NIIMBL Protein A Model User Guide

Contents

1 Introduction

The protein A model is intended to be used for simulation of protein A load-wash-elute processes. The model has been implemented in Python using free and open source software including CADET and CADET-Process. The user interface for the protein A model is a Jupyter Notebook which can be opened and edited using Jupyter Notebook or Jupyter Lab.

Note: Click on the hyperlinked names throughout the text for more information on each topic.

This user guide contains information and instructions for the protein A model. Additional instructions are provided within the Jupyter Notebook.

1.1 Modeling overview

The workflow for using the protein A model is diagrammed in Figure 1.

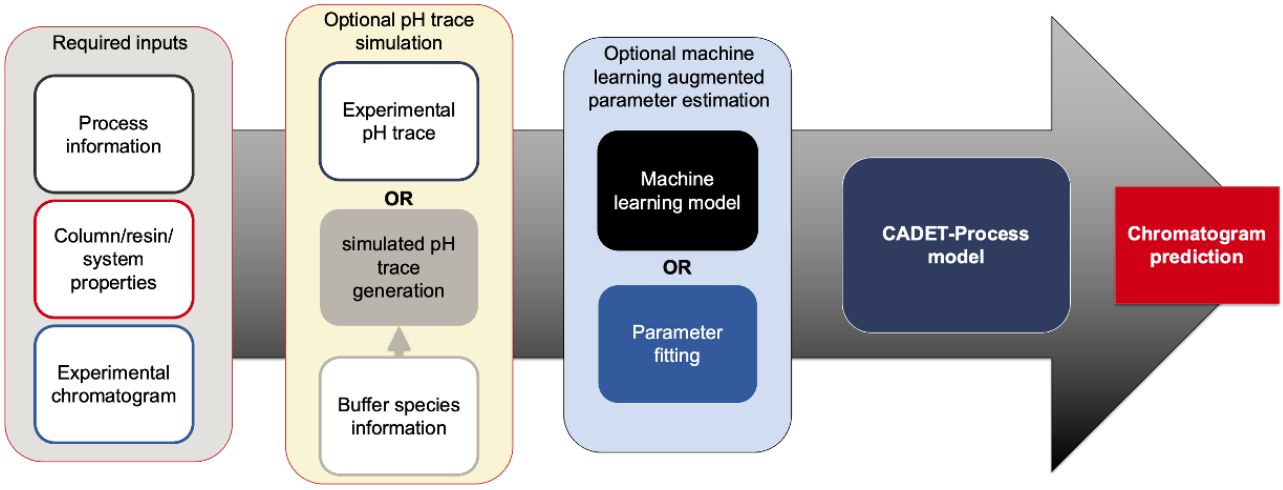

Figure 1: Protein A model workflow

All required inputs must be entered in the Jupyter Notebook. Experimental chromatograms can be supplied as .xlsx files to be read into the model. The user can either supply an experimental pH trace or generate one using the model. Parameters can be estimated by fitting the data directly in CADET-Process or by first passing the experimental chromatogram to a machine learning model. Once parameters are determined, the required inputs, pH trace, and parameters are passed to the CADET-Process model to generate the simulated chromatogram. Each step will be discussed in more detail in the later sections.

> The machine learning model for parameter estimation is not yet operational. See section 6 for the proposed hybrid modeling workflow.

The model has been set up to provide the user flexibility to easily switch between running forward simulations or fitting by changing a single variable setting. Additionally, options have been provided to allow the user to run either a load and wash only or a load-wash-elute process. The user can also set whether to use the pH dependent or non-pH-dependent parameter set with a single option. Within each of these, any subset of variables can be chosen for fitting.

The model can handle multiple experiments simultaneously, as is often desired for fitting. The format for entering data for multiple experiments and multiple protein components is described in the Jupyter Notebook. All chromatogram data must be provided as an excel workbook with data in one worksheet corresponding to one experiment.

The user should create a new model by copying and renaming the folder containing the user_inputs.ipynb Jupyter Notebook. All inputs and outputs for that model should be stored in the copied folder. Each time the model is run, a unique results folder will be generated containing the output excel files and plots.

The folder called 'EXAMPLE_PROA_SIM' contains an example Jupyter Notebook and results folders from several types of runs. The 'figures' and 'modules' folders should not be touched by the user. The 'modules' folder contains the python code comprising the model and the 'figures' folder contains reference images for the Jupyter Notebook.

2 Theory

There are many possible approaches to modeling protein A chromatography. Here, we use the General Rate Model for transport in the column and the multicomponent colloidal binding model for adsorption. We also include unit operations to account for variable extra-column hold up volume due to system tubing.

One of the main challenges of modeling protein A elution has proven to be the need for an accurate representation of the pH transition that occurs between the wash and elution steps. This pH transition is complicated by buffering reactions and interactions of charged species with components of the solid phase including protein A ligands and adsorbed proteins. One approach to supplying the correct pH (or more accurately proton concentration [H+]) trace to the model is to input experimental [H+] data directly. However, any inaccuracy in calibration, delay in response time, or mixing effects due to the nature of the pH probe itself would lead to inaccurate model predictions.

As an alternative to using an experimental pH trace, this model includes the option to generate a simulated pH trace for some buffer combinations. The simulation attempts to account for the effects of buffering reactions and protein A ligand titration on the elution [H+] profile. Simple step or linear gradient [H+] profiles can also be generated, but given the considerations above, these may not adequately represent physical conditions.

A more detailed account of the theory behind each part of the model is given below.

2.1 pH transition simulation

2.1.1 Buffering reactions

Buffering reactions are simulated by first applying a CSTR mixing function to a step input that goes from 0 to 100% B. The volume of the CSTR used here is set by the user. Next, the output of the CSTR function as a fraction of component B is applied to each the user defined buffer components. For example, if the wash buffer contains 0 mM acetate and the elution buffer contains 50 mM acetate, at 50% B the acetate concentration will be 25 mM. Then, with the total concentrations of each species known at each time point, the equilibrium equations with known dissociation constants for each buffer can be solved to determine the amount of dissociation for each buffer and the concentrations for each ionic species. Finally, the proton concentration can be determined by applying the electroneutrality equation. This is an iterative process because the dependence of the pKa is dependent on the ionic strength.

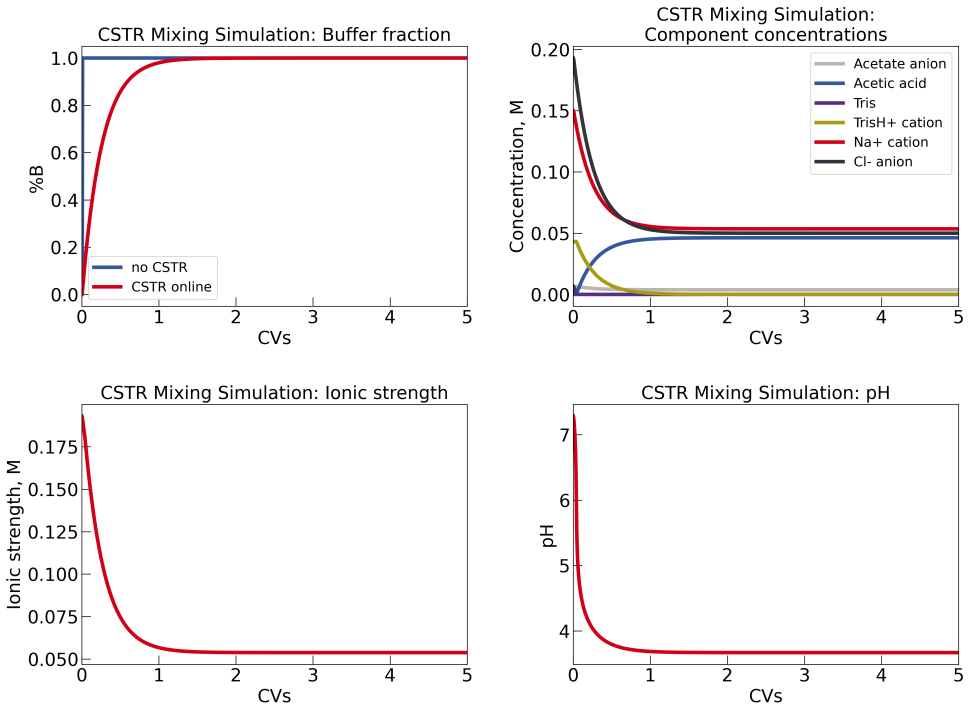

Figure 2: Example output plots from the CSTR buffering reactions model

2.1.2 Resin titration

The resin titration simulation takes the species concentrations at each time point from the output of the buffer reaction simulation and returns the species concentrations at the column outlet. The model accounts for the interaction of sodium ions with the solid phase as the buffer components flow through the column. The amount of sodium ions in solution is reduced as some ions adsorb on the column, and the species concentrations are recalculated based on the new sodium ion concentration, resulting in an adjusted [H+]. The resulting [H+] profile is shifted back by the column hold up volume and used as the inlet concentration profile of the CADET-Process model.

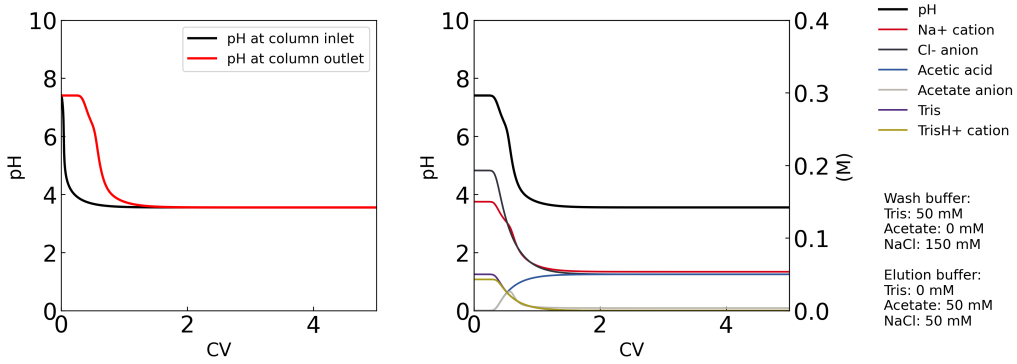

Figure 3: Example resin titration output plots

This approach has several limitations. First, supplying the [H+] inlet profile to the column model neglects the fact that resin titration is occurring at the microscale within the resin pores, and not only resulting in changes to the bulk [H+]. However, accounting for these effects would require more than a 1D column model. Second, interactions of the bound protein with the buffer species are not accounted for. Bound and unbound proteins are also likely to contain titratable groups that could affect the [H+] in solution. However, accounting for this would require knowing the bound and solution phase protein concentrations at the time of solving for pH, meaning the general rate model and binding model would have to be solved for both the interacting species simultaneously with the buffer reaction equilibria equations and electroneutrality, which is outside the scope of this work.

The resin titration simulation section of the model is currently only functional for the tris and acetate buffer system but will be expanded to include other buffer systems.

The resin titration simulation is currently very slow and can take about 15 minutes to run. Improving the efficiency of this calculation is one of the goals for future iterations of the model.

There is currently an offset visible in the plot between the simulated pH input and CADET-Process pH trace. This is because the activity coefficient was accounted for in converting the simulated [H+] but not the CADET-Process output. This is a known issue.

2.2 CADET-Process model

2.2.1 Flowsheet

The model flow sheet consists of five total unit operations (including the inlet and outlet), but at any given time only four are in line. The model flowsheet is shown in Figure 4 .

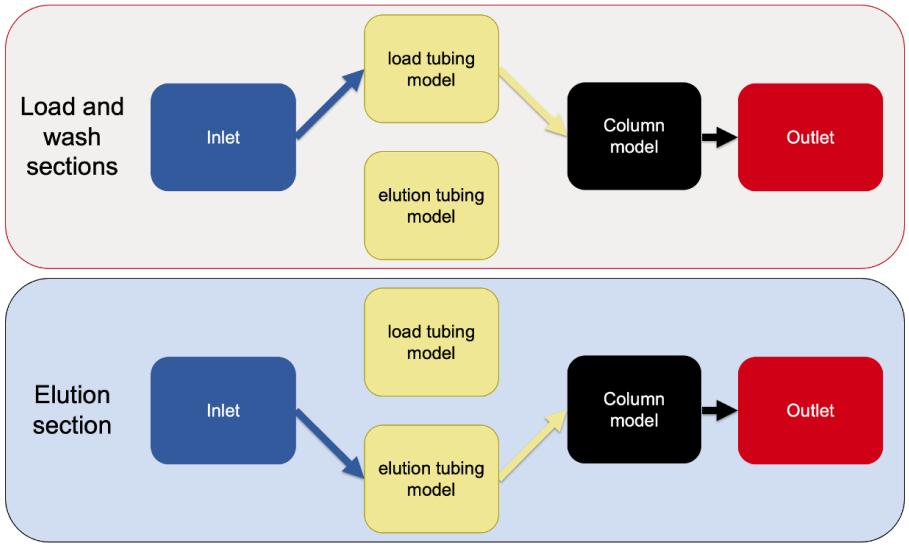

Figure 4: Flow diagram for CADET-Process model

The two tubing units are used to account for the fact that the flow path may be different between the load/wash phase of the process and the elution. Whether or not these units should be utilized depends on how the [H+] profile is supplied. (See section 5)

As the [H+] profile is supplied to the model either as an experimental trace or a simulated trace that already includes mixing effects, no mixer is included as part of the CADET-Process flow sheet.

2.2.2 Column model

The General Rate Model (GRM) is used to model transport in the column and the multicomponent colloidal isotherm is used to model adsorption.

The GRM is one of the most complex models available for column transport because it accounts explicitly for pore diffusion and surface diffusion (also referred to as "homogeneous diffusion"). While inclusion of these parameters separately makes model parameter estimation more difficult, it has been shown that surface diffusion contributes significantly to mass transfer in the intermediate to low pH range of protein A chromatography, which means it likely affects the shape of the elution peak. Therefore, we include a [H+] dependent surface diffusivity parameter in this protein A model. See [1] for more information on [H+] dependent surface diffusivity.

2.2.3 Tubing model

The tubing is modeled as a dispersive plug flow reactor using the Lumped Rate Model Without Pores in CADET with the porosity set to 1 and no binding. Axial dispersion is estimated using the tubing diameter, flow rate, and molecular diffusion as originally described by Taylor [2], so no fitting of parameters is necessary.

It is important to include accurate system tubing lengths and diameters because these contribute to the hold up volume of the system, which shifts the simulated chromatogram to the left or right. If not included, there may be an offset from experimental data.

3 Installation

3.1 Choosing a conda distribution

The protein A model is implemented in Python and requires several packages to be installed. It is recommended to set up a conda environment to install the required packages and run the model. Conda is a package and environment manager that can be installed in a few different forms.

- Anaconda
- Miniconda
- Miniforge

If you already have Anaconda, Miniconda, or Miniforge installed, you can proceed to the next section.

If you do not already have one of these, you will need to choose one to install. There are various reasons why you may choose one or another, but in the context of this project, Miniforge may be the best option because it is free and open-source for all users, whereas Anaconda and Miniconda require licenses for businesses with > 200 employees.

Note: You do not want to have more than one of these installed, so if you want to install Miniforge or Miniconda, you would need to uninstall Anaconda first.

Follow the installation instructions for Miniforge found here: https://kirenz.github.io/codelabs/codelabs/miniforge-setup/#0.

When you get to step 5, rather than setting up an example environment, follow the steps in the next section to set up the environment for the protein A model.

3.2 Create the environment

Let's create a new environment called cadet-env.

```
conda create -n cadet-env
```

when conda asks you to proceed `(proceed ([y]/n?)`, type `y`.

Now you have created a new environment, but nothing is installed in it yet and you are still working in the base environment. To move to the newly created cadet-env, run

```
conda activate cadet-env
```

Now you are in the cadet-env, and you should see (cadet-env) at the start or your command prompt.

3.3 Install packages

Now we will install all of the necessary packages for the protein A model in the new environment.

Install CADET by running

```
conda install -c conda-forge cadet
```

Note: CADET v5.0.0 is required. Using older versions (v4.4.0 or v4.3.0) will result in errors related to the colloidal isotherm parameters.

To install CADET-Process, we need pip. Install pip by running

```
conda install pip
```

Install the specific branch of cadet-process that we need by running

```
pip install git+https://github.com/angelamoser1/CADET-Process.git@add_GRM_surface_diffusion_dependence
```

Install Jupyter by running

```
conda install jupyter
```

Note: If you are using Anaconda, Jupyter is already included and you do not need to install it.

Note: You do not necessarily need to install jupyter in each of your environments, you can install it once and then set up kernels, but here we will just install it in the environment because this is the simplest option to set up (See this post).

4 Running the model

From within cadet-env, open a jupyter notebook by running

```
jupyter notebook
```

This will open Jupyter in your browser and you should see a file tree. Navigate to where you have the protein A model saved and open the file ending in *user_inputs.ipynb*. Here is where you will enter all information and run the model.

Plots will be shown in-line but will also be saved along with excel spreadsheets containing output data in a results folder located within the protein A model folder each time the model is run.

Follow the instructions in *user_inputs.ipynb* to set up and run the model.

After installation is complete, each time you want to open a Jupyter Notebook and run the model you will need to run

`conda activate cadet-env` and then `jupyter notebook`

If you are using miniforge, you may first need to run

```
C:\Users\username\miniforge3\Scripts\activate
```

then run `conda activate cadet-env` and `jupyter notebook`.

4.1 Additional packages

If you try to run the model and then get a module not found error, you can simply install the missing module by running "conda install <packagename>".

For example, if you see something like:

```
ModuleNotFoundError: No module named 'openpyxl'
```

simply run

```
conda install openpyxl
```

within your environment and the problem should be solved.

5 [H+] profile

Four options exist for supplying the [H+] profile to the protein A model. These are summarized in table 1.

user input	description
profile = 'step'	Step elution (no buffering effects)
profile = 'linear'	Linear gradient (no buffering effects)
profile = 'simulated'	Simulated (included buffer mixing and resin titration)
profile = 'user_defined'	User defined (an experimental pH trace converted to [H+], for example)

Table 1: [H+] profile options

The step and linear profiles are created directly by CADET-Process and do not account for any mixing or resin titration effects. These options will likely not reflect realistic elution conditions, but are included as a means to conveniently generate pH traces without requiring inputs for pH trace simulation or a user supplied [H+] profile. This is handy when checking if the simulation is working correctly. The length of the linear gradient is defined by the length of the elution_CV variable, and the starting and ending pHs are defined by the wash_pH and elution_pH variables.

To generate a simulated pH trace, follow the instructions in the pH Trace Simulation section of the Jupyter Notebook. Note that exact sodium and chloride concentrations must be known for the wash and elution buffers. The pH trace simulation only generates the simulated trace for the elution section, and CADET-Process fills in the rest with a spline. If the simulated pH and the wash_pH and elution_pH variable values do not match up there may be sharp jumps in the profile.

A user defined elution pH trace can be input directly from an excel spreadsheet specified by the user. In this spreadsheet, the first column should be volume in mL and the second should be the [H+] concentration in mM. **The inlet profile can only be supplied for the elution step, regions outside of the elution step will be trimmed off before the trace is supplied to CADET-Process.** The supplied trace does not need to be shifted back by the hold up volume because the hold up volume of the column and tubing will be accounted for by the model. Similarly to with the simulated pH trace, CADET-Process will fill in the blanks with a spline, and if the simulated pH and the wash_pH and elution_pH variable values do not match up there may be sharp jumps in the profile.

As with input chromatogram data, it is possible to input elution [H+] profiles for multiple experiments to be simulated at once by including the profile for each experiment in a different sheet in the Excel workbook.

6 Hybrid modeling workflow

One of the main drawbacks of the protein A model is the number of parameters that must be fit for model calibration. This is due to the need for pH dependency in the model. While the calibration method described in section 7 of fitting breakthrough curves at 5 pHs results in an accurate and reliably calibrated model, the experimental burden is high, and fitting can be time consuming. The goal of the proposed hybrid model is to increase the speed of model calibration while potentially also reducing the number of experiments required.

The proposed model attempts to take advantage of the known relationships between the pH dependent and non-pH dependent adsorption and transport parameters. A muti-step workflow is proposed such that the first step requires only one breakthrough curve in the typical loading pH range. Then, information learned in this step will be used to aid in the second step. The proposed two step hybrid modeling workflow is shown in figure 5.

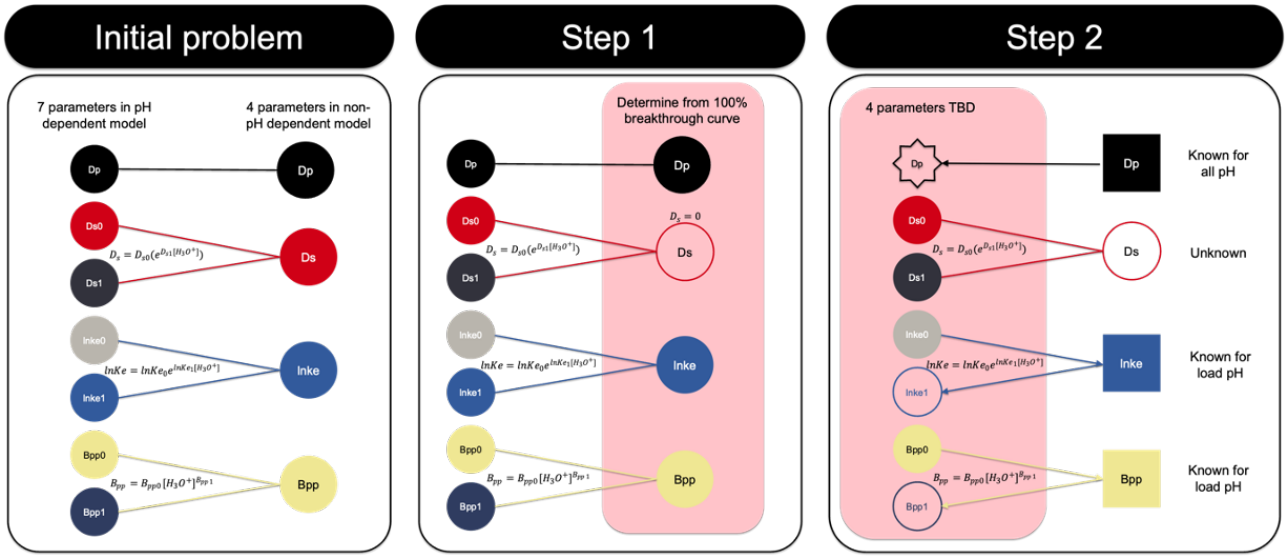

Figure 5: Proposed two step hybrid modeling workflow

In the first step, three parameters are determined based on one breakthrough curve at loading pH. These parameters could also be determined by fitting of the mechanistic model, but using a pre-trained neural network to estimate these parameters could yield parameter values on the order of seconds rather than hours. The neural network could be trained on synthetic data generated earlier by the mechanistic model.

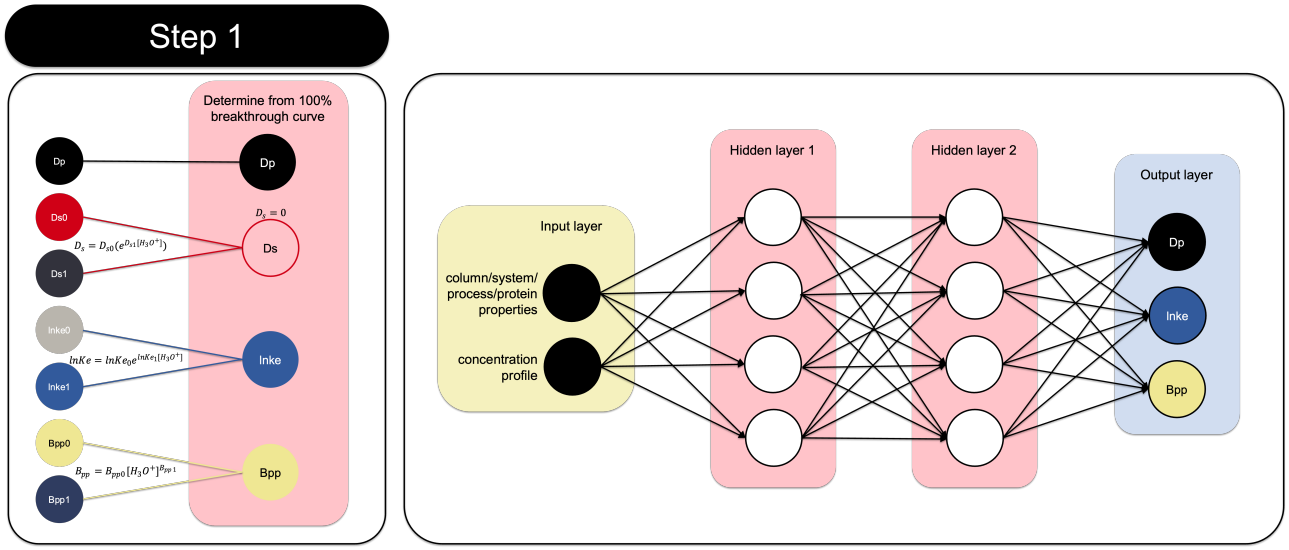

Figure 6: Example neural network for first step of hybrid modeling

In the second step, we attempt to reduce the difficulty of fitting the pH dependent parameter set by taking advantage of what we have learned from the non-pH-dependent parameters. The known parameter values and

known relationships combined with pH dependent chromatogram data (what data specifically is tbd) can be fed to another pre-trained neural network to quickly generate the set of pH dependent parameters.

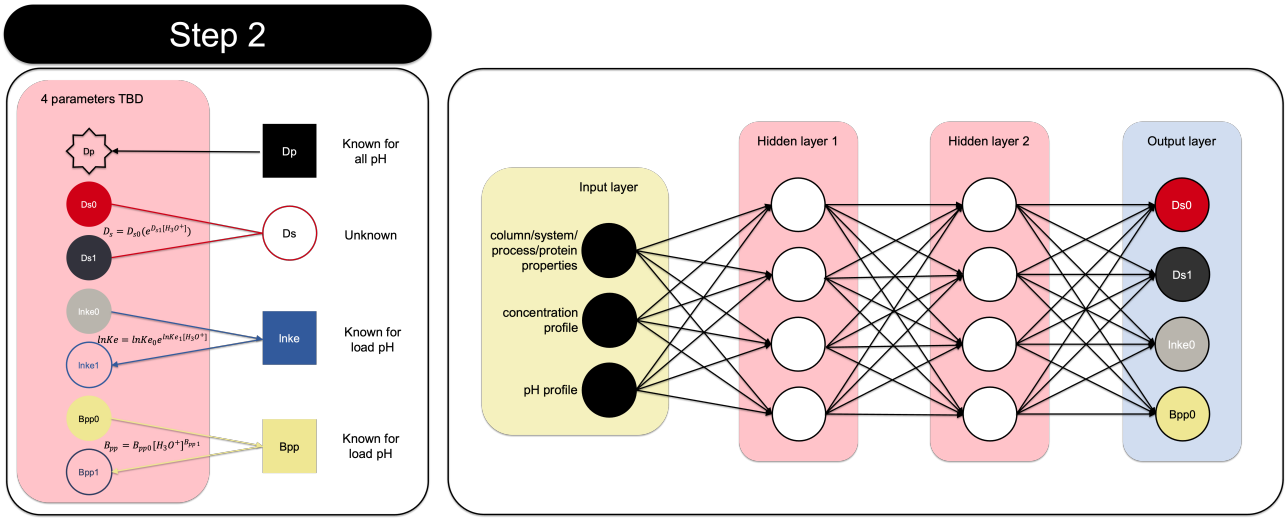

Figure 7: Example neural network for second step of hybrid modeling

If the two models are well trained, these could be the final parameters used in the model. Otherwise, the parameters generated by the machine learning model could serve as initial guesses to be refined by fitting with the mechanistic model.

7 Experimental protocol for parameter estimation using the mechanistic model

7.1 Overview

Chromatogram data from multiple full breakthrough experiments is required for accurate parameter estimation by fitting. For protein A, since the relevant mobile phase modifier is pH, these full breakthrough experiments should cover a range of pHs while keeping ionic strength and protein concentration constant. This allows the pH dependence to be determined from the experimental data. It is recommended that at least 5 full breakthrough experiments are performed. The pH values used should span the range of pHs used in a typical protein A bind and elute experiment, with more data required in the pH range where binding behavior changes rapidly with pH, corresponding to the "cliff" of the isotherm, roughly pH 3.8 – 5.0. The resulting chromatograms will be simultaneously fitted using the Protein A Bind and Elute Model to determine model parameter values. This is similar to the method described in [3].

7.2 Buffers and Sample Preparation

To maintain constant ionic strength during the load and wash sections for all the experiments, the equilibration buffers can be created by titrating a high pH buffer with a low pH buffer of similar ionic strength. A buffer mixture that has sufficient buffering capacity across the required range of pHs should be used. The recommended buffers, as described in [4], are:

1. 100 mM Na2CO3, 50 mM Tris, 20 mM bis-tris, pH 11.5
2. 100 mM acetic acid + 150 mM NaCl, pH 2.5

Recommended target pHs are: 4.0, 4.3, 5.0, 5.8, 7.0.

Note: Although data below pH 4.0 may be useful for fitting, getting good breakthrough curve data below pH 4.0 has proven difficult due to the low capacity of the resin in this range.

The protein of interest must be buffer exchanged into the appropriate buffer for each experiment. The buffer exchanged protein solutions can then be diluted with the appropriate buffer to reach the desired final concentration for loading. All feeds should be at the same concentration. A concentration in the same range as the feed material of interest should be used, as model parameters can vary somewhat with concentration. For example, if clarified

cell culture harvest typically has a concentration around 5 g/L, the buffer exchanged protein solutions for the full breakthrough experiments should be diluted to around 5 g/L.

7.3 Protein A bind and elute method

The column should be loaded at least to 100% breakthrough in each experiment, and the amount of protein required to reach 100% breakthrough will vary depending on the pH. The loading required to reach 100% breakthrough is much higher than a typical protein A dynamic binding capacity at 10% breakthrough. The table below provides a rough guide for how much protein may be required to reach 100% breakthrough at a given pH, but actual values will depend on the specific resin and protein being used.

pH	Loading [g protein / L column]	CVs at 5 g/L feed conc
4.0	35	15
4.3	60	25
5.0	90	40
5.8	90	40
7.0	90	40

Table 2: Estimated loading at recommended pH values.

Because of the high loading, a long wash step should be used. The shape of the tail of the breakthrough curve contains information used for parameter fitting, so it is important that the wash is not cut off too early. A wash step length of 20 column volumes is recommended. The elution and strip sections are not used for parameter fitting, so the pH, buffer compositions, and lengths of these sections are left to the preference of the user.

7.4 UV calibration curve

The breakthrough curve data must be entered into the model in terms of concentration in [g/L]. It is recommended that UV trace data be converted into concentration data using a calibration curve. An extinction coefficient may be sufficient if the entire curve falls within the linear range of the system's UV detector. To create the calibration curve, a series of protein solutions at concentrations ranging between 0.0 g/L and the maximum concentration of the breakthrough curves should be injected in large enough volumes that a clear plateau in the UV signal is distinguishable. The UV signal plateau values and concentrations of each solution can then be used to create the calibration curve. The concentration vs. UV signal can be fitted to a polynomial equation which can be used to convert the UV traces to concentrations.

Collecting UV data at 295 or 300 nm rather than (or in addition to) the typical 280 nm is recommended because the linear range will extend to a higher concentration.

While concentrations measured from fractions can be used for fitting if fractions are collected frequently, with lower resolution the fit may be less accurate.

8 Resin Properties

This section contains suggested input properties for a set of protein A resins.

Ep, Ee, particle diameter, and pore radius values are copied directly from the 2018 paper titled "Evaluation of recent Protein A stationary phase innovations for capture of biotherapeutics" by Pabst et al. [5]. They are included here for convenience.

Multiple values for particle porosity corresponding to different measurement methods were reported in [5]. The Ep listed here corresponds to the one determined by fitting dextran pulses. The Ee and Ep values listed here along with the porosity for a mAb and Kd values reported in [5] were used to calculate the phase ratio for the colloidal isotherm as described below.

First, the Kd values for each resin were squared and divided by the Kd for the smallest tracer, yielding a maximum Kd equal to 1, implying full penetration of the pores. Then the log normal distribution was fit to the dextran probe radii and corresponding Kd values for each resin, where Kd is defined as the ratio of accessible and total volume for each tracer, as given by equation (7) in [6], also shown below.

$$Kd = \frac{\int_{r_m}^{\infty} f(r)[1-(r_m//r)^2]\,dr}{\int_0^{\infty} f(r)\,dr} \tag{1}$$

The accessible surface area for the mAb was then determined using equation (10) from [6]. The mAb hydrodynamic radius used for calculations was 4.49 nm. Equation (10) from [6] is shown below.

$$A(r_{mAb}) = \frac{\int_{r_{mAb}}^{\infty} \frac{2(r-r_{mAb})}{r^2} f(r)\,dr}{\int_0^{\infty} f(r)\,dr} \tag{2}$$

Et_{mAb} was calculated using Ee and Ep_{mAb3} in [5]. The surface area accessible to mAb from equation 2 is in units of m^2/mL pore volume accessible to dextran. It must be converted to m^2/m^3 solid phase volume accessible to mAb as shown.

$$\varphi = A(r_{mAb})\frac{(Et_{dex}-Ee)}{(1-Et_{mAb})}(1E6\,mL/m^3) \tag{3}$$

Resin	Ep	Ee	particle diameter [um]	pore radius [nm]	phase ratio $[m^2/m^3 SP]*1E8$
MabSelect SuRe	0.94	0.31	89.6	41.8	1.78
MabSelect SuRe LX	0.90	0.31	88.4	38.2	1.46
MabSelect SuRe pcc	0.87	0.35	44.4	29.7	1.27
MabSelect PrismA	0.92	0.34	54.1	32.7	1.07
Amsphere A3	0.85	0.34	49.6	46.0	0.998
AF-rProtein A HC-650F	0.76	0.34	54.2	30.8	0.764
KanCapA	0.91	0.36	75.0	51.0	1.15
KanCapA 3G	0.88	0.33	79.2	42.5	1.44
Eshmuno A	0.81	0.34	50.2	40.8	1.59
Praesto AP	0.95	0.25	92.5	37.4	1.57
Monofinity A	0.92	0.29	95.7	36.8	1.32
MabSpeed rP202	0.74	0.37	44.9	60.5	0.423

Table 3: Protein A Resin Properties

References

[1] Soumitra Bhoyar et al. "Predictive mechanistic modeling of loading and elution in protein A chromatography". In: *Journal of Chromatography A* (Dec. 2023), p. 464558. ISSN: 00219673. DOI: 10.1016/j.chroma.2023.464558. URL: https://linkinghub.elsevier.com/retrieve/pii/S0021967323007835 (visited on 12/05/2023).

[2] Geoffrey Taylor. "Dispersion of soluble matter in solvent flowing slowly through a tube". In: *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* 219.1137 (Aug. 25, 1953), pp. 186–203. ISSN: 0080-4630, 2053-9169. DOI: 10.1098/rspa.1953.0139. URL: https://royalsocietypublishing.org/doi/10.1098/rspa.1953.0139 (visited on 09/20/2024).

[3] Vijesh Kumar et al. "Robust mechanistic modeling of protein ion-exchange chromatography". In: *Journal of Chromatography A* 1660 (Dec. 20, 2021). Publisher: Elsevier B.V. ISSN: 18733778. DOI: 10.1016/j.chroma.2021.462669.

[4] Soumitra Bhoyar et al. "Engineering protein A ligands to mitigate antibody loss during high-pH washes in protein A chromatography". In: *Journal of Chromatography A* 1696 (May 2023), p. 463962. ISSN: 00219673. DOI: 10.1016/j.chroma.2023.463962. URL: https://linkinghub.elsevier.com/retrieve/pii/S0021967323001887 (visited on 05/17/2023).

[5] Timothy M. Pabst, Johnny Thai, and Alan K. Hunter. "Evaluation of recent Protein A stationary phase innovations for capture of biotherapeutics". In: *Journal of Chromatography A* 1554 (June 2018), pp. 45–60. ISSN: 00219673. DOI: 10.1016/j.chroma.2018.03.060. URL: https://linkinghub.elsevier.com/retrieve/pii/S002196731830387X (visited on 01/25/2023).

[6] Peter DePhillips and Abraham M. Lenhoff. "Pore size distributions of cation-exchange adsorbents determined by inverse size-exclusion chromatography". In: *Journal of Chromatography A* 883.1 (2000), pp. 39–54. ISSN: 00219673. DOI: 10.1016/S0021-9673(00)00420-9.