

BioCapt® Single-Use

Aktiver Luftkeimmonitor

Testfertiger, aktiver Ersatz für Sedimentationsplatten



- Kontinuierliche Luftprobenahme und Überwachung lebensfähiger Partikel
- Testfertiger Ersatz für Absetzplatten
- Weniger Handhabungsschritte und weniger falsch positive Untersuchungen
- Für den mobilen oder remote Einsatz
- Auswahl verschiedener Agarformulierungen und Flussraten
- Erfüllt ISO 14698-1 und den erwarteten EU-GMP-Annex 1 Einweganforderungen
- Verwendung mit aktiven Luftkeimsammler-Instrumenten für eine vollständig validierte Lösung

Mehr erfahren



**PARTICLE
MEASURING
SYSTEMS®**
a spectris company

Kontaktieren Sie uns für mehr Informationen:


pmeasuring.com/de

T: +49 351 8896 3850

E: pmsgermany@pmeasuring.com

ARTICLE

A multiscale modeling method for therapeutic antibodies in ion exchange chromatography

David Saleh^{1,2}  | Rudger Hess^{1,2} | Michelle Ahlers-Hesse³ |
Federico Rischawy^{1,3} | Gang Wang³ | Jan-Hendrik Grosch² | Thomas Schwab² |
Simon Kluters³ | Joey Studts³ | Jürgen Hubbuch¹

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

²Early Stage Bioprocess Development, Boehringer Ingelheim Pharma GmbH & Co. KG, Biberach an der Riss, Germany

³Late Stage DSP Development, Boehringer Ingelheim Pharma GmbH & Co. KG, Biberach an der Riss, Germany

Correspondence

Jürgen Hubbuch, Karlsruhe Institute of Technology (KIT), Institute of Process Engineering in Life Sciences, Section IV: Biomolecular Separation Engineering, Karlsruhe, Germany.
Email: Juergen.Hubbuch@kit.edu

Abstract

The development of biopharmaceutical downstream processes relies on exhaustive experimental studies. The root cause is the poorly understood relationship between the protein structure of monoclonal antibodies (mAbs) and their macroscopic process behavior. Especially the development of preparative chromatography processes is challenged by the increasing structural complexity of novel antibody formats and accelerated development timelines. This study introduces a multiscale in silico model consisting of homology modeling, quantitative structure–property relationships (QSPR), and mechanistic chromatography modeling leading from the amino acid sequence of a mAb to the digital representation of its cation exchange chromatography (CEX) process. The model leverages the mAbs' structural characteristics and experimental data of a diverse set of 21 therapeutic antibodies to predict elution profiles of two mAbs that were removed from the training data set. QSPR modeling identified mAb-specific protein descriptors relevant for the prediction of the thermodynamic equilibrium and the stoichiometric coefficient of the adsorption reaction. The consideration of two discrete conformational states of IgG4 mAbs enabled prediction of split-peak elution profiles. Starting from the sequence, the presented multiscale model allows in silico development of chromatography processes before protein material is available for experimental studies.

KEYWORDS

biopharmaceutical downstream processing, ion exchange chromatography, mechanistic chromatography modeling, multiscale modeling, quantitative structure–property relationships

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2022 The Authors. *Biotechnology and Bioengineering* published by Wiley Periodicals LLC.

1 | INTRODUCTION

Monoclonal antibodies (mAbs) and bispecific antibody formats (bsAbs) are intriguing treatment options for a wide spectrum of therapeutic areas, including oncology, hematology, inflammatory diseases (Morrison, 2020; Weidle et al., 2013), and more recently for passive immunization or treatment of infectious diseases, such as coronavirus disease 2019 (COVID-19) (Wang et al., 2020). Biopharmaceutical companies strive to accelerate the process development of biologics to bring potentially life-saving medicines to patients as quickly as possible. In light of the ongoing Coronavirus pandemic, Kelley (2020) proposed an antibody-specific development plan that shortened the time from lead identification to start of phase 1 clinical investigation from 12 to 6 months. This strategy combines novel technologies with a drastically templated platform process and a strict parallelization of experimental work packages (Kelley, 2020). While enabling an early start of phase 1 clinical studies, the development plan designed by Kelley results in a higher business risk caused by missing experimental studies during early-stage development. The biopharmaceutical industry demands novel methods supporting process development at pandemic-pace, while achieving highest product quality and robust material supply for clinical investigations.

Due to the structural similarities of different mAb products, their large-scale purification is based on the so-called platform process (Shukla et al., 2007; Shukla & Thömmes, 2010). This platform process consists of a standardized sequence of orthogonal separation mechanisms and the adaptation of process conditions is reduced to a minimum. However, the increasing structural complexity of antibody formats and the poorly understood adsorption mechanisms in preparative chromatography challenge downstream processing (DSP) under standardized conditions. For cation exchange (CEX) chromatography, the antibody format (Luo et al., 2015, 2014) as well as minimal changes in the primary structure of mAbs (Saleh et al., 2021) influence elution profiles and the resulting optimal operating conditions. Multiple authors propose to increase process understanding by using mechanistic models as digital twins of the manufacturing process (Cardillo et al., 2021; Narayanan et al., 2020; Smiatek et al., 2020). Mechanistic chromatography models consist of partial differential equations describing mass transport and thermodynamic adsorption phenomena within the chromatography column. In silico screening of process conditions via mechanistic modeling enables efficient process optimization (Pirring et al., 2017) and robustness analysis (Borg et al., 2014). For CEX chromatography, the stoichiometric displacement model (SDM) allows simulation of protein adsorption and desorption under diluted loading conditions (Boardman & Partridge, 1955; Velayudhan & Horváth, 1988). In light of the Quality by Design (QbD) initiative (Mollerup et al., 2008), regulatory authorities support the use of mechanistic models to increase product quality by understanding the fundamental relationships between process parameters and quality attributes (Mollerup et al., 2008; Rathore & Winkle, 2009). Despite successful case studies and regulatory initiatives, the cumbersome model calibration can

negate the potential benefits of mechanistic models for industrial application. Mechanistic model parameters are not directly measurable and must be determined for each molecule entering the development phase. Especially during the early phases of DSP development, protein material and time for model calibration are limited. Further, the experiments demanded for the determination of SDM parameters, including experiments under diluted loading conditions (Yamamoto et al., 1983), are not performed during conventional development workflows. Understanding the relationships between the protein structure of therapeutic antibodies and their adsorption isotherm parameters would support a model calibration demanding less experiments.

The inability of current models to transfer process knowledge from existing mAbs to new drug candidates has increased interest in machine learning and AI-based methods for bioprocessing (Hutter et al., 2021; Smiatek et al., 2020; von Stosch et al., 2021; Wang et al., 2017). Quantitative structure–property relationships (QSPR) leverage machine learning algorithms and existing data to predict a target property based on the protein structure. When process understanding and experimental data is limited, QSPR models give initial insights into the developability of mAb candidates during early-stage development (Karlberg et al., 2020, 2018). Kizhedath et al. (2019) developed a QSPR model for the prediction of retention times in cross-interaction chromatography that revealed the relevance of local protein descriptors on protein–protein interactions. For the purification of proteins, QSPRs allowed the prediction of protein retention times in ion exchange chromatography (IEX) (Mazza et al., 2001) and mixed-mode chromatography (Chung et al., 2010; Robinson et al., 2020, 2017). Further, Kittelmann et al. (2017b) introduced QSPR models capable of predicting binding orientations of mAbs and model proteins (Kittelmann et al., 2017a) in IEX chromatography. For mAbs in CEX chromatography, Ishihara et al. (2005) found a correlation between the elution salt concentration and positively charged patches in the heavy chain variable region. While these QSPR models are tied to specific process conditions, multiscale models link the protein structure directly to mechanistic model parameters. Thus, multiscale models are able to predict entire chromatograms at varying process conditions. Ladiwala et al. (2005) built a QSPR model for the prediction of adsorption isotherm parameters of small model proteins. Noteworthy, their multiscale model predicted entire IEX chromatograms based on physico-chemical descriptors derived from experimentally determined protein structures (Ladiwala et al., 2005). However, protein structures of full-length mAbs are not available during process development. For relevant applications in the biopharmaceutical industry, we must integrate homology modeling, QSPRs, and mechanistic chromatography modeling in a common in silico method.

In this study, we introduce a multiscale model leading from the amino acid sequence of a therapeutic antibody to the mechanistic chromatography model of the corresponding CEX unit operation. Antibody-specific descriptors are derived from homology models built for a diverse set of full-length antibodies, including bispecific formats. QSPR modeling is then used to build an empirical

relationship between the molecular descriptors and experimentally determined SDM parameters. We perform a rigorous validation of the multiscale model by predicting chromatograms of two molecules not included in the training data set, including a split-peak elution profile observed during IgG4 purification. For new antibody candidates, our *in silico* method yields sequence-based predictions of entire CEX chromatograms at varying process conditions without the need of initial experiments. Thus, our model can accelerate development while providing crucial process understanding at the interface of research and development.

2 | METHODS

2.1 | Antibodies and homology modeling

A set of 21 therapeutic IgGs was used to train and test a QSPR model for the prediction of SDM isotherm parameters. The data set consisted of Fab ($n = 1$), IgG1 mAbs ($n = 10$), IgG4 mAbs ($n = 5$), and bsAbs ($n = 5$). Bispecific antibodies were added to the data set to enable predictions for the increasing number of complex biologics in biopharmaceutical development. Additionally, enzymatic fragmentation of IgG1 and IgG4 mAbs was performed following the protocol by Andrew and Titus (1997) to further increase size and structural diversity of the data set. All molecules investigated in our study are IgG derivatives and thus have sufficient structural similarity to be covered by a single QSPR model. The antibodies were expressed in Chinese hamster ovary cells (Boehringer Ingelheim GmbH & Co. KG). Full-length homology models of investigated mAbs and complex antibody formats were built in Maestro BioLuminate 3.7 (Schrödinger) following the method developed by Zhu et al. (2014). Table 1 lists the resolutions and R -values of the template structures used for homology modeling of full-length antibodies. The mAb crystal structure with the PDB entry 1HZH (Saphire et al., 2001) was used as Fc-template for IgG1 mAbs and bsAbs. Due to the recently discovered conformational diversity of IgG4, two different Fc-templates were used for the homology modeling of each IgG4 mAb (Blech et al., 2019). The PDB entry 5DK3 (Scapin et al., 2015) was used as Y-shaped template, while 6GFE (Blech et al., 2019) was used as template for the λ -shaped IgG4 conformation. Figure 1 shows exemplary homology models for each antibody format, including the two conformations of IgG4 mAbs. Further, two different bispecific

antibody formats were included in the data set, one knob-in-hole format ($n = 2$) with two different Fabs, and one appended IgG(H)-scFv format ($n = 3$). For the variable regions, separate templates for light and heavy chains were selected based on sequence identity of framework regions. All structures were prepared following the method developed by Zhu and Day (2013) before descriptor calculation. Structure preparation included the assignment of polar hydrogen positions, protonation states, and energy minimization using the OPLS3e force field (Roos et al., 2019). The movement of heavy atoms during energy minimization was limited to a maximum root mean square deviation of 0.3 Å between the initial homology model and the energy-minimized structure. As shown in Table 1, the quality of the energy-minimized antibody templates was examined with the MolProbity score (Chen et al., 2010). MolProbity allowed to test for the self-consistency of the final structure templates (Zhu & Day, 2013). The MolProbity score combines multiple quality parameters into a single metric, including a clashscore based on an all atom contact analysis, favored side-chain rotamers, and Ramachandran favored dihedral angles. To achieve a resolution-specific score, the final MolProbity score was compared to a cohort of PDB structures within a resolution range of 0.25 Å. All template structures showed high resolution-specific MolProbity scores (98%, 93%, and 86%) when compared to other structures in their resolution range.

2.2 | Protein descriptors

Maestro BioLuminate 3.7 was used for the calculation of antibody-specific protein descriptors. The recent publication of Sankar et al. (2022) gives detailed information on all protein descriptors that were used in this study. Scalar descriptors were calculated for the amino acid sequences or the three-dimensional protein structures. These descriptors considered a variety of physico-chemical and genetic properties of amino acids, including molecular weight, bulkiness, solvent exposure, flexibility, hydrophobicity, polarity, secondary structure propensity, amino acid composition, transmembrane tendency and more. The sequence-based descriptors were derived from parameterized empirical models for large data sets of proteins. Structure-based descriptors were calculated for the three-dimensional homology models providing scalar values for (1) geometrical magnitudes: solvent accessible surface area (SASA); (2) electrostatic properties: net charge, pI, estimates for the zeta

TABLE 1 Quality assessment of the template structures used for homology modeling of full-length antibodies

Molecule format	PDB	Resolution (Å)	R -value	MolProbity score
IgG1 mAb and bsAb formats	1HZH	2.70	0.229	2.01 (98)
IgG4 Y-conformation	5DK3	2.28	0.184	1.92 (93)
IgG4 λ -conformation	6GFE	1.80	0.181	1.78 (86)

Note: Resolutions and R -values were obtained from the PDB entries for 1HZH (Saphire et al., 2001), 5DK3 (Scapin et al., 2015), and 6GFE (Blech et al., 2019). MolProbity scores (Chen et al., 2010) represent the energy-minimized template structures that were used for homology modeling of the 21 full-length molecules.

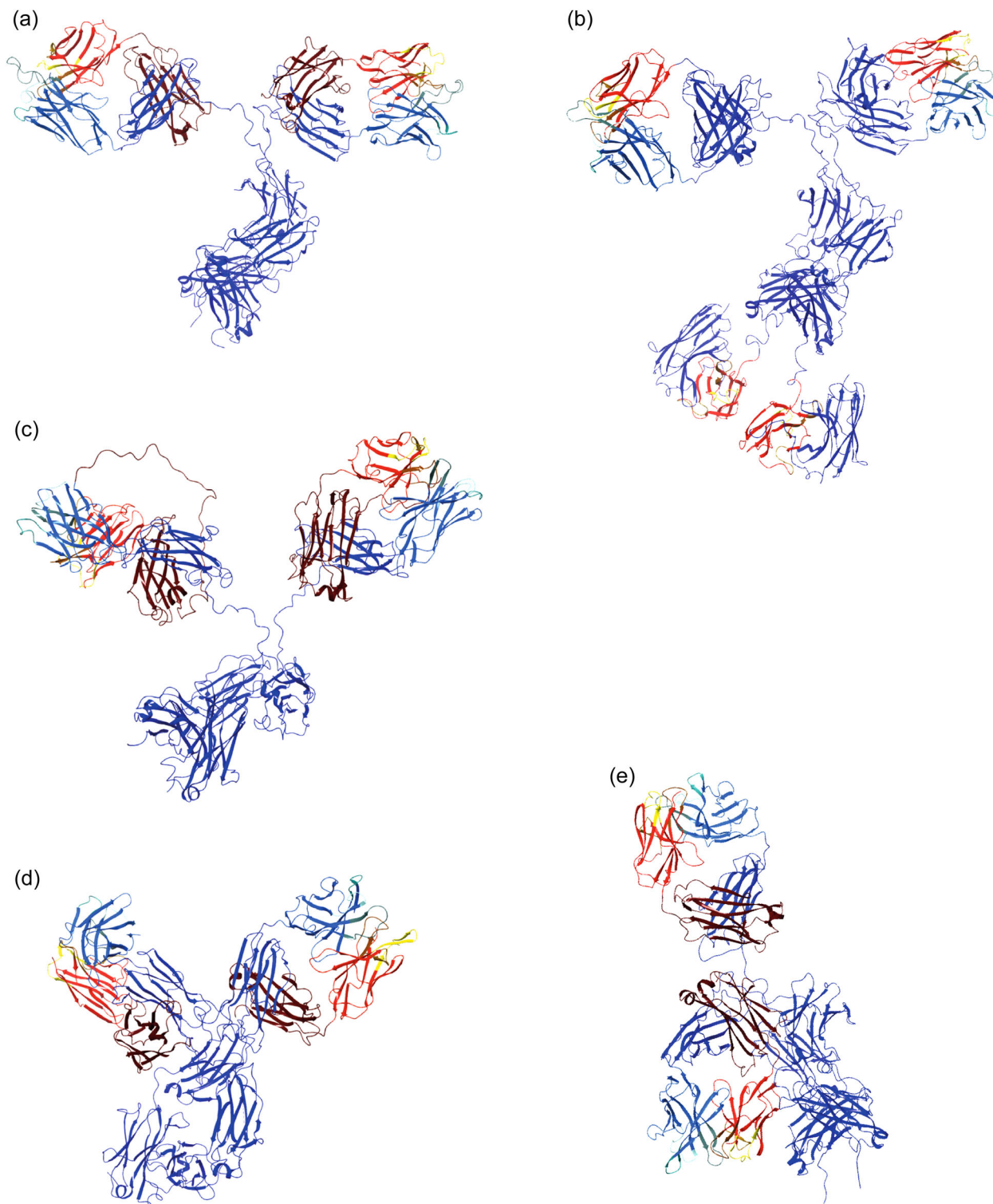


FIGURE 1 Template structures for homology modeling of full-length IgG1, IgG4 mAbs, and bsAbs. (a) displays the IgG1 mAb template, a modified version of the PDB entry 1HZH (Saphire et al., 2001). 1HZH was also used as source structure for both bsAbs formats, the appended IgG(H)-scFv (b), and the knob-in-hole format (c). Two discrete conformations were considered for the homology modeling of IgG4 mAbs. (d) IgG4 Y-conformation (PDB = 5DK3) (Scapin et al., 2015). (e) IgG4 λ -conformation (PDB = 6GFE) (Blech et al., 2019).

potential; (3) bonding information: number of hydrogen bonds; (4) dynamic properties: radius of gyration, hydrodynamic radius, a moment of inertia tensor. In addition to scalar descriptors, hydrophobic and electrostatic surface patch descriptors were

calculated at pH conditions of the respective chromatography experiments using the OPLS3 forcefield (Roos et al., 2019). Multiple patch-specific descriptors were derived that provide values for the intensity, size, and number of hydrophobic as well as electrostatic

patches. Further, the feature vector included aggregation propensity descriptors, such as AggScore, Aggrescan, and Zyggregator (Sankar et al., 2018). The final descriptor set consisted of 166 parameters, including 135 global and 31 local descriptors per region. Protein descriptors calculated for the entire antibody structure were defined as global descriptors. Local descriptors included the patch-specific properties and were defined for 30 individual regions of the antibody. The region-specific descriptors were calculated for the framework regions (FRs), FR1, FR2, FR3, and FR4, the complementarity-determining regions (CDRs) of the heavy (H1, H2, and H3) and light chain (L1, L2, and L3), for the entire CDR and Fv region, as well as for the hinge and constant region (CL, CH1, CH2, and CH3). The 135 global descriptors plus 31 local descriptors for 30 individual mAb regions resulted in a vector containing 1065 features per full-length mAb (Sankar et al., 2018, 2022). While local descriptors give additional insights in the binding orientation and adsorption mechanism of mAbs on CEX ligands, the high dimensionality of the feature vector increases the risk of over-fitting. Thus, QSPR models that utilize the descriptor set of Sankar et al. (2022) must incorporate a rigorous feature elimination methodology.

2.3 | Process conditions

All molecules were captured via Protein A affinity chromatography and further polished using the strong CEX resin POROS XS (Thermo Fisher Scientific). Column characterization experiments were conducted as triplicates. Column dimensions and column-specific model parameters are given in the Appendix (Supporting Information: Table S1). The chemicals used in this study were of pharmaceutical grade. All buffers were prepared with deionized water and filtered with a 0.2 μm sterile filter. Preparative CEX experiments were performed on an ÄKTA Avant 25 controlled using Unicorn 7 (both Cytiva).

An identical set of preparative CEX experiments in bind-elute mode was performed for all 21 molecules and corresponding Fabs of IgG1/IgG4 mAbs. Linear salt gradient elutions (LGE) were conducted at gradient lengths ranging from 10 CV to 30 CV, at three different mobile phase pH values, pH 5.00, pH 5.25, and pH 5.50. The counterion concentration increased from 50 to 500 mM sodium during gradient elution. The LGEs were conducted in the linear region of the adsorption isotherm, at a loading density of $1\text{g}/L_{\text{Resin}}$, allowing for the calculation of the characteristic charge v_i and the equilibrium constant $k_{eq,i}$ at varying pH conditions using the Yamamoto method (Yamamoto et al., 1983). So 1 and 0.1 M sodium hydroxide were used for column regeneration and storage, respectively.

2.4 | Mechanistic chromatography modeling

The transport dispersive model in Equation (1) was selected as the column model (Guiochon et al., 2006). Here, the change of the concentration $c_i(x, t)$ in the mobile phase depends on convective

mass transport in the interstitial volume of the packed bed with the superficial velocity u . Peak broadening effects are described by axial dispersion D_{ax} and interfacial mass transfer between the interstitial volume and the particle phase. The effective mass transfer parameter $k_{eff,i}$ lumps film diffusion effects in the particle boundary layer and pore diffusion in the particle phase. Equation (2) describes the accumulation of component i in the pore volume $c_{p,i}$. Danckwerts' boundary conditions are given in Equations (3) and (4).

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

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

$$\frac{\partial c_i}{\partial x}(0, t) = \frac{u(t)}{\epsilon_{col} D_{ax}} (c_i(0, t) - c_{in,i}(t)) \quad (3)$$

$$\frac{\partial c_i}{\partial x}(L, t) = 0 \quad (4)$$

Protein adsorption was modeled using the SDM (Boardman & Partridge, 1955; Velayudhan & Horváth, 1988). Equation (5) shows the kinetic form of the SDM isotherm, where q_i and $c_{p,i}$ denote the protein concentration in the solid and liquid phase of the particle, respectively. The SDM describes the equilibrium binding behavior of the protein considering the salt concentration in the pore phase c_s , the ionic capacity of the resin Λ , and protein-specific parameters. The protein characteristic charge v_i accounts for the number of charges interacting with the resin, while the constants $k_{eq,i} = k_{ads,i}/k_{des,i}$ and $k_{kin,i} = 1/k_{des,i}$ comprise adsorption and desorption rates of the modeled proteins.

$$k_{kin,i} \frac{\partial q_i}{\partial t} = k_{eq,i} q_{salt}^{v_i} c_{p,i} - q_i c_s^{v_i}, \quad (5)$$

$$q_{salt} = \Lambda - \sum_{j=1}^k v_j q_j. \quad (6)$$

The aim of the multiscale modeling method was the prediction of a mAb's SDM parameters, characteristic charge v_i and the equilibrium constant $k_{eq,i}$, which are the model parameters defining the retention volume during linear salt gradient elution. To generate training and testing data, the Yamamoto method was used for the analytical solution of v_i and $k_{eq,i}$ using a set of LGEs at differing salt gradient slopes for all therapeutic antibodies. Equation (7) describes the linear correlation between the normalized gradient slope GH and the elution salt concentration $c_{s,i}$ of component i at diluted loading conditions (Yamamoto et al., 1988, 1983). Equations (8) and (9) allow for the calculation of the normalized gradient slope GH , where $c_{s,initial}$ is the salt concentration at the gradient begin, $c_{s,final}$ is the salt concentration at the gradient end, and V_G is the gradient length in mL.

$$\log(GH) = (v_i + 1) \log(c_{s,i}) - \log(k_{eq,i} \Lambda^{v_i} (v_i + 1)), \quad (7)$$

$$g = \frac{C_{s,final} - C_{s,initial}}{V_G}, \quad (8)$$

$$GH = g(V_{col} - \epsilon_t V_{col}). \quad (9)$$

2.5 | QSPR modeling

The QSPR is the central building block of the multiscale model and connects molecular-level protein descriptors x with adsorption isotherm parameters y . The QSPR modeling workflow consisted of feature selection, machine learning via Gaussian process regression (GPR), and model validation. Machine learning and visualization was performed with Python 3.8.10 including scikit-learn (Pedregosa et al., 2011). Before feature selection, the data set was split into training and test data using two different sampling techniques. In the first section of this manuscript, 20% of the overall data was randomly removed from the data set for model validation. For external validation in the second part of the manuscript, all isotherm parameters and descriptors associated with a single mAb at different process conditions, fragmentation states, and conformations were removed from the training data set. The initial feature selection is crucial for QSPR modeling, when considering the high dimensionality of the feature vector ($n = 1065$) compared to the relative small number of measured adsorption isotherm parameters ($n = 94$). To avoid over-fitting, two consecutive steps of feature elimination were performed that identified descriptors x relevant for prediction of isotherm parameters y . First, a filter method was employed to reduce the computational costs of the subsequent recursive feature elimination (RFE). The filter method removed descriptors based on low variance and low mutual information (Kraskov et al., 2004; Ross, 2014) between individual descriptors x and isotherm parameters y of the training data set. This simple filter method reduced the feature vector by 97%, from 1065 to approximately 30 descriptors per isotherm parameter y . The dimensionality of the remaining descriptor set was further reduced using RFE as supervised feature selection method (Kumar & Minz, 2014). RFE methods evaluate multiple regression models by consecutively removing descriptors to find an optimal number and combination of descriptors that maximize the predictive power of the model. Therefore, RFE can identify the combinatorial effects of multiple descriptors on a single isotherm parameter. The consideration of combinatorial effects is critical for the prediction of adsorption behavior in CEX chromatography since the final GPR model is a product of multiple regional and global descriptors that define binding orientation, protein adsorption, and isotherm parameters. During each iteration of the RFE, 30 GPR models were trained using a fivefold cross-validation with six repetitions using a subset of 80% of the training data. After each iteration, the feature with the weakest contribution to the GPR model was removed from the descriptor vector x (Breiman, 2001). The optimal number of features used to train the final GPR model was determined via a scoring function considering the mean absolute errors (MAEs) calculated during cross-validation and the log-marginal-likelihood (LML) (Akaike, 1974) obtained from corresponding GPR

models. After filter-based and RFE, the GPR models used for external validation contained less than nine descriptors.

The general aim of the machine learning method is the prediction of an isotherm parameter y based on a high-dimensional vector x of protein descriptors. Here, the main challenge is the potential error propagation originating from homology modeling, wet-lab experiments, and the determination of mechanistic model parameters. The Bayesian approach of Gaussian processes provides estimates on the probability distribution for model predictions based on the underlying training data (Obrezanova et al., 2007; Rasmussen, 2006). Therefore, the GPR allows the visualization of heteroscedastic confidence intervals for model predictions that could be considered in the next step of multiscale modeling, for example, in mechanistic model simulations. A Gaussian process can be defined as generalization of a Gaussian distribution over a vector space to a function space of infinite dimensions (MacKay, 2003). The GPR relies on Bayesian inference, assuming a prior probability distribution for the values of the function $y(x)$ and updates the probability distribution in the presence of observed data to yield a *posterior* probability distribution. Following Obrezanova et al. (2007), the Bayesian update rule is

$$P(y(x)|\mathcal{D}) \propto P(Y|y(x), X)P(y(x)), \quad (10)$$

where $\mathcal{D} = \{X, Y\}$ is the training data, $P(y(x)|\mathcal{D})$ describes the *posterior* distribution, $P(y(x))$ is the *prior*, and $P(Y|y(x), X)$ is the likelihood. The *prior* was specified by addition of three subkernels: a linear term; a nonlinear Matérn class kernel; and a white noise kernel (Pedregosa et al., 2011; Rasmussen, 2006). The GPR was fitted by optimizing the hyperparameters Θ of the kernel functions to maximize the LML with the L-BFGS-B algorithm (Zhu et al., 1997).

The quality of the multiscale model prediction was evaluated by visual comparison of predicted and measured chromatograms. Additionally, QSPR model performance was assessed with the root mean-squared error (RMSE) shown in Equation (11). Chai and Draxler (2014) showed that RMSE is appropriate to represent model quality when the error distribution is expected to be Gaussian. The RMSE was normalized (NRMSE), as shown in Equation (12), which allowed for improved comparability of the individual QSPR models for $k_{eq,i}$ and v_i .

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}}, \quad (11)$$

$$\text{NRMSE} = \frac{\text{RMSE}}{Y_{\max} - Y_{\min}} 100\%. \quad (12)$$

3 | RESULTS AND DISCUSSION

In this manuscript, the multiscale model depicted in Figure 2 is used for the a priori prediction of elution profiles of mAbs in preparative CEX chromatography. SDM isotherm parameters were determined based on experimental data of 21 therapeutic antibodies, including IgG1 mAbs, IgG4 mAbs, Fabs, and bsAbs. Protein descriptors were derived from the corresponding antibody homology models. For

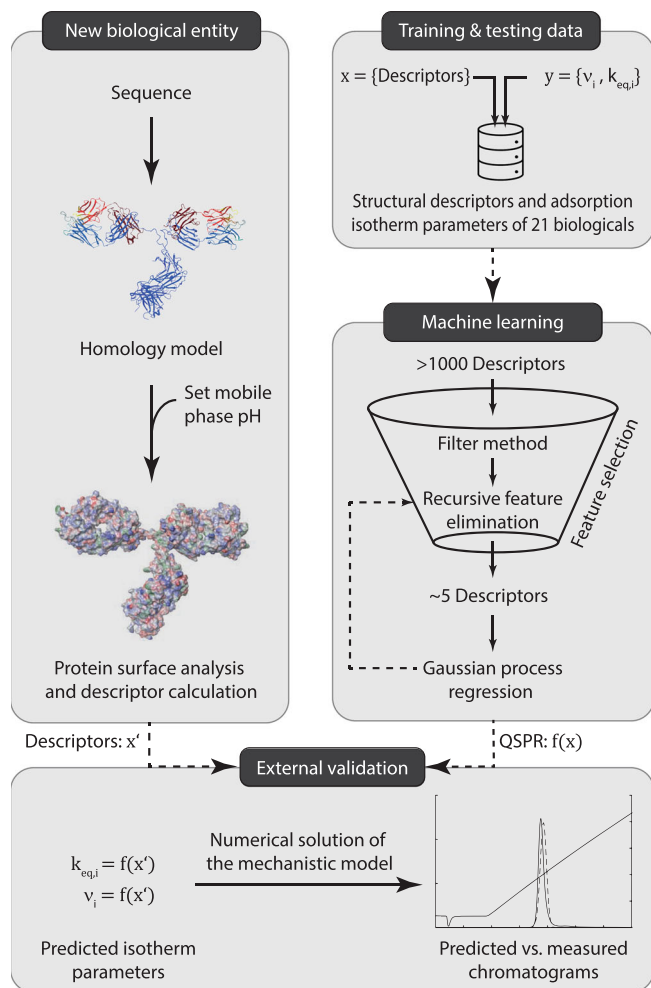


FIGURE 2 Multiscale modeling method for predicting cation exchange chromatography experiments for therapeutic antibody purification.

external validation, data of multiple antibodies was individually removed from the QSPR training data set. The predicted SDM isotherm parameters were used to simulate entire chromatograms and validated with experimental chromatograms.

3.1 | QSPR modeling for the prediction of SDM isotherm parameters

For the training and testing data set, SDM model parameters were determined based on CEX experiments at bench-scale. Purification experiments for all molecules were performed on the identical chromatography column with the strong CEX media POROS XS and an inner diameter of 1 cm, at a linear flow rate of 200 cm/h. Supporting Information: Appendix Table S1 shows experimentally determined column-specific parameters necessary for mechanistic modeling. The measured column porosities, ionic capacity, and axial dispersion coefficient were in accordance with values found in the literature (Hahn et al., 2016; Rodrigues, 1997).

The thermodynamic equilibrium constant $k_{eq,i}$ and characteristic charge v_i of 21 mAbs were determined experimentally using the Yamamoto method described in Equations (7)–(9). The Yamamoto method is based on a set of LGE experiments under diluted loading conditions of 1g/L_{Resin}. IgG1 and IgG4 mAbs were further digested with Papain to increase the structural diversity of the training data set. Thus, for the 10 IgG1 and the five IgG4 mAbs, SDM isotherm parameters were determined for full-length molecules and corresponding Fabs at pH 5.00, pH 5.25, and pH 5.50, respectively. For five bsAbs and the single Fab, SDM isotherm parameters were only determined for full-length molecules at pH 5.25. The resulting distributions of measured parameters in the training and testing data are shown in Figure 3. For reasons of clarity, values of the equilibrium constant are shown as their natural logarithm $\ln(k_{eq,i})$. When considering the differing pH conditions, fragmentation states, and conformational states of IgG4, the overall data set consisted of 94 pairs of isotherm parameters. Equilibrium constants $\ln(k_{eq,i})$ ranged from -6.34 to 3.23 . Characteristic charge parameters v_i ranged from 3.8 for a single Fab at pH 5.50 to 13.9 for a bispecific mAb of the IgG (H)-scFv format at pH 5.25. All mAbs had pI values above the

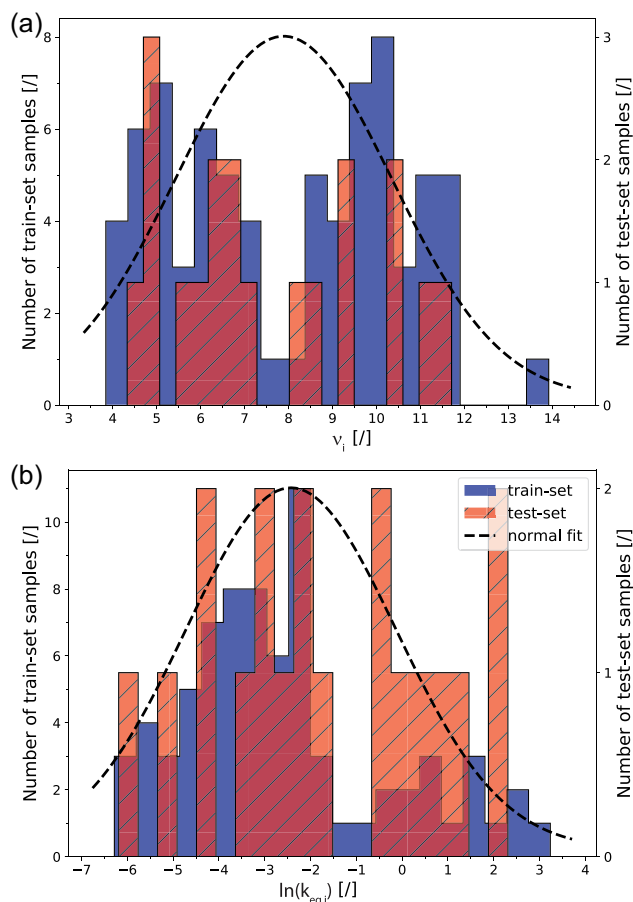


FIGURE 3 Distributions of SDM parameters in the test and training data set. Characteristic charge v_i (a) and equilibrium constant $\ln(k_{eq,i})$ (b) were determined via the Yamamoto method for 21 IgG1 mAbs, IgG4 mAbs, Fabs, and bsAbs at mobile phase pH conditions of pH 5.00, pH 5.25, and pH 5.50.

investigated pH conditions of pH 5.00 to pH 5.50. Consequently, the data in Figure 3 suggests a relationship of the characteristic charge ν_i with the molecular weight and net charge of investigated antibodies. Hunt et al. (2017) reported characteristic charge parameters ranging between 5 and 20 for IgGs on two CEX resins, a strong (SO_3^-) and weak (COO^-) CEX ligand. Binding orientation and isotherm parameters of mAbs depend on the CEX ligand (Müller-Späth et al., 2008). Therefore, the applicability domain of our QSPR model is limited to the strong CEX ligand SO_3^- . Figure 4 shows that $k_{eq,i}$ and ν_i parameters of the investigated molecules increased with decreasing mobile phase pH, which resulted from the increased positive net charge of proteins at low pH conditions (Schmidt et al., 2014). For full-length IgG1 and IgG4 mAbs investigated in this study, characteristic charge parameters had values between 6.2 and 11.79. Further, characteristic charge ν_i parameters of Fabs were reduced by approximately 40% compared to corresponding full-length molecules. Consequently, the bimodal distribution observable in Figure 3a mainly results from fragmentation of IgG1 and IgG4 mAbs. In the formulation of the SDM isotherm (Boardman & Partridge, 1955), the characteristic charge parameter ν_i depicts multipoint binding of a protein to the charged stationary phase. The relationship between the molecular weight and ν_i suggests that the overall protein size is crucial for the stoichiometric equilibrium of the antibody adsorption process.

Noteworthy, all IgG4 molecules showed a distinct split-peak elution behavior similar to the results reported by Luo et al. (2015).

An exemplary split-peak chromatogram of one IgG4 investigated in this study is shown in Figure 6b. When collecting one of the peaks and repeating the identical experiment, the split-peak elution behavior was observed again. Double elution peaks were only observed for full-length IgG4 mAbs and not visible for corresponding Fabs. To understand underlying adsorption phenomena, the Yamamoto method was applied separately to both elution peaks of full-length mAbs. This resulted in two pairs of equilibrium constant $k_{eq,i}$ and characteristic charge ν_i for each IgG4 molecule and mobile phase pH. Both parameters, $k_{eq,i}$ and ν_i , affect retention times in gradient elution experiments. Interestingly, $k_{eq,i}$ values were comparable for both peaks of each IgG4 molecule. In contrast, characteristic charge parameters ν_i of early eluting species were 8%–23% lower compared to the second peak, and thus caused the shift in retention volume. A recent study performed by Blech et al. (2019) revealed that IgG4 mAbs can adopt multiple conformational states coexisting in a dynamic equilibrium, including the typical Y-shaped conformation and a self-associated λ -conformation where one Fab is directed toward the Fc-portion of the mAb. Further, Blech et al. (2019) reported the X-ray crystal structure of the intact, full-length IgG4 mAb in its λ -conformation (PDB = 6GFE) (Blech et al., 2019). Based on the two published structure templates 5DK3 and 6GFE, it was possible to consider the conformational diversity of IgG4 mAbs during homology modeling and the subsequent descriptor calculation. Exemplary homology models of a IgG4 in Y- and λ -conformations are

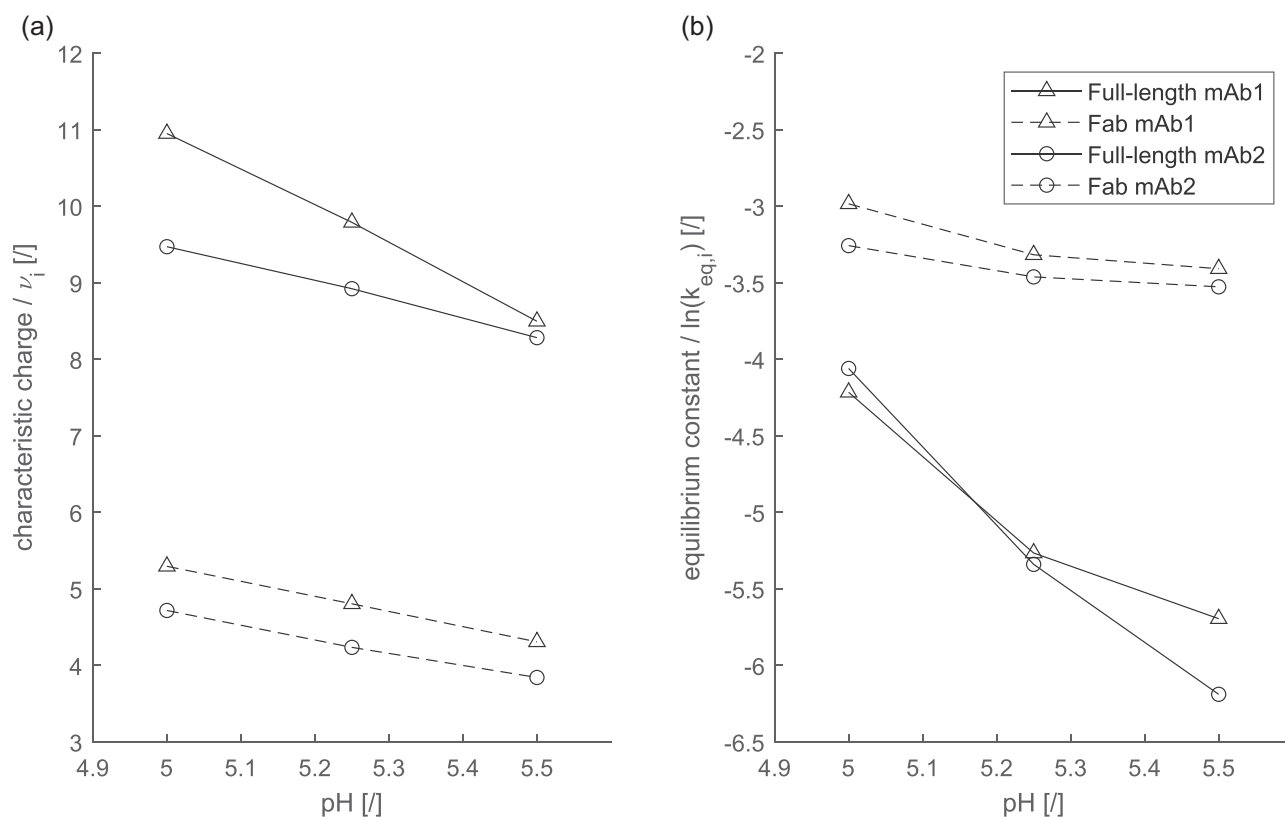


FIGURE 4 pH-dependencies of the characteristic charge ν_i (a) and equilibrium constant $k_{eq,i}$ (b) of two full-length IgG1 mAbs and corresponding Fabs on the strong CEX resin POROS XS.

depicted in Figure 1d,e, respectively. Global and local descriptors were derived from homology models of the 21 full-length antibodies and used as independent variables x in the subsequent QSPR modeling workflow. Global descriptors, such as the surface net charge, of identical IgG4 mAbs in different conformations, changed significantly due to the reduced SASA of the self-associated λ -conformation compared to the open Y-conformation. In contrast, local descriptors associated with the variable region of individual IgG4 mAbs were not affected by conformational diversity. Due to the lower protein net charge of λ -conformations and lower measured v_i parameters of the early eluting peak, we hypothesize that the split-peak phenomenon is caused by the reversible, conformational diversity of IgG4 mAbs as reported by Blech et al. (2019). Thus, the early eluting peak represents the IgG4 λ -conformation, and the second peak is the Y-conformation. To build a predictive QSPR model, SDM parameters y of the first peak were combined with protein descriptors x of the λ -conformation and SDM isotherm parameters of the second peak were combined with descriptors of the Y-conformation. The hypothesis was tested in the second result section, by external validation including the prediction of an IgG4 split-peak chromatogram.

The QSPR models depicted in Figure 5 represent the central building block of the multiscale modeling method. GPR was selected as a machine learning algorithm that maps between the descriptor vector x and the adsorption isotherm parameters y . The two-staged feature elimination aimed to avoid over-determination of the GPR model. Additionally, the results of the feature elimination in Figure 5a,b give deeper insights into binding orientation and dependencies of SDM isotherm parameters on protein structure descriptors. The initial filtering method removed descriptors from the normalized feature vector x showing <0.01 variance across the training data set. Subsequently, assessment of the mutual information between descriptors x and isotherm parameters y in the training set resulted in the elimination of 97% of the 1065 descriptors. Global descriptors, such as the net charge, the hydrodynamic radius, and the total aromatic SASA, shared the highest mutual information with the characteristic charge parameter v_i . The thermodynamic equilibrium constant $k_{eq,i}$ shared the highest mutual information with global as well as local, charge and aggregation propensity estimating descriptors associated with the FR and the CDRs. A total of 15–30 descriptors remained for the second level of feature selection based on RFE. Figure 5a,b depicts the protein structural descriptors selected during supervised feature elimination. During each iteration of RFE, a multivariate GPR was trained, cross-validated, and the feature with the lowest permutation importance was removed until the optimal number of features was reached. A total of $n=5$ descriptors was determined as optimal number of features to train the final GPR models for characteristic charge and equilibrium constant. When comparing the changes of MAE during cross-validation of both SDM isotherm parameters in Figure 5a,b, the model prediction for the equilibrium constant is less accurate compared to the characteristic charge. Further, Figure 5b indicates

that the predictive power of the GPR model for $k_{eq,i}$ significantly decreases when removing descriptors estimating the sum positive surface patch energy in the heavy-chain CDRs and the FR atomic contact energy descriptor. Based on the selected descriptors, it can be assumed that the variable region of mAbs is involved in protein adsorption and “Fab-first” is the preferred binding orientation on strong CEX media.

The final predictions of the GPR QSPR models with a 20% randomly selected test set are plotted in Figure 5c,d. Correlation coefficients of $R^2 > 0.99$ as well as the uniformly dispersed residuals of training and testing data, indicate an adequate accuracy of the model predictions for both SDM isotherm parameters. A reason for the comparably high predictive power of the QSPR model could be the structural connotation of the parameters defined in the SDM isotherm. Characteristic charge describes the average number of charged groups on the protein surface interacting with the chromatography ligands. This is a potential explanation for the clear correlation between global charge descriptors and the v_i parameter. The comparably wide 95% confidence intervals (CIs) in the testing set of $k_{eq,i}$ suggest a higher prediction uncertainty of the equilibrium constant compared to the characteristic charge parameter. Further, the $k_{eq,i}$ test predictions with the furthest distance to the ideal prediction line also showed the widest 95% CIs. This indicates that the GPR model could successfully estimate the *posterior* probability distribution of $y(x)$ based on the underlying data that was used for model training. The higher model uncertainty of $k_{eq,i}$ compared to v_i could be a consequence of inaccurate predictions of CDR loop conformations during homology modeling and subsequent calculation of local descriptors. The QSPR model for the characteristic charge parameter was exclusively trained with global protein descriptors, which are less dependent on the challenging prediction of hyper-flexible CDR loop structures. Recent advances in protein structure prediction based on deep neural networks (Senior et al., 2020) could possibly improve homology models for mAb CDRs and the resulting multiscale model predictions. Nonetheless, the predictive power of the QSPR models in Figure 5 is sufficient for simulation of elution profiles as shown during external model validation. The ability to calculate probability distributions on model predictions is the main benefit of Gaussian process regression as a machine learning method for QSPR modeling. This is especially relevant for multiscale modeling tasks as presented in this study, where the propagation of model uncertainty from homology to mechanistic modeling could affect prediction accuracy. Feature selection and QSPR modeling revealed quantitative relationships between the structure of therapeutic antibodies and their adsorption model parameters. These relationships were also suggested by a previous work of our group, where single amino acid substitution in the CDR of mAbs had a significant impact on $k_{eq,i}$ parameters (Saleh et al., 2021). The three IgG1 mAbs with single amino acid substitution were also included in the present data set of 21 mAbs, increasing the predictive power of the QSPR model for $k_{eq,i}$. It is important to consider that the random sampling method applied in Figure 5 potentially included identical molecules in training and testing data sets, but at differing process conditions or

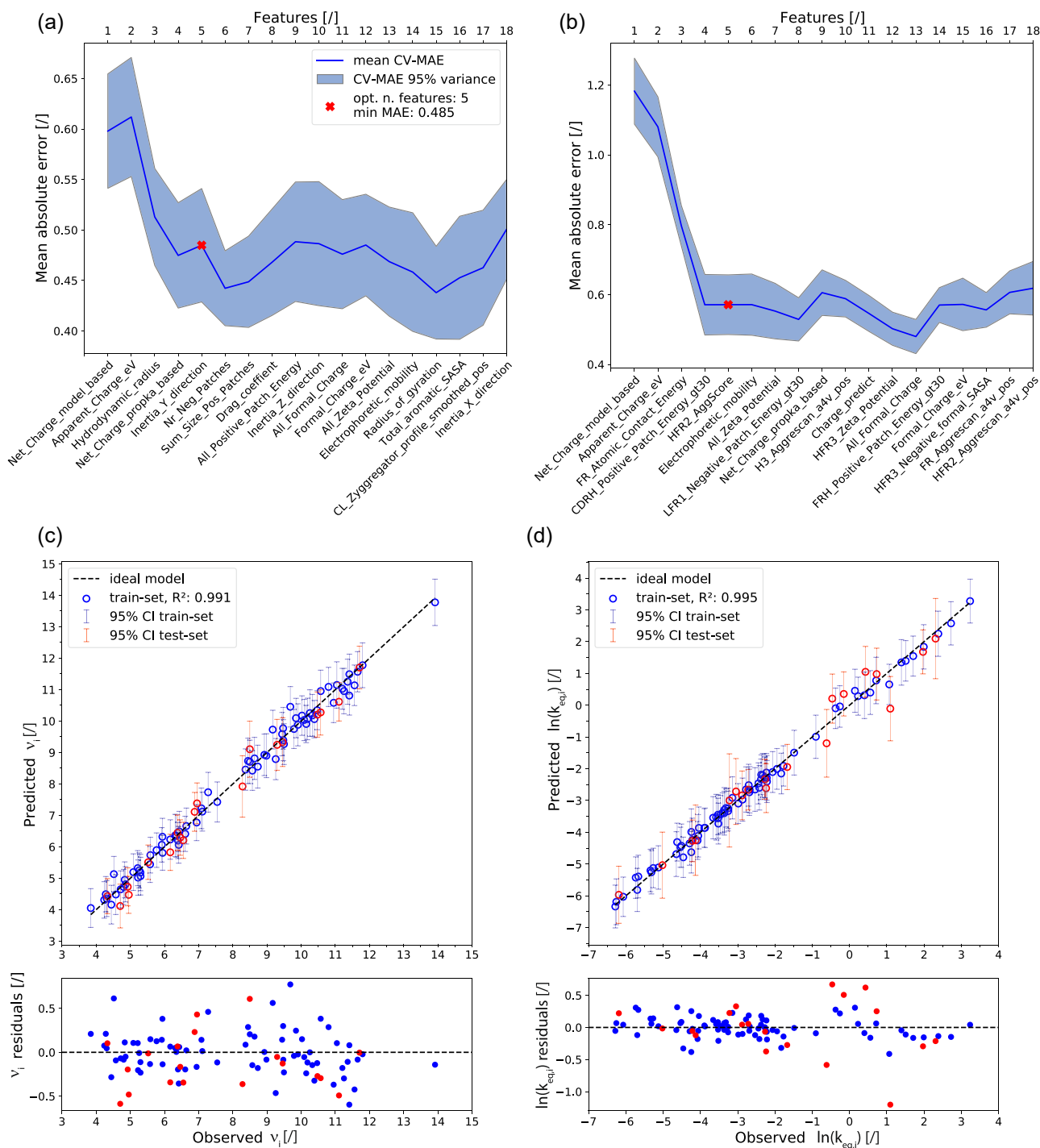


FIGURE 5 Recursive feature elimination and QSPR model for prediction of SDM parameters of mAbs on POROS XS at pH 5.00, 5.25, and 5.50. (a) and (b) show the results of feature selection for characteristic charge v_i and equilibrium constant $k_{eq,i}$, respectively. The corresponding QSPR models for the prediction of v_i and $k_{eq,i}$ of a randomly selected test set representing 20% of the overall data are given in (c) and (d).

fragmentation states. Therefore, the trained GPR model in Figure 5 can be used to simulate process behavior of a mAb at an unknown mobile phase pH, if an initial experiment at a different pH value was already performed. The following chapter explores the possibility to predict entire chromatograms of a new antibody candidate without conducting initial experiments.

3.2 | External validation and prediction of chromatograms

For external validation, all data points associated with a specific mAb product must be removed from the training and cross-validation data set. Consequently, for each of the 16 antibodies

shown in Table 2, all data points at varying pH values, fragmentation states, and conformations were removed from the training data set. The antibodies selected for model validation covered all molecule formats, including the Fab, IgG1, and IgG4 mAbs, as well as both bispecific antibody formats. Specific outliers were excluded from the validation procedure to avoid extrapolation. For the prediction of standard, full-length IgG1, and IgG4 mAbs, NRMSEs were located between 4.0% and 15.0%. $\ln(k_{eq,i})$ and v_i of the single Fab (13) were predicted with NRMSEs of 1.8% and 7.8%, respectively. Due to the fragmentation of full-length IgG1 and IgG4, the training data set also contained experimental information on 15 additional Fabs, which could explain the accurate prediction of Fab (13) isotherm parameters. The prediction error increased for the bispecific formats, as less data was available to train the regression model compared to the IgG1 and IgG4 mAbs and Fabs. The comparably high NRMSE of 33.6% for the characteristic charge of the appended IgG(H)-scFv (15) shows the limitation of the QSPR model. Here, the regression-based QSPR model was used to extrapolate in a numerical range not covered by the training data set. Overall, the data in Table 2 shows that the QSPR model enables predictions for a variety of antibody formats with differing physicochemical properties and elution behavior. However, more data should be added for the bispecific formats to improve the predictive power of the model and to increase the understanding of the adsorption mechanism and binding orientation.

TABLE 2 QSPR predictions of adsorption isotherm parameters of full-length antibodies at pH 5.25

Format (#)	$\ln(k_{eq,i})$			v_i		
	Measured	Predicted	NRMSE (%) ^a	Measured	Predicted	NRMSE (%) ^a
IgG1 (1)	-5.3	-5.3	4.4	8.9	8.8	4.2
IgG1 (2)	-5.3	-5.2	4.8	9.8	10.1	3.9
IgG1 (3)	-3.9	-3.1	8.3	9.8	10.3	3.5
IgG1 (4)	-4.2	-3.7	4.0	10.1	10.4	1.4
IgG1 (5)	-3.4	-3.5	4.3	10.4	10.1	3.8
IgG1 (6)	-0.6	-1.6	15.0	8.5	9.6	11.6
IgG1 (7)	0.7	1.0	12.0	11.7	11.7	3.5
IgG4 (8)	-4.6	-5.5	13.6	6.2	7.9	6.2
IgG4 (9)	-4.5	-4.5	5.1	6.3	6.9	5.1
IgG4 (10)	-4.6	-4.9	6.2	9.5	9.3	7.6
IgG4 (11)	-4.5	-4.5	8.9	9.9	8.5	12.1
IgG4 (12)	-5.7	-5.5	7.0	9.0	8.7	8.3
Fab (13)	-3.6	-3.4	1.8	4.5	5.3	7.8
knob-in-hole (14)	-2.8	-1.8	9.8	11.3	9.0	22.0
IgG(H)-scFv (15)	0.7	0.3	4.2	13.9	10.5	33.6
IgG(H)-scFv (16)	-6.3	-5.2	11.5	10.3	10.0	2.2

Note: For each molecule, all model parameters at varying pH values, fragmentation states, and conformations were removed from the training data set.

^aThe NRMSE calculation is based on the prediction of the entire data set of the respective antibody.

The goal of the multiscale modeling workflow is the simulation of elution profiles. As a final step of the model validation, Figure 6 compares predicted with measured chromatograms for one IgG1 (2) and one IgG4 (12) molecule. The QSPR models for mAbs simulated in Figure 6 are presented in Supporting Information: Appendix Figure S1. Similar to the randomly sampled QSPR model, the QSPR models with individually selected testing data enabled an acceptable prediction accuracy for both isotherm parameters. However, wider CIs of equilibrium constants were observed for the external test set compared to the CIs in Figure 5. This indicates that predictions for unknown mAbs are more challenging compared to isotherm parameters of mAbs that were already investigated at differing pH conditions, conformations, or fragmentation states. For both mAbs in Figure 6, our multiscale model enabled predictions that were in good agreement with measured UV curves at 280 nm wavelength. Two protein species were considered for simulation of the IgG4 elution profile in Figure 6b. The first peak represents the self-associated IgG4 λ -conformation that showed a reduced theoretical surface charge compared to the Y-conformation that elutes in the second peak. The dynamic equilibrium between λ - and Y-conformation discovered by Blech et al. (2019) could explain why the split-peak phenomenon is reproducible when collecting the peaks separately and repeating the experiment (Luo et al., 2015, 2014). The multiscale model is trained with experimental data in the range of pH 5.00 to pH 5.50 for IgG4 mAbs. Therefore, the model considers the pH-dependency of the split-peak phenomenon within the

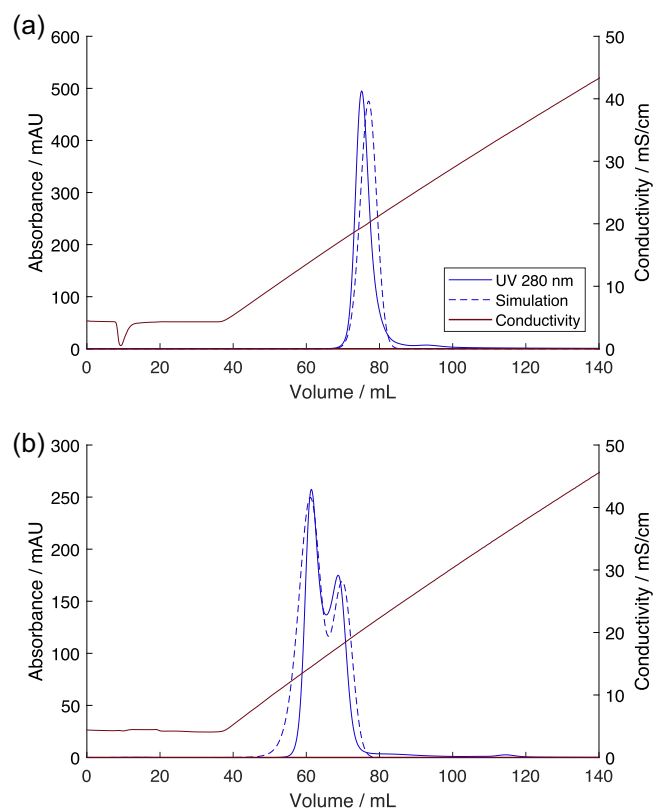


FIGURE 6 Measured and predicted chromatograms of IgG1 (a) and IgG4 (b) mAb on POROS XS at pH 5.25, 10 CV gradient slope, 200 cm/h linear flow rate. The corresponding SDM isotherm parameters for the model prediction are shown in Supporting Information: Appendix Figure S1. Two protein species representing the Y-conformation and the λ -conformation were considered to simulate the IgG4 split peak phenomenon.

investigated pH range. A stronger change of the pH conditions might shift the equilibrium between the conformational states towards Y- or λ -conformation and could therefore avoid split-peak elution. Due to the semi-mechanistic nature of the SDM isotherm, model predictions are not limited to mobile phase conditions applied in the wet-lab experiments shown in Figure 6. Hence, the model could be used to test varying gradient slopes, pH conditions, or step elution salt concentrations to define an initial process design space before protein material is available. In contrast to the mechanistic chromatography model, the GPR-based QSPR model is based on empirical relationships between protein descriptors x and isotherm parameters y . Therefore, extrapolation toward molecule formats and chromatography ligands not included in the training set is not feasible. Continuous improvement of QSPR predictions can be achieved by retraining the GPR model when a new biological entity enters the developmental phase. Ladiwala et al. (2005) introduced the first multiscale QSPR model enabling the prediction of chromatograms based on descriptors derived from published crystal structures of smaller model proteins. To the best of our knowledge, the present work is the first to predict isotherm parameters and chromatograms based on protein descriptors derived from mAb

homology models. This is crucial for applications in biopharmaceutical industry, where experimentally determined X-ray crystal structures of full-length mAbs are not available during process development. For mechanistic model calibration, v_i and $k_{eq,i}$ are typically estimated via multiple salt gradient experiments at low loading densities. Therefore, our multiscale model can substitute these time-consuming calibration experiments with sequence-based predictions. Currently, the QSPR model is limited to the linear region of the steric mass-action (SMA) isotherm. An extension of the QSPR model to the nonlinear region would demand additional wet-lab experiments at higher loading density for all investigated molecules. Alternatively, a nonlinear SMA isotherm model can be achieved by combining the presented multiscale model with inverse model calibration using a single experiment at higher loading density (Hahn et al., 2016). Additionally, fraction analysis with offline analytical measurements can be used to include charge- and size-variants of the mAb.

4 | CONCLUSION

The present study introduced a multiscale model leading from the amino acid sequence of a therapeutic antibody to the mechanistic model of its preparative CEX chromatography process. The model includes IgG1 mAbs, IgG4 mAbs, Fabs, and bispecific antibody formats at different pH conditions. Multiple molecules were individually removed from the training data set for external validation. The GPR-based QSPR model predicted the SDM isotherm parameters of the test set molecules with an accuracy that enabled the simulation of CEX chromatograms. Here, the multiscale model predicted complex split-peak elution curves of the IgG4 mAb by considering two discrete conformations coexisting in a dynamic equilibrium. The closed λ -conformation of the IgG4 mAb had a reduced positive surface charge and characteristic charge v_i compared to the Y-conformation leading to an early elution during gradient elution experiments. The two-staged feature selection method via regression-based filtering and RFE effectively avoided over-fitting. Further, the final protein descriptors selected for GPR modeling gave insights in the relationships between antibody structure and adsorption isotherm parameters. With the knowledge gained on antibody adsorption, mAb candidates can be selected that integrate into a standardized platform process. Most importantly, our multiscale model allows simulation of different process conditions and elution modes before protein material is available for wet-lab experiments. This enables *in silico* process optimization and robustness analysis to shorten the time from antibody drug discovery to start of clinical investigations. Based on sufficient data, the multiscale model can be extended to predict other properties relevant for biopharmaceutical development, for example, mAb charge heterogeneity, product stability, or aggregation propensity.

AUTHOR CONTRIBUTIONS

David Saleh, Federico Rischawy, Gang Wang, Simon Kluters, and Jürgen Hubbuch designed the research; David Saleh, Rudger Hess, and Michelle Ahlers-Hesse performed the research; David Saleh,

Rudger Hess, and Michelle Ahlers-Hesse analyzed the data; David Saleh wrote the paper; Thomas Schwab, Joey Studts, and Jürgen Hubbuch supervised the research.

ACKNOWLEDGMENT

The authors would like to acknowledge experimental and scientific support of their colleagues within the DSP, BTB, and PDB departments at Boehringer Ingelheim. Open Access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

Research data are not shared.

ORCID

David Saleh  <http://orcid.org/0000-0002-1212-3679>

REFERENCES

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716–723.
- Andrew, S. M., & Titus, J. A. (1997). Fragmentation of immunoglobulin g. *Current Protocols in Immunology*, 21(1), 2–8.
- Blech, M., Hörer, S., Kuhn, A. B., Kube, S., Göddeke, H., Kiefer, H., Zang, Y., Alber, Y., Kast, S. M., Westermann, M., Tully, M. D., Schäfer, L. V., & Garidel, P. (2019). Structure of a therapeutic full-length anti-nprag igg4 antibody: Dissecting conformational diversity. *Biophysical Journal*, 116(9), 1637–1649.
- Boardman, N., & Partridge, S. (1955). Separation of neutral proteins on ion-exchange resins. *Biochemical Journal*, 59(4), 543–552.
- Borg, N., Brodsky, Y., Moscariello, J., Vunnum, S., Vedantham, G., Westerberg, K., & Nilsson, B. (2014). Modeling and robust pooling design of a preparative cation-exchange chromatography step for purification of monoclonal antibody monomer from aggregates. *Journal of Chromatography A*, 1359, 170–181.
- Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5–32.
- Cardillo, A. G., Castellanos, M. M., Desailly, B., Dessoy, S., Mariti, M., Portela, R. M., Scutella, B., von Stosch, M., Tomba, E., & Varsakelis, C. (2021). Towards in silico process modeling for vaccines. *Trends in Biotechnology*, 39(11), 1120–1130.
- Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, 7(3), 1247–1250.
- Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., & Richardson, D. C. (2010). Molprobity: All-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D: Biological Crystallography*, 66(1), 12–21.
- Chung, W. K., Freed, A. S., Holstein, M. A., McCallum, S. A., & Cramer, S. M. (2010). Evaluation of protein adsorption and preferred binding regions in multimodal chromatography using nmr. *Proceedings of the National Academy of Sciences of the United States of America*, 107(39), 16811–16816.
- Guiochon, G., Felinger, A., & Shirazi, D. G. (2006). *Fundamentals of preparative and nonlinear chromatography*. Elsevier.
- Hahn, T., Baumann, P., Huuk, T., Heuveline, V., & Hubbuch, J. (2016). Uv absorption-based inverse modeling of protein chromatography. *Engineering in Life Sciences*, 16(2), 99–106. <https://doi.org/10.1002/elsc.201400247>
- Hahn, T., Huuk, T., Osberghaus, A., Doninger, K., Nath, S., Hepbildikler, S., Heuveline, V., & Hubbuch, J. (2016). Calibration-free inverse modeling of ion-exchange chromatography in industrial antibody purification. *Engineering in Life Sciences*, 16(2), 107–113. <https://doi.org/10.1002/elsc.201400248>
- Hunt, S., Larsen, T., & Todd, R. J. (2017). Modeling preparative cation exchange chromatography of monoclonal antibodies. In A. Staby & S. Ahuja (Eds.), *Preparative chromatography for separation of proteins*. John Wiley & Sons, Incorporated.
- Hutter, C., von Stosch, M., Bournazou, M. N. C., & Butté, A. (2021). Knowledge transfer across cell lines using hybrid Gaussian process models with entity embedding vectors. *Biotechnology and Bioengineering*, 118(11), 4389–4401.
- Ishihara, T., Kadoya, T., Yoshida, H., Tamada, T., & Yamamoto, S. (2005). Rational methods for predicting human monoclonal antibodies retention in protein affinity chromatography and cation exchange chromatography. *Journal of Chromatography A*, 1093(1), 126–138.
- Karlberg, M., de Souza, J. V., Fan, L., Kizhedath, A., Bronowska, A. K., & Glassey, J. (2020). Qsar implementation for hic retention time prediction of mabs using fab structure: A comparison between structural representations. *International Journal of Molecular Sciences*, 21(21), 8037.
- Karlberg, M., von Stosch, M., & Glassey, J. (2018). Exploiting mab structure characteristics for a directed qbd implementation in early process development. *Critical Reviews in Biotechnology*, 38(6), 957–970.
- Kelley, B. (2020). Developing therapeutic monoclonal antibodies at pandemic pace. *Nature Biotechnology*, 38(5), 540–545.
- Kittelmann, J., Lang, K. M., Ottens, M., & Hubbuch, J. (2017a). Orientation of monoclonal antibodies in ion-exchange chromatography: A predictive quantitative structure-activity relationship modeling approach. *Journal of Chromatography A*, 1510, 33–39.
- Kittelmann, J., Lang, K. M., Ottens, M., & Hubbuch, J. (2017b). An orientation sensitive approach in biomolecule interaction quantitative structure-activity relationship modeling and its application in ion-exchange chromatography. *Journal of Chromatography A*, 1482, 48–56.
- Kizhedath, A., Karlberg, M., & Glassey, J. (2019). Cross-interaction chromatography-based qsar model for early-stage screening to facilitate enhanced developability of monoclonal antibody therapeutics. *Biotechnology Journal*, 14(8), 1800696.
- Kraskov, A., Stögbauer, H., & Grassberger, P. (2004). Estimating mutual information. *Physical Review E*, 69(6), 066138.
- Kumar, V., & Minz, S. (2014). Feature selection: A literature review. *SmartCR*, 4(3), 211–229.
- Ladiwala, A., Rege, K., Breneman, C. M., & Cramer, S. M. (2005). A priori prediction of adsorption isotherm parameters and chromatographic behavior in ion-exchange systems. *Proceedings of the National Academy of Sciences of the United States of America*, 102(33), 11710–11715.
- Luo, H., Cao, M., Newell, K., Afdahl, C., Wang, J., Wang, W. K., & Li, Y. (2015). Double-peak elution profile of a monoclonal antibody in cation exchange chromatography is caused by histidine-protonation-based charge variants. *Journal of Chromatography A*, 1424, 92–101.
- Luo, H., Macapagal, N., Newell, K., Man, A., Parupudi, A., Li, Y., & Li, Y. (2014). Effects of salt-induced reversible self-association on the elution behavior of a monoclonal antibody in cation exchange chromatography. *Journal of Chromatography A*, 1362, 186–193.
- MacKay, D. J. (2003). *Information theory, inference and learning algorithms*. Cambridge University Press.
- Mazza, C., Sukumar, N., Breneman, C., & Cramer, S. (2001). Prediction of protein retention in ion-exchange systems using molecular descriptors obtained from crystal structure. *Analytical Chemistry*, 73(22), 5457–5461.
- Mollerup, J. M., Hansen, T. B., Kidal, S., & Staby, A. (2008). Quality by design—thermodynamic modelling of chromatographic separation of proteins. *Journal of Chromatography A*, 1177(2), 200–206.

- Morrison, C. (2020). Fresh from the biotech pipeline-2019. *Nature Biotechnology*, 38, 126–131.
- Müller-Spáth, T., Aumann, L., Melter, L., Ströhlein, G., & Morbidelli, M. (2008). Chromatographic separation of three monoclonal antibody variants using multicolumn countercurrent solvent gradient purification (MCSGP). *Biotechnology and Bioengineering*, 100(6), 1166–1177.
- Narayanan, H., Luna, M. F., von Stosch, M., Cruz Bournazou, M. N., Polotti, G., Morbidelli, M., Butté, A., & Sokolov, M. (2020). Bioprocessing in the digital age: The role of process models. *Biotechnology Journal*, 15(1), 1900172.
- Obrezanova, O., Csányi, G., Gola, J. M., & Segall, M. D. (2007). Gaussian processes: A method for automatic QSAR modeling of adme properties. *Journal of Chemical Information and Modeling*, 47(5), 1847–1857.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Pirung, S. M., van der Wielen, L. A., van Beckhoven, R. F., van de Sandt, E. J., Eppink, M. H., & Ottens, M. (2017). Optimization of biopharmaceutical downstream processes supported by mechanistic models and artificial neural networks. *Biotechnology Progress*, 33(3), 696–707.
- Rasmussen, C. (2006). CKI Williams Gaussian processes for machine learning.
- Rathore, A. S., & Winkle, H. (2009). Quality by design for biopharmaceuticals. *Nature Biotechnology*, 27(1), 26–34.
- Robinson, J., Roush, D., & Cramer, S. M. (2020). The effect of pH on antibody retention in multimodal cation exchange chromatographic systems. *Journal of Chromatography A*, 1617, 460838.
- Robinson, J. R., Karkov, H. S., Woo, J. A., Krogh, B. O., & Cramer, S. M. (2017). Qsar models for prediction of chromatographic behavior of homologous fab variants. *Biotechnology and Bioengineering*, 114(6), 1231–1240.
- Rodrigues, A. E. (1997). Permeable packings and perfusion chromatography in protein separation. *Journal of Chromatography B: Biomedical Sciences and Applications*, 699(1-2), 47–61.
- Roos, K., Wu, C., Damm, W., Reboul, M., Stevenson, J. M., Lu, C., Dahlgren, M. K., Mondal, S., Chen, W., Wang, L., Abel, R., Friesner, R. A., & Harder, E. D. (2019). Opls3e: Extending force field coverage for drug-like small molecules. *Journal of Chemical Theory and Computation*, 15(3), 1863–1874.
- Ross, B. C. (2014). Mutual information between discrete and continuous data sets. *PLoS One*, 9(2), e87357.
- Saleh, D., Hess, R., Ahlers-Hesse, M., Beckert, N., Schönberger, M., Rischawy, F., Wang, G., Bauer, J., Blech, M., Kluters, S., Studts, J., & Hubbuch, J. (2021). Modeling the impact of amino acid substitution in a monoclonal antibody on cation exchange chromatography. *Biotechnology and Bioengineering*, 118(8), 2923–2933.
- Sankar, K., Krystek Jr., S. R., Carl, S. M., Day, T., & Maier, J. K. (2018). Aggscore: Prediction of aggregation-prone regions in proteins based on the distribution of surface patches. *Proteins: Structure, Function, and Bioinformatics*, 86(11), 1147–1156.
- Sankar, K., Trainor, K., Blazer, L. L., Adams, J. J., Sidhu, S. S., Day, T., Meiering, E., & Maier, J. K. (2022). A descriptor set for quantitative structure-property relationship prediction in biologics. *Molecular Informatics*, 41(9), 2100240.
- Saphire, E. O., Parren, P. W., Pantophlet, R., Zwick, M. B., Morris, G. M., Rudd, P. M., Dwek, R. A., Stanfield, R. L., Burton, D. R., & Wilson, I. A. (2001). Crystal structure of a neutralizing human igg against hiv-1: A template for vaccine design. *Science*, 293(5532), 1155–1159.
- Scapin, G., Yang, X., Prorise, W. W., McCoy, M., Reichert, P., Johnston, J. M., Kashi, R. S., & Strickland, C. (2015). Structure of full-length human anti-pd1 therapeutic igg4 antibody pembrolizumab. *Nature Structural & Molecular Biology*, 22(12), 953–958.
- Schmidt, M., Hafner, M., & Frech, C. (2014). Modeling of salt and pH gradient elution in ion-exchange chromatography. *Journal of Separation Science*, 37(1-2), 5–13.
- Senior, A. W., Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T., Qin, C., Židek, A., Nelson, A. W., Bridgland, A., Penedones, H., Petersen, S., Simonyan, K., Crossan, S., Kohli, P., Jones, D. T., Silver, D., Kavukcuoglu, K., & Hassabis, D. (2020). Improved protein structure prediction using potentials from deep learning. *Nature*, 577(7792), 706–710.
- Shukla, A. A., Hubbard, B., Tressel, T., Guhan, S., & Low, D. (2007). Downstream processing of monoclonal antibodies-application of platform approaches. *J Chromatogr B Analyt Technol Biomed Life Sci*, 848(1), 28–39. <https://doi.org/10.1016/j.jchromb.2006.09.026>
- Shukla, A. A., & Thömmes, J. (2010). Recent advances in large-scale production of monoclonal antibodies and related proteins. *Trends in Biotechnology*, 28(5), 253–261.
- Smiatek, J., Jung, A., & Bluhmki, E. (2020). Towards a digital bioprocess replica: Computational approaches in biopharmaceutical development and manufacturing. *Trends in Biotechnology*, 38(10), 1141–1153.
- Velayudhan, A., & Horváth, C. (1988). Preparative chromatography of proteins: Analysis of the multivalent ion-exchange formalism. *Journal of Chromatography A*, 443, 13–29.
- von Stosch, M., Portela, R. M., & Varsakelis, C. (2021). A roadmap to ai-driven in silico process development: Bioprocessing 4.0 in practice. *Current Opinion in Chemical Engineering*, 33, 100692.
- Wang, C., Li, W., Drabek, D., Okba, N. M., van Haperen, R., Osterhaus, A. D., van Kuppeveld, F. J., Haagmans, B. L., Grosveld, F., & Bosch, B.-J. (2020). A human monoclonal antibody blocking sars-cov-2 infection. *Nature Communications*, 11(1), 1–6.
- Wang, G., Briskot, T., Hahn, T., Baumann, P., & Hubbuch, J. (2017). Estimation of adsorption isotherm and mass transfer parameters in protein chromatography using artificial neural networks. *Journal of Chromatography A*, 1487, 211–217.
- Weidle, U. H., Tiefenthaler, G., Weiss, E. H., Georges, G., & Brinkmann, U. (2013). The intriguing options of multispecific antibody formats for treatment of cancer. *Cancer Genomics & Proteomics*, 10(1), 1–18.
- Yamamoto, S., Nakanishi, K., & Matsuno, R. (1988). *Ion-exchange chromatography of proteins*. CRC Press.
- Yamamoto, S., Nakanishi, K., Matsuno, R., & Kamikubo, T. (1983). Ion exchange chromatography of proteins—prediction of elution curves and operating conditions. I. theoretical considerations. *Biotechnology and Bioengineering*, 25(6), 1465–1483.
- Zhu, C., Byrd, R. H., Lu, P., & Nocedal, J. (1997). Algorithm 778: L-bfgs-b: Fortran subroutines for large-scale bound-constrained optimization. *ACM Transactions on Mathematical Software (TOMS)*, 23(4), 550–560.
- Zhu, K., & Day, T. (2013). Ab initio structure prediction of the antibody hypervariable h3 loop. *Proteins: Structure, Function, and Bioinformatics*, 81(6), 1081–1089.
- Zhu, K., Day, T., Warshaviak, D., Murrett, C., Friesner, R., & Pearlman, D. (2014). Antibody structure determination using a combination of homology modeling, energy-based refinement, and loop prediction. *Proteins: Structure, Function, and Bioinformatics*, 82(8), 1646–1655.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Saleh, D., Hess, R., Ahlers-Hesse, M., Rischawy, F., Wang, G., Grosch, J.-H., Schwab, T., Kluters, S., Studts, J., & Hubbuch, J. (2022). A multiscale modeling method for therapeutic antibodies in ion exchange chromatography. *Biotechnology and Bioengineering*, 1–14. <https://doi.org/10.1002/bit.28258>