SUPPORTING INFORMATION

Quantitative analysis of ligand induced hetero-dimerization of two distinct receptors

Chang Lu[1] and Zhi-Xin Wang[1, *]

1. Key Laboratory of Ministry of Education for Protein Science, School of Life Sciences, Tsinghua University Beijing 100084 P.R. China

[*] Corresponding author.

E-mail: zhixinwang@mail.tsinghua.edu.cn

Phone: 86-10-62785505

FAX: 86-10-62792826

Contents

Section I. Theory

Time course Model

Consider a mixture of free proteins P_1 and P_2, with total concentrations $[P_1]_0$ and $[P_2]_0$ respectively. When hetero-bifunctional ligand L is added to the mixture, the reactions described in Scheme S1 (or Figure 1B) would occur.

$$P_1 + L \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} P_1L$$
$$L + P_2 \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} LP_2$$
$$P_1 + LP_2 \underset{k_{-3}}{\overset{k_3}{\rightleftharpoons}} P_1LP_2$$
$$P_1L + P_2 \underset{k_{-4}}{\overset{k_4}{\rightleftharpoons}} P_1LP_2$$

Scheme S1

Let $[P_1], [P_2]$ be the concentrations of free proteins P_1 and P_2, $[P_1L], [LP_2]$ *and* $[P_1LP_2]$ be concentrations of ligand-bound form and hertero-dimerized ternary complex form of proteins, and $[L]$ be the free ligand concentration. The differential equations describing the model in Scheme S1 are:

$$\begin{aligned}
\frac{d[L]}{dt} &= k_{-1}[P_1L] + k_{-2}[LP_2] - k_1[P_1][L] - \mathrm{k}_2[\mathrm{P}_2][L] \\
\frac{d[\mathrm{P}_1L]}{dt} &= k_1[P_1][L] + \mathrm{k}_{-4}[P_1LP_2] - k_{-1}[P_1L] - k_4[\mathrm{P}_1L][P_2] \\
\frac{d[LP_2]}{dt} &= k_2[P_2][L] + \mathrm{k}_{-3}[P_1LP_2] - k_{-2}[LP_2] - k_3[\mathrm{P}_1][\mathrm{L}P_2] \\
\frac{d[\mathrm{P}_1LP_2]}{dt} &= k_3[P_1][LP_2] + \mathrm{k}_4[P_1L][P_2] - (k_{-3} + k_{-4})[P_1LP_2]
\end{aligned} \tag{S1}$$

And law of mass conservation holds:

$$\begin{aligned}
[P_1]_0 &= [P_1] + [P_1L] + [P_1LP_2] \\
[P_2]_0 &= [P_2] + [LP_2] + [P_1LP_2] \\
[L]_0 &= [L] + [P_1L] + [LP_2] + [P_1LP_2]
\end{aligned} \tag{S2}$$

The above equations can be combined to form a set of time course model if we start with two proteins and no ligand, in other words, at t=0 [L]=[P_1L]=[LP_2]=0. As the four reactions forms a circle, detailed balance principle applies to the model and $k_1[P_1]k_4[P_2]k_{-3}k_{-2} = k_2[P_2]\mathrm{k}_3[P_1]k_{-4}k_{-1}$, which can be simplified to $\frac{k_{-3}k_{-2}}{k_2\,\mathrm{k}_3} = \frac{k_{-4}k_{-1}}{k_1k_4}$.

Equilibrium state Model

In accordance with the principle of detailed balance, when the system described by Scheme S1 and eqns. S1 reaches steady state and the concentration of each species in the mixture are approximately constant; it can be also derived that every step of the reactions in Scheme S1 is also in equilibrium, or in other words eqns. S2 hold.

$$\begin{aligned} k_{-1}[P_1L] &= k_1[P_1][L] \\ k_{-2}[LP_2] &= k_2[P_2][L] \\ k_{-3}[P_1LP_2] &= k_3[P_1][LP_2] \\ k_{-4}[P_1LP_2] &= k_4[P_1L][P_2] \end{aligned} \tag{S3}$$

Define dissociation constants $K_1 = {}^{k_{-1}}\!/_{k_1}, K_2 = {}^{k_{-2}}\!/_{k_2}, K_3 = {}^{k_{-3}}\!/_{k_3}, K_4 = {}^{k_{-4}}\!/_{k_4}$, which satisfies the relationship: $K_1K_4 = K_2K_3$. The cooperativity factor is defined as $\alpha = {}^{K_1}\!/_{K_3} = {}^{K_2}\!/_{K_4}$. We are interested the dependence of equilibrium concentrations on the initial (or total) concentrations of proteins and ligand ($[P_1]_0$, $[P_2]_0$, $[\mathrm{L}]_0$) and kinetic constants K_1, K_2 and α. Equations (S4)-(S7) characterize the kinetic equilibrium, and equations (S8)-(S10) account for the mass balance of the proteins and the dimerization-inducing ligand respectively.

$$[P_1][L] = K_1[P_1L] \tag{S4}$$

$$[L][P_2] = K_2[LP_2] \tag{S5}$$

$$[P_1][LP_2] = K_3[P_1LP_2] = {}^{K_1}\!/_{\alpha}[P_1LP_2] \tag{S6}$$

$$[P_1L][P_2] = K_4[P_1LP_2] = {}^{K_2}\!/_{\alpha}[P_1LP_2] \tag{S7}$$

$$[P_1]_0 = [P_1] + [P_1L] + [P_1LP_2] \tag{S8}$$

$$[P_2]_0 = [P_2] + [LP_2] + [P_1LP_2] \tag{S9}$$

$$[L]_0 = [L] + [P_1L] + [LP_2] + [P_1LP_2] \tag{S10}$$

Note that one of the first four equations is redundant. The rest six non-redundant equations constitute a set of quadratic equations of six variables, namely the equilibrium concentrations $[P_1], [P_2], [P_1L], [LP_2], [L]$ and $[P_1LP_2]$, given total concentrations $[L]_0, [P_1]_0, [P_2]_0$ and parameters K_1, K_2, α.

Generally speaking, there are two possible approaches to derive the equilibrium state determined by a set of chemical reactions. The first one is to solve the set of differential equations, which are derived from chemical reaction schemes by law of mass action (e.g. eqns.(S1)), as a numerical *initial value problem* with, for instance Runge-Kutta method. The complexity of such method is $O(p^N)$, where p is the chosen step-size and N is the number of equations involved. This is a general approach applicable for deriving steady state expressions by computationally simulating the chemical reactions, but is limited in its ability to provide analytical results owing to the difficulty in sampling a representative set of parameter space. The other one is to solve the system of multivariate quadratic (MQ) equations (e.g. eqns. (S4) to (S10)). However, solving systems of MQ equations is hard in general, and its associated MQ-problem is known to be NP-complete [1]. With a small number of variables, in this case six, the equations can be numerically solved using MATLAB or other software.

Here, for the specific problem of ligand-induced hetero-dimerization, MQ equations (S4) to (S10) can be combined into a fifth-order polynomial equation. Solving roots of single variate polynomial is equivalent to finding the eigenvalues of the associated companion matrix, which can be solved with generalized schur decomposition (QZ) algorithm with complexity $O(N^3)$, N being the order of polynomial [2]. This approach reduces the time-complexity in computing steady state, and is conducive to providing an efficient least-square fitting algorithm to experimental data. In the next section, we combine the six-variable quadratic equations (S4) to (S10) to a quintic equation of $[L]$ and provide the analytical expressions of the other five variables in the form of $[L]$.

Derivation of the Binding Polynomial.

Eqs(S4)-(S5), (S8)-(S10) can be arranged to yield expression of equilibrium $[P_1]$, $[P_2]$, $[P_1L]$, $[LP_2]$ and $[P_1LP_2]$ as a function of free inducer concentration L, total inducer concentration $[L]_0$, total protein concentrations $[P_1]_0$, $[P_2]_0$, and equilibrium parameters K_1, K_2 and cooperativity α.

$$
\begin{aligned}
[P_1] &= \frac{K_1}{K_1K_2-[L]^2}\left[[L]([L]-[L]_0+[P_2]_0)+K_2([L]-[L_0]+[P_1]_0)\right]\\
[P_2] &= \frac{K_2}{K_1K_2-[L]^2}\left[[L]([L]-[L]_0+[P_1]_0)+K_1([L]-[L_0]+[P_2]_0)\right]\\
[P_1L] &= \frac{[L]}{K_1K_2-[L]^2}\left[[L]([L]-[L]_0+[P_2]_0)+K_2([L]-[L_0]+[P_1]_0)\right]\\
[LP_2] &= \frac{[L]}{K_1K_2-[L]^2}\left[[L]([L]-[L]_0+[P_1]_0)+K_1([L]-[L_0]+[P_2]_0)\right]\\
[P_1LP_2] &= [P_2]_0-\frac{K_2+[L]}{K_1K_2-[L]^2}\left[[L]([L]-[L]_0+[P_1]_0)+K_1([L]-[L]_0+[P_2]_0)\right]
\end{aligned}
\qquad \text{(S11)}
$$

We next substitute expressions of $[P_1L], [P_2]$ and $[P_1LP_2]$ from eqn.(S11) into Eq. (S6), and rearrange it to yield a quintic equation of $[L]$ (Eq.(S12) with coefficients written out in full). Hence the equilibrium concentrations of $[P_1]$, $[P_2]$, $[P_1L]$, $[LP_2]$ and $[P_1LP_2]$ can be derived using expressions in equations(S11).

$$
\begin{aligned}
&\tfrac{1}{\alpha}(K_1K_2-[L]^2)\{(K_1+[L])(K_2+[L])([L]-[L]_0)+[L]((K_2+[L])[P_1]_0+(K_1+[L])[P_2]_0)\}\\
&+[L]\{[L]([L]-[L]_0+[P_1]_0)+K_1([L]-[L]_0+[P_2]_0)\}\{[L]([L]-[L]_0+[P_2]_0)+K_2([L]-[L]_0+[P_1]_0)\}=0
\end{aligned}
\quad \text{or}
$$

$$
a_0[L]^5+a_1[L]^4+a_2[L]^3+a_3[L]^2+a_4[L]+a_5=0 \qquad \text{(S12)}
$$

Where

$$
\begin{aligned}
a_0 &= 1-\tfrac{1}{\alpha}\\
a_1 &= -[L]_0+(1-\tfrac{1}{\alpha})(K_1+K_2-[L]_0+[P_1]_0+[P_2]_0)\\
a_2 &= (K_1-[L]_0+[P_1]_0)(K_2-[L]_0+[P_2]_0)-(1-\tfrac{1}{\alpha})[K_2([L]_0-[P_1]_0)+K_1([L]_0-[P_2]_0)]\\
a_3 &= K_2(K_1-[L]_0+[P_1]_0)(\tfrac{K_1}{\alpha}-[L]_0+[P_1]_0)+K_1(K_2-[L]_0+[P_2]_0)(\tfrac{K_2}{\alpha}-[L]_0+[P_2]_0)+2\tfrac{K_1K_2}{\alpha}[L]_0\\
a_4 &= K_1K_2(\tfrac{K_1}{\alpha}-[L]_0+[P_1]_0)(\tfrac{K_2}{\alpha}-[L]_0+[P_2]_0)+(1-\tfrac{1}{\alpha})\tfrac{1}{\alpha}K_1^2K_2^2\\
a_5 &= -\tfrac{1}{\alpha}K_1^2K_2^2[L]_0
\end{aligned}
$$

Maximum of induced ternary complex during titration of inducer.

Some of the following results (eqs. S19, S20) have been previously obtained by Douglass et. al.[3], here we provide derivation process for the completeness of the paper.

Derivation-Part 1.

In this section, we discuss the situation when the concentration of two proteins and inducer complex $[P_1LP_2]$ reaches its maximum. Theoretically, the amount of total inducer necessary to produce the maximum amount of $[P_1LP_2]$ can be derived with Eq.(S13).

$$\frac{\partial[P_1LP_2]}{\partial[L]_0}=0 \tag{S13}$$

However, this expression cannot be easily obtained. To simplify the calculation, we tackle the problem through the observation that the concentration of inducer-bound proteins (e.g. $[P_1L],[LP_2]$) would monotonically increase when the amount of inducer is increased ($\frac{\partial[P_1L]}{\partial[L]_0},\frac{\partial[LP_2]}{\partial[L]_0}>0$). Because of the chain-rule, namely $\frac{\partial[P_1LP_2]}{\partial[L]_0}=\frac{\partial[P_1LP_2]}{\partial[P_1L]}\frac{\partial[P_1L]}{\partial[L]_0}$, and solving Eq.(S13) is equivalent to solving Eq.(S14).

$$\frac{\partial[P_1LP_2]}{\partial[P_1L]}=0 \tag{S14}$$

Now we are going to derive an implicit expression of the form $f([P_1LP_2],[P_1L])=0$ to solve Eq.(S14).

Substituting $[P_2]$ and $[LP_2]$ in Eq.(S9) with expressions:

$[P_2]=\frac{K_2/\alpha[P_1LP_2]}{[P_1L]}$ (derived from Eq.(S7))

and $[LP_2]=\frac{K_1/\alpha[P_1LP_2]}{[P_1]}=\frac{K_1/\alpha[P_1LP_2]}{[P_1]_0-[P_1LP_2]-[P_1L]}$ (derived from Eq.(S6) and Eq.(S8))

yields

$$\frac{K_2/\alpha[P_1LP_2]}{[P_1L]}+\frac{K_1/\alpha[P_1LP_2]}{[P_1]_0-[P_1LP_2]-[P_1L]}+[P_1LP_2]=[P_2]_0 \tag{S15}$$

Therefore, the first derivative of Eq.(S15) with respect to $[P_1L]$ can be simplified to yield expression Eq.(S16).

Combined with Eq.(S14):

$$\frac{-K_2[P_1LP_2]}{[P_1L]^2}+\frac{K_1[P_1LP_2]}{([P_1]_0-[P_1LP_2]-[P_1L])^2}=0$$

Recall that $[P_1]_0=[P_1]+[P_1L]+[P_1LP_2]$, thus

$$K_2[P_1]^2=K_1[P_1L]^2 \tag{S16}$$

Derivation-Part 2.

In summary, several interesting relations concerning the equilibrium concentrations of various species, when the induced

$[P_1LP_2]$ is maximized, can be derived.

1) Combining Eq.(S16) and Eq.(S4) ($[P_1][L] = K_1[P_1L]$), we know the free inducer equilibrium concentration when induced complex is maximized (in Eq.(S17)). Apparently, this variable is independent of the protein concentrations and the cooperativity factor.

$$[L]_{[P_1LP_2]_{max}} = \sqrt{K_1K_2} \tag{S17}$$

2) Further, an interesting relationship among monomer proteins and protein-inducer complexes can be obtained when $[P_1LP_2]$ is maximized (Eq.(S18)).

$$[P_1]_{[P_1LP_2]_{max}}[P_2]_{[P_1LP_2]_{max}} = [P_1L]_{[P_1LP_2]_{max}}[LP_2]_{[P_1LP_2]_{max}} \tag{S18}$$

3) Importantly, we can derive the total amount of inducer that is necessary to produce the maximum $[P_1LP_2]$.

Eq.(S16) and Eq.(S4) gives: $[P_1L] = \frac{\sqrt{K_2}([P_1]_0 - [P_1LP_2])}{\sqrt{K_1} + \sqrt{K_2}}$, similarly: $[LP_2] = \frac{\sqrt{K_1}([P_2]_0 - [P_1LP_2])}{\sqrt{K_1} + \sqrt{K_2}}$,

Thus: $[P_1L] + [LP_2] + [P_1LP_2] = \frac{\sqrt{K_2}[P_1]_0 + \sqrt{K_1}[P_2]_0}{\sqrt{K_1} + \sqrt{K_2}}$. Recall that $[L]_0 = [L] + [P_1L] + [LP_2] + [P_1LP_2]$, therefore:

$$[L]_{0,[P_1LP_2]_{max}} = \sqrt{K_1K_2} + \frac{\sqrt{K_1}[P_2]_0 + \sqrt{K_2}[P_1]_0}{\sqrt{K_1} + \sqrt{K_2}} \tag{S19}$$

The expression of $[L]_{0,[P_1LP_2]_{max}}$ obtained by Perelson[4] is similar to Eq.(S17), due to their assumption that total amount of inducer is equivalent to the amount of free inducer. Obviously, the second term in Eq.(S19) accounts for the amount of inducers in the protein-bound form. This equation indicates that the protein-bound form of inducer can be ignored, or equivalently the assumption of free inducer being equal to the total inducer is valid, only when the protein concentrations $[P_1]_0$ *and* $[P_2]_0$ is significantly smaller than dissociation constants K_1 *and* K_2. More importantly, Eq.(S19) implies the possibility to calculate the dissociation constants by varying protein concentrations and observe the amount of inducer necessary to produce maximum complex.

4) The maximum amount of induced-complex $[P_1LP_2]_{max}$ can also be obtained.

Substitute $[P_1] = \frac{\sqrt{K_1}([P_1]_0 - [P_1LP_2])}{\sqrt{K_1} + \sqrt{K_2}}$ and $[LP_2] = \frac{\sqrt{K_1}([P_2]_0 - [P_1LP_2])}{\sqrt{K_1} + \sqrt{K_2}}$ into Eq.(S6) yields a quadratic equation:

$$[P_1LP_2]^2 - \left\{[P_1]_0 + [P_2]_0 + \left(\sqrt{K_1} + \sqrt{K_2}\right)^2/\alpha\right\}[P_1LP_2] + [P_1]_0[P_2]_0 = 0$$

Therefore the maximum amount of $[P_1LP_2]_{max}$ is given by:

$$[P_1LP_2]_{max} = \frac{1}{2}\left\{[P_1]_0 + [P_2]_0 + K_{app} - \sqrt{([P_1]_0 + [P_2]_0 + K_{app})^2 - 4[P_1]_0[P_2]_0}\right\} \tag{S20}$$

where $K_{app} = \left(\sqrt{K_1} + \sqrt{K_2}\right)^2 / \alpha = \left(\sqrt{K_3} + \sqrt{K_4}\right)^2$

5) The equilibrium concentration of all other species when $[P_1LP_2]$ is maximized can be trivially obtained, with each expression presented below (Eq.(S21)).

$$[P_1]_{[P_1LP_2]_{max}} = \frac{\sqrt{K_1}([\mathrm{P}_1]_0 - [P_1LP_2]_{max})}{\sqrt{K_1} + \sqrt{K_2}}, \quad [P_1L]_{[P_1LP_2]_{max}} = \frac{\sqrt{K_2}([\mathrm{P}_1]_0 - [P_1LP_2]_{max})}{\sqrt{K_1} + \sqrt{K_2}}$$
$$[P_2]_{[P_1LP_2]_{max}} = \frac{\sqrt{K_2}([\mathrm{P}_2]_0 - [P_1LP_2]_{max})}{\sqrt{K_1} + \sqrt{K_2}}, \quad [LP_2]_{[P_1LP_2]_{max}} = \frac{\sqrt{K_1}([\mathrm{P}_2]_0 - [P_1LP_2]_{max})}{\sqrt{K_1} + \sqrt{K_2}} \qquad \text{(S22)}$$

Section II. Results and Discussion

Effects of parameters on the response curve.

The bell-shaped response curve of $[P_1LP_2]$ versus $[L]_0$ is evaluated under various parametric conditions (Figure S3) as well as different concentrations of the proteins (Figures S4-S6). For simplicity, it is assumed that all quantities (namely $[P_1]_0$, $[P_2]_0$, K_1, K_2) share the same unit, e.g. nM, and unit notations are omitted in these Figures.

Dependence on K_1 and K_2. Comparison of the 4 curves in each panel of Figure S3-S5 shows that if the magnitudes of K_1 and K_2 are significantly different, the width of the bell-shaped response curve would notably increase. This is because when one of the proteins binds to the inducer with much higher affinity than the other (say $K_1 << K_2$), P_1L is quickly (almost) saturated at low inducer concentration, while LP_2 forms relatively slower at high inducer concentration.

Dependence on α. The effect of different values of α can also be generalized from Figures S3, where the proteins concentrations are comparable. When the system is non-cooperative, or $\alpha=1$, the curves are symmetrical about the maximum (Figures S3A). With positively cooperative systems ($\alpha>1$), the slope of the response curve is positive and steep when $[L]_0<[L]_{0,\,max}$, and is negative and shallow when $[L]_0>[L]_{0,\,max}$. Positive cooperativity postpones the reduction of $[P_1LP_2]$ because the binding of P_1 or P_2 to LP_2 or P_1L is more favored that the binding of P_1 or P_2 to L. Accordingly, with negatively cooperative system ($\alpha<1$), the curve is steeper when $[L]_0>[L]_{0,\,max}$.

Dependence on $[P_1]_0$ and $[P_2]_0$. Further, in order to examine the system behavior when the two protein concentrations are drastically different, we regenerated the curves in Figure S3 using the same set of parametric range while assuming $[P_2]_0$ is in excess over $[P_1]_0$ (Figure S4) or $[P_1]_0$ is in excess over $[P_2]_0$ (Figure S5). One result is that for non-cooperative system ($\alpha=1$, Figure S4A), the curve is unsymmetrical for small K_2, unlike the curves with large K_2. This is because $[P_2]_0$ (200 units) is in excess over K_2 (1 units), as previously pointed out by Douglass et. al.[3]. Further, examination of the corresponding curves in Figure S3, S4 and S5 shows that the bell-shaped response curve tends to be widened when the protein concentrations are very different, as compared with similar protein concentrations.

$[P_1]_0$ and $[P_2]$ << K_1 and K_2. A special circumstance is when both of $[P_1]_0$ and $[P_2]_0$ are significantly smaller than K_1 and K_2 (Figure S6). Under this situation, the amount of hetero-dimer $[P_1LP_2]$ is only a small fraction of $[P_1]_0$ or $[P_2]_0$ due to the exceedingly large dissociation constants. Yet, if the signal produced by the hetero-dimer is still detectable, for instance inside the cell where the dimerized receptors initiates a cascade of reactions which substantially amplifies the signal, the response curve can still be measured. The theoretically generated response curves in Figure S6 show that, unlike previously discussed in Figures S3-S5, all curves appear to be symmetrical about maximum despite the change in α or relative magnitude between K_1 and K_2. In fact, the consequence of the $[P_1]_0$, $[P_2]_0 << K_1$, K_2 is that $[L]_{0,max}$ can be

simplified to $\sqrt{K_1K_2}$ (Eq. (20), and that the ligand is in great excess over both proteins, and $[L]_0 \approx [L]$. In this situation, the ligand-in-excess theory (initially derived by Perelson[1]) can be applied instead of the exact theory presented here.

Simulation of the FKBP-rapamycin-FRB system with different K_1.

In the case of very tight binding, the protein concentrations used in experiments are usually much higher than dissociation constant. As shown in Figures S8C, S8D, when K_1 is on the orders of 0.1~1 nM, the binding curves with different values of K_1 become indistinguishable, suggesting that the concentration of P_1 in the solution is much larger than the dissociation constant, K_1. This may result in the inaccuracy for determining the tight binding constant. In order to further address the issue, we used the equation (4) to generate theoretically simulated data (each with 10% Gaussian noise), and then estimate the three parameters by fitting the data with our non-linear regression program. As shown in Figure S10, the data points for K_1=0.3 nM and 3 nM are rather similar, hence the difference in the dissociation constant cannot be detected. However, the data points generated with K_1 =20 nM or 50 nM are clearly discernable, therefore the p-values for K_1 estimation in these cases are small and the estimations are relatively precise (Figures S10C, S10D).

Theoretical test of the model with another system.

To further validate the new method, we propose to test our model using theoretical simulation. We believe that the well-characterized induced PPI system, dimerization of FKBP and SH2 domain of Fyn induced by a bifunctional molecule SLFpYEEI [5], is appropriate for test for the following reasons. The reactions involved in this system have been quantitatively analyzed using fluorescence polarization and competitive inhibition where the parameters are estimated as: K_1=12 nM, K_2=1.16 μM, K_3=60 nM and K_4=6 μM. As these parameters are different from rapamycin, the system should constitute a satisfying test of our method. To avoid the circular argument, we used the $[P_1LP_2]$ quintic polynomial presented in Douglass et. al.[3] to generate theoretically simulated data (each with 10% Gaussian noise), and then analyze the simulated data by our program for determination of the three parameters (Figure S11).

As previously stated, the choice of protein concentrations used in the experiment is critical for accurate estimation of parameters. When we simulate data with the same $[P_1]_0$ and $[P_2]_0$ concentrations as used in rapamycin system, the parameter estimation for both K_4 and C are extremely inaccurate (Figure S11A). However, if we increase the $[P_2]_0$ concentrations, non-linear fitting results in nicely overall fit and low p-values for all parameter estimations (Figure S11B). The reason for inaccuracy in Figure S11A is that, as the maximal level of hetero-dimer is controlled by both K_3, K_4 (see the following subsection) and the scaling factor C, thus neither can be accurately measured if both $[P_1]_0$ and $[P_2]_0$ are much smaller than K_{app}.

Implications of the expressions of $[L]_{0,\,max}$ and $[P_1LP_2]_{max}$.

Here we further discuss two interesting properties of the bell-shaped response curve.

1. The implication of expression of $[L]_{0,\,max}$ (Eq. S19):

Most notably, the analytical expression for $[L]_{0,\,max}$ demonstrates that the amount of ligand necessary to produce the largest hetero-dimer is not subject to change in cooperativity α. It also shows that $[L]_{0,\,max}$ is directly proportional to both protein concentrations. Therefore, the knowledge of protein concentrations, which could be estimated through fluorescence labeling or immunoblotting, and $[L]_{0,\,max}$, which could be estimated by maximal level of response curve, can be used to derive dissociation constants for in vivo experiments (Figure S12). For instance, for the same $[P_2]_0$, $[L]_{0,\,max}$ is linearly related to $[P_1]_0$ where the slope = $\frac{\sqrt{K_2}}{\sqrt{K_1}+\sqrt{K_2}}$ and the intercept =$\sqrt{K_1}+\frac{\sqrt{K_1}}{\sqrt{K_1}+\sqrt{K_2}}[P_2]_0$ (Eq. S19).

2. The implication of expression of $[P_1LP_2]_{max}$ (Eq. S20):

Interestingly, Eq. S20 resembles the expression of protein-protein complex concentration in a typical bimolecular interaction. It predicts that the maximal amount of heterodimer ($[P_1LP_2]_{max}$) exhibits the same behavior as the $[P_1P_2]$ in a bimolecular association, where $K_{app} = \left(\sqrt{K_3}+\sqrt{K_4}\right)^2$ is the apparent dissociation constant. Therefore, K_{app} can be estimated by fixing one of the protein concentrations, e.g. $[P_1]_0$, while changing $[P_2]_0$.

When $[P_1]_0 << K_{app}$, the relation between the signal $[P_2]_0$ is given by Eq. S23.

$$\begin{aligned} Signal &= C\cdot[P_1LP_2]_{max} \\ &= \frac{C}{2}\cdot\left\{[P_1]_0+[P_2]_0+K_{app}-\sqrt{([P_1]_0+[P_2]_0+K_{app})^2-4[P_1]_0[P_2]_0}\right\} \\ &= C\cdot\frac{[P_1]_0[P_2]_0}{[P_2]_0+K_{app}} \end{aligned} \qquad \text{(S23)}$$

When $[P_2]_0 << K_{app}$, the signal is given by Eq. S24:

$$Signal = \frac{C}{K_{app}}\cdot[P_1]_0[P_2]_0 \qquad \text{(S24)}$$

Therefore, changing $[P_2]_0$ only yields C/ K_{app}, but not C or K_{app}. This explains why the protein concentrations of Figure S11A did not yield accurate estimation of K_4 and C, while those of Figure S11B results in high-confidence estimation of all parameters.

The choice of protein concentrations during experimental design.

To sum up, we believe that the protein concentration choice is crucial in our method. There are two notes for the choice of protein concentrations:

i. The protein concentrations should be on the similar order as or smaller than the primary dissociation constants K_1, K_2.

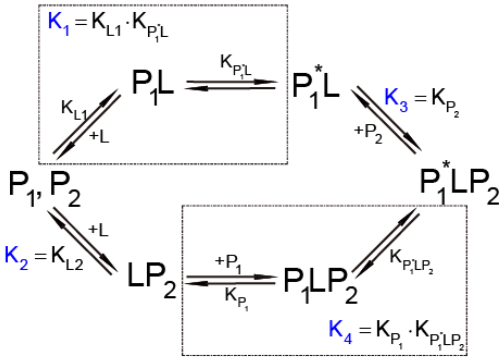
ii. At least one of the protein concentrations should be on the similar order as or larger than the secondary dissociation constants K_3, K_4.

It should be noted that these two rules are not stringent, and that protein concentrations can be one order of magnitude smaller or larger than the dissociation constants and the parameters can still be confidently estimated. We further note that the two rules above are deduced from general concerns for parameter estimation. In actual experiments, the limitations of experimental settings, for instance, the range of signal that can be accurately detected or the maximum protein concentration that is experimentally achievable, must also be taken into consideration.

Finally, in order to arrive at confident estimation of parameter with fewer number of trials, researchers are recommended to use the script provided in our MATLAB program (simulate_data.m) to generate theoretical test data to determine if the parameter estimation would be good enough under certain experimental settings.

Situations with ligand-induced conformal change.

Ligand-induced conformal change of the receptor is quite often observed in nature. For the ligand-mediated mechanism, where ligand induced conformational change in receptors, and then the dimerization is still mediated through direct interaction of the ligand to both receptors (Scheme S2, below), our model is still applicable.



Scheme S2. Ligand-induced conformational change in P_1.

However, in the 'receptor-mediated' model (e.g. EGFR dimerization by EGF), dimerization is mediated entirely by receptor-receptor interactions after the ligand induction [6,7]. In such cases, the ligand binds to their cognate receptor in 1:1 complex, and association is mediated through the induced allosteric conformational change of receptors Such model would produce a saturation response curve to increasing ligand concentrations, instead of a bell-shaped response curve that is observed in our model, and our method would not be applicable to these situations.

References:

(1) Gary, M. R.; Johnson, D. S.; WH Freeman and Company, New York, 1979.

(2) Pan, V. Y.; Chen, Z. Q. In *Proceedings of the thirty-first annual ACM symposium on Theory of computing*; ACM, 1999, pp 507-516.

(3) Douglass, E. F.; Miller, C. J.; Sparer, G.; Shapiro, H.; Spiegel, D. A. *J. Am. Chem. Soc.* **2013**, *135*, 6092-6099.

(4) Perelson, A. S. *Math. Biosci.* **1980**, *49*, 87-110.

(5) Braun, P. D.; Wandless, T. J. *Biochemistry* **2004**, *43*, 5406-5413.

(6) Heldin, C.-H. *Cell* **1995**, *80*, 213-223.

(7) Lemmon, M. A.; Bu, Z.; Ladbury, J. E.; Zhou, M.; Pinchasi, D.; Lax, I.; Engelman, D. M.; Schlessinger, J. *EMBO J.* **1997**, *16*, 281-294.

Section III. Figures.

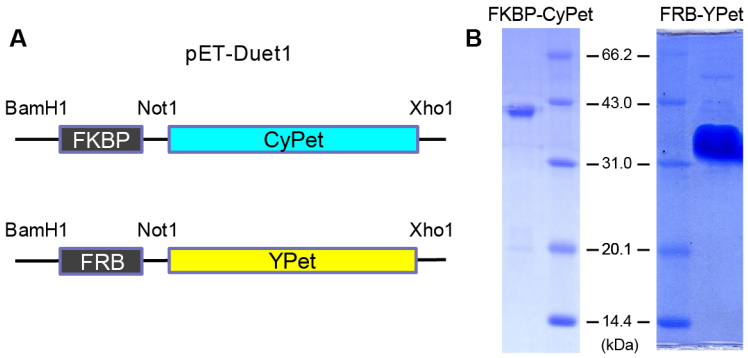


Figure S1. Constructs and SDS-page of fusion proteins. (A) Constructs for FKBP-CyPet and FRB-YPet. (B) SDS-page for purified FKBP-CyPet and FRB-YPet.

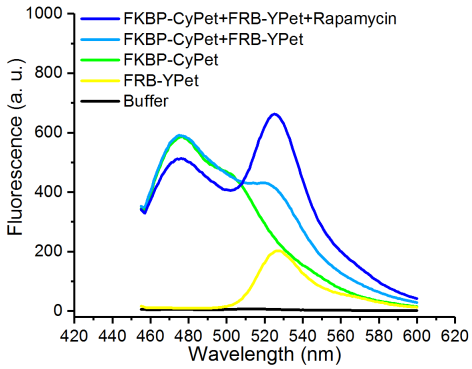


Figure S2. Emission spectra of fusion proteins. Black: PBS buffer (pH 7.4, with 2.5% EtOH), yellow: FRB-YPet (39 nM), green: FKBP-CyPet (47.5 nM), light blue: FKBP-CyPet (47.5 nM) + FRB-YPet (39 nM), blue: FKBP-CyPet (47.5 nM) + FRB-YPet (39 nM) + rapamycin (100 nM). All specimens were excited at the Donor excitation wavelength (λ_{ex}(D) = 438 nm).

Figures S3-S6. The response curve for different values of α, K_1, K_2, $[P_1]_0$ and $[P_2]_0$.

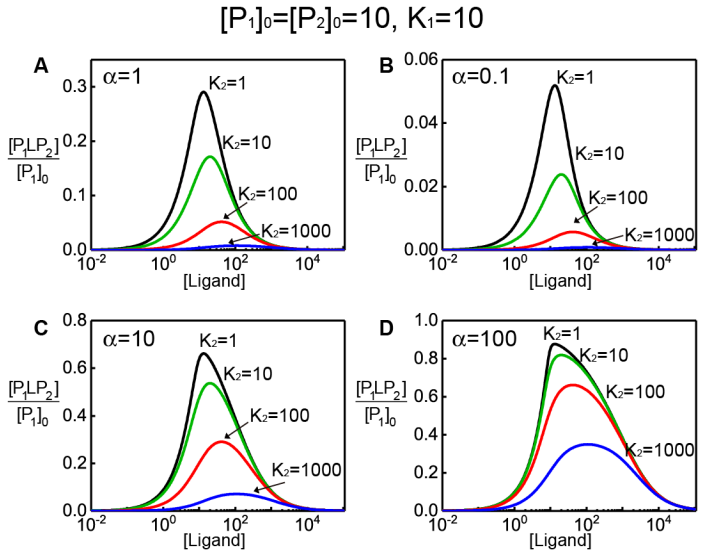


Figure S3. The response curve for different values of α and K_1, K_2 under fixed protein concentrations $[P_1]_0 = [P_2]_0 = 10$. All the theoretically generated hetero-dimer concentrations are normalized by total P_1 concentration. The cooperativity factor is shown as: α=1 (A), α=0.1 (B), α=10 (C), α=100 (D). The dissociation constant K_1=10 is fixed, while K_2 is selected to range 4 orders of magnitudes: K_2=1 (black), K_2=10 (green), K_2=100 (red), K_2=1000 (blue) as indicated in each panel.

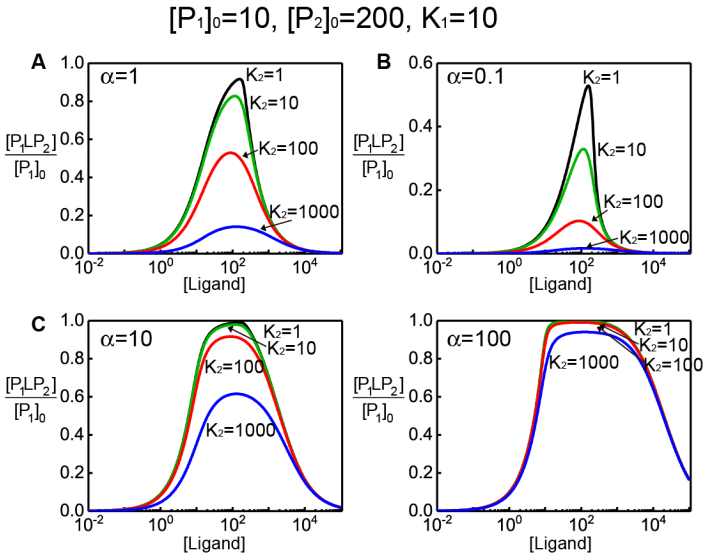


Figure S4. The response curve for different values of α and K_1, K_2 under fixed protein concentrations $[P_1]_0 = 10$, $[P_2]_0 = 200$. As the maximum amount of hetero-dimer $[P_1LP_2]$ is limited by the smaller one of $[P_1]_0$ and $[P_2]_0$, all the theoretical lines generated in this figure are normalized by $[P_1]_0$. The coloring scheme is the same as in Figure S3.

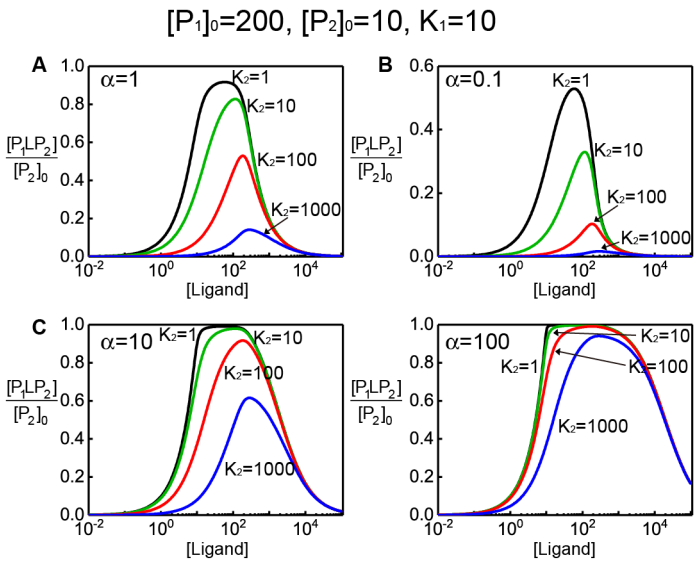


Figure S5. The response curve for different values of α and K_1, K_2 under fixed protein concentrations $[P_1]_0 = 200$, $[P_2]_0 = 10$. As the maximum amount of hetero-dimer $[P_1LP_2]$ is limited by the smaller one of $[P_1]_0$ and $[P_2]_0$, all the theoretical lines generated in this figure are normalized by $[P_2]_0$. The coloring scheme is the same as in Figure S3.

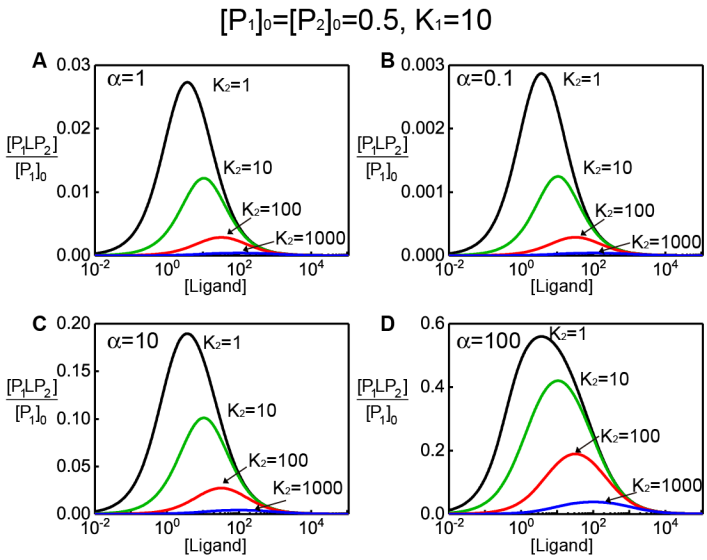


Figure S6. The response curve for different values of α and K_1, K_2 under fixed protein concentrations $[P_1]_0 = [P_2]_0 = 0.5$ ($<<K_1$, K_2). The coloring scheme is the same as in Figure S3.

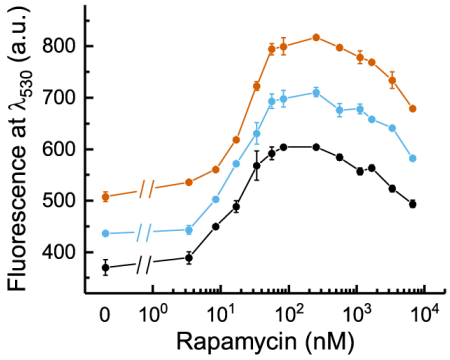


Figure S7. Experimental data before extraction of F_0. Experimental data of rapamycin induced FKBP-CyPet and FRB-YPet before extraction of F_0 (non-FRET fluorescence at λ_{530} without rapamycin).

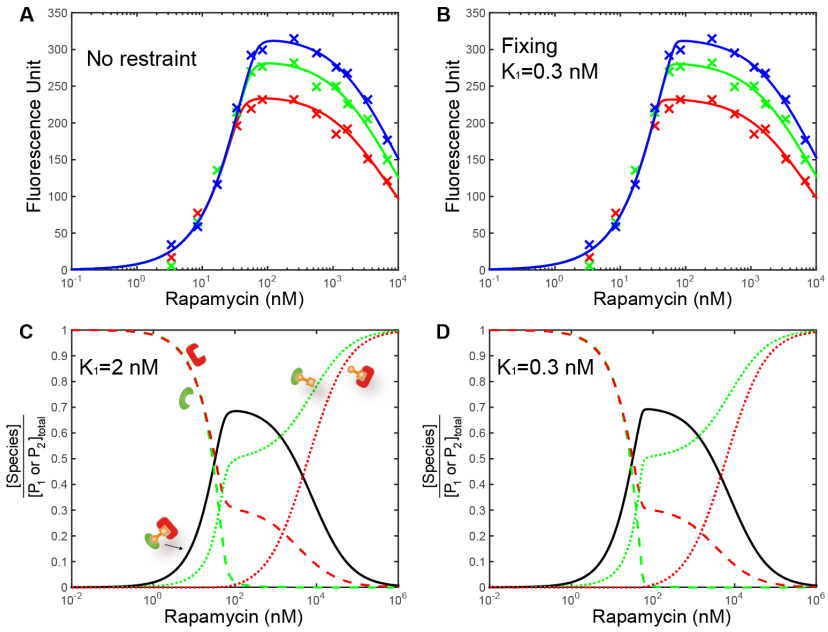


Figure S8. Analysis of parameter K_1. (A-B) The FRET experimental data was analyzed by our method with (A) no restraints and (B) K_1=0.3 nM fixed. (Figure 2C is shown again in (A) for side-to-side comparison.) The curves represent different protein concentrations: $[\text{FRB-Ypet}]_0$ = 39 nM and $[\text{FKBP-CyPet}]_0$ = 39.6 nM (red), 55.4 nM (green), 71.2 nM (blue). The fitting results of K_2, K_4 and C in (B) are similar to those in (A) (see Tables S1, S2 for detail). (C-D) Theoretically generated equilibrium concentrations of all species in the mixture assuming different K_1. The equilibrium concentrations are shown as $\frac{[P_1]}{[P_1]_0}$ (green, dashed line), $\frac{[P_2]}{[P_2]_0}$ (red, dashed line), $\frac{[P_1L]}{[P_1]_0}$ (green, dotted line), $\frac{[LP_2]}{[P_2]_0}$ (green, dashed line) and $\frac{[P_1LP_2]}{[P_2]_0}$ (black, solid line). All lines are generated with: K_2 = 1.4 μM, K_4 = 12 nM, $[\text{FKBP-CyPet}]_0$ ($[P_1]_0$) = 55.4 nM, $[\text{FRB-YPet}]_0$ ($[P_2]_0$) = 39 nM and K_1 = 2 nM (C), K_1 = 0.3 nM (D).

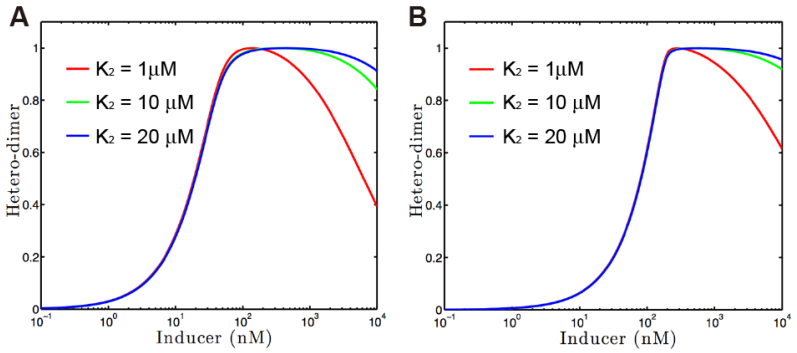


Figure S9. Influence of K_2 on the shape of the curve. The two protein concentrations are assumed to be equal and K_1 is assumed to be 0.3 nM for all theoretical lines. The figure shows percentage of hetero-dimer concentration for (A) α=100 and (B) α =1000, and various level of K_2 as indicated. This figure demonstrates that large K_2 significantly slows down the decomposition process of P_1LP_2. Therefore, our experimental data of rapamycin induced FKBP-CyPet and FRB-YPet dimerization (Figure 2C) defies the possibility for a large K_2 (e.g. >10 μM).

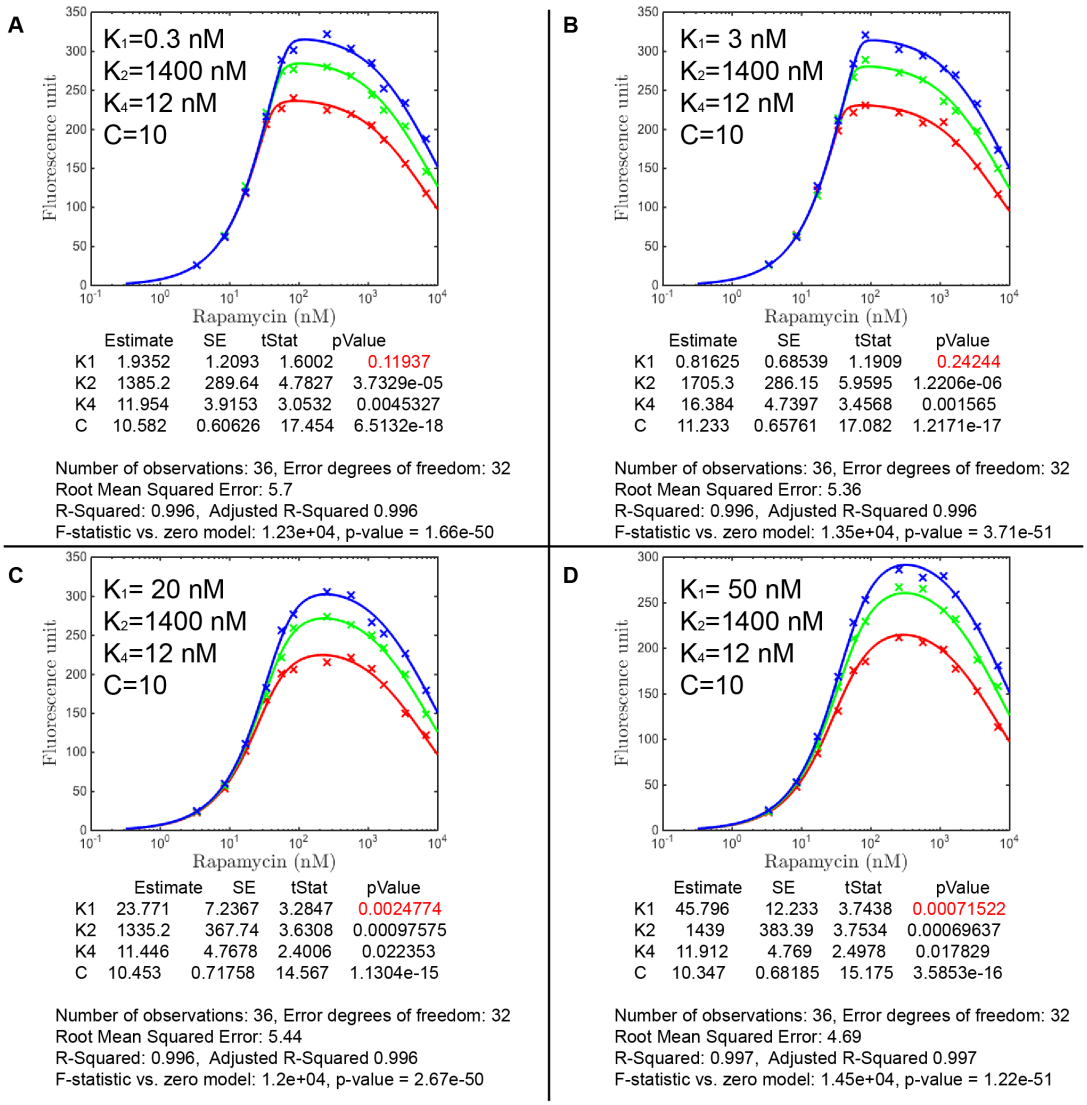
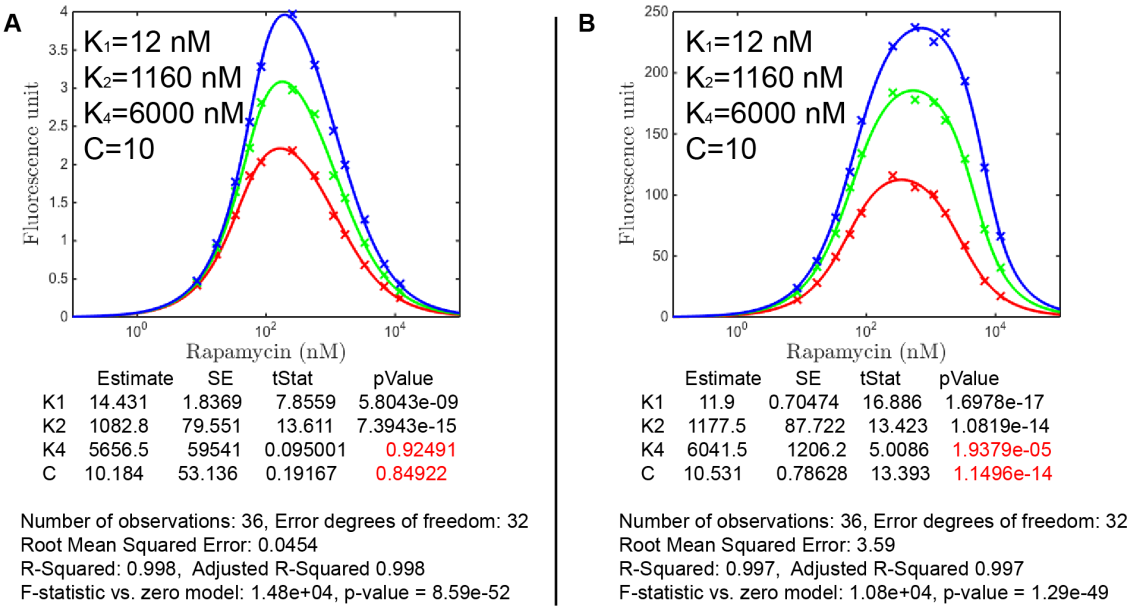


Figure S10. Analysis of theoretically simulated data with various K_1 value. The protein concentrations $[P_2]_0$ = 39 nM and $[P_1]_0$ = 39.6 nM (red), 55.4 nM (green), 71.3 nM (blue). The equilibrium parameters are assumed as indicated in the figure. The simulated data points are labeled as crosses, and fitted curves are plotted. The statistics of non-linear fitting are provided below each figure. Unlike in the cases where K_1=0.3 nM or 3 nM. the p-values for the cases where K_1=20 nM or 50 nM are relatively small, thus these estimations are more precise.



	Estimate	SE	tStat	pValue
K1	14.431	1.8369	7.8559	5.8043e-09
K2	1082.8	79.551	13.611	7.3943e-15
K4	5656.5	59541	0.095001	0.92491
C	10.184	53.136	0.19167	0.84922

Number of observations: 36, Error degrees of freedom: 32
Root Mean Squared Error: 0.0454
R-Squared: 0.998, Adjusted R-Squared 0.998
F-statistic vs. zero model: 1.48e+04, p-value = 8.59e-52

	Estimate	SE	tStat	pValue
K1	11.9	0.70474	16.886	1.6978e-17
K2	1177.5	87.722	13.423	1.0819e-14
K4	6041.5	1206.2	5.0086	1.9379e-05
C	10.531	0.78628	13.393	1.1496e-14

Number of observations: 36, Error degrees of freedom: 32
Root Mean Squared Error: 3.59
R-Squared: 0.997, Adjusted R-Squared 0.997
F-statistic vs. zero model: 1.08e+04, p-value = 1.29e-49

Figure S11. Simulation of SLFpYEEI induced FKBP and Fyn-SH2 dimerization and parameter estimation. The data points are theoretically generated with addition of 10% Gaussian noise. The equilibrium parameters are assumed as indicated in the figure. (A) Data simulated using $[P_2]_0$ = 39 nM and $[P_1]_0$ = 39.6 nM (red), 55.4 nM (green), 71.3 nM (blue), same as previously used in rapamycin experiment. (B) Data simulated using $[FKBP]_0$ ($[P_1]_0$) = 50 nM, $[Fyn]_0$ ($[P_2]_0$) = 2.0 μM (red), 4.0 μM (green) and 6.0 μM (blue). Both sets of data are then used to generate the parameter estimations with our program, with the resultant non-linear fitting statistics shown below. The parameter estimations in (B) are more reliable than in (A).

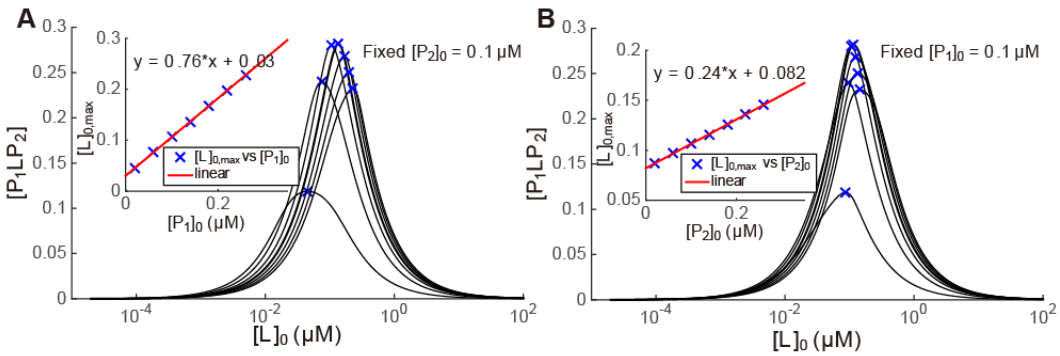


Figure S12. Theoretical experiment that shows an estimation of dissociation constants. For systems where the maximum of response curve is clearly discernible, dissociation constants can be directly estimated from Eq. S19. For the demonstration here, dissociation constants are assumed to be: K_1 = 2 nM , K_2 =20 nM. (A) Theoretical curves generated with $[P_2]$ = 0.1 μM, and different $[P_1]$ (black lines), with the positions where $[P_1LP_2]$ is maximized indicated as blue crosses. Inset: linear fitting to Eq. S19. (B) Theoretical curves generated with $[P_1]$ = 0.1 μM and different $[P_2]$. Both theoretical experiments yield the correct dissociation constants as previously defined.

Section IV. Tables

Table S1. Statistics of parametric fitting to experimental data.

Estimated Coefficients:

	Estimate (nM)	SE	tStat	pValue
K1	2.0589	1.9095	1.0783	0.28898
K2	1434.8	461.48	3.109	0.0039249
K4	12.358	6.2576	1.9749	0.056959
C	10.541	0.94649	11.137	1.5249e-12

Number of observations: 36, Error degrees of freedom: 32
Root Mean Squared Error: 8.72
R-Squared: 0.991, Adjusted R-Squared 0.99
F-statistic vs. zero model: 5.15e+03, p-value = 1.9e-44

Table S2. Statistics of parametric fitting to experimental data while fixing K_1=0.3 nM.

Fixed Coefficients:
K1=0.3 nM
Estimated Coefficients:

	Estimate (nM)	SE	tStat	pValue
K2	1643.8	433.45	3.7923	0.00060427
K4	14.05	6.5253	2.1531	0.038713
C	10.722	0.96532	11.107	1.0951e-12

Number of observations: 36, Error degrees of freedom: 33
Root Mean Squared Error: 8.91
R-Squared: 0.99, Adjusted R-Squared 0.989
F-statistic vs. zero model: 6.57e+03, p-value = 7.05e-46

Table S3. Experimental non-linear fitting results compared with literature value.

K_1 nM	K_2 nM	K_3 nM	K_4 nM	α	source
0.35	26,000	1.6×10^{-4}*	12	2,170*	Banaszynski et al.
0.3	-		24		Tamura et al.
2.1#	1,400	2.5×10^{-3}*	12	117*	This study

*Calculated based on experimental results. #Low confidence parameter.

Notes for usage of MATLAB program.

The MATLAB programs and their respective functions are listed below.

MATLAB Program	Function	Corresponding Figures
demo_fit_fret_data.m	Analysis of FRET experiment	Figures 2C, S8A
demo_fit_fret_data_fixk1.m	Analysis of FRET experiment	Figure S8B
change_K2.m	Analyze response curves for different K_2	Figures S3-S6, S9
generate_theoretical_data.m	Generate all equilibrium concentrations	Figures 1C, S8C and S8D
simulate_data.m	Perform theoretical simulation	Figure S10
simulate_fkbp_fyn.m	Simulation of SLFpYEEI system.	Figure S11