Supporting Information

Homotropic Cooperativity from the Activation Pathway of the Allosteric Ligand-Responsive Regulatory Protein TRAP†

Ian R. Kleckner[a], Craig A. McElroy[b], Petr Kuzmic[c],Paul Gollnick[d], Mark P. Foster[a,b]*

[a]Biophysics Program, The Ohio State University, 484 West 12th Ave, Columbus, OH 43210

[b] Departments of Chemistry and Biochemistry, and Center for RNA Biology, The Ohio State University, 484 West 12th Ave, Columbus, OH 43210

[c]BioKin Ltd., 15 Main Street Suite 232, Watertown, MA 02472

[d] Department of Biological Sciences, University at Buffalo, the State University of New York, Buffalo, NY 14260

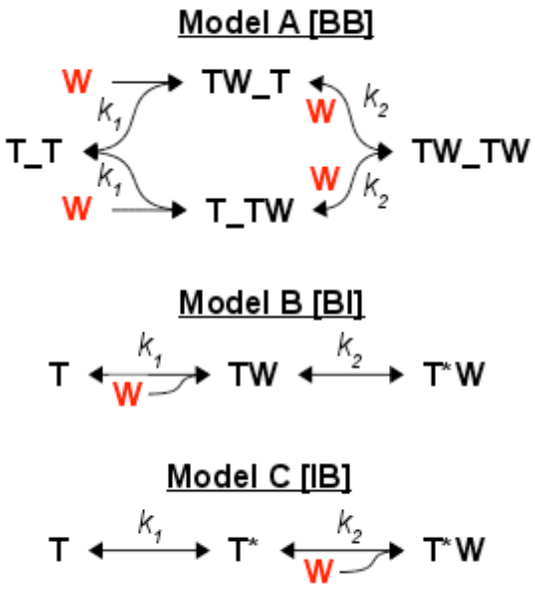
email contacts: ian.kleckner@gmail.com, mcelroy.31@gmail.com, pkuzmic2@biokin.com, gollnick@buffalo.edu, foster.281@osu.edu

† This project was funded by the National Institutes of Health grant R01GM077234, and grant # 1019960 from the National Science Foundation.

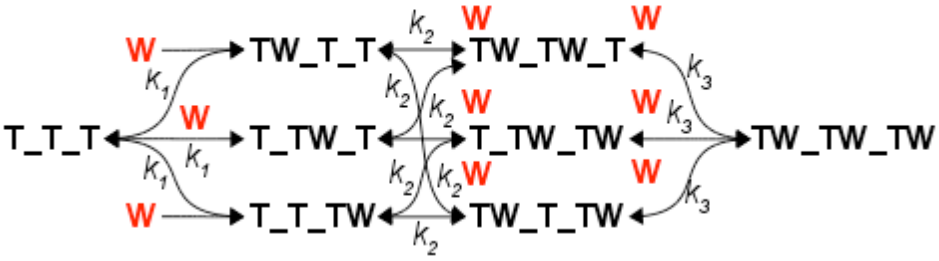
* Corresponding author: 484 West 12th Ave, Columbus, OH, 43210, USA. Phone: (614) 292-1377; Fax: (614) 292-6773; Foster.281@osu.edu

Supporting Information

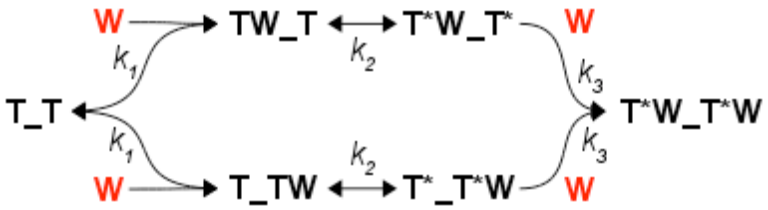
Testing and discriminating candidate models. The program DYNAFIT version 4 (1, 2) was used to compare a variety of models to fit the stopped-flow (SF) data because it permits numerical integration of the rate laws that it automatically derives from an arbitrary reaction mechanism, thus simplifying the comparison process. Ten two- and three-step mechanisms involving binding of Trp to TRAP with optional isomerization are listed below (models A through J). The abbreviations used are [B] = binding, [I] = isomerization, W = Trp, and T = TRAP. An isolated binding site is designated T, a pair of binding sites is designated T_T, and a triplet of binding sites is designated T_T_T. Each isomerization event is designated by an additional star (*) for the set of binding sites. For example, model A is "Bind-Bind [BB]" which describes a pair of binding sites (T_T) with two successive Trp binding steps with distinct binding rates k_1 and k_2. For clarity in the following figures, each rate k designates both forward and backward rate constants, k^F and k^B (i.e., the forward and backward rates are not shown explicitly, but they are used in the fitting process).



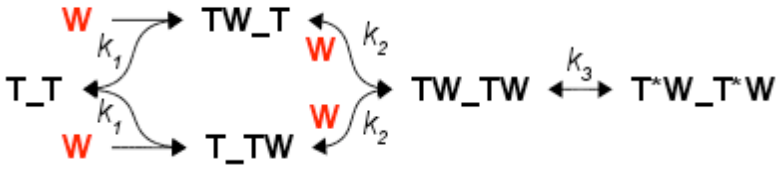
Model D [BBB]



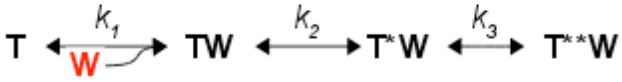
Model E [BIB]



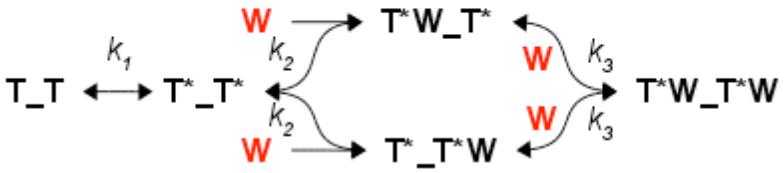
Model F [BBI]



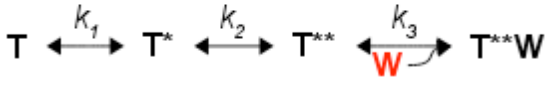
Model G [BII]



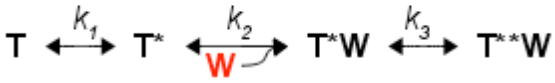
Model H [IBB]



Model I [IIB]



Model J [IBI]



For each mechanism at each temperature, all the 6-7 SF time-courses acquired at that temperature were globally fit by specifying each of the post-mixing concentrations of Trp and TRAP that correspond to each time-course (n.b., the concentration of a pair of binding sites, T_T, is half the concentration of a single binding site, T). The fitted parameters were the forward and backwards rate constants for each kinetic step, k^F and k^B, the baseline offsets for each time-course, and one or more fluorescence response values, R, which were constrained as follows. States without Trp are considered invisible (i.e., $R = 0$). An isomerization event, designated here by adding a star (*) to the state, will change the response value (e.g., R(TW) and R(T*W) can differ). For the initial comparisons of the 10 models, fitting was made more tractable by assuming that a binding event scales the response value by the number of Trp molecules bound (e.g., R(TW_TW) = 2 * R(TW_T) = R(T_TW); otherwise there would be too many adjustable parameters in the model). Subsequent fits comparing the two most promising simpler models, A [BB] and B [BI], did not require this constraint because these models were simple enough to accommodate an extra fitting parameter (i.e., allowing R(TW_TW) to be indepent of R(TW_T)).

As discussed in the main text, model A [BB] appeared most appropriate to describe our data given the following model comparison procedure. First, comparisons of all ten models indicated that the three-step models (D, E, F, G, H, I, and J) were not appropriate because they always returned at least some parameter values that were not intelligible in some way, such as: (1) response values for some Trp-bound states were zero, which is physically unreasonable, (2) temperature-dependences of some parameters were physically unreasonable (e.g., sporadically changing with temperature, not

monotonically changing with temperature), or (3) fitting errors were over 1000% (i.e., undefined parameter value). The result that the three-step models were too complex is consistent with the multi-phase exponential fitting results, wherein although three-phase exponential models fit the data with smaller total residuals, the three-phase exponential model parameter values were poorly constrained (Fig. S1, Table S3). Next, we compared the three two step models (A, B, and C). Model A [BB] was statistically favored with the sum of squared errors (SSE) metric at all temperatures except for 45ºC, where models A [BB] and B [BI] were effectively equivalent (Table S1). Model C [IB] fit the data so poorly that it was obvious the mechanism was inappropriate.

Supporting Figures

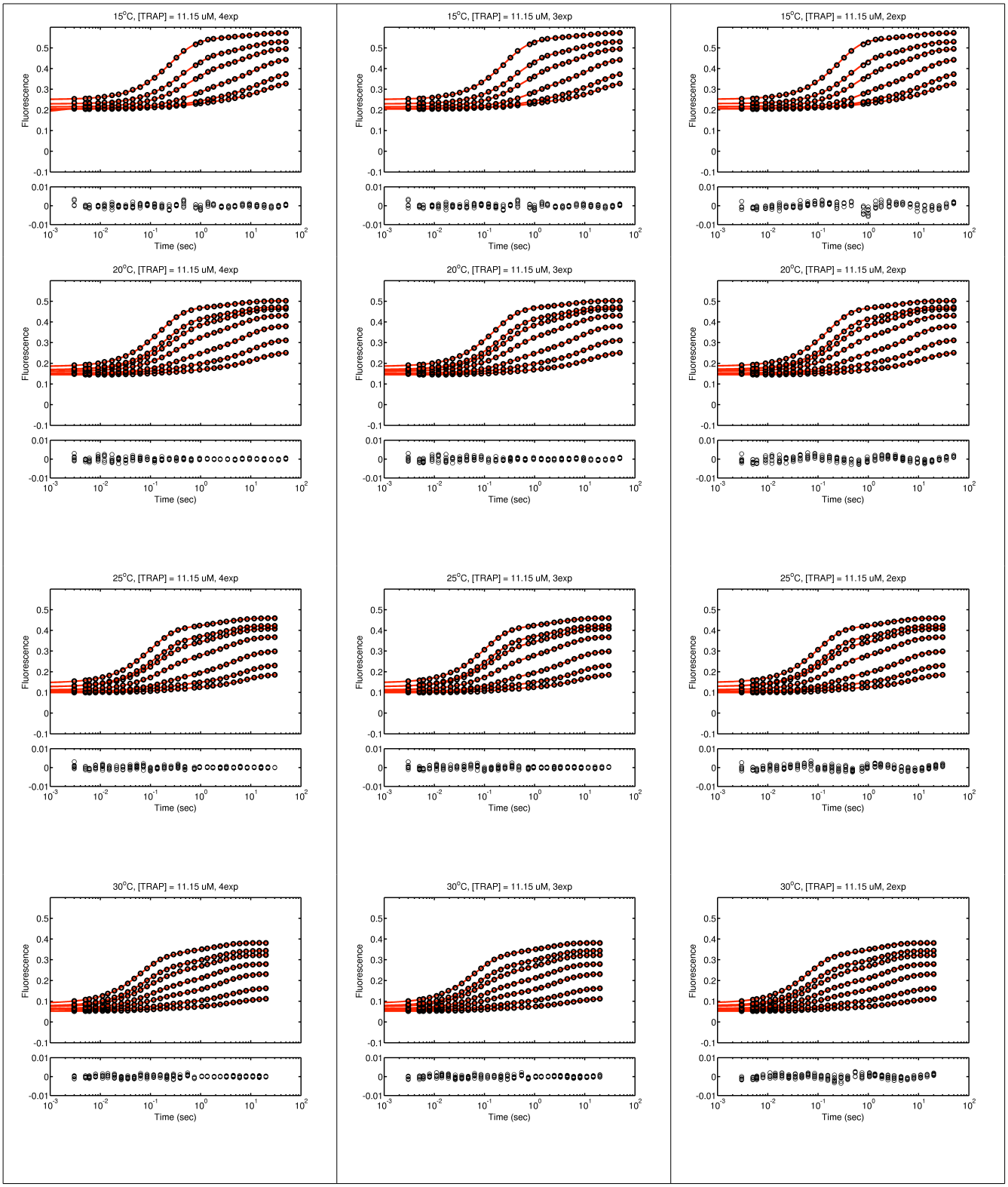


Figure S1 (continued)

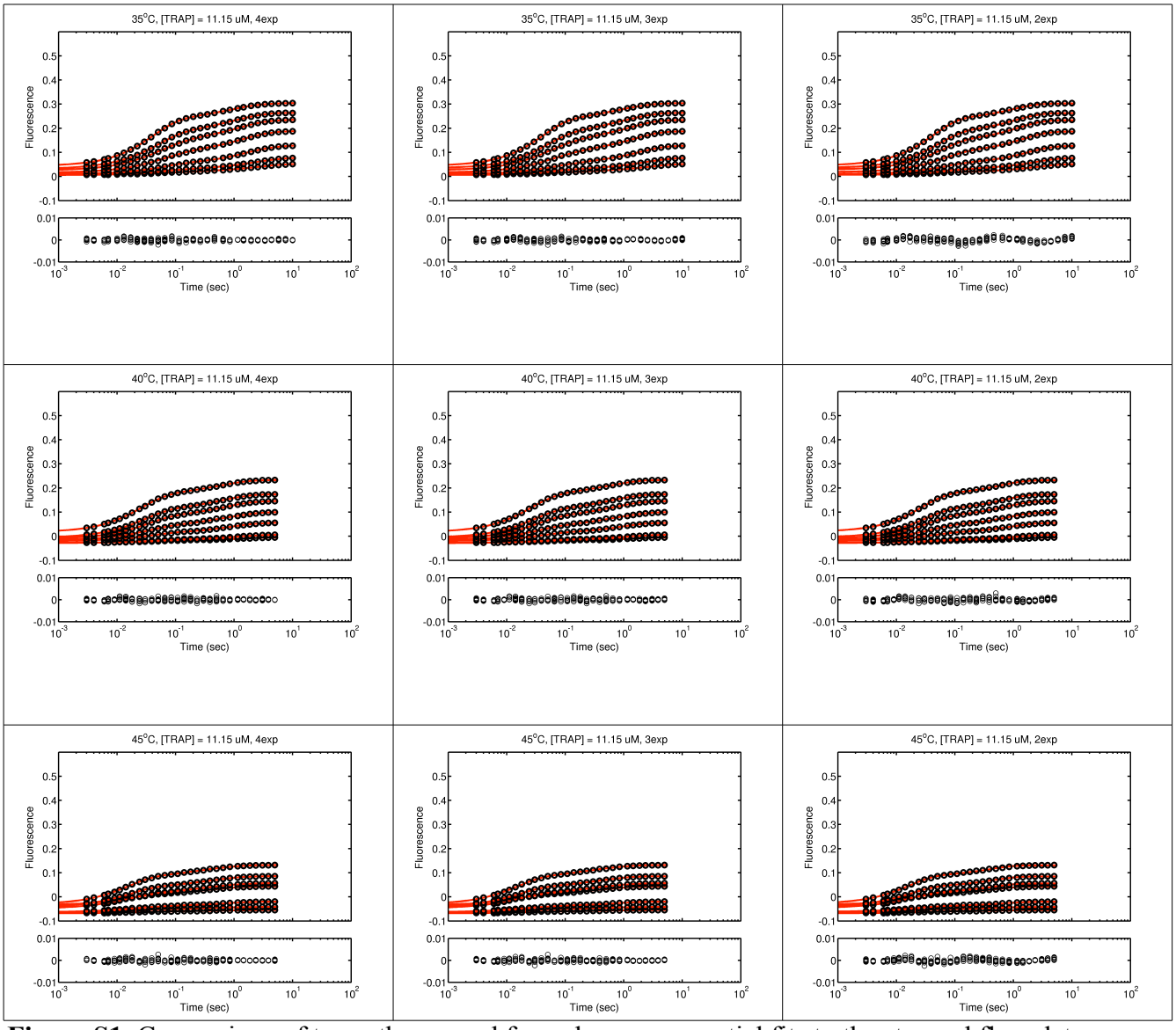


Figure S1. Comparison of two-, three-, and four-phase exponential fits to the stopped flow data. Three-phase exponential fits are generally preferred when fits are compared using the Bayesian Information Criterion (BIC; Table S3). However, the two-phase exponential fit does capture the two major phases of the time course (i.e., the initial fast phase and the final slow phase). The bottom of each panel shows the residuals, which are largest for the two-phase exponential fit.

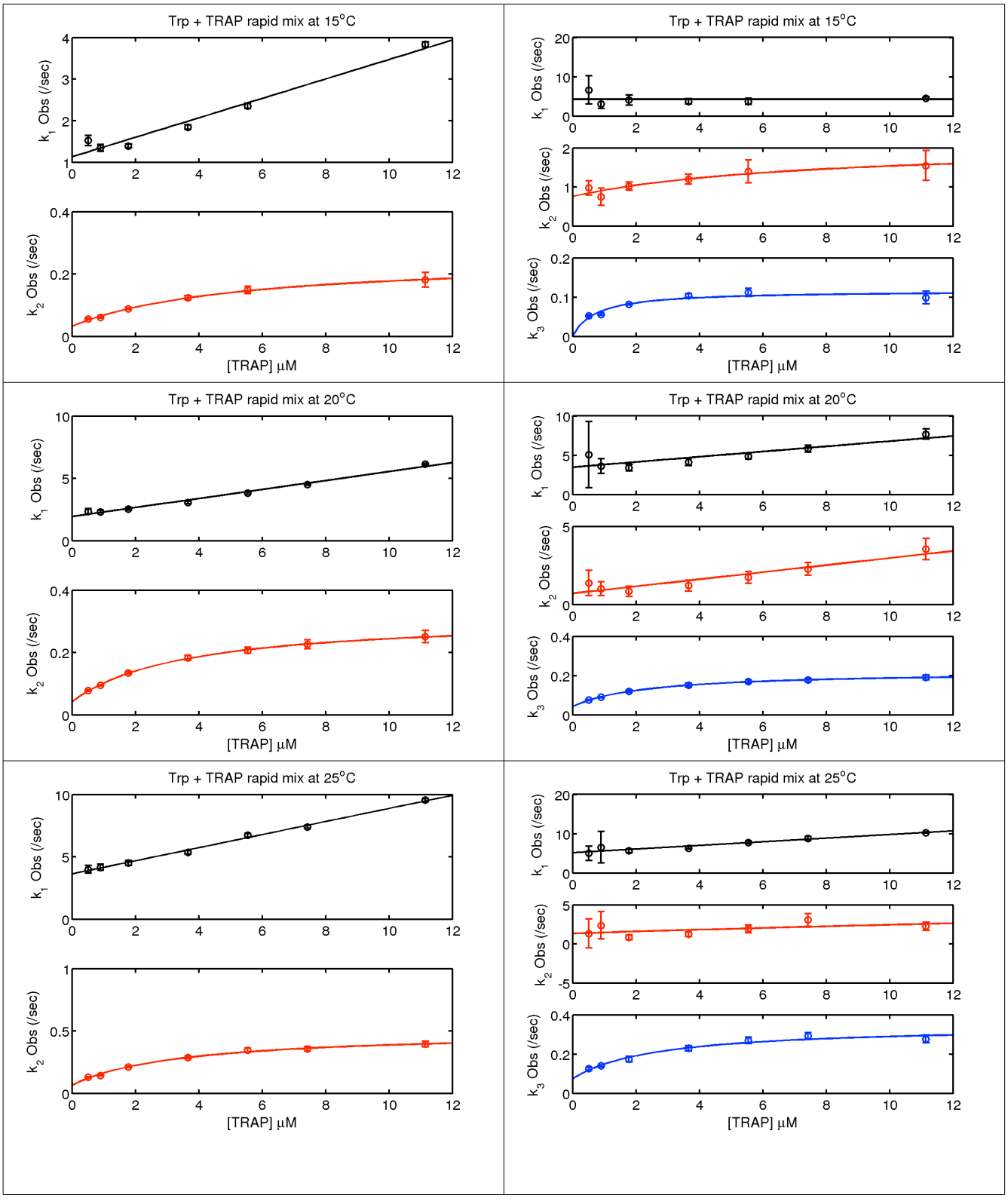

Trp + TRAP rapid mix at 15°C
k1 Obs (/sec)
k2 Obs (/sec)
[TRAP] µM
Trp + TRAP rapid mix at 15°C
k1 Obs (/sec)
k2 Obs (/sec)
k3 Obs (/sec)
[TRAP] µM
Trp + TRAP rapid mix at 20°C
k1 Obs (/sec)
k2 Obs (/sec)
[TRAP] µM
Trp + TRAP rapid mix at 20°C
k1 Obs (/sec)
k2 Obs (/sec)
k3 Obs (/sec)
[TRAP] µM
Trp + TRAP rapid mix at 25°C
k1 Obs (/sec)
k2 Obs (/sec)
[TRAP] µM
Trp + TRAP rapid mix at 25°C
k1 Obs (/sec)
k2 Obs (/sec)
k3 Obs (/sec)
[TRAP] µM

Figure S2 (continued)

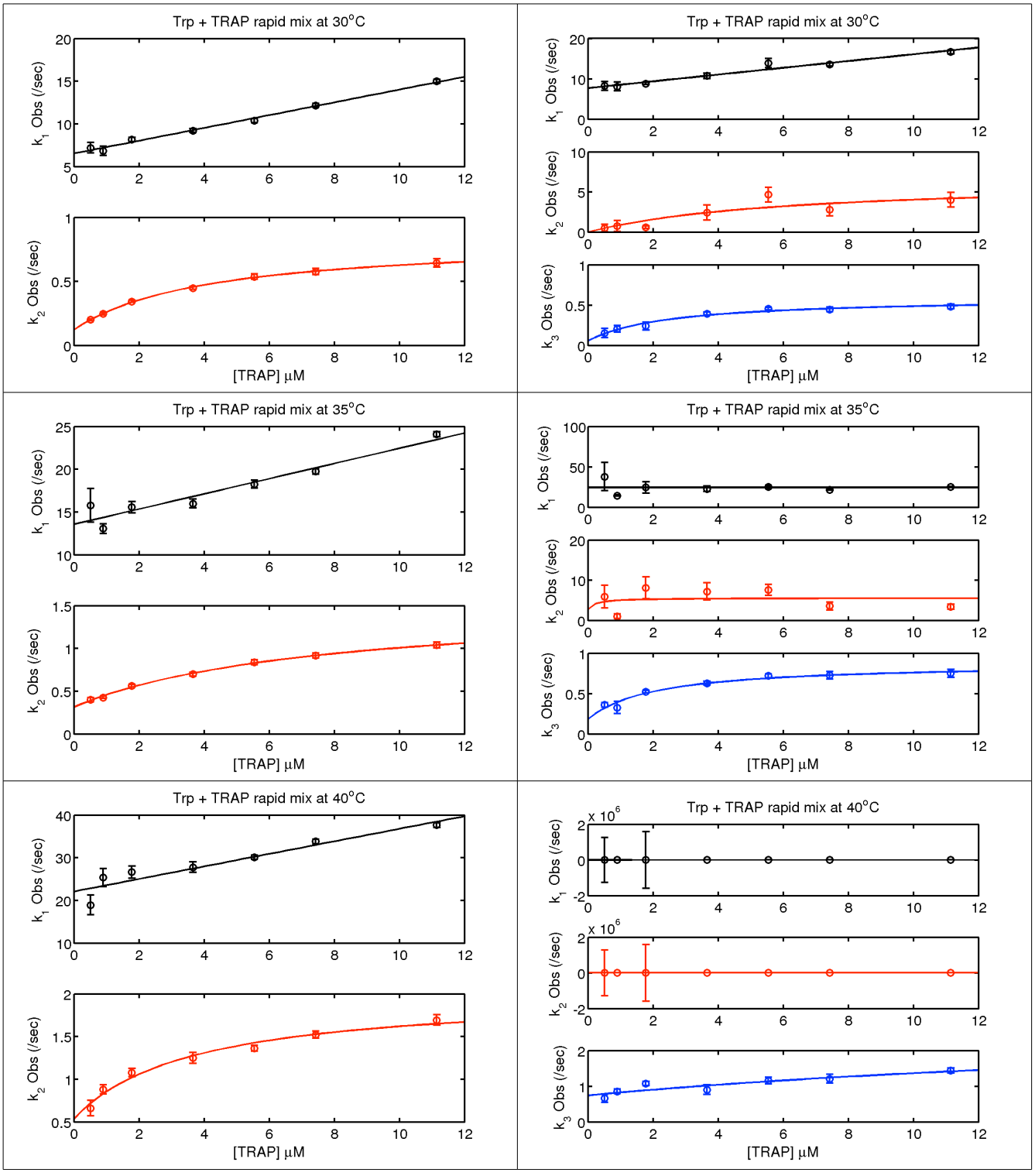


Figure S2 (continued)

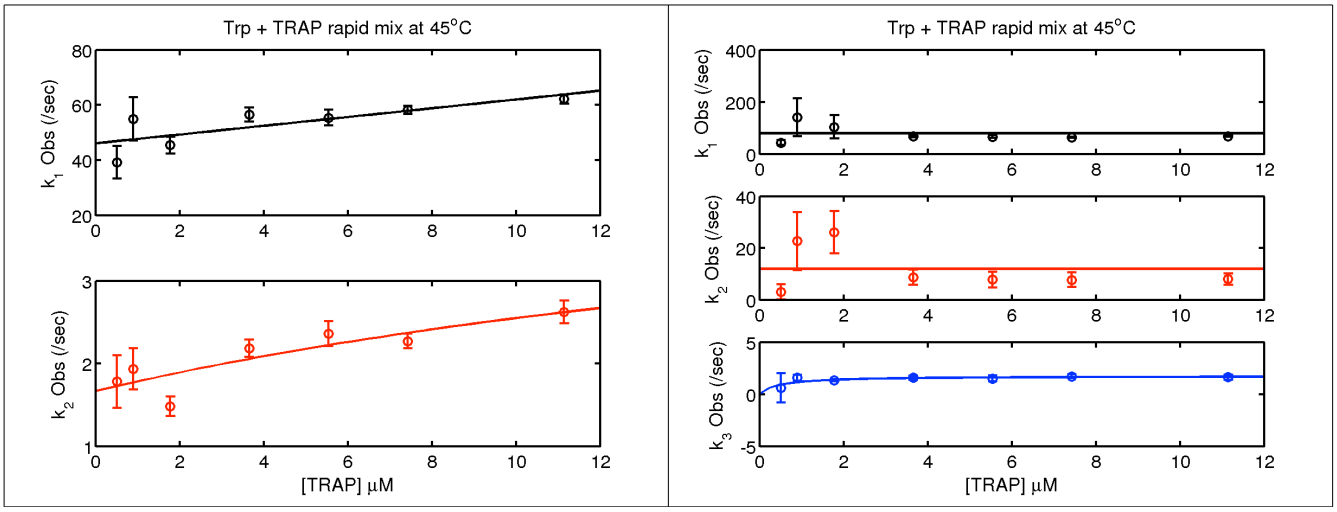


Figure S2. The concentration-dependence of the mechanism-independent apparent rate constants reveal that the second step mechanistically follows the first step. Each sub-figure shows one apparent rate constant obtained from the two- or three-phase exponential fit (left or right, respectively) plotted against the TRAP concentration in the sample where that rate constant was observed. The solid lines in each plot show fits of the k^{Obs} vs. [TRAP] to either a line or a hyperbola. For the two-phase exponential fits (left), the faster (first) step exhibits a linear relationship between k^{Obs} and [TRAP] and the second (slower) step exhibits a hyperbolic relationship between k^{Obs} and [TRAP]. For the three-phase exponential fits (right), there appears to be an additional step that occurs *between* steps 1 and 2 from the two-phase analysis (i.e., two-phase steps 1 and 2 are three-phase steps 1 and 3). Unfortunately, the three-phase exponential fit appears unintelligible as some rate constants are poorly defined (i.e., large error bars as in steps 1 and 2 at 40ºC), and as the data points do not lie near the theoretically expected fit lines. The error bars show the standard errors from fitting the time-dependent SF data to a two-phase or three-phase exponential function and are considered a lower-limit of the uncertainty. Fitted values are shown in Table S2.

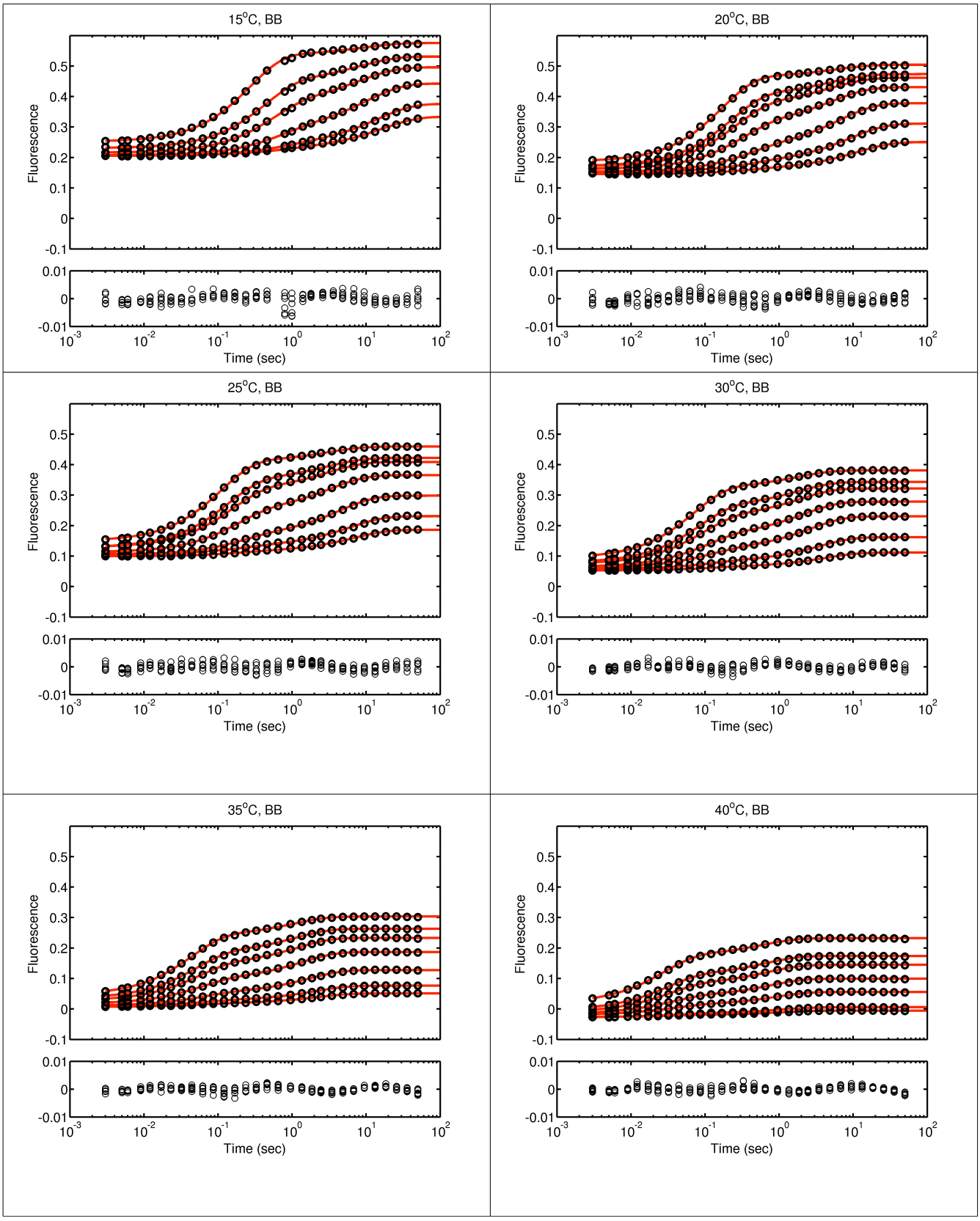

15°C, BB
20°C, BB
25°C, BB
30°C, BB
35°C, BB
40°C, BB
Fluorescence
Time (sec)

Figure S3 (continued)

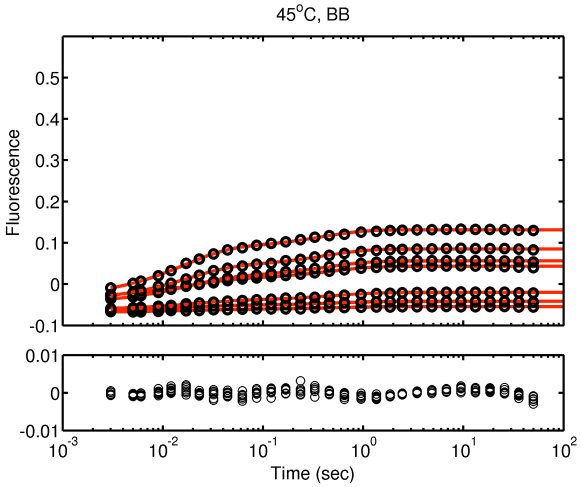


Figure S3. Stopped-flow data and fit to model A [BB] at each of seven temperatures (see above). Data at each of the seven time courses (A -G) correspond to 1 μM Trp mixed with TRAP 11-mer at concentrations of 0.093, 0.164, 0.323, 0.664, 1.008, 1.350, and 2.027 μM (1.022, 1.800, 3.558, 7.306, 11.088, 14.846, and 22.292 μM binding sites). Note, data at 15°C with 1.350 μM TRAP 11-mer were unreliable and were therefore removed.

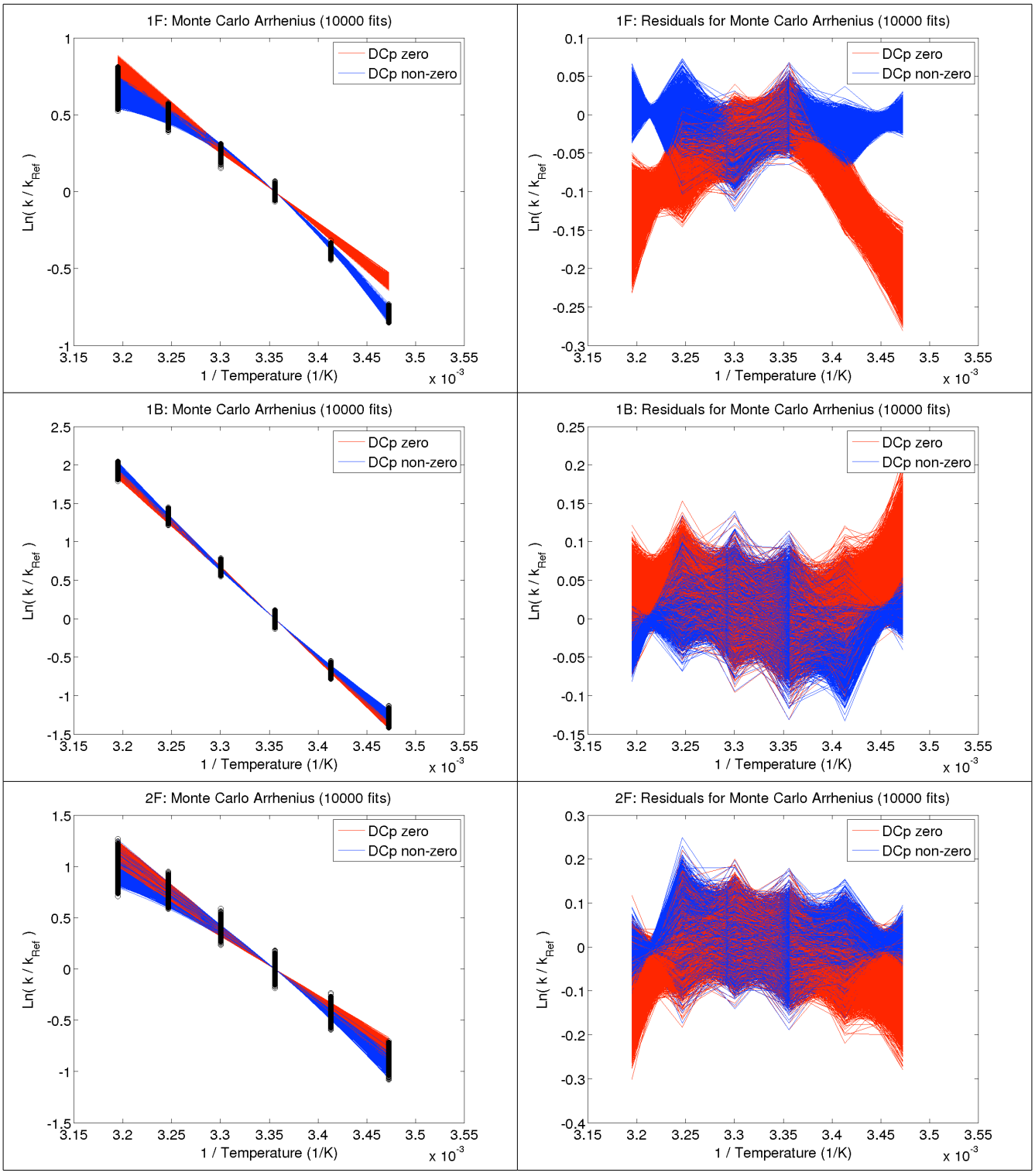

1F: Monte Carlo Arrhenius (10000 fits)
DCp zero
DCp non-zero
Ln(k / k_Ref)
1 / Temperature (1/K)
x 10^-3
1F: Residuals for Monte Carlo Arrhenius (10000 fits)
DCp zero
DCp non-zero
Ln(k / k_Ref)
1 / Temperature (1/K)
x 10^-3
1B: Monte Carlo Arrhenius (10000 fits)
DCp zero
DCp non-zero
Ln(k / k_Ref)
1 / Temperature (1/K)
x 10^-3
1B: Residuals for Monte Carlo Arrhenius (10000 fits)
DCp zero
DCp non-zero
Ln(k / k_Ref)
1 / Temperature (1/K)
x 10^-3
2F: Monte Carlo Arrhenius (10000 fits)
DCp zero
DCp non-zero
Ln(k / k_Ref)
1 / Temperature (1/K)
x 10^-3
2F: Residuals for Monte Carlo Arrhenius (10000 fits)
DCp zero
DCp non-zero
Ln(k / k_Ref)
1 / Temperature (1/K)
x 10^-3

Figure S4 (continued)

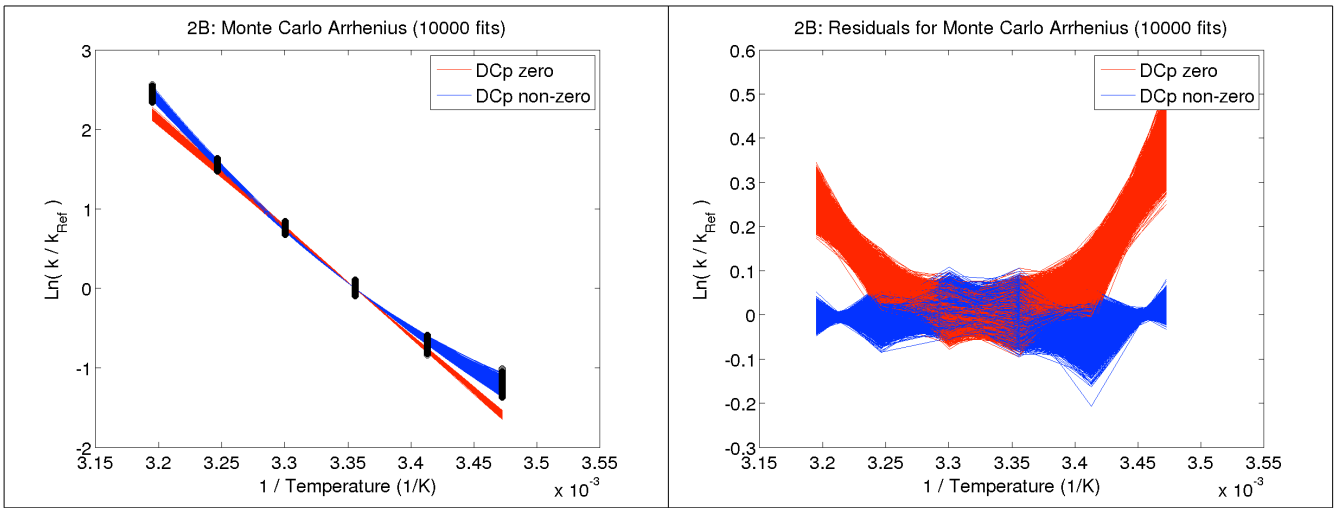


Figure S4. Arrhenius analyses of each of four kinetic rates extracted from the model A [BB] supports non-zero $\Delta C_p^{\ddagger}$ for steps 1F and 2B, whereas steps 1B and 2F are adequately described with zero $\Delta C_p^{\ddagger}$. However, we elected to use non-zero $\Delta C_p^{\ddagger}$ values for all four steps to retain consistency with both the principle of microscopic reversibility and with results from van't Hoff analysis (Fig S5). The Monte Carlo analysis produced two fits to each of 10,000 simulated datasets generated from re-sampling the distribution of residuals; the set of simulated datasets estimates the uncertainty in the measured values on the y-axis (i.e., $\ln(k_1^F / k_1^F{}_{Ref})$, etc.). The fit with $\Delta C_p^{\ddagger}$ set to zero (red line) produces a straight line with zero curvature, and yields curved residuals for steps 1F and 2B (right plots). This curvature is reduced when $\Delta C_p^{\ddagger}$ is non-zero (blue line), but the reduction in residuals is much more significant for steps 1F and 2B than for steps 1B and 2F.

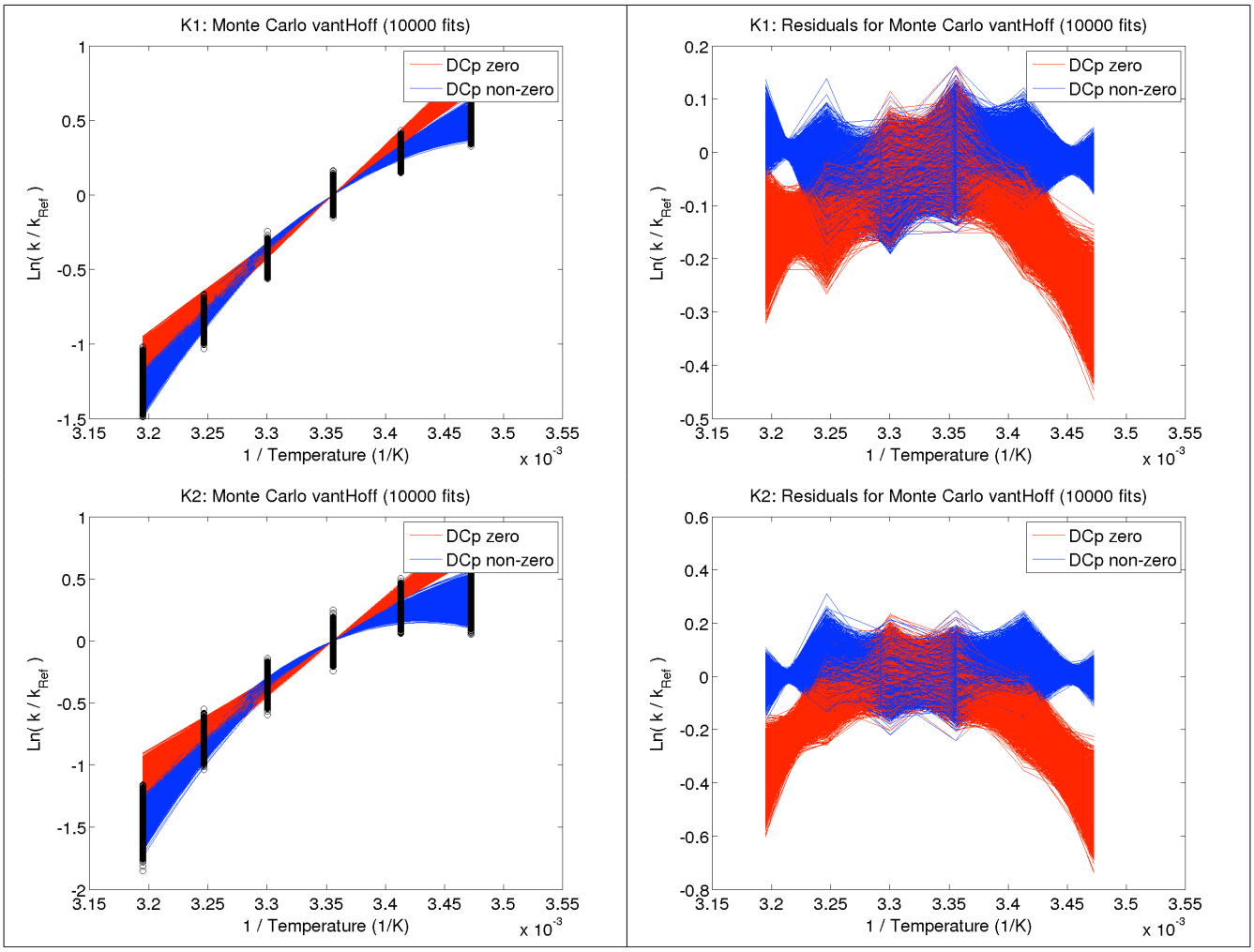


Figure S5. van't Hoff analyses of the two equilibrium constants extracted from the model A [BB] supports non-zero ΔC_p for steps 1 and 2. The Monte Carlo analysis produced two fits to each of 10,000 simulated datasets generated from re-sampling the distribution of residuals; the set of simulated datasets estimates the uncertainty in the measured values on the y-axis (i.e., $\ln(K_1/K_{1Ref})$ and $\ln(K_2/K_{2Ref})$). The fit with ΔC_p set to zero (red line) produces a straight line with zero curvature, and yields curved residuals for steps 1 and 2 (right plots). This curvature is reduced when ΔC_p is non-zero (blue line).

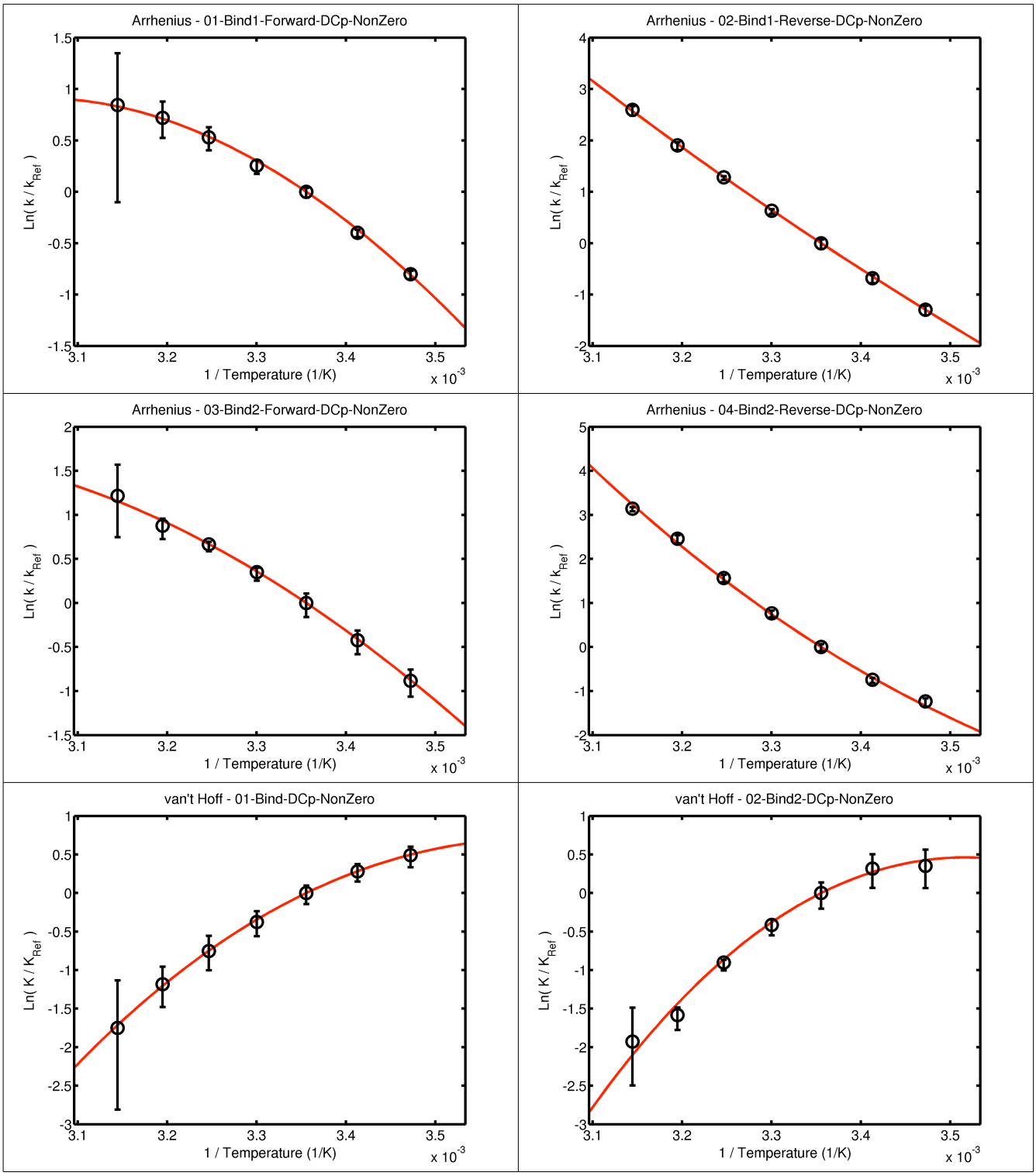


Figure S6. Data at 45°C were not used in the temperature-dependent Arrhenius and van't Hoff analyses because the two forward rate constants yielded Arrhenius and van't Hoff quantities that were outlying

with large uncertainties (45°C is the left-most point in each figure). The red line shows a fit to either the extended Arrhenius or the extended van't Hoff equation (i.e., $\Delta C_p^{\ddagger}$ or ΔC_p is non-zero). Error bars are 95% confidence intervals from a Monte Carlo error analysis propagated from the original stopped flow time courses.

Supporting Tables

Table S1. Sum of squared error (SSE) for comparing models A [BB] and B [BI].

		15ºC	20ºC	25ºC	30ºC	35ºC	40ºC	45ºC	All (15-45ºC)
SSE	**Model A [BB]**	1.000	1.000	1.000	1.000	1.000	1.000	1.002	7.002
	Model B [BI]	1.527	1.928	2.324	2.665	2.100	1.288	1.000	12.832

Model A [BB] is favored over model B [BI] because it has the same number of fitting parameters and its fits exhibit lower SSE at 15-40ºC, whereas it has equivalent SSE to model B [BI] at 45ºC. SSE values are normalized within each temperature so that the sum in the right-most column weights each temperature equally.

Table S2. Fitted parameters to k^{obs} vs. TRAP concentration.

	K^{obs} From Two-Phase Exponential	K^{obs} From Three-Phase Exponential
15ºC	k_1^f = 0.23 [0.06] k_1^b = 1.13 [0.30] K_d = 5.13 [3.07] k_2^f = 0.22 [0.04] k_2^b = 0.03 [0.02]	k_1^f = 0 [Undefined] k_1^b = 4.31 [1.30] K_{d2} = 7.59 [30.96] k_2^f = 1.37 [2.17] k_2^b = 0.76 [0.48] K_{d3} = 0.75 [3.39] k_3^f = 0.12 [0.19] k_3^b = 0.002 [0.22]
20ºC	k_1^f = 0.36 [0.05] k_1^b = 1.95 [0.28] K_d = 3.56 [0.90] k_2^f = 0.27 [0.01] k_2^b = 0.04 [0.01]	k_1^f = 0.33 [0.22] k_1^b = 3.49 [1.25] K_{d2} = 2.91e5 [1.41e10] k_2^f = 6.59e5 [3.19e9] k_2^b = 0.71 [0.92] K_{d3} = 2.40 [0.60] k_3^f = 0.18 [0.01] k_3^b = 0.04 [0.01]
25ºC	k_1^f = 0.52 [0.04] k_1^b = 3.64 [0.22] K_d = 3.50 [2.32] k_2^f = 0.33 [0.06] k_2^b = 0.07 [0.05]	k_1^f = 0.46 [0.15] k_1^b = 5.16 [0.83] K_{d2} = 76.05 [4.57e3] k_2^f = 9.50 [4.98e2] k_2^b = 1.35 [1.90] K_{d3} = 2.67 [4.17]

		k_3^f = 0.27 [0.07] k_3^b = 0.08 [0.09]
30ºC	k_1^f = 0.75 [0.08] k_1^b = 6.54 [0.44] K_d = 4.14 [1.17] k_2^f = 0.71 [0.05] k_2^b = 0.12 [0.03]	k_1^f = 0.84 [0.20] k_1^b = 7.72 [1.17] K_{d2} = 6.23 [16.01] k_2^f = 6.56 [8.33] k_2^b = 0 [Undefined] K_{d3} = 2.56 [3.74] k_3^f = 0.54 [0.13] k_3^b = 0.06 [0.16]
35ºC	k_1^f = 0.89 [0.30] k_1^b = 13.57 [1.73] K_d = 7.59 [3.53] k_2^f = 1.22 [0.22] k_2^b = 0.32 [0.05]	k_1^f = 0 [Undefined] k_1^b = 24.55 [6.46] K_{d2} = 0.25 [35.07] k_2^f = 2.86 [2.71e2] k_2^b = 2.73 [2.77e2] K_{d3} = 2.07 [3.68] k_3^f = 0.70 [0.21] k_3^b = 0.18 [0.30]
40ºC	k_1^f = 1.47 [0.59] k_1^b = 22.12 [3.39] K_d = 3.69 [4.60] k_2^f = 1.48 [0.39] k_2^b = 0.54 [0.29]	k_1^f = 0 [5.91] k_1^b = 38.55 [33.70] K_{d2} = 0 [Undefined] k_2^f = 0 [Undefined] k_2^b = 13.21 [10.38] K_{d3} = 27.00 [2.20e2] k_3^f = 2.30 [13.29]

		k_3^b = 0.75 [0.34]
45ºC	k_1^f = 1.59 [1.48] k_1^b = 46.09 [8.45] K_d = 26.13 [264.33] k_2^f = 3.19 [22.62] k_2^b = 1.66 [0.61]	k_1^f = 0 [Undefined] k_1^b = 79.21 [30.24] K_{d2} = 0.03 [1.06e8] k_2^f = 9e-6 [3.44e4] k_2^b = 12.00 [3.44e4] K_{d3} = 0.45 [0.62] k_3^f = 1.74 [0.43] k_3^b = 0 [Undefined]

Parameters from fitting the concentration-dependence of the mechanism-independent apparent rate constants to either a two-phase or three-phase exponential equation. Fit results are shown in Fig S2. The units of each number are as follows: k_1^f, k_2^f, and k_3^f are /μM/sec, k_1^b, k_2^b, and k_3^b are /sec, and K_d, K_d^2, and K_d^3 are μM. The values in brackets are standard fitting errors from MATLAB, where "Undefined" means MATLAB could not successfully determine the fitted value and/or its error.

Table S3. Bayesian Information Criterion (BIC) for comparing fits of stopped-flow data to 2-, 3-, and 4-phase exponential equations.

Name	2-Exp BIC Weight	3-Exp BIC Weight	4-Exp BIC Weight	Best Model
15C-t1	0.062	0.935	0.003	3
15C-t2	0.071	0.926	0.003	3
15C-t3	0	0.974	0.026	3
15C-t4	0	0.997	0.003	3
15C-t5	0	0.997	0.003	3
15C-t7	0	0.997	0.003	3
20C-t1	0.955	0.045	0	2
20C-t2	0.236	0.759	0.004	3
20C-t3	0.009	0.98	0.011	3
20C-t4	0	0.839	0.161	3
20C-t5	0	0.963	0.037	3
20C-t6	0	0.959	0.041	3
20C-t7	0	0.93	0.07	3
25C-t1	0.976	0.024	0	2
25C-t2	0.958	0.042	0	2
25C-t3	0.005	0.989	0.006	3
25C-t4	0	0.993	0.007	3
25C-t5	0	0.974	0.026	3
25C-t6	0	0.517	0.483	3
25C-t7	0	0.103	0.897	4
30C-t1	0.884	0.116	0	2
30C-t2	0.667	0.091	0.243	2
30C-t3	0	0.996	0.003	3
30C-t4	0.002	0.956	0.042	3
30C-t5	0	0.938	0.062	3
30C-t6	0	0.974	0.026	3
30C-t7	0	0.997	0.003	3
35C-t1	0.574	0.424	0.002	2
35C-t2	0.103	0.894	0.003	3
35C-t3	0.006	0.994	0	3
35C-t4	0.001	0.983	0.016	3
35C-t5	0	0.922	0.078	3
35C-t6	0	0.862	0.138	3
35C-t7	0	0.587	0.413	3
40C-t1	0.993	0.007	0	2
40C-t2	0.983	0.017	0	2
40C-t3	0.993	0.007	0	2
40C-t4	0.007	0.99	0.003	3
40C-t5	0.026	0.966	0.009	3
40C-t6	0.002	0.99	0.008	3
40C-t7	0	0.992	0.007	3
45C-t1	0.985	0.015	0	2
45C-t2	0.574	0.425	0	2
45C-t3	0.234	0.764	0.003	3
45C-t4	0.001	0.985	0.015	3
45C-t5	0.021	0.97	0.009	3
45C-t6	0.002	0.987	0.011	3

The 3-phase model is generally preferred as it has the largest BIC at each temperature and TRAP concentration. Fits are shown in figure S1.

References

1. Kuzmic, P. (1996) Program dynafit for the analysis of enzyme kinetic data: application to hiv proteinase, *Anal. Biochem.* 237, pp. 260-273.

2. Kuzmic, P. (2009) Dynafit--a software package for enzymology., *Meth. Enzymol.* 467, pp. 247-280.