# Curve fitting

This is intended to be an aid-memoir for those involved in fitting biological data to dose response curves in why we do it and how to best do it.

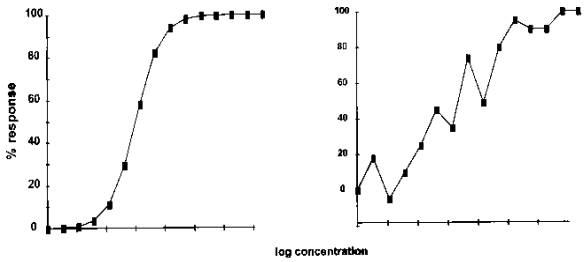
- So why do we use non linear regression?

Much of the data generated in Biology is sigmoidal dose response (E/[A]) curves, the aim is to generate estimates of upper and lower asymptotes, $A_{50}$ and slope parameters and their associated confidence limits.

How can we analyse E/[A] data? Why fit curves? What's wrong with joining the dots?

Fitting

- Uses all the data points to estimate each parameter
- Tell us the confidence that we can have in each parameter estimate
- Estimates a slope parameter

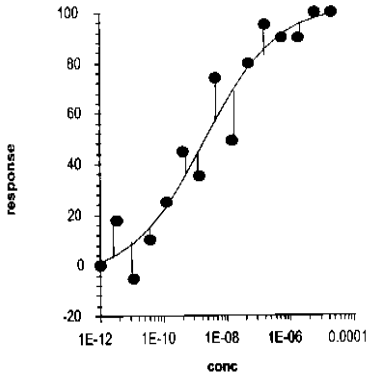



Fitting curves to data is model fitting. The general equation for a sigmoidal dose-response curve is commonly referred to as the 'Hill equation' or the 'four-parameter logistic equation".

$$\text{Response} = Bottom + \frac{(Top - Bottom)}{1 + \frac{EC50^{hillslope}}{[Drug]^{hillslpoe}}}$$

The models most commonly used in Biology for fitting these data are 'model 205 – Dose response one site -4 Parameter Logistic Model or Sigmoidal Dose-Response Model' in XLfit (package used in ActivityBase or Excel) and 'non-linear regression - sigmoidal variable slope' in GraphPad Prism.

<u>What the computer does - Non-linear least squares</u>




- Starts with an initial estimated value for each variable in the equation
- Generates the curve defined by these values
- Calculates the sum of squares
- Adjusts the variables to make the curve come closer to the data points (Marquardt method)
- Re-calculates the sum of squares
- Continues this iteration until there's virtually no difference between successive fittings. Fit has converged.
- Reports the results
- Note that the final values depend on the initial values, so repeated analysis may give slightly different results

XLfit and GraphPad Prism both make their own initial estimates from your raw data.

For more detail refer to 'Fitting Models to Biological Data using Linear and nonlinear regression' by Motulsky and Christopoulos; this is a manual supplied with every copy of GraphPad Prism, for enzyme assays 'Evaluation of Enzyme Inhibitors in Drug Discovery: A Guide for Medicinal Chemists and Pharmacologists' 2005 Copeland or for receptor assays 'A Pharmacology Primer-Theory, Application, and Methods' 2004 Kenakin.

# Curve Fitting best practice and QC

**REMEMBER - Curve fitting packages are not intelligent and results must be scrutinised.**

## Fitting guidelines

When fitting dose response curves the 2 most important things to remember are

### 1. Where ever possible use an unconstrained (4-parameter) fit

If you must constrain parameters, constrain only one parameter, if two parameters need constraining then the data needs regenerating with a more appropriate concentration range.

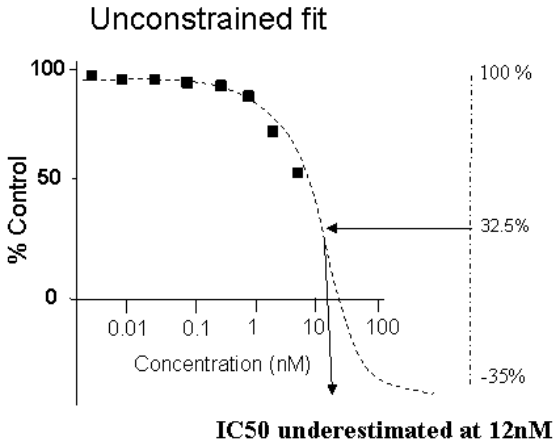
### 2. Do not remove data points

Unless they are obvious outliers or you know you have made a mistake and they need to be excluded. It is far better to show a variable data set, than to try to 'polish' data. If the experiment is inherently variable generate more determinations n=3 or 4. Four pieces of independent information are better than curves with quadruplicate points.

## QC: How to tell if the curve fit is good

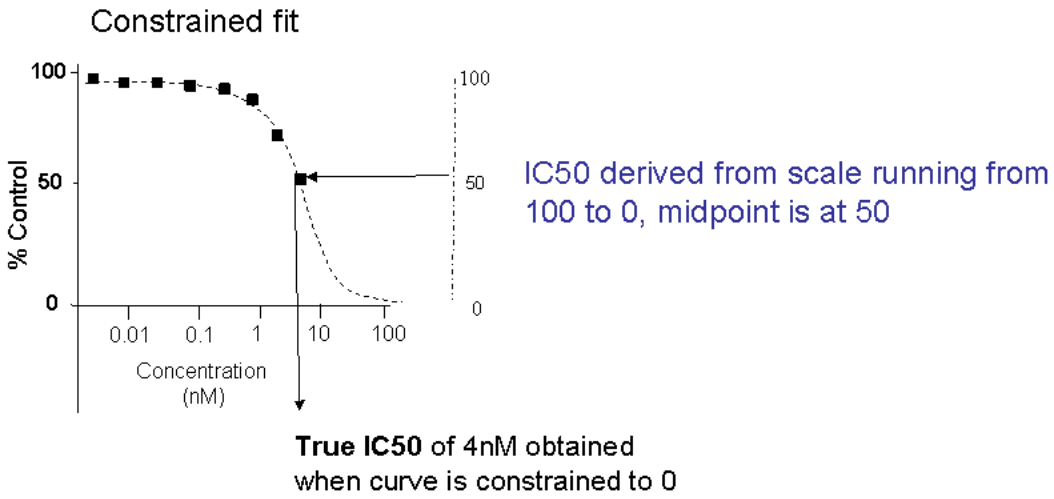
1. Do the max, min and slope make meaningful sense?

Look at the max and min (upper and lower asymptote) are the values sensible? This example of poor lower asymptote estimate underestimates the true IC50 value for the compound.




IC50 derived from scale running from 100 to -35 a scale of 135 units half this response is at 32.5 units

**IC50 underestimated at 12nM**




2. Does the curve go through the points symmetrically?
   - Random scatter
   - Deviation from model
   - F-statistic

3. Goodness of fit statistics: 95% confidence intervals, sum of squares

## For receptor or enzyme inhibition assays:

Data should run from 100% to 0% control, with a slope of 1 when the interaction is simple competitive one-site.

If they don't - check that your controls are constructed properly often the error is due to un-matched solvent concentration.

Analysing and reporting inhibition (receptor or enzyme) assay data:

1. Where ever possible use an unconstrained (4-parameter) fit

2. Constrain to 0% control
   a. **if** lower asymptote poorly defined (>5% to <-5%)
   b. **if** compound reaches greater than 50% inhibition
   c. **otherwise** report activity as greater than the last concentration tested
      i.e. IC50 >10µM, pIC50 <5

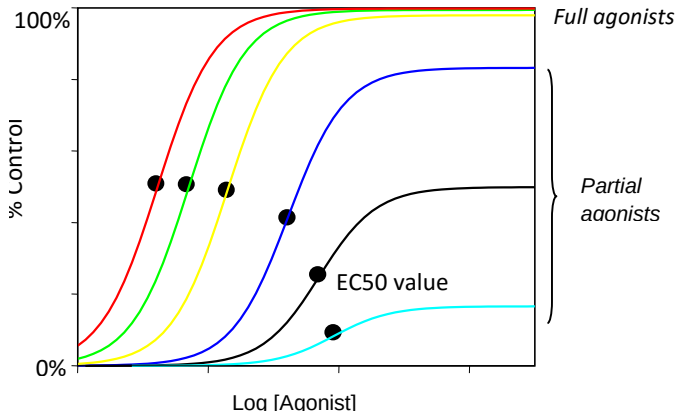
   Run subsequent assays at higher compound concentration

3. Constrain to 100%
   a. if upper asymptote poorly defined <80% at lowest concentration tested

   Run subsequent assays from a lower compound concentration

## For receptor agonist assays:

Data often does not run from 0% to 100% as compounds may be full or partially efficacious agonists.




EC50 is defined as the concentration of agonist that produces a response equivalent to 50% of the maximal response to that agonist. In other words EC50 is the response halfway between the baseline and the highest response achieved, so if you have a partial agonist that starts at 0% and plateaus at 60% of your control reference agonist the halfway response (EC50) is at 30% control.

Analysing and reporting agonist data:

1. If curve has reached >80% control but not formed obvious plateau (no further increase in effect when concentration increased by > ½ log) then constrain to 100%. Run subsequent assays at higher compound concentration.
2. If data are >20% control and curve has reached plateau, allow data to report EC50 for this effect.
3. If data are <20% control then report activity as greater than the last concentration tested i.e. EC50 >10µM, pEC50 <5. Run subsequent assays at higher compound concentration