Rogozina et al. Supplementary Materials.

Data analysis:

Residuals from nonlinear regression.

Nonlinear regression finds parameters that make a model fit the data as closely as possible (given some assumptions). It does not automatically ask whether another model might work better. The graphical residuals analysis is one of the key tools for model validation.

Residuals are the differences between the observed and predicted responses. In other words, residuals are the vertical distances of each corresponding data point from the fit curve. Residuals are positive when the points are above the curve, and are negative when the points are below the curve. By carefully looking at residuals can tell whether the assumptions are reasonable and the choice of model is appropriate. The general assumption applied to the group of residuals is that one expects them to be independently distributed around the fit curve (Supplementary Fig.4). (Motulsky, H. J. and Christopoulos, A. (2003) Fitting Models to Biological Data Using Linear and Nonlinear Regression: A Practical Guide to Curve Fitting. GraphPad Software Inc., San Diego CA.)

Extra sum-of-squares F test.

The goal of nonlinear regression is to find parameter values that make the curve come near the data points or, in other words, to minimize the sum of squares of the vertical distances of the data points from the curve. So it would appear that the model which fits the data with the smallest sum-of-squares is the best. In fact, that approach is too simple. The problem is that a more complicated model (more parameters) can fit the data better just because it can have more inflection points. For instance, a two-phase model almost always fits the data better than a one-phase model. Therefore any method used to compare a simple model with a more complicated model has to balance the decrease in sum-of-squares with the increase in the number of parameters.

The F test (extra sum-of-squares) is one of the key statistical approaches used to compare related models (two models are related when one is a simpler case of the other) (Motulsky, H. J. and Christopoulos, A. (2003) Fitting Models to Biological Data Using Linear and Nonlinear Regression: A Practical Guide to Curve Fitting. GraphPad Software Inc., San Diego CA.)

. It is based on the difference between the sum-of-squares of the two models and also takes into account the number of data points and the number of parameters of each model.

If the simpler model (fewer parameters) is correct, then the relative increase in the sum-of-squares (going from complicated to simple model) would be approximately equal the relative increase in degrees of freedom (DF), which equals the number of data points minus the number of parameters. If the more complicated model is correct, then the relative increase in the sum-of-squares would be greater than the relative increase in degrees of freedom.

The F ratio equals the relative difference in the sum-of-squares divided by the relative difference in degrees of freedom.

$$F = \frac{(SS1-SS2)/SS2}{(DF1-DF2)/DF2}$$

This equation is more commonly shown in the following equivalent form:

$$F = \frac{(SS1-SS2)/(DF1-DF2)}{SS2/DF2}$$

where the numbers 1 and 2 refer to the simple (single exponential) and more complex (double exponential) models, respectively. (DF1-DF2) is the degrees of freedom for the numerator (DFn), and DF2 is the degrees of freedom for the denominator (DFd).

If the simpler model is correct, then an F ratio is expected to be near 1.0. If the F ratio is much greater than 1.0, then there are two possibilities:

- The more complicated model is correct.
- The simpler model is correct, but random scatter in the data led the more complicated model to fit better.

The P value, which can be calculated from the F ratio and the two DF values, tells how frequently the second possibility would happen. P is the probability (ranging from 0 to 1) that the results observed in study could have occurred by chance. If the P value is low (0.05 or below), then one can conclude that the more complicated model is significantly better than the simple model. For the results of the F-test analysis of the fits see Supplementary Table 2.