When should I exclude data point outliers in a kinetic curve fitting?

For kinetics data analysis of a pair of binding interaction, a global fit using at least four analyte concentrations is recommended to obtain the most accurate kinetic and affinity constants. Outlier data points of binding kinetics data can be excluded to improve the fittings results when fitting the data with kinetics models. However, data exclusion is a user's scientific decision and caution must be exercised on not manipulating the data. Here is a suggested workflow and users can adjust as needed:

1. In Octet Data Analysis software, processing data normally and go to **Tab 3 Analysis**. Below the graph you will see a data table showing the information for all data traces.
2. Prior to conducting any curve fitting, inspect the response values of all curves and exclude any curves that have negative or close to zero response values. This is the initial exclusion of obvious bad data points.
3. Perform the first round of curve fitting by choosing an appropriate fitting model and analysis mode, and clicking the **Fit Curves!** button. Inspect fitting parameters such as R^2, $?^2$, k_{on} error, k_{dis} error, and residuals. Evaluate the goodness of fit using the general guideline described in How do I evaluate goodness of fit in a kinetic curve fitting for my Octet data?
4. After the first round curve fitting and data evaluation, remove outliers as appropriate. Please refer to How do I exclude undesirable data traces from analysis in a kinetic curve fitting? for how to remove data points in Octet Data Analysis Software. Removing the data traces that have most deviations from the ideal curves (the fitted curves) often improves curve fitting. Be sure not to manipulate data. Some tips:
    - When visually inspecting the curves, one might encounter obvious fitting deviations. For example, large deviations are often seen at the highest concentrations, and smaller deviations are seen with decreased analyte concentrations. Users can try to exclude the top concentration, and see if it improves data fitting. Often excluding data points far away from the expected K_D will likely be mostly helpful.
    - Do NOT selectively remove a data point in the middle of the analyte concentration titration series. Such "pick-and-choose" is data manipulation, and is not recommended or supported. Removed data points should be either at the top or bottom concentrations, and must be consecutive if multiple concentrations are to be excluded. The goal is to ensure the retained data points form an uninterrupted serial dilution data series. For example, in an analyte dilution series, 100, 50, 25, 12.5, 6.25, 3.63, and 1.81 nM of analyte were tested. The initial curve fitting showed that the 100 nM curve was heterogeneous and could not be fitted well with the kinetics model. Removing the 100 nM data point improved the curve fitting. However, one should not further remove 25 nM data point if the 50 nM is not going to be removed. Removing the 25 nM data point alone without removing the higher concentration data (100nM and 50nM) is considered as data manipulation.
    - Pay attention to residuals. While $?^2$ is indicative of an accumulated error, the residual graphs allow one to judge which curve fitting is least desirable (with most significant residual deviations).
    - While removing outliers can improve fitting, it is important to retain four or more analyte concentrations in a dilution series (for global fitting) to ensure fitted kinetics values are reliable. The retained analyte concentrations should bracket the K_D (typical range: 10-20X of K_D to 0.1X of K_D). It is a user's judgment call on whether exceptions should be applied, but the nature of the interaction and the assay goal should always be considered. For example, in the case of high affinity binders, the binding signal below the K_D may not be measurable. For more information on the strategy for such scenario, please refer to Application Note 14: Biomolecular Binding Kinetics Assays on the Octet Platform
5. Refit the data and confirm that R^2 is over 0.95 for large molecule kinetics.
6. Always visually evaluate if the fitted lines agree with the raw data traces.
7. Most importantly, the result from the kinetics assay should corroborate with the result from an independent approach, such as a biological assay.

Please refer to Application Note 14: Biomolecular Binding Kinetics Assays on the Octet Platform for additional guidance on assay development and data analysis.

Article ID: 44

Posted: Sat, Oct 15, 2016 at 12:55 PM.

Last Updated: Thu, Aug 3, 2017 at 4:29 PM

Filed Under: Data,Assays
