Journal of Chromatography A 1730 (2024) 465121

Contents lists available at ScienceDirect

Journal of Chromatography A

journal homepage: www.elsevier.com/locate/chroma

Modeling multi-component separation in hydrophobic interaction chromatography with improved parameter-by-parameter estimation method

Yu-Xiang Yang [a], Zhi-Yuan Lin [b], Yu-Cheng Chen [a], Shan-Jing Yao [a], Dong-Qiang Lin [a,*]

[a] Key Laboratory of Biomass Chemical Engineering of Ministry of Education, Zhejiang Key Laboratory of Smart Biomaterials, College of Chemical and Biological Engineering, Zhejiang University, Hangzhou 310058, China
[b] Zhejiang University-University of Edinburgh Institute, Zhejiang University, Haining 314400, China

ARTICLE INFO

Keywords:
Hydrophobic interaction chromatography
Multi-component system
Parameter estimation
Mechanistic model
Mollerup isotherm

ABSTRACT

Mechanistic models are powerful tools for chromatographic process development and optimization. However, hydrophobic interaction chromatography (HIC) mechanistic models lack an effective and logical parameter estimation method, especially for multi-component system. In this study, a parameter-by-parameter method for multi-component system (called as mPbP-HIC) was derived based on the retention mechanism to estimate the six parameters of the Mollerup isotherm for HIC. The linear parameters ($k_{s,i}$ and $k_{eq,i}$) and nonlinear parameters (n_i and $q_{max,i}$) of the isotherm can be estimated by the linear regression (LR) and the linear approximation (LA) steps, respectively. The remaining two parameters ($k_{p,i}$ and $k_{kin,i}$) are obtained by the inverse method (IM). The proposed method was verified with a two-component model system. The results showed that the model could accurately predict the protein elution at a loading of 10 g/L. However, the elution curve fitting was unsatisfactory for high loadings (12 g/L and 14 g/L), which is mainly attributed to the demanding experimental conditions of the LA step and the potential large estimation error of the parameter q_{max}. Therefore, the inverse method was introduced to further calibrate the parameter q_{max}, thereby reducing the estimation error and improving the curve fitting. Moreover, the simplified linear approximation (SLA) was proposed by reasonable assumption, which provides the initial guess of q_{max} without solving any complex matrix and avoids the problem of matrix unsolvable. In the improved mPbP-HIC method, q_{max} would be initialized by the SLA and finally determined by the inverse method, and this strategy was named as SLA+IM. The experimental validation showed that the improved mPbP-HIC method has a better curve fitting, and the use of SLA+IM reduces the error accumulation effect. In process optimization, the parameters estimated by the improved mPbP-HIC method provided the model with excellent predictive ability and reasonable extrapolation. In conclusion, the SLA+IM strategy makes the improved mPbP-HIC method more rational and can be easily applied to the practical separation of protein mixture, which would accelerate the process development for HIC in downstream of biopharmaceuticals.

1. Introduction

Hydrophobic interaction chromatography (HIC) is widely used as a polishing step in the downstream of biopharmaceuticals [1–3]. The conventional downstream process development, optimization and robustness analysis rely on researcher experience and extensive experimental work [2,4]. In contrast, the application of mechanistic model-based tools can reduce the number of labor-intensive experiments, shorten process development time and reduce the costs [5–9], thus meeting the demands of strongly competitive biopharmaceutical market and paving the way for digitization and integrated continuous manufacturing [10].

The core of the mechanistic model of HIC lies in the adsorption isotherm [11]. The six-parameter (stoichiometric parameter n, maximum binding capacity q_{max}, equilibrium coefficient k_{eq}, salt-protein interaction parameter k_s, and protein-protein interaction parameter k_p, and kinetic coefficient k_{kin}) isotherm of HIC proposed by Mollerup takes into account the salt dependence of protein adsorption

* Corresponding author.
E-mail address: lindq@zju.edu.cn (D.-Q. Lin).

https://doi.org/10.1016/j.chroma.2024.465121
Received 8 April 2024; Received in revised form 10 June 2024; Accepted 24 June 2024
Available online 29 June 2024

and the steric hindrance effect under nonlinear conditions [12–14], which has been popularly used for process characterization and optimization of HIC [9,15]. However, the HIC Mollerup isotherm lacks a rational and efficient method for model calibration, and the parameter estimation normally relies on the inverse method [9,15]. The inverse method is a standard approach to fitting simulated data to experimental data, depending on a reasonable optimization algorithm and initial guess. Unfortunately, for the parameter optimization problem in multi-dimensional spaces, the inverse method tends to fall into the trap of local optima or requires long time scales to converge [16–18]. This ill-posed parameter estimation problem may be caused by parameter correlation, limitations of experimental information, or inappropriate initial parameter guesses [19]. The consequences may be multiple, physically meaningless, and unstable solutions in parameter estimation [20]. In addition, as the number of component or parameter increases, the mechanistic model may suffer from overfitting due to excessive model complexity [21] or data noise [6,22]. The possible solutions are to remove terms that are expected to have little impact on the model's predictions [23], or fix some model parameters to reduce the parameter number [24]. The potential problems mentioned above imply that the parameters estimated by the inverse method might not be able to accurately predict outside of calibration experiments [20,25], thus certainly hindering the practical application of the HIC Mollerup isotherm.

Fixing model parameters at reasonable values based on prior knowledge of the mechanism can reduce the number of parameters that needed to be estimated [24], thus making the inverse method more reliable. The Yamamoto method is a parameter estimation method based on a physical cognition [26], which has been successfully used for process development and optimization in ion exchange chromatography (IEC) and mixed-model chromatography (MMC) [19,27–29]. This approach has the potential to be rationally applied to the linear region of the HIC Mollerup isotherm [28], where the parameters k_s and k_{eq} (classified as linear parameters) can be calculated by several linear gradient elution experiments (LGEs). However, the Yamamoto method tends to be more effective under dilution condition and cannot be directly transferred to the nonlinear region of the isotherm [30], which means that the parameters n and q_{max} (classified as nonlinear parameters) of the isotherm cannot be straightforwardly obtained through prior knowledge of the method. On the other hand, using the inverse method to determine the nonlinear parameters potentially has large uncertainty and error, making it difficult to obtain a physically meaningful result [31,32]. Therefore, the estimation of nonlinear parameters in the HIC isotherm exists challenging. To solve this problem, a parameter-by-parameter method for single-component HIC (called as PbP-HIC) was developed in our previous study based on the retention model [33]. Through six LGEs of different conditions, the PbP-HIC method allows the four isotherm parameters (k_s, k_{eq}, n, and q_{max}) to be estimated step-by-step with a deep physical understanding. The parameters k_s and k_{eq} can be estimated by the linear regression (LR) step and classified as the linear parameters, while the parameters n and q_{max} can be estimated by the linear approximation (LA) step and classified as the nonlinear parameters. The remaining two parameters (k_p and k_{kin}) are obtained by the inverse method. However, for practical applications to separate protein mixture, previous derivations based on single-component system did not consider the interactions between multiple proteins, which cannot be directly transferred to multi-component parameter estimation. Specifically, the estimation of the parameter q_{max} takes into account the steric hindrance of the protein [12], and q_{max} will change with the existence of other proteins in the multi-component system [32]. Furthermore, it was found in our previous study [33] that the q_{max} has a narrow experimental space for accurate estimation and is sensitive to the variation of loading, which is because the derivation assumption do not hold at high loading, thus potentially presenting a large estimation error. The challenge of q_{max} estimation and the limitation of experiment conditions hinder the application and generalization of the PbP-HIC method for multi-component system.

Based on the analysis above, a new parameter-by-parameter estimation method of HIC Mollerup isotherm for multi-component system (called as mPbP-HIC) would be developed first in this study. Four parameters ($k_{s,i}$, $k_{eq,i}$, n_i, and $q_{max,i}$) of component i would be estimated one-by-one, where the parameter $q_{max,i}$ would be determined by solving a matrix. The remaining two parameters ($k_{p,i}$ and $k_{kin,i}$) would be calibrated by the inverse method. The feasibility of the mPbP-HIC method would be verified by a two-component system containing lysozyme and α-Chymotrypsinogen A as the model proteins. However, the linear approximation step of the mPbP-HIC method requires demanding experimental conditions, and the estimation of the parameter $q_{max,i}$ potentially has a large error. Hence, the mPbP-HIC method would be further improved, and a promising enhancement is to further calibrate $q_{max,i}$ using the inverse method. Furthermore, to obtain the initial guess of $q_{max,i}$ easily and stably, a simplified linear approximation (SLA) would be developed in place of the linear approximation step. The feasibility and advantages of the improved mPbP-HIC method would also be evaluated with the two-component system, and compared with the inverse method. Finally, the process optimization of the two-component system with the estimated parameters would be performed to demonstrate the application potential for practical industrial protein separation.

2. Theoretical section

2.1. Chromatographic column model

The mass transfer within the chromatographic column is described using the lumped rate model (LRM) [34] as:

$$\frac{\partial c_i}{\partial t}+u_{int}\frac{\partial c_i}{\partial z}+k_{eff,i}\frac{3}{r_p}\frac{1-\varepsilon_c}{\varepsilon_c}\left(c_i-c_{p,i}\right)=D_{ax}\frac{\partial^2 c_i}{\partial z^2}, \tag{1}$$

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

where c_i and $c_{p,i}$ represent the mobile phase concentration of component i in the interparticle volume and intraparticle volume (mol/L), respectively. q_i denotes the concentration of component i in the stationary phase (mol/L). t represents time (s) and z represents the axial position (m). u_{int} characterizes the interstitial velocity (m/s). r_p is the particle radius (m). L denotes the column length (m). ε_c and ε_p denote the column porosity and particle porosity, respectively. D_{ax} accounts for the axial diffusion coefficient (m^2/s). $k_{eff,i}$ denotes the effective film diffusion coefficient of component i.

The rectangular pulse injection condition is applied as:

$$c_{inj,i}\left(t_{inj}\right)=\begin{cases} c_{inj}x_i & 0<t\leqslant t_{inj} \\ 0 & t>t_{inj} \end{cases} \tag{3}$$

where c_{inj} is the loading concentration of the protein sample (mol/L) and t_{inj} is the loading time (s). $c_{inj,i}$ and x_i represent the loading concentration (mol/L) and the mole fraction of component i, respectively. The boundary conditions are applied at the column inlet and outlet [35] as following:

$$u_{int}c_{inj,i}(t)=u_{int}c_i(t,0)-D_{ax}\frac{\partial c_i}{\partial z}(t,0), \tag{4}$$

$$\frac{\partial c_i}{\partial z}(t,L)=0. \tag{5}$$

2.2. Adsorption isotherm model

The HIC isotherm developed by Mollerup [12–14] is used as:

$$k_{\mathrm{kin},i}\frac{\partial q_i}{\partial t}=c_{\mathrm{p},i}k_{\mathrm{eq},i}\left(1-\sum_{j=1}^{N}\frac{q_j}{q_{\mathrm{max},j}}\right)^{n_i}\exp\left(k_{\mathrm{s},i}c_{\mathrm{s}}+k_{\mathrm{p},i}c_{\mathrm{p},i}\right)-q_i, \tag{6}$$

where $k_{\mathrm{kin},i}$, $k_{\mathrm{eq},i}$, $q_{\mathrm{max},i}$, n_i, $k_{\mathrm{s},i}$, $k_{\mathrm{p},i}$ are the kinetic coefficient, equilibrium coefficient, maximum binding capacity, stoichiometric parameter (number of ligands bound per protein), salt-protein interaction parameter, and protein-protein interaction parameter of component i, respectively. c_{s} represents the salt concentration in the mobile phase (mol/L). N represents the total number of components.

2.3. Parameter-by-parameter method for multi-component system

The derivation procedure of the PbP-HIC method for the single-component system can be found in the previous study [33]. The following derivations on the parameter estimation would focus on multi-component system, where the parameters $k_{\mathrm{p},i}$ and $k_{\mathrm{kin},i}$ would still be estimated using the inverse method.

2.3.1. Estimating $k_{s,i}$ and $k_{eq,i}$ with linear regression

For linear gradient elution of multiple components at low loading, if the $\exp(-k_{\mathrm{s},i}c_{\mathrm{s,R},i})/\exp(-k_{\mathrm{s},i}c_{\mathrm{s,initial}})\gg 1$ is satisfied for each component, the retention model can be written in logarithmic form as:

$$\ln(-\mathrm{GH})=-k_{\mathrm{s},i}c_{\mathrm{s,R},i}-\ln\left(k_{\mathrm{s},i}\cdot k_{\mathrm{eq},i}\right), \tag{7}$$

$$\mathrm{GH}=\frac{(1-\varepsilon_{\mathrm{t}})\left(c_{\mathrm{s,final}}-c_{\mathrm{s,initial}}\right)}{\mathrm{CV_G}}, \tag{8}$$

where GH represents the normalized gradient slope (mol/L) and $c_{\mathrm{s,R},i}$ describes the salt concentration at the peak maximum of component i. $\mathrm{CV_G}$ represents the length of the elution gradient in terms of column volume. $c_{\mathrm{s,initial}}$ and $c_{\mathrm{s,final}}$ correspond to the salt concentrations at the start and end of the gradient (mol/L), respectively. ε_{t} is the total porosity. Based on Eq. (7), the parameters k_{s} and k_{eq} for each component of the protein mixture can be determined from the slope and intercept of the linear regression.

2.3.2. Estimating n_i and $q_{max,i}$ with linear approximation

Under undiluted conditions, the nonlinear term $\left(1-\sum_{j=1}^{N}\frac{q_j}{q_{\mathrm{max},j}}\right)^{n_i}$ in the HIC isotherm cannot be neglected. In this nonlinear term, q of component j at the retention time of protein i $(t_{\mathrm{R},i})$ can be expressed using the Henry coefficient $H_{ij}=k_{\mathrm{eq},j}\exp\left(k_{\mathrm{s},j}c_{\mathrm{s},i}\right)$ [12,36] as:

$$q_{ij}\left(t_{\mathrm{R},i}\right)\approx H_{ij}\left(t_{\mathrm{R},i}\right)\cdot c_{\mathrm{p},ij}\left(t_{\mathrm{R},i}\right)=\frac{H_{ij}\left(t_{\mathrm{R},i}\right)\cdot c_{\mathrm{p},ij}\left(t_{\mathrm{R},i}\right)}{\bar{c}_{\mathrm{inj}}\left(t_{\mathrm{inj}}\right)}\cdot\bar{c}_{\mathrm{inj}}\left(t_{\mathrm{inj}}\right)=g_{ij}\left(t_{\mathrm{R},i}\right)\cdot\bar{c}_{\mathrm{inj}}\left(t_{\mathrm{inj}}\right), \tag{9}$$

where $c_{\mathrm{p},ij}$ and q_{ij} represent the mean concentration in the mobile and stationary phase (mol/L), respectively. $\bar{c}_{\mathrm{inj}}=\frac{c_{\mathrm{inj}}\cdot u_{\mathrm{int}}\cdot\varepsilon_{\mathrm{c}}\cdot t_{\mathrm{inj}}}{L}$ is the relative loading concentration (mol/L). The nonlinear term can be depicted using the nonlinear coefficient $a_i=g_{ij}/q_{\mathrm{max},i}$ as:

$$\left[1-\sum_{j=1}^{N}\frac{q_{ij}\left(t_{\mathrm{R},i}\right)}{q_{\mathrm{max},j}}\right]^{n_i}\approx\left[1-\sum_{j=1}^{N}\frac{g_{ij}\left(t_{\mathrm{R},i}\right)\cdot\bar{c}_{\mathrm{inj}}\left(t_{\mathrm{inj}}\right)}{q_{\mathrm{max},j}}\right]^{n_i}=\left[1-a_i\left(t_{\mathrm{R},i}\right)\cdot\bar{c}_{\mathrm{inj}}\left(t_{\mathrm{inj}}\right)\right]^{n_i}. \tag{10}$$

For elution under a linear gradient, the retention model can be written as:

$$\frac{1}{-\mathrm{GH}\cdot k_{\mathrm{s},i}\cdot k_{\mathrm{eq},i}}\left[\exp\left(-k_{\mathrm{s},i}c_{\mathrm{s,R},i}\right)-\exp\left(-k_{\mathrm{s},i}c_{\mathrm{s,initial}}\right)\right]=\left(1-a_i\cdot\bar{c}_{\mathrm{inj}}\right)^{n_i}. \tag{11}$$

The left-hand side term in Eq. (11) can be defined as $\boldsymbol{y}$, and Eq. (11) can be reformulated in logarithmic form as:

$$\ln\boldsymbol{y}_i=\boldsymbol{n}_i\cdot\ln\left(1-\boldsymbol{a}_i\cdot\bar{\boldsymbol{c}}_{\mathrm{inj}}\right). \tag{12}$$

Performing a Taylor expansion for the right-hand side term in Eq. (12), the second-order and higher-order terms of the expansion can be ignored when $\boldsymbol{a}_i\cdot\bar{\boldsymbol{c}}_{\mathrm{inj}}\ll 1$. In this case, the right-hand term of Eq. (12) can be simplified to $-\boldsymbol{a}_i\cdot\boldsymbol{n}_i\cdot\bar{\boldsymbol{c}}_{\mathrm{inj}}$, and $\boldsymbol{m}_2$ is defined as the slope ($\boldsymbol{m}_2=-\boldsymbol{a}_i\boldsymbol{n}_i$) of this linear equation. Integrating $\boldsymbol{m}_2$ into Eq. (12), the parameter $\boldsymbol{n}$ for each component can be determined by defining a loss function and minimizing it as expressed as:

$$\mathscr{L}(\boldsymbol{y};\boldsymbol{n})=\min\frac{1}{\boldsymbol{N}_{\mathrm{exp}}}\sum_{i=1}^{\boldsymbol{N}_{\mathrm{exp}}}\left[\ln\boldsymbol{y}_i-\boldsymbol{n}\cdot\ln\left(1+\frac{\boldsymbol{m}_2\cdot\bar{\boldsymbol{c}}_{\mathrm{inj},i}}{\boldsymbol{n}}\right)\right]^2, \tag{13}$$

where $\boldsymbol{N}_{\mathrm{exp}}$ represents the total number of LGEs. Furthermore, the nonlinear coefficient $\boldsymbol{a}$ for each component can be calculated:

$$\boldsymbol{a}_i=-\frac{\boldsymbol{m}_{2,i}}{\boldsymbol{n}_i}. \tag{14}$$

Through Eq. (10), $\boldsymbol{a}_i$ can be expressed as:

$$\sum_{j=1}^{N}\frac{\boldsymbol{g}_{ij}\left(t_{\mathrm{R},i}\right)}{\boldsymbol{q}_{\mathrm{max},j}}=\boldsymbol{a}_i\left(t_{\mathrm{R},i}\right). \tag{15}$$

The linear system of equations, $\mathbf{G}\left(1/\boldsymbol{q}_{\mathrm{max}}\right)=\boldsymbol{a}$, can be expressed as:

$$\begin{bmatrix}\boldsymbol{g}_{11} & \cdots & \boldsymbol{g}_{1j}\\ \vdots & \ddots & \vdots\\ \boldsymbol{g}_{i1} & \cdots & \boldsymbol{g}_{ij}\end{bmatrix}\begin{bmatrix}1/\boldsymbol{q}_{\mathrm{max},1}\\ \vdots\\ 1/\boldsymbol{q}_{\mathrm{max},j}\end{bmatrix}=\begin{bmatrix}\boldsymbol{a}_1\\ \vdots\\ \boldsymbol{a}_i\end{bmatrix}, \tag{16}$$

where the element $\boldsymbol{g}_{ij}=\boldsymbol{H}_{ij}\cdot\boldsymbol{c}_{\mathrm{p},ij}/\bar{\boldsymbol{c}}_{\mathrm{inj}}$ can be obtained from Eq. (9). If the matrix $\mathbf{G}$ is invertible and positive definite, $\boldsymbol{q}_{\mathrm{max}}$ can be calculated as:

$$1/\boldsymbol{q}_{\mathrm{max}}=\mathbf{G}^{-1}\boldsymbol{a}. \tag{17}$$

The above derivation is the linear approximation (LA) step for the mPbP-HIC method in multi-component system.

2.3.3. Simplification of linear approximation

Combining with the definition of $\boldsymbol{g}_{ij}$ in Eq. (9), the matrix $\mathbf{G}$ in Eq. (17) can be expressed as:

$$\det(\mathbf{G})=\prod_{j}^{N}\frac{\boldsymbol{k}_{\mathrm{eq},j}}{\bar{\boldsymbol{c}}_{\mathrm{inj}}}\cdot\det\begin{bmatrix}\exp\left(\boldsymbol{k}_{\mathrm{s},1}\cdot\boldsymbol{c}_{\mathrm{s,R},1}\right)\cdot\boldsymbol{c}_{\mathrm{p},11} & \cdots & \exp\left(\boldsymbol{k}_{\mathrm{s},N}\cdot\boldsymbol{c}_{\mathrm{s,R},1}\right)\cdot\boldsymbol{c}_{\mathrm{p},1N}\\ \vdots & \ddots & \vdots\\ \exp\left(\boldsymbol{k}_{\mathrm{s},1}\cdot\boldsymbol{c}_{\mathrm{s,R},N}\right)\cdot\boldsymbol{c}_{\mathrm{p},N1} & \cdots & \exp\left(\boldsymbol{k}_{\mathrm{s},N}\cdot\boldsymbol{c}_{\mathrm{s,R},N}\right)\cdot\boldsymbol{c}_{\mathrm{p},NN}\end{bmatrix}. \tag{18}$$

In practice, the second term of Eq. (18) is not always nonzero. Therefore, the matrix $\mathbf{G}$ may not be invertible and the solution of the linear system does not exist. To solve this problem, the simplified linear approximation in the PbP-IEC method published previously [32] is extended the present work for HIC. Under the assumption that the concentrations of other protein $\boldsymbol{c}_{\mathrm{p},ij}(\boldsymbol{i}\neq\boldsymbol{j})$ at the retention time of protein $\boldsymbol{i}$ tends to zero, $\boldsymbol{g}_{ij}$ can be simplified as:

$$\boldsymbol{g}_{ii}=\boldsymbol{k}_{\mathrm{eq},i}\cdot\exp\left(\boldsymbol{k}_{\mathrm{s},i}\cdot\boldsymbol{c}_{\mathrm{s,R},i}\right)\cdot\frac{\boldsymbol{c}_{\mathrm{p},ii}}{\bar{\boldsymbol{c}}_{\mathrm{inj}}}. \tag{19}$$

Therefore, the matrix $\mathbf{G}$ can be rewritten as a diagonal matrix:

$$\det(\mathbf{G})=\det\begin{bmatrix}\boldsymbol{k}_{\mathrm{eq},1}\cdot\exp\left(\boldsymbol{k}_{\mathrm{s},1}\cdot\boldsymbol{c}_{\mathrm{s,R},1}\right)\cdot\frac{\boldsymbol{c}_{\mathrm{p},11}}{\bar{\boldsymbol{c}}_{\mathrm{inj}}} & & \\ & \ddots & \\ & & \boldsymbol{k}_{\mathrm{eq},N}\cdot\exp\left(\boldsymbol{k}_{\mathrm{s},N}\cdot\boldsymbol{c}_{\mathrm{s,R},N}\right)\cdot\frac{\boldsymbol{c}_{\mathrm{p},NN}}{\bar{\boldsymbol{c}}_{\mathrm{inj}}}\end{bmatrix}. \tag{20}$$

The diagonal matrix $\mathbf{G}$ is positive definite, hence the linear system has a solution as:

$$q_{\max,i} = \frac{g_{ii}}{a_i} = k_{\mathrm{eq},i} \cdot \exp\left(k_{\mathrm{s},i} \cdot c_{\mathrm{s,R},i}\right) \cdot \frac{c_{\mathrm{p},ii}}{\bar{c}_{\mathrm{inj}}} \cdot \frac{1}{a_i}. \tag{21}$$

The above derivation is named as the simplified linear approximation (SLA) of the mPbP-HIC method. The SLA is computationally equivalent to the LA step for single-component system published previously [33].

3. Materials and methods

3.1. Calibration experiments for HIC

For calibration experiments, a pre-packed column (HiScreen Phenyl FF low sub, Cytiva, Uppsala, Sweden) with the dimensions of 0.77 cm × 10 cm and a column volume of 4.7 ml was used. A two-component system was constructed using lysozyme (abbreviated as Lys, 14.4 kDa, TargetMol, Massachusetts, USA) and α-Chymotrypsinogen A (abbreviated as Chy, 25.6 kDa, Sigma-Aldrich, Saint Louis, USA) as model proteins. All calibration experiments were performed using an ÄKTA Pure chromatography system (Cytiva, Uppsala, Sweden). The parameters of the column, resin and mass transfer were determined in the previous study [33] and were used in the present work as listed in Table 1.

The mPbP-HIC method utilized the six LGEs listed in Table 2. The linear regression step used three LGEs (Runs 1 to 3) with varying gradient lengths (10 CV to 30 CV) and consistent loadings (0.5 g/L) for estimating the parameters $k_{\mathrm{s},i}$ and $k_{\mathrm{eq},i}$. The linear approximation step utilized three LGEs (Runs 4 to 6) with different loading conditions (10 g/L to 14 g/L) and consistent gradient lengths (30 CV) for estimating the parameters n_i, and $q_{\max,i}$.

The total concentration of the protein mixture sample was adjusted to 2.91 g/L (2.0 g/L for Lys and 0.91 g/L for Chy) using the equilibration buffer. All LGEs were first equilibrated for 3 CV with 25 mmol/L phosphate buffer (pH 7.0) containing 2.0 mol/L ammonium sulfate. After injecting the protein mixture under the loading conditions listed in Table 2, the column was washed with equilibration buffer for 3 CV and then eluted at the gradient length listed in Table 2. Finally, the column was washed with 25 mmol/L phosphate buffer (pH 7.0) for 3 CV and regenerated with deionized water for 3 CV.

To analyze Lys and Chy concentrations in the fractions, a TSKgel G3000SW$_{\mathrm{XL}}$ column (Tosoh, Tokyo, Japan) was used with isocratic flow of SEC buffer (50 mmol/L phosphate buffer, 150 mmol/L sodium sulfate, pH 7.0) and run at a flow rate of 1 mL/min for 15 min. The sample was analyzed with a wavelength of 280 nm.

3.2. Process simulation and optimization

The chromatography analysis and design toolkit (CADET) [37,38] was used for chromatographic process simulation. The discretization of the column was set to 150 axial nodes. The absolute and relative tolerances were set to 10^{-6}.

For the estimation of parameter $\boldsymbol{n}$, the least squares minimization problem was tackled by the scipy.optimize module in Python 3.10 [39].

The inverse method estimates the unknown parameters through experimental elution curves and loss function. The loss function is defined as follows:

$$\mathcal{L}(c_{\mathrm{sim}}; \mathbf{P}) = \min \sum_{j=1}^{N_{\mathrm{exp}}} \sum_{i=1}^{N_{comp}} \| c_{\mathrm{exp},i} - c_{\mathrm{sim},i}(z, t; \mathbf{P}) \|_{L^2}^2, \tag{22}$$

where N_{exp} and N_{comp} denote the number of calibration experiments and components. c_{exp} and c_{sim} represent experimental results and numerical approximations of protein concentration, respectively. $\mathbf{P}$ represents the set of parameters that need to be estimated. $\| \cdot \|_{L^2}$ represents the L^2-norm. The parameter search ranges for the inverse method are shown in Table 3.

Genetic algorithm (GA), simulated annealing (SA), and particle swarm optimization (PSO) from the scikit-opt library were used as optimizers for parameter estimation and process optimization. To assess the effectiveness of parameter estimation, the deviation between the experimental curves y_{exp} and the model simulations y_{sim} was measured by the average L^2-error as:

$$\frac{1}{N_{\mathrm{exp}}} \sum_{j=1}^{N_{\mathrm{exp}}} \sum_{i=1}^{N_{comp}} \sqrt{\frac{\| y_{\mathrm{exp},i} - y_{\mathrm{sim},i} \|_{L^2}^2}{\| y_{\mathrm{exp},i} \|_{L^2}^2}}. \tag{23}$$

Table 1
Model parameters of column, resin and mass transfer.

Parameter	Symbol	Value	Unit
Column length	L	100.0	mm
Column volume	V	4.65	mL
Flow rate	u	0.417	mm/s
Column porosity	ε_c	0.37	–
Particle porosity	ε_p	0.97	–
Particle radius	r_p	45.0	μm
Axial dispersion coefficient	D_{ax}	1.50×10^{-7}	m^2/s
Effective film diffusion coefficient	k_{eff}	1.50×10^{-5}	m/s
Salt effective film diffusion coefficient	$k_{\mathrm{eff,s}}$	1.50×10^{-5}	m/s

Table 2
Experimental runs for parameter calibration.

Run	Purpose	CV_G (CV)	Loadings (g/L)	$c_{\mathrm{s,initial}}$ (mol/L)
1	LR	10	0.5	2.0
2	LR	20	0.5	2.0
3	LR	30	0.5	2.0
4	LA	30	10.0	2.0
5	LA	30	12.0	2.0
6	LA	30	14.0	2.0

4. Results and discussion

4.1. Parameter estimation based on the mPbP-HIC method

The feasibility of the mPbP-HIC method for multi-component system was verified with the protein mixture containing Lys and Chy. The parameter $q_{\max}$ was calculated directly from the matrix solution (derived in Section 2.3.2). The procedure of the mPbP-HIC method is depicted in Fig. 1.

Six calibration experiments were carried out with protein mixture following the design strategy proposed in previous study [33]. As listed in Table 2, Run 1 to Run 3 were performed at low loading, when the effect of the nonlinear term was neglected. Through establishing a linear relationship between $c_{\mathrm{s,R}}$ and $\mathbf{GH}$, the linear parameters k_s and k_{eq} for each component can be calculated from the slope and intercept of Eq. (7). Run 4 to Run 6 were conducted under undiluted conditions, and the peak shapes exhibit asymmetry. Through linear approximation, the nonlinear parameters n and $q_{\max}$ can be obtained sequentially using the loss function and matrix solving.

The linear regression and linear approximation results are shown in Fig. 2. In the LR step, the linear correlation coefficients of both proteins approach 1.00. In the LA step, Lys shows a good linear relationship,

Table 3
Parameter search boundaries of the inverse method.

Parameter	Lower bound	Upper bound	Unit
k_s	1.00	10.0	$(kmol/m^3)^{-1}$
k_{eq}	0.100	10.0	–
n	0.100	10.0	–
$q_{\max}$	0.0100	1.00	$kmol/m^3$
k_p	−50,000	50,000	$(kmol/m^3)^{-1}$
k_{kin}	1.00	1000	s

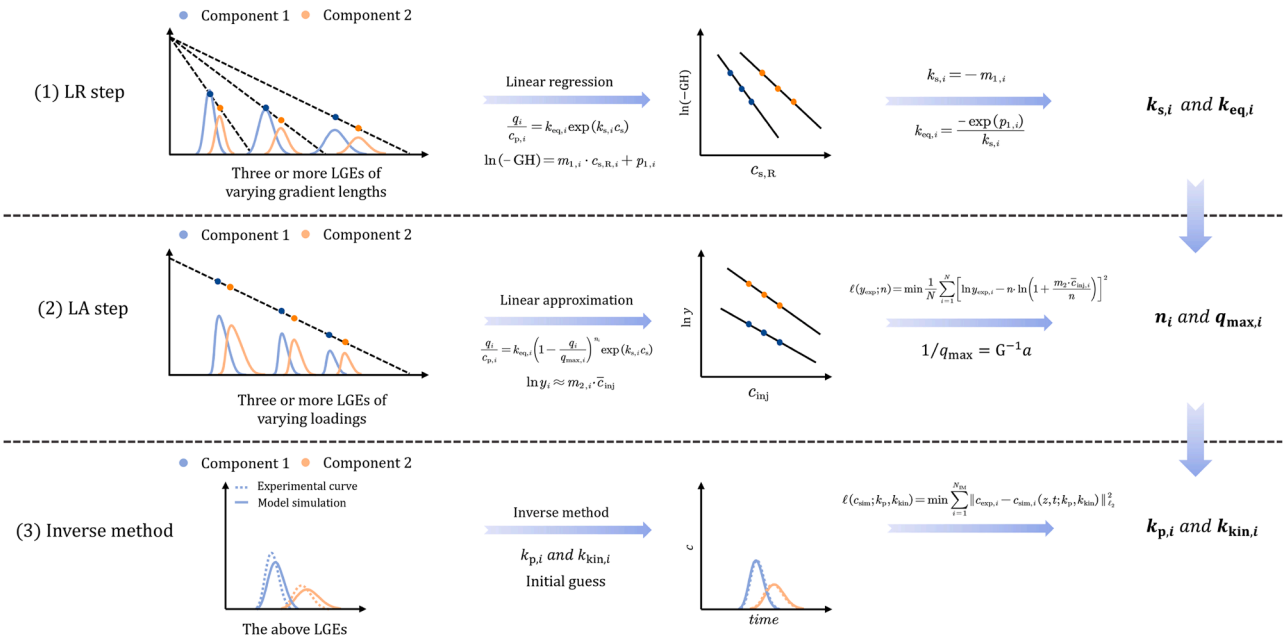

Fig. 1. Flowchart of parameter estimation based on the mPbP-HIC method for multi-component system.

while the correlation coefficient of Chy is 0.93. The relatively weak linear relationship for Chy may be attributed to experimental error or noise [31], which affected the selection of retention time for Chy.

The parameters k_p and k_{kin} were estimated using the inverse method. Table 4 lists all parameters estimated. Compared to the parameters of Lys obtained in the previous study for single-component separation system with same resin [33], k_s increased from 4.58 to 4.67 and k_{eq} decreased from 0.888 to 0.862. The linear parameters of Lys are relatively consistent, and the slight variations may be due to the different conditions of the two sets of calibration experiments. For the nonlinear parameters, n decreased from 0.922 to 0.867 and q_{max} decreased from 0.156 to 0.114. A possible explanation for the decrease in q_{max} is that the competitive binding effect causes a more asymmetric peak shape in the multi-component system than that in the single-component under the same loading, which affects the retention time of the component, resulting in a smaller q_{max}. The general understanding is that the proteins with higher molecular weights have more significant steric hindrance effects [40,41]. In this study, the estimated q_{max} of Chy is lower than that of Lys, which is consistent with the physical understanding.

The model simulations based on the estimated parameters are shown in Fig. 3. Good agreement between the experimental results and the model simulations can be found, and the average L^2-error is 0.465 for six calibration experiments. At low loadings (Runs 1, 2, 3), the components elute earlier in the model simulations compared to the experimental curves. This might be attributed to the estimation errors of the linear parameters, which have a significant influence on the retention time prediction of the elution peak [33]. According to our previous investigation [33], the estimation of parameter k_s is weakly affected by changes in retention position or loading in the LR step and has a superior accuracy, whereas k_{eq} is sensitive to such effects. Hess et al. [28] also observed that k_{eq} has higher estimation uncertainty compared to k_s

Table 4
Isotherm parameters estimated by the mPbP-HIC method.

Parameter	Lys	Chy	Unit
k_s	4.67	5.78	$(kmol/m^3)^{-1}$
k_{eq}	0.862	0.944	–
n	0.867	1.36	–
q_{max}	0.114	0.0361	$kmol/m^3$
k_p	715.3	1090.5	$(kmol/m^3)^{-1}$
k_{kin}	47.2	54.6	s

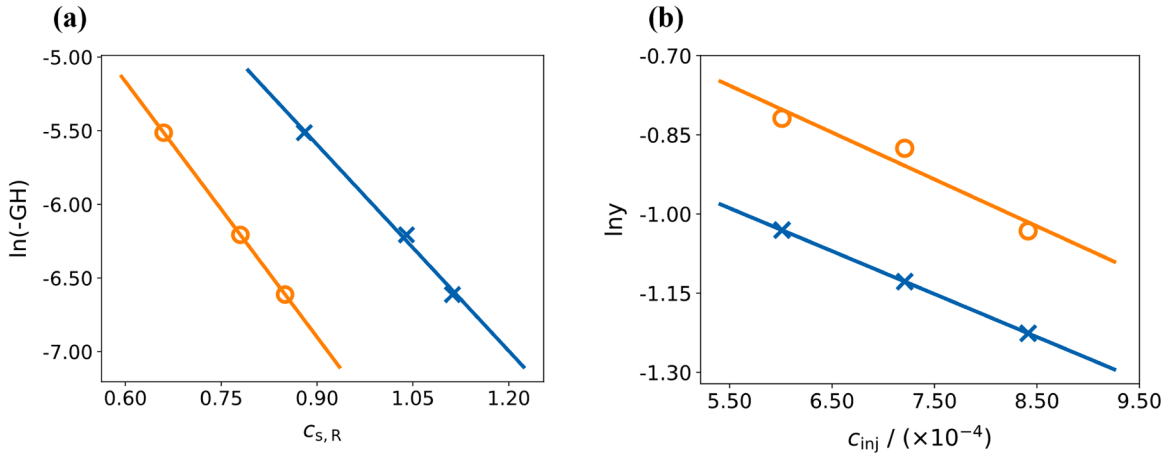

Fig. 2. Linear regression based on Runs 1–3 (a) and linear approximation based on Runs 4–6 (b). ◯: Lys, × : Chy.

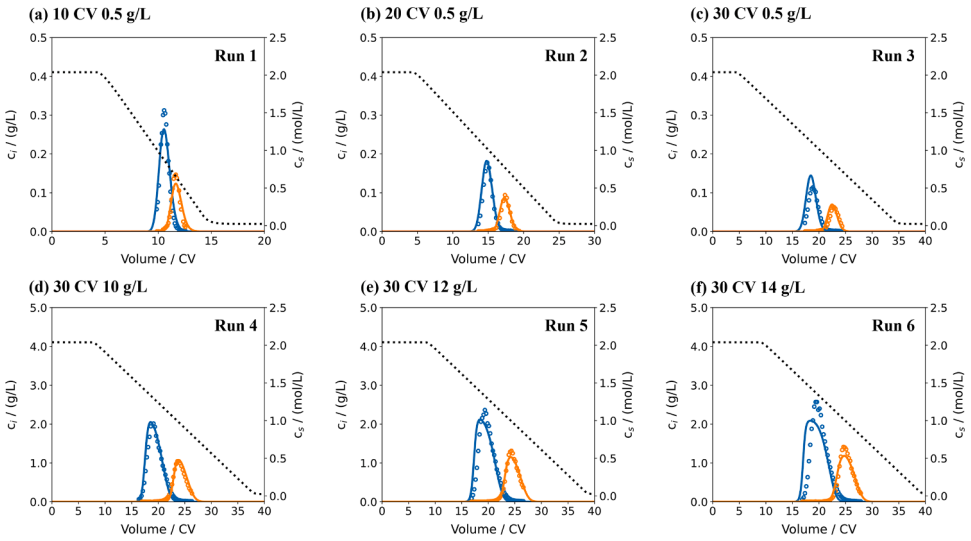

Fig. 3. Elution curves of HIC experiments (scatters, blue: Lys, orange: Chy) and model calculations based on the mPbP-HIC method (solid line, blue: Lys, orange: Chy). The simulated salt concentrations at the column outlet are presented at the dotted line. (a) to (f) correspond to six calibration experiments in Table 2.

through the confidence ellipse. Therefore, the bias in retention time may be mainly caused by the error of $\boldsymbol{k_{eq}}$ in this case. For the three LGEs at high loading (Runs 4, 5, 6), the agreement between simulations and experiments gradually worse with increasing loading, with an average $\boldsymbol{L}^2$-error of 0.427. In Runs 5 and 6, the peak heights of the model simulations are obviously lower than those of the experiment curves, and the asymmetry of the simulated curves increases more significantly. This discrepancy at high loading may be attributed to the estimation deviation of the nonlinear parameter caused by the error accumulation effect. For the mPbP-HIC method, because the parameters are determined sequentially one-by-one, the parameter estimation errors in the former step will accumulate to the latter step, which may lead to unreasonable nonlinear parameters, especially the parameter $\boldsymbol{q_{max}}$ obtained in the last step. Due to the estimation bias of the parameter $\boldsymbol{q_{max}}$, the model cannot reliably evaluate the peak shape of protein elution, which affects the prediction of the retention time. To enhance the fitting degree of the elution curves and to reduce the error accumulation effect, it is necessary to improve the mPbP-HIC method.

4.2. Improvement of the mPbP method for parameter estimation

Based on the above results, the mPbP-HIC method has limitations and needs to be improved. When the isotherm is in the transition or partly nonlinear region and the slope change of the isotherm is relatively small, the assumption of $\boldsymbol{a_i \cdot \bar{c}_{inj} \ll 1}$ and the approximation of Eq. (9) would hold. However, the above assumption and approximation gradually fail as the loading increases, which leads to the possibility that the LA step of the mPbP-HIC method may give unreasonable nonlinear parameters, resulting in a poor curve fitting. Especially for the parameter $\boldsymbol{q_{max}}$, the calculation of $\boldsymbol{q_{max}}$ from Eq. (17) will be unreliable at high loading, since the matrix **G** is based on the approximation of Eq. (9), which fails gradually as the loading increases.

In our previous study [33], the range of operating conditions (loading and salt concentration) that make the LA step work well was explored by numerical experiments. It was found that the parameter $\boldsymbol{n}$ can be estimated with high accuracy over a broader range of operating conditions (4 % < loading factor < 20 %) by the LA step, while the $\boldsymbol{q_{max}}$ estimate is more sensitive to the experiment conditions (estimation accurately only at 3.5 % < loading factor < 8.5 %), and the error may be high in practical experiments due to the assumption failure or the error accumulation effect. Thus, it is worth to consider further calibrating the parameter $\boldsymbol{q_{max}}$ obtained from LA step by the inverse method to reduce the estimation error and improve the fitting degree of elution curve. Similar strategies have been used for the determination of the steric mass action (SMA) isotherm parameters, aiming to streamline the existing calibration techniques [19,32,42]. Furthermore, as analyzed in Section 2.3.3, another potential problem with the LA step is that it might be impossible to solve the matrix to obtain $\boldsymbol{q_{max}}$, for example in the presence of the singular matrix. Despite the fact that the LA linear system has a unique solution, it is still cumbersome to solve when the number of components increases in the complex multi-component system [32]. Obviously, it is not necessary to accurately compute the initial guess of $\boldsymbol{q_{max}}$ by solving a complex linear system. Therefore, the SLA step in the mPbP-HIC method was proposed to easily obtain the initial value of $\boldsymbol{q_{max}}$. The improved mPbP-HIC method and the procedure is illustrated in Fig. 4. Main improvement is to replace the LA step with the SLA step in the mPbP-HIC method for the parameter $\boldsymbol{q_{max}}$ estimation, and to take the estimated $\boldsymbol{q_{max}}$ as an initial guess for the inverse method.

The parameters were estimated following the improved mPbP-HIC method mentioned above. The linear parameters ($\boldsymbol{k_s}$ and $\boldsymbol{k_{eq}}$) and $\boldsymbol{n}$ are consistent as Section 4.1, while $\boldsymbol{q_{max}}$ would be initialized by the SLA as:

$$\begin{bmatrix} 92.52 & 0.2047 \\ 0.4254 & 19.76 \end{bmatrix} \\ \begin{bmatrix} 103.0 & 0.5041 \\ 0.4486 & 21.87 \end{bmatrix} \times \begin{bmatrix} 1/q_{max,Lys} \\ 1/q_{max,Chy} \end{bmatrix} = \begin{bmatrix} 937.7 \\ 654.8 \end{bmatrix}. \\ \begin{bmatrix} 113.7 & 0.6663 \\ 0.7060 & 23.61 \end{bmatrix} \tag{24}$$

In Eq. (24), the diagonal elements of the matrix **G** are much larger than the other elements in the same row, thus the simplification of the non-diagonal elements in the SLA is reasonable. The initial guess for $\boldsymbol{q_{max}}$ ($\boldsymbol{q_{max,Lys}}$=0.109 and $\boldsymbol{q_{max,Chy}}$=0.0331) can be derived directly from Eq. (21).

As the SLA could provide reasonable initial guesses for $\boldsymbol{q_{max}}$, the search boundaries for $\boldsymbol{q_{max}}$ could be narrowed to the vicinity of the initial guess ($0.545 < \boldsymbol{q_{max,Lys}} < 0.218$ and $0.016 < \boldsymbol{q_{max,Chy}} < 0.066$). The boundaries for $\boldsymbol{k_p}$ and $\boldsymbol{k_{kin}}$ are listed in Table 3. The three parameters were further estimated by the inverse method, and the results are presented in Table 5. The $\boldsymbol{q_{max}}$ obtained by SLA+IM with the improved mPbP-HIC method are higher than those obtained by LA with the classical mPbP-HIC method, which implies that the LA step may have

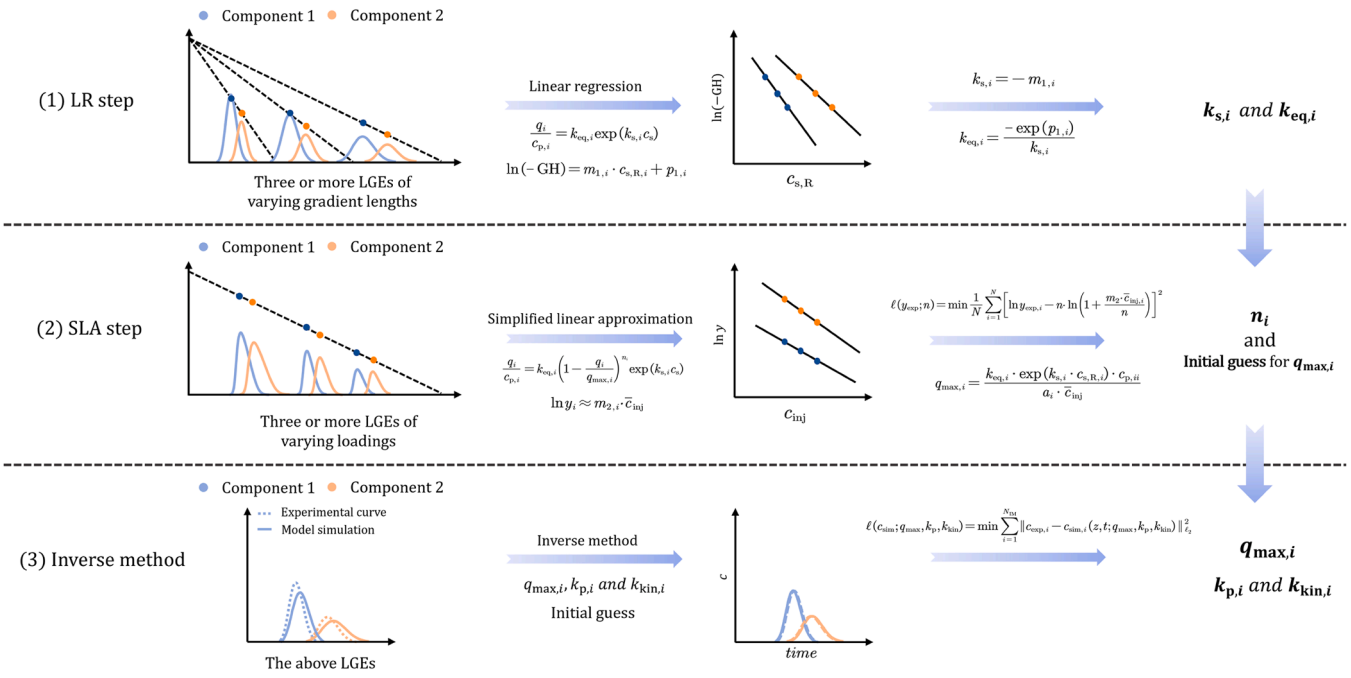

Fig. 4. Flowchart of parameter estimation based on the improved mPbP-HIC method for multi-component system.

overestimated the steric hindrance effect. It can be found that when the parameter q_{max} was further estimated by the inverse method, the re-estimated result of the k_p in Table 5 changed relatively large compared to that in Table 4. This may be due to the comparatively large confidence interval for this parameter [43], and the high uncertainty in the inverse method, which can be further evaluated with Markov Chain Monte Carlo [44]. With the improved mPbP-HIC method, the elution curve fitting is significantly improved under undiluted condition (Runs 4, 5, 6) and the average L^2-error decreases from 0.427 (Section 4.1) to 0.346 as shown in Fig. 5. It is interesting to found that the loading factor for lysozyme ($\mathbf{LF} = \bar{c}_{inj}/\left(q_{max,LYS}\cdot(1-\varepsilon_t)\right)\times$ 100%) [32,33] was 30.1% in Run 6, at which point the experiment was in relatively poor agreement with the model simulation. This is due to the fact that the current understanding of the HIC adsorption mechanism is limited [45], and the mechanistic model of Mollerup isotherm might not reasonably characterize the high loading or overloading conditions [9].

Compared to the LA step, the SLA step in the improved mPbP-HIC method can avoid the risk of unsolvable matrix and reduces the time complexity for solving linear system. For a protein mixture system containing n components, the time complexity of solving Eq. (17) in the LA step is $O(n^3)$ using the classical Gaussian elimination method. However, the time complexity required to compute Eq. (21) of SLA is only $O(n)$. Obviously, the solution time of SLA increases linearly only with the growth of the component. For practical separation of three-component system, the solving speed of the SLA step will be 27 times faster than the LA step (using Gaussian elimination method).

The initial guess of q_{max} obtained using the SLA step is based on the retention mechanism that provides a comprehensive physical knowledge of the parameters. This would narrow the search parameter space for the inverse method, thus avoiding the ill-posed problem of multiple solutions for nonlinear parameter q_{max} estimation. As reported in our previous study for the PbP-IEC method [32], the SLA+IM results were also closer to the ground truths and are more reasonable than the directly determining with the LA estimation. The results of better elution curve fitting with the SLA+IM strategy indicate that the error accumulation effect caused by the stepwise parameter estimation might be effectively reduced. Importantly, to simultaneously estimate the parameters n and q_{max} with high accuracy, the LA step of the classical mPbP-HIC method tends to be demanding on the calibration experimental conditions. In contrast, the use of the SLA+IM strategy makes the improved mPbP-HIC method more rational and convenient, and allows the parameters to be calibrated over a wider operating space.

However, for practical application of the improved mPbP-HIC method, the error accumulation effect still exists, and the experimental noises (e.g., conductivity shifts at high salt concentrations and UV detector noise) are complex and miscellaneous, which may have impacts on the parameter estimation [31,46]. The mechanistic model [47] and the improved mPbP-HIC method are difficult to account for these effects, and thus may lead to poor fitting of calibration experiments. It is worthwhile to consider using the parameters obtained from the improved mPbP-HIC method as initial values, and then further estimating some or all of the parameters by inverse method. Since the improved mPbP-HIC method could provide more reasonable initial guess, the parameter search range of the inverse method will be narrowed, allowing for faster convergence and eliminating the challenges for multiple optima [48].

Table 5
Isotherm parameters estimated by the improved mPbP-HIC method.

Parameter	Lys	Chy	Unit
k_s	4.67	5.78	$(kmol/m^3)^{-1}$
k_{eq}	0.862	0.944	–
n	0.867	1.36	–
q_{max}	0.140	0.0393	$kmol/m^3$
k_p	−878.3	−492.1	$(kmol/m^3)^{-1}$
k_{kin}	49.5	56.2	s

4.3. Parameter estimation with inverse method

In order to compare the proposed mPbP-HIC method, all isotherm parameters were estimated using the inverse method, and the parameter search boundaries are listed in Table 3.

Table 6 summarizes the results of the parameter estimation. In this study, the inverse method faced a 12-dimension optimization problem (six parameters per protein). During the inverse procedure, the assignment of unreasonable parameters or parameter combinations may tend to result in steep concentration fronts, which can abort or fail the

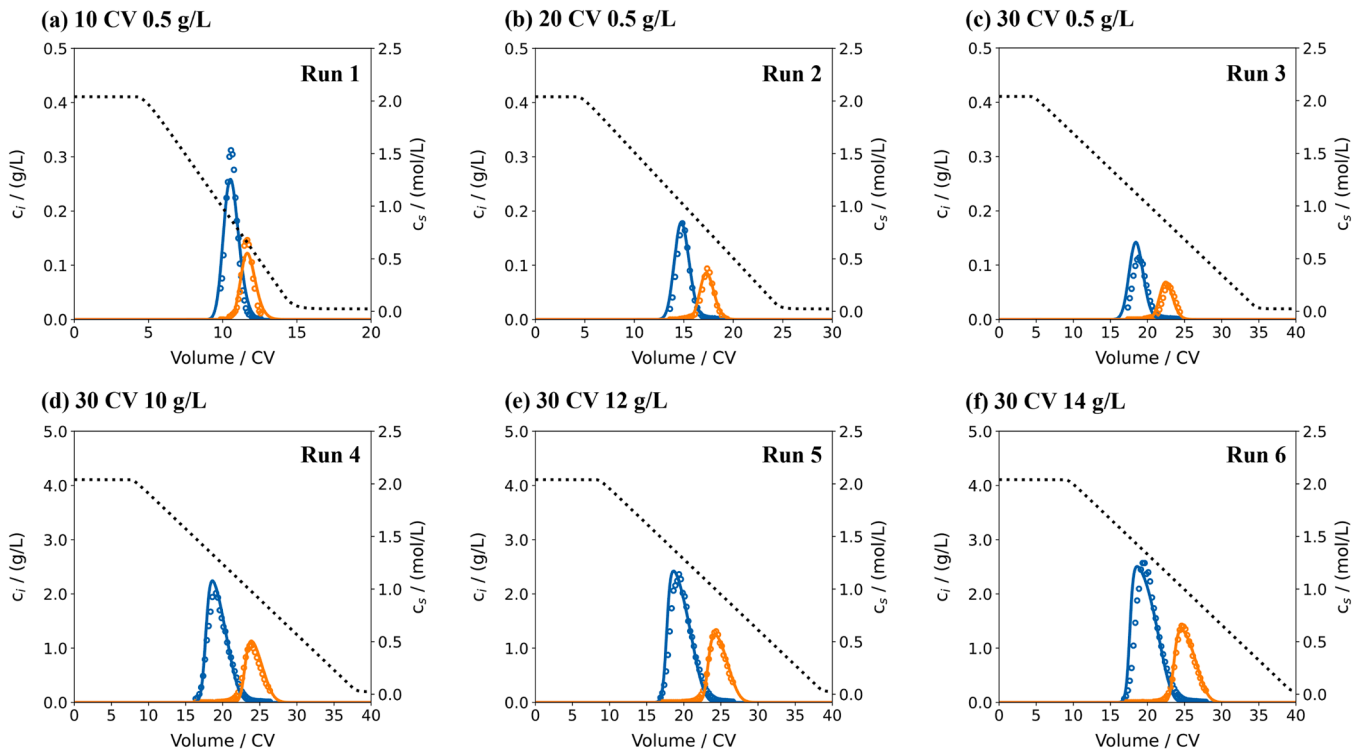

Fig. 5. Elution curves of HIC experiments (scatters; blue: Lys, orange: Chy) and model calculations based on the improved mPbP-HIC method (solid line; blue: Lys, orange: Chy). The simulated salt concentrations at the column outlet are presented at the dotted line. (a) to (f) correspond to six calibration experiments in Table 2.

optimization process [49]. In contrast, the improved mPbP-HIC method requires only six parameters (three parameters per protein) to be estimated by the inverse method and has reasonable initial values for q_{max}, thus would reduce the difficulty and time of parameter search exponentially.

Fig. 6 shows the experimental curves and the model simulations based on the estimated parameters. Compared to Fig. 5, the parameters obtained by the inverse method more precisely fitted the protein concentration. The average L^2-error for the six experiments was 0.375, while the L^2-error was 0.429 for the improved mPbP-HIC method in Fig. 5. The inverse method is based on direct fitting of curves, which would clearly lead to smaller L^2-error. However, the result of the inverse method varies depending on the algorithm and the optimization objectives (e.g., peak shape, peak position, and peak height) [16,50], and there may be ill-posed problem in parameter estimation [19]. This will lead to multiple or unstable solutions for the parameters, which may cause the model unable to predict accurately outside the calibration experiments [20,25].

4.4. Process optimization and model validation

The estimated parameters were first validated by the process optimization for a sample loading of 10 g/L, with Lys defined as the main product. The optimization objective was to maximize the sum of purity and yield of Lys. After each simulation, the protein concentration profiles were automatically integrated to calculate the area. The pooling interval started from the beginning of Lys elution and ended when the optimization objective was maximized.

Table 6
Isotherm parameters estimated by the inverse method.

Parameter	Lys	Chy	Unit
k_s	5.06	5.87	$(kmol/m^3)^{-1}$
k_{eq}	0.585	0.905	–
n	2.03	3.15	–
q_{max}	0.224	0.0651	$kmol/m^3$
k_p	−286.0	1385.6	$(kmol/m^3)^{-1}$
k_{kin}	51.0	39.5	s

The process optimization was based on two-gradient chromatography with a salt concentration of 2.0 mol/L at the start of the first gradient. The salt concentration at the end of the first gradient was defined as the decision variable. The optimization was performed using parameters obtained from the improved mPbP-HIC method (listed in Table 5). After the optimization, the final value of the decision variable (the salt concentration at the end of the first gradient) was 1.44 mol/L, and predicted purity of Lys was 99.9 % with the yield of 99.7 %.

To compare the two mPbP-HIC method and the inverse method, Fig. 7 shows the model predictions and the experimental data for the optimized process with the gray area representing the pooling interval. The obtained purity of the experiment was close to 100 % with the yield of 97.1 %. The decrease in yield compared to the predicted value was attributed to the fact that Lys eluted earlier in the experiment than the model prediction and the peak shape was more trailing. It can be observed that the parameters provided by both mPbP-HIC methods accurately predict the retention time of the main product with acceptable errors. However, the parameters obtained by the improved mPbP-HIC method have better prediction for the peak height and peak shape of Chy. In addition, although the parameters obtained by the inverse method had a lower L^2-error in fitting the calibration experiments, there was a larger deviation in the process optimization. This indicated the potential risk of applying the inverse method to estimate all parameters in practical process development, and a good fit achieved by parameter estimation would not guarantee the predictive ability of the model [48].

In order to compare the two mPbP-HIC methods that performed good in the loading of 10 g/L optimization experiment, further process optimization was carried out under the extrapolation condition at the loading of 16 g/L.

The optimization process was kept consistent with the above. After optimization, the decision variable was 1.56 mol/L, and the predicted purity and yield were 99.9 % and 99.9 %, respectively. The process

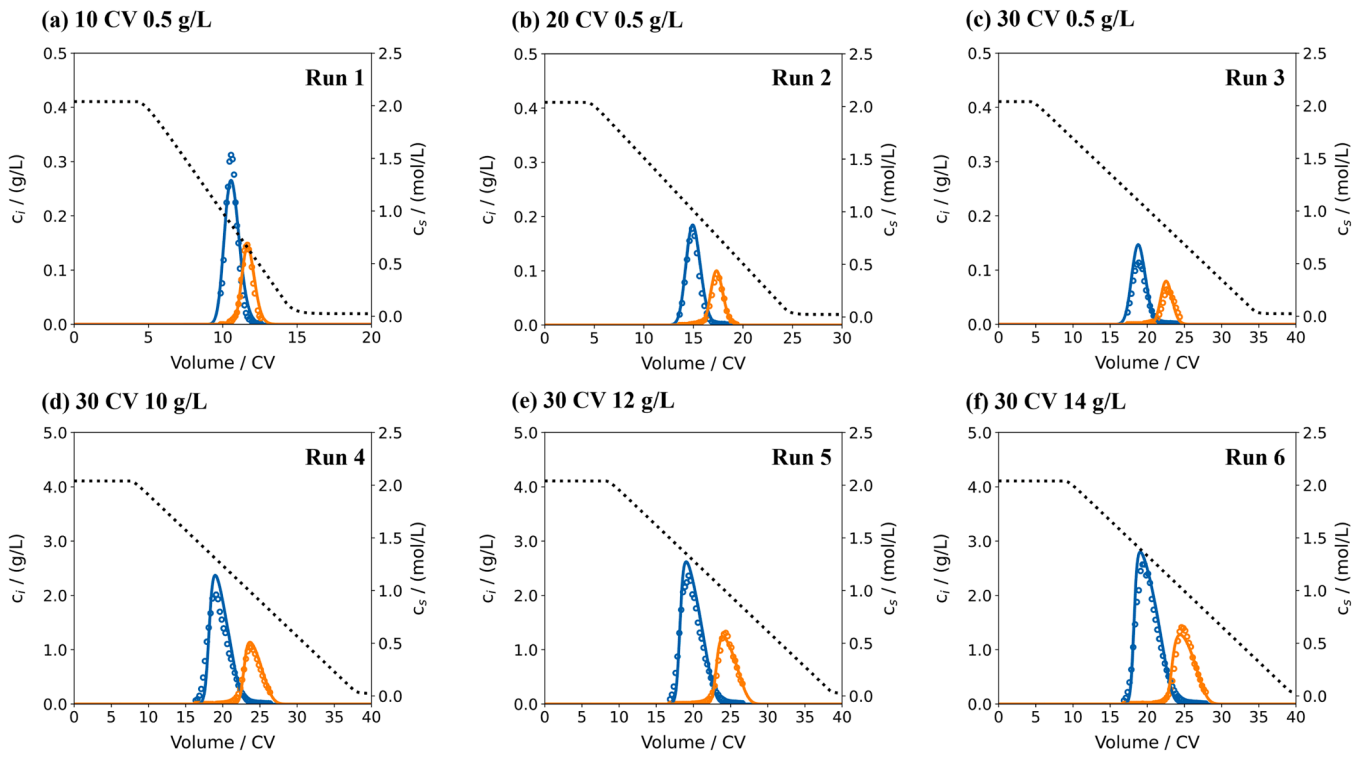

Fig. 6. Elution curves of HIC experiments (scatters; blue: Lys, orange: Chy) and model calculations based on the inverse method (solid line; blue: Lys, orange: Chy). The simulated salt concentrations at the column outlet are presented at the dotted line. (a) to (f) correspond to six calibration experiments in Table 2.

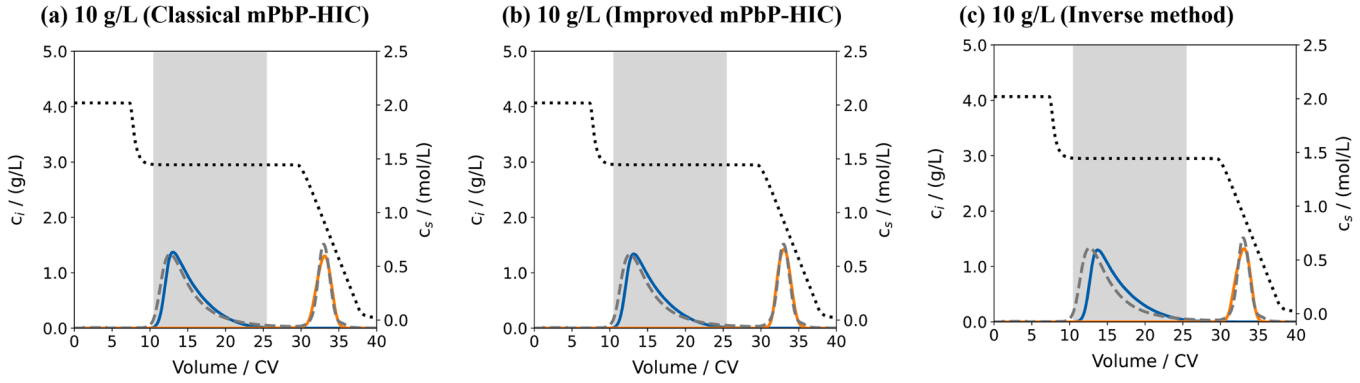

Fig. 7. Elution curves of the optimization experiments (dashed line) and model calculations (solid line, blue: Lys, orange: Chy) for 10 g/L loading. (a) the classical mPbP-HIC method; (b) the improved mPbP-HIC method; (c) the inverse method. The simulated salt concentrations at the column outlet are presented with the dotted line. The gray area represents the pooling interval.

simulation and experimental result are shown in Fig. 8. The purity of the experiment was 97.1 % with the yield of 99.8 %. The yield was relatively consistent with the model prediction, while the purity was slightly lower than the prediction. This might due to a slight overlap of Lys and Chy in the experiment, for which the model failed to accurately predict and gave a wider pooling interval. It should be noted that in practical process optimization, a more reasonable optimization objective would be to maximize productivity, or to maximize yield while ensuring the purity requirements [17,51]. For the loading of 16 g/L, the deviations between model calculations and experiments are non-negligible for both sets of parameters. This deviation may be attributed to a combination of parameter uncertainty, data quality, and model bias [9,17]. However, the model prediction of the parameters provided by the improved mPbP-HIC method was more accurate, with a L^2-error reduction of 0.07 compared to the classical mPbP-HIC method. Moreover, the improved method showed a smaller prediction bias for the peak shape and retention time of Lys. In combination with the calibration experiments, the improved mPbP-HIC method also performed better prediction, indicating that the parameters determined by this method are more physically meaningful, especially for the parameter $\boldsymbol{q_{max}}$. Further determination of $\boldsymbol{q_{max}}$ using the inverse method will enable a more rational evaluation of the steric hindrance effect, allowing the model to effectively characterize the chromatographic process over a wider experimental condition outside the model calibration space.

However, Creasy et al. [52] observed that the protein binding behavior of HIC was expected to be more complicated as the protein loading gradually increased to 30 % of the column equilibrium binding capacity, where the Mollerup HIC isotherm might not represent this binding behavior. In practical industrial separation, despite the commonly unavoidable deviation existing between model predictions and experimental validations, the model-based tool clearly leads in the right direction for process development [8]. The impact of deviation can

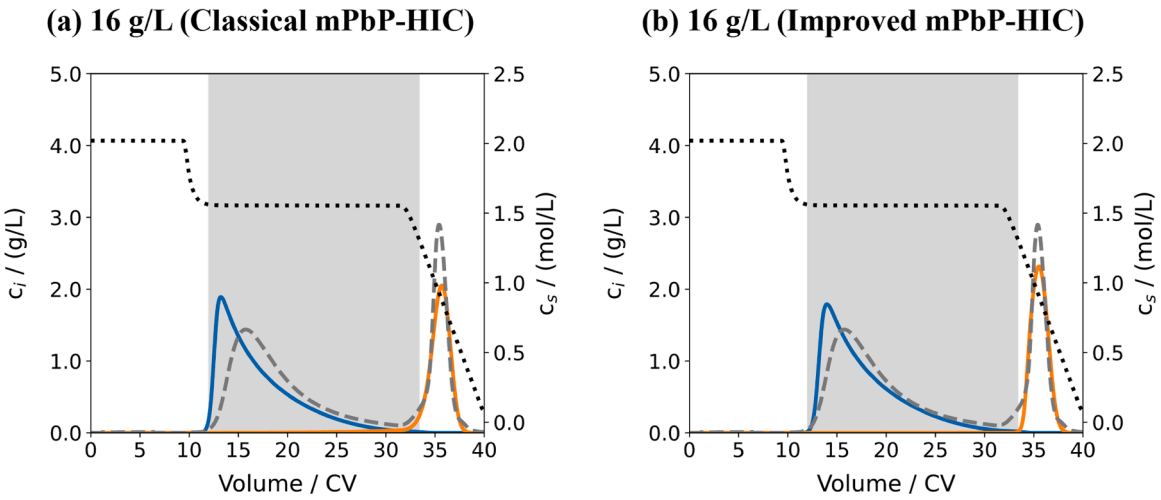

Fig. 8. Elution curves of the optimization experiments (dashed line) and model calculations based on different PbP-HIC methods (solid line, blue: Lys, orange: Chy) for 16 g/L loading. (a) the classical mPbP-HIC method; (b) the improved mPbP-HIC method. The simulated salt concentrations at the column outlet are presented with the dotted line. The gray area represents the pooling interval.

be controlled by using rational error modeling methods [53] or by setting safety factors in process optimization [45]. Furthermore, the model uncertainty and boundaries can be further explored through Bayesian inference using Monte Carlo methods [44,54], so that the robust optimization process can be designed and developed to ensure that the product has a predefined quality at the end of the manufacturing process.

5. Conclusions

In this study, a PbP-HIC method for multi-component system (named as the mPbP-HIC method) was derived. Four parameters ($\boldsymbol{k}_{s,i}$, $\boldsymbol{k}_{eq,i}$, $\boldsymbol{n}_i$, and $\boldsymbol{q}_{max,i}$) of component $\boldsymbol{i}$ would be estimated by the linear regression and linear approximation steps, where the parameter $\boldsymbol{q}_{max,i}$ would be determined by solving a matrix. The remaining two parameters ($\boldsymbol{k}_{p,i}$ and $\boldsymbol{k}_{kin,i}$) would be calibrated by the inverse method. The experimental validation was carried out by a two-component system containing lysozyme and α-Chymotrypsinogen A. The results indicated that the nonlinear parameters ($\boldsymbol{n}$ and $\boldsymbol{q}_{max}$) obtained from the linear approximation step potentially had large errors, especially for the parameter $\boldsymbol{q}_{max}$, which resulted in a poor curve fitting at high loading. Therefore, the inverse method was introduced to further determine the $\boldsymbol{q}_{max}$ to reduce the estimation error. Moreover, considering the possible problems of matrix unsolvable and complicated matrix solving in the linear approximation step, the simplified linear approximation (SLA) was further proposed by reasonable assumption. It implies that the parameter $\boldsymbol{q}_{max}$ can be obtained directly by SLA without solving any complex linear system. Based on the above improvements, the $\boldsymbol{q}_{max}$ could be initialized by SLA and finally determined by the inverse method, and this straightforward strategy (SLA+IM) to determine the parameter $\boldsymbol{q}_{max}$ would be applied in the improved mPbP-HIC method. The experimental validation showed that the improved mPbP-HIC method had better elution curve fitting than the classical mPbP-HIC method under undiluted condition (Runs 4, 5, 6), with the average $\boldsymbol{L}^2$-error reduced from 0.427 to 0.346. Finally, the estimated parameters were validated further by the process optimization both within and outside the model calibration range. The optimization experiments verified that the parameters provided by the improved mPbP-HIC method are more reasonable, enabling the model with a good predictive ability and allow reasonable extrapolation. The results demonstrated that the improved mPbP-HIC method can be easily applied to practical separation, and can accelerate the development of HIC process in the downstream of biopharmaceuticals with savings in experimental costs.

CRediT authorship contribution statement

Yu-Xiang Yang: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Zhi-Yuan Lin:** Writing – review & editing. **Yu-Cheng Chen:** Writing – review & editing. **Shan-Jing Yao:** Writing – review & editing. **Dong-Qiang Lin:** Writing – review & editing, Supervision, Resources, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This work was supported by the Huadong Medicine Joint Funds of the Zhejiang Provincial Natural Science Foundation of China (LHDMZ24B060001), National Natural Science Foundation of China (22078286) and Zhejiang Key Science and Technology Project (2023C03116). The authors would also like to thank the Zhejiang University Information Technology Center for the cloud computing service.

References

[1] J.T. McCue, P. Engel, A. Ng, R. Macniven, J. Thömmes, Modeling of protein monomer/aggregate purification and separation using hydrophobic interaction chromatography, Bioprocess Biosyst. Eng. 31 (2008) 261–275, https://doi.org/10.1007/s00449-008-0200-1.
[2] 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.
[3] I.R.A. Pereira Bresolin, N. Lingg, I.T.L. Bresolin, A. Jungbauer, Hydrophobic interaction chromatography as polishing step enables obtaining ultra-pure recombinant antibodies, J. Biotechnol. 324 (2020) 100020, https://doi.org/10.1016/j.btecx.2020.100020.
[4] R.C. Barrientos, G.L. Losacco, M. Azizi, H. Wang, A.N. Nguyen, V. Shchurik, A. Singh, D. Richardson, I. Mangion, D. Guillarme, E.L. Regalado, I.A. Haidar Ahmad, Automated hydrophobic interaction chromatography screening combined with in silico optimization as a framework for nondenaturing analysis and

purification of biopharmaceuticals, Anal. Chem. 94 (2022) 17131–17141, https://doi.org/10.1021/acs.analchem.2c03453.
[5] D. Keulen, G. Geldhof, O.L. Bussy, M. Pabst, M. Ottens, Recent advances to accelerate purification process development: a review with a focus on vaccines, J. Chromatogr. A 1676 (2022) 463195, https://doi.org/10.1016/j.chroma.2022.463195.
[6] C.R. Bernau, M. Knödler, J. Emonts, R.C. Jäpel, J.F. Buyel, The use of predictive models to develop chromatography-based purification processes, Front. Bioeng. Biotechnol. 10 (2022), https://doi.org/10.3389/fbioe.2022.1009102.
[7] L.K. Shekhawat, M. Chandak, A.S. Rathore, Mechanistic modeling of hydrophobic interaction chromatography for monoclonal antibody purification: process optimization in the quality by design paradigm, J. Chem. Technol. Biotechnol. 92 (2017) 2527–2537, https://doi.org/10.1002/jctb.5324.
[8] B.K. Nfor, P.D.E.M. Verhaert, L.A.M. van der Wielen, J. Hubbuch, M. Ottens, Rational and systematic protein purification process development: the next generation, Trends Biotechnol. 27 (2009) 673–679, https://doi.org/10.1016/j.tibtech.2009.09.002.
[9] E. Lietta, A. Pieri, A.G. Cardillo, M. Vanni, R. Pisano, A.A. Barresi, An experimental and modeling combined approach in preparative hydrophobic interaction chromatography, Processes 10 (5) (2022), https://doi.org/10.3390/pr10051027.
[10] H. Narayanan, M. Sponchioni, M. Morbidelli, Integration and digitalization in the manufacturing of therapeutic proteins, Chem. Eng. Sci. 248 (2022) 117159, https://doi.org/10.1016/j.ces.2021.117159.
[11] L.K. Shekhawat, A.S. Rathore, An overview of mechanistic modeling of liquid chromatography, Prep. Biochem. Biotechnol. 49 (2019) 623–638, https://doi.org/10.1080/10826068.2019.1615504.
[12] J.M. Mollerup, Applied thermodynamics: a new frontier for biotechnology, Fluid Phase Equilib. 241 (2006) 205–215, https://doi.org/10.1016/j.fluid.2005.12.037.
[13] 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.
[14] J.M. Mollerup, T.B. Hansen, S. Kidal, A. Staby, Quality by design—Thermodynamic modelling of chromatographic separation of proteins, J. Chromatogr. A 1177 (2008) 200–206, https://doi.org/10.1016/j.chroma.2007.08.059.
[15] 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.
[16] W. Heymann, J. Glaser, F. Schlegel, W. Johnson, P. Rolandi, E. von Lieres, Advanced score system and automated search strategies for parameter estimation in mechanistic chromatography modeling, J. Chromatogr. A 1661 (2022) 462693, https://doi.org/10.1016/j.chroma.2021.462693.
[17] K. Meyer, M. Søes Ibsen, L. Vetter-Joss, E. Broberg Hansen, J. Abildskov, Industrial ion-exchange chromatography development using discontinuous Galerkin methods coupled with forward sensitivity analysis, J. Chromatogr. A 1689 (2023) 463741, https://doi.org/10.1016/j.chroma.2022.463741.
[18] D. Åsberg, M. Leśko, M. Enmark, J. Samuelsson, K. Kaczmarski, T. Fornstedt, Fast estimation of adsorption isotherm parameters in gradient elution preparative liquid chromatography II: the competitive case, J. Chromatogr. A 1314 (2013) 70–76, https://doi.org/10.1016/j.chroma.2013.09.003.
[19] 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.
[20] D.C. López C, T. Barz, S. Körkel, G. Wozny, Nonlinear ill-posed problem analysis in model-based parameter estimation and experimental design, Comput. Chem. Eng. 77 (2015) 24–42, https://doi.org/10.1016/j.compchemeng.2015.03.002.
[21] A.N. Prybutok, J.Y. Cain, J.N. Leonard, N. Bagheri, Fighting fire with fire: deploying complexity in computational modeling to effectively characterize complex biological systems, Curr. Opin. Biotechnol. 75 (2022) 102704, https://doi.org/10.1016/j.copbio.2022.102704.
[22] V. Rajamanickam, H. Babel, L. Montano-Herrera, A. Ehsani, F. Stiefel, S. Haider, B. Presser, B. Knapp, About model validation in bioprocessing, Processes 9 (6) (2021), https://doi.org/10.3390/pr9060961.
[23] D. Degenring, M. Röhl, A.M. Uhrmacher, Discrete event, multi-level simulation of metabolite channeling, Biosystems 75 (2004) 29–41, https://doi.org/10.1016/j.biosystems.2004.03.008.
[24] K.A.P. McLean, K.B. McAuley, Mathematical modelling of chemical processes—Obtaining the best model predictions and parameter estimates using identifiability and estimability procedures, Can. J. Chem. Eng. 90 (2012) 351–366, https://doi.org/10.1002/cjce.20660.
[25] 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.
[26] 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.
[27] 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.
[28] 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.
[29] D. Saleh, R. Hess, M. Ahlers-Hesse, N. Beckert, M. Schönberger, F. Rischawy, G. Wang, J. Bauer, M. Blech, S. Kluters, J. Studts, J. Hubbuch, Modeling the impact of amino acid substitution in a monoclonal antibody on cation exchange chromatography, Biotechnol. Bioeng. 118 (2021) 2923–2933, https://doi.org/10.1002/bit.27798.
[30] Y.C. Chen, S.J. Yao, D.Q. Lin, Parameter-by-parameter method for steric mass action model of ion exchange chromatography: theoretical considerations and experimental verification, J. Chromatogr. A 1680 (2022) 463418, https://doi.org/10.1016/j.chroma.2022.463418.
[31] 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.
[32] Y.C. Chen, S.J. Yao, D.Q. Lin, Parameter-by-parameter method for steric mass action model of ion exchange chromatography: simplified estimation for steric shielding factor, J. Chromatogr. A 1687 (2023) 463655, https://doi.org/10.1016/j.chroma.2022.463655.
[33] Y.X. Yang, Y.C. Chen, S.J. Yao, D.Q. Lin, Parameter-by-parameter estimation method for adsorption isotherm in hydrophobic interaction chromatography, J. Chromatogr. A 1716 (2024) 464638, https://doi.org/10.1016/j.chroma.2024.464638.
[34] A. Felinger, G. Guiochon, Comparison of the kinetic models of linear chromatography, Chromatographia 60 (2004), https://doi.org/10.1365/s10337-004-0288-7. S175-S180.
[35] P.V. Danckwerts, Continuous flow systems: distribution of residence times, Chem. Eng. Sci. 2 (1953) 1–13, https://doi.org/10.1016/0009-2509(53)80001-1.
[36] L. Pedersen, J. Mollerup, E. Hansen, A. Jungbauer, Whey proteins as a model system for chromatographic separation of proteins, J. Chromatogr. B 790 (2003) 161–173, https://doi.org/10.1016/S1570-0232(03)00127-2.
[37] S. Leweke, E. von Lieres, Chromatography analysis and design toolkit (CADET), Comput. Chem. Eng. 113 (2018) 274–294, https://doi.org/10.1016/j.compchemeng.2018.02.025.
[38] J. Schmölder, M. Kaspereit, A modular framework for the modelling and optimization of advanced chromatographic processes, Processes 8 (1) (2020) 65, https://doi.org/10.3390/pr8010065.
[39] 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, I. Henriksen, E.A. Quintero, C.R. Harris, A.M. Archibald, A.H. Ribeiro, F. Pedregosa, P. van Mulbregt, A.P. Bardelli, A. Rothberg, A. Hilboll, 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, G. Young, G.A. Price, G.L. Ingold, 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 Miranda Cardoso, J. Reimer, J. Harrington, J.L. C. Rodríguez, J. Nunez-Iglesias, J. Kuczynski, M. Thoma, M. Bolingbroke, M. Tartre, N.J. Smith, N. Nowaczyk, N. Shebanov, O. Pavlyk, P.A. Brodtkorb, 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, C. SciPy, L.A.N.M. Los Alamos National Laboratory, 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.
[40] C.R. Bernau, R.C. Jäpel, J.W. Hübbers, S. Nölting, P. Opdensteinen, J.F. Buyel, Precision analysis for the determination of steric mass action parameters using eight tobacco host cell proteins, J. Chromatogr. A 1652 (2021) 462379, https://doi.org/10.1016/j.chroma.2021.462379.
[41] J. Morgenstern, G. Wang, P. Baumann, J. Hubbuch, Model-based investigation on the mass transfer and adsorption mechanisms of mono-pegylated lysozyme in ion-exchange chromatography, Biotechnol. J. 12 (2017) 1700255, https://doi.org/10.1002/biot.201700255.
[42] Y.C. Chen, H.L. Lu, R.Z. Wang, G. Sun, X.Q. Zhang, J.Q. Liang, A. Jungbauer, S. J. Yao, D.Q. Lin, Standardized approach for accurate and reliable model development of ion-exchange chromatography based on parameter-by-parameter method and consideration of extra-column effects, Biotechnol. J. 19 (2024) 2300687, https://doi.org/10.1002/biot.202300687.
[43] T.C. Huuk, T. Hahn, K. Doninger, J. Griesbach, S. Hepbildikler, J. Hubbuch, Modeling of complex antibody elution behavior under high protein load densities in ion exchange chromatography using an asymmetric activity coefficient, Biotechnol. J. 12 (2017) 1600336, https://doi.org/10.1002/biot.201600336.
[44] 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.
[45] C. Ding, C. Gerberich, M. Ierapetritou, Hybrid model development for parameter estimation and process optimization of hydrophobic interaction chromatography, J. Chromatogr. A 1703 (2023) 464113, https://doi.org/10.1016/j.chroma.2023.464113.
[46] L. Zhang, J. Selker, A. Qu, A. Velayudhan, Numerical estimation of multicomponent adsorption isotherms in preparative chromatography: implications of experimental error, J. Chromatogr. A 934 (2001) 13–29, https://doi.org/10.1016/S0021-9673(01)01297-3.

[47] N. Borg, K. Westerberg, N. Andersson, E. von Lieres, B. Nilsson, Effects of uncertainties in experimental conditions on the estimation of adsorption model parameters in preparative chromatography, Comput. Chem. Eng. 55 (2013) 148–157, https://doi.org/10.1016/j.compchemeng.2013.04.013.
[48] V. Kumar, S. Leweke, E. von Lieres, A.S. Rathore, Mechanistic modeling of ion-exchange process chromatography of charge variants of monoclonal antibody products, J. Chromatogr. A 1426 (2015) 140–153, https://doi.org/10.1016/j.chroma.2015.11.062.
[49] J.M. Breuer, S. Leweke, J. Schmölder, G. Gassner, E. von Lieres, Spatial discontinuous Galerkin spectral element method for a family of chromatography models in CADET, Comput. Chem. Eng. (2023) 108340, https://doi.org/10.1016/j.compchemeng.2023.108340.
[50] R.C. Jäpel, J.F. Buyel, Bayesian optimization using multiple directional objective functions allows the rapid inverse fitting of parameters for chromatography simulations, J. Chromatogr. A 1679 (2022) 463408, https://doi.org/10.1016/j.chroma.2022.463408.
[51] C. Ding, M. Ierapetritou, Machine learning-based optimization of a multi-step ion exchange chromatography for ternary protein separation, Comput. Chem. Eng. 184 (2024) 108642, https://doi.org/10.1016/j.compchemeng.2024.108642.
[52] A. Creasy, J. Lomino, G. Carta, Gradient elution behavior of proteins in hydrophobic interaction chromatography with a U-shaped retention factor curve under overloaded conditions, J. Chromatogr. A 1578 (2018) 28–34, https://doi.org/10.1016/j.chroma.2018.10.003.
[53] W. Heymann, J. Glaser, F. Schlegel, W. Johnson, P. Rolandi, E. von Lieres, Advanced error modeling and Bayesian uncertainty quantification in mechanistic liquid chromatography modeling, J. Chromatogr. A 1708 (2023) 464329, https://doi.org/10.1016/j.chroma.2023.464329.
[54] Y. Yamamoto, T. Yajima, Y. Kawajiri, Uncertainty quantification for chromatography model parameters by Bayesian inference using sequential Monte Carlo method, Chem. Eng. Res. Des. 175 (2021) 223–237, https://doi.org/10.1016/j.cherd.2021.09.003.