Method Sheet 112

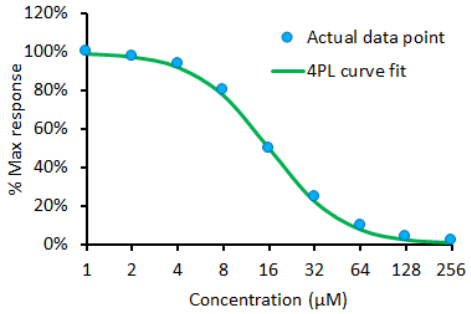
How to fit a 4PL curve to dose response data and calculate IC50

Overview

This method sheet explains how to fit a 4 parameter logistic (4PL) curve to experimental dose response data for the purpose of calculating the IC_{50} of a compound or extract. In drug discovery, we are always keen to learn how potent our new compound or extract may be with respect to inhibiting the target of interest. We do this by calculating the concentration at which the agent causes a 50% inhibition of the target (e.g. bacterial growth, enzyme activity). This concentration is referred to as the Inhibitory Concentration 50%, or **IC_{50}**. The smaller this value, the more potent the drug. IC_{50} is normally measured by first performing a dose response experiment, in which serial doubling dilutions of the agent are tested in a bioassay. Treatment with the negative control should yield no inhibition (100% activity), and the positive control should yield full inhibition (0% activity).

Most compounds display an S-shaped response curve when multiple different concentrations are tested in this way (see example below). Being able to plot a line that appropriately fits through these points is one of the bedrocks of pharmacology, and is an essential starting point for the calculation of IC_{50}. As fitting an S-shape curve to experimental data cannot be achieved with a simple linear equation, we instead use a **four parameter logistic equation**, which is represented as follows:

$$Y = \text{Bottom} + \frac{\text{Top} - \text{Bottom}}{1 + \left(\frac{IC_{50}}{X}\right)^{\text{HillSlope}}}$$



In this example:

- Y = the response we are measuring in the experiment (e.g. cell growth)
- Top = the maximum response in the bioassay (typically 100%)
- Bottom = the minimum response in the bioassay (typically 0%)
- X = the concentration of the drug as tested
- IC_{50} = the concentration at which our drug achieves 50% inhibition of the target
- HillSlope = how steep the middle part of the S-curve is, typically between -1 and -3

Certain software packages, such as GraphPad Prism, are able to automatically fit a 4PL curve and calculate IC_{50} from dose response data. However, as students may not have access to this software, this method sheet will explain how to plot a 4PL curve and calculate IC_{50} manually using Microsoft Excel.

Arrangement of data

1) First, create a data table in a fresh Excel worksheet with the following headings - tested concentrations in the first row, then in the next three cells of the first column, 'Observed data', '4PL curve' and 'Residual', as shown below:

	A	B	C	D	E	F	G	H	I	J
1										
2		256	128	64	32	16	8	4	2	1
3	Observed data									
4	4PL curve fit									
5	Residual									

2) Note that the concentrations begin with the highest dose tested in the experiment, then decrease by a factor of two each step - this is to match the arrangement of data on the experimental plates as given in the downloaded data files.

3) Also note that we do not include the zero concentration, since it is not possible to plot zero values on a log scale.

4) Now, paste as values the data (which should be the mean of at least 3 experiments) for each respective concentration into the 'Observed data' row and format the cells as percentages, as shown below:

	A	B	C	D	E	F	G	H	I	J
1										
2		256	128	64	32	16	8	4	2	1
3	Observed data	2%	4%	10%	25%	50%	80%	94%	98%	100%
4	4PL curve fit									
5	Residual									

5) Now create another table below this one with the following headings, note that these entries must go in the same cell references as shown otherwise the equations given in this example will not work.

	A	B	C
1			
2		256	128
3	Observed data	2%	4%
4	4PL curve fit		
5	Residual		
6			
7	Guess for IC50		10
8	Guess for Hill slope		-1.0
9	Top value		100%
10	Bottom value		0%
11	Residual sum		0.000
12			

6) The squares with yellow highlight are where we will make our **guesses** for the correct IC_{50} and Hill slope values.

7) It can be helpful to begin with guesses of 10 for IC_{50} and -1.0 for the Hill slope.

8) In cell `C7`, type the value 10, and in cell `C8`, type the value `-1`

9) In cell `C9`, type the value 1 then format to percentage to give 100%, and in cell `C10` type the value 0 and format as percentage.

10) In cell `C11`, type the following formula: `=SUM(B5:J5)`

11) Cell `C11` will then calculate the 'Sum of Squares', which is a measure of how closely our line fit goes through all the points on the chart - the lower this value the more accurate the curve fit.

12) Now in cell `B4`, type the following equation, which will calculate our 4PL line fit:

```
=$C$10+($C$9-$C$10)/(1+($C$7/B2)^$C$8)
```

13) Make sure that all the dollar signs shown are correctly included (apart from `B2`) - this formula is the Excel version of the 4PL equation shown earlier in the background.

14) Now drag to copy that equation across the whole row, it should now look like this:

	A	B	C	D	E	F	G	H	I	J
1										
2		256	128	64	32	16	8	4	2	1
3	Observed data	2%	4%	10%	25%	50%	80%	94%	98%	100%
4	4PL curve fit	4%	7%	14%	24%	38%	56%	71%	83%	91%
5	Residual									
6										
7	Guess for IC50		10							
8	Guess for Hill slope		-1.0							
9	Top value		100%							
10	Bottom value		0%							
11	Residual sum		0.000							

15) Now we have to calculate the residuals - in cell B5, type the following formula:

```
=(B3-B4)^2
```

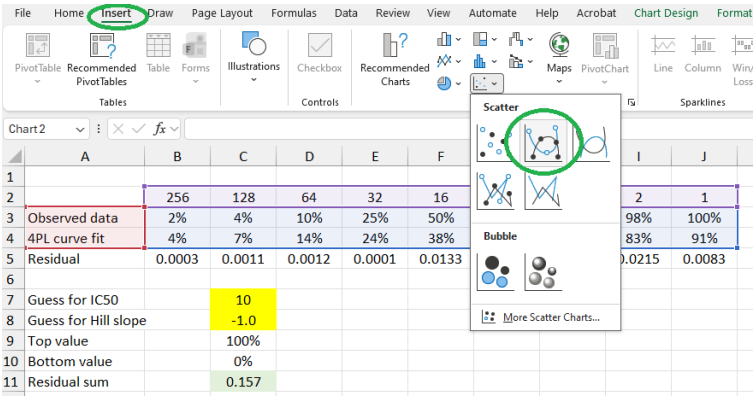
16) This simply takes the square of the difference between the observed data value and the estimated value from the curve equation, for each concentration.

17) Now drag this to copy along the whole row, so you should now see the following:

	A	B	C	D	E	F	G	H	I	J
1										
2		256	128	64	32	16	8	4	2	1
3	Observed data	2%	4%	10%	25%	50%	80%	94%	98%	100%
4	4PL curve fit	4%	7%	14%	24%	38%	56%	71%	83%	91%
5	Residual	0.0003	0.0011	0.0012	0.0001	0.0133	0.0598	0.0509	0.0215	0.0083
6										
7	Guess for IC50		10							
8	Guess for Hill slope		-1.0							
9	Top value		100%							
10	Bottom value		0%							
11	Residual sum		0.157							

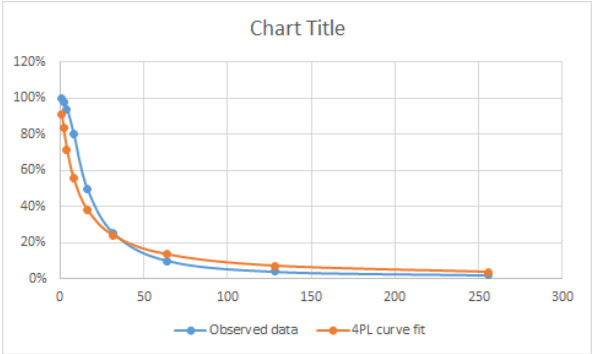
18) Now we must create a chart of both the experimental data points, which we will plot as points only, and the 4PL curve, which we will plot as a line only.

19) Highlight the first three rows of the upper data table (do not include the residuals row), select the 'Insert' tab then 'Scatter plot' with smooth lines, as shown below:

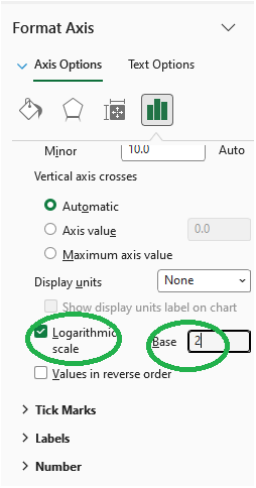


20) Note that you MUST select scatter plot at this stage, the following steps will not work if you accidentally use any other type of chart plotting.

21) You should now see a chart something like this:

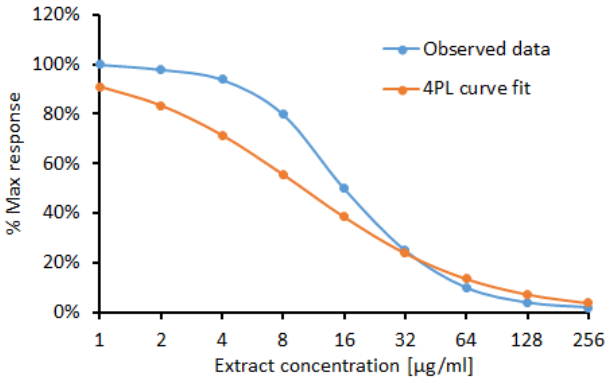


22) Double click on the x-axis, and in the 'Format Axis' toolbar that appears, select 'Logarithmic scale' and 'Base 2'.



23) It is good practice to also format the chart as would be expected for publication quality or for insertion into your dissertation, so give it the following properties:
- Axis lines 1.5 point black with outer major tick marks
- Text black and 12 point
- Add appropriate x- and y-axis titles
- Remove the chart title from the top of the chart, the outer border and the internal gridlines

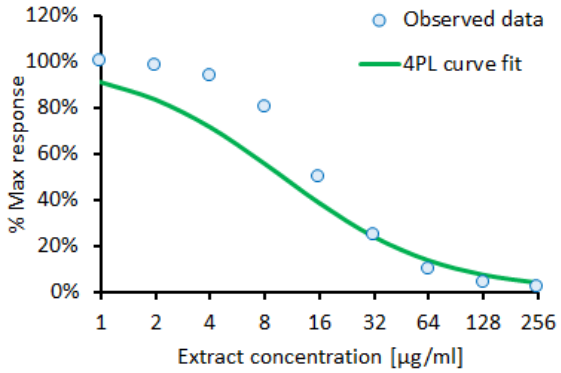
24) Your chart should now look something like this:



25) Now click on the 'Observed data' line and format it so that it shows only points and no line using the 'Format Data Series' toolbar that appears at right of screen.

26) Click on the '4PL curve fit' line, and use the same toolbar to remove the points but keep the line (Marker → Marker Options → None).

27) Now, after changing the colours to make things more clear, you can see that in this example, the 4PL curve line does not fit well with the data points.



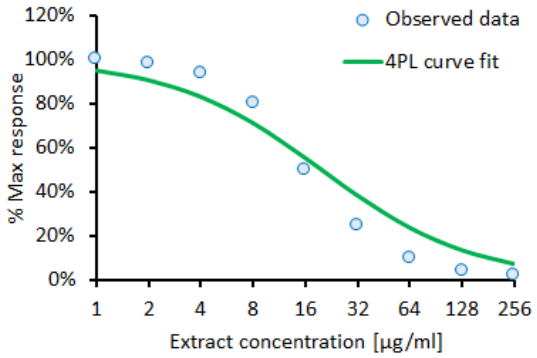
28) To get the curve to fit better, we must improve our guesses of the IC_{50} and Hill slope values in the table below the data values.

29) First, make your best guess of what the IC_{50} might roughly be by drawing an imaginary line across from the 50% value on the y-axis along to the data points, then dropping the line down vertically to where it would land on the x-axis.

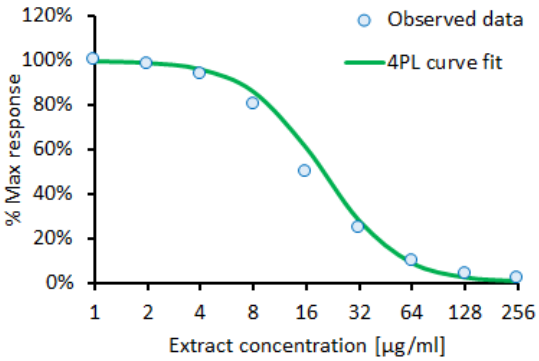
30) Let's imagine that in this example you choose a first guess for IC_{50} of 20, type this in cell `C7` and hit enter, then the curve will slightly shift.

31) The line still does not track well through the points (see left hand chart below), so now try changing the value for the Hill slope in cell `C8`, let's try a value of -2.

32) Now you can see it fits better, but still not quite there (see examples below).



After setting IC_{50} to 20



After setting Hill slope to -2.0

33) What you must now do is keep refining your guesses of IC_{50} and the Hill slope in these cells until the number in the residuals box (cell `C11`) becomes as small as you can get it.

34) Once you have found a residuals number that is as small as you can make it, the curve will be as good as it can be using this method and the process is complete.

35) The IC_{50} value you have arrived at after your final guess is the one you can report in your dissertation.

36) In this example, the best values for IC_{50} and Hill slope are 16 and -1.8, respectively.

37) Your own data could have very different optimum values for IC_{50} and Hill slope, so be prepared to change these to whatever is necessary to fit your own data.

Notes

- To save time and effort in creating the chart from scratch each time, you can save your first working example as a template file, then simply paste as values the data from any additional experiments in the 'Observed data' row shown above.
- If the extract or compound is not sufficiently potent to achieve complete inhibition (i.e. 0% cell growth or 0% enzyme activity), but you can see it is obviously plateauing at an intermediate value, such as 20%, you may need to constrain the bottom parameter to that value (instead of 0%) to obtain a valid IC_{50} concentration.
- We must choose Hill Slope values that are negative when calculating IC_{50} because inhibitors cause the S-shape curve to slope downwards to the right.
- Remember to always plot the chart as a Scatter plot, and set the x-axis to a Log scale - the curve fit line will not plot properly without this.