Journal of Chromatography A 1716 (2024) 464638

Contents lists available at ScienceDirect

Journal of Chromatography A

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

Parameter-by-parameter estimation method for adsorption isotherm in hydrophobic interaction chromatography

Yu-Xiang Yang, Yu-Cheng Chen, Shan-Jing Yao, Dong-Qiang Lin *

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

ARTICLE INFO

Keywords:
Hydrophobic interaction chromatography
Parameter estimation
Mechanistic model
Mollerup isotherm

ABSTRACT

Hydrophobic interaction chromatography (HIC) is used as a critical polishing step in the downstream processing of biopharmaceuticals. Normally the process development of HIC is a cumbersome and time-consuming task, and the mechanical models can provide a powerful tool to characterize the process, assist process design and accelerate process development. However, the current estimation of model parameters relies on the inverse method, which lacks an efficient and logical parameter estimation strategy. In this study, a parameter-by-parameter (PbP) method based on the theoretical derivation and simplifying assumptions was proposed to estimate the Mollerup isotherm parameters for HIC. The method involves three key steps: (1) linear regression (LR) to estimate the salt-protein interaction parameter and the equilibrium constant; (2) linear approximation (LA) to estimate the stoichiometric parameter and the maximum binding capacity; and (3) inverse method to estimate the protein-protein interaction parameter and the kinetic coefficient. The results indicated that the LR step should be used for dilution condition (loading factor below 5%), while the LA step should be conducted when the isotherm is in the transition or nonlinear regions. Six numerical experiments were conducted to implement the PbP method. The results demonstrated that the PbP method developed allows for the systematic estimation of HIC parameters one-by-one, effectively reducing the number of parameters required for inverse method estimation from six to two. This helps prevent non-identifiability of structural parameters. The feasibility of the PbP-HIC method was further validated by real-world experiments. Moreover, the PbP method enhances the mechanistic understanding of adsorption behavior of HIC and shows a promising application to other stoichiometric displacement model-derived isotherms.

1. Introduction

Chromatography stands as the important method for downstream processing of biopharmaceuticals. Among them, hydrophobic interaction chromatography (HIC) is normally used as a polishing step [1–4], known by its excellent capability to remove aggregates [1,5]. Conventional downstream process development and optimization are characterized by their lengthy and costly nature, heavy reliance on the experimentation [6]. In contrast, mechanistic modeling of chromatographic processes rooted in physicochemical principles has been used to assist the process development, such as ion exchange chromatography (IEC) and HIC [4,7,8]. This approach allows for the rational characterization of chromatographic processes [9,10], optimization of operation conditions [11,12] and facilitates the identification of experimental deviations attributed to process variations [13]. Notably, the increasing popularity of model-based tools paves the way for the development of digital twins in the biopharmaceuticals production process [14,15].

Adsorption isotherm models based on the stoichiometric displacement have been widely used in chromatographic processes such as IEC and HIC [4,7]. A representation of the salt-dependent stoichiometric displacement model (SDM) is the steric-mass action (SMA) isotherm [16]. It is a four-parameter model (characteristic charge ν, equilibrium coefficient k_{eq}, shielding factor σ, kinetic coefficient k_{kin}) that elucidates the steric hindrance effects in protein adsorption [17,18]. The above-mentioned salt-dependent adsorption and steric hindrance are also considered in the HIC adsorption isotherm proposed by Mollerup [19,20], which is also based on the stoichiometric displacement process and possesses an outstanding capacity to characterize nonlinear adsorption [10]. The Mollerup isotherm model consists of six parameters, including stoichiometric parameter (number of ligands bound per

* Corresponding author at: College of Chemical and Biological Engineering, Zhejiang University, Hangzhou 310058, China.
E-mail address: lindq@zju.edu.cn (D.-Q. Lin).

https://doi.org/10.1016/j.chroma.2024.464638
Received 21 November 2023; Received in revised form 6 January 2024; Accepted 8 January 2024
Available online 9 January 2024

protein) 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}. Compared to the SMA model, the dependence of salt concentration and protein concentration in the HIC Mollerup isotherm is described by an activity coefficient, leading to two additional parameters (k_s and k_p). Currently, the estimation of Mollerup isotherm parameters lacks an efficient and rational method, relying on the inverse method (IM) [10,11]. The implementation of inverse method involves fitting the simulated data to the elution curves through a well-defined optimization problem, which in turn depends on a reasonable optimization algorithm and initial guesses. Additionally, the practical application of inverse method is time-consuming and often encounters the challenge of local optima trap [21–23], which hinders the application of mechanistic model. To address these issues, it is essential to develop a new and effective method for parameter estimation, particularly for the HIC Mollerup isotherm with numerous parameters.

The Yamamoto method, derived from the retention model, has been successfully used for parameter estimation in both IEC and mixed-model chromatography (MMC) [17,24-26]. This method can efficiently calculate two SMA parameters (ν and k_{eq}) or two MMC parameters (k_s and k_{eq}) through several linear gradient elution experiments (LGEs). However, it was reported that the Yamamoto method tends to be more effective at low loadings [27,28]. Recently, Chen et al. [29,30] proposed an efficient estimation strategy named the parameter-by-parameter (PbP) method for the SMA model (referred as PbP-IEC), which is based on the retention model utilizing LGEs under different conditions to estimate the three SMA parameters (ν, k_{eq} and σ) one-by-one. In contrast to the Yamamoto method, the PbP-IEC method can provide more reliable parameter estimation, particularly under undiluted conditions. The PbP-IEC method is derived from mechanistic models (retention mechanism and stoichiometric displacement), and can be equally applicable to the HIC adsorption isotherm, which is based on the same underlying mechanism.

Compared to the PbP-IEC method published previously [29,30], the HIC isotherm has more parameters with different physical meanings, and the splitting of the retention model will be more challenging. In this study, the PbP method would be extended from IEC to HIC, new parameter-by-parameter method for HIC (referred as PbP-HIC) would be proposed based on the retention theory and reasonable simplifying assumptions for parameter estimation of Mollerup isotherm. Numerical experiments with different elution gradient lengths and sample loadings would be conducted to validate and evaluate the PbP-HIC method. For Mollerup isotherm, four parameters (k_s, k_{eq}, n, and q_{max}) would be estimated one-by-one by the PbP-HIC method, while the remaining two parameters (k_p and k_{kin}) would be calibrated by inverse method. Additionally, the applicability range of the proposed PbP-HIC method would be investigated by evaluating the effects of operating conditions on the parameter estimation. Finally, an experimental strategy would be proposed for the practical implementation of the PbP-HIC method.

2. Theory section

2.1. Chromatographic column model

The mass transfer within the chromatographic column is described using the lumped rate model (LRM), which is upon the assumption of rapid pore diffusion [31]. This simplification allows for the consideration of the average pore concentration, c_p. Consequently, the contributions of the film transfer and pore diffusion can be expressed by the effective film diffusion coefficient, k_{eff}.

The LRM is given by:

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

$$\frac{\partial c_p}{\partial t}+\frac{1-\varepsilon_p}{\varepsilon_p}\frac{\partial q}{\partial t}=\frac{3}{r_p}\frac{k_{eff}}{\varepsilon_p}\left(c-c_p\right), \tag{2}$$

where c and c_p represent the mobile-phase concentration in the interparticle volume and intraparticle volume (mol/L), respectively. q denotes the protein concentration in the stationary (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 are the column porosity and particle porosity, respectively. D_{ax} accounts for the axial diffusion coefficient (m^2/s).

The rectangular pulse injection condition is applied as:

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

where c_{inj} represents the loading concentration of the protein (mol/L) and t_{inj} is the loading time (s). The Danckwerts and Neumann boundary conditions are respectively applied at the column inlet and outlet [32] as following:

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

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

2.2. Adsorption isotherm model

The HIC adsorption isotherm developed by Mollerup [19,20,33] is used in this study and described as:

$$k_{kin}\frac{\partial q}{\partial t}=c_p k_{eq}\left(1-\frac{q}{q_{max}}\right)^n\exp\left(k_s c_s+k_p c_p\right)-q, \tag{6}$$

where k_{kin}, k_{eq}, q_{max}, n, k_s, k_p are six parameters, representing kinetic coefficient, equilibrium coefficient, maximum binding capacity, stoichiometric parameter (number of ligands bound per protein), salt-protein interaction parameter, and protein-protein interaction parameter, respectively. c_s represents the salt concentration in the mobile phase (mol/L). k_{kin} is the reciprocal of the desorption constant k_{des}, and it significantly affects the peak width. k_{eq} is the ratio of the adsorption constants k_{ads} to k_{des}, and it significantly influences the retention time of elution peak [34]. k_s and k_p are parameters describing the effects of salt concentration and protein concentration on the activity coefficient, respectively. The nonlinear term $(1-q/q_{max})^n$ accounts for the effect of steric hindrance, which has a great role in the description of the elution process at high loading.

At the adsorption equilibrium, Eq. (6) can be rewritten as:

$$\frac{q}{c_p}=k_{eq}\left(1-\frac{q}{q_{max}}\right)^n\exp\left(k_s c_s+k_p c_p\right). \tag{7}$$

Furthermore, compared to other parameters, the importance of protein-protein interaction is recognized to be relatively minor [26,35,36], allowing k_p to be approximated as zero. So, Eq. (7) can be simplified as:

$$\frac{q}{c_p}=k_{eq}\left(1-\frac{q}{q_{max}}\right)^n\exp(k_s c_s). \tag{8}$$

For low loading ($q\leqslant q_{max}$), the influence of steric hindrance can be negligible, and consequently $(1-q/q_{max})^n$ can be approximated as one [19,33]. Hence, Eq. (8) can be reformulated into a linear adsorption model as follows:

$$\frac{q}{c_p}=k_{eq}\exp(k_s c_s). \tag{9}$$

Based on the above simplifications, the parameters can be classified

into two linear parameters (k_s and k_{eq}) and two nonlinear parameters (n and q_{max}).

2.3. Parameter-by-parameter method

The core of the PbP-HIC method consists of the linearization of the nonlinear term and the retention mechanism, which have been successfully applied in the derivation of the PbP-IEC method as published previously [29].

For the HIC Mollerup isotherm model, the nonlinear term is $(1-q/q_{max})^n$. When $q_{max} = \Lambda/(\nu+\sigma)$, the nonlinear terms of the Mollerup isotherm and SMA model are equal to each other. This suggests that the linearization method might be judiciously extended from the SMA model to the Mollerup isotherm.

According to the retention mechanism, the retention factor k' can be given by [37]:

$$k'(t) = \frac{1-\varepsilon_t}{\varepsilon_t}\frac{q(t,L)}{c_p(t,L)}. \tag{10}$$

2.3.1. Linear regression

At low loading, in conjunction with Eq. (9), Eq. (10) can be expressed as:

$$k'(t) = \frac{1-\varepsilon_t}{\varepsilon_t}\cdot k_{eq}\cdot\exp(k_s c_s). \tag{11}$$

For elution under a linear gradient, Eq. (11) can be related to the integral form as [28,38,39]:

$$\int_0^{V_R}\frac{dV(t)}{k'(t)\cdot V_{col}\cdot\varepsilon_t} = 1, \tag{12}$$

$$V(t) = \frac{(1-\varepsilon_t)\cdot\left[c_s(t,L)-c_{s,initial}\right]}{GH}\cdot V_{col}, \tag{13}$$

$$GH = \frac{(1-\varepsilon_t)\left(c_{s,final}-c_{s,initial}\right)}{CV_G}, \tag{14}$$

where V_R is the retention volume. GH denotes the normalized gradient slope (mol/L). CV_G represents the length of the elution gradient in terms of column volume. $c_{s,initial}$ and $c_{s,final}$ correspond to the salt concentrations at the start and end of the gradient (mol/L), respectively.

Combining Eqs. (11) - (14), Eq. (12) can be rewritten as:

$$\int_{c_{s,initial}}^{c_{s,R}}\frac{dc_s}{\exp(k_s c_s)} = GH\cdot k_{eq}, \tag{15}$$

where $c_{s,R}$ describes the salt concentration at the peak maximum of the eluted protein. Eq. (15) can be integrated and rewritten as:

$$\frac{1}{k_s}\left[\exp(-k_s c_{s,R}) - \exp(-k_s c_{s,initial})\right] = -GH\cdot k_{eq}. \tag{16}$$

The term $\exp(-k_s c_{s,initial})$ in Eq. (16) can be neglected if the condition of $\exp(-k_s c_{s,R})/\exp(-k_s c_{s,initial}) \gg 1$ is satisfied, and Eq. (16) can be rewritten as the logarithm form as:

$$\ln(-GH) = -k_s c_{s,R} - \ln(k_s\cdot k_{eq}), \tag{17}$$

where k_s and k_{eq} are undetermined parameters. Eq. (17) can be reformulated with the slope $m_1 = -k_s$ and the intercept $p_1 = -\ln(k_s\cdot k_{eq})$:

$$\ln(-GH) = m_1\cdot c_{s,R} + p_1. \tag{18}$$

In this study, Eq. (18) is named as the linear regression (LR) of the PbP-HIC method.

2.3.2. Linear approximation

Under undiluted conditions, the isotherms exhibit nonlinearity. Combining Eqs. (8) and (10), k' can be expressed as:

$$k'(t) = \frac{1-\varepsilon_t}{\varepsilon_t}\cdot k_{eq}\cdot\exp(k_s c_s)\left(1-\frac{q}{q_{max}}\right)^n. \tag{19}$$

For the nonlinear term $(1-q/q_{max})^n$, the key aspect of the PbP-HIC method is the linearization of this term. Specifically, q at the retention time (t_R) can be expressed using the Henry coefficient $H = k_{eq}\exp(k_s c_s)$ [19,40] as:

$$q(t_R) \approx H(t_R)\cdot c_p(t_R) = \frac{H(t_R)\cdot c_p(t_R)}{\bar{c}_{inj}(t_{inj})}\cdot\bar{c}_{inj}(t_{inj}) = g(t_R)\cdot\bar{c}_{inj}(t_{inj}), \tag{20}$$

where $\bar{c}_{inj} = \frac{c_{inj}\cdot u_{int}\cdot\varepsilon_c\cdot t_{inj}}{L}$ is the relative loading concentration. Therefore, the nonlinear term can be depicted using the nonlinear coefficient $a = g/q_{max}$ as:

$$\left[1-\frac{q(t_R)}{q_{max}}\right]^n \approx \left[1-\frac{g(t_R)\cdot\bar{c}_{inj}(t_{inj})}{q_{max}}\right]^n = \left[1-a(t_R)\cdot\bar{c}_{inj}(t_{inj})\right]^n. \tag{21}$$

Utilizing Eqs. (19) and (21), Eq. (12) and its integral form can be rewritten as:

$$\int_{c_{s,initial}}^{c_{s,R}}\frac{dc_s}{\exp(k_s c_s)} = GH\cdot k_{eq}\cdot(1-a\cdot\bar{c}_{inj})^n, \tag{22}$$

$$\frac{1}{-GH\cdot k_s\cdot k_{eq}}\left[\exp(-k_s c_{s,R}) - \exp(-k_s c_{s,initial})\right] = (1-a\cdot\bar{c}_{inj})^n, \tag{23}$$

where n and a denote undetermined parameters, respectively. The left-hand term can be defined as:

$$y = \frac{1}{-GH\cdot k_s\cdot k_{eq}}\left[\exp(-k_s c_{s,R}) - \exp(-k_s c_{s,initial})\right], \tag{24}$$

and Eq. (23) can be reformulated as the logarithm form as:

$$\ln y = n\cdot\ln(1-a\cdot\bar{c}_{inj}). \tag{25}$$

When $a\cdot\bar{c}_{inj} \ll 1$, Eq. (25) can be simplified using linear approximation as the expression as:

$$\ln y \approx -a\cdot n\cdot\bar{c}_{inj}. \tag{26}$$

The slope of Eq. (26) is $m_2 = -an$, and integrating it into Eq. (25) gives:

$$\ln y = n\cdot\ln\left(1+\frac{m_2\cdot\bar{c}_{inj}}{n}\right). \tag{27}$$

The above derivation is named as the linear approximation (LA) of the PbP-HIC method.

2.4. Parameter estimation strategy

(1) Estimating k_s and k_{eq} with linear regression

The parameters k_s and k_{eq} in the Mollerup isotherm exert influence on the retention time. Eq. (18) shows a linear relationship between $c_{s,R}$ and $\ln(-GH)$. Correlating $c_{s,R}$ with $\ln(-GH)$ and using a series of LGEs with different elution gradient lengths at low loadings, k_s can be determined as:

$$k_s = -m_1. \tag{28}$$

According to Eq. (28) and the intercept of Eq. (18), k_{eq} can be determined as:

$$k_{eq} = \frac{-\exp(p_1)}{k_s}. \tag{29}$$

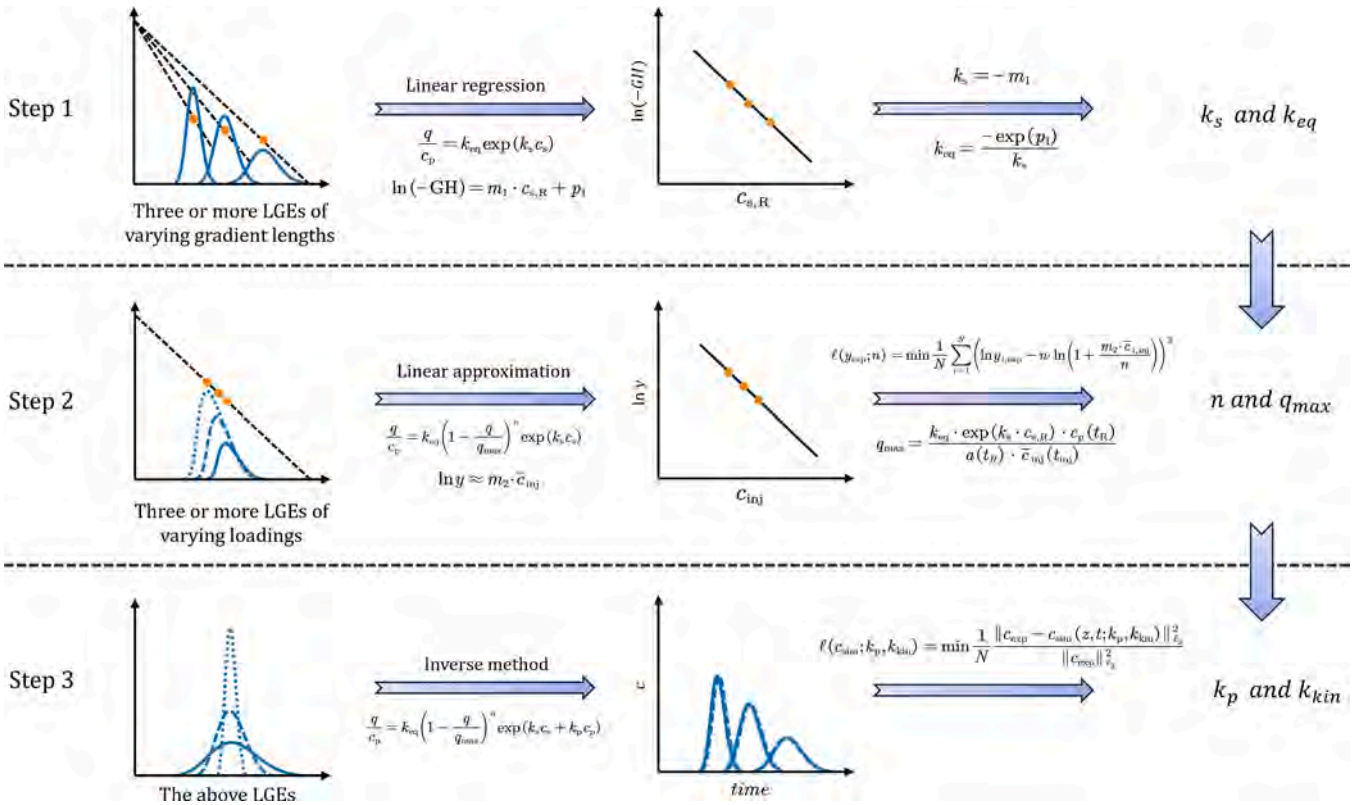

Fig. 1. Flowchart of parameter estimation based on the PbP-HIC method.

(2) Estimating n and q_{max} with linear approximation

The parameters n and q_{max} in the Mollerup isotherm describe the nonlinear behavior. $\bar{c}_{inj}$ characterizes the loading conditions of LGEs and $c_{s,R}$ is affected by the loading conditions. If GH remains constant, $\bar{c}_{inj}$ and y are the independent and dependent variables, respectively. y can be calculated by Eq. (24) using k_s and k_{eq} estimated as mentioned above. Eq. (26) can be rewritten using $m_2 = -a \cdot n$ as:

$$\ln y \approx m_2 \cdot \bar{c}_{inj}. \tag{30}$$

When multiple LGEs with different loadings while consistent elution gradient lengths are performed, using $\bar{c}_{inj}$ and $\ln y$, m_2 can be obtained, which can be applied to figure out n based on $\bar{c}_{inj}$, $\ln y$, and Eq. (27). However, the equation cannot be solved explicitly. Therefore, n can be determined by defining a loss function and minimizing it, as expressed as:

$$\ell(y; n) = \min \frac{1}{N}\sum_{i=1}^{N}\left(\ln y_i - n \cdot \ln\left(1 + \frac{m_2 \cdot \bar{c}_{inj,i}}{n}\right)\right)^2, \tag{31}$$

where N represents the total number of LGEs. Furthermore, a is reformulated as:

$$a = -\frac{m_2}{n}. \tag{32}$$

By combining Eqs. (20), (21), and (32), q_{max} can be obtained as:

$$q_{max} = \frac{k_{eq} \cdot \exp(k_s \cdot c_{s,R}) \cdot c_p(t_R)}{a(t_R) \cdot \bar{c}_{inj}(t_{inj})}. \tag{33}$$

(3) Estimating k_p and k_{kin} with inverse method

The PbP-HIC method for fitting model parameters relies on the retention times. However, in comparison to the salt-protein interactions, the protein-protein interactions were reported to have a limited effect on the asymmetric activity coefficients, resulting in a relatively minor effect on the retention times [26,35,36]. Therefore, the parameter k_p is not taken into account in the PbP-HIC method. Additionally, the parameter k_{kin} is not explicitly shown in Eq. (7). Consequently, the direct estimation of k_p and k_{kin} presents significant challenges. To address this, the inverse method has to be used to estimate the parameters k_p and k_{kin} through the experimental elution curves. The loss function for this purpose is defined as follows:

$$\ell(c_{sim}; k_p, k_{kin}) = \min \frac{1}{N}\frac{\| c_{exp} - c_{sim}(z,t;k_p,k_{kin})\|^2_{\ell_2}}{\| c_{exp}\|^2_{\ell_2}}, \tag{34}$$

where N denotes the number of all calibration experiments, c_{exp} and c_{sim} represent experimental and numerical approximations of protein concentration, respectively. $\| \cdot \|_{\ell_2}$ represents the L^2-norm. Minimizing the loss function allows us to estimate k_p and k_{kin}.

2.5. Parameters estimation procedure

Based on the above derivation, the PbP-HIC method can be applied to estimate the Mollerup isotherm parameters in a step-by-step way with a series of LGEs conducted under different elution gradient lengths and loading conditions. Specifically, the procedure involves the following three steps:

Step 1: Estimating k_s and k_{eq} through LR using multiple LGEs with varying elution gradient lengths while maintaining consistent loadings.
Step 2: Estimating n and q_{max} through LA using multiple LGEs with different loading conditions while maintaining consistent gradient lengths.

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

Parameter	Symbol	Unit	Numerical experiment	Real-world experiment
Column length	L	mm	100.0	100.0
Column volume	V	mL	4.65	4.65
Flow rate	u	mm/s	0.417	0.417
Column porosity	ε_c	–	0.37	0.37
Particle porosity	ε_p	–	0.93	0.97
Particle radius	r_p	μm	17.0	45.0
Axial dispersion coefficient	D_{ax}	m^2/s	9.13×10^{-7}	1.50×10^{-7}
Effective film diffusion coefficient	k_{eff}	m/s	2.59×10^{-5}	1.50×10^{-5}
Salt effective film diffusion coefficient	$k_{eff,s}$	m/s	1.29×10^{-4}	1.50×10^{-5}

Table 2
Mollerup isotherm parameters.

Parameter	BSA	Unit
k_s	3.95	$(kmol/m^3)^{-1}$
k_{eq}	13.1	–
n	9.85	–
q_{max}	0.0376	$kmol/m^3$
k_p	1.00×10^{-6}	$(kmol/m^3)^{-1}$
k_{kin}	100	s

Table 3
HIC runs for numerical calibration experiments.

Run	Purpose	CV_G (CV)	LF (%)	$c_{s,initial}$ (mol/L)
1–1	LR	10	0.5	1.2
1–2	LR	20	0.5	1.2
1–3	LR	30	0.5	1.2
1–4	LA	30	6.0	1.2
1–5	LA	30	7.0	1.2
1–6	LA	30	8.0	1.2
2–1	LR	10	0.5	1.4
2–2	LR	20	0.5	1.4
2–3	LR	30	0.5	1.4
2–4	LA	30	4.0	1.4
2–5	LA	30	4.5	1.4
2–6	LA	30	5.0	1.4

Step 3: Estimating of k_p and k_{kin} with the inverse method using the elution curves of LGEs in Step 1 and Step 2.

The flowchart illustrating the PbP-HIC method for the Mollerup isotherm parameter estimation is given in Fig. 1.

3. Materials and methods

3.1. Numerical experiments

The HIC numerical experiments were performed involving on bovine serum albumin (BSA, 66.5 kDa) separation with Butyl Sepharose HP resin, based on the results reported by Lietta et al. [10]. Column-specific parameters, resin parameters, mass transfer parameters, and the Mollerup isotherm parameters were sourced from the literature reported by Lietta et al. [10]. The model parameters used in this study are presented in Tables 1 and 2.

To assess the effect of loading, the loading factor (LF) is defined as:

$$LF = \frac{\bar{c}_{inj}}{q_{max}\cdot(1-\varepsilon_t)}, \tag{35}$$

where LF is the ratio of the loaded mass to the maximum loading, and ε_t represents the total porosity of the column.

The conditions of two groups of numerical experiments involving different elution gradient lengths and loadings for parameter estimation are listed in Table 3. The salt concentrations at the end of the elution gradient were set as 0.04 mol/L for all LGEs.

Table 4
HIC runs for real-world calibration experiments.

Run	Purpose	CV_G (CV)	Loadings (g/L)	$c_{s,initial}$ (mol/L)
3–1	LR	10	0.5	2.0
3–2	LR	20	0.5	2.0
3–3	LR	30	0.5	2.0
3–4	LA	30	12.0	2.0
3–5	LA	30	13.0	2.0
3–6	LA	30	14.0	2.0

3.2. Real-world experiments

For real-world experiments, a pre-packed column (HiScreen Phenyl FF low sub, Cytiva) with the dimensions of 0.77 cm × 10 cm and a column volume of 4.7 ml was used, and lysozyme (14.4 kDa) purchased from TargetMol was used as a model protein. All LGEs were performed using an ÄKTA Pure chromatography system (Cytiva). The column porosity (ε_c and ε_p) and the axial diffusion coefficient (D_{ax}) were characterized according to the protocols published [41]. The mass transfer was set to unity ($3k_{eff}/r_p = 1$) for all species [18]. The parameters of the column, resin and mass transfer and are presented in Table 1.

The conditions for the LGEs are summarized in Table 4. All LGEs were first equilibrated for 5 CV with 25 mmol/L phosphate buffer (pH 7.0) containing 2.0 mol/L ammonium sulfate. The protein loading sample was adjusted to a concentration of 1.8 g/L using the equilibration buffer. After the sample loading, the column was washed with equilibration buffer for 3 CV and then eluted under the conditions as listed in Table 4. Finally, the column was washed with 25 mmol/L phosphate buffer (pH 7.0) for 5 CV and regenerated with deionized water for 3 CV.

3.3. Process simulation

The chromatography analysis and design toolkit (CADET) was used for chromatographic process simulation, which is an open-source framework and provides numerical tools for discretizing and solving partial differential equations of chromatographic models [42,43].

For the estimation of parameter n, the least squares minimization problem was tackled by the scipy.optimize module in the open-source Python 3.10 [44].

3.4. Model calculation evaluation

To assess the validity of linear derivation assumption of $\exp(-k_s c_{s,R})/\exp(-k_s c_{s,initial}) \gg 1$, the retention position parameter (RP) is introduced as:

$$RP = \frac{c_{s,initial} - c_{s,R}}{c_{s,initial} - c_{s,finish}}. \tag{36}$$

Higher RP indicates that the retention is closer to the end of the gradient, which suggests a higher credibility for the assumption.

To assess the effectiveness of parameter estimation, the concordance between the estimation $\hat{p}$ and ground truth p of the Mollerup isotherm parameters was measured by a L^2-error as:

$$\sqrt{\frac{\| p-\hat{p}\|_{L^2}^2}{\| p\|_{L^2}^2}}. \tag{37}$$

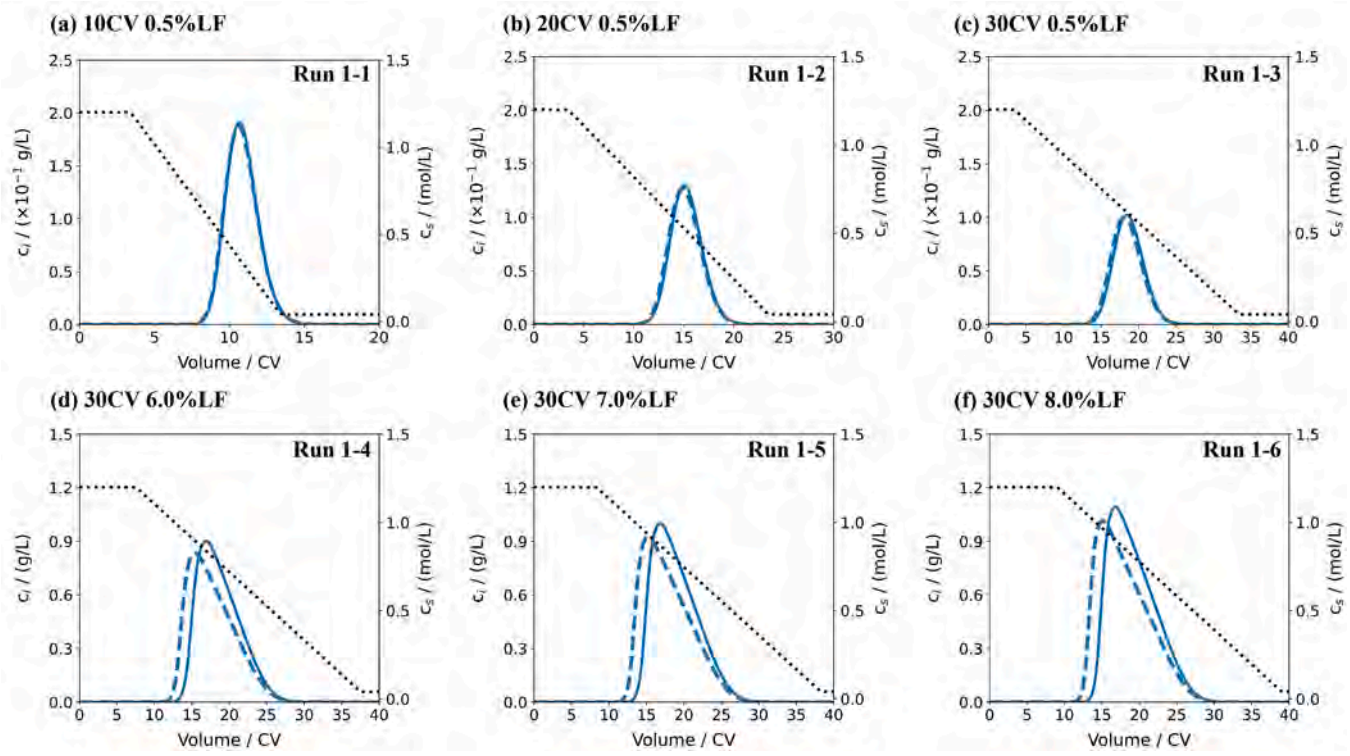

Fig. 2. Elution curves of numerical experiments (dashed line) and model simulations based on the PbP-HIC method (solid line). The salt concentrations at the column outlet are presented at the dotted line. (a) to (f) correspond to six calibration experiments (Run 1–1 to Run 1–6) in Table 3.

4. Results and discussion

4.1. Numerical experiments for parameter estimation

The mechanistic models facilitate a priori prediction of protein separation processes. Numerical experiments can provide a way to assess the congruence between the model predictions and virtual experiments in chromatography [7]. The absence of interference from experimental noise renders the numerical experiments better suited than the real-world experimental data for testing and validating the parameter estimation methods.

To evaluate the feasibility and accuracy of the PbP-HIC method proposed, numerical experiments were performed through the HIC experimental and model parameters from the literature reported by Lietta et al. [10]. The parameters used are listed in Tables 1 and 2, and the first set of numerical experimental conditions is presented in Table 3 for Group 1 (Run 1–1 to Run 1–6). The Step 1 of the PbP-HIC method needs three LGEs with different elution gradient lengths while consistent loadings. Mollerup et al. concluded that the steric hindrance of the free ligand binding sites during adsorption can be neglected under conditions of low loading density [19]. Therefore, the three LGEs (Run 1–1 to Run 1–3) for the LR step were set to LF = 0.5% in order to satisfy the linear isotherm assumption. The Step 2 of the PbP-HIC method needs three LGEs with different loadings while consistent gradient lengths. The LA step is designed to estimate the nonlinear parameters (n and q_{max}), for which the experiments need to be carried out in the nonlinear region

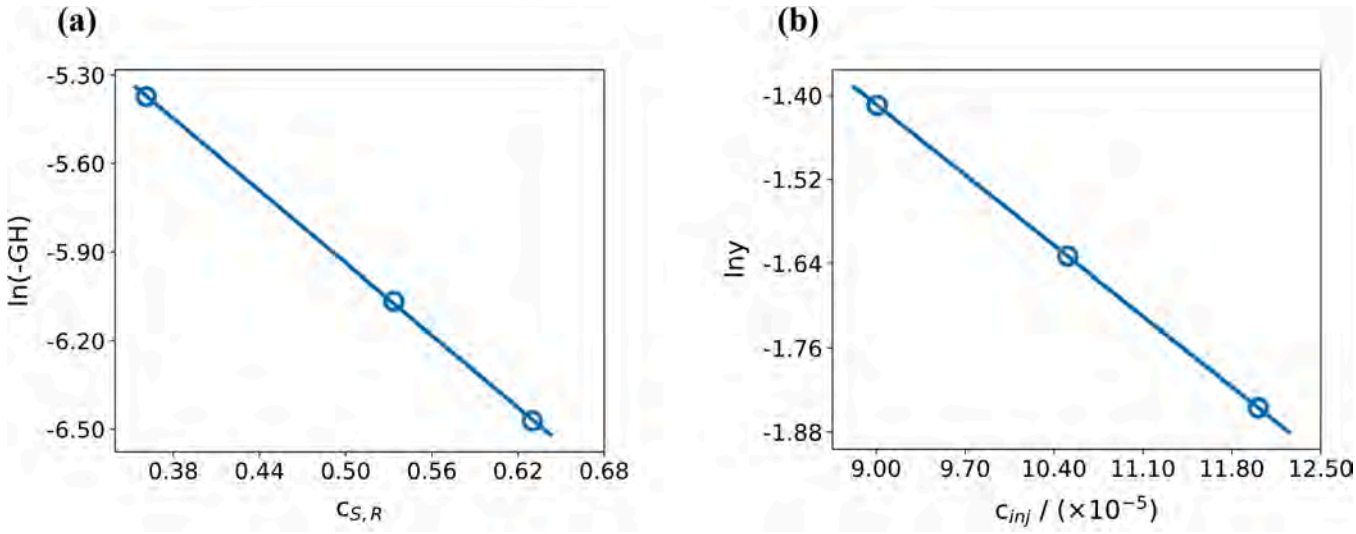

Fig. 3. Linear regression (a) and linear approximation (b) for the PbP-HIC method based the numerical experiment Group 1.

Table 5
Isotherm parameters estimated by the PbP-HIC method based the numerical experiments Group 1.

Parameter	Ground Truth	Estimated value	L^2-error	Unit
k_s	3.95	4.08	0.0323	$(\text{kmol/m}^3)^{-1}$
k_{eq}	13.1	12.1	0.0756	–
n	9.85	11.3	0.149	–
q_{max}	0.0376	0.0536	0.426	kmol/m^3

of the isotherms. Therefore, the LFs of three LGEs (Run 1–4 to Run 1–6) for the LR step were set from 6.0 to 8.0% to ensure non-negligible steric hindrance effect.

The elution curves of six numerical experiments for parameter estimation are shown in Fig. 2, which were generated with the parameters in Table 2 and the operating conditions in Table 3 (Run 1–1 to Run 1–6). It can be observed that the peak shape is symmetric at low loadings (Runs 1–1 to 1–3), indicating that there is very little competition for the binding sites and the elution occurs within the linear region of the adsorption isotherm. As the elution gradient length increases, the protein elutes at higher $c_{s,\,R}$. This is attributed to the fact that $c_{s,\,R}$ in Eq. (17) is an increasing function of the gradient length, and this trend is consistent with experimental observations [26]. At higher loadings (Runs 1–4 to 1–6), an increase in $c_{s,\,R}$ and peak asymmetry can be found as the loading increases. This can be mathematically explained using Eq. (23): as $\bar{c}_{inj}$ increases while GH remains constant, the effect of the nonlinear term $(1-a\cdot\bar{c}_{inj})^n$ becomes more pronounced, resulting in an increase in $c_{s,\,R}$. These results are consistent with the real-world experimental observations [45].

4.2. Parameter estimation by PbP method

Based on the numerical experiments in Fig. 2, the $c_{s,\,R}$ for Run 1–1 to Run 1–6 could be obtained from the elution curves. Combining Eqs. (18) and (26), the linear regression (LR) and linear approximation (LA) can be established, and the results are shown in Fig. 3. A significant negative linear correlation can be found between $\ln(-GH)$ and $c_{s,\,R}$, as well as between $\ln y$ and $\bar{c}_{inj}$, both with correlation coefficients approaching 1. This observation reveals that the theoretical derivations and assumptions are valid under the experimental conditions set above, particularly concerning the linearization of the nonlinear term.

With the LR, k_s and k_{eq} can be determined using the slope and intercept of Eq. (17), while n and q_{max} are determined with the LA by Eqs. (31) and (33). The parameters and L^2-error estimated by the PbP-HIC method are listed in Table 5. Since the inverse method is not the focus of this study, the exact values of the parameters k_p and k_{kin} are used directly in this study.

For the parameters k_s and k_{eq}, the PbP-HIC method can accurately retrieve the parameter values and the L^2-error for k_s and k_{eq} are 0.0323 and 0.0756, respectively. It can be found that the L^2-error of k_{eq} is higher than that of k_s. According to Eq. (17), k_s is derived directly from the slope, while k_{eq} is calculated using the intercept and k_s estimates. Therefore, the error in estimating k_s might accumulate and affect the precise estimation of k_{eq}.

For the parameters n and q_{max}, the PbP-HIC method results in an estimated L^2-error of 0.149 and 0.426, respectively. The relatively large L^2-error of the parameter q_{max} might be attributed to the error accumulation in estimating parameters or the assumption of LA derivation not being satisfied under such operating conditions.

At low loading (Runs 1–1 to 1–3), the numerical experiments and model simulations with the estimated parameter slightly differ in the retention. However, for high loading (Run 1–4 to Run 1–6), the numerical experiments and model simulations show inconsistency in peak heights and peak shapes. This discrepancy might be attributed to a deviation in the nonlinear parameter estimation (n and q_{max}). Therefore, the feasibility of the PbP-HIC method has been verified for estimating the Mollerup isotherm parameters for HIC, but the effect of operating conditions on parameter estimation should be further explored to find better experimental set.

4.3. Effects of operating conditions on parameter estimation

The derivation of the PbP-HIC method relies on certain assumptions regarding loadings and retention times. To evaluate the reliability and robustness of this parameter estimation method and to validate the operating conditions under which these assumptions remain valid, a comprehensive investigation into the effects of loadings and elution gradients is necessary. In this section, model parameters for the mass transfer model and adsorption isotherm are consistent with those presented in Tables 1 and 2. Numerical experiments under various loadings and elution gradients were calculated for the parameter estimation with the PbP-HIC method.

4.3.1. Effects of loadings and salt concentration on k_s and k_{eq}

The effects of loadings on the estimation of k_s and k_{eq} are shown in Fig. 4(a). The experiments were conducted at $c_{s,\,initial} = 1.5$ mol/L, covering the LF range from 0.5% to 10%. Fig. 4(b) describes the effects

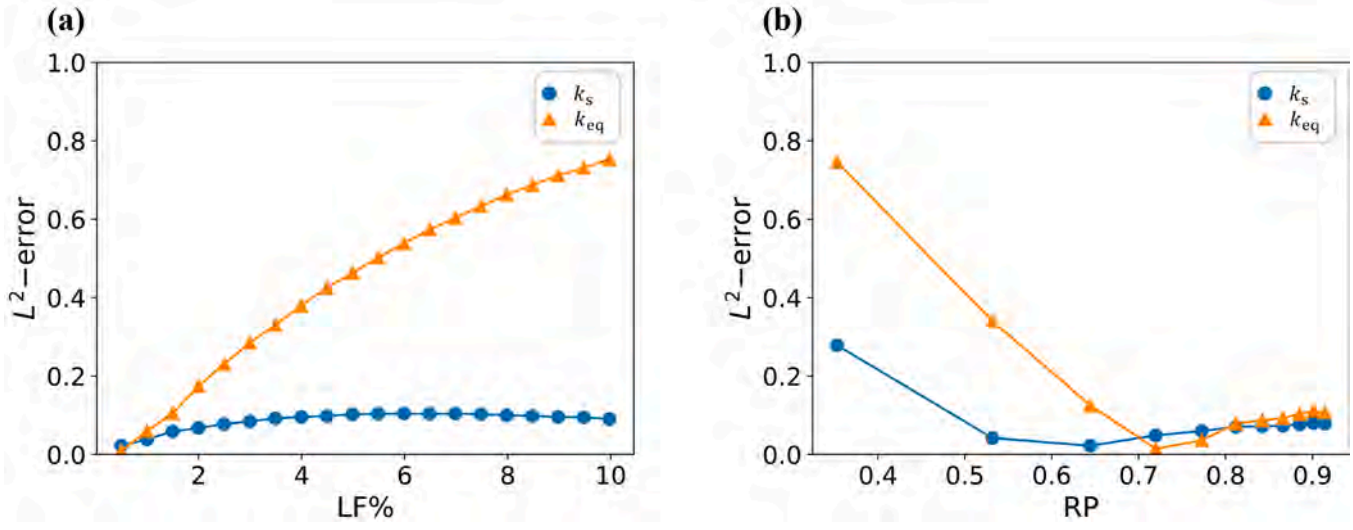

Fig. 4. Effects of loadings (a) and retention position (b) on parameter estimation. ◯: k_s, △: k_{eq}.

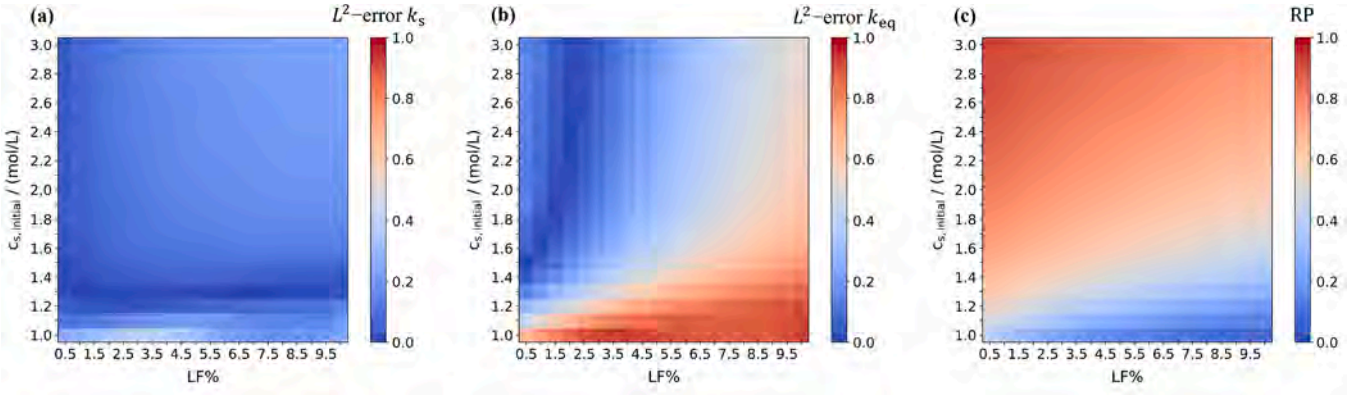

Fig. 5. Effects of loadings and elution gradients on parameter estimation and retention. (a): k_s; (b): k_{eq}; (c): RP.

of gradient slope on k_s and k_{eq} estimation. The experiments were performed with LF = 1% and $c_{s,\ initial}$ ranging from 1.0 mol/L to 3.0 mol/L.

As depicted in Fig. 4(a), the estimated value of k_s closely approached the ground truth as the loadings increased, and the L^2-error remained between 0.0223 and 0.104. However, the L^2-error for k_{eq} experienced a rapid increase, rising from 0.00885 to 0.753 as LF increased from 0.5% to 10%. At $c_{s,\ initial}$ = 1.5 mol/L and LF = 10%, the PbP-HIC method for estimating k_{eq} showed a high L^2-error of 0.753. At this point, it was evident that the effect of the nonlinear term was no longer negligible, as indicated by the clear peak asymmetry (Fig. S1 in the supplementary materials). This peak asymmetry is sufficient to verify that the initial assumption ($q \lesssim q_{max}$) is no longer valid.

The LR step of the PbP-HIC method is based on the assumption of a linear isotherm, which cannot be transplanted to the nonlinear region of the HIC isotherm. Eq. (17) lacks an appropriate term to account for the change in loadings. Therefore, the protein retention should be described as follows:

$$\ln(-\mathrm{GH}) = -k_s c_{s,R} - \ln\left[k_s k_{eq}\left(1 - a\cdot\bar{c}_{inj}\right)^n\right] \tag{38}$$

Comparing Eq. (38) to Eq. (17), the loading change does not affect the slope of the LR and would not affect k_s estimate. However, it does increase the disparity between the second right-hand term of Eq. (38) to Eq. (17), and thus the estimate of k_{eq} is sensitive to the loadings. Although k_s can be estimated at higher loadings, considering the parameter-by-parameter estimation of k_s and k_{eq}together, the desirable experimental conditions for the LR step of the PbP-HIC method should be at low loadings.

The effects of elution gradients on parameter estimation are shown in Fig. 4(b). As the $c_{s,\ initial}$ increased, the hydrophobicity required for protein binding would be enhanced [46–48], resulting in the increased retention. According to the definition of Eq. (36), an increase in RP represents the retention increase. When RP exceeded 0.644, the L^2-errors of k_s and k_{eq} were below 0.125, falling within an acceptable range of error. This implies that the assumptions of the LR, $\exp(-k_s c_{s,R})/\exp(-k_s c_{s,initial}) \gg 1$, could be met.

Fig. 5 visualizes the effects of loadings and gradient conditions together on parameter estimation and retention over a wide range, which aligns with our previous discussion. It can be found that the estimation of k_s is affected relatively weakly by the loading and elution gradient conditions, while the accurate estimation of k_{eq} depends on the appropriate experimental conditions. Comparing Fig. 5(b) and Fig. 5(c), the L^2-error of k_{eq} has a decreasing tendency with the increase of RP. This is due to the fact that the LR rule of the PbP-HIC method is based on the assumptions made in Eq. (17), which tends to be effective at low loading and high RP (LF<5% and RP>0.6).

4.3.2. *Effect of loadings and salt concentration on n and q_{max}*

The effects of loadings and gradient conditions on the estimation of n and q_{max} were investigated deeply. The results are shown in Fig. 6. The horizontal axis $\overline{\mathrm{LF}}$ represents the average of loadings across the three

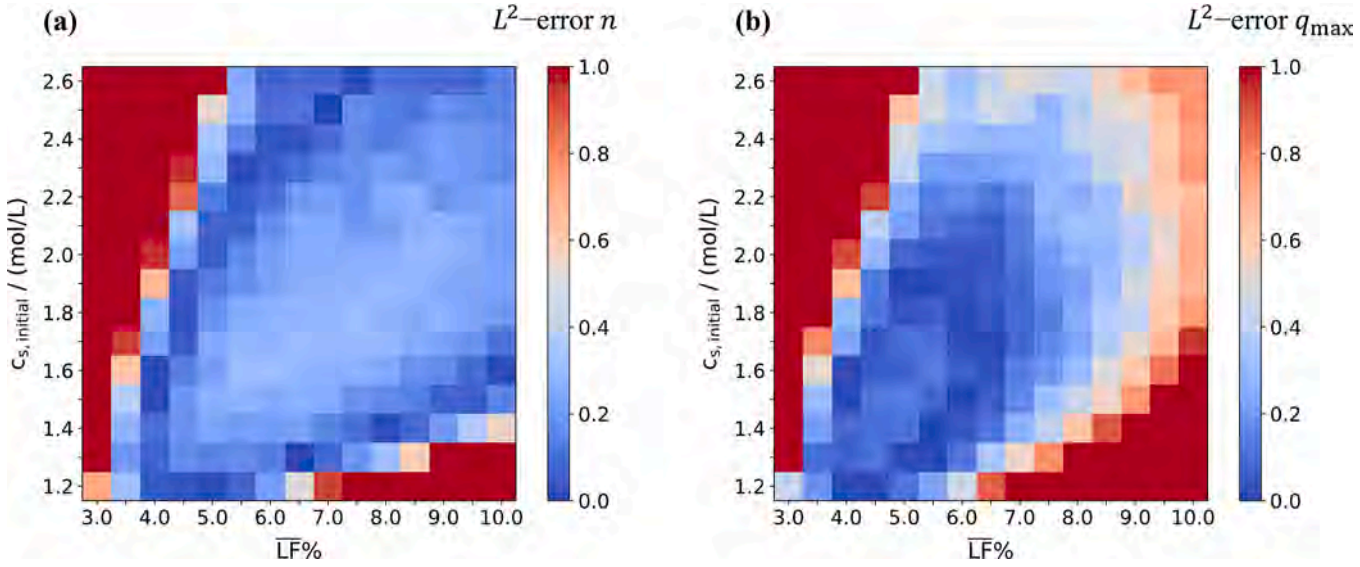

Fig. 6. Effects of loadings and elution gradients on the parameter estimation. (a): n; (b): q_{max}.

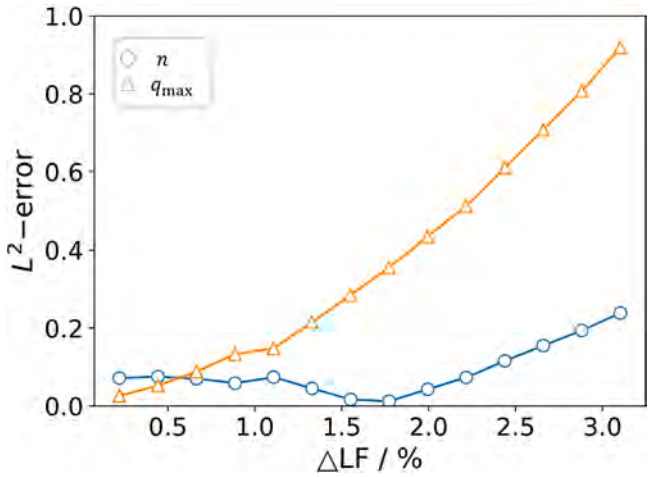

Fig. 7. Effects of the loading interval on parameter estimation of n and q_{max}.

LGEs, while maintaining a 0.5% difference in LF between them.

In Fig. 6(a), the influences of loadings and gradient conditions on the estimation of n are depicted. A notable observation is the higher L^2-error at the upper-left corner of the map, which could be attributed to the challenges in estimating the parameters n and q_{max} due to the limited impact of steric hindrance at lower loadings. As the $\overline{LF}$ increases, the L^2-error of n would reduce, and this trend remains consistent even at higher loadings (LF = 12–20%, as demonstrated in Fig. S2 in the supplementary materials).

For the parameter q_{max}, as shown in Fig. 6(b), an irregular region could be found in the map (3.5%<LF<8.5% and 1.2 mol/L<$c_{s,initial}$<2.4 mol/L) exhibiting a lower L^2-error. This desirable region might correspond to the isothermal transition region or the nonlinear region, implying that the experimental conditions for estimating q_{max} are more demanding than those for the parameter n.

The parameter q_{max} takes into account the effects of steric hindrance and hydrophobic binding sites for a more realistic understanding [19, 49], and it typically requires estimation at higher loading. However, unfortunately, the estimation of q_{max} at high loadings is not desirable in the present work. This may be attributed to the non-satisfaction of the assumption in Eq. (26) ($a\cdot\bar{c}_{inj}\ll1$) as the loadings increase.

4.3.3. *Effect of loadings on linear approximation*

The LA step of the PbP-HIC method uses multiple LGEs with different loading conditions. When the values of $\bar{c}_{inj}$ for multiple LGEs are approaching and the assumption ($a\cdot\bar{c}_{inj}\ll1$) is satisfied, the Eq. (25) in logarithmic form can be simplified to a linear Eq. (26) through linear approximation. However, this linear approximation assumption gradually breaks down as the $\bar{c}_{inj}$ difference between calibration experiments increases. Fig. 7 describes the effects of the loading interval ΔLF between three LGEs on the estimation of n and q_{max}, which was conducted with $c_{s,initial}$ = 2.0 mol/L, and the consistent elution gradient length of 30 CV for all three LGEs. The initial LF was set at 5%, and the loading interval ΔLF between runs was increased from 0.2% to 3.2%.

The results indicate that as the loading interval increased, the L^2-error of n estimation remained within the range of 0.00315 to 0.266, but the error of q_{max} estimation increased rapidly. If ΔLF increased to 2%, the L^2-error of q_{max} would exceed 0.4. This might be primarily attributed to the dual factors that Eq. (25) in logarithmic form cannot be simply transformed into linear form as ΔLF increases, and the progressive invalidation of the assumption ($a\cdot\bar{c}_{inj}\ll1$).

4.4. *Design strategy of calibration experiments*

Based on the above discussion, the first (LR) and second (LA) steps of the PbP-HIC method require different experimental conditions. The LR step for estimating k_s and k_{eq} is most effective under lower loadings and appropriate gradient conditions. In contrast, the LA step for estimating n and q_{max} typically demands higher loadings and smaller loading intervals. Therefore, a reasonable and convenient design strategy of calibration experiments to implement the PbP-HIC method would be proposed as outlined below:

(1) For the LR step, it is recommended to perform three LGEs with different gradient lengths as low LF. It is important to keep the LF below 5% and ensure that the average RP of these experiments is above 0.6. This guarantees the symmetry of the eluted peaks, aligning with the assumptions of Eq. (17) and allowing for an accurate estimation of the parameters k_s and k_{eq}.
(2) For the LA step, it is advisable to conduct at least three LGEs with different loadings. Additionally, the average of loadings across the three LGEs ($\overline{LF}$) should be increased to between 3.5% and 8.5% compared to the LR step to ensure that the steric hindrance effect is not negligible. The loading interval (ΔLF) should be less than 2% to guarantee favorable linearity at LA.
(3) Step3, k_p and k_{kin} are estimated by the inverse method using the elution curves of the above all LGEs.

Table 6
Isotherm parameters estimated by the PbP-HIC method based the numerical experiments Group 2.

Parameter	Ground Truth	Estimated value	L^2-error	Unit
k_s	3.95	3.90	0.0112	$(kmol/m^3)^{-1}$
k_{eq}	13.1	13.5	0.0318	–
n	9.85	10.4	0.0574	–
q_{max}	0.0376	0.0406	0.0813	$kmol/m^3$

Based on the above experimental strategy, the numerical experimental conditions were designed as presented in Table 3 for Run 2–1 to Run 2–6. For the LR step, $c_{s,\ initial}$ = 1.4 mol/L was set to ensure a higher RP. For the LA step, LF was set from 4.0% to 5.0% to ensure that the adsorption is in the transition or nonlinear region and the assumption ($a\cdot\bar{c}_{inj}\ll1$) is valid.

As shown in Fig. 9, good linear relationships could be found for the LR and LA steps. The parameters estimated are summarized in Table 6, and the L^2-error for all the estimated parameter are less than 0.1, implying that the PbP-HIC method worked better under the above design strategy. Fig. 8 displays the comparison between the numerical experiments and the model simulations of the estimated parameters, showing good concordance regarding peak heights and widths.

Both PbP-IEC and PbP-HIC methods are based on the derivation of retention mechanism, and the core of both methods lies in the linearization of the nonlinear terms. The PbP-IEC method uses two linear regressions (LR1-IEC and LR2-IEC) and one linear approximation (LA-IEC) to disentangle the retention model, allowing for stepwise estimation of three SMA isotherm parameters (ν, k_{eq} and σ) through five LGEs. For the calibration experiments for PbP-IEC, three LGEs with varying gradient lengths are used for LR1-IEC to estimate ν, and three LGEs with varying loadings for LR2-IEC and LA-IEC to estimate k_{eq} and σ. Since the LR1-IEC step is not limited by the loading, five calibration experiments can be rationally designed so that the LR1-IEC and LA-IEC steps can co-use one LGE [29]. Compared to the PbP-IEC method, the retention model (Eq. (23)) derived from PbP-HIC has more unknown parameters, which is more challenging to disentangle. Additionally, the LR and LA steps in PbP-HIC have different loading limits that cannot co-use one LGE in most cases. Therefore, for the reasonable estimation of HIC parameters, the design of the calibration experiments for PbP-HIC requires additional assumptions and at least six calibration experiments are proposed.

In order to calibrate the HIC linear isotherm, Andris et al. [11] estimated the linear parameters (k_s and k_{eq}) by the inverse method.

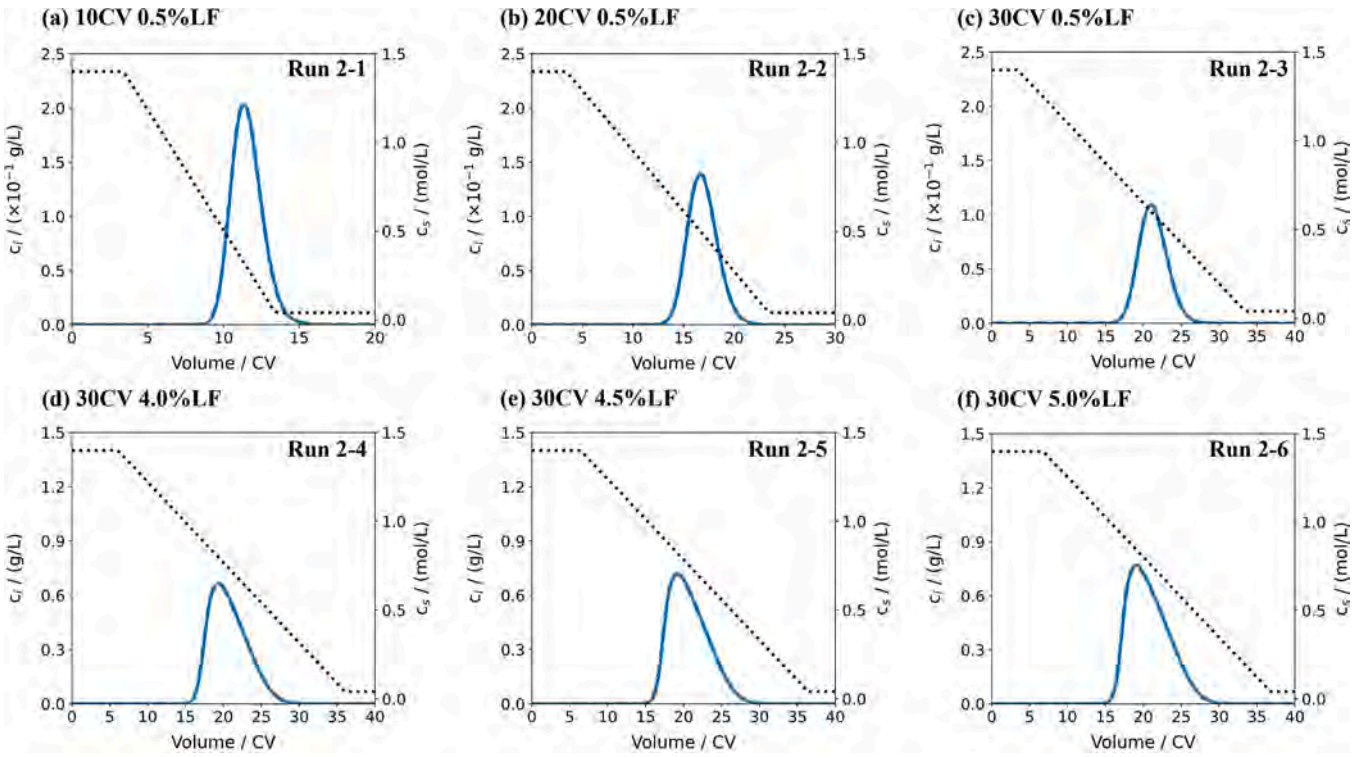

Fig. 8. Elution curves of numerical experiments (dashed line) and model simulations based on the PbP-HIC method (solid line). The salt concentrations at the column outlet are presented at the dotted line. (a) to (f) correspond to six calibration experiments (Run 2–1 to Run 2–6) in Table 3.

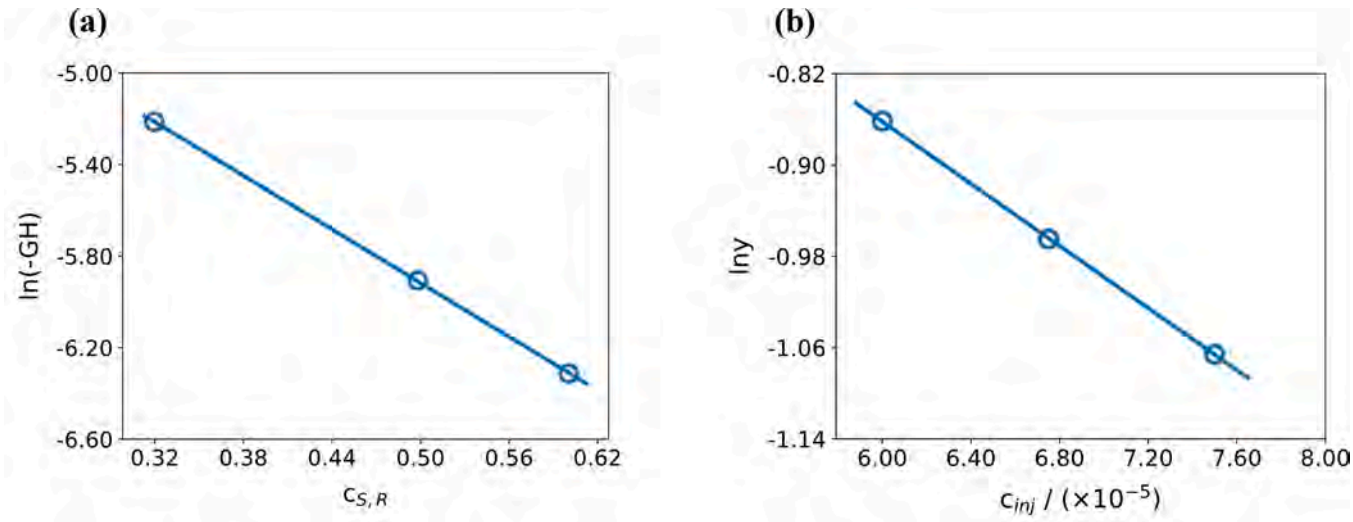

Fig. 9. Linear regression (a) and linear approximation (b) for the PbP-HIC method based the numerical experiment Group 2.

However, applying the LR step of the PbP method reduces the experimental effort required to calibrate the model, and parameter estimation based on the retention mechanism could improve the identifiability of the model parameters. For the parameters n and q_{max} that take into account the nonlinear term of the isotherm at high loadings, they are often determined using the inverse method under overload conditions [10,49,50]. In contrast, in the present work a nonlinear factor a is introduced in the LA step, which can characterize the effect of the nonlinear term under undiluted conditions, Therefore, the LA step offers the strong support for rational estimation of n and q_{max}.

In summary, estimating a large number of model parameters using inverse method can be time-consuming and might yield parameters with limited physical interpretability. Moreover, the results may exhibit instability or struggle to find a unique optimal solution [51], possibly due to possible linear correlations between parameters or incomplete data. which has been observed for the SMA isotherm [18]. In contrast, the PbP-HIC process is more efficient, effectively saving the experimental efforts usually required to estimate HIC parameters by design of experiments (DoE), and reasonably reducing the number of parameters for inverse method fitting from six to two per protein.

4.5. *Validation with real-world experiments and application prospect*

The real-world experiments were designed based on the design strategy proposed above, and the experiment conditions are listed in Table 4. The process of parameter estimation for the real-world

Table 7
Isotherm parameters estimated by the PbP-HIC method based the real-world experiments.

Parameter	Estimated value	Unit
k_s	4.58	$(kmol/m^3)^{-1}$
k_{eq}	0.888	–
n	0.922	–
q_{max}	0.156	$kmol/m^3$
k_p	3.20×10^{-6}	$(kmol/m^3)^{-1}$
k_{kin}	33.3	s

experiments remained consistent with the numerical experiments described above, and Table 7 summarizes all the estimated parameters. The experiment results and model calculations are compared in Fig. 10. The LR and LA steps exhibit good linear relationships for the real-world experiment as shown in Fig. 11. It can be found that the peak shape exhibits asymmetry for Run 3–3 to Run 3–6, indicating that the isotherm is no longer in the linear region at the present loading. The trends of all real-world experiments are similar to that of the numerical experiments described above. Overall, model simulations accurately predicted the retention and peak shape based on the parameters estimated by the PbP-HIC method.

The present studies were only validated for the numerical experiments with bovine serum albumin and the real-world experiments with

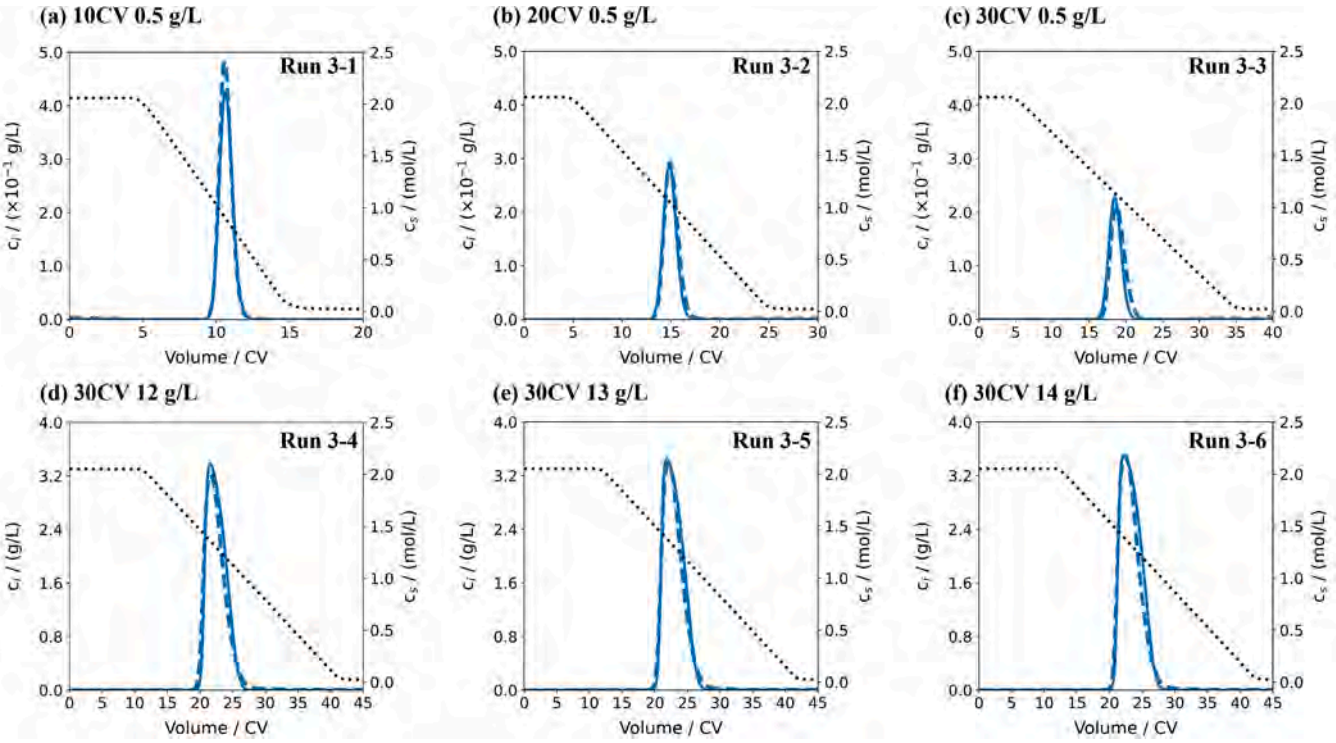

Fig. 10. Elution curves of real-world experiments (dashed line) and model calculations based on the PbP-HIC method (solid line). The salt concentrations at the column outlet are presented at the dotted line. (a) to (f) correspond to six calibration experiments (Run 3–1 to Run 3–6) in Table 4.

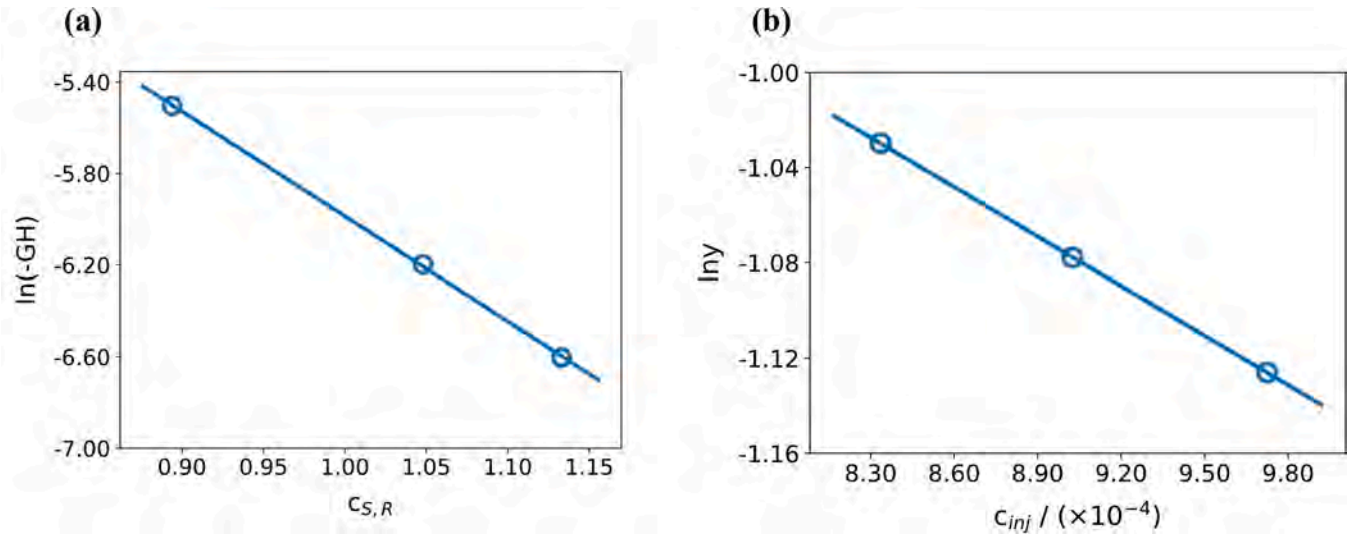

Fig. 11. Linear regression (a) and linear approximation (b) for the PbP-HIC method based the real-world experiments.

lysozyme. However, for real applications the experimental noises (e.g., conductivity shifts at high salt concentrations and UV detector noise) are complicated and miscellaneous, which may have impacts for parameter estimation and process modeling [52]. The model parameters with high sensitivity might be susceptible to experimental errors or noise, thus the estimated parameters would have relatively large deviations. These effects are difficult to be accounted by the mechanism models [53] and the PbP-HIC method. Therefore, additional efforts are required to understand and evaluate the effect of experimental noises on the estimated parameters, for example modeling the experimental error (e.g., pump delay and UV detector noise) [54] and quantifying the parameter uncertainty based on statistical methods [53–55].

Rischawy et al. [18] found that the sensitivity of the model parameters were varying under different operating conditions, and an increase in parameter sensitivity would pose a risk to parameter estimation. The evaluation on the simulation process disturbances can help us identify the key parameters with high sensitivity, thus pointing out the parameters that need to be estimated with good accuracy [18,56,57]. For the SMA isotherm, k_{eq} and ν affect the elution peak position and exhibit high sensitivity in the experiments, while the nonlinear parameter σ exhibits higher sensitivity at high loading, implying that low loading experiments are not suitable for estimating σ [18]. For the HIC Mollerup isotherm, the parameters k_s and k_{eq} affecting peak position might be highly sensitive and should be estimated with high precision, while the nonlinear parameters n and q_{max} may be insensitive at low loading and therefore need to be estimated under undiluted conditions.

Additionally, certain process variations (e.g., column aging and pH change) might also affect the model parameter values thereby reducing the predictive power of the model. Further incorporation of artificial neural networks might be suitable for investigating the root causes of deviations [13]. For the further research work, the sensitivity of model parameters and more applications with the PbP-HIC method will be systematically investigated in order to improve the understanding of the proposed method and apply it to the complex multi-component system for protein purification.

5. Conclusions

In this study, a PbP method was proposed for parameter estimation of Mollerup isotherm for HIC through theoretical derivation and reasonable simplifying assumptions. New method was validated through numerical experiments of the LGEs under varying gradient lengths and loading conditions. Four parameters (k_s, k_{eq}, n and q_{max}) can be estimated step-by-step. Additionally, a detailed evaluation was conducted to understand how changes in loading and elution gradient conditions affect the performance of parameter estimation, shedding light on the applicability of the PbP-HIC method proposed. It was found that the PbP-HIC method had a superior performance when the LR step was used under dilution conditions (loading factor below 5%), and the LA step was conducted when the isotherm lay in the transition or nonlinear region. A strategy containing six LGEs was proposed to make the implementation of the PbP-HIC method simpler and more reasonable. The numerical experiments were performed based on the proposed strategy, and the estimated L^2-error of the four parameters were less than 0.1. The real-world experiments with lysozyme further validated the feasibility of the PbP-HIC method. The proposed method can efficiently save the cost of calibration experiments and reduce the parameters to be fitted by inverse method from six to two per protein, hence preventing non-identifiability of structural parameters. Furthermore, the PbP method is expected to be rationally applied to other SDM-based derived isotherms to accelerate the model-based downstream process development.

CRediT authorship contribution statement

Yu-Xiang Yang: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Data curation, Conceptualization. **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 National Natural Science Foundation of China (22078286), National Key R&D Program of China (2021YFE0113300) 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.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.chroma.2024.464638.

References

[1] J. Chen, J. Tetrault, A. Ley, Comparison of standard and new generation hydrophobic interaction chromatography resins in the monoclonal antibody purification process, J. Chromatogr. A 1177 (2008) 272–281, https://doi.org/10.1016/j.chroma.2007.07.083.
[2] P. Gronemeyer, R. Ditz, J. Strube, Trends in upstream and downstream process development for antibody manufacturing, Bioengineering 1 (2014) 188–212, https://doi.org/10.3390/bioengineering1040188.
[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] 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.
[5] J.A. Queiroz, C.T. Tomaz, J.M.S. Cabral, Hydrophobic interaction chromatography of proteins, J. Biotechnol. 87 (2001) 143–159, https://doi.org/10.1016/S0168-1656(01)00237-1.
[6] A.T. Hanke, M. Ottens, Purifying biopharmaceuticals: knowledge-based chromatographic process development, Trends Biotechnol 32 (2014) 210–220, https://doi.org/10.1016/j.tibtech.2014.02.001.
[7] 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.
[8] 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.
[9] 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.
[10] 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.
[11] 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.
[12] Y. Kawajiri, Model-based optimization strategies for chromatographic processes: a review, Adsorption 27 (2021) 1–26, https://doi.org/10.1007/s10450-020-00251-2.

[13] G. Wang, T. Briskot, T. Hahn, P. Baumann, J. Hubbuch, Root cause investigation of deviations in protein chromatography based on mechanistic models and artificial neural networks, J. Chromatogr. A 1515 (2017) 146–153, https://doi.org/10.1016/j.chroma.2017.07.089.
[14] Y. Chen, O. Yang, C. Sampat, P. Bhalode, R. Ramachandran, M. Ierapetritou, Digital twins in pharmaceutical and biopharmaceutical manufacturing: a literature review, Processes 8 (9) (2020), https://doi.org/10.3390/pr8091088.
[15] A. Schmidt, S. Zobel-Roos, H. Helgers, L. Lohmann, F. Vetter, C. Jensch, A. Juckers, J. Strube, G. Subramanian, Digital twins for continuous biologics manufacturing. Process Control, Intensification, and Digitalisation in Continuous Biomanufacturing, Wiley-VCH, Germany, 2022, pp. 265–350.
[16] C.A. Brooks, S.M. Cramer, Steric mass-action ion exchange: displacement profiles and induced salt gradients, AIChE J 38 (1992) 1969–1978, https://doi.org/10.1002/aic.690381212.
[17] 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.
[18] 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.
[19] 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.
[20] 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.
[21] V. Kumar, S. Leweke, W. Heymann, E. von Lieres, F. Schlegel, K. Westerberg, A. M. Lenhoff, Robust mechanistic modeling of protein ion-exchange chromatography, J. Chromatogr. A 1660 (2021) 462669, https://doi.org/10.1016/j.chroma.2021.462669.
[22] 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.
[23] F.-G. Wieland, A.L. Hauber, M. Rosenblatt, C. Tönsing, J. Timmer, On structural and practical identifiability, Curr. Opin. Syst. Biol. 25 (2021) 60–69, https://doi.org/10.1016/j.coisb.2021.03.005.
[24] T. Ishihara, S. Yamamoto, Optimization of monoclonal antibody purification by ion-exchange chromatography: application of simple methods with linear gradient elution experimental data, J. Chromatogr. A 1069 (2005) 99–106, https://doi.org/10.1016/j.chroma.2004.10.040.
[25] 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.
[26] 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.
[27] A. Creasy, J. Reck, T. Pabst, A. Hunter, G. Barker, G. Carta, Systematic interpolation method predicts antibody monomer-dimer separation by gradient elution chromatography at high protein loads, Biotechnol. J. 14 (2019) 1800132, https://doi.org/10.1002/biot.201800132.
[28] 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.
[29] 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.
[30] 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.
[31] A. Felinger, G. Guiochon, Comparison of the kinetic models of linear chromatography, Chromatographia 60 (2004) S175–S180, https://doi.org/10.1365/s10337-004-0288-7.
[32] 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.
[33] 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.
[34] T. Hahn, A. Sommer, A. Osberghaus, V. Heuveline, J. Hubbuch, Adjoint-based estimation and optimization for column liquid chromatography models, Comput. Chem. Eng. 64 (2014) 41–54, https://doi.org/10.1016/j.compchemeng.2014.01.013.
[35] R.W. Deitcher, J.E. Rome, P.A. Gildea, J.P. O'Connell, E.J. Fernandez, A new thermodynamic model describes the effects of ligand density and type, salt concentration and protein species in hydrophobic interaction chromatography, J. Chromatogr. A 1217 (2010) 199–208, https://doi.org/10.1016/j.chroma.2009.07.068.
[36] B.K. Nfor, M. Noverraz, S. Chilamkurthi, P.D.E.M. Verhaert, L.A.M. van der Wielen, M. Ottens, High-throughput isotherm determination and thermodynamic modeling of protein adsorption on mixed mode adsorbents, J. Chromatogr. A 1217 (2010) 6829–6850, https://doi.org/10.1016/j.chroma.2010.07.069.
[37] S.R. Gallant, A. Kundu, S.M. Cramer, Modeling non-linear elution of proteins in ion-exchange chromatography, J. Chromatogr. A 702 (1995) 125–142, https://doi.org/10.1016/0021-9673(94)00992-I.
[38] 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.
[39] M. Rüdt, F. Gillet, S. Heege, J. Hitzler, B. Kalbfuss, B. Guélat, Combined Yamamoto approach for simultaneous estimation of adsorption isotherm and kinetic parameters in ion-exchange chromatography, J. Chromatogr. A 1413 (2015) 68–76, https://doi.org/10.1016/j.chroma.2015.08.025.
[40] 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.
[41] H. Schmidt-Traub, Preparative Chromatography of Fine Chemicals and Pharmaceutical Agents, Wiley-VCH, Germany, 2005.
[42] 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.
[43] 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.
[44] 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, 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.
[45] 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.
[46] M. Baca, J. De Vos, G. Bruylants, K. Bartik, X. Liu, K. Cook, S. Eeltink, A comprehensive study to protein retention in hydrophobic interaction chromatography, J. Chromatogr. B 1032 (2016) 182–188, https://doi.org/10.1016/j.jchromb.2016.05.012.
[47] X. Geng, L.A. Guo, J. Chang, Study of the retention mechanism of proteins in hydrophobic interaction chromatography, J. Chromatogr. A 507 (1990) 1–23, https://doi.org/10.1016/S0021-9673(01)84176-5.
[48] T.W. Perkins, D.S. Mak, T.W. Root, E.N. Lightfoot, Protein retention in hydrophobic interaction chromatography: modeling variation with buffer ionic strength and column hydrophobicity, J. Chromatogr. A 766 (1997) 1–14, https://doi.org/10.1016/S0021-9673(96)00978-8.
[49] G. Wang, T. Hahn, J. Hubbuch, Water on hydrophobic surfaces: mechanistic modeling of hydrophobic interaction chromatography, J. Chromatogr. A 1465 (2016) 71–78, https://doi.org/10.1016/j.chroma.2016.07.085.
[50] S.H. Altern, J.P. Welsh, J.Y. Lyall, A.J. Kocot, S. Burgess, V. Kumar, C. Williams, A. M. Lenhoff, S.M. Cramer, Isotherm model discrimination for multimodal chromatography using mechanistic models derived from high-throughput batch isotherm data, J. Chromatogr. A 1693 (2023) 463878, https://doi.org/10.1016/j.chroma.2023.463878.
[51] 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.
[52] 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.
[53] 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.
[54] 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.
[55] 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.
[56] N. Jakobsson, M. Degerman, B. Nilsson, Optimisation and robustness analysis of a hydrophobic interaction chromatography step, J. Chromatogr. A 1099 (2005) 157–166, https://doi.org/10.1016/j.chroma.2005.09.009.
[57] M. Degerman, K. Westerberg, B. Nilsson, Determining critical process parameters and process robustness in preparative chromatography – a model-based approach, Chem. Eng. Technol. 32 (2009) 903–911, https://doi.org/10.1002/ceat.200900019.