Advancing Chromatography Simulations: GUI Development & Neural Network-Based Parameter Prediction

Shobhith Vadlamudi – ED21B069

Nadeem Mohammed, Arch Creasy, Leo Kutlowski, Shilpa Ananthakrishnan, Shu Yang

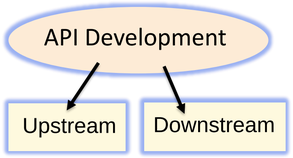
INTRODUCTION

In this internship I am working on two key projects aimed at enhancing downstream process modelling.

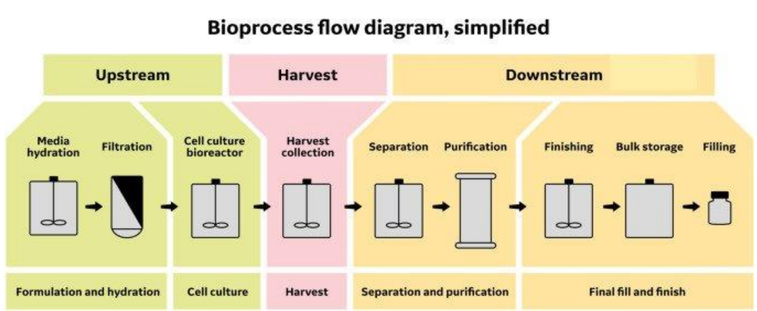
- Development of a user-friendly GUI (Graphical user interface) for chromatography forward simulations
- Neural Network-based parameter prediction for isotherm modelling

The main goals behind these projects is to make modelling more accessible to the laboratory scientists and to save time and increase accuracy while fitting the model parameters

PHARMACEUTICAL INDUSTRY BASICS



- APIs are active pharmaceutical ingredients, they are the biologically active components in drugs, these are derived from living organisms. Examples include mAbs (monoclonal antibodies) used in cancer therapy, and mRNA-based vaccines like the COVID-19 vaccine

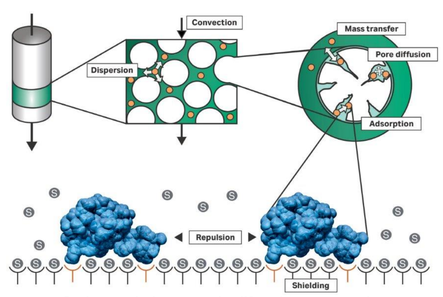


DOWNSTREAM PROCESSES – CHROMATOGRAPHY

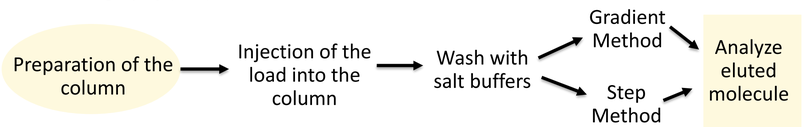
Chromatography is a purification technique used to separate biomolecules

Working – A sample mixture is passed through a stationary phase (resin) while a mobile phase (buffer) carries the components through. Molecules separate based on size, charge, hydrophobicity or affinity

Types – Affinity chromatography, Ion Exchange chromatography (IEX), Hydrophobic Interaction chromatography (HIC)



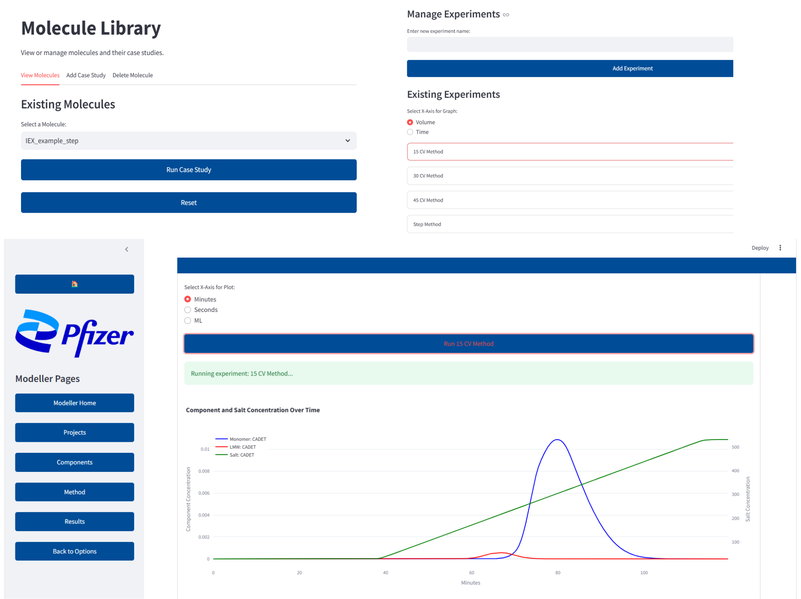
Cytiva web (www.cytivalifesciences.com)



GUI DEVELOPMENT

The developed Graphical User Interface provides an intuitive platform for experiment setup, execution, result visualization and report generation. Designed using Streamlit it allows the users to

- Load case studies designed for the molecule using GoSilico, simulated using CADET
- Modify model parameters if needed and set up experiments with different loading conditions or steps
- Compare the yield and purity for different fractions and export all the results to an excel sheet



HYBRID MODELS FOR CHROMATOGRAPHY MODELLING

- **Column Model** – Describes mechanisms in the mobile phase within an interstitial volume of a column.

$$\frac{\partial c_i}{\partial t}(x,t) = -\frac{u(t)}{\varepsilon_{tot}}\frac{\partial c_i}{\partial x}(x,t) + D_{app}\frac{\partial c_i{}^2}{\partial x^2}(x,t) - \frac{1-\varepsilon_{tot}}{\varepsilon_{tot}}\frac{\partial q}{\partial t}(x,t)$$

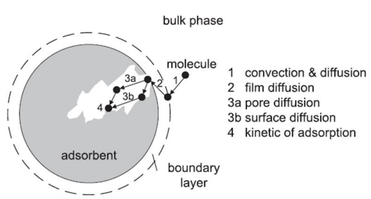
- Equilibrium dispersive model (includes effects of convection and dispersion)

- **Pore Model** – Describes effects in the mobile phase within pores of an adsorber

$$\frac{\partial c_p}{\partial t}(x,t) = \frac{3}{r_{\text{bead}}}\frac{k_{\text{eff}}}{\varepsilon_{\text{bead}}}(c(x,t) - c_p(x,t)) - \frac{1-\varepsilon_{\text{bead}}}{\varepsilon_{\text{bead}}}\frac{\partial q}{\partial t}(x,t)$$

Lumped pore model

- **Adsorption Model** – Describes binding and elution of molecules on a column. Also referred to as an isotherm model

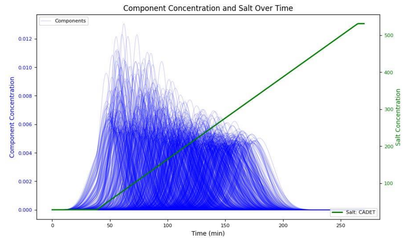


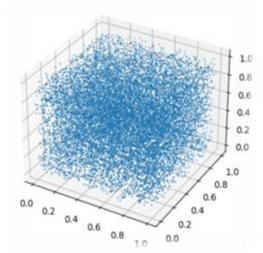
Steric Mass Action

$$k_{kin,i}\frac{\partial q_i}{\partial t} = k_{eq,i}\cdot \bar{q}_{salt}^{\nu_i}\cdot c_{p,i} - q_i\cdot c_{p,salt}^{\nu_i}$$

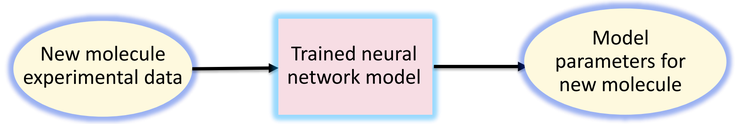
$$\text{with } \bar{q}_{salt} = \Lambda_{IEC} - \Sigma_{j=1}^{N}(\nu_j + \sigma_j)\cdot q_j$$

- **Data Generation –** Modelled the entire variable space for 3 parameters of IEX and HIC chromatography, this was done on GoSilico, a chromatography simulation software by Cytiva. The molecule I used was from the Cytiva case study example given.

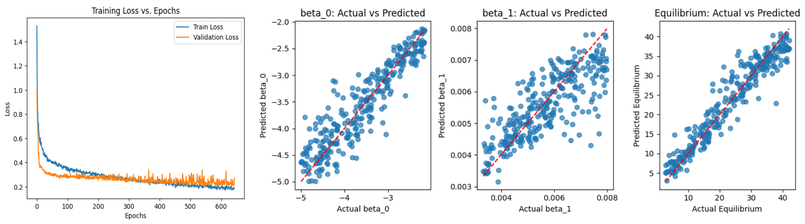




- **Model Creation –** Focused first on overfitting the model, then gradually simplified the network and added dropout and regularization to get the best test accuracy. Fully connected neural network architecture was used. Convolution layers were also introduced.



- **Results -** Achieved a training loss of 0.0125 and validation loss of 0.0620. Traditional numerical methods (e.g., GoSilico) take hours with no guarantee of convergence, while neural networks provide faster, reliable initial guesses for the mechanistic model within minutes



FUTURE PLANS

- Deploy the developed GUI using DataIku - **Dataiku** is a collaborative data science platform that enables building and deploying machine learning models and applications, including **custom GUI-based applications**.

- Adaptive sampling is a technique where data is collected and generated dynamically based on patterns in the data, rather than at fixed intervals or uniformly across the dataset. I plan to use this technique to speed up the training process.

- Improve the accuracy of the neural network model when extended to 5 parameters and use CADET to generate data for both IEX and HIC models and integrate the fitting process with the GUI. Validate the model with real experiment data