{1}{\widetilde{K}'_{\mathrm{eq},i}K_{\mathrm{s},i}}\right) - K_{\mathrm{s},i}c_{\mathrm{s},i}^\mathrm{R} \tag{3.14}$$

The given linearization enables the analytical determination of the isotherm parameters $\widetilde{K}'_{\mathrm{eq},i}$ and $K_{\mathrm{s},i}$ from a semi-logarithmic plot of the negative normalized gradient slope against the ionic strength of the gradient eluant referred to as a Yamamoto plot. The intercept of the linear equation is defined as $y_{0,i} = \log\left(-1/\widetilde{K}'_{\mathrm{eq},i} K_{\mathrm{s},i}\right)$ whereas the slope is defined as $m_i = -K_{\mathrm{s},i}$. If size exclusion effects of the protein species do appear $K_{\mathrm{sec},i} \neq K_{\mathrm{sec,s}}$ and the lower limit of the integral is not neglected, the isotherm parameters have to be numerically estimated by Eq. (3.15).

$$c_{\mathrm{s},i}^{\mathrm{R}} = -\log\left(\frac{\widetilde{K}'_{\mathrm{eq},i}\left(-\exp\left(K_{\mathrm{s},i} GHK_{\Delta,is}\right)\right) + \widetilde{K}'_{\mathrm{eq},i} + K_{\Delta,is}\exp\left(K_{\mathrm{s},i}\left(GHK_{\Delta,is} - c_{\mathrm{s}}^{\mathrm{I}}\right)\right)}{K_{\Delta,is}}\right)\frac{1}{K_{\mathrm{s},i}} \tag{3.15}$$

Within the numerical solution the isotherm parameters are estimated by iteratively minimizing the error of the predicted peak salt concentration. Additionally, the parameter $K_{\Delta,is}\,[-]$ is introduced as the difference between the salt distribution coefficient $K_{\mathrm{sec,s}}$ and the protein specific distribution coefficient $K_{\mathrm{sec},i}$, as defined in Eq. (3.16).

$$K_{\Delta,is} = K_{\mathrm{sec,s}} - K_{\mathrm{sec},i} = \varepsilon_{\mathrm{p,s}}\left(1 - \frac{\varepsilon_{\mathrm{p},i}}{\varepsilon_{\mathrm{p,s}}}\right) \tag{3.16}$$

3.3 Materials and Methods

3.3.1 Resin, Buffers, and Molecules

The multimodal strong anion exchanger Capto adhere (Cytiva, Marlborough, USA) was used in this study. The model calibration and validation were conducted with a prepacked HiScreen column (Cytiva) having a bed height of 10 cm, and column volume of 4.7 mL. The spherical resin beads are comprised of a rigid and highly cross-linked agarose matrix with a median particle size of 75 µm. The N-Benzyl-N-methyl ethanol amine ligand exhibits multimodal functionality through ionic interaction, hydrogen bonding, and hydrophobic interaction [68].

All substances employed in this study were purchased from Sigma-Aldrich Co LLC (Saint Louis, USA). Ultrapure water was filtered with the Milli-Q Advantage A10 (Merck Millipore, Burlington, USA) water purification system and used for buffer preparation. Thereafter, the buffer pH was adjusted with 1 M hydrochloric acid as well as 1 M sodium hydroxide. A multicomponent buffer system at pH 5.0 and 6.0 was utilized for equilibration, elution, and buffer exchange during all chromatographic experiments. The buffer system was comprised of 9.1 mM 1,2-ethanediamine, 6.4 mM 1-methylpiperazine, 13.7 mM 1,4-dimethylpiperazine, 5.8 mM bis-tris, and 7.7 mM hydroxylamine with the addition of 60 mM hydrochloric acid, adapted from Kröner and Hubbuch [137]. The buffer system offered

a broad buffer capacity in the range of pH 4.0 to 9.0 while avoiding the introduction of anionic counterions. Further, 1.5 M sodium chloride was supplemented to the multicomponent buffer that was used for equilibration (at inlet B), adding up to a total of 1.56 M chloride counterions. The adapted buffering capability was validated by titration with sodium hydroxide (data not shown).

Four different human immunoglobulin-based molecules (Boehringer Ingelheim Pharma GmbH & Co. KG, Biberach, Germany) were used as model proteins in this study. A Fab, a bispecific (IgG(H)-scFv) [39], an IgG4, and IgG1 antibody were investigated, which did show a broad variability of physicochemical parameters, as listed in Table 3.1. The antibody formats were expressed using a stably transfected Chinese hamster ovary cell line and captured by Protein A affinity chromatography. After neutralization to pH 5.5 and 0.2 µm sterile filtration the protein solutions were frozen at −70°C. After thawing, the final load material was buffer exchanged with a 5 mL HiTrap Desalting column (Cytiva) and adjusted to a concentration of 5 g/L using the equilibration buffer.

Table 3.1: Overview of the model proteins physicochemical parameters.The hydrodynamic radii of the molecules were taken from literature values.

Molecule	**Molecular weight [kDa]**	**Hydrodynamic radius [nm]**	**Isoelectric point [pH]**
Fab	48	4.1 [138]	8.6
Bispecific	202	6.3 [139]	9.3
IgG4	144	4.5 [140]	7.2
IgG1	146	4.5 [140]	8.3

3.3.2 Instruments and Software

All chromatographic experiments were performed on an ÄKTA Avant 25 (Cytiva) preparative chromatography system using a 0.6 mL mixing chamber. The outlet UV signal was monitored at a wavelength of 280 nm through a flow cell with 0.2 cm pathlength. The system was operated with the Unicorn 7.5 (Cytiva) control software. The offline protein concentration was measured using a NanoDrop 2000c (Thermo Fisher Scientific, Waltham, USA) spectrophotometer. *In silico* workloads were executed on a P620 ThinkStation® (Lenovo, Morrisville, USA) workstation pc equipped with a Ryzen Threadripper PRO 3995WX (AMD, Santa Clara, USA) processor. The software ChromX (Cytiva) [141] was used to perform the numerical simulations of the TDM chromatography models. For all

simulations, a fractional time-stepping scheme with an initial step size of 1 s was used. Further, the chromatographic column was spatially discretized into 60 equidistant axial cells using the linear Streamline-Upwind Petrov–Galerkin method. To account for the mixing chamber, a continuous stirred-tank reactor (CSTR) model was included upfront the chromatographic column. Lastly, the open-source programming language Python 3.8.8 was used as a general environment for the parameter estimation, uncertainty assessment and result visualization.

3.3.3 System and Column Characterization

The extra-column effects and column specific parameters were characterized in terms of sensor dead volumes, column porosities, and ionic capacity. All experiments were performed in triplicate at the set point flow rate. All parameters except the ionic capacity were determined using pulse injections with a pulse volume of 1% of the column volume. The sensor dead volumes and total column porosity were determined using a pore-penetrating, non-adsorbing tracer composed of 1 M sodium chloride and 1% (V/V) acetone. A distinction was made between an isocratic and non-isocratic system characterization to account for mixing effects throughout the system. The isocratic characterizing considered dead volumes starting from the injection valve of the system whereas non-isocratic experiments did require additional consideration of the buffer flow path. The column bed porosity was calculated from the interstitial column volume using 2000 kDA dextran as a non-pore penetrating, non-interacting tracer. The accessible bead porosity was determined by applying salt and protein pulses under non-binding conditions to the chromatographic column, according to methods by Hunt et al. [142]. Non-binding conditions were achieved using a 50 mM acidic acid buffer at pH 3.5. The non-binding behavior was validated by repeating the experiments with a 40 kDA dextran tracer using 50 mM acidic acid buffer at pH 3.5 as well as a buffer at experimental set point conditions (data not shown). Lastly, the ionic capacity of the column was determined by acid-base titration as published by Huuk et al. [143]. In contrast to the published protocol, a 10 mM hydrochloric acid solution was used as the titrant for the anionic acid-base titration. According to Nfor et al. [54], the hydrophobic capacity was set to be equal to the electrostatic capacity as the Capto adhere ligand presents the same number of electrostatic and hydrophobic functional groups.

3.3.4 Column Experiments

The chromatographic experiments consisted of bind-and-elute as well as isocratic experiments with a residence time of 5 min as listed in Table 3.2. The column experiments were conducted at pH 5.0 for the bispecific and IgG4 and at pH 6.0 for the Fab, and IgG1 antibody. All chromatographic experiments were initialized through a 3 CV equilibration phase using the high salt multicomponent buffer, except the isocratic experiments that

utilized a mixture of the high and low salt buffer instead. After the equilibration, 1.25 mL buffer exchanged load material was applied during the loading phase followed by a 0.5 mL post loading step using the equilibration buffer. The loading density was set to 1.0 g/L. Three LGE experiments were conducted for model calibration. During the gradient elution, the counterion concentration was linearly decreased from 1560 to 60 mM over a gradient length of 10, 20, and 30 CV. The model validation included two step elution (SE) experiments with a step length of 10 CV starting from a counterion concentration of 1560 mM to a final concentration of 100 and 300 mM, respectively. Further, a 10 CV isocratic elution (IE) experiment was conducted with a constant counterion concentration of 200 mM during the column equilibration, loading and elution phase. The elution phase was followed by 4 CV of column regeneration with 1M acetic acid and a 5 CV cleaning in place step using 1 M sodium hydroxide. Lastly, the column was stored with a 20% ethanol solution over 4 CV.

Table 3.2: Summary of all chromatographic calibration and validation experiments.

Molecule	**pH**	**Load density [g/L]**	**Objective**	**Elution mode**	**Elution length [CV]**	**Salt concentration gradient end [mM]**
Fab	6.0	0.5 - 1.0	Calibration	LGE	10, 20, 30	75
			Validation	SE / IE	10	100, 200, 300
Bispecific	5.0	0.5 - 1.0	Calibration	LGE	10, 20, 30	75
			Validation	SE / IE	10	100, 200, 300
IgG4	5.0	0.5 - 1.0	Calibration	LGE	10, 20, 30	75
			Validation	SE / IE	10	100, 200, 300
IgG1	6.0	0.5 - 1.0	Calibration	LGE	10, 20, 30	75
			Validation	SE / IE	10	100, 200, 300

3.3.5 Model parameter and Uncertainty Estimation

The mechanistic model variables consisted of mass transfer coefficients and thermodynamic isotherm parameters. Prior to the estimation of the thermodynamic parameters, the axial dispersion coefficient and the effective mass transfer coefficient were determined under consideration of the molecule hydrodynamic radii. The coefficients were determined by

using empirical correlations of Chung and Wen [144] as well as the penetration theory [43]. According to Rischawy et al. [145], the mass transfer of the non-interacting salt component was set to unity by defining $k_{\mathrm{eff,s}} \cong r_{\mathrm{p}}/3$. The derivations of all mass transfer coefficients are listed in Table A1 in the Appendix A.

Subsequently, the three thermodynamic parameters $\widetilde{K}'_{\mathrm{eq},i}$, $K_{\mathrm{s},i}$, and $k_{\mathrm{kin},i}$ were determined from the salt LGE calibration experiments, as depicted in Fig. 3.1. After the salt concentrations at the peak maximum of each molecule species $c_{\mathrm{s},i}^{\mathrm{R}}$ were gathered, the Yamamoto method was employed, as introduced in Section 3.2.3. According to the Yamamoto method, the initial isotherm parameters $\widetilde{K}'_{\mathrm{eq},i,0}$ and $K_{\mathrm{s},i,0}$ were determined using the linearization given in Eq. (3.14). Thereafter, the refined isotherm parameters $\widetilde{K}'_{\mathrm{eq},i}$ and $K_{\mathrm{s},i}$ were numerically estimated using Eq. (3.15) to incorporate resin specific size exclusion effects, as recognized by Creasy et al. [146]. Starting with $\widetilde{K}'_{\mathrm{eq},i,0}$ and $K_{\mathrm{s},i,0}$ as initial values of the unconstrained numerical estimation, the residual sum of squares between the calculated and the experimental peak salt concentration was minimized using the Levenberg-Marquardt gradient descent algorithm [147] as implemented in the python package SciPy 1.7 [148]. The isotherm parameter uncertainty was derived from the numerical estimation, according to published protocols of Rischawy et al. [145].

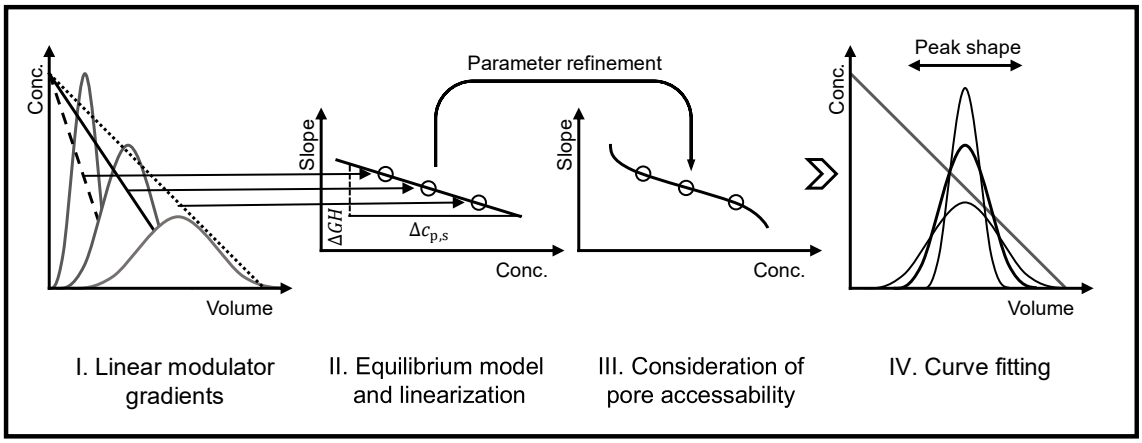

Fig. 3.1: Thermodynamic parameter estimation and uncertainty assessment routine.A four-step workflow for the estimation of the equilibrium constant $\widetilde{\mathrm{K}}'_{\mathrm{eq,i}}$, the salt-protein interaction parameter $\mathrm{K}_{\mathrm{s,i}}$, and the kinetic coefficient $\mathrm{k}_{\mathrm{kin,i}}$ is shown. Starting from linear salt gradient elution experiments (I), the salt concentrations at the peak maxima are used for the determination of the parameters $\widetilde{\mathrm{K}}'_{\mathrm{eq,i}}$ and $\mathrm{K}_{\mathrm{s,i}}$ via the Yamamoto method. The initial equilibrium constant $\widetilde{\mathrm{K}}'_{\mathrm{eq,i,0}}$ and the initial salt-protein interaction parameter $\mathrm{K}_{\mathrm{s,i,0}}$ are determined through linearization (II) and further numerically refined under consideration of the molecule pore accessibility (III). Additionally, the uncertainty of the refined parameters $\widetilde{\mathrm{K}}'_{\mathrm{eq,i}}$ and $\mathrm{K}_{\mathrm{s,i}}$ is estimated from the numerical parameter estimation. Refined parameters are employed for chromatographic simulation to adjust the simulated peak shape by fitting of parameter $\mathrm{k}_{\mathrm{kin,i}}$ (IV).

The covariance matrix of the parameter estimation was given as a Jacobian approximation to the Hessian of the least squares objective function. From the covariance matrix a two-dimensional 95% confidence ellipse was constructed based on the method of Schöneberger et al. [149]. After, the refined isotherm parameters $\widetilde{K}'_{\mathrm{eq},i}$ and $K_{\mathrm{s},i}$ were determined, they were employed for chromatographic simulation using the TDM as introduced in Section 3.2.1. Lastly, the kinetic parameter $k_{\mathrm{kin},i}$ was estimated by simultaneously curve fitting the three LGE calibration experiments for each molecule, as described by Hahn et al. [83].

3.4 Results

3.4.1 System and Column Characterization

The characterization of the chromatographic system was performed without a column installed to account for the mixing effects throughout the buffer flow path. All chromatographic experiments were based on the system characterization and were corrected by their corresponding dead volume. The column characterization did include the specification of the interstitial and the component specific accessible porosity [150] as well as the determination of the ligand density according to published protocols of Huuk et al. [143]. Tracer experiments were conducted with the column connected to the system using dextran, sodium chloride, and the four antibody molecules under non-binding conditions. All system and column specific parameters including their derivation are listed in Table 3.3 and Table 3.4.

3.4.2 Model Calibration

The chromatographic model calibration required the determination of the component specific mass transfer and thermodynamic model parameters. Additionally, a 0.2 mL CSTR was included upstream the column model to account for the mixing chamber of the ÄKTA system. The mass-transfer phenomena of the system were determined based on empirical correlations. The axial dispersion coefficient was calculated as $D_{\mathrm{ax}} = 0.125\ \mathrm{mm^2 s^{-1}}$ assuming a dynamic buffer viscosity of 0.001 $\mathrm{kgm^{-1}s^{-1}}$ as determined by Guichon [43]. The effective mass transfer of the salt component was assumed to be not limiting, and its coefficient was set to 0.0125 $\mathrm{mms^{-1}}$. The molecule specific effective mass transfer coefficients did depend on the hydrodynamic radii of the proteins, listed in Table 3.1. Coupling the Mackie and Meares correlation with the penetration theory correlation [43,44] enabled the calculation of the protein effective mass transfer coefficients to be 0.0028, 0.0021, and 0.0019 $\mathrm{mms^{-1}}$ for the Fab, the bispecific, and the IgG4, and IgG1 antibodies, respectively. The final model parameters and their derivation are listed in Table 3.4 and Table A1 in the Appendix A.

Table 3.3: Parameters of the system and column characterization.

Parameter	Symbol	Value	Unit	Derivation
Length	L_c	100	mm	Manufacturer
Diameter	d_c	7.7	mm	Manufacturer
Column volume	V_c	4.657	mL	$V_c = \pi \frac{{d_c}^2}{4} L_c$
Bead radius	r_p	37.5	µm	Manufacturer
System dead volume	V_0	0.321 1.972	mL	Tracer, no column installed
Interstitial porosity	ε_b	0.364	–	$\frac{V_b}{V_c}$
Total porosity	ε_t	0.872	–	$\frac{V_t}{V_c}$
Particle porosity	ε_p	0.799	–	$\frac{V_t - V_b}{V_c - V_b}$
Ionic capacity	Λ	1.007	$\mathrm{molm^{-3}}$	$\Lambda = \frac{c_{Cl-} V_{Cl-}}{V_{col}(1-\varepsilon_t)}$
Superficial velocity	u	0.333	$\mathrm{mms^{-1}}$	Experimental setup

The thermodynamic model parameters $\tilde{K}'_{eq,i}$ and $K_{s,i}$ were determined at pH 5.0 and 6.0 for the different model proteins. For each molecule, three linear gradient elution experiments with an altering negative slope of the chloride ion concentration were conducted, as shown in Fig. 3.4. Afterwards, the chromatographic experiments were mechanistically modeled and the counterion concentration at the monomer peak retention extracted from the simulated salt signal. The parameter set was derived by analytical and numerical parameter estimation as depicted in Fig. 3.1 and Section 3.3.5. Fig. 3.2 displays the Yamamoto plot as the semi-logarithmic negative normalized gradient slope $\log(-GH)$ against the chloride counterion concentration at the peak retention $c_{s,i}^R$ for each protein. The analytical linearization and the numerical nonlinear solution of the parameter estimation are indicated for the Fab, the bispecific, the IgG4, and the IgG1 antibody. For all molecules, a negative correlation was observed with differences in the slope and the ordinate of the proteins in a range of 0.152 M to 0.705 M chloride counterion concentration at peak retention. Furthermore, the linear as well as the nonlinear solution of the parameter estimation diverged at increased and decreased gradient slopes compared to the gradient slopes of the model calibration experiments.

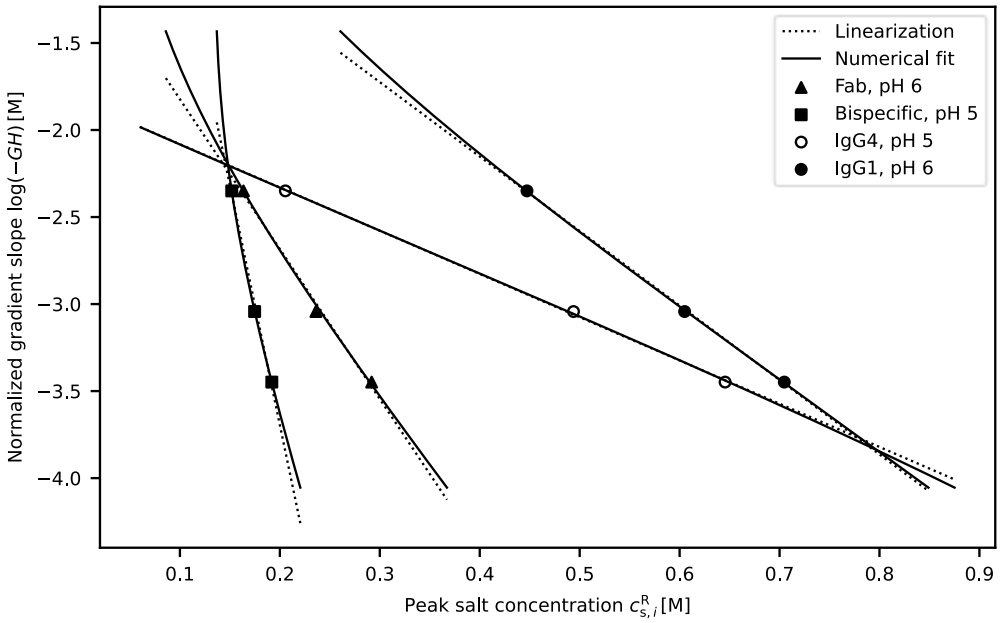

Fig. 3.2: Analytical and numerical isotherm parameter estimation shown for four different antibody formats at pH 5.0 and 6.0.Eq. (3.14) and (3.15) were employed for the analytical and numerical estimation of the isotherm parameters $\widetilde{K}'_{\mathrm{eq},i}$ and $K_{\mathrm{s},i}$. The markers indicate the salt concentrations at the peak maximum of each molecule species $c^{\mathrm{R}}_{\mathrm{s},i}$ to which the linear and numerical model were fitted. The length of the calibration LGEs was set to 10, 20, and 30 CV starting with a chloride ion concentration of 1560 mM to 60 mM at the end of each gradient. Larger $\log(-GH)$ values indicate steeper gradients *vice versa*.

The negative slope of the linear fit enabled the initial estimation of the salt-protein interaction parameter $K_{\mathrm{s},i,0}$ using Eq. (3.14). Further, the y-axis intercept allowed the calculation of the initial equilibrium constant $\widetilde{K}'_{\mathrm{eq},i,0}$ for each protein. Afterwards, the initial parameter pair was refined based on the identical calibration data using Eq. (3.15). The refined isotherm parameters $\widetilde{K}'_{\mathrm{eq},i}$ and $K_{\mathrm{s},i}$ considered the molecule specific size exclusion effect of the resin given by the parameter $K_{\Delta,i\mathrm{s}}$ to compensate for the nonlinear behavior of the Yamamoto plot becoming visible at increased and decreased gradient slopes. Additionally, the lower boundary of the multivariate parameter uncertainty was estimated via numerical optimization. In Table 3.4. all protein specific mass transfer and thermodynamic model parameters are listed including the 95% confidence interval of the lumped equilibrium constant $\mathrm{CI}_{\widetilde{K}'_{\mathrm{eq},i}}[-]$ and the salt-protein interaction parameter $\mathrm{CI}_{K_{\mathrm{s},i}}[\mathrm{mol}^{-1}\mathrm{m}^3]$.

Table 3.4: Protein specific mass transfer and thermodynamic model parameters. The initial thermodynamic parameters were calculated by neglecting the pore accessibility of the resin using the linearization given in Eq. (3.14). Subsequently, the refined thermodynamic parameters were calculated by minimizing the error given in Eq. (3.15) to consider the accessibility of the resin pore system.

	Parameter	**Fab**	**Bispecific**	**IgG4**	**IgG1**
Mass-transfer	Delta distribution coeff.[a] $K_{\Delta,is}$ [−]	0.767	0.797	0.795	0.795
	Eff. mass transfer coeff.[b] $k_{\mathrm{eff},i}$ [$\mu m s^{-1}$]	2.836	1.882	2.143	2.143
Thermo-dynamic	Kinetic coeff.[c] $k_{\mathrm{kin},i}$ [s]	7.124	34.9	33.018	34.705
	Initial equilibrium constant $\widetilde{K}'_{\mathrm{eq},i,0}$ [−]	0.304	0.006	2.516	0.363
	Refined equilibrium constant $\widetilde{K}'_{\mathrm{eq},i}$ [−]	0.566	0.072	3.229	0.569
	CI equilibrium constant $\mathrm{CI}_{\widetilde{K}'_{\mathrm{eq},i}}$ [−]	0.454 – 0.677	0.070 – 0.074	2.985 – 3.477	0.533 – 0.605
	Initial salt-protein interaction parameter $K_{\mathrm{s},i,0}$ [M]	8.612	27.560	2.486	4.279
	Refined salt-protein interaction parameter $K_{\mathrm{s},i}$ [M]	7.312	17.822	2.149	3.817
	CI salt-protein interaction parameter $\mathrm{CI}_{K_{s,i}}$ [M]	6.752 – 7.871	17.701 – 17.943	2.051 – 2.245	3.739 – 3.895

[a] Derived from isocratic pulse experiments, compare Section 3.3.3.

[b] Calculated from empirical correlations listed in the Appendix A, Table A1.

[c] Fitted to peak shape of calibration experiments, compare Section 3.3.5.

The protein specific size exclusion effect $K_{\Delta,is}$ was calculated as the difference between the salt distribution coefficient and the protein distribution coefficient using Eq. (3.16). In general, a large deviation of the apparent pore penetration for the salt component and the proteins was observed. Among the four model proteins the difference between the size exclusion effect of the Fab $(K_{\Delta,is} = 0.767)$ and the bispecific antibody $(K_{\Delta,is} = 0.797)$ was especially pronounced. Similarly, the difference in molecular weight between the Fab and the bispecific antibody of 48 kDa against 202 kDa and more importantly the difference in hydrodynamic radius of 4.1 nm against 6.3 nm was largest between the two model proteins, as listed in Table 3.1. When comparing the initial isotherm parameter pairs $\widetilde{K}'_{\mathrm{eq},i,0}$ and

$K_{s,i,0}$ with the refined parameter pairs $\widetilde{K}'_{eq,i}$ and $K_{s,i}$, an increase of the equilibrium constant and decrease of the salt-protein interaction parameter was indicated. A significant deviation between the initial and the refined isotherm parameters was observed for all model proteins, as the initial parameters laid outside the 95% confidence interval of the refined isotherm parameter estimation.

To investigate the multivariate parameter uncertainty and parameter correlation further, two-dimensional confidence ellipses of the numerically estimated isotherm parameters were calculated, as depicted in Fig. 3.3. The normalized parameters $\widetilde{K}'_{eq,i}$ and $K_{s,i}$ are centered around their refined values with their confidence region being scaled accordingly. The relative size of the 95% confidence region increased in the following order: Bispecific < IgG1 < IgG4 < Fab. The smallest confidence region was observed for the bispecific antibody with a maximum deviation of 0.7% for the $K_{s,i}$ and 2.8% for the $\widetilde{K}'_{eq,i}$ parameter. Meanwhile the largest confidence region was observed for the Fab antibody with a maximum deviation of 7.7% for the $K_{s,i}$ and 19.7% for the $\widetilde{K}'_{eq,i}$ parameter. Overall proteins, the confidence in the $K_{s,i}$ estimation was approximately four times greater compared to confidence in the $\widetilde{K}'_{eq,i}$ estimation. The strength of the correlation between each of the isotherm parameter pairs is described by the slope of the multivariate confidence regions. As an uncorrelated parameter pair would show a horizontal confidence ellipse, parameter correlation was observed for all parameter pairs.

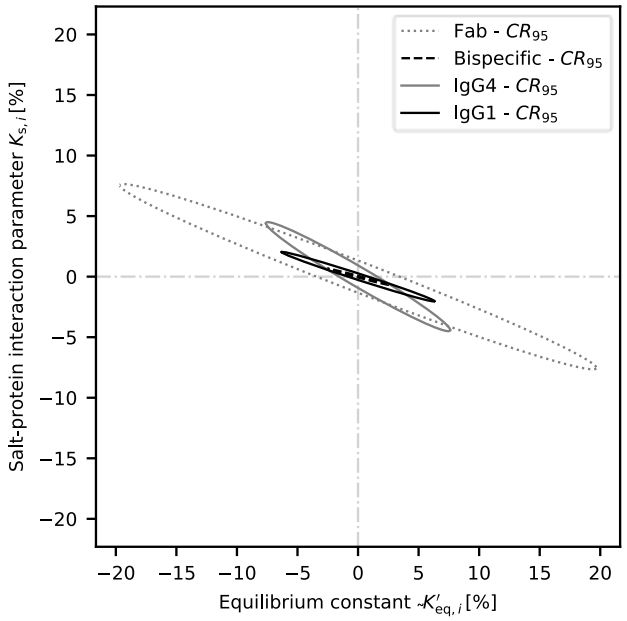

Fig. 3.3: 95% confidence ellipse of the isotherm parameters for four different antibody formats.The relative parameter uncertainty and correlation is estimated from the Fisher information of the numerical parameter estimation. Larger parameter uncertainty regions lead to inaccuracies in model predictions, while steeper slopes of the ellipses indicate increased parameter correlation.

Thereafter, the initial isotherm parameters $\widetilde{K}'_{\mathrm{eq},i,0}$ and $K_{\mathrm{s},i,0}$ as well as the refined parameters $\widetilde{K}'_{\mathrm{eq},i}$ and $K_{\mathrm{s},i}$ were employed in simulating the experimental calibration set. Additionally, the kinetic coefficient $k_{\mathrm{kin},i}$ listed Table 3.4, was fitted to the calibration data. Fig. 3.4 compares the simulated and experimental results of the LGE calibration set, showing a single monomer peak for each molecule. Per row the negative salt LGEs of one of the four proteins are displayed starting with the Fab followed by the bispecific, the IgG4, and the IgG1 antibody. The simulated UV signals were derived according to the Lambert-Beer law using the protein specific theoretical extinction coefficients and their simulated concentration. In general, the retention was shifted to higher ionic strength with increasing gradient slope, also indicated by the Yamamoto plots shown in Fig. 3.2.

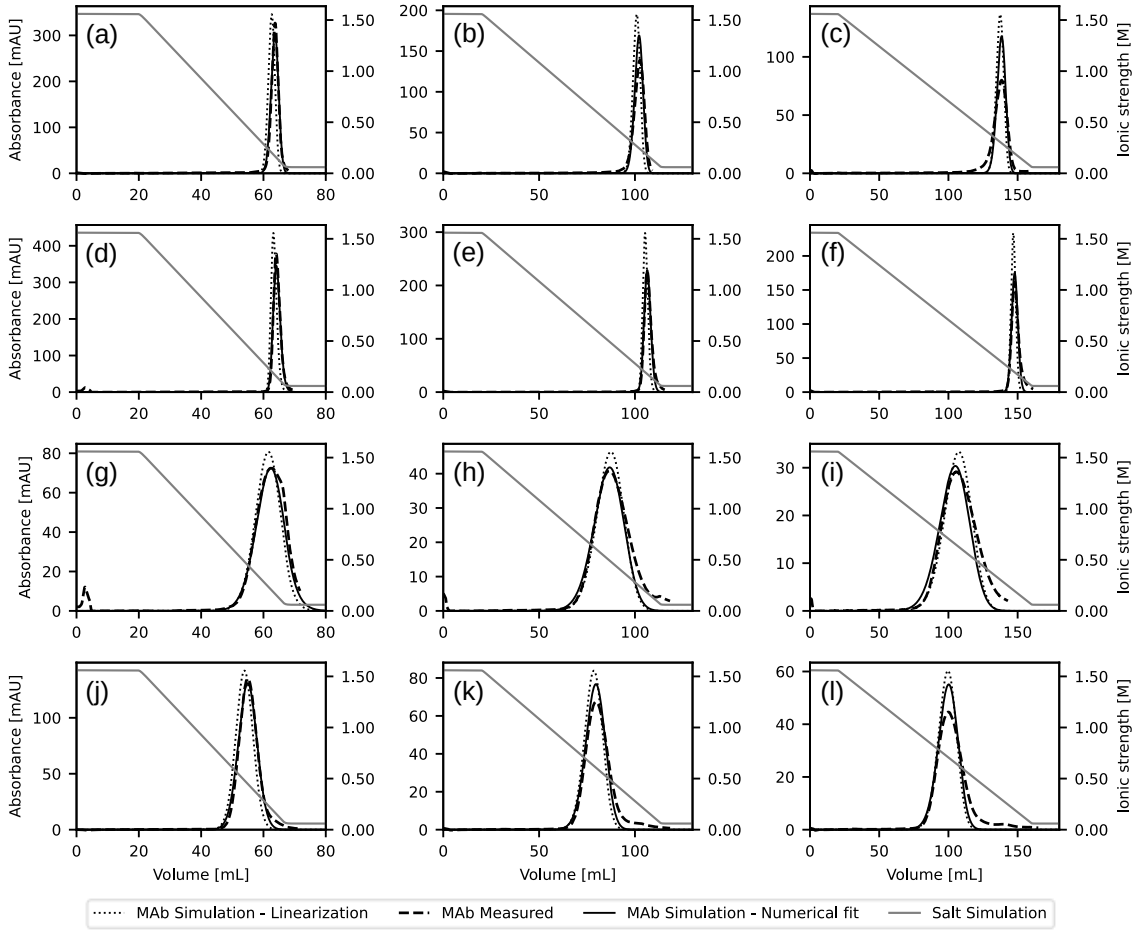

Fig. 3.4: Experimental and simulated chromatograms of the calibration experiments for four different antibody formats at low loading density.The 10, 20, and 30 CV LGE calibration experiments are shown for the Fab (a) to (c), the bispecific (d) to (f), the IgG4 (g) to (i), and the IgG1 (j) to (l) antibody. Two simulated UV signals are depicted per chromatogram using the initial linearized and the numerically refined isotherm parameter pairs. Note, that the upper limits of the left y-axes differ among the experiments.

When examining the simulated and experimental chromatograms closely, an improved fit of the simulated UV signals was observed when using the refined model parameters. Especially, the prediction of the peak retention showed an improved accuracy. Larger deviations between the simulated and experimental data were visible in the elongated peak tail and peak height of the measurement data even when using the refined model parameters. An exception to the experimental peak tailing behavior was visible for the results of the Fab molecule in row one, which showed experimental peak fronting instead.

3.4.3 Model Validation

To validate the calibrated models, three additional experiments were performed per protein beyond the chromatographic conditions that were applied during the LGE calibration. For each molecule, two bind-and-elute step elution and one isocratic flow through experiment were conducted while keeping the processing pH and loading density constant. Fig. 3.5 depicts the validation of the mechanistic model of all proteins by comparing the initial and refined model simulations with the experimental results. In every row a low salt SE is shown on the left side followed by the IE experiment in the middle and a high salt SE on the right side. Per row the chromatograms of one of the four proteins are displayed starting with the Fab followed by the bispecific, the IgG4, and the IgG1 antibody. The elution of the bispecific mAb during the high salt SE shown in Fig. 3.5 (f) did not reach baseline level as the protein binding was increased compared to the other antibodies.

Overall, the peak retention and peak shape were predicted accurately by the TDM. The retention of the molecules increased with elevated salt concentration. The incorporation of the molecule and column specific size exclusion effect using the refined model parameters led to improved model accuracy especially at the increased and decreased gradient slopes that were encountered during step and isocratic elution experiments. In particularly, the prediction of the peak retention and shape of the IE experiments, shown in Fig. 3.5 (b), (e), (h), and (k) was enhanced when using the refined isotherm parameters. The biggest improvement using the refined isotherm parameters was observed for the high salt SE of the bispecific molecule shown in Fig. 3.5 (f). On the other hand, the largest difference between the experimental and simulated elution profile was seen for the high salt SE of the Fab molecule shown in Fig. 3.5 (c) while still being reduced by using the refined model parameters.

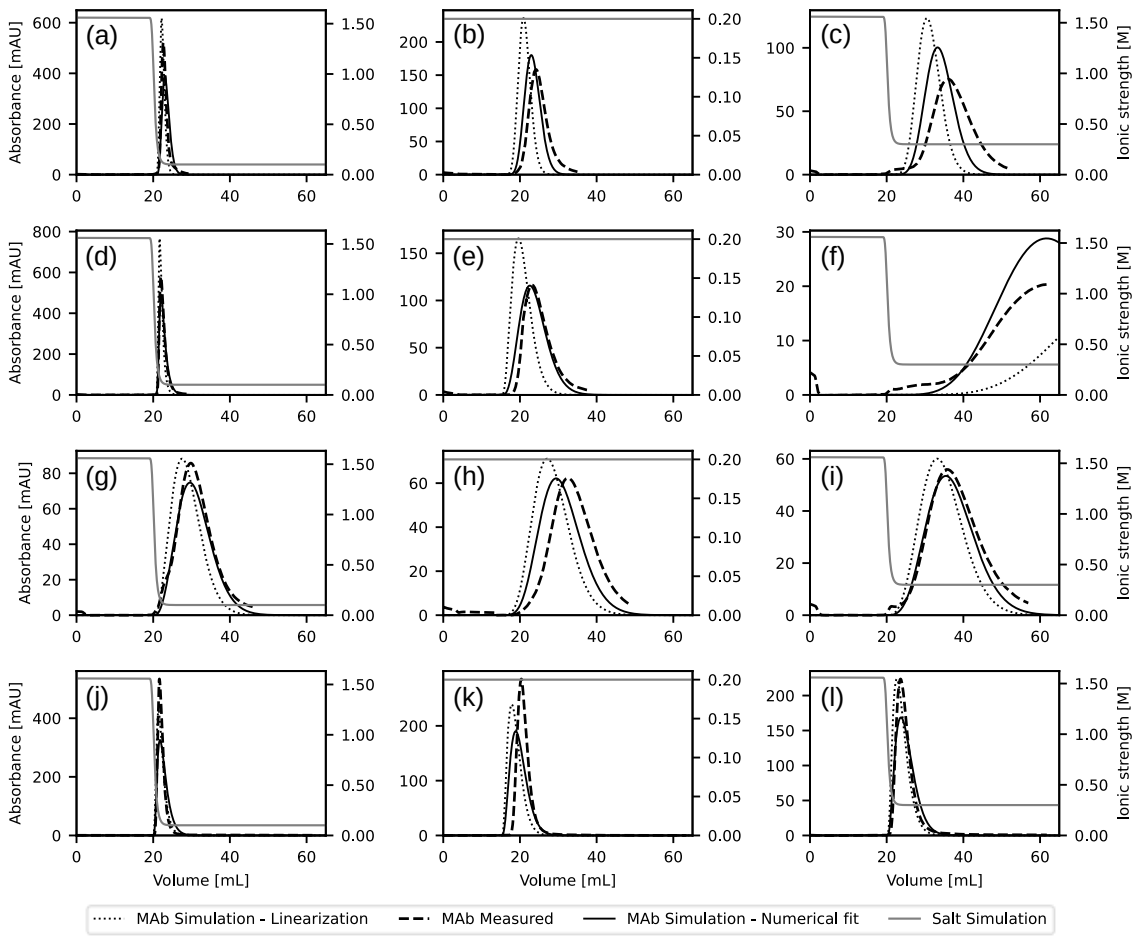

Fig. 3.5: Experimental and simulated chromatograms of the validation experiments for four different antibody formats at low loading density.The low salt SE, IE, and high salt SE experiments are shown for the Fab (a) to (c), the bispecific (d) to (f), the IgG4 (g) to (i), and the IgG1 (j) to (l) antibody. Two simulated UV signals are depicted per chromatogram using the initial linearized and the numerically refined isotherm parameter pairs. Note, that the upper limits of the left and right y-axes differ among the experiments while the initial ionic strengths for low- and high salt SE are identical.

3.5 Discussion

3.5.1 Model Development and Calibration

In this study the mechanistic model development and calibration for four antibodies in anionic mixed-mode chromatography is shown. The model calibration required the determination of the mass transfer and the thermodynamic parameters, of a physicochemical varying set of proteins, listed in Table 3.1. Initial modeling efforts indicated difficulties in calibrating the 11 model parameters of the highly parameterized MMC isotherm developed by Nfor et al. [54]. The difficulties in the model calibration were traced back to structural

and practical parameter non-identifiability [151] of the isotherm model. No global parameter optimum could be identified when curve fitting to the experimental UV signals as the parameter set appeared to be partially correlated and the optimization problem had too many degrees of freedom for a limited amount of data. A similar observation has already been made for the SMA isotherm leading to model overfitting due to parameter correlation [145,152,153] and was overcome by an initial analytical parameter determination [135]. To resolve the difficulties in the MMC isotherm calibration, a systematic model reduction was performed employing multiple principles reviewed by Rischawy et al. [145]. The principles target the reduction of model complexity and include the removal, lumping, and fixation of model parameters that were conducted in Section 3.2.2. Accordingly, the complexity of the MMC isotherm model was reduced to only three parameters while still being able to describe the experimental observations depicted in Fig. 3.4. Furthermore, the model reduction allowed an initial analytical parameter determination of the modified MMC isotherm in accordance with the SMA parameter estimation, hence increasing the identifiability of adsorption isotherm parameters. Lastly, the gradient based experimental design enabled a standardized workflow, which substantially reduced the MMC modeling effort.

To validate the standardized MMC modeling workflow, four multimodal chromatographic models were calibrated at two pH environments. The linear relationship of the Yamamoto plots shown in Fig. 3.2 demonstrates the applicability of the Yamamoto method for the MMC isotherm parameter determination and is a clear indication of a protein ligand interaction governed by hydrophobic attraction as has been found by Xiao et al. [136]. As the column characterization experiments revealed size exclusion effects of the Capto adhere resin towards all investigated antibody formats, the final isotherm parameters were refined numerically to incorporate column specific effects as proposed by Creasy et al. [146]. Recently, similar observations had been made by Voitl et al. [61] and Roberts et al. [154] when investigating mass transfer phenomena of multimodal resins. The refined isotherm parameters of the four model proteins listed in Table 3.4 were in good agreement with published results of the Capto adhere resin by Nfor et al. [54]. Interestingly, there was no clear correlation between the molecule size or conformational flexibility in regard to the strength of the hydrophobic interaction as was suggested by Brian et al. [129]. Indeed, the salt-protein interaction parameter $K_{s,i}$ was highest for the comparably large and flexible bispecific antibody format. On the other hand, the $K_{s,i}$ parameter of the smaller and less flexible Fab was at least two times greater than the $K_{s,i}$ parameter of the IgG1 or IgG4 antibody indicating a complex binding mechanism driven by molecule specific surface characteristics. The model uncertainty was acceptable, when investigating the 95% confidence ellipses of the four model parameter pairs shown in Fig. 3.3, with the exception of the equilibrium constant $\widetilde{K}'_{eq,i}$ of the Fab molecule showing a significantly increased uncertainty compared to the other molecule parameters. The high linearity of the Yamamoto plot paired with the comparably small isotherm parameter uncertainty and correlation indicates that three LGE experiments were sufficient for the parameter determination. One the other hand, the visible deviations from linearity at increased and decreased gradient slopes underline the necessity of incorporating size exclusion effects through numerical estimation. It appears that the parameter determination via linearization is experimentally restricted.

Further increasing the gradient slope of the calibration experiments, which is required for predicting SE experiments leads to inaccuracies in determining the salt concentration at peak retention. *Vice versa* the slope of the calibration experiments would have to converge towards zero to reliable predict IE experiments.

Finally, the initial and the refined model parameters were employed in the simulation of the chromatographic experiments as shown in Fig. 3.4. The similarity between the calibrated model and the experimental data underlined the ability of the modified isotherm to describe the elution behavior in MMC at the investigated process conditions. For all model proteins and gradient slopes an improved accuracy of the predicted peak retention and shape was observed when using the refined model parameters, although not as distinct as when using the refined parameters in the validation experiments. The improved model accuracy demonstrates that column specific effects need to be incorporated in the Yamamoto method when modeling chromatographic systems showing size exclusion behavior.

3.5.2 Model Validation and Application

A reliable way of proving model accuracy is by external validation of uninvolved data. Mechanistic models are able to capture fundamental phenomena, which should allow a prediction beyond the experimental conditions applied for model calibration. In this study the predictive capabilities of a modified MMC isotherm model were validated by simultaneously varying the MMC operation mode and ionic strength. Furthermore, a comparison between the model predictions was made when using initial as well as refined model parameters. The model validation was successful for a broad variety of commercially relevant product classes when predicting different process conditions including flow through purification, as shown in Fig. 3.5. In general, the peak retention as well as the peak shape were predicted accurately considering the simplicity of the applied mechanistic model. The incorporation of size exclusion effects through refined model parameters did yield improved model predictions of the validation experiments. One exception to the model performance was visible in the high-salt step elution experiment of the Fab molecule shown in Fig. 3.5 (c), which could only be resolved in part by using the refined parameter set. As the uncertainty of the Fab isotherm parameter was largest, it is likely that ill-defined parameter values are the reason for the deviating model validation. Furthermore, similar effects, which led to an experimental fronting in the model calibration could be the cause of deviating results in the model validation. These effects could include but are not limited to neglected contributions of product specific impurities, molecule unfolding due to hydrophobic interaction [155], or multiple molecule binding orientations.

Nevertheless, the main focus of this study was the implementation of a standardized modeling workflow for predicting anionic MMC-mAb interaction at commercially relevant process conditions. The generic experimental setup, little material consumption of less than 30 mg per mAb, the efficient model calibration, and accurate chromatographic predictions, enable a model-guided process design during the early stages of process development.

Furthermore, our model could be applied to multicomponent systems to address complex development tasks, including the optimization of impurity removal or robustness analysis. Overall, our model-guided development approach could yield the widespread deployment of multimodal purification in DSP leading to significant savings in development time and resource consumption. Finally, the large similarity between the implemented MMC model and existing hydrophobic interaction chromatography (HIC) isotherms suggests that the standardized modeling workflow could be transferred to HIC processes.

3.6 Conclusion

In this study, the chromatographic behavior of four therapeutic antibody formats in anionic mixed-mode chromatography has been mechanistically modeled. Initially a systematic model reduction of an existing MMC isotherm was accomplished using fundamental knowledge of preparative chromatography. A comprehensive data set revealed a protein adsorption mediated by hydrophobic effects within the investigated pH range, which led to a reduction of the model complexity while increasing the identifiability of the isotherm parameters. The relative simplicity of the modified isotherm model enabled an analytical parameter determination and provided meaningful mechanistic insight into the multimodal adsorption process. The incorporation of size exclusion effects required a numerical parameter estimation that led to a refined model performance. The comparably small parameter uncertainty and correlation of the refined model emphasized the validity of the model reduction and the selected calibration experiments while underlining the necessity of incorporating molecule specific size exclusion effects for model calibration. Lastly, the calibrated mechanistic models were successfully validated by predicting different MMC operation modes at varying ionic strengths for all therapeutic antibody formats.

Our mechanistic modeling approach offers a standardized workflow for predicting the elution profiles of therapeutic antibodies in anionic MMC. The workflow excels in a reduction of material consumption, as well as experimental and computational burden. Further, the straightforward experimental design has the potential to accelerate the development of complex MMC purification processes in the biopharmaceutical DSP development.

Acknowledgements

The authors would like to thank Simon Kluters, Joey Studts, Federico Rischawy, and the rest of the BI modeling team for the constant support as well as the long and fruitful discussions.

4 Antibody Sequence-Based Prediction of pH Gradient Elution in Multimodal Chromatography

Rudger Hess[1, 2], Jan Faessler[2], Doil Yun[2], David Saleh [2], Jan-Hendrik Grosch[2], Thomas Schwab[2], Jürgen Hubbuch[1,*]

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

[2] DSP Development, Boehringer Ingelheim Pharma GmbH & Co. KG, Biberach, Germany

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

Journal of Chromatography A (1711), 2023, p. 464437

DOI: 10.1016/j.chroma.2023.464437

Abstract

Multimodal chromatography has emerged as a promising technique for antibody purification, owing to its capacity to selectively capture and separate target molecules. However, the optimization of chromatography parameters remains a challenge due to the intricate nature of protein-ligand interactions. To tackle this issue, efficient predictive tools are essential for the development and optimization of multimodal chromatography processes. In this study, we introduce a methodology that predicts the elution behavior of antibodies in multimodal chromatography based on their amino acid sequences. We analyzed a total of 64 full-length antibodies, including IgG1, IgG4, and IgG-like multispecific formats, which were eluted using linear pH gradients from pH 9.0 to 4.0 on the anionic mixed-mode resin Capto adhere. Homology models were constructed, and 1312 antibody-specific physicochemical descriptors were calculated for each molecule. Our analysis identified six key structural features of the multimodal antibody interaction, which were correlated with the elution behavior, emphasizing the antibody variable region. The results show that our methodology can predict pH gradient elution for a diverse range of antibodies and antibody formats, with a test set R^2 of 0.898. The developed model can inform process development by predicting initial conditions for multimodal elution, thereby reducing trial and error during process optimization. Furthermore, the model holds the potential to enable an *in silico* manufacturability assessment by screening target antibodies that adhere to standardized purification conditions. In conclusion, this study highlights the feasibility of using structure-based prediction to enhance antibody purification in the biopharmaceutical industry. This approach can lead to more efficient and cost-effective process development while increasing process understanding.

Keywords: *Multispecific monoclonal antibody (mAb) formats; Structure-function analysis; Quantitative structure-activity/property relationship (QSAR/QSPR); In silico process development; Downstream manufacturability assessment*

4.1 Introduction

At present, the monoclonal antibody (mAb) production relies on chromatographic purification, which is integrated into a templated platform process [9]. Multimodal chromatography has emerged as a highly selective separation method compared to using single-mode interaction resins [28,61]. Specifically, the application of multimodal chromatography in the primary capture from harvested cell culture fluid and subsequent polishing steps has demonstrated its effectiveness in separating process and product-related impurities [51,156,157]. The enhanced selectivity of multimodal resins stems from orthogonal physicochemical interactions with the molecule surface [16,27]. In this context, ligands functionalized with electrostatic, hydrophobic, aromatic, and/or hydrogen bonding groups are commonly used, as illustrated by the Capto adhere ligand in Fig. 4.1 (c) [68].

Owing to the intricate multimodal interaction, a broad range of operating conditions must be assessed, as the purification is constraint to a narrow, molecule-specific parameter window of buffer conductivity, pH, modulator concentration, and temperature compared to unimodal chromatography, which can restrict molecule manufacturability [62,66]. To support process development, extensive research has been conducted to enhance process understanding by examining multimodal protein-ligand interaction alongside efficient screening methodologies. Macroscopic effects have been explored through batch and dynamic-binding experiments, which were described using thermodynamic models [54,84,158]. To improve the resolution of macroscopic observations, domain contributions of multimeric proteins and the impact of amino acid substitutions in homologous protein libraries were investigated [70,159–161]. For a molecular level of detail, spectroscopic evaluation of protein-ligand pairs combined with protein labeling techniques such as atomic force microscopy, nuclear magnetic resonance spectroscopy, or mass spectrometry have been employed [64,162–164]. Additionally, molecular dynamics simulations and the calculation of theoretical physicochemical properties shed light on the complex protein-ligand interactions [69,165–167]. To bridge the gap between the molecular-level of detail and macroscopic observations, quantitative structure-property relationship (QSPR) models were developed, restricted by the amount of available data [57,71,159,168–172]. Conversely, automated screening setups using liquid-handling stations or controlled pH-gradients have been utilized to support process development or assess molecule manufacturability profiles [58,173].

Despite recent advancements in increasing the process understanding of multimodal mAb purification, the complexity and sensitivity of multidomain proteins towards multimodal interactions limits their widespread application in biopharmaceutical development [23,29]. Predictive tools to facilitate the integration of multimodal chromatography into the mAb purification platform and assess molecule manufacturability remain scarce, while demonstrating significant success in other areas of the development cycle, such as candidate screening or the prediction of agglomeration propensity during formulation development [174,175].

In this study, we developed a QSPR model to predict mAb retention in multimodal chromatography during linear pH gradient elution, as depicted in Fig. 4.1. Initially, a comprehensive dataset was acquired comprising antibody-specific descriptors calculated from homology models and chromatographic pH retention for 64 full-length mAbs, including multiple IgG-like derivatives. Subsequently, an empirical model was developed using Gaussian process regression (GPR) and thoroughly validated.

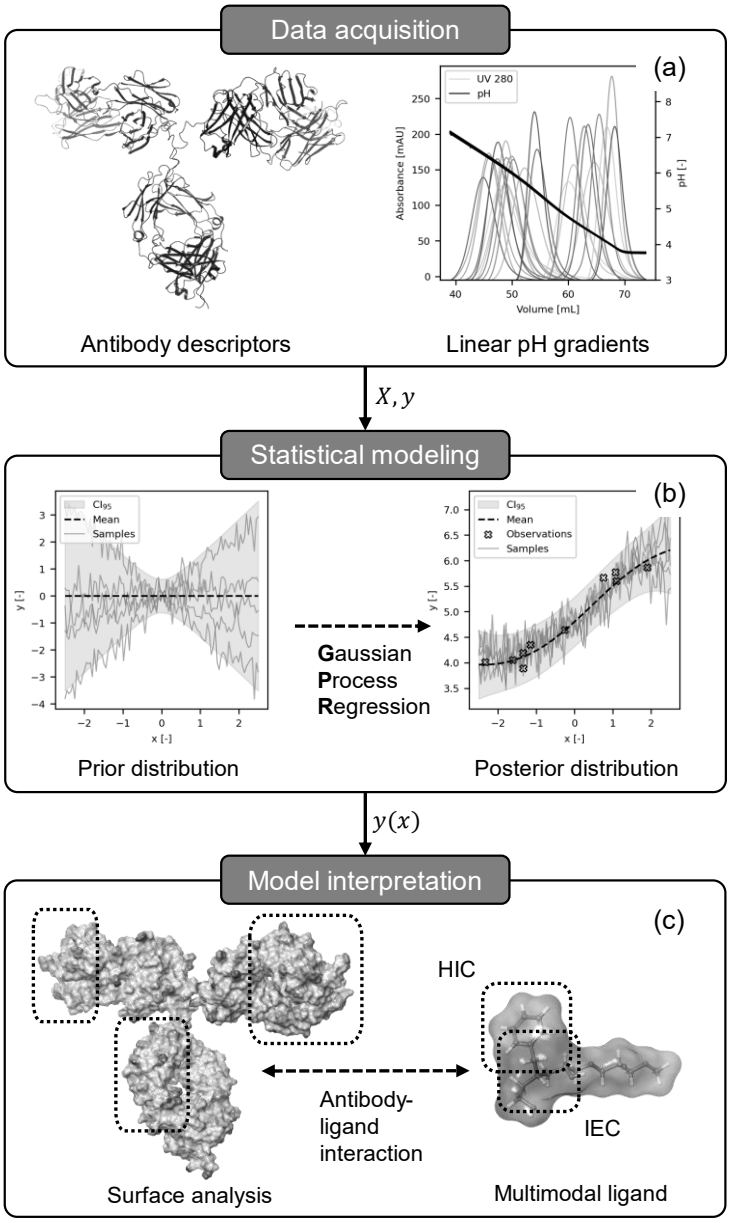

Fig. 4.1: QSPR modeling workflow.A three-step process is shown that includes a) the data acquisition for model training and testing, b) the statistical modeling using Gaussian process regression, and c) the model interpretation by evaluation of the identified descriptor-output correlations. The antibody structure depicts a modification of the PDB entry 1HZH [176].

Finally, the GPR model was interpreted, providing insights into the multimodal interaction mechanism. The validated model can be employed to support process development and enable a candidate manufacturability assessment based solely on sequence information. Moreover, the mechanistic insights can contribute to the development of advanced adsorption models, transitioning from a macroscopic process understanding to the molecular level.

4.2 Material and Methods

4.2.1 Chromatography Resin, Buffers, and Molecules

In this study, the multimodal strong anion exchanger Capto adhere (Cytiva, Marlborough, USA) was utilized during the chromatographic experiments. A prepacked Capto adhere HiScreen column (7.7x100 mm, Cytiva) with a column volume (CV) of 4.7 mL was employed, as detailed in Section 4.2.2. The resin surface is functionalized with the N-Benzyl-N-methyl ethanol amine ligand, as depicted in Fig. 4.1 (c). This ligand exhibits multimodal functionality due to its capacity for ionic interaction, hydrogen bond formation, and hydrophobic interactions [68].

All buffer substances were purchased from Sigma-Aldrich Co LLC (Saint Louis, USA), while ultrapure water was filtered with the Milli-Q Advantage A10 (Merck Millipore, Burlington, USA) water purification system. The linear pH gradients necessitated a multicomponent buffer system compatible with anion exchange chromatography. Consequently, an anionic multicomponent buffer was selected to avoid the introduction of unspecified counterions while providing a broad buffer capacity within the pH range of 9.0 to 4.0 [177]. The buffer system was adapted from Kröner and Hubbuch [137] and consists of 9.1 mM 1,2-ethanediamine, 6.4 mM 1-methylpiperazine, 13.7 mM 1,4-dimethylpiperazine, 5.8 mM bis-tris, and 7.7 mM hydroxylamine. In addition, 125 mM sodium chloride and 75 mM hydrochloric acid were incorporated, resulting in a total of 200 mM chloride counterions and a conductivity of 20 mS/cm. The addition of sodium chloride was required to achieve mixed-mode behavior and increase protein solubility [178]. Furthermore, the increased conductivity values enabled the augmentation of cation exchange elution pool as load material, which is regularly employed prior to the salt tolerant Capto adhere resin within the antibody purification process [42]. Thereafter, the equilibration and the elution buffer were titrated to pH 9.0 and 4.0 using 1 M sodium hydroxide. Other buffers used in the chromatographic experiments included 1 M acetic acid for column regeneration, 1 M sodium hydroxide for column cleaning, and 20% ethanol for column storage.

The study involved 64 full-length IgG derivatives (Boehringer Ingelheim Pharma GmbH & Co. KG, Biberach, Germany), comprising 62 human-origin and 2 humanized murine-origin antibodies. These antibodies displayed an extensive range of physicochemical parameters, as evidenced by their widely distributed elution behavior, shown in Fig. 2a. The antibody set included 33 IgG1s, 20 IgG bispecifics with two single-chain fragment variables

(scFv) appended to each heavy chain C-terminus (IgG(H)-scFv), 8 IgG4s, 2 Knob-in-Hole bispecifics (KiH), and 1 KiH trispecific with a single scFv attached to the C-terminus of the Hole chain (KiH-scFv). The antibody expression was achieved using a stably transfected Chinese hamster ovary cell line, followed by capture through Protein A affinity chromatography. After neutralizing to pH 5.5 and sterile filtration using a 0.2 µm filter, the protein solutions were frozen at -70°C. Upon thawing, the final load material was adjusted to a concentration of 5 g/L, as determined by a NanoDrop 2000c spectrophotometer (Thermo Fisher Scientific, Waltham, USA). Prior to sample application, the load material underwent buffer exchange into the equilibration buffer using a 5 mL HiTrap Desalting column (Cytiva) according to the manufacturer's instructions.

4.2.2 Linear pH Gradient Elution

In this study, all 64 antibodies were eluted by a linear pH gradient using the multicomponent buffer system described in Section 4.2.1. The chromatographic experiments were conducted with an ÄKTA Avant 25 (Cytiva) preparative chromatography system, controlled by the Unicorn 7.5 (Cytiva) control software, and maintained at a residence time of 5 min. Initially, the column was equilibrated for 3 CV with the pH 9.0 buffer. The equilibration was followed by column loading with the antibody solution up to a loading density of 1.0 g/L. Subsequently, the antibodies were eluted by linearly decreasing the pH from 9.0 to 4.0 within 10 CV using the elution buffer, while maintaining a constant conductivity of 20 mS/cm, as illustrated in Fig. 4.1 (a). The retention times of the molecules were determined by measuring the first moments of the elution peaks through the UV trace at a wavelength of 280 nm. Subsequently, the corresponding pH values at peak retention were determined by correcting the online pH trace of the chromatography system with the offline pH measurement of the equilibration and elution buffer, as well as accounting for the pH sensor dead volume. Following column elution, a 4 CV column regeneration step, a 5 CV cleaning in place procedure, and 4 CV column storage step were appended.

4.2.3 Antibody Homology Modeling

The prediction of the antibody structures was achieved through homology modeling, which was required for the subsequent calculation of physicochemical descriptors, as previously described by our group [179]. Hereto, the molecular modeling and visualization environment Maestro Bioluminate 4.9 (Schrödinger Inc., New York, USA) was employed for structure prediction, evaluation, and model refinement.

The initial homology modeling workflow was adapted from Zhu et al. [180] and comprises an automated five-stage process, which includes: (1) framework and complementary-determining region (CDR) template selection; (2) variable region model grafting; (3) CDR loop modeling and sidechain prediction; (4) full-length antibody modeling; (5) energy

minimization. Within the workflow, the antibody numbering scheme Enhanced Chothia [181] was utilized. The modified crystal structure of a human IgG1, 1HZH [176], shown in the Fig. 4.1 (a/c), served as a full-length template for all molecules except for the IgG4 subtypes, which employed the human IgG4 crystal structure 5DK3 [182] instead. For structural prediction of the complex bi- and trispecific formats, intra- and intermolecular linkers were grafted using homology modeling and *ab initio* prediction to append the independently modeled scFv domains to the full-length mAb structures [183]. Hereafter, the initial homology models were further refined according to the protocol of Sastry et al. [184]. In brief, the protocol includes preprocessing steps to modify and validate the hydrogen network, bond order assignment, as well as atom naming and numbering. The preprocessing is followed by energetic optimization of terminal hydrogen atoms from the amino acid side chains and the assignment of protonation states of ionizable groups using PROPKA3 [185]. Lastly an all-atom energy minimization is conducted using the OPLS4 forcefield with nonhydrogen atoms being constrained to a root-mean-square deviation of 0.3 Å [186].

4.2.4 Antibody-Specific Descriptor Calculation

Following the antibody homology modeling, physicochemical descriptors were calculated from the protein structures using the molecular modeling and visualization environment Bioluminate 4.9 (Schrödinger Inc., New York, USA). The utilized descriptor set comprises 165 unique features derived from first principal models, as well as parameterized empirical models [110]. Moreover, the descriptor set can be subdivided into sequence-based descriptors (n=69) devised from bioinformatic scales, structural descriptors (n=59) encoding for geometric and electrostatic properties of the molecule, and patch-specific descriptors (n=37) calculated from the hydrophobic and electrostatic energy of the protein surface. Herein, the surface hydrophobicity is calculated by employing the atomistic Wildman and Crippen logP parameters [187], whereas the electrostatic surface potential is calculated from partial charges based on the OPLS4 [186] forcefield as described by Sankar et al. [188]. Furthermore, proximal hydrophobic and electrostatic surface characteristics are combined into aggregation propensity descriptors. The final patch descriptors are derived by binning the calculated surface properties into quantifiable features based on their interaction type (positive, negative, hydrophobic), size, intensity, and number.

To increase the resolution of the descriptor set further, a region-specific subset (n*=31) of the initial descriptor set is selected and calculated for 37 subdomains of the antibody structure. The antibody-specific subdomains comprise the light and heavy chain variable regions (VL, VL_Fv, VH, VH_Fv), emphasizing the complementarity-determining regions (CDR, CDRL, CDRH) and framework regions (FR, FRL, FRH), which consist of individual loops (L1, L2, L3, H1, H2, H3), and frameworks (LFR1, LFR2, LFR3, LFR4, HFR1, HFR2, HFR3, HFR4). On top of the variable region, the antibody constant regions (CL, CH1, CH2, CH3) and the hinge region (Hinge) are considered. Furthermore, the regional

descriptors are extended by seven custom regions to account for the fragment variable (Fv), the fragment antigen binding (Fab), the fragment crystallizable (Fc), as well as the sum of the constant regions (CR). Additionally, the single-chain fragment variable regions (scFv, VLscFv, VLscFv) of the bi- and trispecific formats are considered. The final descriptor set comprises 1312 features per molecule including the initial 165 global descriptors (All) and 1147 local descriptors, providing detailed information about the physicochemical topology of the IgG-like structures. Given the large quantity of the descriptors, the descriptor naming scheme combines the descriptor location, interaction type, and binning strategy, as exhibited in Section 4.3.2.

4.2.5 QSPR Model Development and Evaluation

A multivariate regression model was established to predict the linear pH gradient retention from the antibody-specific descriptor set. Due to the high dimensionality of the regression problem, with $N = 1312$ features per antibody, paired with the comparably small data set size of $M = 64$ observations, descriptor preprocessing, dimensionality reduction, and model evaluation were required. The QSPR workflow was developed with and evaluated by Python 3.9.12 in conjunction with the machine learning package scikit-learn 1.0.2 [189]. GPR was utilized to predict nonlinear relationships between the target vector $y = \{y^{(m)}\}_{m=1}^{M}$ and the feature matrix $X = \{X^{(n)}\}_{n=1}^{N}$, while providing a heteroscedastic uncertainty estimation [116,190]. GPR is based on Bayesian inference and involves prior assumptions regarding the underlying target function $y(x)$ that can later be updated in course of the Bayesian update rule, shown in Eq. (4.1).

$$P(y(x)|\mathcal{D}) \propto P(y|y(x), X)P\big(y(x)\big) \tag{4.1}$$

Within the Bayesian framework, the model predictions are derived as the posterior $P(y(x)|\mathcal{D})$, which is a gaussian distribution of functions conditioned to fit the training data $\mathcal{D} = \{y, X\}$, as depicted in Fig. 4.1 (b). During model training, the prior $P\big(y(x)\big)$ that defines the similarity and the smoothness between the observations is conditioned by maximizing the likelihood $P(y|y(x), X)$ of the mean and the variance from the posterior distribution to reflect all training data. In this study, the prior is derived as a mixed covariance function by multiplying a linear kernel with a Matérn class kernel and subsequent addition of a white noise kernel [116,189,191]. The addition of a noise kernel is necessary to avoid model overfitting by specifying the uncertainty of the measured data as visualized by the amplitude of the posterior distribution depicted in Fig. 4.1 (b). Subsequently, the model is conditioned by minimizing the log(marginal likelihood) (LML) of the posterior distribution using the L-BFGS-B algorithm [192].

The QSPR workflow is initiated by data preprocessing, where empty, positional, non-informative, and redundant descriptors are discarded. Additionally, several operations are

performed on the descriptors to account for the structural diversity of the IgG-like molecules. The regional antibody descriptors are multiplied by the frequency of the given region within the multimer protein, sparse antibody regions imputed as zero, and the descriptor regions calculated for both KiH knob and hole chains, averaged based on the analysis of Parasnavis et al. [70]. The data set is then randomly split into 80% training data and 20% test data. Lastly, the descriptors are scaled by their standard deviation (SD) and centered based on the training data.

After data preprocessing, dimensionality reduction is performed by removing invariant descriptors, which decreases the risk of model overfitting and increases model interpretability [193]. Low variance ($cov(x,x) \leq 0.01$) features are discarded, and the remaining descriptors are sorted based on the results of a F-test from a univariate linear regression model with the target variable. Hereafter, collinear features are removed based on Pearson correlation ($\rho \geq 0.80$), and ten highest scoring features are selected following the F statistics. Lastly, recursive feature elimination (RFE) is conducted by iteratively removing the lowest-ranked features according to feature permutation importance [194]. Permutation importance is defined as the average increase in model deviation when accessing the model performance after shuffling a single feature one hundred times while keeping the remaining features constant. At each iteration of the RFE procedure, the LML of the current model and the mean absolute error (MAE) of leave-one-out cross-validation are calculated to identify the overall best model.

The last step of the QSPR workflow comprises model evaluation to increase the understanding of the underlying adsorption mechanisms of antibodies in multimodal chromatography, as visualized in (4.1) (c). The model evaluation includes an assessment of the overall model reliability and performance. On top of that, an investigation of the feature interdependence, sensitivity, and their contribution to the model predictions is conducted, to enable mechanistic interpretability of the model [96]. The model performance is assessed through inspection of goodness of fit to the training data and goodness of prediction of the test data, including an estimation of the model 95% confidence interval per observation. Furthermore, fivefold cross-validation with ten repetitions is employed for internal validation of the training data. On the other hand, model reliability is analyzed by y-scrambling the full data set one hundred times with subsequent calculation of the MAE from leave-one-out cross-validation [195]. Feature interdependence is evaluated by investigating the pairwise relationships between model features, as well as the target variable. Finally, the feature sensitivity and contribution are assessed by means of feature permutation importance and partial dependence towards the model prediction [111,194,196].

4.3 Results and Discussion

4.3.1 Elution Behavior of Antibody Formats

In the course of this study, a large and structurally diverse set of IgG-like molecules was examined. All 64 full-length IgGs could be eluted from the anionic mixed-mode resin Capto adhere during linear pH gradients from pH 9.0 to 4.0 at a constant conductivity of 20 mS/cm. Fig. 4.2 depicts the first moments of the elution peaks from each molecule. Upon inspecting the elution distribution, a bimodal trend is apparent separating the molecules in two groups of antibodies. The first group elutes at a lower pH range (pH 4.00-4.99, m=20) compared to the second group that elutes at a higher pH range (pH 5.71-6.81, m=43). In contrast to the overall distribution mean of pH 5.72, the lower and the higher elution groups are centered around pH 4.59 and 6.26, with a single IgG1 laying in-between both groups at a pH of 5.33.

To investigate the elution behavior further, Fig. 4.2 (a) compares the retention of different antibody subclasses and formats that were analyzed in this study, as detailed in Section 4.2.1. The IgG1 subclass depicts the most abundant format within the data set and exhibits the broadest retention distribution across both pH groups. Interestingly, the distribution of the structurally homologue IgG1 antibodies follows the same trend as the full data set, being centered in the higher pH group and skewed towards lower pH values. This observation emphasizes that multimodal elution is not necessarily dependent on the overall size or shape of the molecule but physicochemical properties that are distributed within a structural homologue set of proteins, as observed in multiple studies [159,160,171]. The IgG-scFv formats exhibit a similar trend, while displaying a narrower pH distribution with only two molecules eluting in the lower pH group despite having two additional scFv regions attached to the Fc domain. Moving forward, the IgG4 antibodies exclusively elute in the lower pH group, whereas all KiH formats elute at the upper boundary of the observed pH values with a maximum elution pH of 6.79. As only three molecules are associated with the KiH format, no inference towards this antibody class can be conducted. On the other hand, the main structural distinction between the IgG4 subclass and the other antibodies is given by the Fc domain, as the remaining formats share similar IgG1 backbones. The difference in the elution behavior arising from deviating Fc regions leads to the assumption that multiple binding domains on the whole antibody surface exist, as reported in numerous studies conducted by the Cramer lab involving cationic mixed-mode resins [70,161,163,164,166,167]. In general, IgG4 backbones exhibit fewer acidic residues leading to a lower pI, as well as increased surface hydrophobicity in comparison to IgG1 backbones [197]. These two characteristics align with the experimental observation of the IgG4 formats to elute in the lower pH group and the mechanism of hydrophobic charge induction chromatography (HCIC) [165,198]. According to the HCIC mechanism, molecules adsorb through hydrophobic attraction and desorb with increasing electrostatic repulsion in course of a pH modification. The lower pI of the IgG4 formats would result in a reduced positive surface charge at an acidic pH environment, which in turn would lead to a reduced repulsion towards the anionic multimodal ligand.

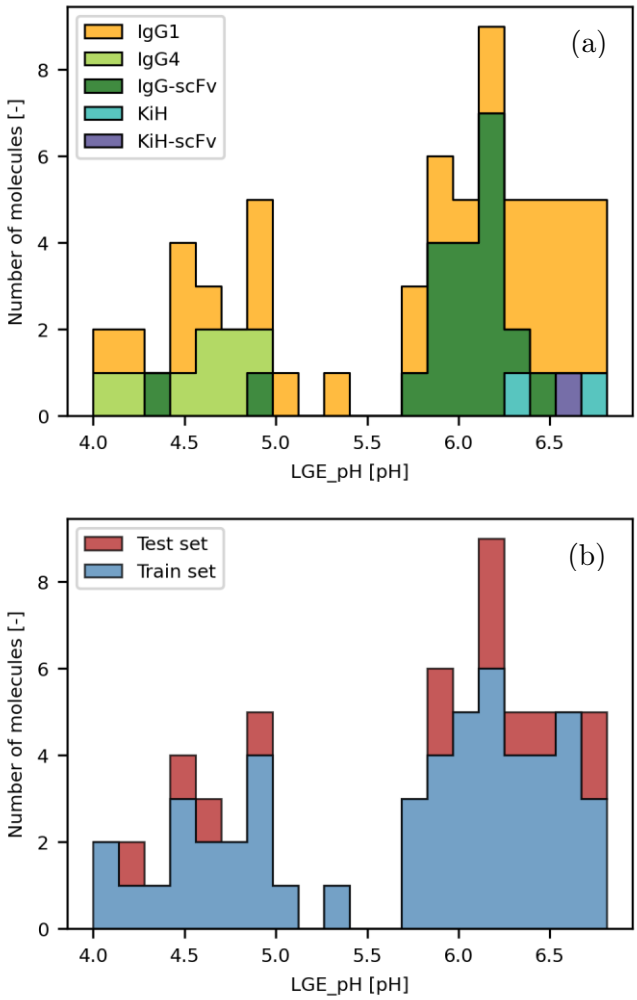

Fig. 4.2: Distribution of pH retentions derived from linear pH elution experiments.a) Histogram illustrating the antibody format-specific retention, and b) Histogram displaying the distribution of model training and testing split data.

4.3.2 QSPR Modeling of pH Gradient Retention

An integral part of statistical model development is the identification of a predictive feature set and a suitable mapping function followed by the assessment of model quality. In this study, a GPR model was used to regress the pH retention of a large antibody set (M=64) to their physiochemical properties, which were encoded into 1312 descriptors per molecule, as detailed in Section 4.2.4 and 4.2.5.

Initially, the data set was divided into training and testing data for the validation of the empirical model. Fig. 4.2 (b) displays the distribution of randomly selected molecules into 20% test set (m=13) and 80% training set (m=51), ensuring a representative distribution of pH retentions and molecule formats. Within the test set, four molecules categorize to the lower pH group and nine molecules to the higher pH group. Furthermore, the test set comprises six IgG1, four IgG-scFv, two IgG4, and one KiH molecule yielding a robust test

set selection, as listed in Table 4.1. Thereafter, a two-staged feature selection was conducted using the training data to reduce 1102 preprocessed descriptors to six features. The first stage involved filter methods to efficiently discard the majority of uninformative or convoluted descriptors. Initially, low variance ($cov(x,x) \leq 0.01$) features were removed, reducing the feature number to 1083. The remaining features were sorted according to their unimodal interaction towards the target variable, using linear regression models. Multicollinear descriptors with significant Pearson correlation ($\rho \geq 0.80$) were removed, as suggested by Sankar et al. [110]. Although collinearity not necessarily diminishes model predictiveness, the removal of multicollinear descriptors was required to improve mechanistic interpretability.

From the remaining 413 features, ten descriptors were selected based on their linear relationship towards the pH retention, as depicted in the bottom x-axis of Fig. 4.3. The selected features provided insights into the multimodal binding mechanism of IgG-like molecules, as well as relevant antibody domains. Mostly charge-related descriptors were selected, with eight out of ten descriptors encoding for electrostatic interactions. The two remaining descriptors gave insight into hydrophobic contributions. The small representation of hydrophobic descriptors could be a result of the weak pH dependency of hydrophobic attraction compared to electrostatic interaction [166,199]. The strongest linear relationship was observed for the isoelectric point (pI) of the molecules termed "pI_model_pKA_based", which categorizes as a global and structure-based descriptor. The other descriptors were either pointing towards the Fab domain or the constant region, as well as local descriptors within both regions. Furthermore, solely structure-based and patch descriptors were identified during the first stage of feature selection, while the relationship towards sequence-based descriptors deemed less significant. Comparable global descriptors have been identified in multiple studies, including the structure-based molecule pI [159,168–172]. Interestingly, no aggregation-propensity descriptors were selected, which in theory should be able to encode for adjacent electrostatic-hydrophobic interactions as observed for Capto adhere in a previous study [172]. On the other hand, proximal interactions have not been reported in a more recent study concerning Capto adhere, selecting from a significantly larger pool of initial descriptors compared to the previous study [72]. It remains an open question whether adjacent patch interactions are not as pronounced in Capto adhere binding as reported for multimodal cationic resins [160,171] or if the current class of proximal descriptors [170,188] is not sufficiently parameterized to describe multimodal anionic interactions. In addition, no descriptors from the custom scFv region were chosen during the filter process, which could indicate a minor role of this region in antibody binding. A possible explanation for the less significant role of the scFv domain compared to the Fv domain of the bispecific mAbs could be related to steric constraints due to the engineered inter- and intrachain linkers limiting the configurational flexibility of this region [200]. To answer these questions, more research is required to investigate local ligand interactions as by using recent labeling methods [164].

In the second stage of feature selection, recursive feature elimination (RFE) was employed based on model performance, as illustrated in Fig. 4.3. A GPR model with a mixed

covariance function was utilized to estimate the target function. The Bayesian method was chosen to account for both linear and nonlinear feature contributions in the multimodal binding mechanism, while accurately estimating heteroscedastic prediction uncertainty, as previously observed in cation exchange chromatography [179]. During the RFE process, GPR models were sequentially fitted to the training data, starting with the ten most significant descriptors on the right-hand side of Fig. 4.3 and ending with a single feature for the final model on the left-hand side of the figure. At each iteration, the feature with the lowest permutation importance or the smallest impact on model accuracy was removed. The calculation of feature permutation importance was required since no meaningful model weights could be extracted from the nonlinear GPR model. Simultaneously, the log(marginal likelihood) of the model conditioning and the cross-validation mean absolute error of model prediction were recorded at each iteration.

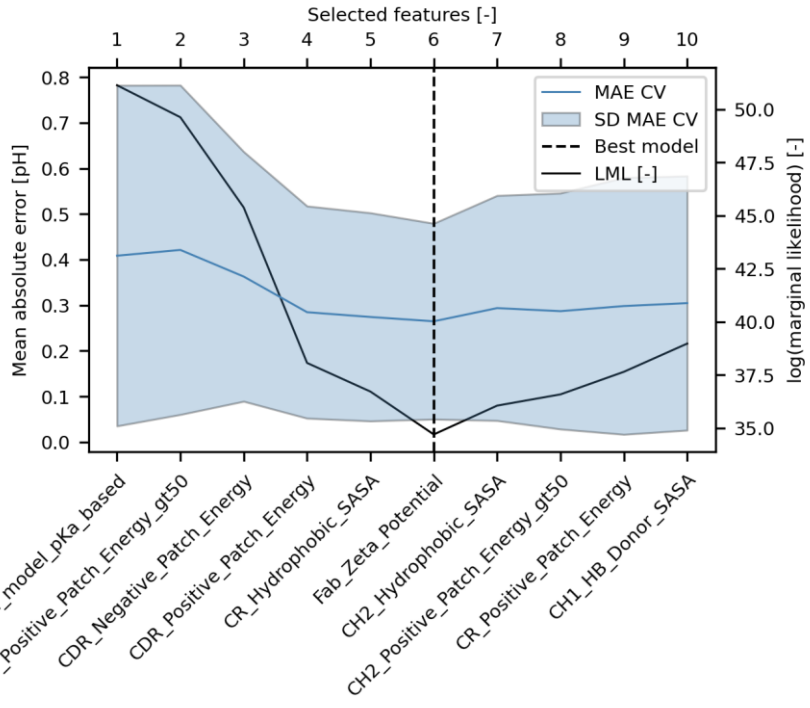

Fig. 4.3: Recursive feature elimination with cross-validation of training data.Starting with ten features on the right-hand side of the figure, a model is initially fitted using all available features. At each iteration, the model is evaluated by calculating the cross-validation mean absolute error (MAE CV), the standard deviation (SD) of cross-validation scores, and the log(marginal likelihood) (LML) of the given model. The feature with the lowest permutation importance is eliminated per iteration, as indicated on the bottom x-axis. Once all but one feature is removed, the best model with the lowest log(marginal likelihood) is selected.

Upon evaluating Fig. 4.3, a distinct trend for both the LML and the cross-validation MAE, along with its standard deviation is displayed, identifying the optimal GPR model as the one using six features. The LML of model conditioning ranged from 34.698 for the sixth to 51.133 for the final RFE iteration. Meanwhile, the cross-validation MAE fell within a pH range of 0.265 pH for the sixth to 0.421 pH for the penultimate iteration. Furthermore, the SD of cross-validation was lowest at the selected model at 0.215 pH, indicating increased model robustness compared to the largest SD observed for the final RFE iteration with a MAE SD of 0.373 pH. Overall, the cross-validation score appeared more susceptible to variation in identifying the most predictive model compared to the Bayesian likelihood, thus demonstrating the advantage of the Bayesian approach, not only for feature mapping, but also for feature selection.

The six selected features comprised the molecule pI as a global charge-related descriptor, three local charge descriptors within the antibody Fv domain, namely "LFR4_Positive_Patch_Energy_gt50", "CDR_Negative_Patch_Energy", and "CDR_Positive_Patch_Energy", as well as electrostatic and hydrophobic contributions from the constant region and the Fab domain through the descriptors "CR_Hydrophobic_SASA", and "Fab_Zeta_Potential". According to the descriptor naming scheme, "LFR4_Positive_Patch_Energy_gt50" is defined as the summation of all positive patches within the antibody LFR4 region with a patch area larger than 50 Å^2. Interestingly, the selected features included two custom descriptors, encoding for the antibody constant region and its Fab domain. In contrast, the local descriptors within the constant region were discarded, despite multiple studies highlighting the importance of the Hinge region and the CH2–CH3 interface in multimodal interaction [64,163–165,167]. Furthermore, the removal of the descriptors from the constant region and the Fab domain had less impact on the LML and the cross-validation MAE compared to the removal of local descriptors within the Fv domains. This finding emphasizes the significance of the Fv domain, particularly the CDRs, for antibody binding to multimodal chromatography resins, as observed by multiple authors for cationic MMC [57,70,160,161,166,171]. A comprehensive evaluation and interpretation of the model features will be provided in Section 4.3.3.

Following the feature selection, the predictiveness of the final QSPR model was evaluated via internal and external model validation. Accordingly, the goodness of fit to the training data and the goodness of prediction for the test data were assessed, as depicted in Fig. 4.4. Both the training set and the test set, consisting of 80% (m=41) and 20% (m=13) of the antibodies from the full data set are displayed. Furthermore, the pH observations and predictions of the test set, as well as molecule residuals and uncertainties are listed in Table 4.1. Again, an agglomeration of data points into a lower eluting and a higher eluting pH group is evident. The upper part of the figure compares the observed and predicted pH retention, while the lower part focuses on the distribution of the molecule residuals. Additionally, linear fits to the scattered training and test set molecules are displayed to enable a quick assessment of model performance against a theoretical ideal model with zero error in predictions.

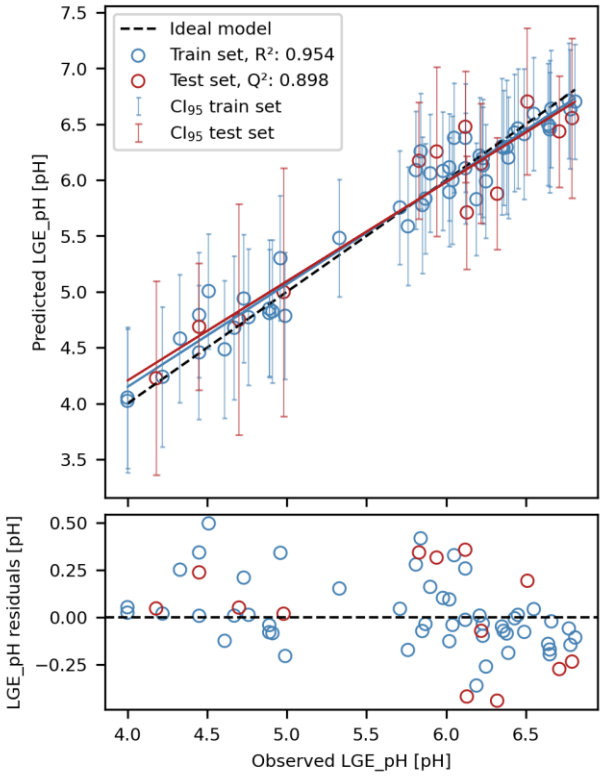

Fig. 4.4: Goodness of pH elution prediction (R^2, Q^2).The upper part of the figure compares the predicted with the experimental pH elution for the molecules in the random training and the testing data. An ideal model is represented by a straight line, where predicted and experimental observations have zero error. For each molecule, the 95% confidence interval (CI_{95}) of prediction is calculated. The bottom part of the figure displays a residual plot, with the y-axis normalized to the absolute deviation of model prediction.

On the top figure, the heteroscedastic 95% CI for each molecule is indicated, while the lower figure presents uniformly dispersed residuals, signifying the absence of systematic model errors. The model's uncertainty varies between one and two pH units, while the overall accuracy demonstrates a maximum deviation of 0.44 pH for IgG1 (10), as outlined in Table 4.1. This increased uncertainty, in comparison to the prediction accuracy, was necessary to avoid model overfitting, as discussed in Section 4.2.5. The source of this uncertainty might be attributed to the descriptor set's inaccuracy to fully capture the pH sensitivity of multimodal protein adsorption [27,28,84,166].

When evaluating the model performance metrics, the fit of the final GPR model to the training data achieved a coefficient of determination of $R^2 = 0.954$, with a MAE of 0.132 pH, and an average SD of model uncertainty of 0.281 pH. Internal model validation of the training data through five-fold cross-validation with ten repetitions resulted in a mean Q^2 of 0.780 and a MAE of 0.279 pH. Lastly, the pH retentions of the external test set could be predicted with a Q^2 of 0.898, a MAE of 0.231 pH, and an average SD of model uncertainty of 0.344 pH, as detailed in Table 4.1. Comparing the results of the model training

and testing, similar scores were achieved for the R^2 and external Q^2, despite showing a decreased model accuracy of approximately 0.1 pH, as well as an increased uncertainty in the model predictions. The similarity of these quality metrics underlines the robustness of the empirical model, as significantly diverging training and test results would suggest model overdetermination. The model's robustness is further supported by the y-scrambling results in the Appendix B, Fig. B1, which indicates a less than 1% probability of achieving the model performance by chance. Continuing with the internal model validation, a divergence between the model fit and cross-validation scores is apparent. Larger divergence during cross-validation is a common phenomenon of empirical models and becomes especially pronounced for small data sets, as the missing data used for subsampling can impair model performance.

Upon examining the 95% CI of the model predictions, comparably large intervals are apparent for the molecules eluting in the lower pH group. When evaluating the SDs listed in Table 4.1., the three largest model uncertainties were observed for the molecules IgG4 (4), IgG4 (3), and IgG1 (1) in a descending order, with SDs of 0.57, 0.53, and 0.44 pH, respectively. The elution pH of both IgG4 formats within test set was predicted with the highest uncertainty among all molecules examined. The substantial 95% CI of the two IgG4s may indicate an antibody format-dependent uncertainty captured by the empirical model, as all formats except the IgG4 type molecules shared IgG1 backbones, as previously discussed. Exemplarily, the KiH antibody was predicted more accurately than the IgG4 formats, despite its bispecific or trispecific functionality. Conversely, the molecule with the third largest uncertainty was a standard IgG1, eluting at the lower end of the pH spectrum. This observation could be attributed to the scarcity of data points in close vicinity to the IgG1 (1), compared to molecules in the higher eluting pH, potentially leading to impaired model performance.

In conclusion, the predictive power of the QSPR model relies on both the density and the overall number of molecules sharing similar physicochemical surface characteristics. In comparison to peer studies on (homologue) proteins libraries using a single resin system, our model demonstrates a superior performance [159,168–171]. However, previous models only employed a third of the number of molecules used in this study. Considering the performance and broad applicability of the validated model in accurately predicting pH retention for a wide range of commercially available antibodies based solely on their amino acid sequence, two potential applications emerge. First, the model can be employed to predict an initial pH set point for the anionic mixed-mode resin Capto adhere during early process development, which is especially advantageous for multi-domain proteins as mAbs concerning the pH-sensitivity of their domain contributions to chromatographic retention [166]. Second, it can serve as an *in silico* screening tool for identifying molecule manufacturability in regards to downstream purification.

Table 4.1: Overview of goodness of test set prediction and model uncertainty.

Molecule	Observed [pH]	Predicted [pH]	Residual [pH]	SD [pH]	CI_{95} [pH]
IgG1 (1)	4.18	4.23	0.05	0.44	3.36 – 5.09
IgG1 (2)	4.45	4.69	0.24	0.29	4.12 – 5.25
IgG4 (3)	4.7	4.75	0.05	0.53	3.72 – 5.79
IgG4 (4)	4.98	5.00	0.02	0.57	3.89 – 6.11
IgG-scFv (5)	5.83	6.17	0.34	0.27	5.65 – 6.70
IgG1 (6)	5.94	6.26	0.32	0.39	5.50 – 7.01
IgG-scFv (7)	6.12	6.48	0.36	0.25	5.98 – 6.97
IgG-scFv (8)	6.13	5.71	0.42	0.26	5.20 – 6.22
IgG-scFv (9)	6.22	6.15	0.07	0.27	5.61 – 6.68
IgG1 (10)	6.32	5.88	0.44	0.26	5.38 – 6.38
IgG1 (11)	6.51	6.70	0.19	0.34	6.05 – 7.36
IgG1 (12)	6.71	6.44	0.27	0.25	5.94 – 6.93
KiH (13)	6.79	6.56	0.23	0.36	5.84 – 7.27

4.3.3 Model Inspection and Interpretation

After the final QSPR model was established, the feature contributions within the model were further analyzed to investigate the interaction mechanism of the antibodies to the anionic multimodal resin illustrated in Fig. 4.1 (c). Throughout this study, in particular charged-based descriptors within the variable region appeared correlated to the multimodal interaction. Therefore, a model inspection was conducted, identifying interdependencies between the selected features, to support the mechanistic interpretability of the final model. Fig. 4.5. depicts the histogram of the standardized and mean centered feature values for the entire data set on the top diagonal axes (column a – f), as well as the pairwise relationship of model features on the off-diagonal axes (row 2 – 6). Additionally, the pairwise relationship of the model features to the target variable is appended to the figure

bottom (row 7) and sorted according to the linearity of the training set features to the pH retention. Furthermore, the different molecule formats are indicated in each subplot and the Pearson correlation coefficient p, calculated from the entire data set.

When evaluating the individual feature distributions shown in the top diagonal axes of Fig. 4.5, similar trends to the pH retention introduced in Fig. 4.2 (a) are evident. In both figures, the IgG1 and IgG-scFv formats display the broadest distribution throughout the parameter space, whereas the IgG4 and KiH formats show a clustered behavior. Notably, the distribution of the molecule pI, shown in subfigure (a1) of Fig. 4.5 is comparable to the distribution of the target variable, shown in Fig. 4.2 (a). The similarity between the molecule pI distribution and the pH retention distribution is supported by their strong positive linear correlation of $p = 0.74$ shown in Fig. 4.5, subfigure (a7). The skewed and bimodal distribution of the pH retention shown in Fig. 4.2 (a) appears to be primarily influenced by the molecule pI. In this context, the abundance of elevated pI values within the full data set, in contrast to the significantly lower pI values of the IgG4 formats, could explain the bimodal appearance of the pH retention distribution. The global pI, however, depends on individual contributions from the antibody subdomains. Consequently, further insight into the antibody interactions can be obtained by inspecting the zeta potential distribution of the Fab domain in Fig. 4.5, subfigure (c3), as well as the hydrophobic accessible surface area (SASA) distribution of the constant region shown in subfigure (d4). The IgG4 formats exhibit a reduced zeta potential of their Fab regions, while the three KiH formats define the upper limit of zeta potential. The surface charge of the Fab not only translates to an increased pI, as assessed by their positive correlation in subfigure (a3), but also correlates positively to the pH retention, as depicted in subfigure (c7). Moreover, the Fab zeta potential is correlated to the electrostatic surface potential of the CDRs encoded via their positive and negative patch energy, shown in subfigure (b3) and (c6). Upon inspecting the distribution of hydrophobic SASA from the antibody constant region in subfigure (d4), again, two groups are apparent. The first group represents less hydrophobic molecules with a IgG1 backbone, while the second group is comprised of IgG4 antibodies exhibiting increased surface hydrophobicity, which it is expected considering the conserved nature of the Fc domain. In this context, the hydrophobic clustering behavior might obscure the true importance of the constant region to the multimodal binding by weakening the linear relationship of its contribution.

In conclusion, the strong correlation of the pH retention to Fab domain, and particularly, the antibody CDRs does not explicitly imply a Fab-first binding orientation but suggests a complex interaction mechanism that could depend on multiple binding domains on the entire antibody surface. This assumption is supported by the strong correlation of the global pI to the pH retention and the results published by Robinson et al., studying domain contributions and pH dependency of the multimodal antibody interaction [161,166]. Intriguingly, the energy of large positive patches in the LFR4 region shown in Fig. 4.5, subfigure (e7) exhibited no significant correlation with the target variable, other features, or clustering behavior. However, it was removed second to last during the RFE process depicted in Fig. 4.3, indicating the potential relevance of this region.

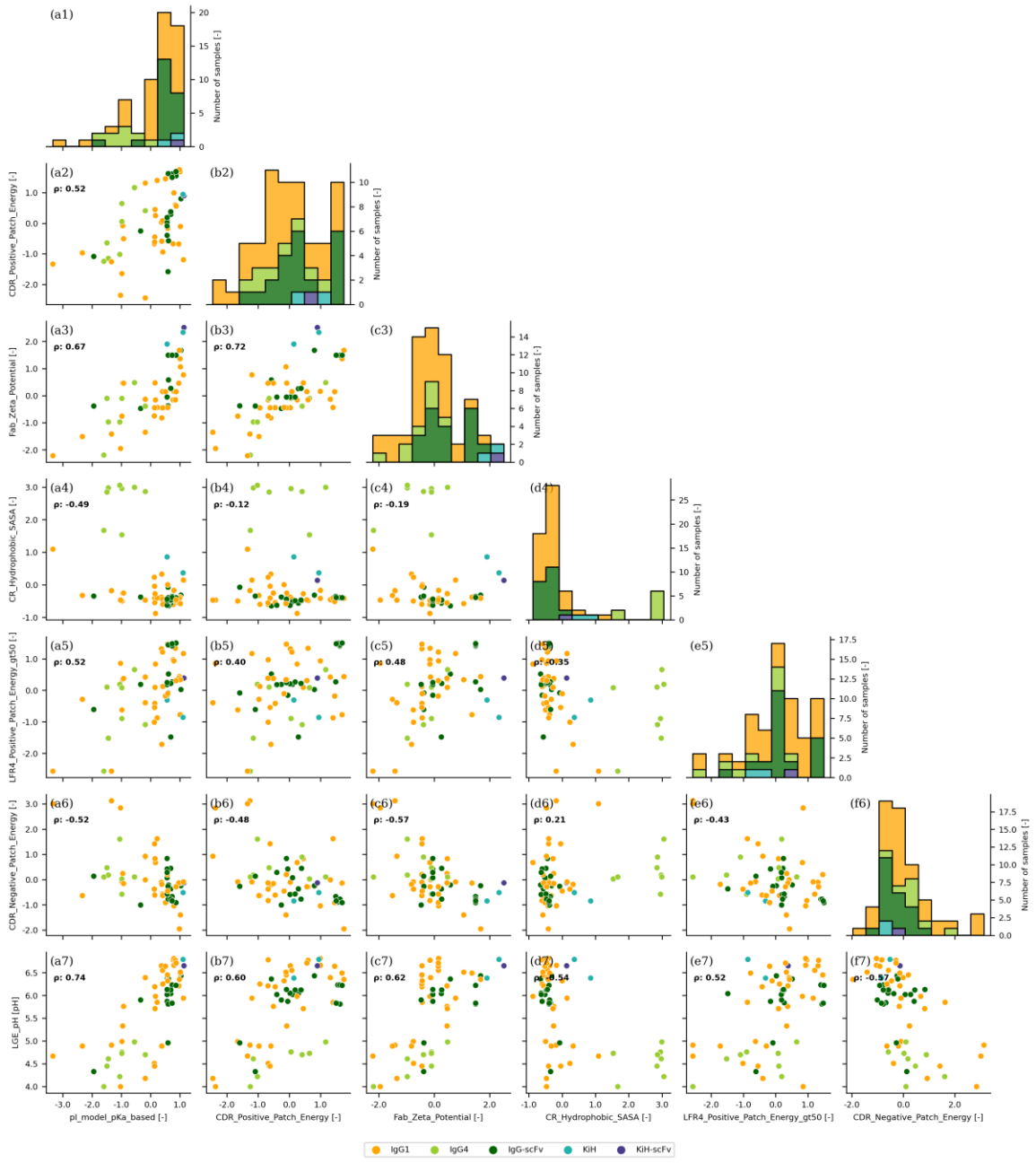

Fig. 4.5: Pairwise relationship and distribution of model features and pH elution.The diagonal subfigures display the univariate histograms of the model features, as shown on the bottom x-axis. The off-diagonal subfigures illustrate the bivariate relationship between model features, while the bottom row presents the relationship between the target variable and the model features. Features are sorted in descending order based on their absolute Pearson correlation coefficient (p) towards the target variable within the training data, indicated in the upper left corner of each subplot. All subfigures differentiate between antibody formats.

To further investigate the descriptor contributions and finalize the mechanistic interpretation, the partial dependences of the features within the GPR model were analyzed, as depicted in Fig. 4.6. Partial dependence allows for examining the strength and form of nonlinear feature contributions to a multivariate model, as well as identifying feature interdependencies. In brief, a single factor perturbation is performed by marginalizing all but one feature and permutating it within the full feature range. Subsequent recording of output predictions for a single molecule result in its individual conditional expectation (ICE), while averaging over all ICEs yields the partial dependence of the feature. Consequently, heterogeneous behavior of the ICEs can reveal feature interdependencies. However, during the interpretation of partial dependence, it is crucial to consider multicollinearity within the inspected feature set, as implausible parameter pairs can form, such as the simultaneous occurrence of strong positive and strong negative patch energies in a specific region [201].

The partial dependences of the model features shown in Fig. 4.6 share the same order as displayed in the bottom row of Fig. 4.5 and follow similar relationships suggested by the scattered data. The feature distributions of each descriptor are indicated by deciles lines on the x-axis. Additionally, comparable permutation importance's are calculated in a range in between 0.12 and 0.18 pH for each feature and are displayed on the upper left corner of the subplots already employed during the RFE. Furthermore, a distinction between the training and test set is made, which implies robust model performance based on comparable functions of the partial dependences throughout the full feature space. Upon inspecting the ICE lines, all but the molecule pI and the hydrophobic SASA of the constant region shown in subfigure (d) display pronounced heterogeneous behavior. This behavior can be explained for the Fab and the CDR descriptors based on their mutual linearity, as discussed earlier. Moreover, the feature collinearity leads to underestimation of permutation importance, which implies a dominating role of the variable region during the multimodal interaction when adding up their individual permutation importance's. Only the heterogeneous ICE lines of the energy from large positive patches in the LFR4 region shown in subfigure (e) provide further insight into a potential interaction mechanism. It appears that for early eluting molecules during the pH gradient, increasing the positive patch energy in the LFR4 region, an area in-between the L3 loop and the VL-CL interface, has a negligible effect on the pH retention. Conversely, an increase of positive patch energy in the LFR4 region shifts the retention of strong binding molecules more than one pH unit. In conclusion, the ICE lines of the large positive patches in the LFR4 region suggest that the LFR4 region can contribute to a significant binding domain, as recently identified by Parasnavis et al. [70]. Furthermore, the contribution of the LFR4 region to the antibody binding is dependent from the LFR4 surface charge, leading to an increased adsorption in the absence of strongly repelling positive patches.

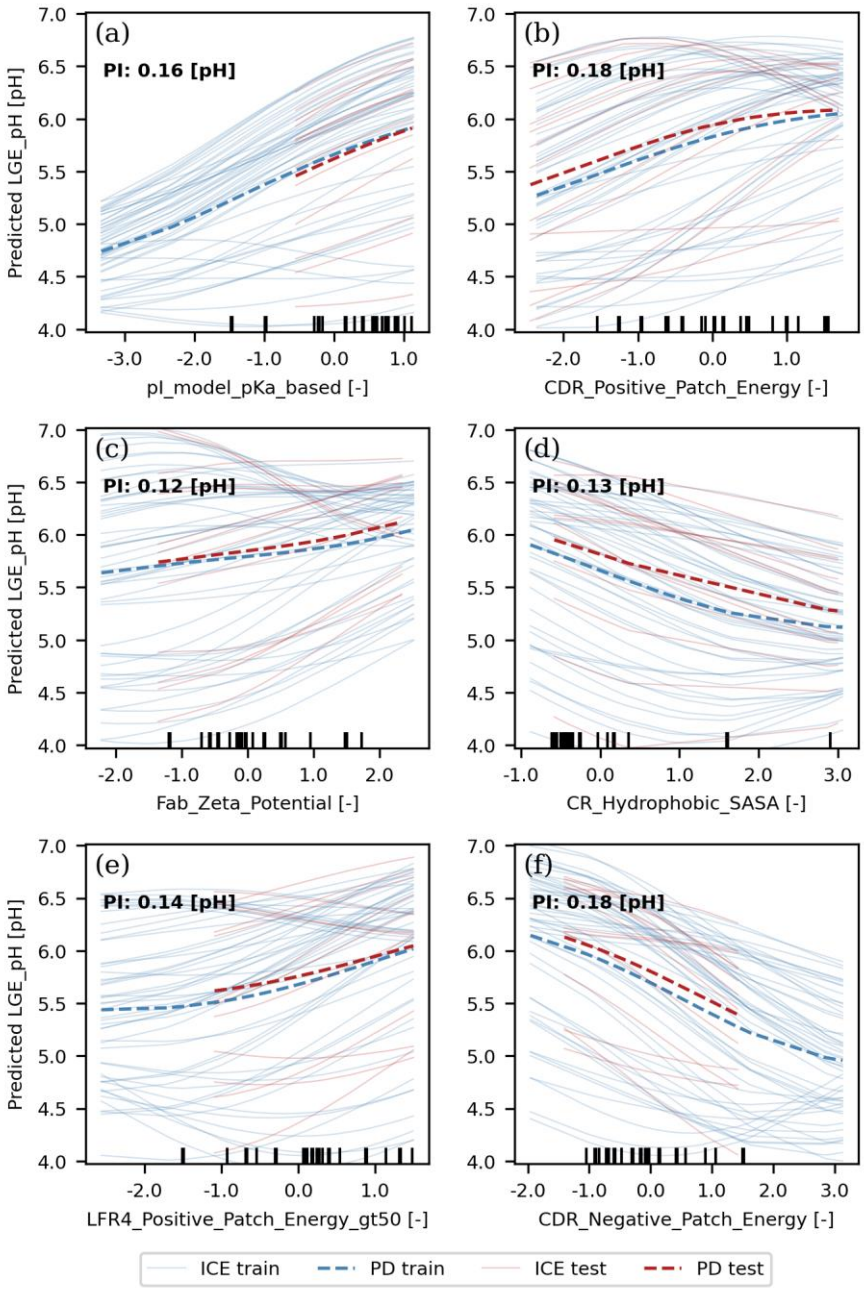

Fig. 4.6: Partial dependence of pH elution from model features of training and test data.Each subfigure depicts the model response when permutating the features in their normalized feature range. The individual conditional expectation lines represent the target prediction for each molecule, while the partial dependence of the feature is given by their mean values. The feature permutation importance (PI) towards the target prediction is shown in the upper left corner of each subplot. The decile lines at the bottom of each subplot indicate the frequency of the feature values within the data set.

Considering all feature contributions, the multimodal antibody-ligand binding appears to be driven by linear contributions of electrostatic attraction and repulsion from the CDRs but further depends on significant binding domains throughout the entire antibody surface, as implied by the strong relationship to the global pI and the constant region. A significant role of adjacent electrostatic-hydrophobic patches in Capto adhere binding was not observed but might be attributed to inadequate descriptors to encode for multimodal anionic interaction, as discussed in Section 4.3.2.

4.4 Conclusion

In this study, we developed a QSPR model to predict pH gradient elution in multimodal chromatography using a diverse set of therapeutic antibodies. In total 64 IgG-derivates were included, categorized into five therapeutic antibody formats, which exhibited distinctive chromatographic behavior. Throughout the QSPR workflow, physiochemical characteristics tailored for antibody description were derived from homology modeling and regressed to the pH retention. A rigorous feature selection was conducted, reducing the initial descriptor count from 1312 to six in course of a two-staged feature selection process. The utilization of GPR as a Bayesian modeling approach proved advantageous due to its strong model performance, heteroscedastic uncertainty estimation, and nonlinear feature identification.

Our experimental results demonstrate that the IgG backbone significantly impact chromatographic retention, as indicated by the comparable elution behavior of the IgG4 molecules. However, the main driver of multimodal interaction is presumed to be in the antibody Fv domain, as homologous IgG1 derivatives showed diverging elution trends. The feature dependencies of our QSPR model support these findings and shed light on a complex adsorption mechanism in multimodal chromatography. The proposed mechanism originates from the CDR region but involves the entire antibody structure due to a combination of electrostatic and hydrophobic contributions.

The overall model performance and its mechanistic interpretability allow for its application in an accelerated process development of IgG-like purification based solely on sequence information. Our model can replace the experimental screening of initial process pH in multimodal chromatography, which is particularly beneficial for the material and time-constrained early process development. Furthermore, our model can serve as an *in silico* screening approach to identify candidates suitable for purification by multimodal chromatography. Lastly, the identified feature dependencies could aid in the development of improved mechanistic chromatography models by considering a molecular level of detail. In this context, multiscale modeling through the correlation of mechanistic isotherm parameters to molecular-level descriptors could be considered as an intermediate step. To enhance current QSPR models and address the structural diversity of engineered biologics, future

research should target the development of global descriptors to encompass multimodal surface interactions, as well as protein topology, independent from molecule structure.

Acknowledgements

The authors would like to thank the department of Development Biologicals (Boehringer Ingelheim, Biberach, Germany) and the Biotherapeutics Discovery department Biologics Research CMC (Boehringer Ingelheim, Ridgefield, USA) and especially Michael Marlow, Helen Wu, and James Hu, for provision of therapeutic antibodies. Furthermore, we would like to thank the BI modeling team for their constant support, as well as the long and fruitful discussions.

5 Predicting Multimodal Chromatography of Therapeutic Antibodies using Multiscale Modeling

Rudger Hess[1,2], Jan Faessler[2], Doil Yun[2], Ahmed Mama[2], David Saleh[2], Jan-Hendrik Grosch[2], Gang Wang[2], Thomas Schwab[2], Jürgen Hubbuch[1,*]

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

[2] DSP Development, Boehringer Ingelheim Pharma GmbH & Co. KG, Biberach, Germany

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

Journal of Chromatography A (1718), 2024, p. 464706

DOI: 10.1016/j.chroma.2024.464706

Abstract

Multimodal chromatography has emerged as a powerful method for the purification of therapeutic antibodies. However, process development of this separation technique remains challenging because of an intricate and molecule-specific interaction towards multimodal ligands, leading to time-consuming and costly experimental optimization.

This study presents a multiscale modeling approach to predict the multimodal chromatographic behavior of therapeutic antibodies based on their sequence information. Linear gradient elution (LGE) experiments were performed on an anionic multimodal resin for 59 full-length antibodies, including five different antibody formats at pH 5.0, 6.0, and 7.0 that were used for parameter determination of a linear adsorption model at low loading density conditions. Quantitative structure-property relationship (QSPR) modeling was utilized to correlate the adsorption parameters with up to 1374 global and local physicochemical descriptors calculated from antibody homology models. The final QSPR models employed less than eight descriptors per model and demonstrated high training accuracy ($R^2 > 0.93$) and reasonable test set prediction accuracy ($Q^2 > 0.83$) for the adsorption parameters. Model evaluation revealed the significance of electrostatic interaction and hydrophobicity in determining the chromatographic behavior of antibodies, as well as the importance of the HFR3 region in antibody binding to the multimodal resin. Chromatographic simulations using the predicted adsorption parameters showed good agreement with the experimental data for the vast majority of antibodies not employed during the model training.

The results of this study demonstrate the potential of sequence-based prediction for determining chromatographic behavior in therapeutic antibody purification. This approach leads to more efficient and cost-effective process development, providing a valuable tool for the biopharmaceutical industry.

Keywords: *Multispecific monoclonal antibody (mAb) formats; Mechanistic chromatography modeling; Mixed-mode antibody descriptors; Quantitative structure-activity/property relationship (QSAR/QSPR); In silico / a priori process development*

5.1 Introduction

Liquid column chromatography has established itself as a prominent technique for the purification of biotherapeutics and is virtually indispensable for the production of monoclonal antibodies (mAb) [16]. Current research targets to improve this separation method by combining multiple forms of physicochemical interactions into a single chromatographic ligand, referred to as multimodal chromatography [52]. The improved ligand structure enables increased selectivity and salt tolerant adsorption compared to unimodal interaction resins [46].

The increased selectivity of multimodal chromatography originates from orthogonal interactions towards the molecule surface, such as complex electrostatic and hydrophobic effects and leads to increased sensitivity towards molecule characteristics and process conditions [19]. However, this increased sensitivity often necessitates a molecule-specific process development, which involves extensive experimentation for process optimization and characterization to ensure adequate product quality and process yield [64]. To expedite this time and resource-intensive process development, high-throughput experimentation and *in silico* tools have been deployed. Mechanistic and statistical modeling serve as valuable tools for the acceleration, optimization, and characterization of chromatography process development by either simulating the mass-transport and accumulation within the chromatographic column or by capturing convoluted interrelations where mechanistic understanding is scarce [29,202]. While reliable mass transfer models for chromatographic simulation exist, current isotherm equations struggle to capture the complex adsorption and desorption kinetics of multimodal chromatography [203]. Deviations of chromatographic simulation from experimental data are frequently accounted to the anisotropic shape, charge, and hydrophobicity distribution of the molecule surface, leading to a conformation- and orientation-dependency of the adsorption process that is especially pronounced for complex multidomain proteins such as mAbs [165,204]. Nevertheless, electrostatic, and hydrophobic effects are combined into a complex but lumped parameter model to reflect stoichiometric displacement [54] or colloidal particle interaction [205] of the molecule-adsorber system in a continuum representation, leading to semi-mechanistic isotherm equations. Additional factors like dewetting phenomena or pH dependence can be incorporated into the mechanistic description, further increasing model complexity [84]. Elevated complexity and semi-mechanistic model structure increase the difficulty in calibrating isotherm parameters as they cannot be directly measured but must be fitted to experimental observations [86]. In contrast, model applicability is enhanced through simplification, by focusing on either electrostatic [124] or hydrophobic effects [177] as the governing form of physicochemical interaction.

To gain a deeper understanding of the multimodal antibody binding, various biophysical methods have recently been employed and include: thermodynamic modeling of macroscopic observations [158], investigation of single amino acid substitutions in homologous fragment antigen binding (Fab) libraries [160], assessment of full-length antibody domain contributions [70,161,166], and identification of preferred ligand-binding regions of the

fragment crystallizable (Fc) by using solution-phase nuclear magnetic resonance spectroscopy in concert with molecular dynamics (MD) simulations [163,167], as well as covalent labeling and mass spectrometry [164]. This experimental data can then be correlated with physicochemical antibody descriptors to infer quantitative structure-property relationships and improve mechanistic interpretability [171,206].

Despite recent advancements in isotherm discrimination and biophysical investigation, current models lack mechanistic insight into the multimodal adsorption process and are impeded by an elaborate and resource-intensive model calibration. A significant gap remains in linking macroscopic adsorption parameters with antibody structure to debottleneck the expensive model calibration and enhance mechanistic understanding, to facilitate an *a priori* process development of multimodal chromatography [93,203].

In this study, we aim to bridge this gap by introducing a multiscale modeling approach to predict multimodal chromatographic behavior of therapeutic antibodies based on their sequence information, as illustrated in Fig. 5.1. Previous studies have demonstrated the potential of multiscale modeling in predicting unimodal interactions within cation exchange chromatography for model proteins, both with [207] and without [208] pH dependency, as well as mAbs [179]. To the best of our knowledge, this is the first approach of predicting adsorption parameters of therapeutic antibodies in multimodal chromatography at various pH conditions.

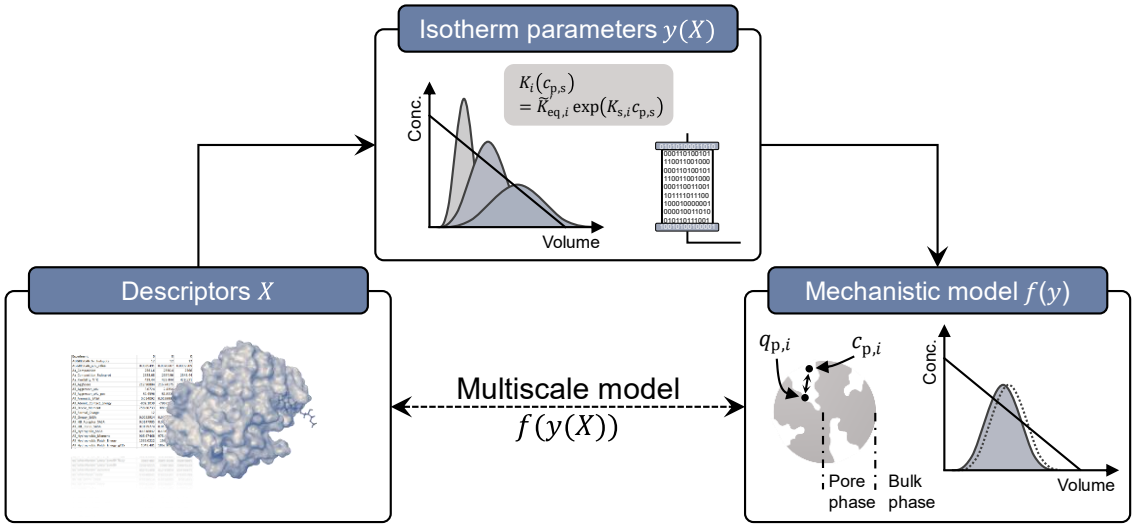

Fig. 5.1: Multiscale modeling workflow.This figure illustrates how QSPR models predict mechanistic adsorption parameters y from protein descriptors X. These parameters are then used in *a priori* process development through chromatographic simulation $f(y)$. Deviations in these simulations from standard operating conditions can guide molecule design to mitigate manufacturing risks. The depicted antibody structure is a modification of PDB entry 1HZH [176].

Our approach involves conducting LGE experiments with a multimodal anion exchange resin for a diverse set of therapeutic antibodies, determining linear adsorption parameters, and employing QSPR modeling to correlate these parameters with physicochemical descriptors calculated from antibody homology models. The accuracy of our multiscale predictions, provide valuable insights into the physicochemical aspects of multimodal chromatography by segregating contributions of the adsorption parameters. Ultimately, our multiscale approach paves the way for a more efficient and cost-effective development process in the purification of therapeutic antibodies.

5.2 Material and Methods

5.2.1 Chromatography Resin, Buffers, and Molecules

In this study, we employed the multimodal strong anion exchanger Capto adhere (Cytiva, Marlborough, USA) for all chromatographic experiments. Model calibration and validation were conducted using a prepacked Capto adhere HiScreen column (7.7x100 mm, Cytiva) with a column volume (CV) of 4.7 mL [68].

All chemicals utilized in this study were purchased from Sigma-Aldrich Co LLC (Saint Louis, USA). Highly purified water was used for buffer preparation, while 1 M hydrochloric acid and 1 M sodium hydroxide were employed for pH adjustment. A multicomponent buffer system at pH 5.0, 6.0, and 7.0 was used for equilibration, elution, and buffer exchange during all chromatographic experiments. This buffer system, adapted from Kröner and Hubbuch [137], consisted of 9.1 mM 1,2-ethanediamine, 6.4 mM 1-methylpiperazine, 13.7 mM 1,4-dimethylpiperazine, 5.8 mM bis-tris, and 7.7 mM hydroxylamine, with the addition of 60 mM hydrochloric acid [177]. Furthermore, 1.5 M sodium chloride was added to the multicomponent buffer used for equilibration (at inlet B), resulting in a total of 1.56 M chloride counterions. Other buffers employed in the chromatographic experiments included 1 M acetic acid for column regeneration, 1 M sodium hydroxide for column cleaning, and 20% ethanol for column storage.

This study utilized a diverse set of 59 full-length immunoglobulin G (IgG) derivatives, provided by Boehringer Ingelheim Pharma GmbH & Co. KG (Biberach, Germany) as introduced in our previous study [206]. This set included 57 human-origin and 2 humanized murine-origin antibodies, exhibiting a broad range of physicochemical characteristics. The antibody collection comprised 30 IgG1s, 19 IgG bispecifics with two single-chain fragment variables (scFv) appended to each heavy chain C-terminus (IgG(H)-scFv), 6 IgG4s, 3 Knob-in-Hole bispecifics (KiH), and 1 KiH trispecific with a single scFv attached to the C-terminus of the Hole chain (KiH-scFv). Antibody expression was performed using Chinese hamster ovary cell lines. Subsequently, the antibodies were captured through Protein A affinity chromatography. The protein solutions were then neutralized to pH 5.5, sterile-filtered using a 0.2 µm filter, and frozen at -70°C. Upon thawing, the final load material

was adjusted to a concentration between 1.25 and 5 g/L, as determined by a NanoDrop 2000c spectrophotometer (Thermo Fisher Scientific, Waltham, USA). Prior to sample application, the load material underwent buffer exchange into the equilibration buffer using a 5 mL HiTrap Desalting column (Cytiva), following the manufacturer's instructions.

5.2.2 Homology Modeling and Descriptor Calculation

The antibody structure prediction was conducted using homology modeling in Maestro Bioluminate 4.9 (Schrödinger Inc., New York, USA). The homology modeling protocol involved a five-stage process adapted from Zhu et al. [206], followed by further structure refinement based on the protocol of Sastry et al. [184]. The Enhanced Chothia numbering scheme [181] was employed, utilizing the human IgG1 1HZH [176] as full-length template for all molecules, except for IgG4 subtypes, which used 5DK3 [182] instead. Complex bi- and trispecific formats were modeled by grafting intra- and intermolecular linkers to append independently modeled scFv domains to the full-length mAb structures [183].

Subsequently, physicochemical descriptors were calculated using Bioluminate at the experimental pH value. A set of 165 unique features was calculated for each antibody structure, divided into sequence-based (n=69), structural (n=59), and patch-specific (n=37) categories [110]. Additionally, a region-specific subset (n*=31) of the initial descriptor set was calculated for 37 antibody subdomains. These subdomains are specific to the antibody structure and encompass the light and heavy chain variable regions (VL, VL_Fv, VH, VH_Fv) focusing on the complementarity-determining regions (CDR, CDRL, CDRH) and framework regions (FR, FRL, FRH). These regions consist of individual loops (L1, L2, L3, H1, H2, H3) and frameworks (LFR1, LFR2, LFR3, LFR4, HFR1, HFR2, HFR3, HFR4). In addition to the variable region, the antibody constant regions (CL, CH1, CH2, CH3) and the hinge region (Hinge) are considered. The regional descriptors were further expanded by incorporating seven custom regions to accommodate the fragment variable (Fv), Fab, Fc, and the sum of the constant regions (CR). Moreover, the single-chain fragment variable regions (scFv, VLscFv, VLscFv) of the bi- and trispecific formats were taken into account. A more comprehensive explanation of the structure prediction and descriptor calculation can be found in our previous study [206].

In addition to the descriptors provided by Bioluminate, 62 supplementary global descriptors were calculated to complement the final feature set. These custom descriptors combine electrostatic, hydrophobic, and topological characteristics of the molecule solvent-accessible surface, calculated by MSMS [209] as implemented into the open-source tool SURFMAP [210]. SURFMAP provides the electrostatic potential calculated by the Adaptive Poisson–Boltzmann Solver [109], surface hydrophobicity using the residue-based Wimley-White hydrophobicity scale [211], and surface exposure, quantified through circular variance calculation [212] to consider surface cavity formation. To further characterize the surface topography, the normalized distance to the molecule center of mass was calculated for each surface grid point. Hereafter, these protein surface characteristics (electrostatic

potential, hydrophobicity, circular variance, and normalized protein radius), were combined and summated, averaged, or binned into positive and negative contributions. An overview of the custom descriptors is detailed in the Appendix C, Table C1. The final descriptor set consists of 1374 features per molecule, including 227 global descriptors and 1147 local descriptors.

5.2.3 Chromatographic Experimentation

The chromatographic experiments involved a system and column characterization, as well as salt LGE experiments for the standardized calibration of the semi-mechanistic chromatography model as performed in our previous study [177]. These experiments were conducted using an ÄKTA Avant 25 (Cytiva) preparative chromatography system, controlled by Unicorn 7.5 (Cytiva) software, and maintained at a residence time of 5 minutes.

During the system and column characterization, extra-column effects and column-specific parameters were assessed by evaluating sensor dead volumes, column porosities, and ionic capacity. The characterization protocols were based on published protocols of Hunt et al. [213] and Huuk et al. [143]. Hereto, pulse injections of tracer solutions were applied to the system and the column at set flow rate. Subsequently, salt LGEs were performed to elute a total of 59 antibodies. Linear pH gradients performed in a previous study [206] identified a molecule-specific pH range in which salt elution was viable, reducing excessive experimentation. Three salt LGEs were then performed per molecule with a gradient length of 10, 20, and 30 CV at up to three distinct pH values (5.0, 6.0, 7.0). The column was initially equilibrated for 3 CV with equilibration buffer. After equilibration, the antibody solution was loaded onto the column up to a loading density of 1.0 g/L. Thereafter, the antibodies were eluted by linearly decreasing the sodium chloride concentration from 1560 to 60 mM within 10, 20, and 30 CV using the elution buffer. Retention times of the molecules were determined by measuring the first moments of the elution peaks through the UV trace at a wavelength of 280 nm, while recording the corresponding conductivity corrected by sensor dead volume. The column was then subjected to a 4 CV regeneration step, a 5 CV cleaning in place procedure, and a 4 CV storage step.

5.2.4 Chromatographic Simulation

Following wet-lab experimentation, the chromatographic behavior was numerically simulated using the software ChromX (Cytiva) [141]. The column mass transfer was simulated by a transport dispersive model [43,44] to account for dispersion and diffusion effects, while the adsorption-desorption mechanism was described with a linear adsorption model introduced by Hess et al. [177], shown in Eq. (5.1).

$$K_i(c_{\mathrm{p,s}}) = \widetilde{K}'_{\mathrm{eq},i} \exp(K_{\mathrm{s},i} c_{\mathrm{p,s}}) \tag{5.1}$$

The utilized adsorption model follows an exponential form comparable to the hydrophobic interaction chromatography (HIC) isotherm introduced by Mollerup [79] and effectively neglects electrostatic attractive contributions that would be observed in ion-exchange operation. The equilibrium ratio $K_i(c_{\mathrm{p,s}})$ $[-]$ represents the ratio between the adsorbed protein $i \in [1, n]$ to the solute protein and is dependent from the modulator $s \in [1, m]$ concentration within the pore volume of the resin bead $c_{\mathrm{p,s}}$ $[\mathrm{molm^{-3}}]$. The modified equilibrium constant $\widetilde{K}'_{\mathrm{eq},i} \approx k_{\mathrm{ads},i}/k_{\mathrm{des},i}$ $[-]$ is derived from the law of mass action at thermodynamic equilibrium and equals the ratio of the adsorption rate $k_{\mathrm{ads},i}$ to the desorption rate $k_{\mathrm{des},i}$ of the protein species from the ligand binding site and should be dependent on the ligand density of the resin. The salt-protein interaction parameter $K_{\mathrm{s},i}$ $[\mathrm{mol^{-1}m^3}]$ introduced by Mollerup [79] describes the thermodynamic activity of the solute protein and expresses the excess Gibbs free energy of the thermodynamic association during salt-protein interaction. Conceptually, this parameter should represent the intermolecular force of hydrophobic interaction caused by the salt-protein association.

The kinetic material balance of the multimodal adsorption process at low load density conditions can be expressed by Eq. (5.2), following the kinetic notation often employed in chromatographic simulation [83,86].

$$k_{\mathrm{kin},i} \frac{\partial q_{p,i}}{\partial t}(x,t) = \widetilde{K}'_{\mathrm{eq},i} \exp(K_{\mathrm{s},i} c_{\mathrm{p,s}})\, c_{\mathrm{p},i}(x,t) - q_{p,i}(x,t) \tag{5.2}$$

The kinetic adsorption model describes the time t [s] and location $x \in [0, L_{\mathrm{c}}]$ [m] dependent binding of the protein to the chromatographic resin. Within the pore phase of the resin, the exchange of the unbound protein concentration $c_{\mathrm{p},i}$ $[\mathrm{molm^{-3}}]$ with adsorbed protein concentration $q_{\mathrm{p},i}[\mathrm{molm^{-3}}]$ is expressed, as depicted in Fig. 5.1. Consistent with Hahn et al. [128], the kinetic coefficient is defined as $k_{\mathrm{kin},i} = k_{\mathrm{des},i}^{-1}$ $[\mathrm{s\ molm^{-3}}]$.

Linear adsorption parameters were determined according to the protocol recently published by Hess et al. [177], based on the theoretical foundation of Yamamoto et al. [76]. The estimation method relies on linear modulator gradients and requires the determination of the normalized gradient slope GH $[\mathrm{molm^{-3}}]$, defined in Eq. (5.3).

$$GH = \frac{c_{\mathrm{s}}^{\mathrm{F}} - c_{\mathrm{s}}^{\mathrm{I}}}{V_{\mathrm{G}}} (1 - \varepsilon_{\mathrm{b}}) V_{\mathrm{col}} \tag{5.3}$$

The slope of the gradient is given by the initial $c_{\mathrm{s}}^{\mathrm{I}}$ $[\mathrm{molm^{-3}}]$ and the final $c_{\mathrm{s}}^{\mathrm{F}}$ $[\mathrm{molm^{-3}}]$ salt concentrations at the start and the end of the gradient, divided by the linear gradient volume V_{G} $[\mathrm{m^3}]$. The slope is than normalized by considering the interstitial porosity of the column bed ε_{b} $[-]$ and the volume of the column V_{col} $[\mathrm{m^3}]$. According to Yamamoto [133], the salt concentration at the peak maximum $c_{\mathrm{s},i}^{\mathrm{R}}$ $[\mathrm{molm^{-3}}]$ and GH are employed in an

equilibrium chromatography model, which considers the zone movement of the components through the eluting column, shown in Eq. (5.4).

$$GH = \int_{c_s^I}^{c_{s,i}^R} \frac{dc_s}{\widetilde{K}'_{eq,i} \exp\left(K_{s,i} c_s\right) + K_{sec,i} - K_{sec,s}} \tag{5.4}$$

Eq. (5.4), includes Eq. (5.1) and is integrated between the peak and the initial salt concentration. Furthermore, the distribution coefficients $K_{sec,i} = \varepsilon_{p,i}\,[-]$ and $K_{sec,s} = \varepsilon_{p,s}\,[-]$ are considered, denoting the fraction of the pore volume accessible to the protein and the salt component at non-interacting conditions. The adsorption parameters $\widetilde{K}'_{eq,i}$ and $K_{s,i}$ can be numerically estimated when inserting the first moments of the elution peaks from the salt LGEs into Eq. (5.4). In this study, the first moments were set equal to the peak maxima due to peak symmetry and were determined by using the python package SciPy [148]. Lastly, the kinetic parameter $k_{kin,i}$ is estimated by simultaneously curve fitting the three LGE experiments for each molecule and pH value, as described by Hahn et al. [83].

5.2.5 Multiscale Model Development and Evaluation

In this study, we developed a multiscale model for predicting adsorption parameters from protein sequences using quantitative structure-property relationship modeling. Due to high dimensionality of the antibody-specific descriptor set (n=1374) and the small dataset size (m=59), descriptor preprocessing, dimensionality reduction, and model evaluation were required. The QSPR workflow was implemented using Python 3.9.12 and scikit-learn 1.0.2 [189], and modified from our previous study [206]. Gaussian process regression (GPR) was employed for nonlinear predictions and heteroscedastic uncertainty estimation [116,190].

GPR, is rooted in Bayesian inference, and operates on the assumption that similar input variables (X) will yield similar target values (y). The primary objective of GPR is to identify a Gaussian distribution of mapping functions, commonly referred to as a kernel. This kernel is used to predict the target variable from the input features with minimal uncertainty, leveraging prior knowledge of their interrelationship. Once the joint distribution of mapping functions is updated or conditioned based on the training information, the model generates predictions derived from the mean of this updated distribution. The uncertainty, on the other hand, is represented by the distribution's standard deviation. This process allows GPR to make accurate predictions while providing an individual measure of confidence in each prediction.

According to the Bayesian framework, the adsorption parameters were predicted by conditioning a prior distribution of functions $P\big(y(x)\big)$ with the training data $D = \{y, X\}$ to derive the posterior distribution $P(y(x)|\mathcal{D})$ using the Bayesian update rule, shown in Eq. (5.5).

$$P(y(x)|\mathcal{D}) \propto P(y|y(x), X)P\big(y(x)\big) \tag{5.5}$$

A mixed covariance function was chosen as the prior by multiplying a linear kernel with a Matérn class kernel and subsequently adding a white noise kernel [116,189,191]. Posterior conditioning was achieved by maximizing the likelihood $P(y|y(x), X)$ through minimization of log(marginal likelihood) using the L-BFGS-B algorithm [59].

The QSPR workflow begins with data preprocessing, where irrelevant descriptors were discarded, and several operations are performed to account for structural diversity of the IgG-like molecules. Furthermore, inaccurately determined adsorption parameters with a p-value above 0.1 are removed from the data set according to a two-sided Wald test proving the linearity of the initial parameter determination [177]. Thereafter, the dataset is randomly divided into 80% training and 20% test data, scaling the descriptors by their standard deviation (SD), and centering them based on the training data, while the adsorption parameters are transformed with the natural logarithm to assure normality of the target distribution. Two QSPR models are then created for both adsorption parameters using an identical workflow, training, and test set. Due to the strong correlation between the two adsorption parameters $\widetilde{K}'_{\mathrm{eq},i}$ and $K_{\mathrm{s},i}$, a chained regression is performed, with $\widetilde{K}'_{\mathrm{eq},i}$ predictions serving as possible feature for the $K_{\mathrm{s},i}$ parameter model, which is referred to as multioutput model. Dimensionality reduction is conducted by removing invariant descriptors, discarding low variance features, sorting remaining descriptors based on F-test results, removing collinear features, and selecting the 20 highest scoring features. Recursive feature elimination is then conducted using feature permutation importance [194]. Permutation importance is calculated as the average increase in model deviation when assessing model performance after shuffling a single feature one hundred times while keeping the remaining features constant. Model evaluation involved assessing overall model reliability and performance, feature interdependence, sensitivity, and contribution. Goodness of fit and prediction are inspected, and a 95% confidence interval estimated. Fivefold cross-validation with ten repetitions is used for internal validation, while model reliability is analyzed through y-scrambling and MAE calculation [195]. Lastly, feature interdependence, sensitivity, and contribution are assessed using pairwise relationships, feature permutation importance, and partial dependence [111,194,196].

5.3 Results and Discussion

5.3.1 Elution Behavior of Antibodies and Antibody Formats

LGE experiments were conducted using a multimodal anion exchange resin to evaluate 59 full-length antibodies across various formats and three pH values (5.0, 6.0, 7.0). The antibodies were eluted by linearly decreasing the sodium chloride concentration. An inverse

relationship between salt concentration and molecule retention was observed, indicating a hydrophobically driven adsorption process. Fig. 5.2 presents the ionic strength [M] at peak elution during a 30 CV gradient, arranged in ascending order based on pH 5.0 retention (decreasing ionic strength indicates increased multimodal retention during negative salt LGEs). Despite the structural similarities among the examined formats, antibody elution varied significantly, ranging from 1.347 [M] for IgG1 (24) at pH 6.0 to 0.069 [M] for IgG1 (16) at pH 5.0. This variability within the structurally conserved IgG1 antibodies indicates a weak format-dependent elution behavior. In contrast to the weak format-dependent elution behavior, a strong relationship was observed between antibody retention and elution pH, with retention increasing under elevated pH conditions.

While most antibodies exhibited minimal variation in retention between pH 5.0 and 6.0, a substantial increase in retention was observed from pH 6.0 to 7.0. Similar observations have been reported for model proteins on a cation exchange resin [208] and a mAb on multimodal-cation exchange resins [166]. These findings were attributed to multiple, pH-dependent binding domains on the protein surface, leading to an altered protein binding orientation [178]. The pronounced pH dependency of antibody elution resulted in selectivity reversals, exemplified by IgG1 (13) and IgG-scFv (3), enclosed in a dashed box shown in Fig. 5.2. Upon further examination of pH trends, all antibodies eluting at a salt concentration above 0.7 [M] at pH 5.0 displayed a retention decrease from pH 5.0 to 6.0, followed by a larger increase from pH 6.0 to 7.0. Conversely, antibodies eluting below 0.7 [M] demonstrated a monotonous increase in retention with rising pH. The observed phenomena suggest a pH dependency in the multimodal chromatographic retention that is sensitive to salt concentration. To further elucidate these observations, additional experimentation will be required.

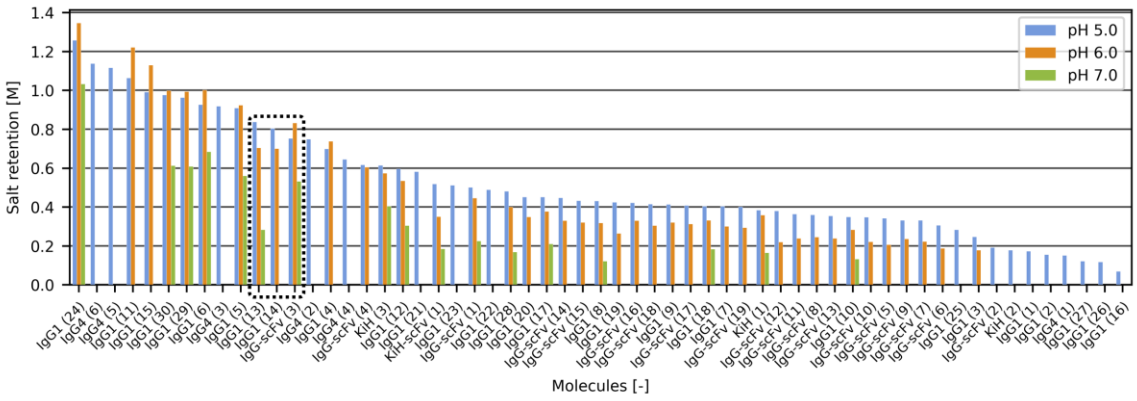

Fig. 5.2: Peak salt concentration of molecules in LGE elution experiments.The ionic strength at peak elution is shown for all molecules during 30 CV long negative sodium chloride gradients at pH 5.0, 6.0, and 7.0, arranged in ascending order based on pH 5.0 retention. A limited number of pH conditions were evaluated for some mAbs, as determined by linear pH gradients from a previous study [206]. The dashed black box highlights the pH-dependent selectivity reversal, exemplified by IgG1 (13) and IgG-scFv (3).

Following the gradient elution experiments, the equilibrium constant $\widetilde{K}'_{\mathrm{eq},i}$ and salt-protein interaction parameter $K_{\mathrm{s},i}$ of the multimodal adsorption model were calculated using Eq. (5.4). Initially, 118 parameter pairs were determined from 59 molecules at various pH levels. Out of the initial parameters, 103 parameter pairs from 55 molecules met the accuracy requirements as defined by the p-value in the parameter determination workflow, detailed in Section 5.2.5 and listed in the Appendix C, Table C4. Fig. 5.3. displays the natural logarithm of the scattered parameter pairs, providing format and pH information (a), as well as the distribution of training and test sets (b). Unimodal histograms of the adsorption parameters suggest an approximately normal distribution of the transformed parameters, framing the joint data as marginal axes. The parameter pairs appear to be broadly distributed and format independent. The logarithmic parameter $\log\left(\widetilde{K}'_{\mathrm{eq},i}\right)$ ranges from -3.848 [-] for IgG1 (1) at pH 5.0 to 1.814 [-] for IgG1 (24) at pH 7.0 ($\widetilde{K}'_{\mathrm{eq},i}$: 0.021 - 6.132 [-]). Conversely, the parameter $\log\left(K_{\mathrm{s},i}\right)$ ranges from -0.251 [-] for IgG1 (11) at pH 6.0 to 3.229 [-] for IgG1 (1) at pH 5.0 ($K_{\mathrm{s},i}$: 3.229 - 25.243 [-]). A strong negative correlation between the two parameters is evident, with a Pearson correlation coefficient p of -0.837. Parameter correlation has been observed in similar adsorption equations and is an inherent feature of the model structure used, which can impede parameter estimation [145,177]. Upon closer examination of the scattered adsorption parameters, a cluster of IgG4 antibodies is noticeable at higher $\log\left(\widetilde{K}'_{\mathrm{eq},i}\right)$ and lower $\log\left(K_{\mathrm{s},i}\right)$ values. This observation may indicate a format-specific relationship between the IgG4 structure and the multimodal elution behavior, as previously reported [206]. Furthermore, a strong and positive pH dependency of the $\log\left(\widetilde{K}'_{\mathrm{eq},i}\right)$ values is indicated by perpendicular bands of parameter pairs at constant pH on the $\log\left(\widetilde{K}'_{\mathrm{eq},i}\right)$ plane. This pattern could imply electrostatic contributions to the equilibrium constant within the multimodal adsorption model. This hypothesis is supported by the strong and negative correlation of $\log\left(\widetilde{K}'_{\mathrm{eq},i}\right)$ values with the calculated net charge of the protein surface, as shown in the Appendix C, Fig. C3, Subfigure (a8), with a Pearson correlation coefficient of -0.767. Consequently, a reduced positive surface charge at higher pH levels could decrease electrostatic repulsion towards the partially cationic ligand, thereby increasing the adsorption-to-desorption ratio, as expressed by the equilibrium constant $\widetilde{K}'_{\mathrm{eq},i}$. However, a similar pH dependency was not observed for the salt-protein interaction parameter $K_{\mathrm{s},i}$.

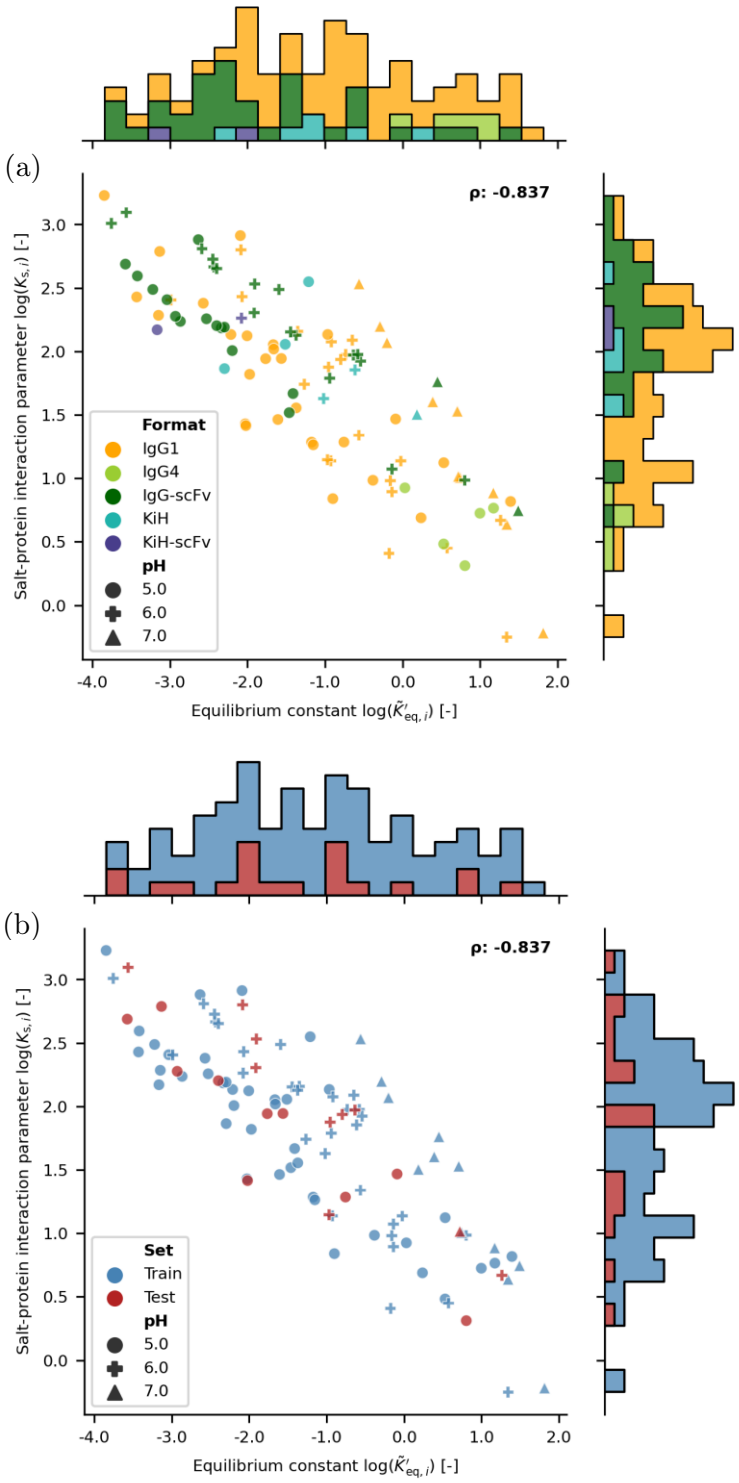

Fig. 5.3: Adsorption parameter-pair distributions from salt LGE experiments.a) Scatterplot showing antibody format- and pH-specific adsorption parameters, and b) Scatterplot presenting the distribution of model training and testing data, including pH information. The marginal axes of each subplot exhibit the unimodal histogram for each logarithmic adsorption parameter. The Pearson correlation coefficient (p) between both adsorption parameters is displayed in the upper right corner of the joint axes.

5.3.2 QSPR Modeling of Thermodynamic Adsorption Parameters

To realize the multiscale model, pH sensitive and molecule-specific parameters of the multimodal adsorption model had to be predicted from antibody structure. QSPR modeling was employed to correlate 103 parameter pairs from 55 molecules at three distinct pH values with 1374 global and local physicochemical descriptors, calculated from antibody homology models.

Initially, the data set was randomly divided into 80% training (n=44) and 20% testing (n=11) molecules. To prevent sharing structural information between the training and testing sets, different pH conditions of the molecules were grouped accordingly. Fig. 5.3 (b) displays the scattered distribution of the logarithmic equilibrium constant $\log(\widetilde{K}'_{\mathrm{eq},i})$ and salt-protein interaction parameter $\log(K_{\mathrm{s},i})$ for the training and testing data. The parameter pairs of the test set molecules, including IgG1, IgG4, and IgG-scFv formats at pH 5.0, 6.0, and 7.0, are evenly distributed within the training data, as listed in Table 5.1. Two consecutive QSPR models were built from the training data for the two adsorption parameters. The prediction of $\log(\widetilde{K}'_{\mathrm{eq},i})$ served as a potential input feature for the subsequent $\log(K_{\mathrm{s},i})$ model. This multioutput model, also known as chained regression, aimed to enhance the accuracy of the $\log(K_{\mathrm{s},i})$ predictions, given their strong correlation displayed in Fig. 5.3. During the two-staged feature selection, informative descriptors were identified by reducing the initial 1374 descriptors to seven for the $\log(\widetilde{K}'_{\mathrm{eq},i})$ model and five for the $\log(K_{\mathrm{s},i})$ model, as can be assessed by the results of the recursive feature elimination shown in the Appendix C, Fig. C1.

After model training, the two QSPR models were externally validated by predicting 22 parameter pairs of the test set molecules at multiple pH values, as listed in Table 5.1. The goodness of fit to the training data and the prediction accuracy for the test data is shown for $\log(\widetilde{K}'_{\mathrm{eq},i})$ (a) and $\log(K_{\mathrm{s},i})$ (b) in Fig. 5.4. High training accuracy was achieved for both logarithmic adsorption parameters, with R^2 values of 0.933 and 0.941 for $\log(\widetilde{K}'_{\mathrm{eq},i})$ and $\log(K_{\mathrm{s},i})$, respectively. The test set predictions exhibited reasonable accuracy, with Q^2 values of 0.834 and 0.835, and mean absolute errors (MAE) of 0.453 and 0.237 for the logarithmic parameters. Minor deviations in model prediction could originate from inaccuracies in the underlying descriptor models or adsorption parameter determination. Experimental parameter uncertainties from the standardized calibration workflow of up to 91% for $\widetilde{K}'_{\mathrm{eq},i}$ determination and 20% for $K_{\mathrm{s},i}$ determination had to be accounted for, as observed for IgG-scFv (6) at pH 6.0, listed in the Appendix C, Table C4.

Table 5.1: Observed and predicted thermodynamic model parameters of test set molecules after back transformation.

Molecule	pH	Obs. $\widetilde{K}'_{eq,i}$ [−]	Pred. $\widetilde{K}'_{eq,i}$ [−]	Residual $\widetilde{K}'_{eq,i}$ [%]	Obs. $K_{s,i}$ [M]	Pred. $K_{s,i}$ [M]	Residual $K_{s,i}$ [%]
IgG1 (3)	5.0	0.044	0.073	67.434	16.239	12.782	-21.287
	6.0	0.124	0.157	26.094	16.452	12.188	-25.918
IgG1 (14)	5.0	0.469	0.406	-13.375	3.617	3.558	-1.649
	6.0	3.538	1.346	-61.954	1.950	2.597	33.196
IgG1 (20)	5.0	0.209	0.278	32.769	6.987	6.489	-7.130
	6.0	0.449	0.818	82.156	6.923	6.004	-13.278
IgG1 (22)	5.0	0.914	1.199	31.247	4.337	7.558	74.274
IgG1 (28)	5.0	0.171	0.273	59.693	6.976	5.496	-21.213
	6.0	0.385	0.921	139.177	6.536	3.374	-48.377
IgG1 (30)	5.0	0.132	0.152	14.642	4.118	4.200	2.008
	6.0	0.380	0.457	20.211	3.143	3.112	-0.992
	7.0	2.054	2.222	8.200	2.760	2.312	-16.243
IgG4 (6)	5.0	2.233	1.492	-33.199	1.366	1.809	32.455
IgG-scFv (6)	5.0	0.028	0.036	30.259	14.698	11.399	-22.443
	6.0	0.028	0.087	206.069	22.055	13.005	-41.032
IgG-scFv (11)	6.0	0.149	0.092	-37.920	12.564	15.669	24.712
IgG-scFv (15)	5.0	0.053	0.070	31.117	9.742	9.608	-1.382
	6.0	0.147	0.226	53.548	10.008	7.507	-24.993
IgG-scFv (18)	5.0	0.091	0.074	-18.133	9.050	10.110	11.716
	6.0	0.531	0.168	-68.399	7.183	9.867	37.356

Improving the standardized calibration workflow to increase data quality could significantly enhance the model predictions. However, the 95% confidence interval (CI_{95}) of the training and test set predictions captured the model deviations, demonstrating the reliability of GPR to capture model uncertainty based on its training data. The reliability of the two QSPR models was further emphasized by the results of the internal model validation with a cross-validated Q^2 of 0.815 and 0.835 for $\log(\tilde{K}'_{eq,i})$ and $\log(K_{s,i})$, as well as the results of y-scrambling, shown in the Appendix C, Fig. C2. The absence of systematic model deviations was indicated by homogeneously dispersed residuals, as shown in the lower sections of Fig. 5.4, (a) and (b).

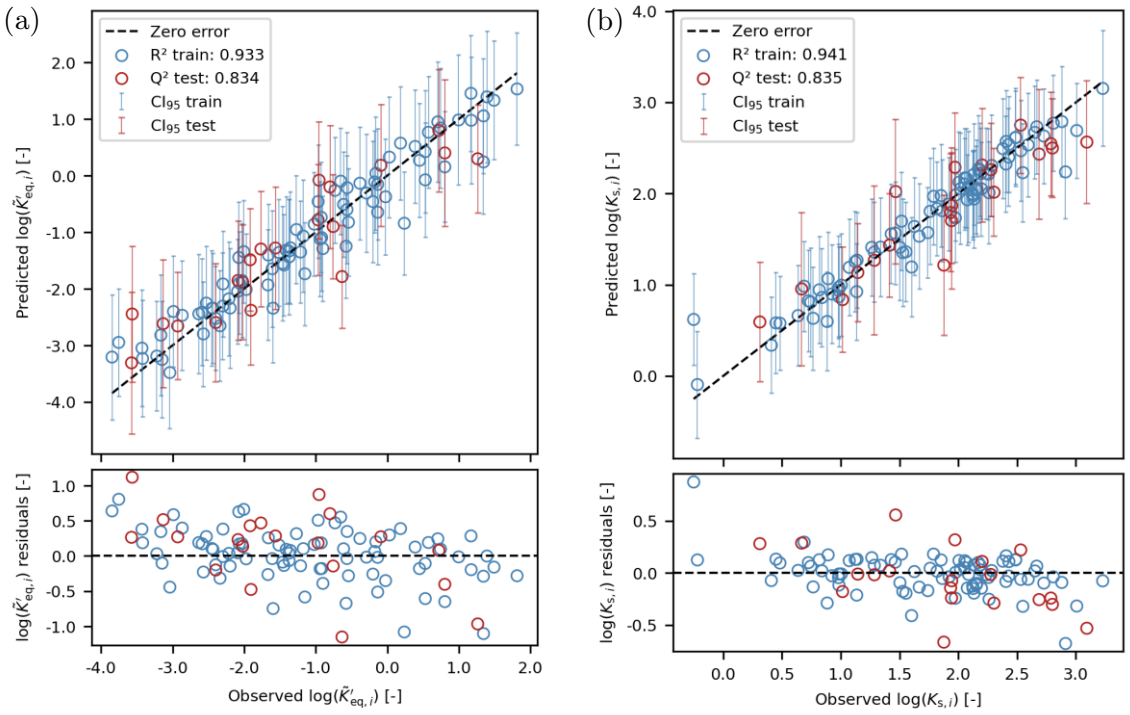

Fig. 5.4: Evaluation of adsorption parameter fit and prediction (R^2, Q^2).a) Logarithmic equilibrium constant $log(\tilde{K}'_{eq,i})$ model, and b) Logarithmic salt-protein interaction parameter $log(K_{s,i})$ model. The top section of each subfigure compares predicted and experimentally derived adsorption parameters for the molecules in the random training and testing data post-logarithmic transformation. An ideal model is depicted by a straight line, indicating zero error between predicted and experimental observations. The 95% confidence interval (CI_{95}) of prediction is calculated for each molecule. The figure's lower section presents a residual plot, with the y-axis normalized to the absolute deviation of model prediction.

After validation of the QSPR models, an inspection of the descriptor contributions to the model predictions was conducted. The descriptor contributions were expressed through their partial dependencies, a method for examining nonlinear feature interactions within a multivariate model. Fig. 5.5 displays the partial dependencies for $\log(\widetilde{K}'_{\mathrm{eq},i})$ (a) and $\log(K_{\mathrm{s},i})$ (b), which capture the target relationships depicted in the Appendix C, Fig. C3 and Fig. C4. In brief, a single factor perturbation is performed by marginalizing all features except one and permutating it within the full feature range. Recording the output predictions yields the partial dependence of the descriptor. Additionally, feature permutation importance was calculated for each descriptor, as displayed on the upper right of the subfigures. Permutation importance measures model deterioration by shuffling individual features, enabling an assessment of the feature significance within a multivariate model.

In general, the partial dependencies of the test set molecules shown in Fig. 5.5 were comparable to the training data for both QSPR models, covering dense regions within the feature space as indicated by the decile lines on the bottom of the subfigures. The descriptors of the $\log(\widetilde{K}'_{\mathrm{eq},i})$ model primarily exhibited linear relationships towards the target variable and consisted of global (a1, a2, a7) and local (a5, a6) charge characteristics that emphasize the antibody variable region (a3, a4). In contrast, the descriptors of the $\log(K_{\mathrm{s},i})$ model displayed nonlinear contributions and considered the molecule's global hydrophobicity, expressed by a supplemented descriptor (b4). The descriptors of the $\log(K_{\mathrm{s},i})$ model displayed similarities to the $\log(\widetilde{K}'_{\mathrm{eq},i})$ descriptors (b2, b3, b5) and even included the $\log(\widetilde{K}'_{\mathrm{eq},i})$ predictions (b1), which is unsurprising given the strong mutual correlation between the adsorption parameters discussed earlier. Moreover, the significance of the HFR3 region, an area in-between the H2 and H3 loop of the antibody variable region, was identified for both adsorption parameters (a3, b2). A detailed derivation of the utilized descriptors is given by Sankar et al. [110] and in the Appendix C, Table C1.

Interpreting the descriptor contributions of the adsorption parameters, the back-transformed equilibrium constant $\widetilde{K}'_{\mathrm{eq},i}$ may capture an exponential electrostatic effect of positive repulsion or reduced negative attraction of the entire protein surface towards the anionic multimodal resin with emphasis on the antibody variable region. This effect is indicated by strong negative electrostatic descriptor correlations and importance in Fig. 5.5 (a1, a2, a5, a6, a7) and was already speculated about in Section 5.3.1. Interestingly, the results of our previous QSPR study [179] conducted for cation exchange chromatography, identified similar features for predicting the logarithmic equilibrium constant of the stoichiometric displacement model [81,214]. This finding could indicate the generalizability of the identified antibody characteristics to be relevant for other chromatographic resins utilizing electrostatic interactions. Therefore, improvements to the multimodal adsorption model should incorporate a pH extension of the $\widetilde{K}'_{\mathrm{eq},i}$ parameter, already introduced for ion exchange adsorption models [213,215].

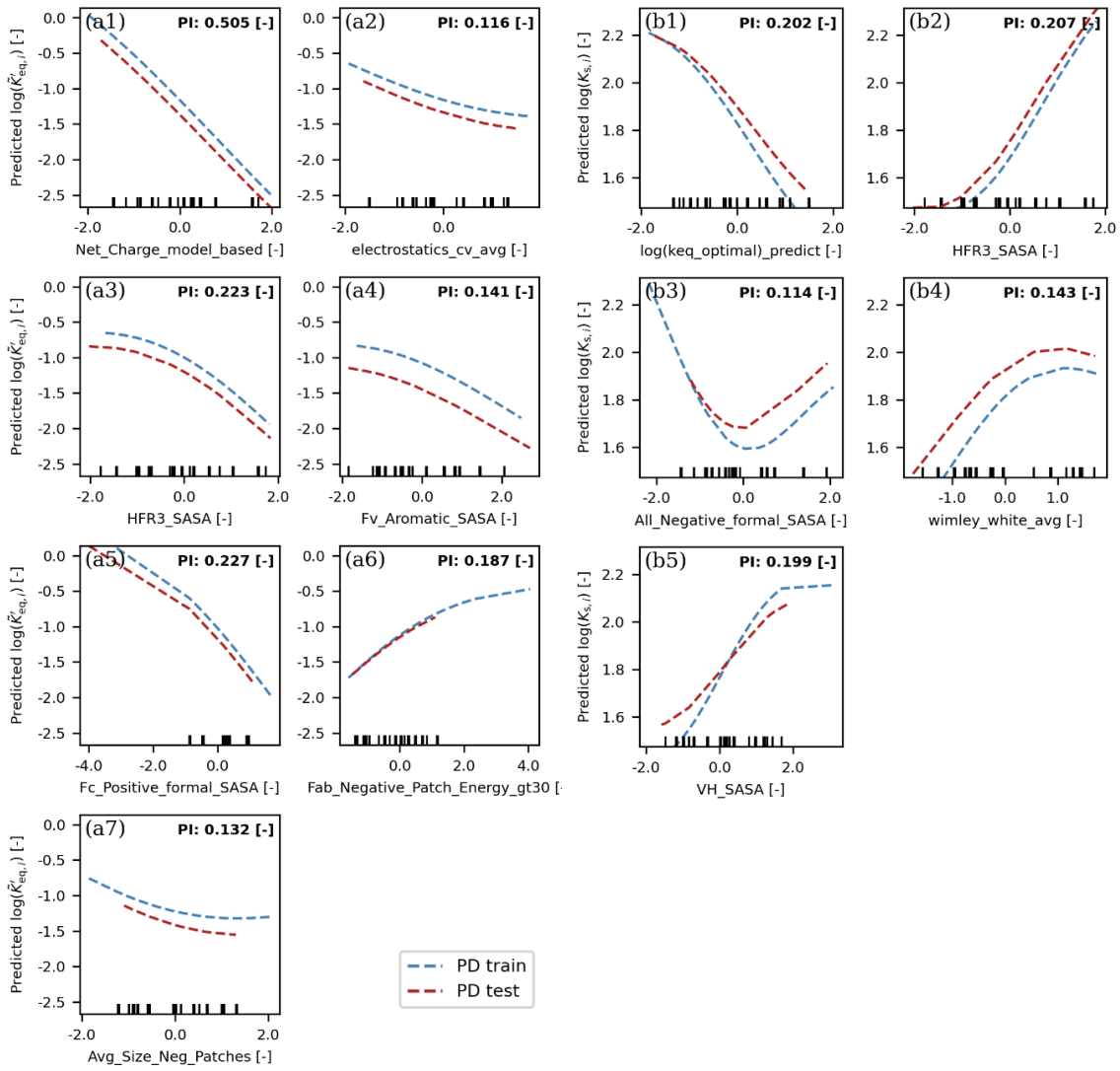

Fig. 5.5: Analysis of model features' partial dependence and permutation importance for training and testing data.(a) Logarithmic equilibrium constant $\log(\tilde{K}'_{\mathrm{eq},i})$ model, and (b) Logarithmic salt-protein interaction parameter $\log(K_{\mathrm{s},i})$ model. Each subfigure illustrates the model's response to feature permutations within their normalized range. The partial dependence (PD) displays the average target prediction for each feature. Permutation importance (PI) for target prediction is presented in the top-right corner of each subplot. Decile lines at each subplot's base indicate the frequency of feature values in the dataset.

Conversely, the salt-protein interaction parameter $K_{\mathrm{s},i}$ suggests a more complex mechanism, potentially lumping multiple characteristics of significant binding domains into a single parameter. One such characteristic is the average surface hydrophobicity calculated after Wimley and White [211], which has recently proven to be a reliable predictor for HIC retention [216] and should be included as a standard descriptor in future QSPR studies. Lastly, the significance of the HFR3 region for antibody binding to multimodal resins has not yet been reported previously and warrants investigation in future studies.

5.3.3 Chromatographic Process Prediction

To validate the multiscale modeling method, the predicted adsorption parameters $\widetilde{K}'_{\mathrm{eq},i}$ and $K_{\mathrm{s},i}$ of the test set molecules were employed in chromatographic simulations and compared against the measurement data. A macroscopic transport dispersive model was utilized for the chromatographic simulation, as implemented in our previous study [177], with $k_{\mathrm{kin},i}$ being determined by curve fitting to the experimental chromatograms. The results of the system and column characterization, as well as mass transfer parameters are provided in the Appendix C, Table C2 and Table C3. Fig. 5.6 presents an overlay of the simulated and the measured elution of IgG1 (30) at pH 5.0, 6.0, and 7.0 during three different salt gradients with gradient lengths of 10, 20, and 30 CV, as shown in Subfigure a) to c). IgG1 (30) was selected for representation as the only test set molecule with experimental data available at all pH values. The chromatograms of the remaining test set molecules can be found in the Appendix C, Fig. C5 to Fig. C14. A good agreement between the experimental and the simulated UV traces is evident in Fig. 5.6 across all operating conditions.

Interestingly, the retention of IgG1 (30) was overpredicted at pH 5.0 and 6.0 and underpredicted at pH 7.0. Upon evaluating the observed and predicted adsorption parameters in Table 5.1, an underprediction of the salt-protein interaction parameter $K_{\mathrm{s},i}$ at pH 7.0 of -16.2% is evident in comparison to the other pH conditions, which explains the deviating model simulations. On the other hand, the chromatographic simulations appeared less sensitive to the equilibrium constant $\widetilde{K}'_{\mathrm{eq},i}$, as previously observed by Altern et al. [86] employing a similar adsorption equation. The $K_{\mathrm{s},i}$ underprediction might be associated with the strong increase of antibody retention from pH 6.0 to 7.0, as discussed in Section 5.3.1. This increase could result from a transition of significant binding domains on the protein surface towards the multimodal ligands, which are challenging to be captured using static physicochemical descriptors and could translate to deteriorated model predictions. MD simulation of protein-ligand interaction could be one way of resolving this limitation as demonstrated by Banerjee et al. [217], but are currently not feasible for large multi-domain proteins. The validation of the remaining test set molecules, displayed in the Appendix C, Fig. C5 to Fig. C14, revealed good agreement for all antibodies and process conditions, except for IgG1 (22) and IgG1 (28) at pH 6.0, as shown in Fig. C8 and Fig. C9. Notably, IgG1 (22) and IgG1 (28) also constituted the largest deviations in $K_{\mathrm{s},i}$ predictions, with residuals of 74.3% and -48.4% listed in Table 5.1, explaining the offset to the experimental chromatograms.

In conclusion, the presented multiscale model effectively predicts chromatographic behavior for a diverse set of therapeutic antibodies in a linear adsorption regime based solely on sequence information. Moreover, the incorporation of pH-sensitive descriptors into a mechanistic adsorption model enables the prediction of a broad range of chromatographic operating conditions, representing a significant step towards *a priori* process development for multimodal chromatography.

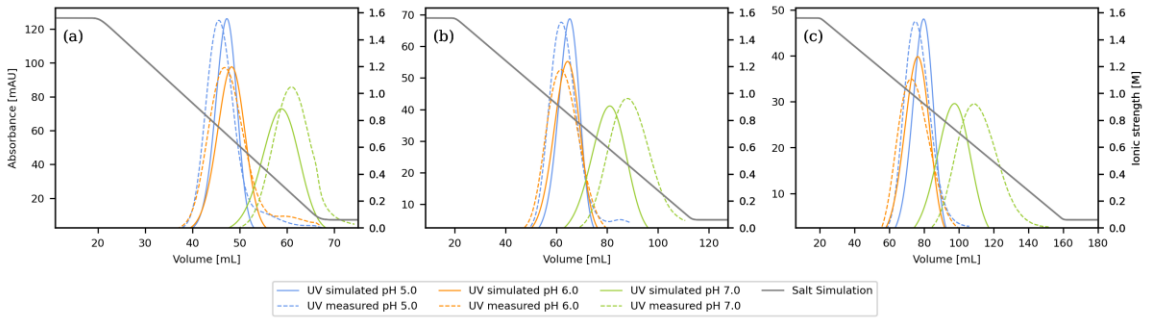

Fig. 5.6: Comparison of experimental and predicted chromatograms for IgG1 (30) under low loading density conditions.Subfigure (a) to (c) show the 10, 20, and 30 CV LGE experiments for the test set molecule at pH levels of 5.0, 6.0, and 7.0, respectively. These chromatograms are generated using multiscale modeling, which involves predicting thermodynamic parameters from antibody structure through QSPR models. These parameters are then incorporated into a transport dispersive model for chromatographic simulation.

5.4 Conclusion

In this study, we successfully demonstrated the potential of a multiscale approach to predict the chromatographic behavior of a broad variety of therapeutic antibodies based solely on sequence information. LGE experiments were conducted on the multimodal resin Capto adhere for 59 full-length antibodies including five different formats and multispecific functionality. Multimodal adsorption parameters were determined across multiple pH values, consisting of the equilibrium constant $\widetilde{K}'_{\mathrm{eq},i}$ and the salt-protein interaction parameter $K_{\mathrm{s},i}$ in a linear adsorption regime. QSPR modeling was employed to associate the adsorption parameters with physicochemical descriptors calculated from antibody homology models. The predicted adsorption parameters were incorporated into a semi-mechanistic chromatography model and validated experimentally.

The results revealed a strong correlation between antibody retention and elution pH as well as molecule specific interactions to the multimodal resin while the importance of the antibody format appeared less significant. The QSPR models exhibited high training accuracy and good test set prediction accuracy. The descriptor contributions of the adsorption parameters suggested the equilibrium constant to capture the electrostatic effects of the protein surface towards the anionic multimodal resin, while the salt-protein interaction parameter could represent a more intricate mechanism involving global surface hydrophobicity and significant binding domains. These finding could be generalizable to other chromatographic modes given the similarity of the selected antibody characteristics to previous QSPR studies. The chromatographic process predictions of the test set molecules showed good agreement with experimental data over a wide range of antibody formats and operating conditions.

This multiscale approach, incorporating pH-sensitive descriptors into an adsorption model, enables the prediction of multiple chromatographic operations and pH conditions (e.g., linear gradient, step gradient, or flow through chromatography). Thus, the presented model facilitates *a priori* process development for a variety of therapeutic antibodies in multimodal chromatography. As the underlining adsorption model is limited to low-loading density conditions, the multiscale model predictions are particularly useful for earlier rather than later stages of process development by specifying likely operating windows.

Increasing the amount of high-quality data and exploring the transition of significant binding domains on the protein surface could lead to potential model improvements. Future research should explore methods of quantifying the interaction of significant binding domains in multimodal chromatography like the impact of the HFR3 region or the role of adjacent surface patches in antibody binding. Next to these model improvements our results could facilitate the development of improved adsorption models for multimodal chromatography. Given the complexity of multimodal chromatography, our multiscale approach can serve as a template for unraveling complex physicochemical phenomena using structural information.

Acknowledgements

The authors would like to thank the department of Development Biologicals (Boehringer Ingelheim, Biberach, Germany) and the Biotherapeutics Discovery department Biologics Research CMC (Boehringer Ingelheim, Ridgefield, USA) and especially Michael Marlow, Helen Wu, and James Hu, for providing the therapeutic antibodies. Furthermore, we would like to thank the BI DSP modeling team for their constant support, as well as the long and fruitful discussions.

6 Conclusion and Outlook

The biopharmaceutical industry continuously seeks to expedite its process development and reduce production costs by adopting innovative technologies. MMC has emerged as a promising technique for purifying therapeutic antibodies. However, its development often requires labor-intensive screening experiments due to limited insight of the complex protein-ligand interaction. This dissertation aimed to facilitate the accelerated adoption of MMC in antibody purification by developing a multiscale chromatographic model to increase its process understanding and accelerate its process development. The study's findings demonstrate the feasibility of *a priori* process development in MMC by linking molecular information with the molecule's chromatographic behavior. For the first time, MMC behavior of antibodies in low loading conditions could be predicted using only the information provided by their amino acid sequence. This enhanced insight could enable a knowledge-driven DSP development, focusing on the physicochemical characteristics of target molecules rather than relying on empirical observations.

6.1 A Priori Development in Biopharmaceutical Downstream Processing

A priori process development in biotherapeutic purification offers the potential for *in silico* process characterization and optimization without the need for wet lab experimentation. In the context of chromatographic purification, this could involve selecting the optimal resin matrix, separation mode, and operating conditions (e.g., buffer composition, loading density, and temperature) to achieve the highest product purity and process yield. However, developing such a model is currently unfeasible and may require significant advancements in mechanistic understanding, computational performance, and data availability. Despite the limitations discussed in the following, multiscale modeling represents a promising path towards an *a priori* process development.

Multiscale modeling can be achieved by extending macroscopic process models with molecular information through a QSPR framework. The extent of this modeling approach is limited by the applicability domain of the training data and the capabilities of the macroscopic model in place. Constraints on the AD may include examining a structurally homologous class of molecules and restricting the analysis to a single absorber and buffer system. Conversely, the capabilities of the macroscopic model might be limited to constant

process conditions, such as loading density, temperature, or operation mode, as each additional parameter increases the effort required for model calibration. This study focused on a common chromatographic ligand used in purifying full-length antibodies with a multicomponent buffer system and an operation principle broadly applicable in antibody purification. To develop the multiscale model, several constraints were introduced, and various challenges were addressed that will be critically discussed below.

Initially, a standardized method for mechanistic modeling of MMC had to be established to accelerate the generation of training data. Although highly parameterized equations for describing mass transfer and thermodynamic adsorption in MMC had already been published, a comparably large number of calibration experiments were needed to obtain macroscopic parameters. Moreover, the complex and semi-mechanistic structure of these models resulted in poor parameter identifiability and model performance, disqualifying them from gathering sufficient training data with limited sample material and within a reasonable time frame. Consequently, a new methodology for modeling MMC was developed in Chapter 3, targeting a balance between calibration efficiency and predictability of process conditions. The complexity of the underlying isotherm equation was significantly reduced by constraining the model applicability to its context of use, specifically the simulation of multimodal anion exchange chromatography in flow through operation. As a result, the model was confined in simulating low loading density conditions at various pH levels in an electrostatically repulsive environment. This approach allowed for an analytical model calibration using a standardized set of chromatographic experiments, which was validated by simulating chromatographic modes and buffer ionic strengths for different antibody formats at a given set point pH.

Subsequently, a structure-dependent tool had to be established for deriving the set point pH, enabling *a priori* development within a relatively broad pH window. Due to the lack of physics-informed models of the structure-pH relationship in MMC, an empirical framework had to be utilized. A sequence-based QSPR workflow was developed in Chapter 4, predicting the optimal process pH from antibody structure. Additionally, the results of this chapter facilitated the automated chromatographic experimentation and prediction of structural models and physicochemical descriptors, laying the groundwork for the final multiscale model. While using an empirical approach allowed for model application in a wide pH range without prior knowledge of the input-output relationship, its applicability was confined to the boundaries of available training data. This necessitated the acquisition of a large and structurally diverse set of antibodies to address future purification challenges and enhance model reliability. Furthermore, an antibody-specific descriptor set was employed to elucidate the intricate pH relationship while improving model performance. For this purpose, Gaussian process regression was used, a Bayesian approach that enables model prediction with improved uncertainty estimation compared to frequentist regression algorithms. Despite reliable uncertainty estimations, the black-box characteristics, and inaccuracies of QSPR models limit their applicability to earlier stages of process development or *in silico* manufacturability assessment.

Finally, the mechanistic and statistical approaches were integrated within Chapter 5 by predicting chromatographic processes from antibody structure using multiscale modeling. The standardized approach facilitated the automated acquisition of isotherm parameters and physicochemical descriptors to generate the necessary training data at different pH environments. The sequence-based QSPR workflow introduced in Chapter 4 was expanded to generate multioutput predictions of two isotherm parameters, while the standardized mechanistic modeling presented in Chapter 3 enabled prediction of MMC operation. In conjunction with the structurally derived set point pH, the current model enables *a priori* process development of biotherapeutics. However, it is important to note that this approach is only valid for predicting low-load chromatographic processes of antibody-like molecules on one specific multimodal resin and within a confined pH window. These constraints limit the model's use case to assessing protein manufacturability or as an *in silico* screening tool for deriving initial set point conditions. Accordingly, the model is more applicable for earlier stages of process development, as later process optimization would require the simulation of unknown impurity profiles and high loading density conditions.

Despite these limitations, the final multiscale model significantly contributes to process development by exploring a wide range of operating conditions for unknown antibody structures without the need for extensive wet lab experimentation. Simultaneously, it substantially increases the understanding of a highly complex mechanism in pharmaceutical manufacturing, multimodal protein-ligand interaction. As a result, the implementation of MMC in antibody purification can be facilitated. Lastly, this approach sets the foundation for future multiscale studies, given the exceptional complexity of MMC compared to other biopharmaceutical unit operations.

6.2 Future Research Perspectives

Leveraging structural information in the development of biological products presents a promising path towards a more efficient pharmaceutical industry. Utilizing molecule structure has already demonstrated significant success in the development of chemical compounds. As a result, research efforts are now focused on creating structure-based applications across all stages of the biologics development cycle, with DSP being a relatively late adopter. While empirical methods like QSPR offer straightforward solutions for harnessing structural information, the goal should be the implementation of first principal models driven by process understanding. To achieve a mechanistic understanding of biophysical phenomena, statistical modeling can serve as an intermediate step to unravel complex interrelations, such as multimodal protein-ligand interactions. Thus, multiscale modeling can enhance existing process models by validating and expanding conceptual parameters of mechanistic models through the incorporation of protein characteristics, beyond its immediate benefits for *a priori* development.

As the development of multiscale models and other structure-based solutions progresses, the increasing accessibility of statistical modeling and structure prediction tools shifts the bottleneck towards data acquisition. Future research should address the challenge of extracting maximum information from experimental and structural data with minimal resource expenditure. The aim should be to broaden the AD of QSPR models and improve model predictions. In the context of chromatographic purification, this may involve informative experimental design and high-throughput screening for automated assessment of binding behavior across various chromatographic resins, buffer environments, protein concentrations, and other parameters. Simultaneously, advancements in *in silico* data acquisition should aim to increase the physicochemical information derived from structural models. This can be achieved by developing innovative descriptor sets in conjunction with molecular or quantum dynamics simulations. Specifically, descriptors should be created to quantify binding hotspots, such as adjacent and multimodal surface attributes or individual domain contributions. This would enable a generalizable yet high-resolution description of physicochemical characteristics for a wide range of protein classes. Improved data acquisition is expected to advance mechanistic understanding of buffer-protein relationships, as well as the effects of resin matrix and ligand structure on protein interactions. This could lift the current constraints of multiscale modeling discussed in the previous section like the limitation to a single component system in a linear adsorption regime.

Given the substantial resources required for implementing multiscale models, alternative applications for utilizing structural information in DSP development should be explored. A focus on broadly applicable QSPR models may be a viable option during early process development. For instance, such models could enhance the parameter space of traditional HTS workflows by narrowing down potential resins or buffer conditions, considering the low accuracy requirements of the initial screening experimentation. Conversely, utilizing structural information to improve mechanistic process models could enhance the accuracy and confidence of structure-based predictions in later stages of process development. This would facilitate *a priori* optimization of impurity clearance but would also necessitate structural information of potential contaminants. The availability of such structures has been partially addressed by deep learning tools like AlphaFold that can be considered a step forward in solving the protein folding problem [104]. However, identifying and characterizing the physicochemical properties of process and product-related impurities remains a challenge for enabling a generic process optimization that should be a focus of future studies.

Moving forward, the distinction between statistical AD and mechanistic model capabilities should become less pronounced, potentially enabling physics-informed models of protein behavior at various buffer environments or ligand structures. We can envision a future where DSP design adopts a bottom-up approach, relying solely on product and contaminant structures to enable a true *a priori* process development.

Bibliography

[1] R.A. Rader, (Re)defining biopharmaceutical, Nat. Biotechnol. 26 (2008) 743–751. https://doi.org/10.1038/nbt0708-743.

[2] G. Walsh, E. Walsh, Biopharmaceutical benchmarks 2022, Nat. Biotechnol. 40 (2022) 1722–1760. https://doi.org/10.1038/s41587-022-01582-x.

[3] C. Morrison, Fresh from the biotech pipeline—2019, Nat. Biotechnol. 38 (2020) 126–131. https://doi.org/10.1038/s41587-019-0405-7.

[4] M.S. Castelli, P. McGonigle, P.J. Hornby, The pharmacology and therapeutic applications of monoclonal antibodies, Pharmacol. Res. Perspect. 7 (2019) e00535. https://doi.org/10.1002/prp2.535.

[5] N.S. Cauchon, S. Oghamian, S. Hassanpour, M. Abernathy, Innovation in Chemistry, Manufacturing, and Controls—A Regulatory Perspective From Industry, J. Pharm. Sci. 108 (2019) 2207–2237. https://doi.org/10.1016/j.xphs.2019.02.007.

[6] H.F. Liu, J. Ma, C. Winter, R. Bayer, Recovery and purification process development for monoclonal antibody production, MAbs. 2 (2010) 480–499. https://doi.org/10.4161/mabs.2.5.12645.

[7] S. Goswami, W. Wang, T. Arakawa, S. Ohtake, Developments and Challenges for mAb-based Therapeutics, Antibodies. 2 (2013) 452–500. https://doi.org/10.3390/antib2030452.

[8] S. Oshinbolu, L.J. Wilson, W. Lewis, R. Shah, D.G. Bracewell, Measurement of impurities to support process development and manufacture of biopharmaceuticals, TrAC Trends Anal. Chem. 101 (2018) 120–128. https://doi.org/10.1016/j.trac.2017.10.026.

[9] A.A. Shukla, B. Hubbard, T. Tressel, S. Guhan, D. Low, Downstream processing of monoclonal antibodies—Application of platform approaches, J. Chromatogr. B. 848 (2007) 28–39. https://doi.org/10.1016/j.jchromb.2006.09.026.

[10] S.S. Farid, M. Baron, C. Stamatis, W. Nie, J. Coffman, Benchmarking biopharmaceutical process development and manufacturing cost contributions to R&D, MAbs. 12 (2020) 1754999. https://doi.org/10.1080/19420862.2020.1754999.

[11] J.A. DiMasi, H.G. Grabowski, R.W. Hansen, Innovation in the pharmaceutical industry: New estimates of R&D costs, J. Heal. Econ. 47 (2016) 20–33. https://doi.org/10.1016/j.jhealeco.2016.01.012.

[12] H. Kaplon, S. Crescioli, A. Chenoweth, J. Visweswaraiah, J.M. Reichert, Antibodies to watch in 2023, MAbs. 15 (2023) 2153410. https://doi.org/10.1080/19420862.2022.2153410.

[13] E. Valeur, S.M. Guéret, H. Adihou, R. Gopalakrishnan, M. Lemurell, H. Waldmann, T.N. Grossmann, A.T. Plowright, New Modalities for Challenging Targets in Drug Discovery, Angew. Chem. Int. Ed. 56 (2017) 10294–10323. https://doi.org/10.1002/anie.201611914.

[14] M.S. Kinch, Z. Kraft, T. Schwartz, 2021 in review: FDA approvals of new medicines, Drug Discov. Today. 27 (2022) 2057–2064. https://doi.org/10.1016/j.drudis.2022.04.010.

[15] D.N. Lakdawalla, Economics of the Pharmaceutical Industry, J. Econ. Lit. 56 (2018) 397–449. https://doi.org/10.1257/jel.20161327.

[16] S.M. Cramer, M.A. Holstein, Downstream bioprocessing: recent advances and future promise, Curr. Opin. Chem. Eng. 1 (2011) 27–37. https://doi.org/10.1016/j.coche.2011.08.008.

[17] A.G. Lopes, Single-use in the biopharmaceutical industry: A review of current technology impact, challenges and limitations, Food Bioprod. Process. 93 (2015) 98–114. https://doi.org/10.1016/j.fbp.2013.12.002.

[18] J. Pollock, J. Coffman, S.V. Ho, S.S. Farid, Integrated continuous bioprocessing: Economic, operational, and environmental feasibility for clinical and commercial antibody manufacture, Biotechnol. Prog. 33 (2017) 854–866. https://doi.org/10.1002/btpr.2492.

[19] L. Yu, L. Zhang, Y. Sun, Protein behavior at surfaces: Orientation, conformational transitions and transport, J. Chromatogr. A. 1382 (2015) 118–134. https://doi.org/10.1016/j.chroma.2014.12.087.

[20] R. Bhambure, K. Kumar, A.S. Rathore, High-throughput process development for biopharmaceutical drug substances, Trends Biotechnol. 29 (2011) 127–135. https://doi.org/10.1016/j.tibtech.2010.12.001.

[21] B.K. Nfor, T. Ahamed, G.W.K. van Dedem, P.D.E.M. Verhaert, L.A.M. van der Wielen, M.H.M. Eppink, E.J.A.X. van de Sandt, M. Ottens, Model-based rational methodology for protein purification process synthesis, Chem. Eng. Sci. 89 (2013) 185–195. https://doi.org/10.1016/j.ces.2012.11.034.

[22] F.K. Insaidoo, M.A. Rauscher, S.J. Smithline, N.C. Kaarsholm, B.P. Feuston, A.D. Ortigosa, T.O. Linden, D.J. Roush, Targeted purification development enabled by computational biophysical modeling, Biotechnol. Prog. 31 (2015) 154–164. https://doi.org/10.1002/btpr.2023.

[23] D.K. Babi, J. Griesbach, S. Hunt, F. Insaidoo, D. Roush, R. Todd, A. Staby, J. Welsh, F. Wittkopp, Opportunities and challenges for model utilization in the biopharmaceutical industry: current versus future state, Curr. Opin. Chem. Eng. 36 (2022) 100813. https://doi.org/10.1016/j.coche.2022.100813.

[24] A.S. Rathore, R. Bhambure, V. Ghare, Process analytical technology (PAT) for biopharmaceutical products, Anal. Bioanal. Chem. 398 (2010) 137–154. https://doi.org/10.1007/s00216-010-3781-x.

[25] A.S. Rathore, Quality by Design (QbD)-Based Process Development for Purification of a Biotherapeutic, Trends Biotechnol. 34 (2016) 358–370. https://doi.org/10.1016/j.tibtech.2016.01.003.

[26] R.M. Haleem, M.Y. Salem, F.A. Fatahallah, L.E. Abdelfattah, Quality in the pharmaceutical industry – A literature review, Saudi Pharm. J. 23 (2015) 463–469. https://doi.org/10.1016/j.jsps.2013.11.004.

[27] G. Zhao, X.-Y. Dong, Y. Sun, Ligands for mixed-mode protein chromatography: Principles, characteristics and design, J. Biotechnol. 144 (2009) 3–11. https://doi.org/10.1016/j.jbiotec.2009.04.009.

[28] K. Kallberg, H. Johansson, L. Bulow, Multimodal chromatography: An efficient tool in downstream processing of proteins, Biotechnol. J. 7 (2012) 1485–1495. https://doi.org/10.1002/biot.201200074.

[29] A.T. Hanke, M. Ottens, Purifying biopharmaceuticals: knowledge-based chromatographic process development, Trends Biotechnol. 32 (2014) 210–220. https://doi.org/10.1016/j.tibtech.2014.02.001.

[30] P. Baumann, J. Hubbuch, Downstream process development strategies for effective bioprocesses: Trends, progress, and combinatorial approaches, Eng. Life Sci. 17 (2017) 1142–1158. https://doi.org/10.1002/elsc.201600033.

[31] S. Singh, N.K. Kumar, P. Dwiwedi, J. Charan, R. Kaur, P. Sidhu, V.K. Chugh, Monoclonal Antibodies: A Review, Curr. Clin. Pharmacol. 13 (2018) 85–99. https://doi.org/10.2174/1574884712666170809124728.

[32] G. Vidarsson, G. Dekkers, T. Rispens, IgG Subclasses and Allotypes: From Structure to Effector Functions, Front. Immunol. 5 (2014) 520. https://doi.org/10.3389/fimmu.2014.00520.

[33] A. Antibody, Antibody Structure | Absolute Antibody, (2024). https://absoluteantibody.com/antibody-resources/antibody-overview/antibody-structure/ (accessed March 9, 2024).

[34] W. Wang, S. Singh, D.L. Zeng, K. King, S. Nema, Antibody Structure, Instability, and Formulation, J. Pharm. Sci. 96 (2007) 1–26. https://doi.org/10.1002/jps.20727.

[35] L. Liu, Antibody Glycosylation and Its Impact on the Pharmacokinetics and Pharmacodynamics of Monoclonal Antibodies and Fc-Fusion Proteins, J. Pharm. Sci. 104 (2015) 1866–1884. https://doi.org/10.1002/jps.24444.

[36] S.M. Lippow, K.D. Wittrup, B. Tidor, Computational design of antibody-affinity improvement beyond in vivo maturation, Nat. Biotechnol. 25 (2007) 1171–1176. https://doi.org/10.1038/nbt1336.

[37] M.L. Chiu, D.R. Goulet, A. Teplyakov, G.L. Gilliland, Antibody Structure and Function: The Basis for Engineering Therapeutics, Antibodies. 8 (2019) 55. https://doi.org/10.3390/antib8040055.

[38] W.R. Strohl, Current progress in innovative engineered antibodies, Protein Cell. 9 (2017) 86–120. https://doi.org/10.1007/s13238-017-0457-8.

[39] C. Spiess, Q. Zhai, P.J. Carter, Alternative molecular formats and therapeutic applications for bispecific antibodies, Mol. Immunol. 67 (2015) 95–106. https://doi.org/10.1016/j.molimm.2015.01.003.

[40] J.B.B. Ridgway, L.G. Presta, P. Carter, 'Knobs-into-holes' engineering of antibody CH3 domains for heavy chain heterodimerization, Protein Eng., Des. Sel. 9 (1996) 617–621. https://doi.org/10.1093/protein/9.7.617.

[41] P. Holliger, P.J. Hudson, Engineered antibody fragments and the rise of single domains, Nat. Biotechnol. 23 (2005) 1126–1136. https://doi.org/10.1038/nbt1142.

[42] A.S. Rathore, D. Kumar, N. Kateja, Recent developments in chromatographic purification of biopharmaceuticals, Biotechnol. Lett. 40 (2018) 895–905. https://doi.org/10.1007/s10529-018-2552-1.

[43] G. Guiochon, D.G. Shirazi, A. Felinger, A.M. Katti, Fundamentals of preparative and nonlinear chromatography, 2nd ed., Academic Press, Boston, 2006.

[44] H. Schmidt-Traub, M. Schulte, A. Seidel-Morgenstern, Preparative chromatography, Second, completely revised and updated edition /, Wiley-VCH Verlag GmbH &Co, KGaA, Weinheim, Germany, 2012.

[45] A. Jungbauer, Chromatographic media for bioseparation, J. Chromatogr. A. 1065 (2005) 3–12. https://doi.org/10.1016/j.chroma.2004.08.162.

[46] V. Halan, S. Maity, R. Bhambure, A.S. Rathore, Multimodal Chromatography for Purification of Biotherapeutics – A Review, Curr. Protein Pept. Sci. 20 (2018) 4–13. https://doi.org/10.2174/1389203718666171020103559.

[47] N. Alt, T.Y. Zhang, P. Motchnik, R. Taticek, V. Quarmby, T. Schlothauer, H. Beck, T. Emrich, R.J. Harris, Determination of critical quality attributes for monoclonal antibodies using quality by design principles, Biologicals. 44 (2016) 291–305. https://doi.org/10.1016/j.biologicals.2016.06.005.

[48] C. Hakemeyer, N. McKnight, R.St. John, S. Meier, M. Trexler-Schmidt, B. Kelley, F. Zettl, R. Puskeiler, A. Kleinjans, F. Lim, C. Wurth, Process characterization and Design Space definition, Biologicals. 44 (2016) 306–318. https://doi.org/10.1016/j.biologicals.2016.06.004.

[49] I.H.T. Guideline, International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use considerations (ICH) guideline Q8 (R2) on pharmaceutical development, (2009). https://www.ema.europa.eu/en/ich-q8-r2-pharmaceutical-development-scientific-guideline (accessed January 3, 2024).

[50] L.X. Yu, G. Amidon, M.A. Khan, S.W. Hoag, J. Polli, G.K. Raju, J. Woodcock, Understanding Pharmaceutical Quality by Design, AAPS J. 16 (2014) 771–783. https://doi.org/10.1208/s12248-014-9598-3.

[51] I.F. Pinto, M.R. Aires-Barros, A.M. Azevedo, Multimodal chromatography: debottlenecking the downstream processing of monoclonal antibodies, Pharm. Bioprocess. 3 (2015) 263–279. https://doi.org/10.4155/pbp.15.7.

[52] D. Sýkora, P. Řezanka, K. Záruba, V. Král, Recent advances in mixed-mode chromatographic stationary phases, J. Sep. Sci. 42 (2019) 89–129. https://doi.org/10.1002/jssc.201801048.

[53] L.A. Kennedy, W. Kopaciewicz, F.E. Regnier, Multimodal liquid chromatography columns for the separation of proteins in either the anion-exchange or hydrophobic-interaction mode, J. Chromatogr. A. 359 (1986) 73–84. https://doi.org/10.1016/0021-9673(86)80063-2.

[54] B.K. Nfor, M. Noverraz, S. Chilamkurthi, P.D.E.M. Verhaert, L.A.M. van der Wielen, M. Ottens, High-throughput isotherm determination and thermodynamic modeling of protein adsorption on mixed mode adsorbents, J. Chromatogr. A. 1217 (2010) 6829–6850. https://doi.org/10.1016/j.chroma.2010.07.069.

[55] J. Pezzini, C. Cabanne, R. Gantier, V.N. Janakiraman, X. Santarelli, A comprehensive evaluation of mixed mode interactions of HEA and PPA HyperCel™ chromatographic media, J. Chromatogr. B. 976 (2015) 68–77. https://doi.org/10.1016/j.jchromb.2014.11.020.

[56] K.A. Kaleas, M. Tripodi, S. Revelli, V. Sharma, S.A. Pizarro, Evaluation of a multimodal resin for selective capture of CHO-derived monoclonal antibodies directly from harvested cell culture fluid, J. Chromatogr. B. 969 (2014) 256–263. https://doi.org/10.1016/j.jchromb.2014.08.026.

[57] L.E. Crowell, C. Goodwine, C.S. Holt, L. Rocha, C. Vega, S.A. Rodriguez, N.C. Dalvie, M.K. Tracey, M. Puntel, A. Wigdorovitz, V. Parreño, K.R. Love, S.M. Cramer, J.C. Love, Development of a platform process for the production and purification of single-domain antibodies, Biotechnol. Bioeng. 118 (2021) 3348–3358. https://doi.org/10.1002/bit.27724.

[58] J. Pezzini, G. Joucla, R. Gantier, M. Toueille, A.-M. Lomenech, C.L. Sénéchal, B. Garbay, X. Santarelli, C. Cabanne, Antibody capture by mixed-mode chromatography: A comprehensive study from determination of optimal purification conditions to identification of contaminating host cell proteins, J. Chromatogr. A. 1218 (2011) 8197–8208. https://doi.org/10.1016/j.chroma.2011.09.036.

[59] L.I. Sakhnini, A.K. Pedersen, I.R. León, P.J. Greisen, J.J. Hansen, M.B. Vester-Christensen, L. Bülow, M.B. Dainiak, Optimizing selectivity of anion hydrophobic multimodal chromatography for purification of a single-chain variable fragment, Eng. Life Sci. 19 (2019) 490–501. https://doi.org/10.1002/elsc.201800207.

[60] J. Chen, J. Tetrault, Y. Zhang, A. Wasserman, G. Conley, M. DiLeo, E. Haimes, A.E. Nixon, A. Ley, The distinctive separation attributes of mixed-mode resins and their application in monoclonal antibody downstream purification process, J. Chromatogr. A. 1217 (2010) 216–224. https://doi.org/10.1016/j.chroma.2009.09.047.

[61] A. Voitl, T. Müller-Späth, M. Morbidelli, Application of mixed mode resins for the purification of antibodies, J. Chromatogr. A. 1217 (2010) 5753–5760. https://doi.org/10.1016/j.chroma.2010.06.047.

[62] D. Gao, L.-L. Wang, D.-Q. Lin, S.-J. Yao, Evaluating antibody monomer separation from associated aggregates using mixed-mode chromatography, J. Chromatogr. A. 1294 (2013) 70–75. https://doi.org/10.1016/j.chroma.2013.04.018.

[63] R.A. Chmielowski, S. Meissner, D. Roush, T.O. Linden, E. Glowacki, J. Konietzko, J. Nti-Gyabaah, Resolution of heterogeneous charged antibody aggregates via multimodal chromatography: A comparison to conventional approaches, Biotechnol. Prog. 30 (2014) 636–645. https://doi.org/10.1002/btpr.1908.

[64] E. O'Connor, M. Aspelund, F. Bartnik, M. Berge, K. Coughlin, M. Kambarami, D. Spencer, H. Yan, W. Wang, Monoclonal antibody fragment removal mediated by mixed mode resins, J. Chromatogr. A. 1499 (2017) 65–77. https://doi.org/10.1016/j.chroma.2017.03.063.

[65] W.R. Keller, M. Wendeler, Using multimodal chromatography for post-conjugation antibody-drug conjugate purification: A methodology from high throughput screening to in-silico process development, J. Chromatogr. A. 1653 (2021) 462378. https://doi.org/10.1016/j.chroma.2021.462378.

[66] L.S. Wolfe, C.P. Barringer, S.S. Mostafa, A.A. Shukla, Multimodal chromatography: Characterization of protein binding and selectivity enhancement through mobile phase modulators, J. Chromatogr. A. 1340 (2014) 151–156. https://doi.org/10.1016/j.chroma.2014.02.086.

[67] N. Yoshimoto, D. Itoh, Y. Isakari, A. Podgornik, S. Yamamoto, Salt tolerant chromatography provides salt tolerance and a better selectivity for protein monomer separations, Biotechnol. J. 10 (2015) 1929–1934. https://doi.org/10.1002/biot.201400550.

[68] Cytiva, Multimodal Chromatography, (2021). https://www.cytivalifesciences.com/en/us/support/handbooks (accessed April 27, 2022).

[69] C.L. Bilodeau, E.Y. Lau, D.J. Roush, M.A. Snyder, S.M. Cramer, Behavior of Water Near Multimodal Chromatography Ligands and Its Consequences for Modulating Protein–Ligand Interactions, J. Phys. Chem. B. 125 (2021) 6112–6120. https://doi.org/10.1021/acs.jpcb.1c01549.

[70] S.S. Parasnavis, B. Niu, M. Aspelund, W.K. Chung, M. Snyder, S.M. Cramer, Systematic workflow for studying domain contributions of bispecific antibodies to selectivity in multimodal chromatography, Biotechnol. Bioeng. 119 (2022) 211–225. https://doi.org/10.1002/bit.27967.

[71] J.A. Woo, H. Chen, M.A. Snyder, Y. Chai, R.G. Frost, S.M. Cramer, Defining the property space for chromatographic ligands from a homologous series of mixed-mode ligands, J. Chromatogr. A. 1407 (2015) 58–68. https://doi.org/10.1016/j.chroma.2015.06.017.

[72] S. Koley, S.H. Altern, M. Vats, X. Han, D. Jang, M.A. Snyder, C. Belisle, S.M. Cramer, Evaluation of guanidine-based multimodal anion exchangers for protein selectivity and orthogonality, J. Chromatogr. A. 1653 (2021) 462398. https://doi.org/10.1016/j.chroma.2021.462398.

[73] A. Staby, A.S. Rathore, S. Ahuja, Preparative chromatography for separation of proteins, Wiley, Hoboken, New Jersey, 2017.

[74] T. Gu, G. Tsai, G.T. Tsao, New approach to a general nonlinear multicomponent chromatography model, AIChE J. 36 (1990) 784–788. https://doi.org/10.1002/aic.690360517.

[75] T. Gu, G.T. Tsao, G. Tsai, M.R. Ladisch, Displacement effect in multicomponent chromatography, AIChE J. 36 (1990) 1156–1162. https://doi.org/10.1002/aic.690360805.

[76] S. Yamamoto, K. Nakanishi, R. Matsuno, T. Kamikubo, Ion exchange chromatography of proteins—prediction of elution curves and operating conditions. I. Theoretical considerations, Biotechnol. Bioeng. 25 (1983) 1465–1483. https://doi.org/10.1002/bit.260250605.

[77] C.A. Brooks, S.M. Cramer, Steric mass-action ion exchange: Displacement profiles and induced salt gradients, AIChE J. 38 (1992) 1969–1978. https://doi.org/10.1002/aic.690381212.

[78] J.M. Mollerup, Preparative Chromatography for Separation of Proteins, (2017) 11–79. https://doi.org/10.1002/9781119031116.ch2.

[79] J.M. Mollerup, A Review of the Thermodynamics of Protein Association to Ligands, Protein Adsorption, and Adsorption Isotherms, Chem. Eng. Technol. 31 (2008) 864–874. https://doi.org/10.1002/ceat.200800082.

[80] N.K. Boardman, S.M. Partridge, Separation of neutral proteins on ion-exchange resins, Biochem. J. 59 (1955) 543–552. https://doi.org/10.1042/bj0590543.

[81] W. Kopaciewicz, M.A. Rounds, J. Fausnaugh, F.E. Regnier, Retention model for high-performance ion-exchange chromatography, J. Chromatogr. A. 266 (1983) 3–21. https://doi.org/10.1016/s0021-9673(01)90875-1.

[82] J.M. Mollerup, Applied thermodynamics: A new frontier for biotechnology, Fluid Phase Equilibria. 241 (2006) 205–215. https://doi.org/10.1016/j.fluid.2005.12.037.

[83] T. Hahn, P. Baumann, T. Huuk, V. Heuveline, J. Hubbuch, UV absorption-based inverse modeling of protein chromatography, Eng. Life Sci. 16 (2016) 99–106. https://doi.org/10.1002/elsc.201400247.

[84] Y.F. Lee, H. Graalfs, C. Frech, Thermodynamic modeling of protein retention in mixed-mode chromatography: An extended model for isocratic and dual gradient elution chromatography, J. Chromatogr. A. 1464 (2016) 87–101. https://doi.org/10.1016/j.chroma.2016.08.026.

[85] G. Wang, T. Hahn, J. Hubbuch, Water on hydrophobic surfaces: Mechanistic modeling of hydrophobic interaction chromatography, J. Chromatogr. A. 1465 (2016) 71–78. https://doi.org/10.1016/j.chroma.2016.07.085.

[86] S.H. Altern, J.P. Welsh, J.Y. Lyall, A.J. Kocot, S. Burgess, V. Kumar, C. Williams, A.M. Lenhoff, S.M. Cramer, Isotherm model discrimination for multimodal chromatography using mechanistic models derived from high-throughput batch isotherm data, J. Chromatogr. A. 1693 (2023) 463878. https://doi.org/10.1016/j.chroma.2023.463878.

[87] E.N. Muratov, J. Bajorath, R.P. Sheridan, I.V. Tetko, D. Filimonov, V. Poroikov, T.I. Oprea, I.I. Baskin, A. Varnek, A. Roitberg, O. Isayev, S. Curtalolo, D. Fourches, Y. Cohen, A. Aspuru-Guzik, D.A. Winkler, D. Agrafiotis, A. Cherkasov, A. Tropsha, QSAR without borders, Chem. Soc. Rev. 49 (2020) 3525–3564. https://doi.org/10.1039/d0cs00098a.

[88] F. Grisoni, D. Ballabio, R. Todeschini, V. Consonni, Computational Toxicology, Methods and Protocols, Methods Mol. Biol. 1800 (2018) 3–53. https://doi.org/10.1007/978-1-4939-7899-1_1.

[89] Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment, 1st ed., Elsevier Science, 2015. https://doi.org/10.1016/c2014-0-00286-9.

[90] M. Dehmer, K. Varmuza, D. Bonchev, Statistical Modelling of Molecular Descriptors in QSAR/QSPR, Wiley-Blackwell, Weinheim, 2012. https://doi.org/10.1002/9783527645121.

[91] A. Cherkasov, E.N. Muratov, D. Fourches, A. Varnek, I.I. Baskin, M. Cronin, J. Dearden, P. Gramatica, Y.C. Martin, R. Todeschini, V. Consonni, V.E. Kuz'min, R. Cramer, R. Benigni, C. Yang, J. Rathman, L. Terfloth, J. Gasteiger, A. Richard, A. Tropsha, QSAR Modeling: Where Have You Been? Where Are You Going To?, J. Med. Chem. 57 (2014) 4977–5010. https://doi.org/10.1021/jm4004285.

[92] M.I.J. Raybould, C. Marks, K. Krawczyk, B. Taddese, J. Nowak, A.P. Lewis, A. Bujotzek, J. Shi, C.M. Deane, Five computational developability guidelines for therapeutic antibody profiling, Proc. Natl. Acad. Sci. 116 (2019) 4025–4030. https://doi.org/10.1073/pnas.1810576116.

[93] F. Insaidoo, S. Banerjee, D. Roush, S. Cramer, Preparative Chromatography for Separation of Proteins, (2017) 177–225. https://doi.org/10.1002/9781119031116.ch6.

[94] Y.-C. Lo, S.E. Rensi, W. Torng, R.B. Altman, Machine learning in chemoinformatics and drug discovery, Drug Discov. Today. 23 (2018) 1538–1546. https://doi.org/10.1016/j.drudis.2018.05.010.

[95] P. Polishchuk, Interpretation of Quantitative Structure–Activity Relationship Models: Past, Present, and Future, J. Chem. Inf. Model. 57 (2017) 2618–2639. https://doi.org/10.1021/acs.jcim.7b00274.

[96] A. Tropsha, Best Practices for QSAR Model Development, Validation, and Exploitation, Mol. Inform. 29 (2010) 476–488. https://doi.org/10.1002/minf.201000061.

[97] P. De, S. Kar, P. Ambure, K. Roy, Prediction reliability of QSAR models: an overview of various validation tools, Arch. Toxicol. 96 (2022) 1279–1295. https://doi.org/10.1007/s00204-022-03252-y.

[98] S. Kar, K. Roy, J. Leszczynski, Computational Toxicology, Methods and Protocols, Methods Mol. Biol. 1800 (2018) 141–169. https://doi.org/10.1007/978-1-4939-7899-1_6.

[99] P. Zhou, Q. Liu, T. Wu, Q. Miao, S. Shang, H. Wang, Z. Chen, S. Wang, H. Wang, Systematic Comparison and Comprehensive Evaluation of 80 Amino Acid Descriptors in Peptide QSAR Modeling, J. Chem. Inf. Model. 61 (2021) 1718–1731. https://doi.org/10.1021/acs.jcim.0c01370.

[100] Molecular Modeling of Proteins, 2015. https://doi.org/10.1007/978-1-4939-1465-4.

[101] A.J. Cruz-Cabeza, J. Bernstein, Conformational Polymorphism, Chem. Rev. 114 (2014) 2170–2191. https://doi.org/10.1021/cr400249d.

[102] M.T. Muhammed, E. Aki-Yalcin, Homology modeling in drug discovery: Overview, current applications, and future perspectives, Chem. Biol. Drug Des. 93 (2019) 12–20. https://doi.org/10.1111/cbdd.13388.

[103] J. Pereira, A.J. Simpkin, M.D. Hartmann, D.J. Rigden, R.M. Keegan, A.N. Lupas, High-accuracy protein structure prediction in CASP14, Proteins: Struct., Funct., Bioinform. 89 (2021) 1687–1699. https://doi.org/10.1002/prot.26171.

[104] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S.A.A. Kohl, A.J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A.W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction with AlphaFold, Nature. 596 (2021) 583–589. https://doi.org/10.1038/s41586-021-03819-2.

[105] M. Varadi, S. Anyango, M. Deshpande, S. Nair, C. Natassia, G. Yordanova, D. Yuan, O. Stroe, G. Wood, A. Laydon, A. Žídek, T. Green, K. Tunyasuvunakool, S. Petersen, J. Jumper, E. Clancy, R. Green, A. Vora, M. Lutfi, M. Figurnov, A. Cowie, N. Hobbs, P. Kohli, G. Kleywegt, E. Birney, D. Hassabis, S. Velankar, AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models, Nucleic Acids Res. 50 (2021) D439–D444. https://doi.org/10.1093/nar/gkab1061.

[106] J. Emonts, J.F. Buyel, An overview of descriptors to capture protein properties – Tools and perspectives in the context of QSAR modeling, Comput. Struct. Biotechnol. J. 21 (2023) 3234–3247. https://doi.org/10.1016/j.csbj.2023.05.022.

[107] H. Butt, K. Graf, M. Kappl, Physics and Chemistry of Interfaces, 2003. https://doi.org/10.1002/3527602313.

[108] P. Grochowski, J. Trylska, Continuum molecular electrostatics, salt effects, and counterion binding—A review of the Poisson–Boltzmann theory and its modifications, Biopolymers. 89 (2008) 93–113. https://doi.org/10.1002/bip.20877.

[109] E. Jurrus, D. Engel, K. Star, K. Monson, J. Brandi, L.E. Felberg, D.H. Brookes, L. Wilson, J. Chen, K. Liles, M. Chun, P. Li, D.W. Gohara, T. Dolinsky, R. Konecny, D.R. Koes, J.E. Nielsen, T. Head-Gordon, W. Geng, R. Krasny, G. Wei, M.J. Holst, J.A. McCammon, N.A. Baker, Improvements to the APBS biomolecular solvation software suite, Protein Sci. 27 (2018) 112–128. https://doi.org/10.1002/pro.3280.

[110] K. Sankar, K. Trainor, L.L. Blazer, J.J. Adams, S.S. Sidhu, T. Day, E. Meiering, J.K.X. Maier, A Descriptor Set for Quantitative Structure-property Relationship Prediction in Biologics, Mol. Inform. 41 (2022) e2100240. https://doi.org/10.1002/minf.202100240.

[111] T. Hastie, The elements of statistical learning : data mining, inference, and prediction, 2nd ed., New York, N.Y. Springer, 2009.

[112] A. Varnek, I. Baskin, Machine Learning Methods for Property Prediction in Chemoinformatics: Quo Vadis?, J. Chem. Inf. Model. 52 (2012) 1413–1437. https://doi.org/10.1021/ci200409x.

[113] I.H. Sarker, Machine Learning: Algorithms, Real-World Applications and Research Directions, SN Comput. Sci. 2 (2021) 160. https://doi.org/10.1007/s42979-021-00592-x.

[114] J.B.O. Mitchell, Machine learning methods in chemoinformatics, Wiley Interdiscip. Rev.: Comput. Mol. Sci. 4 (2014) 468–481. https://doi.org/10.1002/wcms.1183.

[115] J. Wang, An Intuitive Tutorial to Gaussian Process Regression, Comput. Sci. Amp Eng. 25 (2023) 4–11. https://doi.org/10.1109/mcse.2023.3342149.

[116] C.E. Rasmussen, C.K.I. Williams, Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning), The MIT Press, 2005.

[117] U. Gottschalk, K. Brorson, A.A. Shukla, The need for innovation in biomanufacturing, Nat. Biotechnol. 30 (2012) 489–492. https://doi.org/10.1038/nbt.2263.

[118] N. Singh, S. Herzer, New Bioprocessing Strategies: Development and Manufacturing of Recombinant Antibodies and Proteins, Adv Biochem Eng Biotechnology. (2017) 115–178. https://doi.org/10.1007/10_2017_12.

[119] D. Gao, D.-Q. Lin, S.-J. Yao, Mechanistic analysis on the effects of salt concentration and pH on protein adsorption onto a mixed-mode adsorbent with cation ligand, J. Chromatogr. B. 859 (2007) 16–23. https://doi.org/10.1016/j.jchromb.2007.08.044.

[120] D.H. Cebulla, C. Kirches, A. Potschka, Parameter Identifiability in a Novel Kinetic Adsorption Isotherm for Multi-Modal Chromatography, 2019 IEEE 58th Conf. Decis. Control (CDC). 00 (2019) 4755–4760. https://doi.org/10.1109/cdc40024.2019.9029813.

[121] M. Zhu, G. Carta, Protein adsorption equilibrium and kinetics in multimodal cation exchange resins, Adsorption. 22 (2016) 165–179. https://doi.org/10.1007/s10450-015-9735-z.

[122] P.F. Gomes, J.M. Loureiro, A.E. Rodrigues, Adsorption of Human Serum Albumin (HSA) on a mixed-mode adsorbent: equilibrium and kinetics, Adsorption. 23 (2017) 491–505. https://doi.org/10.1007/s10450-017-9861-x.

[123] S. Kluters, M. Hafner, T. von Hirschheydt, C. Frech, Solvent modulated linear pH gradient elution for the purification of conventional and bispecific antibodies: Modeling and application, J. Chromatogr. A. 1418 (2015) 119–129. https://doi.org/10.1016/j.chroma.2015.09.053.

[124] H.S. Karkov, L. Sejergaard, S.M. Cramer, Methods development in multimodal chromatography with mobile phase modifiers using the steric mass action model, J. Chromatogr. A. 1318 (2013) 149–155. https://doi.org/10.1016/j.chroma.2013.10.004.

[125] L. Sejergaard, H.S. Karkov, J.K. Krarup, A.B.B. Hagel, S.M. Cramer, Model-based process development for the purification of a modified human growth hormone using multimodal chromatography, Biotechnol. Prog. 30 (2014) 1057–1064. https://doi.org/10.1002/btpr.1923.

[126] Y.F. Lee, M. Schmidt, H. Graalfs, M. Hafner, C. Frech, Modeling of dual gradient elution in ion exchange and mixed-mode chromatography, J. Chromatogr. A. 1417 (2015) 64–72. https://doi.org/10.1016/j.chroma.2015.09.032.

[127] L. Zhang, S. Parasnavis, Z. Li, J. Chen, S. Cramer, Mechanistic modeling based process development for monoclonal antibody monomer-aggregate separations in multimodal cation exchange chromatography, J. Chromatogr. A. 1602 (2019) 317–325. https://doi.org/10.1016/j.chroma.2019.05.056.

[128] T. Hahn, A. Sommer, A. Osberghaus, V. Heuveline, J. Hubbuch, Adjoint-based estimation and optimization for column liquid chromatography models, Comput. Chem. Eng. 64 (2014) 41–54. https://doi.org/10.1016/j.compchemeng.2014.01.013.

[129] B.C.S. To, A.M. Lenhoff, Hydrophobic interaction chromatography of proteins II. Solution thermodynamic properties as a determinant of retention, J. Chromatogr. A. 1141 (2007) 235–243. https://doi.org/10.1016/j.chroma.2006.12.022.

[130] R.W. Deitcher, J.E. Rome, P.A. Gildea, J.P. O'Connell, E.J. Fernandez, A new thermodynamic model describes the effects of ligand density and type, salt concentration and protein species in hydrophobic interaction chromatography, J. Chromatogr. A. 1217 (2010) 199–208. https://doi.org/10.1016/j.chroma.2009.07.068.

[131] S. Andris, J. Hubbuch, Modeling of hydrophobic interaction chromatography for the separation of antibody-drug conjugates and its application towards quality by design, J. Biotechnol. 317 (2020) 48–58. https://doi.org/10.1016/j.jbiotec.2020.04.018.

[132] J. Mollerup, T. Hansen, S. Frederiksen, A. Staby, Advances in Chromatography Volume 48, Adv Chromatogr. 48 (2009) 57–97. https://doi.org/10.1201/9781420084542-c2.

[133] S. Yamamoto, Electrostatic Interaction Chromatography Process for Protein Separations: Impact of Engineering Analysis of Biorecognition Mechanism on Process Optimization, Chem Eng Technol. 28 (2005) 1387–1393. https://doi.org/10.1002/ceat.200500199.

[134] M. Rüdt, F. Gillet, S. Heege, J. Hitzler, B. Kalbfuss, B. Guélat, Combined Yamamoto approach for simultaneous estimation of adsorption isotherm and kinetic parameters in ion-exchange chromatography, J. Chromatogr. A. 1413 (2015) 68–76. https://doi.org/10.1016/j.chroma.2015.08.025.

[135] D. Saleh, G. Wang, B. Müller, F. Rischawy, S. Kluters, J. Studts, J. Hubbuch, Straightforward method for calibration of mechanistic cation exchange chromatography models for industrial applications, Biotechnol. Prog. 36 (2020) e2984. https://doi.org/10.1002/btpr.2984.

[136] Y. Xiao, A. Rathore, J.P. O'Connell, E.J. Fernandez, Generalizing a two-conformation model for describing salt and temperature effects on protein retention and stability in hydrophobic interaction chromatography, J. Chromatogr. A. 1157 (2007) 197–206. https://doi.org/10.1016/j.chroma.2007.05.009.

[137] F. Kröner, J. Hubbuch, Systematic generation of buffer systems for pH gradient ion exchange chromatography and their application, J. Chromatogr. A. 1285 (2013) 78–87. https://doi.org/10.1016/j.chroma.2013.02.017.

[138] S.K. Li, M.R. Liddell, H. Wen, Effective electrophoretic mobilities and charges of anti-VEGF proteins determined by capillary zone electrophoresis, J. Pharm. Biomed. Anal. 55 (2011) 603–607. https://doi.org/10.1016/j.jpba.2010.12.027.

[139] L.K. Kimerer, T.M. Pabst, A.K. Hunter, G. Carta, Chromatographic behavior of bivalent bispecific antibodies on cation exchange columns. I. Experimental observations and phenomenological model, J. Chromatogr. A. 1601 (2019) 121–132. https://doi.org/10.1016/j.chroma.2019.04.012.

[140] J.M. Reck, T.M. Pabst, A.K. Hunter, X. Wang, G. Carta, Adsorption equilibrium and kinetics of monomer–dimer monoclonal antibody mixtures on a cation exchange resin, J Chromatogr A. 1402 (2015) 46–59. https://doi.org/10.1016/j.chroma.2015.05.007.

[141] T. Hahn, T. Huuk, V. Heuveline, J. Hubbuch, Simulating and Optimizing Preparative Protein Chromatography with ChromX, J Chem Educ. 92 (2015) 1497–1502. https://doi.org/10.1021/ed500854a.

[142] S. Hunt, T. Larsen, R.J. Todd, Preparative Chromatography for Separation of Proteins, in: 2017: pp. 399–427. https://doi.org/10.1002/9781119031116.ch13.

[143] T.C. Huuk, T. Briskot, T. Hahn, J. Hubbuch, A versatile noninvasive method for adsorber quantification in batch and column chromatography based on the ionic capacity, Biotechnol. Prog. 32 (2016) 666–677. https://doi.org/10.1002/btpr.2228.

[144] S.F. Chung, C.Y. Wen, Longitudinal dispersion of liquid flowing through fixed and fluidized beds, AIChE J. 14 (1968) 857–866. https://doi.org/10.1002/aic.690140608.

[145] F. Rischawy, D. Saleh, T. Hahn, S. Oelmeier, J. Spitz, S. Kluters, Good modeling practice for industrial chromatography: Mechanistic modeling of ion exchange chromatography of a bispecific antibody, Comput. Chem. Eng. 130 (2019) 106532. https://doi.org/10.1016/j.compchemeng.2019.106532.

[146] A. Creasy, G. Barker, Y. Yao, G. Carta, Systematic interpolation method predicts protein chromatographic elution from batch isotherm data without a detailed mechanistic isotherm model, Biotechnol J. 10 (2015) 1400–1411. https://doi.org/10.1002/biot.201500089.

[147] J.J. Moré, Numerical Analysis, Proceedings of the Biennial Conference Held at Dundee, June 28–July 1, 1977, Lect. Notes Math. (2006) 105–116. https://doi.org/10.1007/bfb0067700.

[148] P. Virtanen, R. Gommers, T.E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S.J. van der Walt, M. Brett, J. Wilson, K.J. Millman, N. Mayorov, A.R.J. Nelson, E. Jones, R. Kern, E. Larson, C.J. Carey, İ. Polat, Y. Feng, E.W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E.A. Quintero, C.R. Harris, A.M. Archibald, A.H. Ribeiro, F. Pedregosa, P. van Mulbregt, S. 1 0 Contributors, A. Vijaykumar, A.P. Bardelli, A. Rothberg, A. Hilboll, A. Kloeckner, A. Scopatz, A. Lee, A. Rokem, C.N. Woods, C. Fulton, C. Masson, C. Häggström, C. Fitzgerald, D.A. Nicholson, D.R. Hagen, D.V. Pasechnik, E. Olivetti, E.

Martin, E. Wieser, F. Silva, F. Lenders, F. Wilhelm, G. Young, G.A. Price, G.-L. Ingold, G.E. Allen, G.R. Lee, H. Audren, I. Probst, J.P. Dietrich, J. Silterra, J.T. Webber, J. Slavič, J. Nothman, J. Buchner, J. Kulick, J.L. Schönberger, J.V. de M. Cardoso, J. Reimer, J. Harrington, J.L.C. Rodríguez, J. Nunez-Iglesias, J. Kuczynski, K. Tritz, M. Thoma, M. Newville, M. Kümmerer, M. Bolingbroke, M. Tartre, M. Pak, N.J. Smith, N. Nowaczyk, N. Shebanov, O. Pavlyk, P.A. Brodtkorb, P. Lee, R.T. McGibbon, R. Feldbauer, S. Lewis, S. Tygier, S. Sievert, S. Vigna, S. Peterson, S. More, T. Pudlik, T. Oshima, T.J. Pingel, T.P. Robitaille, T. Spura, T.R. Jones, T. Cera, T. Leslie, T. Zito, T. Krauss, U. Upadhyay, Y.O. Halchenko, Y. Vázquez-Baeza, SciPy 1.0: fundamental algorithms for scientific computing in Python, Nat. Methods. 17 (2020) 261–272. https://doi.org/10.1038/s41592-019-0686-2.

[149] J.C. Schöneberger, H. Arellano-Garcia, G. Wozny, S. Körkel, H. Thielert, Model-Based Experimental Analysis of a Fixed-Bed Reactor for Catalytic SO2 Oxidation, Ind. Eng. Chem. Res. 48 (2009) 5165–5176. https://doi.org/10.1021/ie801288d.

[150] D. Pfister, F. Steinebach, M. Morbidelli, Linear isotherm determination from linear gradient elution experiments, J. Chromatogr. A. 1375 (2015) 33–41. https://doi.org/10.1016/j.chroma.2014.11.067.

[151] F.-G. Wieland, A.L. Hauber, M. Rosenblatt, C. Tönsing, J. Timmer, On structural and practical identifiability, Curr. Opin. Syst. Biol. 25 (2021) 60–69. https://doi.org/10.1016/j.coisb.2021.03.005.

[152] A. Osberghaus, S. Hepbildikler, S. Nath, M. Haindl, E. von Lieres, J. Hubbuch, Determination of parameters for the steric mass action model—A comparison between two approaches, J. Chromatogr. A. 1233 (2012) 54–65. https://doi.org/10.1016/j.chroma.2012.02.004.

[153] T. Briskot, F. Stückler, F. Wittkopp, C. Williams, J. Yang, S. Konrad, K. Doninger, J. Griesbach, M. Bennecke, S. Hepbildikler, J. Hubbuch, Prediction uncertainty assessment of chromatography models using Bayesian inference, J. Chromatogr. A. 1587 (2019) 101–110. https://doi.org/10.1016/j.chroma.2018.11.076.

[154] J.A. Roberts, L. Kimerer, G. Carta, Effects of molecule size and resin structure on protein adsorption on multimodal anion exchange chromatography media, J. Chromatogr. A. 1628 (2020) 461444. https://doi.org/10.1016/j.chroma.2020.461444.

[155] B. Beyer, A. Jungbauer, Conformational changes of antibodies upon adsorption onto hydrophobic interaction chromatography surfaces, J. Chromatogr. A. 1552 (2018) 60–66. https://doi.org/10.1016/j.chroma.2018.04.009.

[156] M.A. Holstein, A.A.M. Nikfetrat, M. Gage, A.G. Hirsh, S.M. Cramer, Improving selectivity in multimodal chromatography using controlled pH gradient elution, J. Chromatogr. A. 1233 (2012) 152–155. https://doi.org/10.1016/j.chroma.2012.01.074.

[157] K. Zhang, X. Liu, Mixed-mode chromatography in pharmaceutical and biopharmaceutical applications, J. Pharm. Biomed. Anal. 128 (2016) 73–88. https://doi.org/10.1016/j.jpba.2016.05.007.

[158] R.B. Gudhka, D.J. Roush, S.M. Cramer, A thermodynamic evaluation of antibody-surface interactions in multimodal cation exchange chromatography, J. Chromatogr. A. 1628 (2020) 461479. https://doi.org/10.1016/j.chroma.2020.461479.

[159] W.K. Chung, Y. Hou, M. Holstein, A. Freed, G.I. Makhatadze, S.M. Cramer, Investigation of protein binding affinity in multimodal chromatographic systems using a homologous protein library, J. Chromatogr. A. 1217 (2010) 191–198. https://doi.org/10.1016/j.chroma.2009.08.005.

[160] H.S. Karkov, B.O. Krogh, J. Woo, S. Parimal, H. Ahmadian, S.M. Cramer, Investigation of protein selectivity in multimodal chromatography using in silico designed Fab fragment variants, Biotechnol. Bioeng. 112 (2015) 2305–2315. https://doi.org/10.1002/bit.25642.

[161] J. Robinson, D. Roush, S. Cramer, Domain contributions to antibody retention in multimodal chromatography systems, J. Chromatogr. A. 1563 (2018) 89–98. https://doi.org/10.1016/j.chroma.2018.05.058.

[162] K. Srinivasan, S. Banerjee, S. Parimal, L. Sejergaard, R. Berkovich, B. Barquera, S. Garde, S.M. Cramer, Single Molecule Force Spectroscopy and Molecular Dynamics Simulations as a Combined Platform for Probing Protein Face-Specific Binding, Langmuir. 33 (2017) 10851–10860. https://doi.org/10.1021/acs.langmuir.7b03011.

[163] R.B. Gudhka, C.L. Bilodeau, S.A. McCallum, M.A. McCoy, D.J. Roush, M.A. Snyder, S.M. Cramer, Identification of preferred multimodal ligand-binding regions on IgG1 FC using nuclear magnetic resonance and molecular dynamics simulations, Biotechnol. Bioeng. 118 (2021) 809–822. https://doi.org/10.1002/bit.27611.

[164] K. Dhingra, R.B. Gudhka, S.M. Cramer, Evaluation of preferred binding regions on ubiquitin and IgG1 FC for interacting with multimodal cation exchange resins using DEPC labeling/mass spectrometry, Biotechnol. Bioeng. 120 (2023) 1592–1604. https://doi.org/10.1002/bit.28361.

[165] H.-F. Tong, C. Cavallotti, S.-J. Yao, D.-Q. Lin, Molecular insight into protein binding orientations and interaction modes on hydrophobic charge-induction resin, J. Chromatogr. A. 1512 (2017) 34–42. https://doi.org/10.1016/j.chroma.2017.06.071.

[166] J. Robinson, D. Roush, S.M. Cramer, The effect of pH on antibody retention in multimodal cation exchange chromatographic systems, J. Chromatogr. A. 1617 (2020) 460838. https://doi.org/10.1016/j.chroma.2019.460838.

[167] R.B. Gudhka, M. Vats, C.L. Bilodeau, S.A. McCallum, M.A. McCoy, D.J. Roush, M.A. Snyder, S.M. Cramer, Probing IgG1 FC–Multimodal Nanoparticle Interactions: A Combined Nuclear Magnetic Resonance and Molecular Dynamics Simulations Approach, Langmuir. 37 (2021) 12188–12203. https://doi.org/10.1021/acs.langmuir.1c02114.

[168] Y. Hou, S.M. Cramer, Evaluation of selectivity in multimodal anion exchange systems: A priori prediction of protein retention and examination of mobile phase modifier effects, J. Chromatogr. A. 1218 (2011) 7813–7820. https://doi.org/10.1016/j.chroma.2011.08.080.

[169] J.F. Buyel, J.A. Woo, S.M. Cramer, R. Fischer, The use of quantitative structure–activity relationship models to develop optimized processes for the removal of tobacco host cell proteins during biopharmaceutical production, J. Chromatogr. A. 1322 (2013) 18–28. https://doi.org/10.1016/j.chroma.2013.10.076.

[170] J. Woo, S. Parimal, M.R. Brown, R. Heden, S.M. Cramer, The effect of geometrical presentation of multimodal cation-exchange ligands on selective recognition of hydrophobic regions on protein surfaces, J. Chromatogr. A. 1412 (2015) 33–42. https://doi.org/10.1016/j.chroma.2015.07.072.

[171] J.R. Robinson, H.S. Karkov, J.A. Woo, B.O. Krogh, S.M. Cramer, QSAR models for prediction of chromatographic behavior of homologous Fab variants, Biotechnol. Bioeng. 114 (2017) 1231–1240. https://doi.org/10.1002/bit.26236.

[172] J. Robinson, M.A. Snyder, C. Belisle, J. Liao, H. Chen, X. He, Y. Xu, S.M. Cramer, Investigating the impact of aromatic ring substitutions on selectivity for a multimodal anion exchange prototype library, J. Chromatogr. A. 1569 (2018) 101–109. https://doi.org/10.1016/j.chroma.2018.07.049.

[173] S.M. Timmick, N. Vecchiarello, C. Goodwine, L.E. Crowell, K.R. Love, J.C. Love, S.M. Cramer, An impurity characterization based approach for the rapid development of integrated downstream purification processes, Biotechnol. Bioeng. 115 (2018) 2048–2060. https://doi.org/10.1002/bit.26718.

[174] M. Bailly, C. Mieczkowski, V. Juan, E. Metwally, D. Tomazela, J. Baker, M. Uchida, E. Kofman, F. Raoufi, S. Motlagh, Y. Yu, J. Park, S. Raghava, J. Welsh, M. Rauscher, G. Raghunathan, M. Hsieh, Y.-L. Chen, H.T. Nguyen, N. Nguyen, D. Cipriano, L. Fayadat-Dilman, Predicting Antibody Developability Profiles Through Early Stage Discovery Screening, MAbs. 12 (2020) 1743053. https://doi.org/10.1080/19420862.2020.1743053.

[175] J.T. Heads, S. Kelm, K. Tyson, A.D.G. Lawson, A computational method for predicting the aggregation propensity of IgG1 and IgG4(P) mAbs in common storage buffers, Mabs. 14 (2022) 2138092. https://doi.org/10.1080/19420862.2022.2138092.

[176] E.O. Saphire, P.W.H.I. Parren, R. Pantophlet, M.B. Zwick, G.M. Morris, P.M. Rudd, R.A. Dwek, R.L. Stanfield, D.R. Burton, I.A. Wilson, Crystal Structure of a Neutralizing Human IgG Against HIV-1: A Template for Vaccine Design, Science. 293 (2001) 1155–1159. https://doi.org/10.1126/science.1061692.

[177] R. Hess, D. Yun, D. Saleh, T. Briskot, J.-H. Grosch, G. Wang, T. Schwab, J. Hubbuch, Standardized method for mechanistic modeling of multimodal anion exchange chromatography in flow through operation, J. Chromatogr. A. 1690 (2023) 463789. https://doi.org/10.1016/j.chroma.2023.463789.

[178] R. Wälchli, M. Ressurreição, S. Vogg, F. Feidl, J. Angelo, X. Xu, S. Ghose, Z.J. Li, X.L. Saoût, J. Souquet, H. Broly, M. Morbidelli, Understanding mAb aggregation during low pH viral inactivation and subsequent neutralization, Biotechnol. Bioeng. 117 (2020) 687–700. https://doi.org/10.1002/bit.27237.

[179] D. Saleh, R. Hess, M. Ahlers-Hesse, F. Rischawy, G. Wang, J. Grosch, T. Schwab, S. Kluters, J. Studts, J. Hubbuch, A multiscale modeling method for therapeutic antibodies in ion exchange chromatography, Biotechnol. Bioeng. 120 (2023) 125–138. https://doi.org/10.1002/bit.28258.

[180] K. Zhu, T. Day, D. Warshaviak, C. Murrett, R. Friesner, D. Pearlman, Antibody structure determination using a combination of homology modeling, energy-based refinement, and loop prediction, Proteins: Struct., Funct., Bioinform. 82 (2014) 1646–1655. https://doi.org/10.1002/prot.24551.

[181] K.R. Abhinandan, A.C.R. Martin, Analysis and improvements to Kabat and structurally correct numbering of antibody variable domains, Mol. Immunol. 45 (2008) 3832–3839. https://doi.org/10.1016/j.molimm.2008.05.022.

[182] G. Scapin, X. Yang, W.W. Prosise, M. McCoy, P. Reichert, J.M. Johnston, R.S. Kashi, C. Strickland, Structure of full-length human anti-PD1 therapeutic IgG4 antibody pembrolizumab, Nat. Struct. Mol. Biol. 22 (2015) 953–958. https://doi.org/10.1038/nsmb.3129.

[183] K. Zhu, T. Day, Ab initio structure prediction of the antibody hypervariable H3 loop, Proteins: Struct., Funct., Bioinform. 81 (2013) 1081–1089. https://doi.org/10.1002/prot.24240.

[184] G.M. Sastry, M. Adzhigirey, T. Day, R. Annabhimoju, W. Sherman, Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments, J. Comput.-Aided Mol. Des. 27 (2013) 221–234. https://doi.org/10.1007/s10822-013-9644-8.

[185] M.H.M. Olsson, C.R. Søndergaard, M. Rostkowski, J.H. Jensen, PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pK a Predictions, J. Chem. Theory Comput. 7 (2011) 525–537. https://doi.org/10.1021/ct100578z.

[186] C. Lu, C. Wu, D. Ghoreishi, W. Chen, L. Wang, W. Damm, G.A. Ross, M.K. Dahlgren, E. Russell, C.D.V. Bargen, R. Abel, R.A. Friesner, E.D. Harder, OPLS4: Improving Force Field Accuracy on Challenging Regimes of Chemical Space, J. Chem. Theory Comput. 17 (2021) 4291–4300. https://doi.org/10.1021/acs.jctc.1c00302.

[187] S.A. Wildman, G.M. Crippen, Prediction of Physicochemical Parameters by Atomic Contributions, J. Chem. Inf. Comput. Sci. 39 (1999) 868–873. https://doi.org/10.1021/ci9903071.

[188] K. Sankar, S.R. Krystek, S.M. Carl, T. Day, J.K.X. Maier, AggScore: Prediction of aggregation-prone regions in proteins based on the distribution of surface patches, Proteins: Struct., Funct., Bioinform. 86 (2018) 1147–1156. https://doi.org/10.1002/prot.25594.

[189] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, G. Louppe, P. Prettenhofer, R. Weiss, R.J. Weiss, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine Learning in Python, J. Mach. Learn. Res. 12 (2011) 2825–2830.

[190] O. Obrezanova, G. Csányi, J.M.R. Gola, M.D. Segall, Gaussian Processes: A Method for Automatic QSAR Modeling of ADME Properties, J. Chem. Inf. Model. 47 (2007) 1847–1857. https://doi.org/10.1021/ci7000633.

[191] P. Zhou, F. Tian, X. Chen, Z. Shang, Modeling and prediction of binding affinities between the human amphiphysin SH3 domain and its peptide ligands using genetic algorithm-Gaussian processes, Pept. Sci. 90 (2008) 792–802. https://doi.org/10.1002/bip.21091.

[192] C. Zhu, R.H. Byrd, P. Lu, J. Nocedal, Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization, ACM Trans. Math. Softw. (TOMS). 23 (1997) 550–560. https://doi.org/10.1145/279232.279236.

[193] V. Kumar, Feature Selection: A literature Review, Smart Comput. Rev. 4 (2014) 211–229. https://doi.org/10.6029/smartcr.2014.03.007.

[194] L. Breiman, Random Forests, Mach. Learn. 45 (2001) 5–32. https://doi.org/10.1023/a:1010933404324.

[195] M. Ojala, G.C. Garriga, Permutation Tests for Studying Classifier Performance, 2009 Ninth IEEE Int. Conf. Data Min. (2009) 908–913. https://doi.org/10.1109/icdm.2009.108.

[196] A. Goldstein, A. Kapelner, J. Bleich, E. Pitkin, Peeking Inside the Black Box: Visualizing Statistical Learning With Plots of Individual Conditional Expectation, J Comput Graph Stat. 24 (2015) 44–65. https://doi.org/10.1080/10618600.2014.907095.

[197] J.T. Heads, R. Lamb, S. Kelm, R. Adams, P. Elliott, K. Tyson, S. Topia, S. West, R. Nan, A. Turner, A.D.G. Lawson, Electrostatic interactions modulate the differential aggregation propensities of IgG1 and IgG4P antibodies and inform charged residue substitutions for improved developability, Protein Eng Des Sel. 32 (2019) 277–288. https://doi.org/10.1093/protein/gzz046.

[198] S. Ghose, B. Hubbard, S.M. Cramer, Protein Interactions in Hydrophobic Charge Induction Chromatography (HCIC), Biotechnol Progr. 21 (2005) 498–508. https://doi.org/10.1021/bp049712+.

[199] P. Baumann, K. Baumgartner, J. Hubbuch, Influence of binding pH and protein solubility on the dynamic binding capacity in hydrophobic interaction chromatography, J. Chromatogr. A. 1396 (2015) 77–85. https://doi.org/10.1016/j.chroma.2015.04.001.

[200] L.K. Kimerer, T.M. Pabst, A.K. Hunter, G. Carta, Role of configurational flexibility on the adsorption kinetics of bivalent bispecific antibodies on porous cation exchange resins, J. Chromatogr. A. 1655 (2021) 462479. https://doi.org/10.1016/j.chroma.2021.462479.

[201] C. Molnar, Interpretable Machine Learning. A Guide for Making Black Box Models Explainable., 2019. https://christophm.github.io/interpretable-ml-book.

[202] M. Karlberg, M. von Stosch, J. Glassey, Exploiting mAb structure characteristics for a directed QbD implementation in early process development, Crit. Rev. Biotechnol. 38 (2018) 957–970. https://doi.org/10.1080/07388551.2017.1421899.

[203] V. Kumar, A.M. Lenhoff, Mechanistic Modeling of Preparative Column Chromatography for Biotherapeutics, Annu. Rev. Chem. Biomol. Eng. 11 (2020) 235–255. https://doi.org/10.1146/annurev-chembioeng-102419-125430.

[204] J. Kittelmann, K.M.H. Lang, M. Ottens, J. Hubbuch, Orientation of monoclonal antibodies in ion-exchange chromatography: A predictive quantitative structure–activity relationship modeling approach, J. Chromatogr. A. 1510 (2017) 33–39. https://doi.org/10.1016/j.chroma.2017.06.047.

[205] T. Hahn, N. Geng, K. Petrushevska-Seebach, M.E. Dolan, M. Scheindel, P. Graf, K. Takenaka, K. Izumida, L. Li, Z. Ma, N. Schuelke, Mechanistic modeling, simulation, and optimization of mixed-mode chromatography for an antibody polishing step, Biotechnol. Prog. 39 (2023) e3316. https://doi.org/10.1002/btpr.3316.

[206] R. Hess, J. Faessler, D. Yun, D. Saleh, J.-H. Grosch, T. Schwab, J. Hubbuch, Antibody sequence-based prediction of pH gradient elution in multimodal chromatography, J. Chromatogr. A. 1711 (2023) 464437. https://doi.org/10.1016/j.chroma.2023.464437.

[207] A. Ladiwala, K. Rege, C.M. Breneman, S.M. Cramer, A priori prediction of adsorption isotherm parameters and chromatographic behavior in ion-exchange systems, Proc. Natl. Acad. Sci. 102 (2005) 11710–11715. https://doi.org/10.1073/pnas.0408769102.

[208] T. Yang, M.C. Sundling, A.S. Freed, C.M. Breneman, S.M. Cramer, Prediction of pH-Dependent Chromatographic Behavior in Ion-Exchange Systems, Anal. Chem. 79 (2007) 8927–8939. https://doi.org/10.1021/ac071101j.

[209] M.F. Sanner, A.J. Olson, J. Spehner, Reduced surface: An efficient way to compute molecular surfaces, Biopolymers. 38 (1996) 305–320. https://doi.org/10.1002/(sici)1097-0282(199603)38:3<;305::aid-bip4>3.0.co;2-y.

[210] H. Schweke, M.-H. Mucchielli, N. Chevrollier, S. Gosset, A. Lopes, SURFMAP: A Software for Mapping in Two Dimensions Protein Surface Features, J. Chem. Inf. Model. 62 (2022) 1595–1601. https://doi.org/10.1021/acs.jcim.1c01269.

[211] W.C. Wimley, S.H. White, Experimentally determined hydrophobicity scale for proteins at membrane interfaces, Nat. Struct. Biol. 3 (1996) 842–848. https://doi.org/10.1038/nsb1096-842.

[212] M. Mezei, A new method for mapping macromolecular topography, J. Mol. Graph. Model. 21 (2003) 463–472. https://doi.org/10.1016/s1093-3263(02)00203-6.

[213] S. Hunt, T. Larsen, R.J. Todd, Preparative Chromatography for Separation of Proteins, (2017) 399–427. https://doi.org/10.1002/9781119031116.ch13.

[214] A. Velayudhan, C. Horváth, Preparative chromatography of proteins Analysis of the multivalent ion-exchange formalism, J. Chromatogr. A. 443 (1988) 13–29. https://doi.org/10.1016/s0021-9673(00)94779-4.

[215] M. Schmidt, M. Hafner, C. Frech, Modeling of salt and pH gradient elution in ion-exchange chromatography, J. Sep. Sci. 37 (2014) 5–13. https://doi.org/10.1002/jssc.201301007.

[216] F. Waibl, M.L. Fernández-Quintero, F.S. Wedl, H. Kettenberger, G. Georges, K.R. Liedl, Comparison of hydrophobicity scales for predicting biophysical properties of antibodies, Front. Mol. Biosci. 9 (2022) 960194. https://doi.org/10.3389/fmolb.2022.960194.

[217] S. Banerjee, S. Parimal, S.M. Cramer, A molecular modeling based method to predict elution behavior and binding patches of proteins in multimodal chromatography, J. Chromatogr. A. 1511 (2017) 45–58. https://doi.org/10.1016/j.chroma.2017.06.059.

List of Figures

List of Tables

Abbreviations

AC	Affinity chromatography
AD	Applicability domain
AEX	Anion-exchange chromatography
APBS	Adaptive Poisson-Boltzmann Solver
CASP	Critical Assessment of Techniques for Protein Structure Prediction
CDR	Complementarity-determining region
CEX	Cation-exchange chromatography
CH1-3	Heavy chain constant domain 1-3
CI	Confidence interval
CIP	Cleaning-in-place
CL	Light chain constant domain
CMC	Chemistry, Manufacturing, and Controls
CPP	Critical process parameter
CQA	Critical quality attribute
CR	Constant region
cryoEM	Cryogenic electron microscopy
CSTR	Continuous stirred-tank reactor
CV	Column volume
DNA	Deoxyribonucleic acid
DoE	Design of experiments
DSP	Downstream processing
EMA	European Medicines Agency

Fab	Fragment antigen-binding
Fc	Fragment crystallizable
FDA	U.S. Food and Drug Administration
FR	Framework region
FRH	Framework region heavy chain
FRL	Framework region light chain
FT	Flow-through
GPR	Gaussian process regression
GRM	General rate model
H1-3	Heavy chain loop 1-3
HC	Heavy chain
HCIC	Hydrophobic charge induction chromatography
HCP	Host cell protein
HFR1-4	Heavy chain framework region 1-4
HIC	Hydrophobic interaction chromatography
HT	Hydroxyapatite
HTS	High-throughput screening
ICE	Individual conditional expectation
IE	Isocratic elution
IEX	Ion-exchange chromatography
Ig	Immunoglobulin
KiH	Knob-in-Hole
L1-3	Light chain loop 1-3
LC	Light chain
LDF	Linear driving force
LFR1-4	Light chain framework region 1-4
LGE	Linear gradient elution

LML	Log(marginal likelihood)
mAb	Monoclonal antibody
MAE	Mean absolute error
MD	Molecular dynamics
MMA	Multimodal anion-exchange
MMC	Multimodal / Mixed-mode chromatography
NMR	Nuclear magnetic resonance
PD	Partial dependence
PDB	Protein Data Bank
pI	Isoelectric point
PI	Permutation importance
QbD	Quality by design
QM	Quantum dynamics
QSAR	Quantitative structure-activity relationship
QSPR	Quantitative structure-property relationship
RBF	Radial basis function
RFE	Recursive feature elimination
RP-LC	Reversed-phase liquid chromatography
SAS	Solvent accessible surface
SASA	Solvent-accessible surface area
scFv	Single-chain fragment variables
SD	Standard deviation
SE	Step elution
SEC	Size exclusion chromatography
SES	Solvent excluded surface
SMA	Steric mass action
SPP	Similarity-property principle

TDM	Transport-dispersive model
USP	Upstream processing
UV	Ultraviolet
vdW	Van der Waals
VH	Heavy chain variable domain
VL	Light chain variable domain
VR	Variable region
WP	Weak partitioning

Notation

Latin Symbols

Symbol	Description	Units
a	Thermodynamic interaction constant	Pam^6mol^{-2}
c	Mobile phase concentration	$molm^{-3}$
c_∞	Ion bulk concentration	$molm^{-3}$
c_b	Interstitial volume concentration	$molm^{-3}$
c_in	Column inlet concentration	$molm^{-3}$
c_p	Particle liquid phase concentration	$molm^{-3}$
c_s	Salt concentration	$molm^{-3}$
c_s^F	Final gradient salt concentration	$molm^{-3}$
c_s^I	Initial gradient salt concentration	$molm^{-3}$
c_s^R	Retention gradient salt concentration	$molm^{-3}$
c_v	Liquid phase molar density	$molm^{-3}$
d_c	Column diameter	m
d_p	Resin particle diameter	m
D	Data	-
D_ax	Axial dispersion coefficient	m^2s^{-1}
D_m	Molecular diffusion coefficient	mm^2s^{-1}

D_{pore}	Pore diffusion coefficient	m^2s^{-1}
e	Elementary charge	C
E	Evidence	-
f'	Predictive distribution of functions	-
g	Gradient slope	$molm^{-6}$
G	Normalized gradient slope	$molm^{-3}$
GH	Normalized gradient slope with phase ratio	$molm^{-3}$
H	Hypothesis	-
H	Phase ratio	-
I	Ionic strength	$molm^{-3}$
k_{ads}	Adsorption rate constant	s^{-1}
k_{B}	Boltzmann constant	JK^{-1}
k_{des}	Desorption rate constant	s^{-1}
k_{eff}	Effective mass transfer coefficient	ms^{-1}
k_{film}	Film transfer coefficient	ms^{-1}
k_{kin}	Adsorption kinetic parameter	s^{-1}
$k(x, x')$	Covariance function	-
K	Equilibrium ratio	-
K_{eq}	Equilibrium constant	-
$\widetilde{K}_{\mathrm{eq}}$	Redefined equilibrium constant	-
$\widetilde{K}'_{\mathrm{eq},i}$	Modified equilibrium constant	-

K_{H}	Henry coefficient	-
K_{p}	Protein-protein interaction parameter	$mol^{-1}m^3$
K_{s}	Salt-protein interaction parameter	$mol^{-1}m^3$
K_{sec}	Distribution coefficient at non-interacting conditions	-
K_{Δ}	Delta distribution coefficient at non-interacting conditions	-
l	Characteristic length scale	-
L	Ligand species	-
L_{c}	Column length	m
m	Slope	-
$m(x)$	Mean function	-
n	Hydrophobic stoichiometric coefficient	-
N_{A}	Avogadro's constant	mol^{-1}
P	Protein component	-
$P\big(y(x)\big)$	Probability distribution	-
$\mathrm{P}_{\mathrm{ads}}$	Adsorbed protein-ligand complex	-
$\mathrm{P}_{\mathrm{sol}}$	Solute protein component	-
PL	Protein-ligand complex	-
q	Concentration bound per resin backbone volume	$molm^{-3}$
$q^{\max}$	Saturation capacity of resin backbone	$molm^{-3}$
r	Protein radius	nm
r_{p}	Resin particle radius	m

R	Ideal gas constant	$Jmol^{-1}K$
s	Hydrophobic steric shielding factor	-
S	Counterion	
t^{R}	Retention time	s
t_0	Dead time	s
T	Absolute temperature	K
u	Superficial velocity	ms^{-1}
u_{int}	Interstitial velocity	ms^{-1}
V_0	Dead volume	m^3
V_{col}	Column volume	m^3
V_G	Gradient volume	m^3
w	Regression weight (slope)	-
w_p	Propagation velocity	ms^{-1}
W	Water molecule	-
x	Independent variable vector	-
x'	Unobserved and independent variable vector	-
X	Independent variable matrix	-
y	Dependent variable vector	-
y_0	Intercept	-
z_s	Effective counterion charge	-
z_p	Effective protein charge	-

Greek Symbols

Symbol	Description	Units
γ	Chemical activity coefficient	-
γ_{L}	Chemical activity coefficient ligand	-
γ_{LS}	Chemical activity coefficient ligand-salt	-
$\gamma_{\mathrm{P}_{\mathrm{ads},i}}$	Chemical activity coefficient protein-ligand	-
$\gamma_{\mathrm{P}_{\mathrm{sol},i}}$	Chemical activity coefficient protein	-
$\tilde{\gamma}_{\mathrm{P}_{\mathrm{sol},i}}$	Asymmetric activity coefficient protein	-
γ_{S}	Chemical activity coefficient salt	-
∇^2	3D Laplace operator	-
ε	Error term	-
ε_{r}	Relative permittivity	-
ε_0	Vacuum permittivity	$C^2N^{-1}m^{-1}$
ε_{b}	Chromatography interstitial porosity	-
ε_{p}	Chromatography particle porosity	-
ε_{t}	Chromatography total porosity	-
η	Dynamic viscosity	$kgm^{-1}s^{-1}$
μ	Mean	-
ν	Electrostatic stoichiometric coefficient (characteristic charge)	-
Λ_{HIC}	Hydrophobic ligand density	$molm^{-3}$
Λ_{IEX}	Ionic ligand density	$molm^{-3}$

ρ	Density	kgm^{-3}
ρ_e	Local electric charge density	Cm^{-3}
σ	Electrostatic steric shielding factor	-
Σ	Covariance matrix	-
ψ	Electric potential	V

Subscripts

Symbol	Description
i	Component i
j	Counterion j
l	Ligand l
p	Protein
s	Salt

Appendix A Supplementary Material Chapter 3

Table A1: Column and component specific mass transfer parameters.

Parameter	Symbol	Value	Unit	Derivation
Interstitial velocity	u_{int}	0.915	mms^{-1}	$\frac{u}{\varepsilon_{\text{b}}}$ [43,44,73]
Axial dispersion	D_{ax}	0.125	mm^2s^{-1}	Chung and Wen correlation: $D_{\text{ax}} = \frac{u_{\text{int}} d_{\text{p}} \varepsilon_{\text{b}}}{0.2+0.011\left(\varepsilon_{\text{b}} \frac{u_{\text{int}} d_{\text{p}} \rho}{\eta}\right)^{0.48}}$ [43,44,144]
Molecule specific hydrodynamic radius	r_i	4.1, 5.5, 6.3	nm	Experimental (DLS) literature [138–140]
Dynamic viscosity	η	0.001	$\text{kgm}^{-1}\text{s}^{-1}$	Literature [43]
Molecular diffusion	$D_{\text{m},i}$	5.3e-6, 4.0e-6, 3.5e-6	mm^2s^{-1}	Stokes-Einstein equation: $D_{\text{m}} = \frac{RT}{6\pi\eta N r_i}$ [43]
Pore diffusion	$D_{\text{p},i}$	3.0e-6, 2.2e-6, 1.9e-6	mm^2s^{-1}	Mackie and Meares correlation: $D_{\text{p}} = \left(\frac{\varepsilon_{\text{p}}}{2-\varepsilon_{\text{p}}}\right)^2 D_{\text{m}}$ [43,44]
Film transfer parameter	$k_{\text{f},i}$	2.9e-2, 2.5e-2, 2.3e-2	mms^{-1}	Penetration theory correlation: $k_{\text{f}} = \sqrt{\frac{4 D_{\text{m}} u_{\text{int}}}{\pi 2 r_i}}$ [43]
Effective mass transfer coefficient	$k_{\text{eff},i}$	2.8e-3, 2.1e-3, 1.9e-3	mms^{-1}	Series connection: $\frac{1}{k_{\text{eff}}} = \frac{1}{k_{\text{p}}} + \frac{1}{k_{\text{f}}} = \frac{2 r_i}{10 \varepsilon_{\text{p}} D_{\text{p}}} + \frac{1}{k_{\text{f}}}$ [44]
Effective mass transfer coefficient salt	$k_{\text{eff},s}$	12.5e-3	mms^{-1}	Transport dispersive model: $k_{\text{eff},s} \cong r_{\text{p}}/3$ [43,44]

Appendix B Supplementary Material Chapter 4

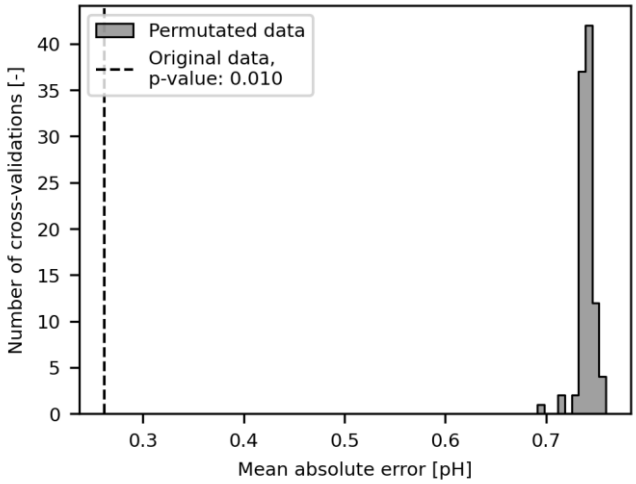

Fig. B1: Significance of y-scrambled cross-validation for pH elution prediction.The cross-validated error of the original data set is plotted against the cross-validated error during 100 permutations of the target variable. The p-value represents the probability of obtaining the original cross-validation score by chance, serving as an indicator of the true dependency between the target variable and the model features.

Appendix C Supplementary Material Chapter 5

C.1 Figures

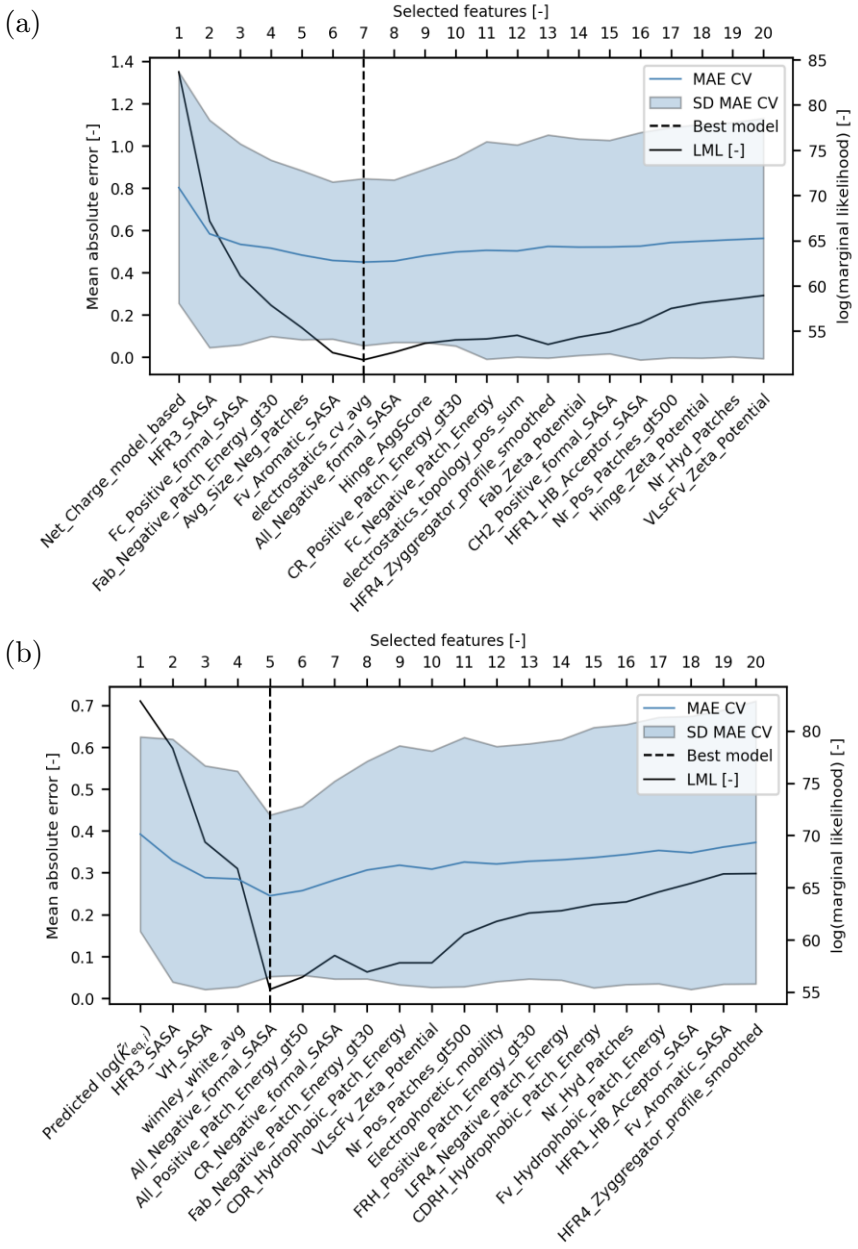

Fig. C1: Recursive feature elimination with cross-validation for training data.a) Logarithmic equilibrium constant, and b) Logarithmic salt-protein interaction parameter. Starting with ten features (right-hand side of each subfigure), an initial model is fitted using all

features. In each iteration, the model is assessed by calculating the cross-validation mean absolute error (MAE CV), standard deviation (SD) of cross-validation scores, and log(marginal likelihood) (LML) of the model. The feature with the lowest permutation importance is removed per iteration (bottom x-axis). The best model, with the lowest LML, is chosen when only one feature remains.

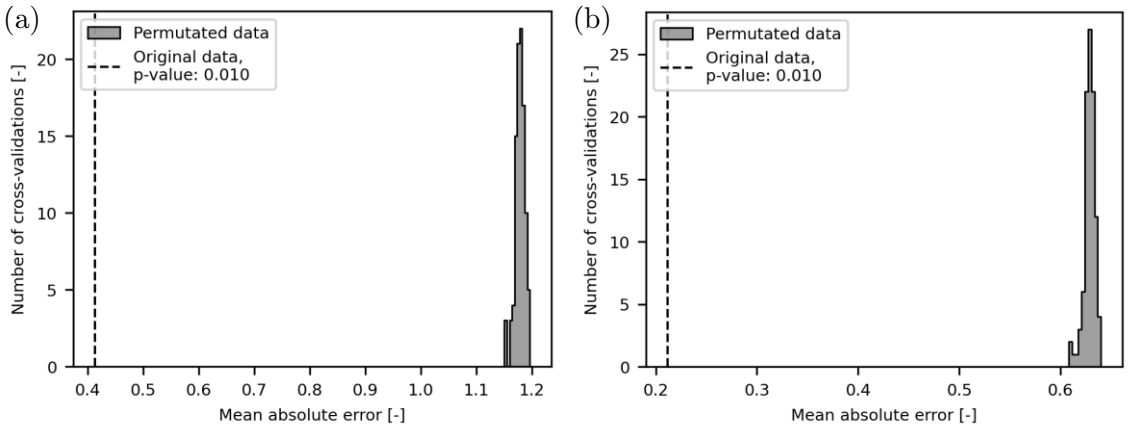

Fig. C2: Evaluating the significance of y-scrambled cross-validation in adsorption parameter prediction.a) Logarithmic equilibrium constant, and b) Logarithmic salt-protein interaction parameter. The graph compares the cross-validated error of the original dataset with the error after 100 permutations of the target variable. The p-value represents the likelihood of obtaining a cross-validation score by chance, suggesting the absence of a genuine correlation between the target variable and the model features.

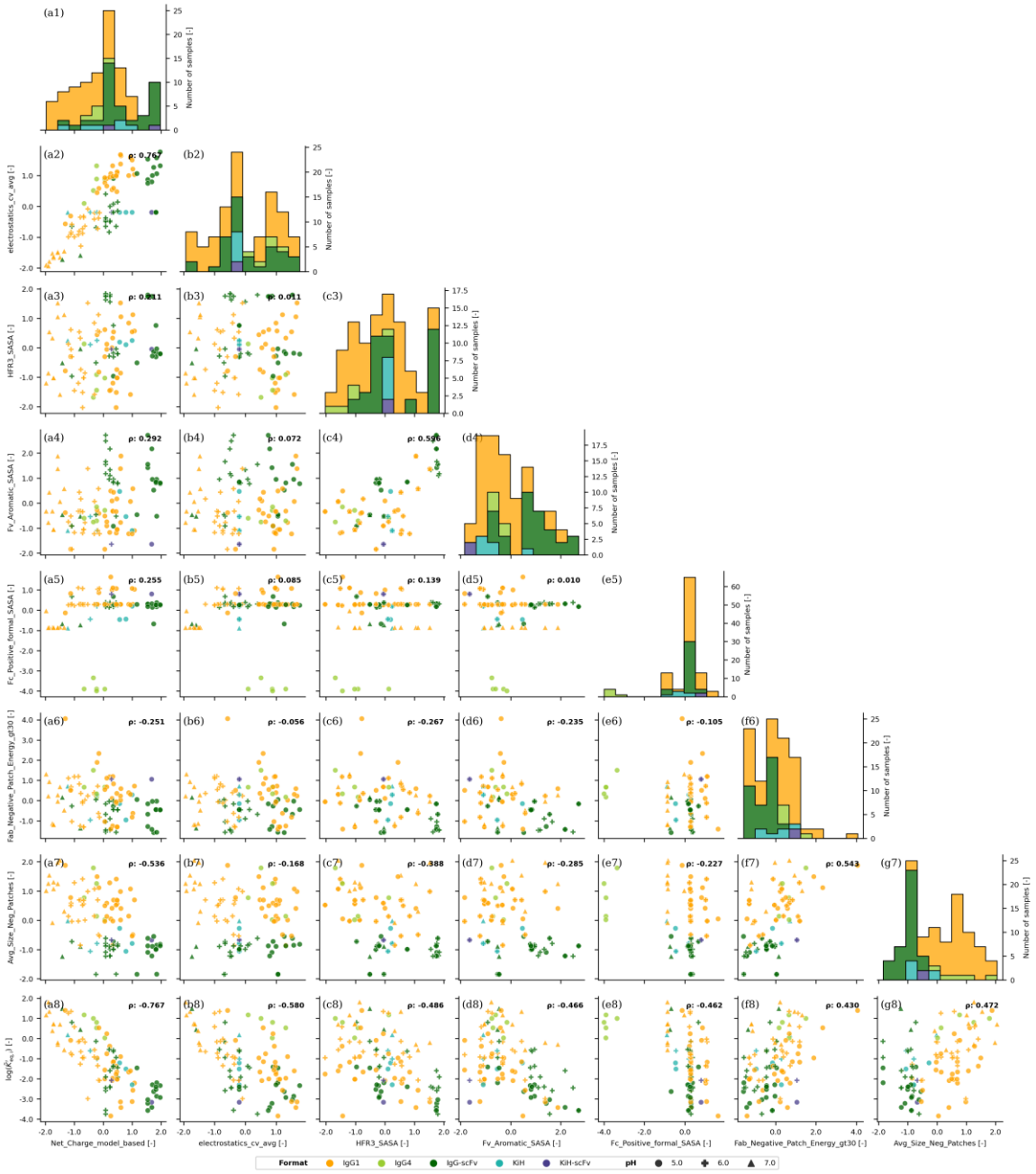

Fig. C3: Pairwise relationship and distribution of model features and logarithmic equilibrium constant $log(\tilde{K}'_{eq,i})$.The diagonal subfigures display the univariate histograms of the model features, as shown on the bottom x-axis. The off-diagonal subfigures illustrate the bivariate relationship between model features, while the bottom row presents the relationship between the target variable and the model features. Features are arranged in descending order based on their absolute Pearson correlation coefficient (p) towards the target variable within the training data, as displayed in the upper right corner of each subplot. All subfigures distinguish between antibody formats.

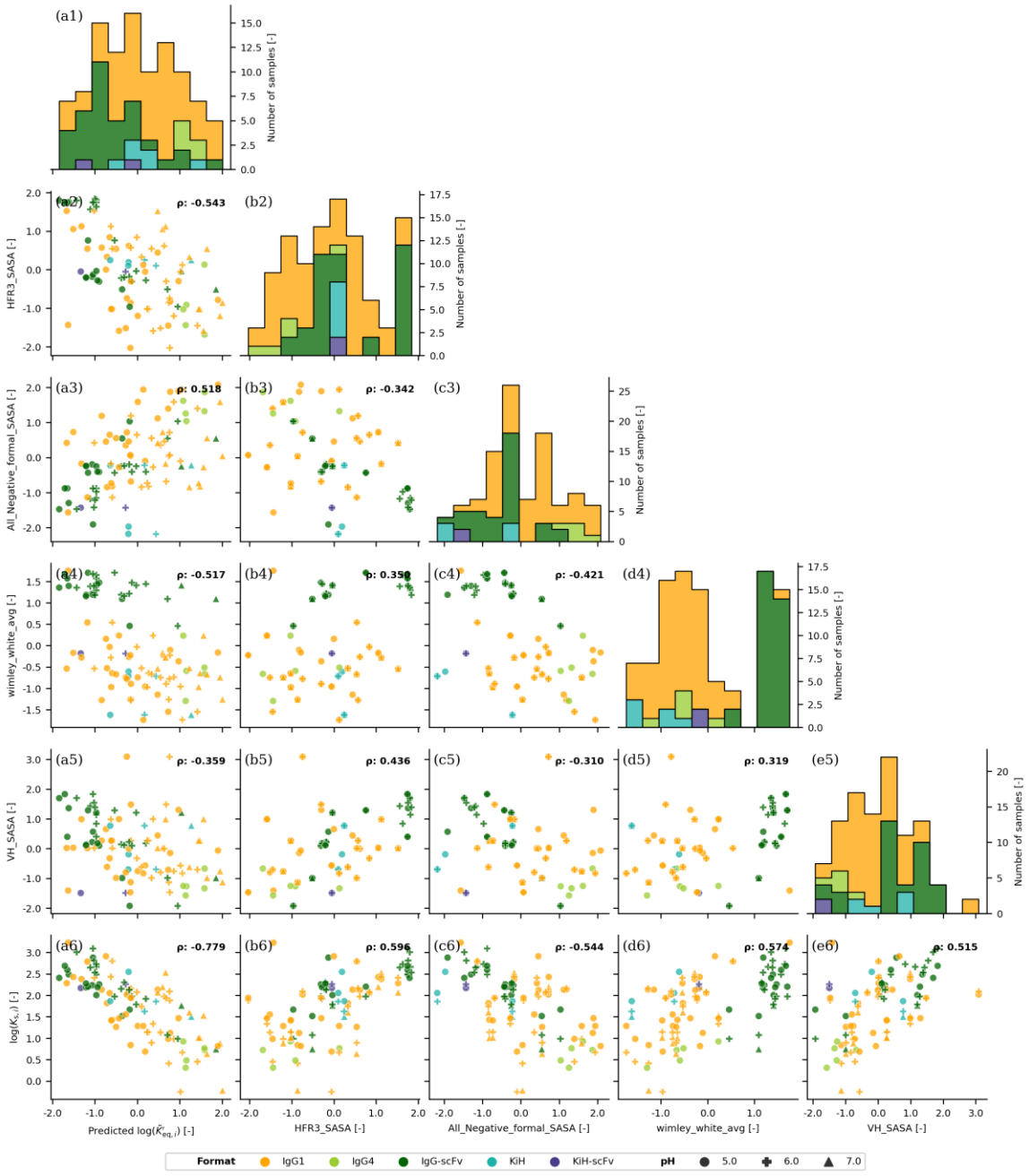

Fig. C4: Pairwise relationship and distribution of model features and logarithmic salt-protein interaction parameter $\log(K_{s,i})$.The diagonal subfigures display the univariate histograms of the model features, as shown on the bottom x-axis. The off-diagonal subfigures illustrate the bivariate relationship between model features, while the bottom row presents the relationship between the target variable and the model features. Features are arranged in descending order based on their absolute Pearson correlation coefficient (p) towards the target variable within the training data, as displayed in the upper right corner of each subplot. All subfigures distinguish between antibody formats.

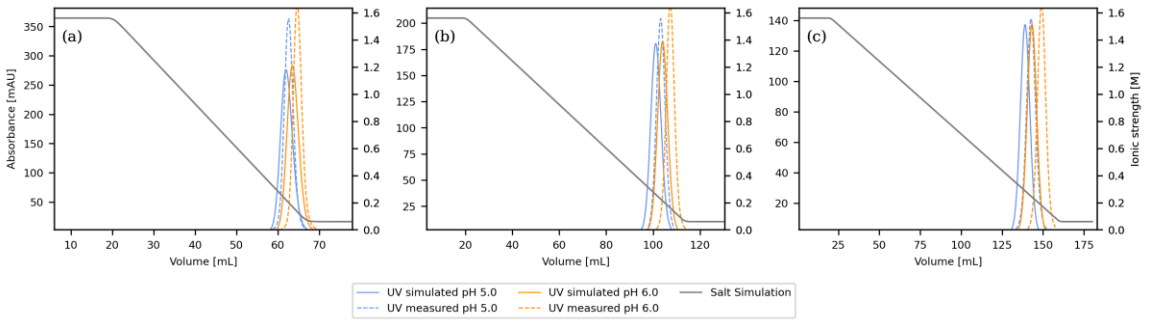

Fig. C5: Comparison of experimental and predicted chromatograms for IgG1 (3) under low loading density conditions.Subfigure (a) to (c) show the 10, 20, and 30 CV LGE experiments for the test set molecule at pH levels of 5.0 and 6.0, respectively. These chromatograms are generated using multiscale modeling, which involves predicting thermodynamic parameters from antibody structure through QSPR models. These parameters are then incorporated into a transport dispersive model for chromatographic simulation.

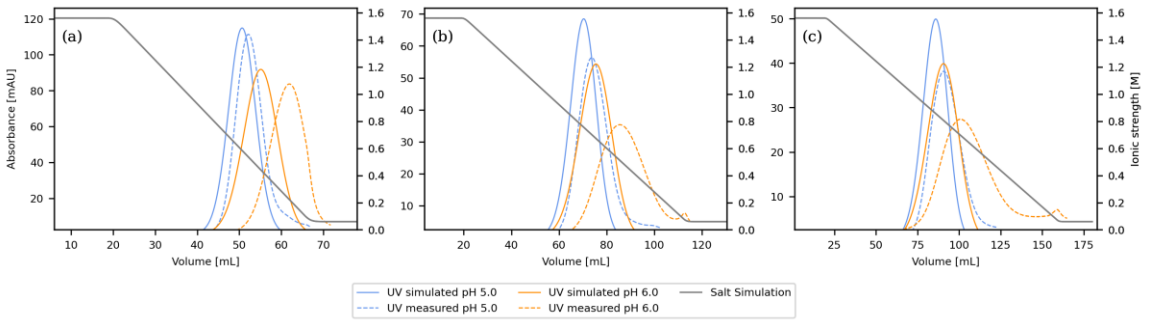

Fig. C6: Comparison of experimental and predicted chromatograms for IgG1 (14) under low loading density conditions.Subfigure (a) to (c) show the 10, 20, and 30 CV LGE experiments for the test set molecule at pH levels of 5.0 and 6.0, respectively. These chromatograms are generated using multiscale modeling, which involves predicting thermodynamic parameters from antibody structure through QSPR models. These parameters are then incorporated into a transport dispersive model for chromatographic simulation.

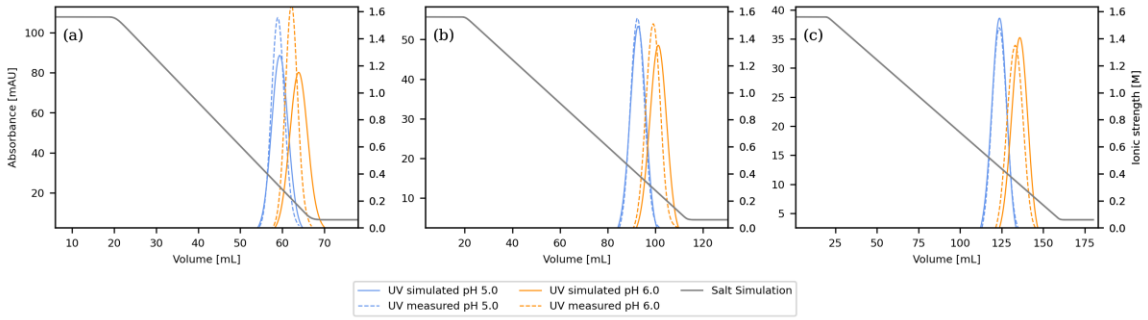

Fig. C7: Comparison of experimental and predicted chromatograms for IgG1 (20) under low loading density conditions.Subfigure (a) to (c) show the 10, 20, and 30 CV LGE experiments for the test set molecule at pH levels of 5.0 and 6.0, respectively. These chromatograms are generated using multiscale modeling, which involves predicting thermodynamic parameters from antibody structure through QSPR models. These parameters are then incorporated into a transport dispersive model for chromatographic simulation.

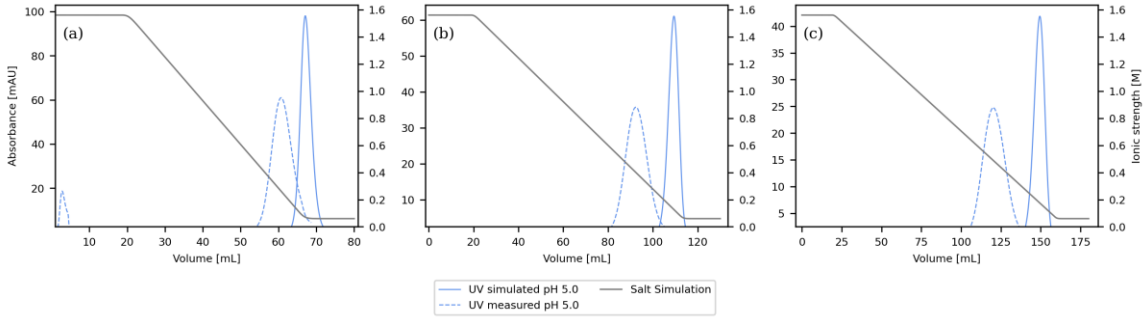

Fig. C8: Comparison of experimental and predicted chromatograms for IgG1 (22) under low loading density conditions.Subfigure (a) to (c) show the 10, 20, and 30 CV LGE experiments for the test set molecule at pH 5.0. These chromatograms are generated using multiscale modeling, which involves predicting thermodynamic parameters from antibody structure through QSPR models. These parameters are then incorporated into a transport dispersive model for chromatographic simulation.

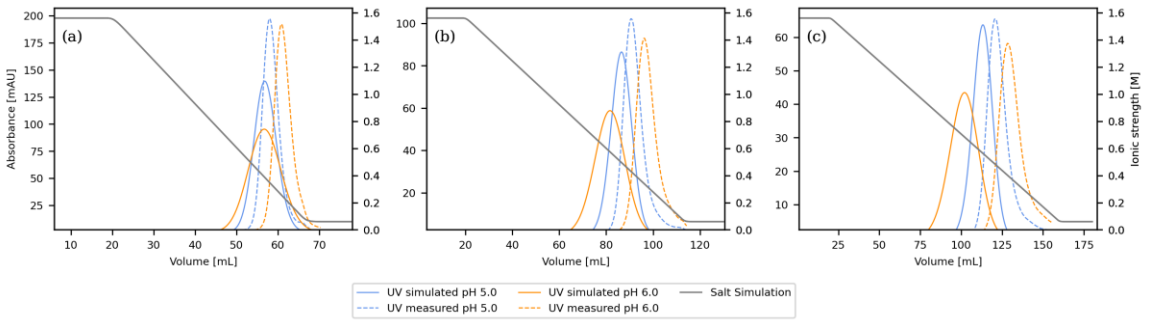

Fig. C9: Comparison of experimental and predicted chromatograms for IgG1 (28) under low loading density conditions.Subfigure (a) to (c) show the 10, 20, and 30 CV LGE experiments for the test set molecule at pH levels of 5.0 and 6.0, respectively. These chromatograms are generated using multiscale modeling, which involves predicting thermodynamic parameters from antibody structure through QSPR models. These parameters are then incorporated into a transport dispersive model for chromatographic simulation.

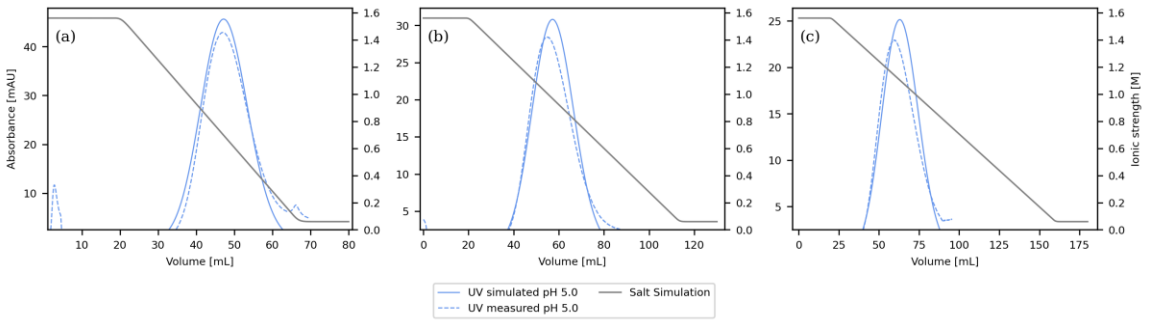

Fig. C10: Comparison of experimental and predicted chromatograms for IgG4 (6) under low loading density conditions.Subfigure (a) to (c) show the 10, 20, and 30 CV LGE experiments for the test set molecule at pH 5.0. These chromatograms are generated using multiscale modeling, which involves predicting thermodynamic parameters from antibody structure through QSPR models. These parameters are then incorporated into a transport dispersive model for chromatographic simulation.

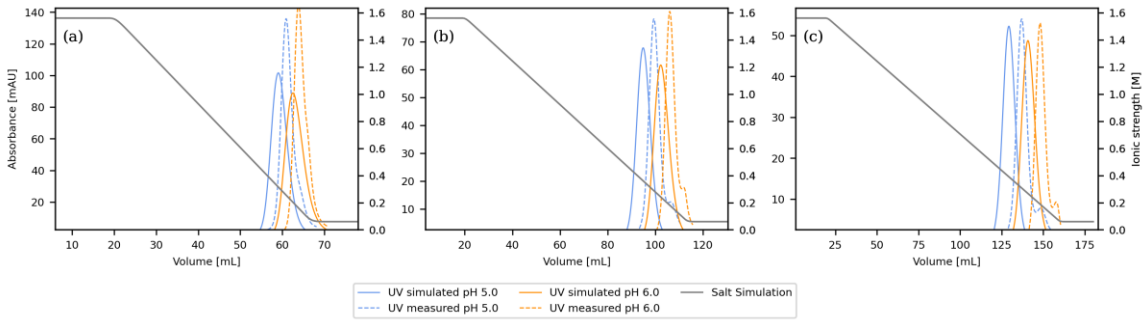

Fig. C11: Comparison of experimental and predicted chromatograms for IgG-scFv (6) under low loading density conditions.Subfigure (a) to (c) show the 10, 20, and 30 CV LGE experiments for the test set molecule at pH levels of 5.0 and 6.0, respectively. These chromatograms are generated using multiscale modeling, which involves predicting thermodynamic parameters from antibody structure through QSPR models. These parameters are then incorporated into a transport dispersive model for chromatographic simulation.

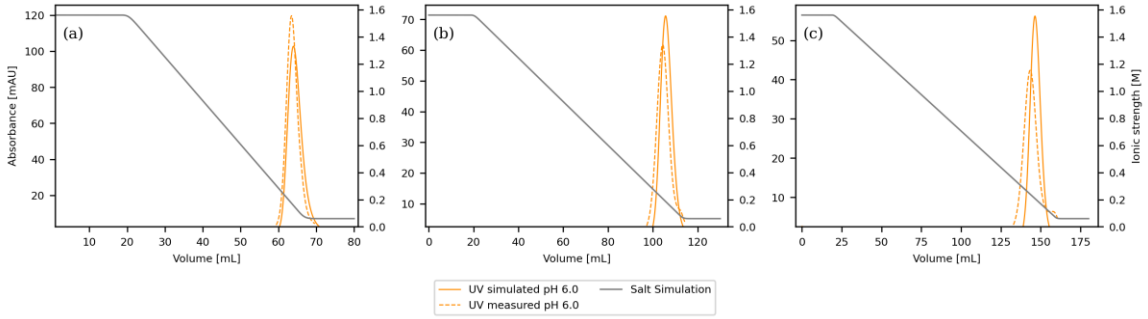

Fig. C12: Comparison of experimental and predicted chromatograms for IgG-scFv (11) under low loading density conditions.Subfigure (a) to (c) show the 10, 20, and 30 CV LGE experiments for the test set molecule at pH 6.0. These chromatograms are generated using multiscale modeling, which involves predicting thermodynamic parameters from antibody structure through QSPR models. These parameters are then incorporated into a transport dispersive model for chromatographic simulation.

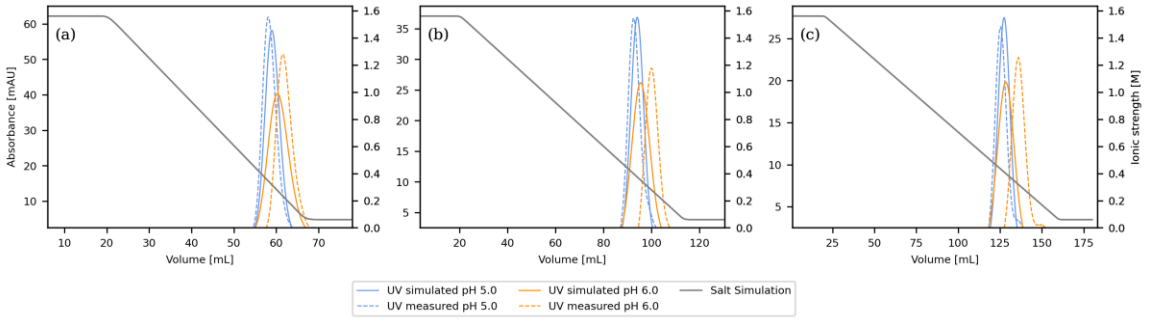

Fig. C13: Comparison of experimental and predicted chromatograms for IgG-scFv (15) under low loading density conditions.Subfigure (a) to (c) show the 10, 20, and 30 CV LGE experiments for the test set molecule at pH levels of 5.0 and 6.0, respectively. These chromatograms are generated using multiscale modeling, which involves predicting thermodynamic parameters from antibody structure through QSPR models. These parameters are then incorporated into a transport dispersive model for chromatographic simulation.

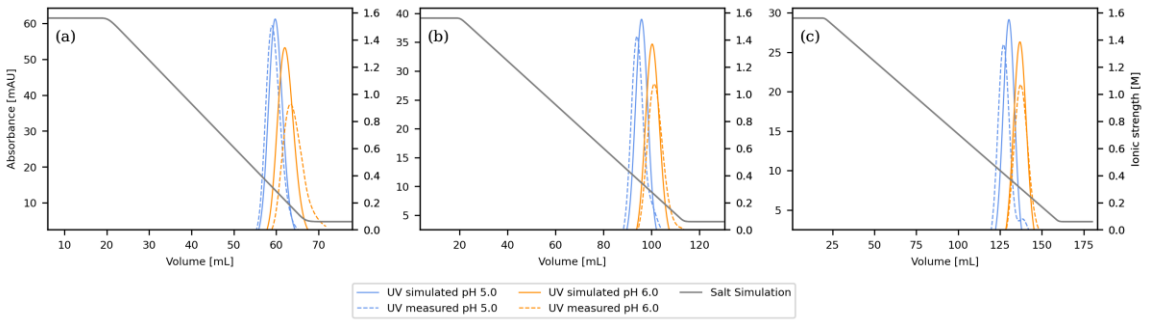

Fig. C14: Comparison of experimental and predicted chromatograms for IgG-scFv (18) under low loading density conditions. Subfigure (a) to (c) show the 10, 20, and 30 CV LGE experiments for the test set molecule at pH levels of 5.0 and 6.0, respectively. These chromatograms are generated using multiscale modeling, which involves predicting thermodynamic parameters from antibody structure through QSPR models. These parameters are then incorporated into a transport dispersive model for chromatographic simulation.

C.2 Tables

Table C1: Overview of custom descriptors based on SURFMAP analysis [210].

Descriptor	Explanation
r_sum [d]	Summation of normalized distance to molecule center of mass.
circular_variance_atom_sum [e]	Summation of atom-based inverted circular variance values.
electrostatics_sum [f]	Summation of electrostatic potential.
wimley_white_sum [g]	Summation of residue-based Wimley-White hydrophobicities.
topology_sum [h]	Summation of normalized radius times inverted circular variance.
MMA_pos_sum	Summation of additive values of the "multimodal anionic" score.
electrostatics_r_sum	Summation of electrostatic potential times normalized radius.
electrostatics_cv_sum	Summation of electrostatic potential times inverted circular variance.
electrostatics_topology_sum	Summation of electrostatic potential times topology score (see above).
wimley_white_r_sum	Summation of Wimley-White values times normalized radius.
wimley_white_cv_sum	Summation of Wimley-White values times inverted circular variance.
wimley_white_topology_sum	Summation of Wimley-White values times topology score (see above).
MMA_pos_r_sum	Summation of MMA_pos score times normalized radius.

MMA_pos_cv_sum	Summation of MMA_pos score times inverted circular variance.
MMA_pos_topology_sum	Summation of MMA_pos score times topology score (see above).
electrostatics_pos_sum	Summation of strictly positive electrostatic potential.
electrostatics_neg_sum	Summation of strictly negative electrostatic potential.
wimley_white_pos_sum	Summation of strictly positive Wimley-White values.
wimley_white_neg_sum	Summation of strictly negative Wimley-White values.
electrostatics_r_pos_sum	Summation of strictly positive electrostatic potential times normalized radius.
electrostatics_r_neg_sum	Summation of strictly negative electrostatic potential times normalized radius.
electrostatics_cv_pos_sum	Summation of strictly positive electrostatic potential times inverted circular variance.
electrostatics_cv_neg_sum	Summation of strictly negative electrostatic potential times inverted circular variance.
electrostatics_topology_pos_sum	Summation of strictly positive electrostatic potential times topology score (see above).
electrostatics_topology_neg_sum	Summation of strictly negative electrostatic potential times topology score (see above).
wimley_white_r_pos_sum	Summation of strictly positive Wimley-White values times normalized radius.
wimley_white_r_neg_sum	Summation of strictly negative Wimley-White values times normalized radius.
wimley_white_cv_pos_sum	Summation of strictly positive Wimley-White values times inverted circular variance.
wimley_white_cv_neg_sum	Summation of strictly negative Wimley-White values times inverted circular variance.

wimley_white_topology_pos_sum	Summation of strictly positive Wimley-White values times topology score (see above).
wimley_white_topology_neg_sum	Summation of strictly negative Wimley-White values times topology score (see above).
r_avg	Average of normalized distance to molecule center of mass.
circular_variance_atom_avg	Average of atom-based inverted circular variance values.
electrostatics_avg	Average of electrostatic potential.
wimley_white_avg	Average of residue-based Wimley-White hydrophobicities.
topology_avg	Average of normalized radius times inverted circular variance.
MMA_pos_avg	Average of additive values of the "multimodal anionic" score.
electrostatics_r_avg	Average of electrostatic potential times normalized radius.
electrostatics_cv_avg	Average of electrostatic potential times inverted circular variance.
electrostatics_topology_avg	Average of electrostatic potential times topology score (see above).
wimley_white_r_avg	Average of Wimley-White values times normalized radius.
wimley_white_cv_avg	Average of Wimley-White values times inverted circular variance.
wimley_white_topology_avg	Average of Wimley-White values times topology score (see above).
MMA_pos_r_avg	Average of MMA_pos score times normalized radius.
MMA_pos_cv_avg	Average of MMA_pos score times inverted circular variance.

MMA_pos_topology_avg	Average of MMA_pos score times topology score (see above).
electrostatics_pos_avg	Average of strictly positive electrostatic potential.
electrostatics_neg_avg	Average of strictly negative electrostatic potential.
wimley_white_pos_avg	Average of strictly positive Wimley-White values.
wimley_white_neg_avg	Average of strictly negative Wimley-White values.
electrostatics_r_pos_avg	Average of strictly positive electrostatic potential times normalized radius.
electrostatics_r_neg_avg	Average of strictly negative electrostatic potential times normalized radius.
electrostatics_cv_pos_avg	Average of strictly positive electrostatic potential times inverted circular variance.
electrostatics_cv_neg_avg	Average of strictly negative electrostatic potential times inverted circular variance.
electrostatics_topology_pos_avg	Average of strictly positive electrostatic potential times topology score (see above).
electrostatics_topology_neg_avg	Average of strictly negative electrostatic potential times topology score (see above).
wimley_white_r_pos_avg	Average of strictly positive Wimley-White values times normalized radius.
wimley_white_r_neg_avg	Average of strictly negative Wimley-White values times normalized radius.
wimley_white_cv_pos_avg	Average of strictly positive Wimley-White values times inverted circular variance.
wimley_white_cv_neg_avg	Average of strictly negative Wimley-White values times inverted circular variance.
wimley_white_topology_pos_avg	Average of strictly positive Wimley-White values times topology score (see above).

wimley_white_topology_neg_avg	Average of strictly negative Wimley-White values times topology score (see above).

[d] Surface radius, normalized by maximum distance to molecule center of mass [209].

[e] Circular variance, inverted like (1-circular_variance) [212].

[f] Electrostatic potential, calculated at process pH and an ionic strength of 0.2 M [109].

[g] Wimley-White hydrophobicity, centered to glycine as -0.01 kcal/mol [211].

[h] MMA_pos score is defined as negative electrostatic potential multiplied with positive (hydrophobic) Whimley-White values followed by a change of sign. Should represent increased binding interaction in anionic multimodal chromatography.

Table C2: Parameters of the system and column characterization, adapted from Hess et al. [177].

Parameter	**Symbol**	**Value**	**Unit**	**Derivation**
Length	L_c	100	mm	Manufacturer
Diameter	d_c	7.7	mm	Manufacturer
Column volume	V_c	4.657	mL	$V_c = \pi \frac{{d_c}^2}{4} L_c$
Bead radius	r_p	37.5	µm	Manufacturer
System dead volume	V_0	0.321 1.972	mL	Tracer, no column installed
Interstitial porosity	ε_b	0.364	–	$\frac{V_b}{V_c}$
Total porosity	ε_t	0.872	–	$\frac{V_t}{V_c}$
Particle porosity	ε_p	0.799	–	$\frac{V_t - V_b}{V_c - V_b}$
Ionic capacity	Λ	1.007	$\mathrm{molm^{-3}}$	$\Lambda = \frac{c_{Cl-} V_{Cl-}}{V_{col}(1-\varepsilon_t)}$
Superficial velocity	u	0.333	$\mathrm{mms^{-1}}$	Experimental setup

Table C3: Column and component specific mass transfer parameters, adapted from Hess et al. [177].

Parameter	Symbol	Value	Unit	Derivation
Interstitial velocity	u_{int}	0.915	mms^{-1}	$\frac{u}{\varepsilon_{\text{b}}}$ [43,44]
Axial dispersion	D_{ax}	0.125	mm^2s^{-1}	Chung and Wen correlation: $D_{\text{ax}} = \frac{u_{\text{int}} d_{\text{p}} \varepsilon_{\text{b}}}{0.2+0.011\left(\varepsilon_{\text{b}} \frac{u_{\text{int}} d_{\text{p}} \rho}{\eta}\right)^{0.48}}$ [43,44]
Molecule-specific hydrodynamic radius	r_i	4.1, 5.5, 6.3	nm	Experimental (DLS) literature [138–140]
Dynamic viscosity	η	0.001	$\text{kgm}^{-1}\text{s}^{-1}$	Literature [43]
Molecular diffusion	$D_{\text{m},i}$	5.3e-6, 4.0e-6, 3.5e-6	mm^2s^{-1}	Stokes-Einstein equation: $D_{\text{m}} = \frac{RT}{6\pi\eta N r_i}$ [43]
Pore diffusion	$D_{\text{p},i}$	3.0e-6, 2.2e-6, 1.9e-6	mm^2s^{-1}	Mackie and Meares correlation: $D_{\text{p}} = \left(\frac{\varepsilon_{\text{p}}}{2-\varepsilon_{\text{p}}}\right)^2 D_{\text{m}}$ [43,44]
Film transfer parameter	$k_{\text{f},i}$	2.9e-2, 2.5e-2, 2.3e-2	mms^{-1}	Penetration theory correlation: $k_{\text{f}} = \sqrt{\frac{4 D_{\text{m}} u_{\text{int}}}{\pi 2 r_i}}$ [43]
Effective mass transfer coefficient	$k_{\text{eff},i}$	2.8e-3, 2.1e-3, 1.9e-3	mms^{-1}	Series connection: $\frac{1}{k_{\text{eff}}} = \frac{1}{k_{\text{p}}} + \frac{1}{k_{\text{f}}} = \frac{2 r_i}{10 \varepsilon_{\text{p}} D_{\text{p}}} + \frac{1}{k_{\text{f}}}$ [44]
Effective mass transfer coefficient salt	$k_{\text{eff},s}$	12.5e-3	mms^{-1}	Transport dispersive model: $k_{\text{eff},s} \cong r_{\text{p}}/3$ [43,44]

Table C4: Protein specific mass transfer and thermodynamic model parameters based on Hess et al. [177].The initial thermodynamic parameters were estimated by linearization and neglecting resin pore accessibility, followed by refinement considering pore system accessibility.

Molecule	pH	$K_{\Delta,is}$ [−] [i]	$k_{\mathrm{eff},i}$ [$\mu m s^{-1}$] [j]	$k_{\mathrm{kin},i}$ [s] [k]	p [−] [l]	$\tilde{K}'_{\mathrm{eq},i,0}$ [−] [m]	$\tilde{K}'_{\mathrm{eq},i}$ [−] [n]	$\mathrm{CI}_{\tilde{K}'_{\mathrm{eq},i}}$ [−] [o]	$K_{s,i,0}$ [M] [p]	$K_{s,i}$ [M] [q]	$\mathrm{CI}_{K_{s,i}}$ [M] [r]
IgG1 (3)	5	0.795	2.143	40	0.067	0.004	0.044	0.033-0.055	23.955	16.239	15.290-17.188
IgG1 (3)	6	0.795	2.143	35	0.021	0.016	0.124	0.112-0.137	24.622	16.452	15.942-16.962
IgG1 (14)	5	0.795	2.143	40	0.015	0.284	0.469	0.419-0.519	4.105	3.617	3.499-3.736
IgG1 (14)	6	0.795	2.143	40	0.022	2.742	3.538	2.945-4.131	2.308	1.950	1.746-2.154
IgG1 (20)	5	0.795	2.143	30	0.020	0.092	0.209	0.199-0.219	8.256	6.987	6.893-7.081
IgG1 (20)	6	0.795	2.143	35	0.041	0.231	0.449	0.369-0.528	8.141	6.923	6.493-7.353
IgG1 (22)	5	0.795	2.143	5	0.020	0.618	0.914	0.846-0.981	4.832	4.337	4.212-4.462
IgG1 (28)	5	0.795	2.143	55	0.017	0.073	0.171	0.167-0.175	8.244	6.976	6.937-7.015
IgG1 (28)	6	0.795	2.143	70	0.038	0.199	0.385	0.317-0.454	7.619	6.536	6.155-6.917
IgG1 (30)	5	0.795	2.143	10	0.001	0.059	0.132	0.127-0.138	4.821	4.118	4.079-4.156
IgG1 (30)	6	0.795	2.143	15	0.001	0.183	0.380	0.370-0.390	3.810	3.143	3.119-3.167
IgG1 (30)	7	0.795	2.143	20	0.024	1.570	2.054	1.778-2.329	3.079	2.760	2.579-2.941

IgG4 (6)	5	0.795	2.143	90	0.051	0.743	2.233	1.852-2.614	2.467	1.366	1.219-1.513
IgG-scFv (6)	5	0.797	1.882	65	0.006	0.002	0.028	0.018-0.038	21.066	14.698	13.586-15.810
IgG-scFv (6)	6	0.797	1.882	100	0.066	0.001	0.028	0.002-0.054	37.676	22.055	17.571-26.539
IgG-scFv (11)	6	0.797	1.882	80	0.087	0.036	0.149	0.086-0.211	16.765	12.564	10.998-14.130
IgG-scFv (15)	5	0.797	1.882	40	0.012	0.012	0.053	0.049-0.058	12.330	9.742	9.562-9.923
IgG-scFv (15)	6	0.797	1.882	70	0.066	0.045	0.147	0.096-0.199	12.623	10.008	9.034-10.982
IgG-scFv (18)	5	0.797	1.882	40	0.008	0.026	0.091	0.082-0.100	11.261	9.050	8.831-9.269
IgG-scFv (18)	6	0.797	1.882	55	0.013	0.273	0.531	0.441-0.621	8.534	7.183	6.722-7.644

[i] Delta distribution coefficient $K_{\Delta,is}$, derived from isocratic pulse experiments.

[j] Effective mass transfer coefficient $k_{\mathrm{eff},i}$, calculated from empirical correlations listed in the Appendix C, Table C4.

[k] Kinetic coefficient $k_{\mathrm{kin},i}$, fitted to peak shape.

[l] p-value for two-sided Wald test, accessing the linearity of the initial parameter determination, derived from Eq. (5.4).

[m] Initial equilibrium constant $\widetilde{K}'_{\mathrm{eq},i,0}$, derived from Eq. (5.4) by neglecting protein-specific pore accessibility.

[n] Refined equilibrium constant $\widetilde{K}'_{\mathrm{eq},i}$, derived from Eq. (5.4) by considering protein-specific pore accessibility.

[o] Confidence interval (95%) of $\widetilde{K}'_{\mathrm{eq},i}$ determination.

[p] Initial salt-protein interaction parameter $K_{s,i}$, derived from Eq. (5.4) by neglecting protein-specific pore accessibility.

[q] Refined salt-protein interaction parameter $K_{s,i}$, derived from Eq. (5.4) by considering protein-specific pore accessibility.

[r] Confidence interval (95%) of $K_{s,i}$ determination.