ARTICLE OPEN ACCESS

Hybrid Modeling of the Reversed-Phase Chromatographic Purification of an Oligonucleotide: Few-Shot Learning From Differentiable Physics Solver-in-the-Loop

Yu-Cheng Chen[1,2] 🆔 | Ismaele Fioretti[2] 🆔 | Dong-Qiang Lin[1] 🆔 | Mattia Sponchioni[2] 🆔

[1]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, China | [2]Department of Chemistry, Materials and Chemical Engineering "Giulio Natta", Politecnico di Milano, Milano, Italy

Correspondence: Mattia Sponchioni (mattia.sponchioni@polimi.it)

ABSTRACT

Hybrid models integrate mechanistic and data-driven components, effectively addressing the challenges of limited process understanding and data availability typical of biopharmaceutical processes. In this study, we applied a hybrid modeling framework named differentiable physics solver-in-the-loop (DP-SOL) to describe the reversed-phase chromatographic purification of an oligonucleotide, overcoming the mentioned limitations of purely mechanistic and data-driven models. The framework establishes a connection between neural networks (NNs) and mechanistic models through differentiable physical operators and their gradients. We first collected a data set comprising six linear gradient elution experiments at different resin loadings and gradient slopes, split in three experiments each for training and testing, for few-shot learning. The hyperparameters were determined through a grid search, resulting in a NN with two hidden layers and 14 nodes. Compared to a calibrated mechanistic model used for initialization of NN, the DP-SOL hybrid model showed significant performance improvement on both training and testing sets, with $R^2 > 0.97$ for the former. The good predictivity of DP-SOL is attributed to the combination of mechanistic models and NNs at the solver level. As a novel and versatile hybrid modeling paradigm, DP-SOL has the potential to significantly impact modeling approaches in the downstream processing field and the broader biopharmaceutical sector.

1 | Introduction

Mathematical models of chemical/biochemical processes can be divided into two main classes (Rizki and Ottens 2023): mechanistic models, expressing fundamental principles on the process in the form of partial differential equations (PDEs) (Chen et al. 2024), and data-driven models predominantly based on deep learning (DL). While mechanistic models can provide a deeper insight into the process (Huuk et al. 2014; Kozorog et al. 2023; Qian et al. 2023),

they require a comprehensive understanding of the system under investigation, which is often out of reach. In contrast, data-driven models require large data sets to establish statistical correlations between input and output variables (Velioğlu et al. 2024) and may lack interpretability and process insights.

Hybrid models, often referred to as gray-box models, have emerged as a valuable alternative as they combine the strengths of purely mechanistic and purely data-driven models, compensating the

limits of both (Jungbauer et al. 2024; Mahanty 2023; Malinov et al. 2024; Narayanan et al. 2022; Narayanan et al. 2023; Roush et al. 2020; Wittkopp et al. 2024). This integration, defined as physics-based DL by Thuerey et al. (2021), encompasses three categories: supervised learning, physics-informed neural networks (PINNs), and differentiable numerical simulations of physical systems (referred to as differentiable physics (DP) in this study). Thuerey et al. (2021) suggested that these methods are evolutionary, with DP representing the ultimate form of physics-based DL.

Recent attempts at supervised learning (first-category of physics-based DL) in the biopharmaceutical field include (Ding et al. 2023; Narayanan et al. 2022; Narayanan et al. 2022; Narayanan et al. 2021). In cell culture processes (upstream), these hybrid models have shown superior performance compared to purely mechanistic and purely data-driven models (Narayanan et al. 2022; Narayanan et al. 2022). In downstream processing, neural networks (NNs) were used to describe the adsorption dynamics, which are poorly understood, while retaining a mechanistic description of the molecule transport in the column, thus obtaining effective hybrid NN-PDE models (Ding et al. 2023; Narayanan et al. 2021). However, this supervised learning has two major limitations. First, its performance heavily depends on the integration of the first-principles and data-driven component, which lacks a unified approach and requires extensive trial and error. For example, in chromatographic modeling (Ding et al. 2023; Narayanan et al. 2021), the different mass transfer contributions can be modeled using either NNs or PDEs, and the integration of these terms needs careful consideration. Second, this model was trained using least squares optimizations and disregards the differentiability and backpropagation of NNs, making it effective only with shallow NNs. When NNs are deepened to model more complex processes, this method becomes inapplicable, easily falling into local optima.

The second category, PINNs, has also been reported in chromatography modeling for both forward and inverse problems (Santana et al. 2022; Söderström 2022; Subraveti et al. 2022a, 2022b, 2023; Tang et al. 2023; Zou et al. 2024). However, the primary issue with PINNs is that they build independently of traditional numerical methods, discarding decades of accumulated knowledge in numerical computation.

To overcome these limitations, a novel and versatile hybrid modeling framework, DP (third-category physics-based DL), was proposed. Unlike supervised learning, which introduces NNs at the model level, DP integrates NNs at the solver level (Ramsundar et al. 2021). This solver-level integration allows existing numerical solvers to compute gradients with respect to their inputs, leading to a DP solver. Once this integration is achieved for all computations within a simulation, the automatic differentiation (AD) functionality of DL frameworks combined with backpropagation enables gradient-based model training, which significantly outperforms the least squares optimization methods (first category physics-based DL). Therefore, DP overcomes the limitations of the first category of physics-based DL by eliminating the need for a careful integration of NN and PDE and by leveraging the differentiability and backpropagation of NNs. Additionally, it addresses the shortcomings of the second category of physics-based DL by retaining numerical computation knowledge.

A classical approach involves the cyclic evaluation of DP using differentiability and backpropagation within the solver time-stepping process, termed as DP solver-in-the-loop (DP-SOL), which can significantly reduce numerical simulation errors. Since its introduction in 2020, DP-SOL has been employed to develop various hybrid models in computational fluid dynamics, covering the Navier-Stokes equations (Um et al. 2020), and convection-dispersion problems (Wiewel et al. 2020). Given its potential, chromatographic models, a variant of convection-dispersion systems, can greatly benefit in terms of accuracy and computational time by the implementation of DP-SOL. In particular, reversed-phase chromatography (RPC) is one of the most frequently used processes for the purification of biopharmaceuticals (Catani et al. 2020; De Luca et al. 2020), but it is still predominantly described through mechanistic models, including the Mollerup (2007) model based on thermodynamic properties and activity (Chen et al. 2024), as well as stoichiometric displacement models (SDM) (Arkell et al. 2017; Arkell et al. 2018; Arkell et al. 2018) with their simplified version (Nozaki et al. 2024). However, the lack of consensus on the underlying mechanisms of RPC, particularly concerning the use of activity models for organic solvents, has limited the applications of these mechanistic approaches in this field.

Given the limited understanding of RPC processes, hybrid models represent a promising alternative for a reliable description of the operation. However, hybrid models require more experimental data for training compared to purely mechanistic models, due to the incorporation of data-driven components. At the same time, the generation of a sufficient amount of high-quality data for model training remains a challenge in biotechnology (Jungbauer et al. 2024; Ou et al. 2024; Saleh et al. 2022; Wu et al. 2024). Consequently, this field often resorts to few-shot learning (Liu et al. 2024) to reduce the data requirement and align with mechanistic models, thus facilitating the transition from a more conventional approach to the hybrid one without incurring into additional experimental effort.

In this direction, this study aims at developing a hybrid model using DP-SOL that can accurately describe the RPC purification of an oligonucleotide with limited experimental data. First, we introduced the concept and derivation of DP-SOL and discussed how to construct it for RPC processes. Then, DP-SOL was implemented in the modeling of six fractionated linear gradient separations obtained at different oligonucleotide loadings on the resin and gradient durations. Out of these experiments, three were used for training the model and three for testing it, in a 1:1 ratio. A comparison between the DP-SOL and a mechanistic model was conducted using the same data set, showing the great potential of the hybrid model in improving the accuracy and predictivity of the simulations and paving the way to a new paradigm in chromatography modeling.

2 | Theory

Figure 1 schematically illustrates the concept of DP-SOL for the hybrid modeling of RPC processes and the different panels are described in details in the following sections.
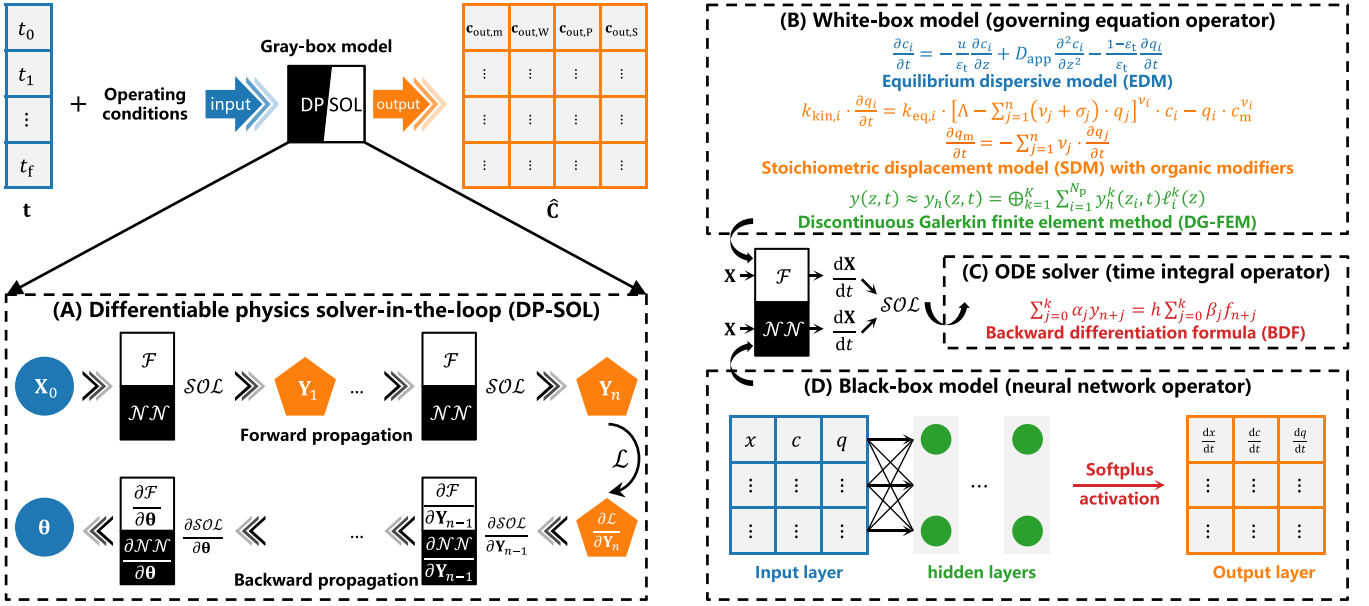
FIGURE 1 | Differentiable physics solver-in-the-loop (DP-SOL) for hybrid modeling (gray-box model). (A) Forward and backward propagation of DP-SOL with solver-in-the-loop using differentiable operators $\mathscr{F}$, $\mathscr{NN}$, and $\mathscr{SOL}$. (B) White-box model for operator $\mathscr{F}$. (C) Solver for operator $\mathscr{SOL}$. (D) Black-box model for operator $\mathscr{NN}$.

2.1 | Differentiability and Backpropagation of Neural Networks

As shown in Figure 1D, DP-SOL is implemented by a fully connected NN (FCNN) comprising one input layer, $n_\mathcal{L}$ hidden layers, and one output layer, with a forward propagation scheme:

$$\mathbf{Y}^l = \mathbf{W}^l \mathbf{X}^{l-1} + \mathbf{b}^l, \tag{1}$$

$$\mathbf{X}^l = f^l(\mathbf{Y}^l), \tag{2}$$

where $\mathbf{Y}^l$, $\mathbf{X}^l$, $\mathbf{W}^l$, $\mathbf{b}^l$, and f^l denote the output, input, weights, bias, and activation function of the l-th layer, respectively.

The training of NNs can be summarized as the tuning of the parameters θ to minimize the loss function $\mathcal{L}$, given the input and output, so that the NN output can approximate the given output as closely as possible. This can be expressed as the following optimization problem:

$$\min_{\theta} \mathcal{L}, \tag{3}$$

$$\theta = \begin{bmatrix} \mathbf{W}^0 & \mathbf{W}^1 & \cdots & \mathbf{W}^{n_\mathcal{L}} & \mathbf{b}^0 & \mathbf{b}^1 & \cdots & \mathbf{b}^{n_\mathcal{L}} \end{bmatrix}. \tag{4}$$

The differentiability and backpropagation of NNs allow for updating θ by gradient descent methods:

$$\mathbf{W}^l = \mathbf{W}^l - \alpha \frac{\partial \mathcal{L}}{\partial \mathbf{W}^l}, \tag{5}$$

$$\mathbf{b}^l = \mathbf{b}^l - \alpha \frac{\partial \mathcal{L}}{\partial \mathbf{b}^l}, \tag{6}$$

where α denotes the learning rate, while the gradients $\frac{\partial \mathcal{L}}{\partial \mathbf{W}^l}$ and $\frac{\partial \mathcal{L}}{\partial \mathbf{b}^l}$ can be obtained by:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{W}^l} = (\mathbf{X}^{l-1})^\mathcal{T} \times \frac{\partial \mathcal{L}}{\partial \mathbf{Y}^l}, \tag{7}$$

$$\frac{\partial \mathcal{L}}{\partial \mathbf{b}^l} = 1 \times \frac{\partial \mathcal{L}}{\partial \mathbf{Y}^l}, \tag{8}$$

$$\frac{\partial \mathcal{L}}{\partial \mathbf{X}^{l-1}} = \frac{\partial \mathcal{L}}{\partial \mathbf{Y}^l} \times (\mathbf{W}^l)^\mathcal{T}. \tag{9}$$

2.2 | Solver for Addressing the Initial Value Problem in Chromatography

The unknowns $\mathbf{y}$ determined by chromatographic models include the concentrations of each specie in the mobile phase (c) and in the stationary phase (q) at different times (t) and longitudinal position (z):

$$y(z, t) = [c(z, t) \quad q(z, t)]^\mathcal{T}. \tag{10}$$

The governing equation of Equation (10) can be written as:

$$\frac{\partial y}{\partial t}(z, t) = \left[\frac{\partial c}{\partial t}(z, t) \quad \frac{\partial q}{\partial t}(z, t) \right]^\mathcal{T}, \tag{11}$$

whose right-hand side can be reduced to a spatially semi-discretized form, resulting in the approximate solution y_h, re-written as:

$$\frac{dy_h(t)}{dt} = f(y_h(t), t). \tag{12}$$

Given the initial value $y_h(0) = y_0$ and the residual term f, one can integrate over time using a solver to obtain y_h for any given time t. Equation (12) can be efficiently solved by the backward differentiable formula (BDF), which uses high-order accuracy whenever possible, resorting to lower order when necessary to maintain stability, with the order ranging from one to five. The sth-order BDF method (Figure 1C) can be expressed as:

$$\alpha_0 y_n + \alpha_1 y_{n+1} + \cdots + \alpha_s y_{n+s,} = h(\beta_0 f_n + \beta_1 f_{n+1} + \cdots + \beta_s f_{n+s}),$$
(13)

where α and β are the coefficients of the BDF method, and h is the time step. y_{n+j} and f_{n+j} represent $y_h(t_{n+j})$ and $f(y_{n+j}, t_{n+j})$, respectively. The BDF method requires iteration to determine the solution at the following time step. If Equation (13) is expressed as a linear system, each iteration involves solving Equation (14)

$$\mathbf{Ay} = \mathbf{f},$$
(14)

where $\mathbf{y} = [y_n \quad y_{n+1} \quad \cdots \quad y_{n+s}]^T$, the coefficient matrix $\mathbf{A}$ depends on s and α, and the right-hand vector $\mathbf{f}$ depends on $h \sum_{j=0}^{s} \beta_j f_{n+j}$.

2.3 | Solver With Differentiable Operators

This section shows how to integrate the solver with the differentiability and backpropagation of NNs to construct DP-SOL for hybrid modeling. The temporal integration in Equation (13) approximates the unknown y_{n+1} using the known y_n, which can be understood as applying a series of operators $\mathcal{P}_1, \mathcal{P}_2, \cdots, \mathcal{P}_m$ to y_n such that

$$\mathcal{P}_1 \circ \mathcal{P}_2 \circ \cdots \circ \mathcal{P}_m \circ y_n = y_{n+1},$$
(15)

where $\circ$ represents function decomposition, i.e., $\mathcal{P}_m \circ y_h(t) = \mathcal{P}_m(y_h(t))$. These operators can be some form of transformation or operation. For instance, a matrix is considered a linear operator that maps one vector space to another. The BDF temporal integration in Equation (14) is an operator, which maps $\mathbf{y}$ to $\mathbf{f}$ by operator $\mathbf{A}$. Similarly, the forward propagation of NNs in Equation (1), is also an operator that maps $\mathbf{X}^{l-1}$ to $\mathbf{Y}^l$ by operators $\mathbf{W}^l$ and $\mathbf{b}^l$. Therefore, if we combine the temporal integration and the forward propagation into one solver, all computation processes in this solver can be expressed in the form of operators.

We refer to these processes as the forward propagation of DP-SOL (Figure 1A), which is expressed by the operator $\mathcal{F}$ for the spatial discretization of the governing equation according to Equtaion (12), by the operator $\mathcal{SOL}$ for the temporal integration of the BDF solver through Equation (14), and by the operator $\mathcal{NN}$ for the forward propagation of NNs in Equations (1) and (2):

$$\mathbf{Y}_n = \mathcal{SOL}[\mathcal{F}(\mathbf{X}_{n-1}) \cdot \mathcal{NN}(\mathbf{X}_{n-1})],$$
(16)

$$\mathbf{X}_n = \mathbf{Y}_n.$$
(17)

Similarly, to implement the backpropagation of DP-SOL (Figure 1A), it is necessary that both NNs and the BDF solver

can perform backpropagation. The NN backpropagation is based on its differentiability. So, the BDF solver also needs differentiability. Since the BDF solver can be expressed by differentiable operators, it is differentiable by taking the partial derivative of both sides of Equation (15) with respect to y_n:

$$\frac{\partial(\mathcal{P}_1 \circ \mathcal{P}_2 \circ \cdots \circ \mathcal{P}_m \circ y_n)}{\partial y_n} = \frac{\partial y_{n+1}}{\partial y_n},$$
(18)

which is rewritten by the definition of differentiable operators and the chain rule as:

$$\frac{\partial \mathcal{P}_m(y_n)}{\partial y_n} \frac{\partial \mathcal{P}_{m-1}(\mathcal{P}_m)}{\partial y_n} \cdots \frac{\partial \mathcal{P}_1(\mathcal{P}_2)}{\partial y_n} = \frac{\partial y_{n+1}}{\partial y_n},$$
(19)

where $\partial \mathcal{P}_m / \partial y_n$ represents the Jacobian matrix. Thus, $\partial y_{n+1} / \partial y_n$ can be obtained.

Combining the derivations of forward and backward propagation reported above, we obtain DP-SOL as shown in Figure 1A, where all calculations can be represented in terms of differentiable operators. DP-SOL is iteratively evaluated during time stepping. When the contribution of $\mathcal{NN}$ in Equation (16) is equal to one, DP-SOL degenerates into a general numerical method. Therefore, the contribution of $\mathcal{NN}$ can be understood as a correction to the output of governing equation $\mathcal{F}$ or the input of $\mathcal{SOL}$. DP-SOL has strong generalization abilities and can be applied to model various processes by altering the governing equations, discretization forms, solvers, and NN structures.

2.4 | Mechanistic Models and Numerical Solution

DP-SOL involves three operators. $\mathcal{SOL}$ (Figure 1C) and $\mathcal{NN}$ (Figure 1D) have been identified above, leaving only $\mathcal{F}$ (Figure 1B), which comprises contributions from both the governing equation and its discrete form. In this study, $\mathcal{F}$ is adapted to the description of RPC processes. The equilibrium-dispersive and dispersed plug flow models are used for the adsorbate components i and the modifier m (organic solvents), respectively:

$$\frac{\partial c_i}{\partial t}(z, t) = -\frac{u}{\varepsilon_t} \frac{\partial c_i}{\partial z}(z, t) + D_{\text{app}} \frac{\partial^2 c_i}{\partial z^2}(z, t) - \frac{1 - \varepsilon_t}{\varepsilon_t} \frac{\partial q_i}{\partial t}(z, t),$$
(20)

$$\frac{\partial c_m}{\partial t}(z, t) = -\frac{u}{\varepsilon_t} \frac{\partial c_m}{\partial z}(z, t) + D_{\text{app}} \frac{\partial^2 c_m}{\partial z^2}(z, t),$$
(21)

where u, ε_t, and D_{app} are the superficial velocity, total porosity, and apparent axial dispersion coefficient, respectively. A simplified SDM is employed for the stationary phase (Arkell et al. 2017; Mollerup 2007):

$$k_{\text{kin},i} \cdot \frac{\partial q_i}{\partial t}(z, t) = k_{\text{eq},i} \cdot \left[\Lambda - \sum_{j=1}^{n} (\nu_j + \sigma_j) \cdot q_j(z, t)\right]^{\nu_i} \cdot c_i(z, t) - q_i(z, t) \cdot c_m^{\nu_i}(z, t),$$
(22)

$$\frac{\partial q_m}{\partial t}(z, t) = -\sum_{j=1}^{n} \nu_j \cdot \frac{\partial q_j}{\partial t}(z, t), \tag{23}$$

where Λ represents the ligand density. $k_{\mathrm{kin}}, k_{\mathrm{eq}}, \nu,$ and σ represent the model parameters, namely the kinetic coefficient, equilibrium constant, characteristic charge, and shielding factor. For both adsorbate components i and the modifier m, Danckwerts and Neumann boundary conditions are applied at the inlet ($z = 0$) and outlet ($z = L$) of the chromatographic column, respectively:

$$-\frac{\varepsilon_t D_{\mathrm{app}}}{u} \frac{\partial c}{\partial z}(0, t) + c(0, t) = \begin{cases} c_{\mathrm{inj}} & 0 < t \le t_{\mathrm{inj}} \\ 0 & t > t_{\mathrm{inj}} \end{cases}, \tag{24}$$

$$\frac{\partial c}{\partial z}(L, t) = 0, \tag{25}$$

where c_{inj} and t_{inj} represent the injection concentration and time, respectively. As initial conditions, we assumed a fully regenerated and equilibrated column.

Another essential element of operator $\mathcal{F}$ is the discrete form (obtaining Equation (12) from Equation (11)). A discontinuous Galerkin finite element method (DG-FEM) with order n_p, number of elements K, and Legendre polynomial basis functions ℓ is used:

$$y(z, t) \approx y_h(z, t) = \bigoplus_{k=1}^{K} \sum_{i=1}^{n_p} y_h^k(z_i, t) \ell_i^k(z). \tag{26}$$

Details of the DG-FEM method for solving chromatographic models can be found in Breuer et al. (2023); Javeed et al. (2011); Meyer et al. (2020).

2.5 | Integration of DP-SOL With RPC Models

DP-SOL is coupled with the RPC mechanistic models. For a ternary separation system, including a weakly adsorbed impurity (W), a main product (P), and a strongly adsorbed impurity (S), and linear gradient elution performed with an organic modifier, DP-SOL can be represented as:

$$\hat{\mathbf{C}} = \mathcal{DPSOL}(\mathbf{t}), \tag{27}$$

$$\mathbf{t} = \begin{bmatrix} t_0 & t_1 & \cdots & t_f \end{bmatrix}_{1 \times n_t}^{T}, \tag{28}$$

$$\hat{\mathbf{C}} = \begin{bmatrix} \mathbf{c}_{\mathrm{out},m} & \mathbf{c}_{\mathrm{out},\mathcal{W}} & \mathbf{c}_{\mathrm{out},\mathcal{P}} & \mathbf{c}_{\mathrm{out},\mathcal{S}} \end{bmatrix}_{n_t \times 4}, \tag{29}$$

where the total inputs are a time series $\mathbf{t}$ and other operating conditions. t_f is the final time of the simulation, and the series length n_t depends on the sampling frequency of the fraction collector. The total output is the mobile-phase concentration at the column outlet $\hat{\mathbf{C}}$, where each element is a vector to t.

In each iteration of DP-SOL, the input to $\mathcal{F}$ is:

$$\begin{bmatrix} \mathbf{c}_m & \mathbf{c}_{\mathcal{W}} & \mathbf{c}_{\mathcal{P}} & \mathbf{c}_{\mathcal{S}} & \mathbf{q}_m & \mathbf{q}_{\mathcal{W}} & \mathbf{q}_{\mathcal{P}} & \mathbf{q}_{\mathcal{S}} \end{bmatrix}_{[K \cdot (n_p+1)] \times 8}, \tag{30}$$

where each element is a vector with respect to z. $K \cdot (n_p + 1)$ implies its length is determined by the discrete order and the number of elements. The output of $\mathcal{F}$ is:

$$\begin{bmatrix} \dfrac{d\mathbf{c}_m}{dt} & \dfrac{d\mathbf{c}_{\mathcal{W}}}{dt} & \dfrac{d\mathbf{c}_{\mathcal{P}}}{dt} & \dfrac{d\mathbf{c}_{\mathcal{S}}}{dt} & \dfrac{d\mathbf{q}_m}{dt} & \dfrac{d\mathbf{q}_{\mathcal{W}}}{dt} & \dfrac{d\mathbf{q}_{\mathcal{P}}}{dt} & \dfrac{d\mathbf{q}_{\mathcal{S}}}{dt} \end{bmatrix}_{[K \cdot (n_p+1)] \times 8}. \tag{31}$$

The input to $\mathcal{NN}$ is obtained by excluding $\mathbf{q}_m$ in Equation (30), and normalized by:

$$\tilde{\mathbf{c}}_i = \frac{1}{c_{\mathrm{inj},i}} \cdot \mathbf{c}_i, \tag{32}$$

$$\tilde{\mathbf{q}}_i = \frac{1}{c_{\mathrm{inj},i}} \cdot \mathbf{q}_i, \tag{33}$$

where the normalization of modifier concentration is unnecessary. The output of $\mathcal{NN}$ retains only the last three terms of Equation (31), indicating that the correction of DP-SOL is only for the adsorption isotherm.

The output of $\mathcal{SOL}$ is similar to Equation (30), while its input is the product of $\mathcal{F}$ output and $\mathcal{NN}$ output. The same normalization of Equations (32) and (33) is applied to the total output of DP-SOL, resulting in the normalized total output $\hat{\mathbf{C}}$. This, along with the observed $\mathbf{C}$, constitutes the mean squared error (MSE):

$$\mathcal{L} = \frac{\|\hat{\mathbf{C}} - \mathbf{C}\|_{L^2}^2}{\|\mathbf{C}\|_{L^2}^2}. \tag{34}$$

3 | Materials and Methods

3.1 | Experiments

The oligonucleotide utilized was a 20-mer single-stranded DNA (5'-ATA CCG ATT AAG CGA AGT TT-3') provided by YMC Japan. The RPC experiments were conducted on a ContiChrom CUBE30+ (YMC ChromaCon), which included an external Blue-Shadow 40D UV/Vis detector set at 300 nm and an external Azura CT2.1 column thermostat with solvent pre-heat cartridge operated at 50°C, both from Knauer. A YMC Triart C18-S column, 100 mm in length, 4.6 mm in internal diameter, particle size of 10 μm and pore size of 12 nm with a total porosity of 0.54 was used for elution experiments. The column specific parameter Λ was determined via silanol titration experiments (Arkell et al. 2017), ε_t was determined by pulse injection experiments using NaCl as a nonbinding tracer (Chen et al. 2024), while D_{app} was calculated from column efficiency via u, column length L and stage number N:

$$D_{\mathrm{app}} = \frac{uL}{2\varepsilon_t N}. \tag{35}$$

For elution experiments, buffer A consisted of 99% 0.2 M sodium acetate and 1% acetonitrile, while buffer B comprised 90% 0.2 M sodium acetate and 10% acetonitrile. The operating conditions (buffer composition, velocity, and volume) are detailed in Table 1, and loading volumes and gradient lengths are

reported in Table 2. The measured conductivity was converted to %B using the Kohlrausch's law (Carta and Jungbauer 2020).

All fractions collected from the elution experiments were analyzed by reversed-phase high-performance liquid chromatography. The analytical runs were performed using a YMC Triart C18 column (100×2 mm, particle size $= 1.9 \mu$m, porosity $= 12$ nm) on an Agilent 1200 HPLC system, with detection at 300 nm using a diode-array detector. The temperature and flow rate were constant to 50°C and 0.2 mL/min, respectively. The equilibration buffer was a 100 mM hexafluoro isopropanol + 4 mM triethylamine solution. The elution buffer was pure methanol. The buffers were filtered through 0.2 μm PVDF membranes and degassed before usage. The system was equilibrated at 5% elution buffer for 1 min, before a gradient to 10% elution buffer in 2 min, from 10% to 15% in 22 min and to 90% in additional 2 min. Finally, re-equilibration at 5% elution buffer was conducted for 15 min.

Impurities eluted within 8 min after the gradient start were considered negligible, as they did not co-elute with the main product (P) during preparative experiments. Impurities co-eluting in the front of P during preparative runs were grouped as weakly adsorbed impurities (W), and were separated via HPLC between 8 and 30.2 min. The product P eluted at 30.2 min. Species eluting after the product P were identified as strongly adsorbed impurities (S). More details about data collection and analytical method can be found in the single column methods of our previous work (Fioretti et al. 2024).

3.2 | Model Implementation in Pytorch

The three operators ($\mathcal{F}$, $\mathcal{NN}$, and $\mathcal{SOL}$) were implemented using PyTorch-supported tensor formats, with their corresponding gradients computed through PyTorch AD framework. All computations were executed on an Intel i9-13900K CPU utilizing single-precision floating-point arithmetic, without GPU acceleration.

The operator $\mathcal{NN}$ involved linear layers from PyTorch as FCNNs, with seven input nodes, three output nodes, and $n_{\mathcal{L}}$ hidden layers (each with n_{ℓ} nodes). Weight initialization for the linear layers was performed using the Kaiming method (He et al. 2015). ReLU served as the activation function between linear layers, while Softplus was utilized as the activation function for the output layer. The MSE function, Equation (34), was adopted as the loss function. For the operator $\mathcal{SOL}$, the linear system in each iteration of the BDF method was solved using the LU decomposition method provided by PyTorch.

3.3 | Network Initialization Based on Mechanistic Models

The NNs were initialized using a mechanistic model. This was preliminarily calibrated through the standardized approach proposed by Chen et al. (2024), including a parameter-by-parameter method followed by an inverse method (Chen et al. 2022, 2023; Yang et al. 2024; Yang et al. 2024). It is worth underlying that this training set used to calibrate the mechanistic model was identical to the one used afterwards to train DP-SOL. The loss function, normalization methods, data set partitioning for mechanistic modeling were consistent with those used in DP-SOL. Once the network initialization was completed, the mechanistic model parameters remained unchanged throughout the NN training process.

3.4 | Hyperparameter Selection

Before formal NN training, we utilized grid search to select the relevant hyperparameters. The NN underwent 20 epochs on the training set during each iteration. The random seed for layer initialization was consistent across iterations.

3.5 | Learning Procedure

The learning procedure included: (1) partitioning of the experimental datasets into training and testing sets; (2) independent obtainment of a calibrated mechanistic model from the training set for initializing NNs; (3) selection of the relevant

TABLE 1 | Operating conditions of linear gradient elution experiments. CV: column volume.

Step	Buffer (%B)	Velocity (cm/h)	Volume (CV)
Equilibration	30	400	3
Loading	Feed mixture	300	As in Table 2
Wash	30	150	2
Gradient	30–100	200	As in Table 2
Strip	100	150	2

TABLE 2 | Loading and gradient lengths of linear gradient elution experiments. CV: column volume.

Exp no.	Label	Load (g/L$_{resin}$)	Percentage of benchmark	Gradient length (CV)	Percentage of benchmark	Purpose
1	Load15_GL6p5	15	100%	6.5	100%	Training
2	Load7p5_GL6p5	7.5	50%	6.5	100%	Training
3	Load15_GL8p5	15	100%	8.5	130%	Training
4	Load15_GL4p6	15	100%	4.6	70%	Testing
5	Load22p5_GL6p5	22.5	150%	6.5	100%	Testing
6	Load22p5_GL8p5	22.5	150%	8.5	130%	Testing

hyperparameters; (4) training of DP-SOL on the training set for 2000 iterations under the selected hyperparameters; (5) evaluation of the extrapolability of the trained DP-SOL to the testing set.

4 | Results

4.1 | Data Preparation and Partitioning

A conscious data set partitioning in training and test sets can ensure the model performs well on unseen data. To prepare the data set, we followed a one-factor-at-a-time approach starting from the experimental conditions reported in Table 1 with a gradient length of 6.5 CV and a loading of $15\,g/L_{resin}$. A total of six experiments, varying the gradient length and loading by $\pm30\%$ and $\pm50\%$, respectively, were divided into a training set (experiments No. 1, 2, and 3 in Table 2) and a testing set (experiments No. 4, 5, and 6 in Table 2) in a 1:1 ratio. From these runs, we collected fractions during the gradient and analyzed them via HPLC to detect and quantify the different species. Among the different components identified, we lumped together those species showing similar behavior. In particular, the impurities eluting earlier then the product were grouped in the pseudo-component W. These are mainly represented by shortmers, and in particular by the *n-1* oligonucleotide. On the other side, the impurities eluting later than the product were grouped in the pseudo-component S. Therefore, the multi-component system was reduced to a ternary mixture comprising W, P and S.

4.2 | Neural Network Initialization Based on Mechanistic Model

NNs were initialized by a calibrated mechanistic model, which also serves as a benchmark for comparison with the hybrid model. The parameters used for the mechanistic model were derived according to the methods reported in Section 3.3 and are listed in Table 3. These parameters are physically meaningful. For example, ν increases moving from W to P and S, and in turn with their retention times.

The simulated elution curves based on these model parameters are shown in Figure S1 in Supporting Information. Due to significant differences in the content of product and impurities in the crude mixture, a more convenient evaluation of the model performance was provided in Figure S2, where the concentration at the column outlet is normalized by the corresponding one in the feed. Through this model, the averaged losses were $\mathcal{L}_{train,MM} \approx 1.1 \times 10^{-1}$ and $\mathcal{L}_{test,MM} \approx 2.6 \times 10^{-1}$ for

TABLE 3 | Mechanistic model parameters of weakly adsorbed impurities (W), main product (P), and strongly adsorbed impurities (S).

Component	ν	k_{eq}	σ	k_{kin}
W	6.08	7.38E-5	29.17	2.89E-7
P	7.70	7.31E-6	12.49	4.49E-10
S	12.63	8.15E-9	0	1.56E-12

the training and test set, respectively. These values suggest that the mechanistic model can already provide a good representation of the RPC process for this oligonucleotide. Indeed, the retention time is accurately captured.

However, the peak shapes were not precisely modeled. Specifically, for the impurity S, the measured peak is tailing (Langmuirian adsorption) while the simulated one is fronting (anti-Langmuirian adsorption), resulting in its R^2 being lower than those for the impurity W and the product P (especially for Figure S2D). According to our previous work (Chen et al. 2025), interpreting adsorption types based on peak symmetry (tailing vs. fronting) requires constant flow conditions. However, as shown by the flow rate profiles in Figures S1 and S2, the elution flow rate was not constant during the elution of the impurity S. Therefore, peak symmetry alone cannot be used to reliably determine the adsorption mechanism in this case.

Additionally, as the product P eluted, a significant variance is observed in the measured conductivity due to displacement effects (Chen et al. 2024; Fioretti et al. 2022). SDM can simulate this variance (Arkell et al. 2017; Arkell et al. 2018), but fails to be fully consistent with experiments.

4.3 | Hyperparameter Selection

Starting from the results of the mechanistic model, we initialized DP-SOL and investigated the effects of the hyperparameters (the number of hidden layers $n_{\mathcal{L}}$, the number of nodes n_ℓ, and the learning rate α) introduced by the operator $\mathcal{NN}$.

For the operator $\mathcal{F}$, also the order n_p and the number of elements K have an effect on the accuracy and computational cost of the discretization (Breuer et al. 2023; Javeed et al. 2011; Meyer et al. 2020). However, the investigation of n_p and K are beyond the scope of this study, and they were fixed to 3 and 7, respectively. For the operator $\mathcal{SOL}$, the integration step size was fixed at 5‰ of the total simulation time.

For the operator $\mathcal{NN}$, the search ranges of $n_{\mathcal{L}}$ and n_ℓ were determined as follows: a NN with $n_{\mathcal{L}} = 2$ can approximate any continuous function, hence $n_{\mathcal{L}} \in \{1, 2, 3\}$; the number of input and output nodes of the operator $\mathcal{NN}$ are 7 and 3, respectively, so n_ℓ was set from 3 (equal to the output node number) to 21 (three times the input node number). The activation function was not considered as a hyperparameter. We expected the sign of the input to the operator $\mathcal{SOL}$ to be controlled solely by the operator $\mathcal{F}$. This requirement can be met by the constantly positive Softplus, which equals to ReLU for the sufficiently large input. Adam optimizer performs better for few-shot learning compared to other optimizers and was not considered as a hyperparameter. PyTorch recommends $\alpha = 10^{-4}$ for the Adam optimizer, hence $\alpha \in \{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$.

The hyperparameter search results in Figure 2A–D demonstrate the effects of NN structures ($n_{\mathcal{L}}$ and n_ℓ) on $\mathcal{L}$ at different α. The optimal α is 10^{-3} as identified in Figure 2C. A comparison of Figure 2A–D reveals that both excessively large or small α may cause a high $\mathcal{L}$. In fact, a too large α may cause drastic fluctuations in $\mathcal{L}$, making the training process unstable and
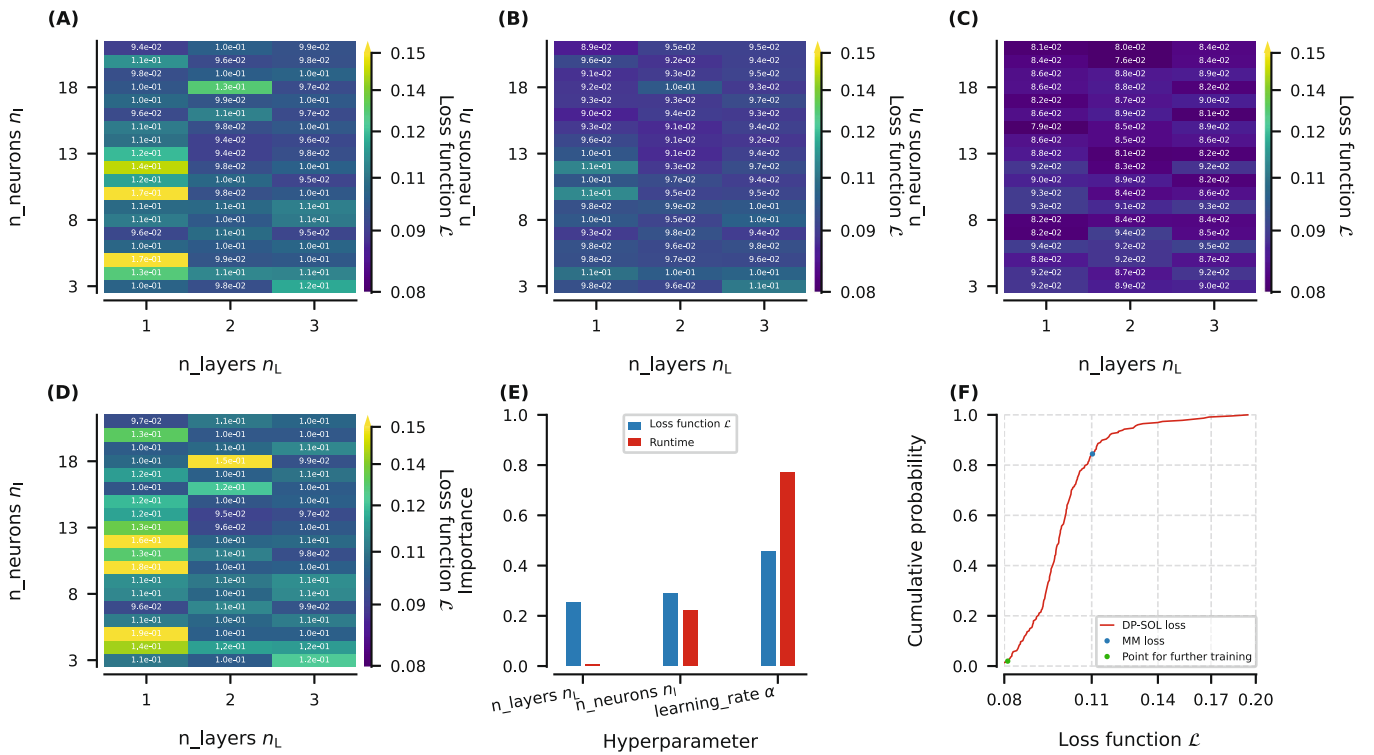
FIGURE 2 | Hyperparameter search results: Contours of loss function between layer number and neuron number at learning rates of: (A) 10^{-1}, (B) 10^{-2}, (C) 10^{-3}, and (D) 10^{-4}. (E) Hyperparameter importance for loss function and runtime. (F) Cumulative probability of empirical distribution.

convergence difficult to reach; while a too small α may result in very slow model convergence or in getting trapped in local optima.

Additionally, increasing $n_{\mathcal{L}}$ does not bring much performance improvement, consistent with Ding et al. (2023) findings. The NN with 14 nodes and 2 layers shows nearly optimal performance across various α. According to Linoff and Berry (2011), NNs can achieve optimal performance when n_{ℓ} does not exceed twice the input node number (which is 7 in this study, making the upper limit 14).

The effects of these three hyperparameters are quantitatively evaluated in Figure 2E, exploring the importance and sensitivity of hyperparameters. This exploration was implemented through AD, which has been employed to explore the sensitivity of chromatographic model parameters (Püttmann et al. 2016; Püttmann et al. 2013). Similarly, it can extend to hyperparameters. Figure 2E indicates that the effects of $n_{\mathcal{L}}$ and n_{ℓ} on both $\mathcal{L}$ and training time are relatively low, while α has a dominant effect.

Figure 2F shows the cumulative probability of the empirical distribution of these grid searches. The blue dot represents the calibration error of the mechanistic model, indicating that during the search process, approximately 85% of the hybrid model outperforms the mechanistic models after only 20 epochs. This demonstrates the significant improvement brought by DP-SOL.

Considering NN performance, prevention of overfitting, computational effort, and training time, we considered $n_{\mathcal{L}} = 2$, $n_{\ell} = 14$, and $\alpha = 10^{-3}$ as a good starting point for further training. Its corresponding $\mathcal{L}$ is shown as a green dot in

Figure 2F, where it is possible to observe that it was lower than the values obtained with 98% of other sets of hyperparameters. It is worth mentioning that this optimal set does not lie on the boundaries, hence the search ranges of hyperparameters were reasonable.

4.4 | Model Training

Using the selected hyperparameters ($n_{\mathcal{L}} = 2$, $n_{\ell} = 14$, and $\alpha = 10^{-3}$), DP-SOL was trained 2000 epochs (100 times the number of hyperparameter selections) with the evolution of $\mathcal{L}_{\text{train,DPSOL}}$ with the number of iterations shown in Figure 3A. The initial $\mathcal{L}$ after NN initialization is 1.1×10^{-1}, which is close to $\mathcal{L}_{\text{train,MM}}$ obtained with the mechanistic model. This indicates that NNs have received good initialization and normalization, without significant fluctuations introduced by NNs. From the second epoch, $\mathcal{L}_{\text{train,DPSOL}}$ remains below $\mathcal{L}_{\text{train,MM}}$ and steadily decreases with increasing iterations until convergence. Training for 2000 epochs is reasonable, as the model reaches the optimal at the 988th epoch with $\mathcal{L}_{\text{train,DPSOL}} = 6.8 \times 10^{-3}$, which is significantly smaller than $\mathcal{L}_{\text{train,MM}}$.

Figure 3B illustrates the contributions of the three experiments used for training to the total $\mathcal{L}_{\text{train,DPSOL}}$. With increasing iterations, $\mathcal{L}_{\text{train,DPSOL}}$ is dominated by the experiment at a loading of $15\,\text{g/L}_{\text{resin}}$ and a gradient length of 6.5 CV (orange shaded area). This is likely due to the relatively high error associated with the impurity S in this experiment.

The elution curves simulated by the trained DP-SOL are depicted in Figure 4. The normalized chromatograms
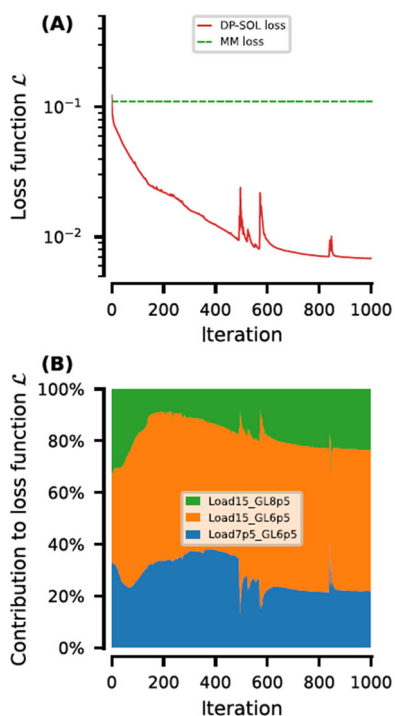
FIGURE 3 | (A) Loss function from DP-SOL and mechanistic model (MM) during consecutive iterations. (B) Contribution to the loss function from the three different experiments during model training.

(Figure 4A–C) show that DP-SOL can accurately reproduce the experimental measurements in terms of both conductivity curves and elution curves of oligonucleotides during RPC processes (overall R^2 exceeds 0.97, and individual R^2 were also provided in brackets for each component in the legend).

Regarding the conductivity curves (black lines in Figure 4), the developed model exhibits similar variations with the measurements, in spite of failing to precisely replicate the magnitude.

Regarding the elution curves of oligonucleotides (colored lines in Figure 4), satisfactory fitting results were achieved in peak height, shape, and retention time, across the different loadings and gradient lengths. For the impurities W and S, despite their low concentrations, the model can remarkably simulate their elution curves. This ideal outcome is closely related to the normalization, including inputs of the operator $\mathcal{NN}$, total outputs of DP-SOL, and $\mathcal{L}$. The normalization effectively scales features to similar ranges, mitigating the dominance of certain features in gradient updates and thus reducing the risk of gradient vanishing or exploding. In unnormalized chromatograms (Figure 4D–F), the overall R^2 almost equals the individual R^2 of the product P because its concentration is higher than those of impurities (W and S) by an order of magnitude. Conversely, in normalized chromatograms (Figure 4A–C), all peaks are scaled to have roughly the same peak areas, enhancing the reflection of overall fitting in R^2.

A comparison based on the same experimental data set, considering chromatograms normalized following identical approach, reveals that the hybrid model (Figure 4A–C) inherits the learning capability of NNs and can capture the adsorption behavior of the impurity S under varying velocities, despite the simplicity of both the column

model and NN architecture, and the hybridization being applied solely to the adsorption model. This was not possible instead with the mechanistic model (Figure S2 in Supporting Information), showing large inaccuracy for this pseudo-component.

4.5 | Model Testing

The trained DP-SOL was then applied to the simulation of the elution conditions selected for testing (experiments No. 4, 5 and 6 in Table 2). Figure 5 reveals that the trained DP-SOL can excellently predict the elution curves at gradient lengths and loading conditions different from those considered for model training, despite some oscillations in certain curves (peak position of product P).

During the hybrid model data set preparation, a one-factor-at-a-time design principle (i.e., varying either elution gradient length or loading) was adopted to ensure that the model could capture the individual effects of these variables. As summarized in Table 3, different velocities were applied across the various chromatographic steps. The effect of changing velocity was accounted for into the hybrid model through the convection and dispersion terms in Equation (20), where D_{app} was modeled as flow-dependent in accordance to Equation (35).

However, R^2 in Figure 5A (15 g/L_{resin} loading and 4.6 CV gradient length) is only 0.784, primarily due to the poor predictivity for impurity S, as it can be deduced from its individual R^2. The impurity S was experimentally detected in only two fractions, because analytical results showed that the impurity S accounted for only 1.2% of the crude mixture, and the low concentration may limit the accuracy of quantification. This can be observed in the unnormalized chromatogram (Figure 5D), where the elution curve of impurity S approaches a horizontal line. The loss in model performance due to anomalous datasets does not justify poor extrapolation of DP-SOL. This experiment was used as a testing rather than for model training, effectively minimizing the impact of analytical accuracy on model development. Despite having only two data points, the hybrid model was still able to predict the peak shape reasonably well.

5 | Discussion

In this study, the training set comprises only three linear gradient elution experiments. In contrast to purely data-driven models relying on extensive data set, this few-shot learning strategy of DP-SOL offers convenience to users in data preparation. This is particularly important in the biopharmaceutical field, characterized by limited data availability and high experimental costs. The cost for preparing the data set of DP-SOL is comparable to traditional mechanistic models, enabling users to seamlessly transition from conventional mechanistic models to hybrid ones without incurring additional experimental costs. Furthermore, it allows users to compare the performance of different models on the same data set.

Unlike most data-driven models (interpolative prediction), DP-SOL is assessed through extrapolative prediction. The model is trained using gradient lengths ranging from 6.5 to 8.5 CV and
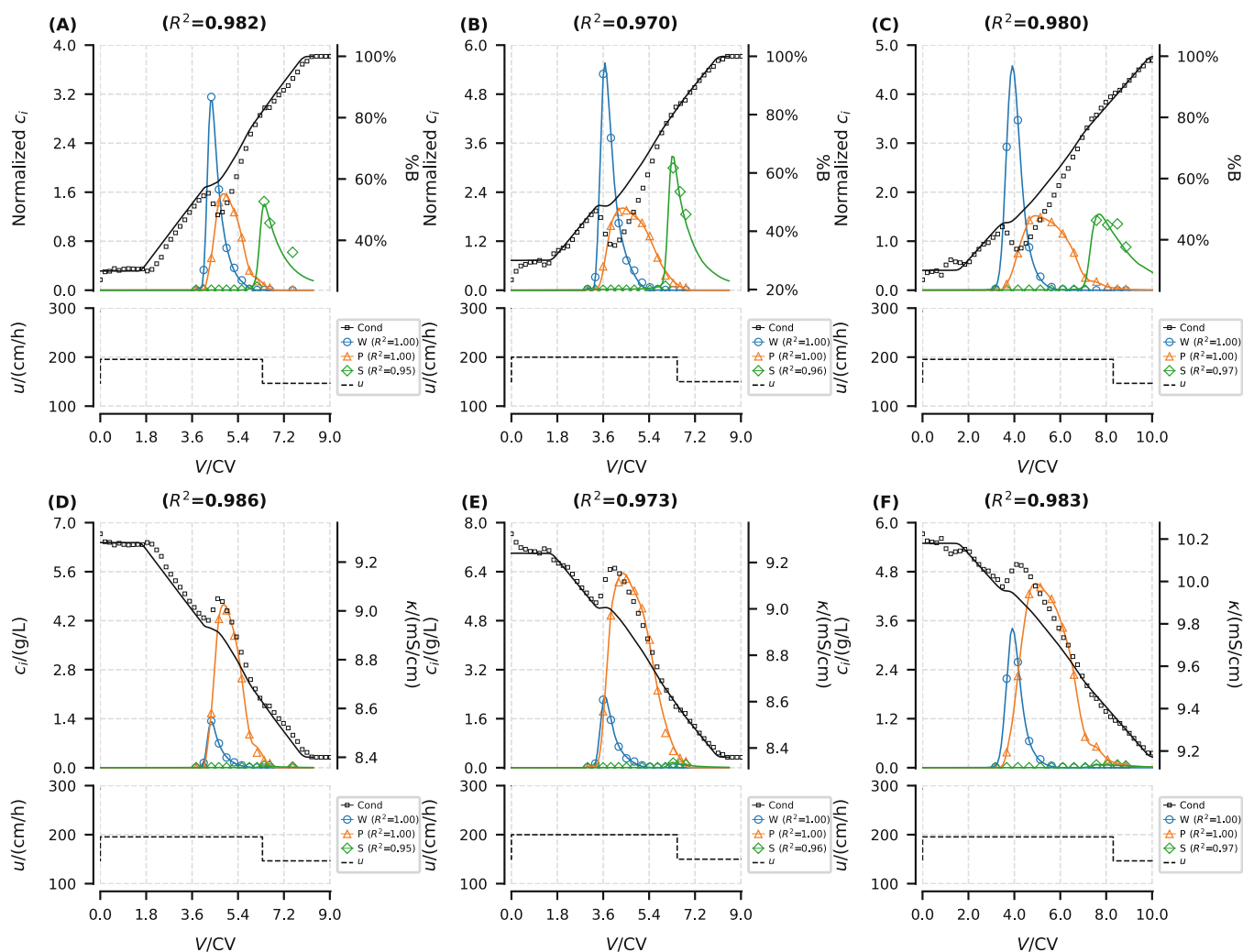
FIGURE 4 | Trainig results of DP-SOL hybrid models: normalized (A–C) and unnormalized (D–F) chromatograms from model simulations (solid lines) and experimental measurements (scatters) for weakly adsorbed impurities (W), main product (P), and strongly adsorbed impurities (S). (A) and (D): 7.5 g/L$_{resin}$ loading and 6.5 CV gradient length. (B) and (E): 15 g/L$_{resin}$ loading and 6.5 CV gradient length. (C) and (F): 15 g/L$_{resin}$ loading and 8.5 CV gradient length. CV: column volume. Overall and individual R^2 are provided in the brackets in the titles and legend, respectively.

loadings from 7.5 to 15.0 g/L$_{resin}$, while predictions are made at a gradient length of 4.6 CV and a loading of 22.5 g/L$_{resin}$. Therefore, we ensure that there is no overlap between the training and testing sets, making them mutually independent.

In the model training, the gradient-based optimization strategy using AD and backpropagation led to efficient NN training. This method can be equally effective in higher-dimensional parameter spaces and deeper NNs, where the least squares method is inadequate. In the model testing, DP-SOL exhibited a significant performance improvement compared to the purely mechanistic model. We attributed the excellent extrapolability of DP-SOL to its two components.

On one hand, contribution stems from the mechanistic (white-box) model, for example, the operator $\mathcal{F}$. SDM performs well in predicting different gradient lengths in the presence of organic solvents, an advantage that has been inherited by DP-SOL. However, recent research indicates that SDM fails in predicting high loadings (Koch et al. 2022; Seelinger et al. 2022; Seelinger et al. 2023; Seelinger et al. 2022), as corroborated in Figure S2E,F

(high loadings), which exhibits larger testing errors compared to other experiments at low loadings.

On the other hand, the contribution of the data-driven (black-box) model, for example, the operator $\mathcal{NN}$. The introduction of NNs enables DP-SOL to interpret unknown mechanisms (Park et al. 2023; Schiemer et al. 2023), such as elution behavior at high loadings, despite employing a very simple column model and shallow NNs, while only considering the contribution of oligonucleotides to the stationary phase. When Ding et al. (2023) utilized NNs to develop a hybrid model for hydrophobic interaction chromatography, the authors considered separate NNs for mass transfer resistance and adsorption isotherms, coupling them to obtain the final output. In comparison, DP-SOL serves as a general modeling paradigm at a solver level. In fact, it eliminates the need for multiple NNs or careful investigation of integration between NNs and PDEs. Therefore, DP-SOL is simpler and more user-friendly.

In terms of interpretability, DP-SOL retains the physical transparency of key transport phenomena, such as mass transfer,
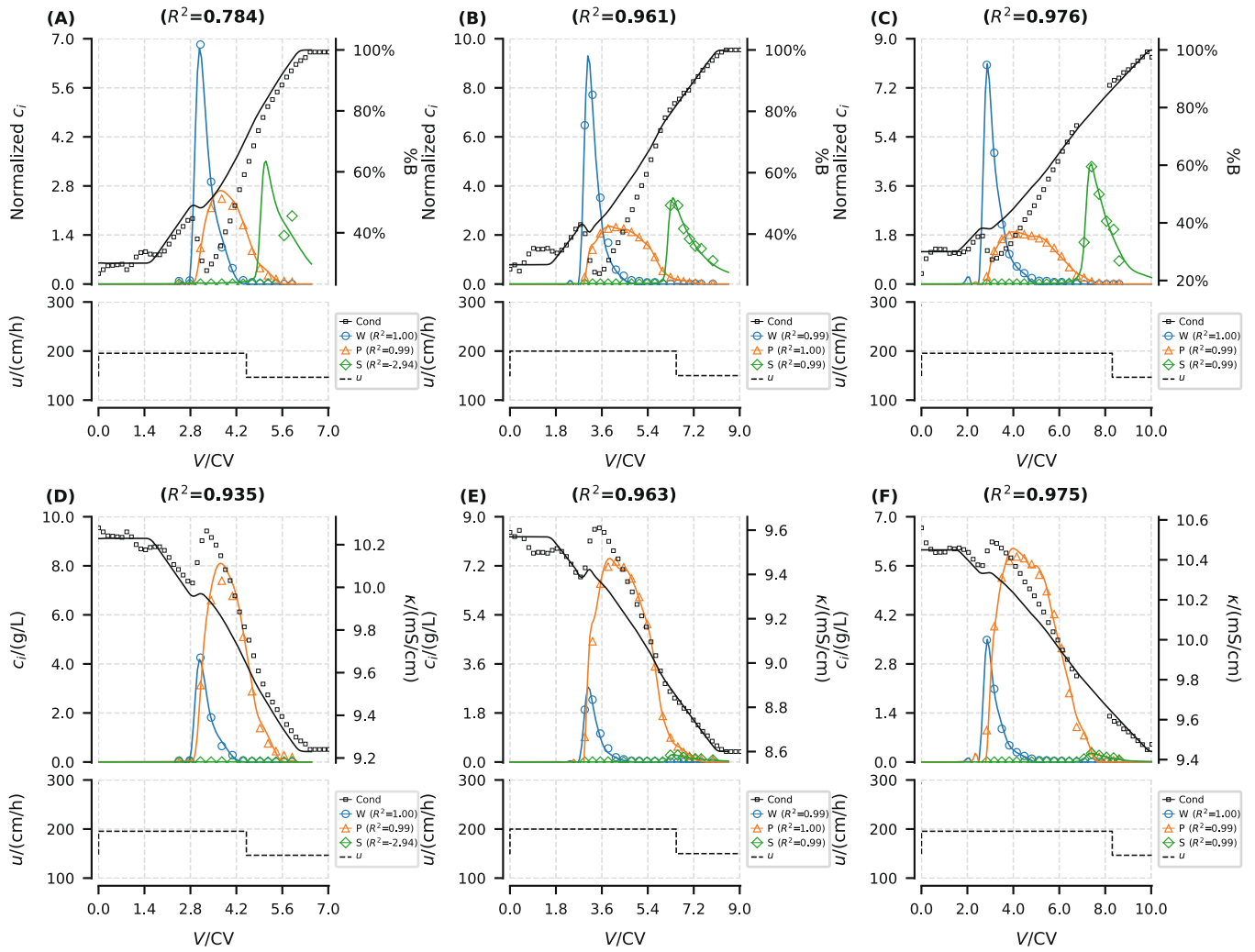
FIGURE 5 | Testing results of DP-SOL hybrid model: normalized (A–C) and unnormalized (D–F) chromatograms of model simulations (solid lines) and measurements (scatters) for weakly adsorbed impurities (W), main product (P), and strongly adsorbed impurities (S). (A) and (D): 15 g/L_{resin} loading and 4.6 CV gradient length. (B) and (E): 22.5 g/L_{resin} loading and 6.5 CV gradient length. (C) and (F): 22.5 g/L_{resin} loading and 8.5 CV gradient length. CV: column volume. Overall and individual R^2 are provided in the brackets in the titles and legend, respectively.

dispersion, and convection, inherited from the mechanistic model. Only the adsorption behavior is learned via the NN operator, enabling the model to generalize beyond its training conditions while maintaining partial interpretability. This structure allows users to trace how input conditions propagate through physical mechanisms to yield model outputs, which is particularly valuable in biopharmaceutical process development where understanding of critical process parameters is essential. From a Quality by Design perspective, DP-SOL might support process optimization, risk assessment, robustness evaluation, and regulatory acceptance more effectively than purely data-driven models.

Overall, the hybridization of DP-SOL leverages the advantages of mechanistic and data-driven models. DP-SOL possesses the capability of data-driven models to interpret unknown mechanisms, which supplements known mechanistic models. A mechanistic model with poor performance can be transformed into a perfect hybrid model using DP-SOL. Actually, DP-SOL falls within the realm of operator learning. By counting the

number of three operators, we observed that for each time step in the loop of DP-SOL, there are 12 operators contributed by NNs for learning and 1920 operators determined by mechanistic models. We calculated the degree of hybridization (proposed by Narayanan et al. [2022]) in this study to be approximately 0.6%. This implies that DP-SOL leans more towards mechanistic modeling rather than data-driven modeling. This elucidates why DP-SOL exhibits characteristics such as few-shot learning, interpretability, and excellent extrapolability, typical of mechanistic models.

However, DP-SOL also inherits the drawbacks of both mechanistic and data-driven models, primarily evident in its implementation. All issues that need to be considered in the implementation of mechanistic and data-driven models must be reevaluated in DP-SOL, including governing equations, discretization, solver selection for mechanistic models, determination of NN structure, and hyperparameter selection for data-driven models. Nevertheless, this also implies that knowledge of numerical methods from known mechanistic models will not be

discarded compared to purely data-driven models. Another drawback lies in computational costs. The computational time of DP-SOL approaches the sum of computational times for mechanistic and data-driven models. Benefitting from efficient numerical methods like DG-FEM and advancements in distributed learning and computing power, DP-SOL will unleash more applications in the future.

As a novel and versatile hybrid modeling paradigm, DP-SOL has the potential to significantly impact modeling approaches in the downstream processing field and the broader biopharmaceutical field. By altering the governing equations, discretization forms, solvers, and NN structures, DP-SOL can be applied to various processes (Andersson et al. 2023; Ding et al. 2022; Gomis-Fons et al. 2024; Tiwari et al. 2023). For example, it can be used in the process design and optimization of multi-column counter-current solvent gradient purification (MCSGP) of oligonucleotides. The gradient information in DP-SOL can conveniently establish the mapping between optimization objectives and decision variables, thereby solving MCSGP optimization problems more efficiently.

6 | Conclusion

This study introduced a hybrid modeling framework named DP-SOL, which integrates traditional mechanistic models with NNs. The framework utilizes differentiable physical operators and their gradients, and was applied to describe the RPC purification of an oligonucleotide. We first collected a small data set for few-shot learning by performing six linear gradient elution experiments at different resin loadings and gradient slopes, with three experiments each for training and testing in a 1:1 ratio. The NN was initialized using a calibrated mechanistic model. Hyperparameters were determined through a grid search, resulting in a NN with 2 hidden layers and 14 nodes. Compared to a purely mechanistic model, the DP-SOL hybrid model showed significant performance improvement, invariably achieving a higher R^2 on both training and testing sets, with $R^2 > 0.97$ for the former. We attributed the robust predictive capability of DP-SOL to the combination of mechanistic models and NNs at the solver level. As a novel and versatile hybrid modeling paradigm, DP-SOL has the potential to significantly impact the chromatographic purification of biologics, establishing digital tools as key resources for reduced process development times and costs.

Author Contributions

Yu-Cheng Chen: methodology, software, investigation, visualization, writing – original draft. **Ismaele Fioretti:** investigation, validation, data curation. **Dong-Qiang Lin:** supervision, funding acquisition, writing – review and editing. **Mattia Sponchioni:** project administration, supervision, funding acquisition, writing – review and editing.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

Andersson, N., J. G. Fons, M. Isaksson, et al. 2023. "Methodology for Fast Development of Digital Solutions in Integrated Continuous Downstream Processing." *Biotechnology and Bioengineering* 121, no. 8: 2378–2387. https://doi.org/10.1002/bit.28501.

Arkell, K., M. P. Breil, S. S. Frederiksen, and B. Nilsson. 2017. "Mechanistic Modeling of Reversed-Phase Chromatography of Insulins With Potassium Chloride and Ethanol as Mobile-Phase Modulators." *ACS Omega* 2, no. 1: 136–146. https://doi.org/10.1021/acsomega.6b00248.

Arkell, K., M. P. Breil, S. S. Frederiksen, and B. Nilsson. 2018. "Mechanistic Modeling of Reversed-Phase Chromatography of Insulins Within the Temperature Range 10–40°C." *ACS Omega* 3, no. 2: 1946–1954. https://doi.org/10.1021/acsomega.7b01527.

Arkell, K., H. K. Knutson, S. S. Frederiksen, M. P. Breil, and B. Nilsson. 2018. "Pareto-Optimal Reversed-Phase Chromatography Separation of Three Insulin Variants With a Solubility Constraint." *Journal of Chromatography A* 1532: 98–104. https://doi.org/10.1016/j.chroma.2017.11.065.

Breuer, J. M., S. Leweke, J. Schmölder, G. Gassner, and E. von Lieres. 2023. "Spatial Discontinuous Galerkin Spectral Element Method for a Family of Chromatography Models in CADET." *Computers & Chemical Engineering* 177: 108340. https://doi.org/10.1016/j.compchemeng.2023.108340.

Carta, G., and A. Jungbauer. 2020. "Introduction to Protein Chromatography." In *Protein Chromatography*, 2nd ed., 63–91. Wiley. https://doi.org/10.1002/9783527824045.ch2.

Catani, M., C. De Luca, J. Medeiros Garcia Alcântara, et al. 2020. "Oligonucleotides: Current Trends and Innovative Applications in the Synthesis, Characterization, and Purification." *Biotechnology Journal* 15, no. 8: 1900226. https://doi.org/10.1002/biot.201900226.

Chen, Y.-C., H. L. Lu, R. Z. Wang, et al. 2024. "Standardized Approach for Accurate and Reliable Model Development of Ion-Exchange Chromatography Based on Parameter-by-Parameter Method and Consideration of Extra-Column Effects." *Biotechnology Journal* 19, no. 3: 2300687. https://doi.org/10.1002/biot.202300687.

Chen, Y.-C., G. Recanati, F. De Mathia, D. Q. Lin, and A. Jungbauer. 2024. "Residence Time Distribution in Continuous Virus Filtration." *Biotechnology and Bioengineering* 121, no. 6: 1876–1888. https://doi.org/10.1002/bit.28696.

Chen, Y.-C., S.-J. Yao, and D.-Q. Lin. 2022. "Parameter-by-Parameter Method for Steric Mass Action Model of Ion Exchange Chromatography: Theoretical Considerations and Experimental Verification." *Journal of Chromatography A* 1680: 463418. https://doi.org/10.1016/j.chroma.2022.463418.

Chen, Y.-C., S.-J. Yao, and D.-Q. Lin. 2023. "Parameter-by-Parameter Method for Steric Mass Action Model of Ion Exchange Chromatography: Simplified Estimation for Steric Shielding Factor." *Journal of Chromatography A* 1687: 463655. https://doi.org/10.1016/j.chroma.2022.463655.

Chen, Y.-C., S.-J. Yao, and D.-Q. Lin. 2024. "Enhancing Thermodynamic Consistency: Clarification on the Application of Asymmetric Activity Model in Multi-Component Chromatographic Separation." *Journal of Chromatography A* 1731: 465156. https://doi.org/10.1016/j.chroma.2024.465156.

Chen, Y.-C., X.-Z. Zhong, C. Shi, et al. 2025. "Mechanistic Modeling of Anti-Langmuirian to Langmuirian Behavior of Fc-Fusion Proteins in

Cation Exchange Chromatography." *Journal of Chromatography A* 1741: 465602. https://doi.org/10.1016/j.chroma.2024.465602.

Ding, C., H. Ardeshna, C. Gillespie, and M. Ierapetritou. 2022. "Process Design of a Fully Integrated Continuous Biopharmaceutical Process Using Economic and Ecological Impact Assessment." *Biotechnology and Bioengineering* 119, no. 12: 3567–3583. https://doi.org/10.1002/bit.28234.

Ding, C., C. Gerberich, and M. Ierapetritou. 2023. "Hybrid Model Development for Parameter Estimation and Process Optimization of Hydrophobic Interaction Chromatography." *Journal of Chromatography A* 1703: 464113. https://doi.org/10.1016/j.chroma.2023.464113.

Fioretti, I., T. Müller-Späth, L. Aumann, and M. Sponchioni. 2024. "UV-Based Dynamic Control Improves the Robustness of Multicolumn Countercurrent Solvent Gradient Purification of Oligonucleotides." *Biotechnology Journal* 19, no. 7: 2400170. https://doi.org/10.1002/biot.202400170.

Fioretti, I., T. Müller-Späth, R. Weldon, S. Vogg, M. Morbidelli, and M. Sponchioni. 2022. "Continuous Countercurrent Chromatographic Twin-Column Purification of Oligonucleotides: The Role of the Displacement Effect." *Biotechnology and Bioengineering* 119, no. 7: 1861–1872. https://doi.org/10.1002/bit.28093.

Gomis-Fons, J., B. Zee, D. Hurwit, J. Woo, J. Moscariello, and B. Nilsson. 2024. "Mechanistic Modeling of Empty-Full Separation in Recombinant Adeno-Associated Virus Production Using Anion-Exchange Membrane Chromatography." *Biotechnology and Bioengineering* 121, no. 2: 719–734. https://doi.org/10.1002/bit.28595.

He, K., X. Zhang, S. Ren, and J. Sun. 2015. Delving Deep Into Rectifiers: Surpassing Human-Level Performance on Imagenet Classification. *arXiv.* https://arxiv.org/abs/1502.01852.

Huuk, T. C., T. Hahn, A. Osberghaus, and J. Hubbuch. 2014. "Model-Based Integrated Optimization and Evaluation of a Multi-Step Ion Exchange Chromatography." *Separation and Purification Technology* 136: 207–222. https://doi.org/10.1016/j.seppur.2014.09.012.

Javeed, S., S. Qamar, A. Seidel-Morgenstern, and G. Warnecke. 2011. "A Discontinuous Galerkin Method to Solve Chromatographic Models." *Journal of Chromatography A* 1218, no. 40: 7137–7146. https://doi.org/10.1016/j.chroma.2011.08.005.

Jungbauer, A., G. Ferreira, M. Butler, et al. 2024. "Status and Future Developments for Downstream Processing of Biological Products." *Biotechnology and Bioengineering* 121, no. 8: 2524–2541. https://doi.org/10.1002/bit.28738.

Jungbauer, A., P. Satzer, A. Duerauer, et al. 2024. "Continuous Downstream Processing." *Separation and Purification Technology* 338: 126439. https://doi.org/10.1016/j.seppur.2024.126439.

Koch, J., D. Scheps, M. Gunne, O. Boscheinen, M. Hafner, and C. Frech. 2022. "Mechanistic Modeling and Simulation of a Complex Low and High Loading Elution Behavior of a Polypeptide in Cation Exchange Chromatography." *Journal of Separation Science* 45, no. 12: 2008–2023. https://doi.org/10.1002/jssc.202200098.

Kozorog, M., S. Caserman, M. Grom, F. A. Vicente, A. Pohar, and B. Likozar. 2023. "Model-Based Process Optimization for mAb Chromatography." *Separation and Purification Technology* 305: 122528. https://doi.org/10.1016/j.seppur.2022.122528.

Linoff, G. S., and M. J. A. Berry. 2011. *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management*, 3rd ed. Wiley. https://books.google.it/books?id=AyQfVTDJypUC.

Liu, Y., X. Zhou, T. Wang, et al. 2024. "Genetic Algorithm-Based Semisupervised Convolutional Neural Network for Real-Time Monitoring of *Escherichia coli* Fermentation of Recombinant Protein Production Using a Raman Sensor." *Biotechnology and Bioengineering* 121, no. 5: 1583–1595. https://doi.org/10.1002/bit.28661.

De Luca, C., S. Felletti, G. Lievore, et al. 2020. "Modern Trends in Downstream Processing of Biotherapeutics Through Continuous Chromatography: The Potential of Multicolumn Countercurrent Solvent Gradient Purification." *TrAC Trends in Analytical Chemistry* 132: 116051. https://doi.org/10.1016/j.trac.2020.116051.

Mahanty, B. 2023. "Hybrid Modeling in Bioprocess Dynamics: Structural Variabilities, Implementation Strategies, and Practical Challenges." *Biotechnology and Bioengineering* 120, no. 8: 2072–2091. https://doi.org/10.1002/bit.28503.

Malinov, N. G., K. Raudenbush-Henderson, C. Ding, J. V. Reddy, and M. G. Ierapetritou. 2024. "End-to-End Process Flowsheet Modeling for Biopharmaceutical Production: Current State and Future Potential." *Current Opinion in Chemical Engineering* 45: 101044. https://doi.org/10.1016/j.coche.2024.101044.

Meyer, K., S. Leweke, E. von Lieres, J. K. Huusom, and J. Abildskov. 2020. "Chromatech: A Discontinuous Galerkin Spectral Element Simulator for Preparative Liquid Chromatography." *Computers & Chemical Engineering* 141: 107012. https://doi.org/10.1016/j.compchemeng.2020.107012.

Mollerup, J. M. 2007. "The Thermodynamic Principles of Ligand Binding in Chromatography and Biology." *Journal of Biotechnology* 132, no. 2: 187–195. https://doi.org/10.1016/j.jbiotec.2007.05.036.

Narayanan, H., M. N. Cruz Bournazou, G. Guillén Gosálbez, and A. Butté. 2022. "Functional-Hybrid Modeling Through Automated Adaptive Symbolic Regression for Interpretable Mathematical Expressions." *Chemical Engineering Journal* 430: 133032. https://doi.org/10.1016/j.cej.2021.133032.

Narayanan, H., M. Luna, M. Sokolov, A. Butté, and M. Morbidelli. 2022. "Hybrid Models Based on Machine Learning and an Increasing Degree of Process Knowledge: Application to Cell Culture Processes." *Industrial & Engineering Chemistry Research* 61, no. 25: 8658–8672. https://doi.org/10.1021/acs.iecr.1c04507.

Narayanan, H., T. Seidler, M. F. Luna, M. Sokolov, M. Morbidelli, and A. Butté. 2021. "Hybrid Models for the Simulation and Prediction of Chromatographic Processes for Protein Capture." *Journal of Chromatography A* 1650: 462248. https://doi.org/10.1016/j.chroma.2021.462248.

Narayanan, H., M. Sponchioni, and M. Morbidelli. 2022. "Integration and Digitalization in the Manufacturing of Therapeutic Proteins." *Chemical Engineering Science* 248: 117159. https://doi.org/10.1016/j.ces.2021.117159.

Narayanan, H., M. von Stosch, F. Feidl, M. Sokolov, M. Morbidelli, and A. Butté. 2023. "Hybrid Modeling for Biopharmaceutical Processes: Advantages, Opportunities, and Implementation." *Frontiers in Chemical Engineering* 5: 1157889. https://doi.org/10.3389/fceng.2023.1157889.

Nozaki, S., N. Yoshimoto, and S. Yamamoto. 2024. "A Method for Determining the Optimum Temperature for a Polyphenol Separation Process by Reversed Phase Chromatography." *Japan Journal of Food Engineering* 25, no. 1: 15–22. https://doi.org/10.11301/jsfe.23636.

Ou, J., Y. Tang, J. Xu, J. Tucci, M. C. Borys, and A. Khetan. 2024. "Recent Advances in Upstream Process Development for Production of Recombinant Adeno-Associated Virus." *Biotechnology and Bioengineering* 121, no. 1: 53–70. https://doi.org/10.1002/bit.28545.

Park, S. Y., S. J. Kim, C. H. Park, J. Kim, and D. Y. Lee. 2023. "Data-Driven Prediction Models for Forecasting Multistep Ahead Profiles of Mammalian Cell Culture Toward Bioprocess Digital Twins." *Biotechnology and Bioengineering* 120, no. 9: 2494–2508. https://doi.org/10.1002/bit.28405.

Püttmann, A., S. Schnittert, S. Leweke, and E. von Lieres. 2016. "Utilizing Algorithmic Differentiation to Efficiently Compute Chromatograms and Parameter Sensitivities." *Chemical Engineering Science* 139: 152–162. https://doi.org/10.1016/j.ces.2015.08.050.

Püttmann, A., S. Schnittert, U. Naumann, and E. von Lieres. 2013. "Fast and Accurate Parameter Sensitivities for the General Rate Model of Column Liquid Chromatography." *Computers & Chemical Engineering* 56: 46–57. https://doi.org/10.1016/j.compchemeng.2013.04.021.

Qian, W., M. Hu, Y. Su, et al. 2023. "Insight Into Mass Transfer During Adsorption of Geniposidic Acid Onto a Fixed-Bed Column by Numerical Simulation Considering Influence of Operating Conditions on Column Adsorption Performance." *Separation and Purification Technology* 319: 124021. https://doi.org/10.1016/j.seppur.2023.124021.

Ramsundar, B., D. Krishnamurthy, and V. Viswanathan. 2021. Differentiable Physics: A Position Piece. *arXiv*. https://arxiv.org/abs/2109.07573.

Rizki, Z., and M. Ottens. 2023. "Model-Based Optimization Approaches for Pressure-Driven Membrane Systems." *Separation and Purification Technology* 315: 123682. https://doi.org/10.1016/j.seppur.2023.123682.

Roush, D., D. Asthagiri, D. K. Babi, et al. 2020. "Toward in Silico CMC: An Industrial Collaborative Approach to Model-Based Process Development." *Biotechnology and Bioengineering* 117, no. 12: 3986–4000. https://doi.org/10.1002/bit.27520.

Saleh, D., R. Hess, M. Ahlers-Hesse, et al. 2022. "A Multiscale Modeling Method for Therapeutic Antibodies in Ion Exchange Chromatography." *Biotechnology and Bioengineering* 120, no. 1: 125–138. https://doi.org/10.1002/bit.28258.

Santana, V. V., M. S. Gama, J. M. Loureiro, et al. 2022. "A First Approach Towards Adsorption-Oriented Physics-Informed Neural Networks: Monoclonal Antibody Adsorption Performance on an Ion-Exchange Column as a Case Study." *Chemengineering* 6, no. 2: 21. https://doi.org/10.3390/chemengineering6020021.

Schiemer, R., J. T. Weggen, K. M. Schmitt, and J. Hubbuch. 2023. "An Adaptive Soft-Sensor for Advanced Real-Time Monitoring of an Antibody-Drug Conjugation Reaction." *Biotechnology and Bioengineering* 120, no. 7: 1914–1928. https://doi.org/10.1002/bit.28428.

Seelinger, F., F. Wittkopp, T. von Hirschheydt, and C. Frech. 2022. "Application of the Steric Mass Action Formalism for Modeling Under High Loading Conditions: Part 2. Investigation of High Loading and Column Overloading Effects." *Journal of Chromatography A* 1676: 463266. https://doi.org/10.1016/j.chroma.2022.463266.

Seelinger, F., F. Wittkopp, T. von Hirschheydt, and C. Frech. 2023. "Anti-Langmuir Elution Behavior of a Bispecific Monoclonal Antibody in Cation Exchange Chromatography: Mechanistic Modeling Using a pH-Dependent Self-Association Steric Mass Action Isotherm." *Journal of Chromatography A* 1689: 463730. https://doi.org/10.1016/j.chroma.2022.463730.

Seelinger, F., F. Wittkopp, T. von Hirschheydt, M. Hafner, and C. Frech. 2022. "Application of the Steric Mass Action Formalism for Modeling Under High Loading Conditions: Part 1. Investigation of the Influence of pH on the Steric Shielding Factor." *Journal of Chromatography A* 1676: 463265. https://doi.org/10.1016/j.chroma.2022.463265.

Söderström, P. 2022. "Physics-Informed Neural Networks for Liquid Chromatography." MSc Thesis, Umeå University. https://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-196783.

Subraveti, S. G., Z. Li, V. Prasad, and A. Rajendran. 2022a. "Can a Computer 'Learn' Nonlinear Chromatography?: Physics-Based Deep Neural Networks for Simulation and Optimization of Chromatographic Processes." *Journal of Chromatography A* 1672: 463037. https://doi.org/10.1016/j.chroma.2022.463037.

Subraveti, S. G., Z. Li, V. Prasad, and A. Rajendran. 2022b. "Physics-Based Neural Networks for Simulation and Synthesis of Cyclic Adsorption Processes." *Industrial & Engineering Chemistry Research* 61, no. 11: 4095–4113. https://doi.org/10.1021/acs.iecr.1c04731.

Subraveti, S. G., Z. Li, V. Prasad, and A. Rajendran. 2023. "Can a Computer 'Learn' Nonlinear Chromatography?: Experimental Validation of Physics-Based Deep Neural Networks for the Simulation of Chromatographic Processes." *Industrial & Engineering Chemistry Research* 62, no. 14: 5929–5944. https://doi.org/10.1021/acs.iecr.2c04355.

Tang, S.-Y., Y.-H. Yuan, Y.-C. Chen, S.-J. Yao, Y. Wang, and D.-Q. Lin. 2023. "Physics-Informed Neural Networks to Solve Lumped Kinetic Model for Chromatography Process." *Journal of Chromatography A* 1708: 464346. https://doi.org/10.1016/j.chroma.2023.464346.

Thuerey, N., P. Holl, M. Mueller, P. Schnell, F. Trost, and K. Um, 2021. *Physics-Based Deep Learning*. https://doi.org/10.48550/arXiv.2109.05237.

Tiwari, A., V. S. Masampally, A. Agarwal, and A. S. Rathore. 2023. "Digital Twin of a Continuous Chromatography Process for mAb Purification: Design and Model-Based Control." *Biotechnology and Bioengineering* 120, no. 3: 748–766. https://doi.org/10.1002/bit.28307.

Um, K., R. Brand, Y. R. Fei, P. Holl, and N. Thuerey. 2020. "Solver-in-the-Loop: Learning From Differentiable Physics to Interact With Iterative Pde-Solvers." *Advances in Neural Information Processing Systems* 33: 6111–6122. https://doi.org/10.48550/arXiv.2007.00016.

Velioğlu, S., H. E. Karahan, and Ş. B. Tantekin-Ersolmaz. 2024. "Predictive Transport Modelling in Polymeric Gas Separation Membranes: From Additive Contributions to Machine Learning." *Separation and Purification Technology* 340: 126743. https://doi.org/10.1016/j.seppur.2024.126743.

Wiewel, S., B. Kim, V. C. Azevedo, B. Solenthaler, and N. Thuerey. 2020. "Latent Space Subdivision: Stable and Controllable Time Predictions for Fluid Flow." *Computer Graphics Forum* 39, no. 8: 15–25. https://doi.org/10.1111/cgf.14097.

Wittkopp, F., J. Welsh, R. Todd, et al. 2024. "Current State of Implementation of In Silico Tools in the Biopharmaceutical Industry—Proceedings of the 5th Modeling Workshop." *Biotechnology and Bioengineering* 121, no. 9: 2952–2973. https://doi.org/10.1002/bit.28768.

Wu, C. H., B. Chan, Z. Sarich, et al. 2024. "Accelerating Attribute-Focused Process and Product Development Through the Development and Deployment of Autonomous Process Analytical Technology Platform System." *Biotechnology and Bioengineering* 121, no. 4: 1256–1269. https://doi.org/10.1002/bit.28649.

Yang, Y.-X., Y.-C. Chen, S.-J. Yao, and D.-Q. Lin. 2024. "Parameter-by-Parameter Estimation Method for Adsorption Isotherm in Hydrophobic Interaction Chromatography." *Journal of Chromatography A* 1716: 464638. https://doi.org/10.1016/j.chroma.2024.464638.

Yang, Y.-X., Z.-Y. Lin, Y.-C. Chen, S.-J. Yao, and D.-Q. Lin. 2024. "Modeling Multi-Component Separation in Hydrophobic Interaction Chromatography With Improved Parameter-by-Parameter Estimation Method." *Journal of Chromatography A* 1730: 465121. https://doi.org/10.1016/j.chroma.2024.465121.

Zou, T., T. Yajima, and Y. Kawajiri. 2024. "A Parameter Estimation Method for Chromatographic Separation Process Based on Physics-Informed Neural Network." *Journal of Chromatography A* 1730: 465077. https://doi.org/10.1016/j.chroma.2024.465077.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.