Mechanistic chromatography modeling — general model calibration workflow

A mechanistic chromatography modeling workflow typically requires a model calibration step which includes system and column characterization, protein experiments, and parameter estimation (Fig 1).

This document describes a general workflow to perform model calibration after you have performed all characterization experiments for the system, column and protein-ligand-interaction, and the suitable column, pore and isotherm models were selected. In particular, the parameter estimation process is outlined in this document.

Although it is impractical to provide a precise guide which is applicable to your project requirements, helpful general information to speed up the modeling process and to increase the modeling quality is provided.

The workflow described is a standardized workflow that may vary based on your type of modeling project. This document should be treated only as a high-level guideline to plan the steps of the calibration workflow of your modeling project.

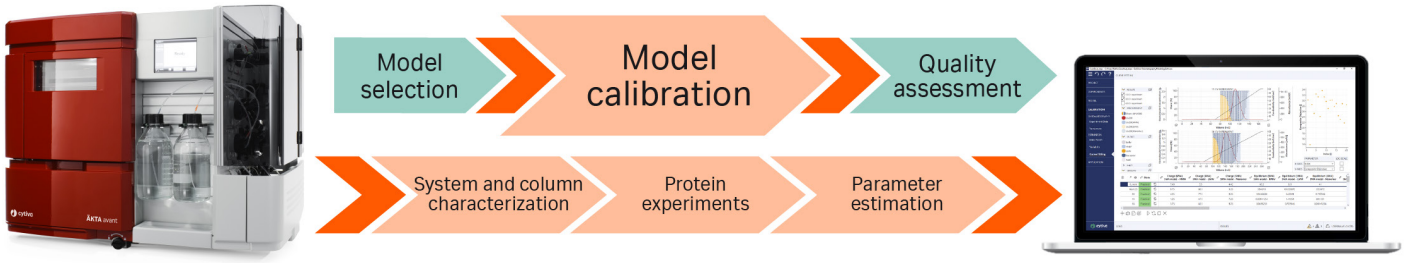

Fig 1. Outline of a mechanistic chromatography modeling workflow, with a focus on model calibration.

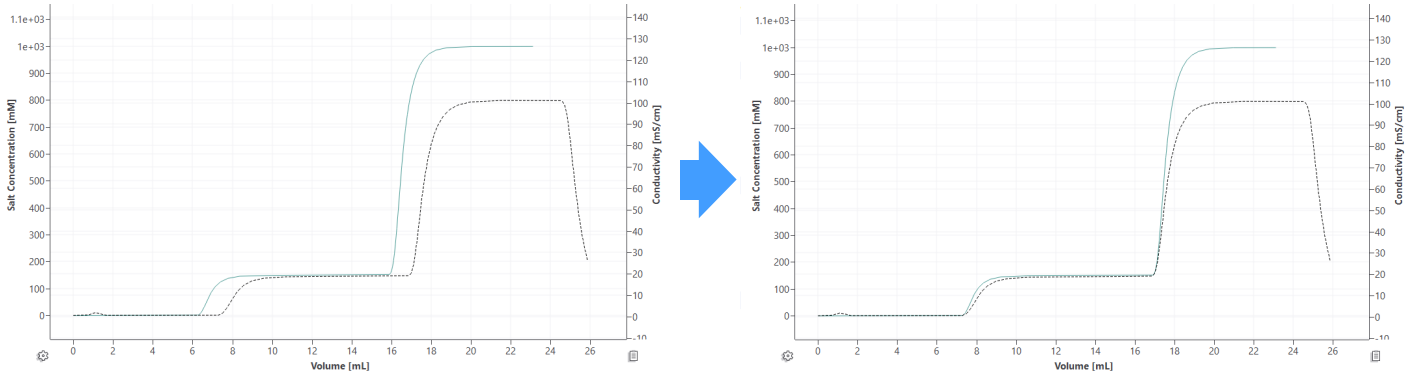

Fig 2. Exemplary chromatograms with simulated conductivity trace (green) and measured conductivity trace (black dotted line) before (left) and after (right) the conductivity check.

Check of conductivity trace

You can use the recorded conductivity to support the selection of a suitable column model. Even if the column model is already selected, we recommend importing the conductivity to GoSilico™ Chromatography Modeling Software and comparing it to the modeled salt trace before starting the modeling of the protein components. This ensures that all salt events are set up correctly (Fig 2).

Moreover, this will enable to detect offsets and backmixing effects for example, due to a missing pump wash. Adjust the simulated salt trace to as close as possible to the measured conductivity signal, to take accidently introduced effects into account.

Use of available data

All available data should be used to increase the number of data points and help to derive the unknown model parameters. Different UV signals can be imported as different “sensors” in GoSilico™ Chromatography Modeling Software. During parameter estimation all sensors will be used for inverse peak fitting.

For parameter estimation, at least one UV signal, for example at 280 nm, is required. As the absorbance at 280 nm is rather high for proteins, the detector limit at ~3000 mAU is often reached for runs with high column loading as required for model building. Consequently, the peak tops are cut off at the detection limit. To capture the peak shape in case of sensor saturation at 280nm, an additional wavelength with a lower absorbance (e.g., 300nm) should be included. This will also double the number of data points available for a more exact parameter estimation.

In addition, offline fraction analysis is equally important for model calibration. In most cases the elution results in a single peak containing both the product and different impurities. To locate the impurities within the elution peak and to define their peak shape, it is essential to include analytical information. The fraction information can be either included as percent of measurement (using size exclusion or cation exchange chromatography analytics) or absolute values for process-related impurities such as host-cell protein (HCP) or DNA.

If the fraction information is included as percent of measurement, the wavelength used in the analytics should be the same as for the recorded UV signal. In some cases, it makes sense to include the fraction information also for different wavelengths, but it should be kept in mind that the offline analytics are may not be 100% valid for additional wavelengths.

General workflow

Note: Before starting the actual model calibration for capture chromatography where typically a large flowthrough of, for example, HCP occurs, it is important to consider the non-binding species. The easiest way to do so is to introduce a non-binding component (e.g., HCP) to cover the flowthrough peak area. This will prevent the solver from considering this part of the UV signal when fitting the product and product related impurities.

In general, the modeling should be done following a "bottom-up" approach. First, the parameters of the main component – often the mAb monomer – are estimated. It is important to start with the highly concentrated protein species, as the main component has the most influence on the less concentrated impurities. Therefore, the impurities should be only modeled after the main component is fixed. In some cases, charge variants also need to be considered. As the interactions between the charge variants are rather strong in the most cases, all charge variants should be estimated within the same estimation run using offline analytics data.

After the main component is set up, the HMW and LMW species can be estimated consecutively. Again, importance must be given to the offline analytics rather than only the UV signal.

Once the product and the product related impurities are modeled, the process related impurities can be added to the model. The concentration units of these components are usually different from the units of the other species. Therefore, they need to be set up in separate sensors. The fraction data is in absolute values instead of percent of UV measurements.

The general workflow to set up a model including various components is shown in Figure 3.

Parameter estimation

The actual procedure of inverse peak fitting is of course strongly dependent on your protein component. Therefore, it is not possible to provide a precise step-by-step guide on parameter estimation. In the following, we describe ways to speed up the parameter estimation and increase the model quality.

Following the "bottom-up" approach described above, there are still some important steps that might be relevant in almost every modeling project.

Before starting the parameter estimation for a protein component, you should consider the parameters that might be irrelevant for your individual component. As an example, the shielding parameter for trace impurities may not play an important role. Therefore, the shielding of the impurities should be avoided unless during special requirements.

In general, the parameter estimation should be started with a heuristic algorithm (e.g., ASA). These algorithms contain a stochastic part that allows random jumps within the search space. Consequently, different areas in the search space are scanned rather quickly. The risk to end up in a local minimum and never go on to the global optimum is not very high in contrast to deterministic algorithms. Thus, heuristic algorithms should be used until alternative solutions are found as a starting point for the deterministic algorithms.

To increase the rate of convergence, the estimation should be carried on with a deterministic algorithm (e.g., Levenberg-Marquardt).

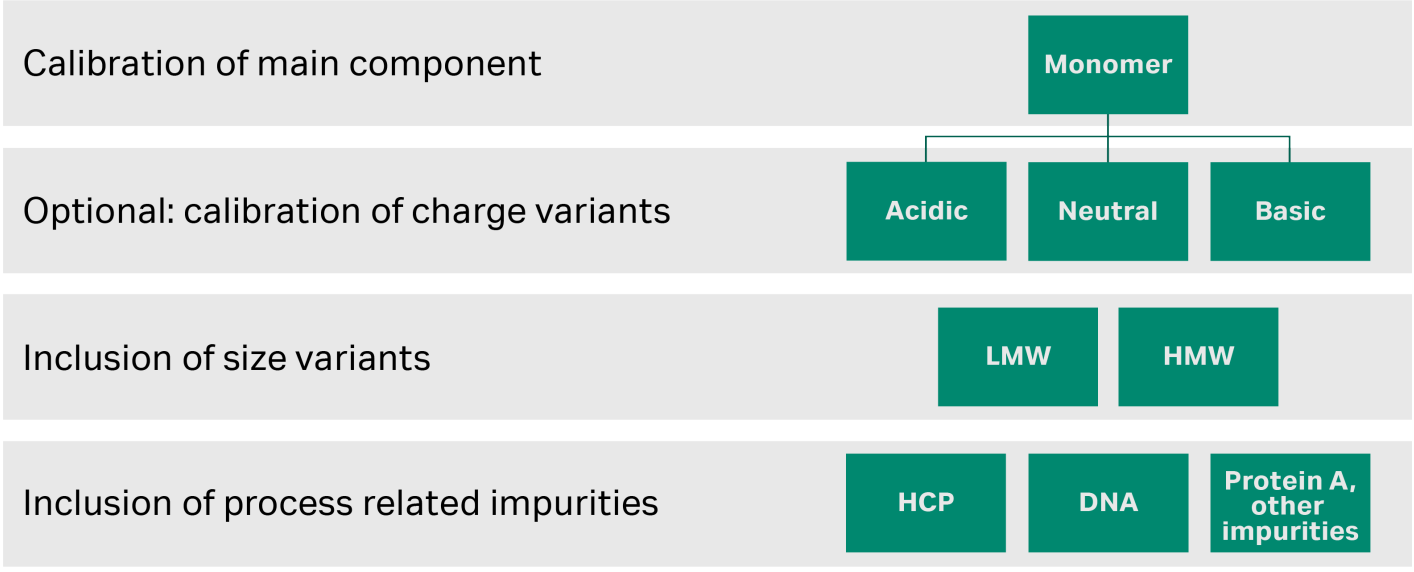

Fig 3. Exemplary overview of biomolecule species and impurities which are introduced in a modeling project.

Please note that the percentage of the parameter change (=jacobian calculation precision) can be adapted in the settings depending on the fitting problem. If the fit derived with the heuristic algorithm is already good, a small value (e.g., 0.001) should be used to change the parameters slowly. If the starting point is not that good, the parameter changes can be increased by using a larger value for the jacobian calculation precision.

Another important point is to also use other error norms if the least square error (L2Error) does not lead to a satisfactory parameter estimation. GoSilico™ Chromatography Modeling Software offers wide variety of different error norms.

If multiple chromatograms and/or components with different peak areas are used for model building, the normalized root-mean-square error (NRMSE) is more suited. Here, the individual residuals of the different components and processes are normalized prior to the calculation of the overall residual. Therefore, components with smaller peak areas are not neglected during model building.

The L2Error Integral, as an integrated form of the least square error, is helpful for finding the correct peak position as the integration of the residual provides a large error if there is no overlap between the measured and modeled chromatograms.

Dynamic time warping (DTW) is an error algorithm that compresses and stretches the signal along the column to minimize the error between the measured and modeled chromatograms. Therefore, a peak shift along the column is less important for the resulting residual. DTW is helpful if the correct peak shape is preferred over the peak position.

It may be sometimes necessary to perform the heuristic parameter estimation for multiple proteins at once because of their protein-protein-interactions (e.g., charge variants). Otherwise, we can do the inverse peak fitting for one component at a time. When switching to the deterministic algorithms, it could be helpful to alternate between refining all components at once and to find out witch option works better for your project. After model calibration, the model quality needs to be investigated. The essential indicator for model quality is the visual fit. A high-quality model describes all calibration runs accurately (Fig 4).

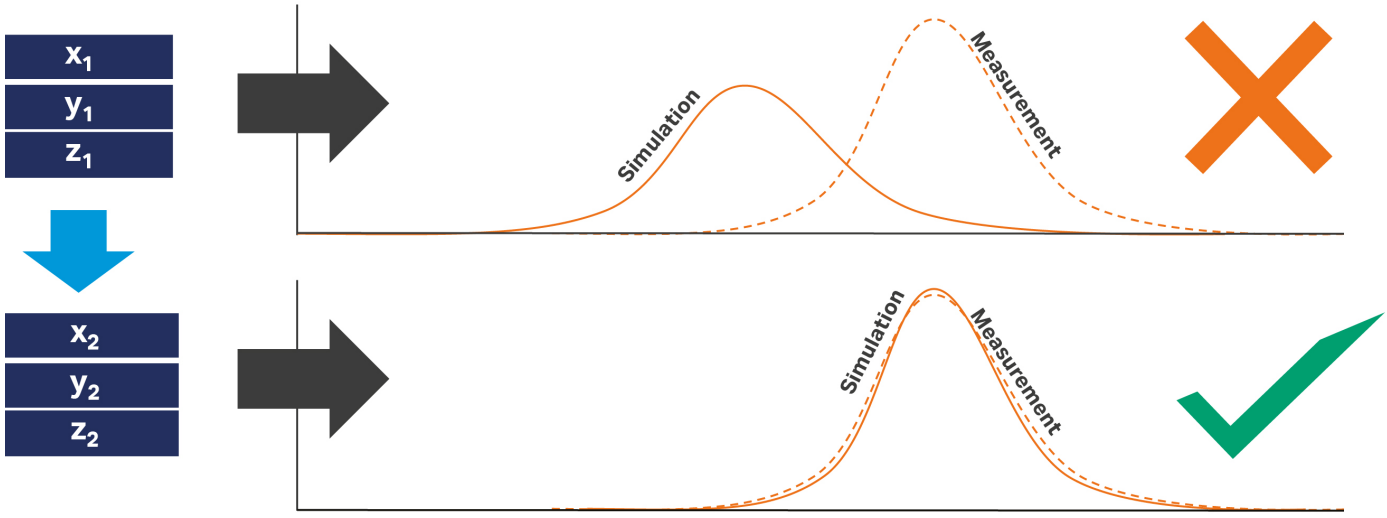

Fig 4. Schematic representation of simulated signal (bold line) and measurement (dotted line) before (top) and after (bottom) calibration. Calibration is achieved by estimating the unknown model parameters (left).

cytiva.com/modeling

Cytiva and the Drop logo are trademarks of Life Sciences IP Holdings Corp. or an affiliate doing business as Cytiva.

GoSilico is a trademark of Global Life Sciences Solutions USA LLC or an affiliate doing business as Cytiva.

Any use of software may be subject to one or more end user license agreements, a copy of, or notice of which, are available on request

© 2022 Cytiva

For local office contact information, visit cytiva.com/contact.

CY25244-02MAR22-PD