FITTING DOSE RESPONSE CURVES

An excerpt from a forthcoming book:

Fitting models to biological data using linear and nonlinear regression. A practical guide to curve fitting.

Harvey Motulsky
GraphPad Software Inc.
Hmotulsky@graphpad.com

Arthur Christopoulos
Dept. Pharmacology
University of Melbourne
arthurc1@unimelb.edu.au

© 2002, GraphPad Software, Inc.
www.graphpad.com

(Draft of October 10, 2002)
Introduction to dose-response curves

What is a dose-response curve?

Dose-response curves can be used to plot the results of many kinds of experiments. The X-axis plots concentration of a drug or hormone. The Y-axis plots response, which could be almost any measure of biological function. For example, the response might be enzyme activity, accumulation of an intracellular second messenger, membrane potential, secretion of a hormone, change in heart rate or contraction of a muscle.

The term "dose" is often used loosely. In its strictest sense, the term only applies to experiments performed with animals or people, where you administer various doses of drug. You don't know the actual concentration of drug at its site of action—you only know the total dose that you administered. However, the term "dose-response curve" is also used more loosely to describe in vitro experiments where you apply known concentrations of drugs. The term "concentration-response curve" is therefore a more precise label for the results of these types of experiments. The term "dose-response curve" is occasionally used even more loosely to refer to experiments where you vary levels of some other variable, such as temperature or voltage.

An agonist is a drug that binds to a receptor and causes a response. If you administer various concentrations of an agonist that causes a stimulatory response, the dose-response curve will go uphill as you go from left (low concentration) to right (high concentration); for an agonist that causes an inhibitory response, the opposite curve profile is observed, i.e., a downhill curve with increasing agonist concentrations. A full agonist is a drug that appears able to produce the maximum cellular or tissue response. A partial agonist is a drug that provokes a response, but the maximum response is less than the maximum response to a full agonist in the same cell or tissue. An inverse agonist is a drug that reduces a pre-existing basal response that is due to constitutive activation of a system in the absence of other ligands, e.g., perhaps due to an activating mutation in a receptor.

An antagonist is a drug that does not provoke a response itself, but blocks agonist-mediated responses. If you vary the concentration of antagonist (in the presence of a fixed concentration of agonist), the antagonist dose-response curve (also called an "antagonist inhibition curve") will run in the opposite direction to that of the agonist dose-response curve. It should be noted that the classification of drugs as full agonists, partial agonists, inverse agonists and antagonists is highly dependent on the biological system in which they are tested. For example, if drug binding is strongly coupled to response in one system and only weakly coupled to response in another system, then a full agonist in the first system may appear as a partial agonist in the second system. Similarly, if a system is not constitutively active in the absence of ligands, then an inverse agonist in such a system would appear indistinguishable from a simple antagonist.

The shape of dose-response curves

Many steps can occur between the binding of the agonist to a receptor and the production of the response. So depending on which drug you use and which response you measure, dose-response curves can have almost any shape. However, in very many systems, dose-response curves follow a standard shape. While a plot of response vs. the amount of drug is typically a rectangular hyperbola, the dose range for the full relationship may span several orders of magnitude, so it is more common to plot response vs. logarithm of the dose.
curves have shapes almost identical to receptor binding curves. The simplest explanation is that the link between receptor binding and response is direct, so response is proportional to receptor binding. However, in most systems one or more second-messengers can link receptor binding to the final, measured, response. For example, the binding of some agonists can activate adenyl cyclase, which creates the second-messenger cAMP. The second messenger then binds to an effector (such as a protein kinase) and initiates or propagates a response.

What do you expect a dose-response curve to look like if a second messenger mediates the response? If you assume that the production of second messenger is proportional to receptor occupancy, the graph of agonist concentration vs. second messenger concentration will have the same shape as receptor occupancy (a hyperbola if plotted on a linear scale, a sigmoid curve with a slope factor of 1.0 if plotted as a semilog graph). If the second messenger works by binding to an effector, and that binding step follows the law of mass action, then the graph of second messenger concentration vs. response will also have that same standard shape. It isn’t obvious, but Black and Leff (see "" on page 23) have shown that the graph of agonist concentration vs. response will also have that standard shape (provided that both binding steps follow the law of mass action). In fact, it doesn’t matter how many steps intervene between agonist binding and response. So long as each messenger binds to a single binding site according to the law of mass action, the dose-response curve will follow the same hyperbolic/sigmoid shape as a receptor binding curve.

The EC$_{50}$

A standard dose-response curve is defined by four parameters: the baseline response (Bottom), the maximum response (Top), the slope (Hill slope), and the drug concentration that provokes a response halfway between baseline and maximum (EC$_{50}$).

It is easy to misunderstand the definition of EC$_{50}$. It is defined quite simply as the concentration of agonist that provokes a response halfway between the baseline (Bottom) and maximum response (Top). It is impossible to define the EC$_{50}$ until you first define the baseline and maximum response. Depending on how you have normalized your data, this may not be the same as the concentration that provokes a response of Y=50. For example, in the example below, the data are normalized to percent of maximum response, without subtracting a baseline. The baseline is about 20%, and the maximum is 100%, so the EC$_{50}$ is the concentration of agonist that evokes a response of about 60% (half way between 20% and 100%).

Don’t over interpret the EC$_{50}$. It is simply the concentration of agonist required to provoke a response halfway between the baseline and maximum responses. Because the EC$_{50}$ defines the location of the dose-response curve for a particular drug, it is the most commonly used measure of an agonist’s potency. However, the EC$_{50}$ is usually not the same as the $K_d$ for the binding of agonist to its receptor, i.e., it is not a direct measure of drug affinity.

The steepness of a dose-response curve

Many dose-response curves follow exactly the shape of a receptor binding curve. As shown below, 81 times more agonist is needed to achieve 90% response than a 10% response.
Some dose-response curves however, are steeper or shallower than the standard curve. The steepness is quantified by the Hill slope, also called a slope factor. A dose-response curve with a standard slope has a Hill slope of 1.0. A steeper curve has a higher slope factor, and a shallower curve has a lower slope factor. If you use a single concentration of agonist and varying concentrations of antagonist, the curve goes downhill and the slope factor is negative. The steeper the downhill slope, the more negative the Hill slope.

The general equation for a sigmoidal dose-response curve is also commonly referred to as the “Hill equation”, the “four-parameter logistic equation”, or the “variable slope sigmoid equation”. One form of this equation is as follows:

$$\text{Response} = \text{Bottom} + \frac{(\text{Top-Bottom})}{1 + \left(\frac{\text{EC}_{50}}{[\text{Drug}]}\right)^\text{HillSlope}}$$

When you fit this equation, you want to find the best fit value of the logEC50, rather than the EC50 itself (see next section). Making that change, as well as defining Y to be the response and X to be the logarithm of [Drug] gives us:

$$\text{Response} = \text{Bottom} + \frac{(\text{Top-Bottom})}{1 + 10^{\log\text{EC}_{50}} \cdot X^{\text{HillSlope}}}$$

If you wish to manually enter this equation into your nonlinear regression program, the following syntax is standard for many different software packages:

$$Y = \text{Bottom} + \frac{(\text{Top-Bottom})}{1 + 10^{((\log\text{EC}_{50}) - X) \cdot \text{HillSlope}}$$

Dose-response curves where X is concentration, not log(concentration)

Dose-response curves are generally performed with concentrations that are equally spaced on a log scale, and are usually fit to find the best-fit value of the logEC50 (see below). It is also possible to make the concentrations equally spaced on a linear scale, and fit to find the EC50.

Start with the standard equation for the dose-response curve:

$$\text{Response} = \text{Bottom} + \frac{(\text{Top-Bottom})}{1 + \left(\frac{\text{EC}_{50}}{[\text{Drug}]}\right)^\text{HillSlope}}$$

Define Y to be response, and X to be [Drug], and simplify.

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

Written as a user-defined equation for most nonlinear regression programs:
Fitting dose-response curves

Y = Bottom + (Top - Bottom) / (1 + (EC50/X)^HillSlope)

When the Hill Slope is set to 1.0, this is the same as the one-site binding hyperbola (except this equation adds a bottom baseline term).

When the Hill Slope is much greater than 1.0, the dose-response curve has a sigmoidal shape.

Note the confusing point here. A standard dose response curve, with a Hill Slope equal to 1.0, has a sigmoidal shape when $X$ is the log of concentration or dose. The same standard dose response curve, with a Hill slope equal to 1.0) has a hyperbolic shape when $X$ is concentration (or dose). Nothing sigmoidal about it.

If the Hill slope is greater than 1.0, the curve has a sigmoidal shape either way – when $X$ is concentration (or dose) or when $X$ is the logarithm of concentration (or dose).

Tip: When you see a sigmoidal dose-response curve, look carefully at the X axis to see if X is concentration (or dose) or the logarithm of concentration (or dose).

In general, you should avoid fitting dose-response curves on a linear scale, for two reasons. First, if the curve spans many orders of drug dose magnitude, then it becomes graphically difficult to present. Second, the error associated with the EC50 parameter (linear scale) of the standard dose-response model does not follow a Gaussian distribution and therefore cannot be used in standard statistical analyses that require the parameters follow a Gaussian distribution. This is discussed next.

Why you should fit the logEC50 rather than EC50

As shown above, you can write an equation for a dose-response curve either in terms of EC50 or log EC50. Curve fitting finds the curve that minimizes the sum-of-squares of the vertical distance from the points. Rewriting the equation to change between EC50 and log EC50 isn’t going to make a different curve fit better. All it does is change the way that the best-fit EC50 is reported.

However, rewriting the equation to change between EC50 and log EC50 has a major effect on standard error and confidence interval of the best-fit values. Consider these sample results:

These data were fit to a dose-response curve with a Hill slope of 1. The best-fit value for logEC50 is -6.059. Converting to the EC50 is no problem – simply take the antilog. The EC50 is $10^{-6.059}$ M, about 0.87 µM.

The standard error of the logEC50 is 0.2717. It is used as an intermediate result to calculate a confidence interval, which ranges from -6.657 to -5.461. This means that the 95%CI of the EC50 extends from $10^{6.657}$ to $10^{-5.461}$ -- from 0.22 to 3.46 µM. Expressed as concentrations (rather than log of concentration) the interval is not centered on the best-fit value (0.87 µM). Switching from linear to log scale turned a symmetrical confidence interval into a very asymmetrical interval, which you can report.
If you fit the same data to an equation describing a dose-response curve in terms of the EC$_{50}$ rather than the logEC$_{50}$, the EC$_{50}$ remains 0.87 µM. But now the program computes the SE of the EC$_{50}$ (0.5459 µM), and uses this to compute the 95% confidence interval of the EC$_{50}$, which ranges from -0.3290 to +2.074 µM. Note that the lower limit of the confidence interval is negative! Since the EC$_{50}$ is a concentration, negative values are nonsense. Even setting aside the negative portion of the confidence interval, it includes all values from zero on up, which isn't terribly useful.

The problem is that the uncertainty of the EC$_{50}$ really isn't symmetrical, especially when you space your doses equally on a log scale. Nonlinear regression (from Prism and most other programs) always reports a symmetrical confidence interval. In cases like this – fitting dose response data to a model written in terms of the EC$_{50}$ – the confidence interval is not very helpful.

When some people see the SE of the logEC$_{50}$, they are tempted to convert this to the standard error of the EC$_{50}$ by taking the antilog. In the example, the SE of the logEC$_{50}$ is 0.2717. The antilog of 0.2717 equals $10^{0.2717}$ or 1.869. What does this mean? It certainly is NOT the SE of the EC$_{50}$. The SE does not represent a point on the axis; rather it represents a distance along the axis. A distance along a log axis does not represent a consistent distance along a linear (standard) axis. For example, increasing the logEC$_{50}$ 1 unit from -9 to -8 increases the EC$_{50}$ 9nM; increasing the logEC$_{50}$ 1 unit from -3 to -2 increases the EC$_{50}$ by 9 mM (which equals 9,000,000 nM). So you cannot interpret the number 1.869 as a concentration. You can interpret it as a multiplier – a factor you multiply by or divide into the EC$_{50}$. To calculate the 95% CI, first multiply 1.869 by a constant from the t distribution for 95% confidence and the appropriate number of degrees of freedom (11 degrees of freedom in this example, so $t = 2.201$). The result is 4.113 Then compute the 95% CI of the EC$_{50}$, It extends from the best-fit EC$_{50}$ divided by 4.113 to the best-fit EC$_{50}$ times 4.113, from 0.21 µM to 3.58 µM.

### Decisions when fitting sigmoid dose-response curves

From the preceding discussion, it is clear that most of the time you should enter your X values as logarithms of concentration or dose if you want to perform a standard sigmoidal dose-response curve fit. If you entered actual concentrations, most standard data analysis programs can transform those values to logarithms for you.

### Before fitting a dose-response curve, you will need to make these decisions:

<table>
<thead>
<tr>
<th>Decision</th>
<th>Discussion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Choose Hill slope of 1 or variable slope?</td>
<td>If you have plenty of data points, you should choose to fit the Hill Slope along with the other parameters. If data are scanty, you may wish to consider fixing the Hill Slope to a value of 1.</td>
</tr>
<tr>
<td>Set Top to a constant value?</td>
<td>Ideally, the top part of the curve is defined by several data points. In this case, the nonlinear regression program will be able to fit the top plateau of the curve. If this plateau is not well defined by data, then you'll need to make the top plateau be a constant based on controls.</td>
</tr>
<tr>
<td>Set Bottom to a constant value?</td>
<td>Ideally, the bottom part of the curve is defined by several data points. In this case, the nonlinear regression program will be able to fit the bottom plateau of the curve. If this plateau is not well defined by data, you'll need to make the bottom plateau be a constant based on controls. If you have subtracted a background value, then the bottom plateau of the curve must be 0. The program won't know this unless you tell it. Make Bottom a constant equal to zero in this case.</td>
</tr>
<tr>
<td>Absolute or relative weighting?</td>
<td>See &quot;Error! Reference source not found.&quot; on page Error! Bookmark not defined..</td>
</tr>
<tr>
<td>Fit each replicate or averages?</td>
<td>See &quot;Error! Reference source not found.&quot; on page Error! Bookmark not defined..</td>
</tr>
</tbody>
</table>

### Checklist. Interpreting a dose-response curve.

After fitting a dose-response model to your data, ask yourself these questions:

Note: Since the logarithm of zero is undefined, you cannot enter a concentration of zero as a logarithm. If you enter a concentration of zero and then transform to logarithms, Prism will leave that result blank. Instead of entering a dose of zero, enter a low concentration, e.g., one log unit below your lowest non-zero concentration.
Advanced analyses of dose-response curves

Other measures of potency

The pEC$_{50}$

The pEC$_{50}$ is defined as the negative logarithm of the EC$_{50}$. If the EC$_{50}$ equals 1 micromolar ($10^{-6}$ molar), the log EC$_{50}$ is –6 and the pEC$_{50}$ is 6. There is no particular advantage to expressing potency this way, but it is customary in some fields.

Note: Expressing potency as the pEC$_{50}$ is a similar practice to quantifying acidity with the pH, which is the negative logarithm of [H$^+$].

If you want to fit the pEC$_{50}$ directly rather than fitting the logEC$_{50}$, use the following equation syntax.

\[ Y = Bottom + (Top - Bottom) / (1 + 10^{*(X - pEC50)*HillSlope}) \]

Calculating any EC value from the EC$_{50}$ and Hill slope

The potency of a drug is commonly quantified as the EC$_{50}$ or the logarithm of the EC$_{50}$. But in some systems you might be more interested in the EC$_{50}$ or the EC$_{90}$ or some other value. You can compute the EC$_{80}$ or EC$_{90}$ (or any other EC value) from the EC$_{50}$ and Hill slope. Or you can fit data to determine any EC value directly. If you express response as a percentage, a standard dose-response curve is described by this equation:

\[ F = 100 \times \frac{[A]^H}{[A]^H + EC_{50}^H} \]

[A] is the agonist concentration, EC$_{50}$ is the concentration that gives half-maximal effect, and H is the Hill constant or slope factor that defines the steepness of the curve. [A] and EC$_{50}$ are expressed in the same units of concentration, so the units cancel out. F is the fractional response, expressed as a percentage.
If you set F to any fractional response you want, and define EC_F as the agonist concentration necessary to achieve that response, then by substitution in the equation above,

\[ F = 100 \times \frac{EC_F^H}{EC_F^H + EC_{50}^H} \]

and rearranging yields this equation:

\[ EC_F = \left( \frac{F}{100-F} \right)^H \times EC_{50} \]

If you know the EC_{50} and Hill slope (H), you can easily compute the EC_{50} or EC_{10} or any other value you want. For example, if the Hill slope equals 1, the EC_{90} equals the EC_{50} times nine. If H equals 0.5, the curve is shallower and the EC_{90} equals the EC_{50} times 81.

**Determining any EC value directly**

You can also fit data directly to an equation written in terms of the EC_F. The advantage of this approach is that Prism will report the 95% confidence value for EC_F. Use the equation below, where X is the log of concentration and Y is response, which ranges from Bottom to Top. In the example below, F is set to a value of 80, but you can set it to be any desired value between 0 and 100.

\[
\text{F} = 80
\]
\[
\log EC_{50} = \log EC_F - \left( \frac{1}{\text{HillSlope}} \right) \log \left( \frac{9}{100-F} \right)
\]
\[
Y = \text{Bottom} + \frac{(\text{Top} - \text{Bottom})}{\left(1 + 10^{(\log EC_{50} - X) \times \text{HillSlope}} \right)}
\]

To fit data to this equation, you’ll need to consider reasonable initial values for your parameters. We suggest setting Top equal to your maximal Y value and Bottom equal to your minimum Y value, as determined from your datapoints. For HillSlope, simply pick a value, probably +1.0 or –1.0. For logEC, enter the logarithm of your middle X value as a crude initial value, or enter a value based on the range of concentrations you use.

Here is a simplified equation for fitting the EC_{90}. Here, the response is expressed as a percent ranging from zero to one hundred, so we dispense with the variables Top and Bottom.

\[
\log EC_{50} = \log EC_{90} - \left( \frac{1}{\text{HillSlope}} \right) \log (9)
\]
\[
Y = \frac{100}{\left(1 + 10^{(\log EC_{50} - X) \times \text{HillSlope}} \right)}
\]

Bell-shaped dose-response curves

When plotted on a logarithmic axis, dose-response curves usually have a sigmoidal shape, as discussed in the previous chapter. However, some drugs may cause an inhibitory response at low concentrations, and a stimulatory response at high concentrations, or vice-versa. The net result is a bell-shaped dose-response curve.

Bell-shaped dose-response curves have been observed experimentally, for example, for many receptors that couple to both stimulation and inhibition of the enzyme, adenylyl cyclase (see S. Tucek et al., Trends Pharmacol. Sci., 23, 171-176, 2002).

**Tip:** Unless you know that the reason for a non-standard dose-response curve shape in your experiment is due to an experimental error, avoid excluding data points simply to make a non-sigmoid curve fit a sigmoid model. Instead, it is relatively easy to extend the standard model of dose-response curves to accommodate different nonlinear and saturating curve shapes.

Combining two sigmoid equations

The following equation combines two sigmoid dose-response relationships to describe a bell-shaped dose-response curve. In the figure below, the curve begins at Plateau1, turns over at the Dip and then approaches Plateau2. The two different values for the LogEC_{50} and nH parameters denote the midpoint potency and the slope factors, respectively, of each phase of the curve. The variable, [A], denotes the agonist concentration.

\[
Y = \text{Dip} + \frac{(\text{Plateau1} - \text{Dip})}{1 + 10^{\left(\log EC_{50} - \log[A]\right) \times nH_1}} + \frac{(\text{Plateau2} - \text{Dip})}{1 + 10^{\left(\log[A] \times \log EC_{50} \times nH_2\right)}}
\]
Here is one way the equation can be typed directly into a computer program:

\[
\begin{align*}
\text{Span1} &= \text{Plateau1} - \text{Dip} \\
\text{Span2} &= \text{Plateau2} - \text{Dip} \\
\text{Section1} &= \frac{\text{Span1}}{1 + 10^{((\log \text{EC50}_1 - X) \times nH1)}} \\
\text{Section2} &= \frac{\text{Span2}}{1 + 10^{((X - \log \text{EC50}_2) \times nH2)}} \\
\text{Y} &= \text{Dip} + \text{Section1} + \text{Section2}
\end{align*}
\]

Because this model is more complicated than the standard monotonic sigmoid dose-response curve, there are a number of practical considerations when it comes to using the model to fit data. First, it is important that there are sufficient data points to adequately define both phases of the response; otherwise the model will fail to converge because it will have too many parameters relative to the number of points. Second, it can be seen from the graph that there are two general types of bell-shaped relationships possible, one where the dip occurs at the highest level of response, and one where the dip occurs at the lowest level of response. In order for the model to converge successfully, you need to be careful with your choice of initial parameter values. Of particular importance is the sign of the slope parameter, \( nH \). As can be seen in the graph, the slope factors are positive for one kind of curve, but negative for the other.

**Using the Gaussian distribution equation**

Sometimes, dose-response curves exhibit a dip after reaching a maximal response level, but the response after this dip is not sufficiently defined for the investigator to conclude whether the curve is truly bell-shaped or not. This is often observed, for instance, when agonists cause desensitization of the tissue at the highest concentrations of drug used. Another common cause of these kinds of curve shapes is insufficient drug to fully define the entire dose-response relationship, perhaps due to solubility issues.

When the data don’t follow a standard sigmoid shape, you should fit the data to a model that more closely approximates the shape of the curve yet still gives you measures of agonist potency, maximal response range and slope factor.

**Tip.** If your dose-response data shows this kind of dip, beware of fitting your data to the standard sigmoid dose-response curve. The best-fit values for the maximal response and the \( \text{EC}_{50} \) will not be very accurate.

One possibility for fitting these kinds of data is combining two sigmoidal shape curves, as described above. However, this approach is only useful when you have sufficient data points to fully define both phases of a curve.

An alternative is to fit the data to the Gaussian distribution. While this distribution is rarely used to fit dose-response data, the figure below shows that a portion of the Gaussian distribution (solid) looks like a dose-response curve with a dip at the top. This is not a mechanistic model, but is a way to empirically fit your data and get parameters that you can compare between treatments.
This Gaussian distribution has been used successfully to fit dose-response data (see A. Christopoulos et al., J. Pharmacol. Exp. Ther., 298, 1260-1268, 2001). Shown below is how this equation can be rewritten to define a bell-shaped dose-response curve.

\[
E = \text{Basal} + \text{Range} \times e^{\left(\frac{\text{midA} - \text{Basal}}{\text{slope}}\right)}
\]

where

\[
\text{midA} = \text{LogEC}_{50} + \text{slope} \times \sqrt{-\ln(0.5)}
\]

In the original formulation of the Gaussian equation, the “midA” parameter would normally define the mean of the distribution, i.e., the x-axis value corresponding to the midpoint of the distribution – the peak. For fitting dose-response curves, this is not useful because the x-axis value we want is the logarithm of drug causing the response halfway between the Basal and the top of the dip, i.e., the LogEC_{50}. Thus, in the revised equation shown above, midA is corrected to allow for the estimation of the LogEC_{50}.

The reparameterized Gaussian dose-response equation defines a bell-shaped dose-response curve in terms of only four parameters, the Basal response, the LogEC_{50}, the maximal response Range (which is the maximal response range from Basal to the dip in the curve) and the Slope. Because the Gaussian equation requires fewer parameters than the bell-shaped equation described at the start of this section (4 vs. 7), it can be used to fit dose-response curves with fewer data points.

Here is how you type the equation into a computer program.

```plaintext
midA = \text{LogEC}_{50} + \text{slope} \times \sqrt{-\ln(0.5)}

Y = \text{Basal} + (\text{Range} \times \exp(-1*(\text{X-}\text{midA})/\text{slope}^2))
```

The figure below shows a curve fit based on the modified Gaussian equation. The dotted line shows the fit of the standard sigmoid equation to the same dataset. Neither the EC_{50}, nor the maximum response, would be correct if the data were fit to the standard sigmoidal dose response curve.

Troubleshooting tip. What does it mean if the curve fits the data well, but the LogEC_{50} is obviously way too high (beyond the range of your data)? Because of its symmetric nature, the Gaussian equation actually has two LogEC_{50} values, a LogEC_{50} for responses to the left of the dip, and a higher LogEC_{50} for responses to the right of the dip. If the best-fit LogEC_{50} is too high, your program probably fit the wrong one. Enter a smaller initial estimate of the LogEC_{50}, and the program will fit the LogEC_{50} for the part of the curve where you actually have data.

What are the advantages and disadvantages of using this equation to fit bell-shaped data compared to the previous equation (combining two sigmoidal curves)? The main advantage of using the Gaussian is that you are dealing with a model containing fewer parameters, and thus increase your chances of obtaining a satisfactory fit with fewer data points. The main disadvantage of the model is in its symmetric nature. In the Gaussian, the “down” phase of the bell-shape is a mirror image of the “up” phase of the bell-shape. If you have a complete dataset that fully defines both phases of the bell-shape, then the Gaussian will only provide a satisfactory fit if the two phases are practically mirror images, which is not that common. In this latter instance, you are better off using the more complicated bell-shaped model described earlier, which accommodates different slopes and plateaus for the two different phases. The Gaussian is best reserved for those datasets where one of the phases of the curve is well-defined, but the other is not, as shown in the figure above.

Note that the slope parameter is not equivalent to the Hill slope (n_H) found in the sigmoid dose-response equations. Although the slope parameter of the Gaussian allows for curves of varying degrees of steepness, its actual value changes
opposite to that of the Hill slope in a sigmoid fit. That is, for steep curves, the value of the Gaussian slope gets smaller, whereas for shallow curves it gets larger, in contrast to the Hill slope. As with the other bell-shaped equation, therefore, you need to be careful when entering the initial values for the Gaussian equation.

### Biphasic dose-response curves

Another common deviation from the standard monotonic sigmoid shape is the biphasic sigmoid shape. An example of an equation for a biphasic dose-response curve is shown below.

\[
Y_{\text{Bottom}} + \frac{(\text{Top}-\text{Bottom}) \cdot \text{Frac}}{1+10^{\frac{\text{LogEC50}_1-\text{Log[A]}}{nH_1}}} + \frac{(\text{Top}-\text{Bottom}) \cdot (1-\text{Frac})}{1+10^{\frac{\text{LogEC50}_2-\text{Log[A]}}{nH_2}}}
\]

Here \text{Top} and \text{Bottom} are the maximal and minimal responses, respectively, \text{LogEC50}_1 and \text{LogEC50}_2 are the midpoint potency parameters for the two different phases, respectively, \(nH_1\) and \(nH_2\) are their corresponding Hill slopes, and \text{Frac} is the fraction of the curve comprising the more potent phase. The equation syntax is shown below, as is a figure illustrating a fit of the equation to a simulated (with random error) dataset.

As with the preceding equations, successful curve-fitting with this model relies on the number of datapoints, the quality of the data and your Initial values. Our experience with this model is that it is especially sensitive to changes in the slope parameters, which often turn out to be significantly different from 1. In the example above, for instance, the value for \(nH_2\) (for the less potent, right-most phase) was greater than 2, and the model had difficulty converging unless an estimate greater than 1 was entered as the initial value for that parameter. For the \text{Frac} parameter, you can use the Constraints feature in the nonlinear regression dialog box to constrain this value to always be between 0 and 1.

```plaintext
Span=Top-Bottom
Section1=Span\cdot Frac/(1+10^{\text{LogEC50}_1-\text{Log[A]}\cdot nH_1})
Section2=Span\cdot (1-\text{Frac})/(1+10^{\text{LogEC50}_2-\text{Log[A]}\cdot nH_2})
Y_{\text{Bottom}} + \text{Section1} + \text{Section2}
```
The operational model of agonist action

Limitations of dose-response curves

Fitting a standard sigmoidal (logistic) equation to a dose-response curve to determine EC50 (and perhaps slope factor) doesn’t tell you everything you want to know about an agonist. The problem is that the EC50 is determined by two properties of the agonist:

- How well it binds to the receptor, quantified by the affinity of the drug for binding to its receptor.
- How well it causes a response once bound. This property is known as the agonist’s efficacy. Since efficacy depends on both agonist and tissue, a single drug acting on a single kind of receptor can have different efficacies, and thus different EC50 values, in different tissues.

A single dose-response experiment cannot untangle affinity from efficacy. Two very different drugs could have identical dose-response curves, with the same EC50s and maximal responses (in the same tissue). One drug binds tightly with high affinity but has low efficacy, while the other binds with low affinity but has very high efficacy. Since the two dose-response curves are identical there is no data analysis technique that can tell them apart. You need to analyze a family of curves, not an individual curve, to determine the affinity and efficacy. The rest of the chapter explains how.

Derivation of the operational model

Black and Leff (Proc. R. Soc. Lond. B, 220:141-162, 1983) developed the operational model of agonism to help understand the action of agonists and partial agonists, and to develop experimental methods to determine the affinity of agonist binding and a systematic way to measure relative agonist efficacy based on an examination of the dose-response curves.

Start with a simple assumption: Agonists bind to receptors according to the law of mass action. At equilibrium, the relationship between agonist concentration ([A]) and agonist-occupied receptor ([AR]) is described by the following hyperbolic equation:

\[ [AR] = \frac{[R_T] \cdot [A]}{[A] + K_A} \]

[R_T] represents total receptor concentration and K_A represents the agonist-receptor equilibrium dissociation constant.

What is the relationship between agonist-occupied receptor (AR) and receptor action? We know biochemical details in some cases, but not in others. This lack of knowledge about all the steps between binding and final response prevents the formulation of explicit, mechanistic equations that completely describe a dose-response curve. However, Black and Leff derived a “practical” or “operational” equation that encompasses the behavior of all of these unknown biochemical cascades. They began with the observation that many dose-response curves have a sigmoidal shape with a Hill Slope of 1.0. (the curves are hyperbolas when response is plotted against agonist concentration, sigmoidal when response is plotted against the log of agonist concentration). They then proved mathematically that if agonist binding is hyperbolic and the dose-response curve has a Hill slope of 1.0, the equation linking the concentration of agonist-occupied receptors to response must also be hyperbolic. This second equation, shown below, has been termed the “transducer function”, because it is a mathematical representation of the transduction of receptor occupation into a response:

\[ \text{Effect} = \frac{\text{Effect}_{\text{max}} \cdot [AR]}{[AR]+K_E} \]

The parameter, Effect_{max} is the maximum response possible in the system. This may not be the same as the maximum response that a particular agonist actually produces. The parameter K_E is the concentration of AR that elicits half the maximal tissue response. The efficacy of an agonist is determined by both K_E and the total receptor density of the tissue ([R_T]). Black and Leff combined those two parameters into a ratio ([R_T]/K_E) and called this parameter tau (τ), the “transducer constant”. Combining the hyperbolic occupancy equation with the hyperbolic transducer function yields an explicit equation describing the effect at any concentration of agonist:

\[ \text{Effect} = \frac{\text{Effect}_{\text{max}} \cdot \tau \cdot [A]}{(K_A+\tau \cdot [A])} \]

This equation can be rewritten as follows, to make it easier to compare the operational model with the standard sigmoid equation for an agonist dose-response curve.

\[ \text{Effect} = \frac{\text{Effect}_{\text{max}} \cdot \tau \cdot [A]}{K_A+\tau \cdot [A]} \]

\[ \text{Effect}_{\text{max}} \cdot \frac{\tau + 1}{\tau + 1} \]
This form of the equation makes it clear that the maximum effect in the dose-response relationship seen with a particular agonist is not $\text{Effect}_{\text{max}}$, but rather is $\text{Effect}_{\text{max}} \times \frac{\tau}{(\tau + 1)}$. Only a full agonist in a tissue with plenty of receptors (high values of $\tau$) will yield a maximum response that approaches $\text{Effect}_{\text{max}}$.

The $\text{EC}_{50}$ does not equal $K_A$ (the equilibrium dissociation constant for agonist binding to the receptors) but rather $K_A/(1 + \tau)$. With a strong agonist (large $\tau$ value), you'll get half-maximal response by binding fewer than half the receptors, so the $\text{EC}_{50}$ will be much less than $K_A$.

This figure shows a dose-response curve for a partial agonist, and shows the relationship between $\text{EC}_{50}$ and maximum response to terms in the operational model.

The parameter, $\tau$, is a practical measure of efficacy. It equals the total concentration of receptors in the system divided by the concentration of receptors that need to be occupied by agonist to provoke a half-maximal tissue response. The $\tau$ value is the inverse of the fraction of receptors that must be occupied to obtain the half-maximal response. If $\tau$ equals 10, that means that occupation of only 10% of the receptors leads to a half-maximal response. If $\tau$ equals 1.0, that means that it requires occupation of all the receptors to give a half-maximal response. This would happen with a partial agonist or with a full agonist in a tissue where the receptors had been significantly depleted. Because $\tau$ is a property of both the tissue and receptor system, it is not a direct measure of intrinsic efficacy, which is commonly defined as a property belonging only to an agonist-receptor pair, irrespective of the assay system in which it is measured.

The equations above show agonist stimulated response, so the curves all begin at zero. It is easy to add a $\text{Basal}$ parameter to model observed response in the absence of drug, so the response with no agonist equals $\text{Basal}$ rather than zero.

**Shallower and steeper dose-response curves**

Some sigmoid dose-response curves are steeper or shallower than a curve with a standard slope factor of 1. The operational model can be extended to analyze these curves.

If you assume the initial binding of the agonist to the receptor follows the law of mass-action (Hill slope equals 1 for the binding step), then transduction step(s) between occupancy and final response must follow an equation that allows for variable slope. If the dose-response curve is still sigmoid, then the operational model can be extended fairly simply, by including a slope parameter, $n$. The extended form of the operational model is:

$$ \text{Effect} = \frac{\text{Effect}_{\text{max}} \cdot \tau^n \cdot [\text{A}]^n}{(K_A + [\text{A}])^n + \tau^n \cdot [\text{A}]^n} $$

The relationship between this operational model and the variable slope sigmoid equation are as follows:

$$ \text{EC}_{50} = \frac{K_A}{(2 + \tau^n)^{1/n}} $$

$$ \text{E}_{\text{max}} = \frac{\text{Effect}_{\text{max}} \cdot \tau^n}{\tau^n + 1} $$

When $n$ equals 1, the equation is the same as those shown earlier, describing dose-response curves with Hill slopes of 1.0. However, $n$ is not the same as the Hill Slope (but the two values will be very close for full agonists).

**Designing experiments to fit to the operational model**

A single dose-response curve does not define both the affinity and efficacy of an agonist. If you try to fit the operational model equation to a single dose-response curve, you'll run into a problem. Either the curve-fitting program will report an error message, or it will report best-fit values with enormously wide confidence intervals.

Any symmetrical dose-response curve is defined by four parameters: $\text{Bottom}$ (response with no agonist), $\text{Top}$ (response at very high concentrations), $\text{EC}_{50}$ (concentration of agonist needed to provoke a response halfway between $\text{Bottom}$ and $\text{Top}$) and the $\text{Hill Slope}$. However, the operational model equation has five
parameters: Basal (response with no agonist), $K_A$ (dissociation constant of agonist binding), $E_{f_{max}}$ (maximum possible effect with a full agonist and plenty of receptors), $\tau$ (a measure of agonist efficacy), and $n$ (transducer slope).

Since the operational model has more parameters than are needed to describe a sigmoid dose-response curve, any curve can be defined by an infinite combination of operational model parameters. Even if a curve-fitting program could find best-fit values (rather than report an error message), the best-fit parameter estimates may not be correct.

To fit the operational model to data, therefore, you cannot analyze just a single dose-response curve. Instead you must fit a family of dose-response curves. Use one of these experimental approaches:

- One approach is to reduce the number of accessible receptors in a tissue or cell line to such an extent that a full agonist can no longer produce the maximal cellular response, no matter how high a concentration is used. A common method for reducing the number of functional receptors is to treat the tissue or cell line with a drug (e.g., alkylating agent) that binds irreversibly to the agonist binding site on the receptor, and thus permanently occludes that site. The agonist curve before alkylation is then compared to the curve after alkylation. This is the experimental method of choice for generating data that will allow affinity and efficacy estimates for drugs that are full agonists.

- A second approach that works only for partial agonist drugs is to directly compare the dose-response curve of a partial agonist with the dose-response curve of the full agonist. This method does not require receptor alkylation, but does require a known full agonist for the receptor of interest.

**Note:** All the dose-response curves should be obtained with the same tissue or cell line, in order to minimize variability in $E_{f_{max}}$ between preparations. This also applies to the use of recombinant expression systems (e.g., cell lines) with genetically engineered differences in receptor density; simultaneous analysis of curves obtained across different cell-lines will introduce between-tissue variability into the analysis, which can lead to problems with parameter estimation. In contrast, receptor depletion experiments using the same preparation of cells before and after treatment should only be subject to within-tissue variability.

---

**Fitting the operational model to find the affinity and efficacy of a full agonist**

**Theory of fitting receptor depletion data to the operational model**

To fit the operational model, one must account for data from two (or more) dose-response curves. For a full agonist, you must compare the dose-response curve in the absence or presence of receptor depletion. Experimentally, the key is to ensure that the receptors have been sufficiently alkylated such that the full agonist can no longer yield a maximal tissue response at saturating concentrations. These conditions are especially important to ensure that fitting the operational model to the data will yield a good estimate of the $E_{f_{max}}$ model parameter, which is crucial for successful estimation of the remaining model parameters. You can do this by globally fitting all the dose-response curves at one time, sharing model parameters across all the datasets.

**GraphPad Note:** Prism version 4 onwards offers the ability to globally share parameters across multiple datasets when fitting curves. This feature is also available in some other computer programs as well, but not in all of them.

Let’s first consider an experiment where the dose-response curve to a full agonist is determined in the absence or presence of progressive depletion of accessible receptor binding sites (this is also referred to as reducing “receptor reserve”). Because $\tau = [R_T]/K_E$, irreversibly occluding agonist binding will reduce $[R_T]$ and thus reduce the value of $\tau$. This will lower the maximal response and shift the dose-response curve to the right.

The operational model assumes that irreversibly occluding some receptors does not change the other three parameters. It assumes that the affinity of the agonist for the remaining receptors ($K_A$), the value of the transducer slope ($n$), and the value $E_{f_{max}}$ are properties of the tissue, not the drug, so have one value for all curves (note that $E_{f_{max}}$ refers to the maximum possible effect when no receptors are occluded, not the maximum effect attained in a particular dose-response curve). To fit the operational model, therefore, we want to globally fit all the data sets, sharing the value of $K_A$, $n$, and $E_{f_{max}}$ but finding separate best-fit values of $\tau$ for each dataset.

**Fitting receptor depletion data to the operational model with Prism**

Follow these steps:
1. Since concentrations are equally spaced on a log scale, enter data with \( X \) equal to the logarithm of the agonist concentration. Or transform your data to make \( X \) equal to the log of concentration if necessary.

2. Enter the operational model into your program. Here is one version:

\[
\text{operate} = \left(\frac{(10^{\log KA}) + (10^X)}{10^{(\log tau + X)}}\right)^n \]

\[
Y = \text{Basal} + \frac{\text{Effectmax} - \text{Basal}}{1 + 10^\text{operate}}
\]

3. Fit the model. For \( \log KA \), \( n \) and \( \text{Effectmax} \), make sure that you have told the program to share these values for all datasets. Leave the \( \log \tau \) to remain non-shared, as this will be unique for each curve. For the \( \text{Basal} \) parameter, either share it for all data sets or constrain it to a constant value of zero (depending on how you have normalized your data).

4. Consider the following recommendations for initial parameter values:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Initial Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{Effectmax} )</td>
<td>1 x maximum Y value for the full agonist curve in the absence of receptor depletion</td>
</tr>
<tr>
<td>( n )</td>
<td>Set to 1 (initial value to be fit)</td>
</tr>
<tr>
<td>( \text{Basal} )</td>
<td>1 x minimum Y value for the full agonist curve. (If there is no basal response in the absence of agonist, then set this value as a constant of zero, or omit it from the equation).</td>
</tr>
<tr>
<td>( \log KA )</td>
<td>1 x the X value corresponding to the response half way between the highest and lowest Y values for the agonist curve after receptor depletion.</td>
</tr>
<tr>
<td>( \log \tau )</td>
<td>Set to 0.0 (initial value to be fit). Since ( \log \tau ) starts at zero, this means that the initial value for ( \tau ) is 1.0. This value of ( \tau ) corresponds to a dose-response curve that plateaus at half ( \text{Effect}_{\text{max}} ), and usually results in successful convergence.</td>
</tr>
</tbody>
</table>

Why fit \( \text{Basal} \)? You may measure a "response" even in the absence of agonist. So include a basal parameter in the model. \( \text{Basal} \) is the measured response in the absence of agonist. If there is no basal activity, or if you have subtracted away basal before analyzing your data, then constrain \( \text{Basal} \) to a constant value of zero.

Example of fitting receptor depletion data to the operational model

In this example, we fit the operational model to two datasets. One data set is the response of human M\(_1\) muscarinic receptors, stably transfected into Chinese hamster ovary cells, to the agonist, acetylcholine in the absence of receptor alkylation, and the other shows the response to the same agonist in the same cells after receptor alkylation with the irreversible alkylating agent, phenoxybenzamine. The actual response being measured is agonist-mediated \(^{[3]}\text{H}\)phosphoinositide hydrolysis (A. Christopoulos, University of Melbourne, unpublished). Shown below are the actual data (d.p.m.):

<table>
<thead>
<tr>
<th>X Values</th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>516,033</td>
<td>240,733</td>
</tr>
<tr>
<td>100</td>
<td>945,069</td>
<td>422,503</td>
</tr>
<tr>
<td>200</td>
<td>2,165,260</td>
<td>121,000</td>
</tr>
<tr>
<td>400</td>
<td>5,000</td>
<td>326,456</td>
</tr>
<tr>
<td>500</td>
<td>7,320,167</td>
<td>527,486</td>
</tr>
<tr>
<td>1,000</td>
<td>11,777,600</td>
<td>746,035</td>
</tr>
<tr>
<td>1,500</td>
<td>14,437,000</td>
<td>325,456</td>
</tr>
<tr>
<td>1,750</td>
<td>16,722,300</td>
<td>591,350</td>
</tr>
<tr>
<td>2,000</td>
<td>17,751,800</td>
<td>496,638</td>
</tr>
<tr>
<td>2,500</td>
<td>19,606,000</td>
<td>613,033</td>
</tr>
</tbody>
</table>

Analysis of these data according to the operational model yielded the curve fits shown below:

Why fit the log of \( K_A \) and the log of \( \tau \)? When writing any model for data analysis, you should arrange the parameters so that the uncertainty is symmetrical and Gaussian. See "Why you should fit the logEC\(_{50}\) rather than EC\(_{50}\)" on page **. If you fit to the logarithm of \( K_A \) and \( \tau \), the uncertainty is more symmetrical (and more Gaussian) than it would be if you fit to \( K_A \) and \( \tau \) directly (see A. Christopoulos, Trends Pharmacol. Sci, 19:351-357, 1998).
The table below shows some of the output from the Results page of the analysis (using GraphPad Prism). Note that the best-fit value for logKA is close to -5, a concentration that gives a full response in control conditions, and is a log unit away from the logEC₅₀ of the control dose-response curve (which is close to -6).

<table>
<thead>
<tr>
<th>X Labels</th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 Operational Model of Agonism (Direct Fit)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 Best-fit values</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 logKA</td>
<td>-4.981</td>
<td>-4.981</td>
<td>-4.981</td>
</tr>
<tr>
<td>4 logTAU</td>
<td>1.024</td>
<td>-0.251</td>
<td>1.024</td>
</tr>
<tr>
<td>5 n</td>
<td>1.279</td>
<td>1.279</td>
<td>1.279</td>
</tr>
<tr>
<td>6 Basal</td>
<td>217.9</td>
<td>217.9</td>
<td>217.9</td>
</tr>
<tr>
<td>7 EMAX</td>
<td>15981</td>
<td>15981</td>
<td>15981</td>
</tr>
</tbody>
</table>

It is impossible to determine the logKA of a full agonist without inactivating receptors (see method above). However, for a full agonist (τ > 10), the Top and HillSlope parameters obtained from the standard sigmoid dose-response equation are very good approximations of the Effectmax and n parameters, respectively, of the operational model. This fact is exploited in the current method for obtaining operational model parameters for partial agonists. Specifically, the dose-response curve for the full agonist is fit to the standard sigmoidal dose-response equation, while the dose-response curves for the partial agonist(s) are fit to the operational model. The Top and HillSlope parameters of the full agonist curve are used by the operational model as Effectmax and n, respectively, when fitting the partial agonist curve.

GraphPad note: Prism 4 lets you share parameters across datasets, even when different datasets are fit to different equations. This feature is also available in some other computer programs, but not in all of them.

Fitting partial agonist data to the operational model with Prism

Follow these steps:

- Since concentrations are equally spaced on a log scale, enter data with X equal to the logarithm of the agonist concentration. Or transform your data to make X equal to the log of concentration.
- Enter the data for the full agonist into your first dataset column (e.g., column A), and the data for the partial agonist in column B.
- Choose nonlinear regression, and enter a user-defined equation. The following example is specific for GraphPad Prism. Depending on your particular program, you will probably need to modify it to suit your software.

Fitting the operational model to find the affinity and efficacy of a partial agonist

Theory of fitting partial agonist data to the operational model

A second application of the operational model is to obtain affinity and efficacy estimates for one or more partial agonists by comparing their responses to a full agonist in the same tissue. The analysis is different than fitting to the operational model after receptor inactivation. With receptor inactivation, the goal is to get a single value of logKA for the agonist. With partial agonists, we expect the get different values of logKA for each partial agonist used.
For Basal, n and Effectmax, choose to share the values for all datasets. Leave logtau, logKA, and logEC50 to be individually fitted.

Consider the following recommendations for initial parameter values:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Initial Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effectmax</td>
<td>1 x maximum Y value for the full agonist curve in the absence of receptor depletion</td>
</tr>
<tr>
<td>n</td>
<td>Set to 1 (initial value to be fit)</td>
</tr>
<tr>
<td>Basal</td>
<td>1 x minimum Y value for the full agonist curve. (If there is no basal response in the absence of agonist, then set this value as a constant of zero, or omit it from the equation).</td>
</tr>
<tr>
<td>logKA</td>
<td>1 x the X value corresponding to the response half way between the highest and lowest Y values for the partial agonist curve.</td>
</tr>
<tr>
<td>logTau</td>
<td>Set to 0.0 (initial value to be fit). Since logtau starts at zero, this means that the initial value for ( \tau ) is 1.0. This value of ( \tau ) corresponds to a dose-response curve that plateaus at half ( \text{Effect}_{\text{max}} ), and usually results in successful convergence.</td>
</tr>
<tr>
<td>logEC50</td>
<td>1 x the X value corresponding to the response half way between the highest and lowest Y values for the full agonist curve.</td>
</tr>
</tbody>
</table>

Example of fitting partial agonist data to the operational model

In this example, we wish to obtain affinity and efficacy estimates for the partial agonist, pilocarpine, by comparing its responses to those of the full agonist, oxotremorine-M, in Chinese hamster ovary cells transfected with the human M₃ muscarinic acetylcholine receptor (A. Christopoulos, University of Melbourne, unpublished). The response being measured is the same as that for the previous example. Note that the full agonist properties of oxotremorine-M were confirmed separately in receptor depletion experiments:
Dose-response curves in the presence of antagonists

Competitive antagonists

The term antagonist refers to any drug that will block, or partially block, a response. When investigating an antagonist, the first thing to check is whether the antagonism is surmountable by increasing the concentration of agonist. The next thing to ask is whether the antagonism is reversible. After washing away antagonist, does agonist regain response? If an antagonist is surmountable and reversible, it is likely to be competitive (see next paragraph). Investigations of antagonists that are not surmountable or reversible are beyond the scope of this manual.

A competitive antagonist binds reversibly to the same receptor as the agonist. A dose-response curve performed in the presence of a fixed concentration of antagonist will be shifted to the right, with the same maximum response and the same shape.

The dose ratio

Gaddum (J. Physiol. (Lond.), 89, 7P-9P, 1936) derived the equation that describes receptor occupancy by agonist in the presence of a competitive antagonist. The agonist is drug A, its concentration is \([A]\) and its dissociation constant is \(K_a\). The antagonist is called drug B, so its concentration is \([B]\) and dissociation constant is \(K_b\). If the two drugs compete for the same receptors, fractional occupancy by agonist \((f)\) equals:

\[
f = \frac{[A]}{[A] + K_a \left(1 + \frac{[B]}{K_b}\right)}
\]

In the above equation, the occupancy of agonist \([A]\) is determined by its \(K_a\). It can therefore be seen that the presence of a competitive antagonist multiplies the \(K_a\) value by a factor equal to \(1 + [B]/K_b\). In other words, the only effect of a simple competitive antagonist on an agonist is to shift the occupancy of the agonist by this constant factor; it has no other effects on the properties of the agonist. This theoretical expectation forms the basis of all currently used methods for quantifying agonist-antagonist interactions, which therefore rely on the determination of agonist dose-response curve shifts in the presence of antagonists.
Because a competitive antagonist does not alter the relationship between agonist occupancy and final response, it is unnecessary for you to know this relationship for the Gaddum equation above to be useful in analyzing dose-response curves. Thus, the equation can just as easily be written in terms of an agonist’s EC$_{50}$ value in the dose-response curve, rather than its $K_a$ value in the occupancy curve. You don’t have to know what fraction of the receptors is occupied at the EC$_{50}$ (and it doesn’t have to be 50%). The key to the usefulness of the equation is that whatever the initial agonist occupancy, you’ll get the same occupancy (and thus the same response) in the presence of antagonist when the agonist concentration is multiplied by $1+\frac{[B]}{K_b}$. Here is what the equation looks like when it is written in terms of the classic sigmoid dose-response curve relationship.

\[
\text{Response} = \text{Bottom} + \frac{(\text{Top-Bottom})^{\text{HillSlope}}}{1 + \left\{ \frac{\text{EC}_{50} \left[1 + \frac{[B]}{K_b}\right]}{[A]} \right\}^{\text{HillSlope}}}
\]

The graph below illustrates this relationship. If concentration A of agonist gives a certain response in the absence of competitive antagonist, but concentration A’ is needed to achieve the same response in the presence of a certain concentration of the antagonist, then A’/A represents the factor $1+\frac{[B]}{K_b}$. The ratio, A’/A, is called the “dose ratio”, and is most conveniently (although not exclusively) determined using EC$_{50}$ values. You’ll get a different dose ratio if you use a different concentration of antagonist, but the shift will always reflect the constant ($1+\frac{[B]}{K_b}$) if the interaction is truly competitive. Thus, if you know [B] and can determine the dose ratio, you should be able to derive a value for $K_b$.

The Schild slope factor

In theory, the principal goal behind measuring agonist dose-response curve shifts in the presence of competitive antagonists is to use the relationship embodied in the dose ratio to obtain an estimate of the antagonist’s $K_b$ value, i.e., its equilibrium dissociation constant. In practice, this can only be accurately derived when you determine the effects of more than one concentration of antagonist on the dose-response curve to agonist. If the interaction were truly competitive, then the shift of the agonist concentration-response curve in the presence of antagonist will always correspond to $1+\frac{[B]}{K_b}$, irrespective of the value of [B]; using different concentrations of B, therefore, allows you to check if the relationship holds. This procedure was first extensively developed by the pharmacologist, Heinz O. Schild (Arunlakshana and Schild, Br. J. Pharmac., 14, 48-57, 1959), and it is thus commonly associated with his name (i.e., “Schild analysis”).

In his studies, Schild also asked the question: what happens if the relationship between antagonist concentration and agonist dose-response curve shift doesn’t follow the factor $1+\frac{[B]}{K_b}$? For example, some non-competitive antagonists can shift agonist dose-response curves to the right without changing agonist maximal response, minimal response and slope, but the degree of the shift doesn’t follow the competitive relationship of $1+\frac{[B]}{K_b}$; in some cases the shift is greater than expected, whereas in others it is less than expected. Alternatively, an antagonist may be competitive, but the tissue or cellular preparation in which it is tested may contain more than one subtype of receptor with equal affinity and responsiveness to the agonist, but different affinities for the antagonist. Again, this latter situation may not result in changes of agonist dose-response curve shape in the presence of antagonist, but it can often result in
agonist curve shifts that do not follow \(1 + [B]/K_b\). In order to accommodate agonist curve shifts in the presence of antagonists that were either greater than or less than expected for simple competition for a single receptor, Schild modified Gaddum’s equation by introducing a slope factor, commonly referred to as the “Schild slope”:

\[
f = \frac{[A]}{[A] + K_b \left[ 1 + \frac{[B]^s}{K} \right]}
\]

In this equation, the antagonist term, \([B]\), is now raised to the power \(S\), where \(S\) denotes the Schild slope factor. Thus, if the antagonist shifts the agonist dose-response curve to the right in a parallel fashion, but greater than that predicted for simple competition, then the value of \(S\) will be greater than 1. In contrast, smaller than expected agonist curve shifts can be accommodated by a value of \(S\) less than 1. Notice that we have also changed the “\(K_b\)” parameter from the previous equation to a “\(K\)” in the above equation. This is because a \(K_b\) value, i.e., an antagonist equilibrium dissociation constant, cannot be derived from the above equation if \(S\) does not equal 1, so by convention, we shouldn’t call it the \(K_b\) in the above model. In practice, the \(K\) parameter should actually be estimated as a negative logarithm, so the equation can be re-written as follows:

\[
f = \frac{[A]}{[A] + K \left[ 1 + \frac{[B]^s}{10^{pK}} \right]}
\]

where \(pK\) is defined as the negative logarithm of \(K\). Hence, the parameter, \(pK\), represents a simple fitting constant that has no mechanistic meaning except when \(S\) = 1, in which case \(pK = pK_b\).

In Schild analysis, therefore, the determination of agonist dose-response curve shifts in the presence of different concentrations of antagonist allows you to first assess the conformity of the data to a model of simple competition, by determining whether the Schild slope is significantly different from 1 or not, and then quantify the antagonism (if \(S\) = 1) by determining the \(pK_b\) value (the dissociation constant of the antagonist binding).

**\(pK_b\) vs \(pA_2\)**

By convention, a \(pK_b\) value can only be derived when \(S\) = 1; under this circumstance, the data are deemed to be consistent with a simple mechanism of one-to-one competition between agonist and antagonist for the receptor, and the \(pK_b\) is thus a mechanistic estimate of the negative logarithm of the antagonist’s equilibrium dissociation constant. In practice, this is done by first fitting the Gaddum/Schild model to experimental data in order to obtain the estimate of \(S\), and then performing a statistical test to determine whether this estimate of \(S\) is different from a value of 1. If \(S\) is not significantly different from 1, then the equation is re-fitted to the data with \(S\) fixed as a constant value of 1, and the resulting estimate of \(pK\) is the \(pK_b\) value.

What happens if \(S\) is significantly different from 1? In this case, the resulting estimate of \(pK\) is not the \(pK_b\), and cannot be quoted as such. It is not the negative logarithm of the dissociation constant of antagonist binding. You will have to conclude that your experimental data are not consistent with a model of simple competition between agonist and antagonist. Nevertheless, you may still wish to quote an empirical estimate of the potency of your antagonist for the sake of comparison with other drugs. By convention, the most common estimate of antagonist potency that is independent of any particular mechanism is the “\(pA_2\) value”. The \(pA_2\) is defined as the negative logarithm of the concentration of antagonist required to cause a 2-fold rightward shift of the agonist dose-response curve. It can readily be seen that for a competitive antagonist, \(pA_2 = pK_b\), because a \(K_b\) concentration of a competitive antagonist will shift an agonist’s curve by a factor of two \((1 + [B]/K_b = 2\) when \([B]=K_b)\). For a non-competitive antagonist, a \(pA_2\) value is not a \(pK_b\), but simply a measure of the potency of the antagonist to shift the curve of an agonist to the right by a factor of two. Historically, the \(pA_2\) was determined from the x-intercept of the Schild plot (see page XX), when \(S\) was not fixed to a value of 1, but it can easily be calculated from the following relationship between Gaddum/Schild model parameters:

\[
pA_2 = \frac{pK}{S}
\]

Thus, even if an antagonist is not competitive, a \(pA_2\) value can be quoted as an empirical estimate of the antagonist’s potency.

**An alternative to the classic Schild slope factor**

Although the Gaddum/Schild equation in its original form (see above) is still the most commonly used model for fitting agonist-antagonist dose-response data, we and others have noted that the two most relevant parameters of interest, the slope, \(S\), and the \(pK\), are highly correlated with one another. That is, when the nonlinear regression algorithm changes one parameter while trying to find its best-fit value, the other parameter also changes to try and compensate. Parameters are always somewhat correlated, but these two are especially correlated, making the results less adequate. We need to find a way to minimize this problem, while still allowing for the derivation of appropriate estimates of the Schild slope and \(pK\). One modification of the Gaddum/Schild equation that overcomes this problem is shown below (see D. R. Waud et al., Life Sci., 22, 1275-1286, 1978; Lazareno and Birdsall, Br. J. Pharmac., 109, 1110-1119, 1993):
It can be seen that the difference between this modified equation and the original Gaddum/Schild equation is that the entire $[B]/K$ term is now raised to the power $S$, rather than just the $[B]$ term in the original equation.

What effect does this have on the model and its ability to fit agonist-antagonist interaction data? We have performed many simulations to investigate the properties of the modified Gaddum/Schild model, and have found the following results. First, the value of $S$ is the same if we use the modified equation compared to the original form of the equation, so the $S$ parameter can still be quoted as an estimate of the Schild slope. Second, and most importantly, the parameters $S$ and $pK$ are far less correlated in the modified equation. Third, if the value of $S$ is significantly different from 1 in the modified equation, then the estimate of the $pK$ is not valid as an estimate of the antagonist’s $pK_a$, but it is a valid estimate of the $pA_2$. In contrast, if the original form of the Gaddum/Schild equation were to be used, then the estimate of $pK$ when the value of $S$ is not 1 is meaningless; it cannot be used as an estimate of the $pA_2$ unless it was first divided by the value of $S$ (see above). Obviously, it is far better to use a model that allows separate estimates of $S$ and $pA_2$ to be obtained directly from the curve-fitting process, rather than having to indirectly calculate a $pA_2$ value from two previously estimated parameters that are each associated with their own standard errors. Based on these findings, therefore, we can rewrite the modified Gaddum/Schild model as follows:

$$ f = \frac{[A]}{[A]+K_s \left[1+\left(\frac{[B]}{10^{pK}}\right)^S\right]} $$

The remaining sections of this chapter describe analyses based on this modified Gaddum/Schild model of agonist-antagonist interactions.
The following syntax can be used to program the model:

\[
\text{EC50} = 10^{\text{LogEC50}}
\]

\[
\text{Antag} = 1 + \left( B / (10^{-1 \cdot pA2}) \right)^{\text{SchildSlope}}
\]

\[
\text{LogEC} = \text{Log} (\text{EC50} \cdot \text{Antag})
\]

\[
Y = \text{Bottom} + \left( \text{Top} - \text{Bottom} \right)/(1+10^{(\text{LogEC} - X) \cdot \text{HillSlope}})
\]

Here are some guides to the choice of Initial Values for the analysis:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Initial Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bottom</td>
<td>1 x minimum Y value for the agonist curves. (If there is no basal response in the absence of agonist, then set this value as a constant of zero, or omit it from the equation).</td>
</tr>
<tr>
<td>Top</td>
<td>1 x maximum Y value for the agonist curves.</td>
</tr>
<tr>
<td>LogEC50</td>
<td>1 x the X value corresponding to the response half way between the highest and lowest Y values for the full agonist curve.</td>
</tr>
<tr>
<td>HillSlope</td>
<td>Set to 1.</td>
</tr>
<tr>
<td>SchildSlope</td>
<td>Set to 1.</td>
</tr>
<tr>
<td>pA2</td>
<td>Enter a rough estimate of the negative logarithm of the antagonist’s K_b based on the minimum concentration of antagonist that you observed experimentally to cause a discernible shift in the control agonist dose-response curve.</td>
</tr>
</tbody>
</table>

Ideally, the most appropriate approach for analyzing the entire family of dose-response curves is to fit them to two versions of the above equation, one where the SchildSlope parameter is set as a constant equal to 1, and the other where it is a shared value for all datasets, and then compare the two different forms of the equation using the F-test. If the simpler model (Schild slope=1) is the better fit, then estimate the pA2 is in fact the pK_b, and may be quoted as such. If the equation where the Schild slope does not equal 1 is the better fit, then the estimate of pA2 is not the pK_b. Shown below are the results of this analysis for the interaction between acetylcholine and N-methylscopolamine at the M_1 muscarinic receptor, based on the table above. Also shown are selected screen shots from the GraphPad Prism Results page, where it can be seen that the equation in which the Schild slope has been set as a constant of 1 (Equation 1) is a better fit of the data, yielding a pA_2 = pK_b value for N-methylscopolamine of 9.68 (K_b = 0.21 nM).
Although the procedure described in the previous section is the preferred method for analyzing agonist-antagonist interaction data, there are some situations where you are not able to use this method, or simply don’t need the level of rigor associated with it. For instance, you may be doing experiments where only a few agonist concentrations are tested, such that only the linear portion of the sigmoid dose-response curve (on a logarithmic scale) is determined in the absence and presence of each antagonist concentration. Alternatively, you may be using a nonlinear regression program that doesn’t allow you to use global parameter-sharing and/or two independent variables. In these instances, you can’t fit the complete sigmoid model presented above, but you can still determine equieffective agonist concentrations, perhaps as EC\textsubscript{50} values. Lew and Angus (Trends Pharmacol. Sci., 16:328-337, 1995) have presented a simple method for analyzing agonist-antagonist interactions using nonlinear regression of agonist EC\textsubscript{50} values obtained in the absence or presence of antagonist.

Start with the Gaddum equation for occupancy as a function of agonist and antagonist concentrations:

\[
f = \frac{[A]}{[A] + K_a + \frac{B + K_a}{K_b}}
\]

Simple algebra expresses the equation this way:

\[
f = \frac{1}{1 + \frac{[B] + K_a}{K_b}}
\]

Thus you can obtain any particular occupancy \(f\) with any concentration of antagonist \([B]\) so long as you adjust \(A\) to keep the quantity in the parentheses constant \((C)\).

\[
\frac{[B] + K_a}{[A]} = C
\]

Rearrange to show how you must change the agonist concentration to have the same response in the presence of an antagonist.

\[
\frac{[B] + K_a}{C} = [A]
\]

The EC\textsubscript{50} is the concentration needed to obtain 50% of the maximal response. You don’t know the fraction of receptors occupied at that concentration of agonist, but you can assume that the same fractional occupancy by agonist leads to the same response regardless of the presence of antagonist. So you can express the equation above to define EC\textsubscript{50} as a function of the antagonist concentration \([B]\).

\[
EC_{50} = \frac{[B] + K_a}{C}
\]

You determined the EC\textsubscript{50} at several concentrations of antagonist (including 0), so you could fit this equation to your data to determine a best-fit value of \(K_b\) (and \(C\), which you don’t really care about). But it is better to write the equation in terms of the logarithm of EC\textsubscript{50}, because the uncertainty is more symmetrical on a log scale. See “Why you should fit the logEC\textsubscript{50} rather than EC\textsubscript{50}” on page \(X\). By tradition, we use the negative logarithm of EC\textsubscript{50}, called the pEC\textsubscript{50}. For similar reasons, you want to determine the best-fit value of log \(K_b\) (logarithm of the dissociation constant of the antagonist) rather than \(K_b\) itself.

\[
pEC_{50} = \log \left(\frac{[B] + 10^{-pK_b}}{C}\right)
\]

Define \(Y\) to be the pEC\textsubscript{50}, \(X\) to be the antagonist concentration \([B]\), and a new constant \(P\) to be log \(C\). Now you have an equation you can use to fit data:

\[
Y = \log \left(\frac{X}{10^{-pK_b}}\right) - P
\]
**Determining the $K_b$ using nonlinear regression of agonist $pEC_{50}$ values**

1. Determine the $EC_{50}$ of the antagonist in the presence of several concentrations of antagonist, including zero concentration. Enter these values into a data table as follows: Into the X column, enter the antagonist concentrations in micromolar. Into the Y column, enter the negative logarithm of the $EC_{50}$ values.

2. Use nonlinear regression to fit this equation.

$$Y = -1 \times \log (X + (10^{-1 \times pK_b}))) - P$$

**Dealing with Schild slopes that do not equal 1**

Here is how the equation above can be re-cast to find the Schild slope and the $pA_2$ using nonlinear regression analysis of $EC_{50}$ values. The equation is again based on the modified Gaddum/Schild model presented earlier in this chapter:

$$pEC_{50} = -\log [B^f + 10^{pA_2 \times S}] - \log(c)$$

Note that the parameter $pK_b$ in the original Lew and Angus equation has been replaced with $pA_2$ in the above equation. This is because if the value for $S$ is significantly different from 1, then the antagonist fitting parameter is *not* the $pK_b$, but will be the $pA_2$.

Enter the following user-defined equation:

$$K = -1 \times pA_2$$

$$Y = -1 \times \log (X^S + (10^{K \times S})) - P$$

When performing this analysis, it is a good idea to fit the data to both equations at the same time and use the F-test to decide which one is the more appropriate equation. If the simpler equation is the better equation, then the $pK_b$ estimate may be quoted. Otherwise, you must conclude that your data are not consistent with a model of simple competition; you can still quote the $pA_2$, however, as an empirical estimate of antagonist potency.

**The Schild plot**

The oldest method for analyzing agonist-antagonist interactions from functional experiments is the original linear regression method developed by Schild. This method relies explicitly on the determination of agonist dose ratios in the absence and presence of antagonist. If you perform experiments with several concentrations of antagonist, you can create a graph with $\log(\text{Antagonist})$ on the X-axis and $\log(\text{Dose Ratio} - 1)$ on the Y-axis; this is commonly referred to as the Schild Plot. If the antagonist is competitive, you expect a slope of 1.0 and an X-intercept of $\log K_b$ for the antagonist.

In comparison to the nonlinear regression methods outlined above, the linear regression method of the Schild plot is potentially flawed. The problem is that the $EC_{50}$ of the control agonist dose-response curve is used to compute dose ratios for all other curves. Any error in that control value shows up in all the data points. The Schild plot was developed in an era when nonlinear regression methods were not available.
was unavailable, so it was necessary to transform data to a linear form. This is no longer an advantage, and Schild plots can be thought of in the same category as Scatchard plots. That is, they are useful for graphical representations of agonist-antagonist interaction data, but for analytical purposes the nonlinear regression methods outlined above are superior.

**Antagonist inhibition curves**

There are often instances where complete agonist dose-response curves in the absence or presence of antagonist cannot be readily determined to fully define the effects of the antagonist over more than one or two orders of magnitude of antagonist concentrations. For example, there may be solubility problems with the agonist, or it may only be available in very small quantities such that large concentrations cannot be prepared, or it may rapidly desensitize the preparation when used at high, but not low, concentrations. These practical difficulties with the agonist, in turn, limit the investigator’s ability to accurately discriminate whether the antagonist is competitive or non-competitive, because non-competitive antagonists may appear competitive when tested at low concentrations, but reveal their non-competitive nature when tested at high concentrations.

One approach to overcoming these limitations that has become increasingly popular is to test the effects of increasing, graded, concentrations of antagonist on a single, fixed, concentration of agonist. This kind of experimental design is referred to as the “antagonist inhibition curve” design, and can readily test the effects of antagonist concentrations that span many orders of magnitude. This method is particularly widespread in the measurement of biochemical responses using cell-based or tissue extract-based assays. Shown below is an example of an agonist dose-response curve as well as the corresponding antagonist inhibition curve determined in the presence of a fixed agonist concentration (3 x 10^-8 M) that produces the response denoted by the dotted line.

It can be seen that the shape of the antagonist inhibition curve appears similar to that of an antagonist competition binding curve obtained from a standard radioligand binding assay. Indeed, the concentration of antagonist that reduces the initial level of agonist response by 50% is usually called the IC_{50}, just like the concentration of antagonist that reduces specific radioligand binding by 50% in a binding assay. However, this is where the similarities end. Although it is relatively straightforward to obtain the LogK_{d} of a competitive antagonist from a competition binding assay using the IC_{50} and the Cheng-Prusoff equation, you generally cannot obtain an equivalent estimate of LogK_{b} from a functional antagonist inhibition curve using the same method (see Leff and Dougall, Trends Pharmacol. Sci., 14, 110-112, 1993). This is because the shape of the antagonist inhibition curve in a functional assay is dependent on the shape of the agonist dose-response curve. If an agonist produces steep dose-response curves in a given tissue or cell line, then the resulting antagonist inhibition curve will be very different than if the agonist produces shallow curves, or curves with a slope of 1.

In order to properly analyze functional antagonist inhibition curve experiments, you need to include information about the control agonist dose-response curve in the analysis. The appropriate experimental design requires that you construct a control agonist curve and the antagonist inhibition curve in the same tissue or cell line. You can then analyze your data as follows:

1. Enter your agonist dose-response data into the first column of a new Data table.
2. Enter your antagonist inhibition curve data into the second column of your Data table.
3. Analyze your data according to the following user-defined equation. The syntax is specific for GraphPad Prism, and will need to be slightly modified if you are using a different program:

\[
\text{Control} = \text{Bottom} + \frac{(\text{Top} - \text{Bottom})}{(1 + 10^{((\text{LogEC50} - X) \times \text{HillSlope})})}
\]

\[
\text{Antag} = (10^{\text{LogEC50}}) \times \frac{(1 + ((10^{X}) / (10^{(-1 \times pA2)})))^{\text{SchildSlope}}}{\text{SchildSlope}}
\]

\[
\text{WithAntag} = \text{Bottom} + \frac{(\text{Top} - \text{Bottom})}{(1 + (\text{Antag} / \text{FixedAg})^{\text{HillSlope}})}
\]

\(<A>Y = \text{Control}\)

\(<B>Y = \text{WithAntag}\)

**GraphPad Note:** The second last line in the equation is preceded by \(<A>\) so it only applies to data set A. It is a standard sigmoidal dose-response curve. The last line is preceded by \(<B>\) so it applies to data set B. In this equation, therefore, it matters which dataset you enter into Column A, and which dataset is entered into Column B, so make sure that the control agonist dose-response data go into the first column and the antagonist inhibition curve data go into the second column.

For the analysis, you will need to globally share the values of all the parameters across all datasets, except for the parameter, \(\text{FixedAg}\), which represents the initial fixed concentration of agonist used in the determination of the antagonist inhibition curve. Choose to set \(\text{FixedAg}\) as a constant value equal to the fixed agonist concentration (Molar) used in your antagonist inhibition curve assay; for the above example, this value would be set as 3e-8. The nonlinear regression algorithm will then work its way through the equation. The desired parameters determined by the algorithm will then reflect the best-fit values that describe both agonist and antagonist curves. Because the Schild slope=1 in our example, the estimate of \(pA2\) is the \(pK_b\). As with the previous examples, however, you should also fit this model with the Schild slope fixed to 1 and compare it using the F test with the model where the Schild slope is shared (but allowed to vary) by both datasets.

Shown below is the same example dataset from above fitted to the model, as well as some of the output from the GraphPad Prism Results page.

An important consideration with fitting antagonist inhibition data is with respect to your initial parameter values. Because this is a relatively complicated model, it is probably best to enter the initial values for each dataset manually, based on reasonable first guesses. For example, the \(\text{HillSlope}\) and \(\text{SchildSlope}\) values can each be initially set to 1. The \(\text{LogEC50}\) and \(pA2\) parameters can be assigned the X values corresponding to the approximate midpoints of the control agonist curve and the antagonist inhibition curve, respectively. Note that the \(\text{Top}\) and \(\text{Bottom}\) parameters must be assigned according to the basal and maximal responses, respectively, of the control agonist dose-response curve.