Supporting Information

Volkán-Kacsó and Marcus 10.1073/pnas.1611601113

Estimation of λ and $\alpha(0)$

Here we provide a summary of the procedure described in ref. 16 to calculate λ and $\alpha(0)$ from available free rotation, stalling, and ensemble biochemical experimental data. In equation 12 of ref. 16 $\alpha(0)$ [i.e., $\alpha(\theta)$ at $\theta=0$], and in equation 18 λ is given as a function of three energy quantities, shown in bold in the energy diagram in Fig. S1: the work term for the weak binding in the reactant (R) state, W^r, the free energy barrier in the collision-based theory of binding, $\Delta G_0^\dagger$, and the standard free energy of reaction defined relative to the R state, ΔG_0^0. These quantities involved in the binding of ATP are defined relative to the states given in Fig. S1, a state (S) in which the ATP molecule is found in the solution, a reactant state (R), which involves the weak binding of ATP at the entrance site, the transition state (TS), and a product state (P) in which the ATP occupies the binding pocket. We now summarize the procedure to estimate the value of the three quantities, also given in Fig. S1.

i) The W^r is estimated to be −9.1 kcal/mol for ATP binding, calculated from the binding affinity of −6.3 kcal/mol to the empty open-conformation subunit [from a dissociation constant (25) of 25 μM] minus an entropic contribution to the free energy change $kT\ln kT/hZ=2.8$ kcal/mol.

ii) A $\Delta G_0^\dagger=5.0$ kcal/mol estimate is yielded by collision theory, assuming a collision frequency of $Z=10^{11}$ $\mathrm{M^{-1}\cdot s^{-1}}$. The free rotation rate constant for ATP binding is $k_{f0}=2\times10^7$ $\mathrm{M^{-1}\cdot s^{-1}}$ (1), which is used in equation 8 of ref. 16 to give the above estimate for $\Delta G_0^\dagger$.

iii) The $\Delta G_0^0=-6.0$ kcal/mol estimate for ATP binding can be calculated, according to Fig. S1, by subtracting the binding affinity of −6.3 kcal/mol for the R state from the standard free energy of reaction defined relative to the solution state (S), $\Delta G_{s0}^0=-kT\ln K_0=-12.3$ kcal/mol (3). This method, based on Fig. S1 for evaluating ΔG_0^0, is simpler than that used in ref. 16.

Correcting the Reported Controlled Rotation Data

Here we describe a theoretical method to calculate corrections for the data reported by Adachi et al. (4) due to the effects of (*i*) missed events and (*ii*) replacing T_0 with T.

If θ_j is the rotor angle at the center of the jth interval, then, according to Adachi et al. (4), we use for the reported ("rep") rate constant estimates in the 36 intervals the values used in their analysis,

$$k_f^{\text{rep}}(\theta_j)=N(0\to1,j)/T, \qquad \textbf{[S1]}$$

$$k_b^{\text{rep}}(\theta_j)=N(1\to0,j)/T_1(j), \qquad \textbf{[S2]}$$

where $N(0\to1,j)$ and $N(1\to0,j)$ are the number of 0 → 1 and 1 → 0 events in the interval j, counted over the whole trajectory (j has a periodicity of 36). The reported equilibrium constant is then calculated as $K^{\text{rep}}=k_f^{\text{rep}}/k_b^{\text{rep}}$. The rate constants estimated using this procedure, obtained by digitizing figure 5 of ref. 4, are reproduced in Fig. S2. Interestingly, the rates for Cy3-ATP binding in the presence and absence of P_i are indistinguishable from each other within the scatter in the data, as seen in Fig. S2. So, we treat them as essentially identical and compare them later with the same theoretical prediction. A similar remark applies to Cy3-ADP in Fig. S2.

The reported uncorrected experimental rate estimates defined in Eqs. **S1** and **S2** are calculated, for sufficiently long trajectories, from the averaged values,

$$k_f^{\text{rep}}(\theta_j)=N(0\to1,j)/T=\frac{\langle N(0\to1)\rangle_j}{t}, \qquad \textbf{[S3]}$$

$$k_b^{\text{rep}}(\theta_j)=\frac{\langle N(1\to0)\rangle_j}{\langle t(\sigma=1)\rangle_j}, \qquad \textbf{[S4]}$$

where the average $\langle\cdot\rangle_j$ is calculated over a single interval j, and all quantities depend on τ, t, k_f, and k_b. The k_f and k_b are the actual θ_j -dependent rate constants, the subject of the present theoretical predictions. In Eq. **S3**, the error due to using T instead of T_0 is also taken into account in the theoretical calculation of uncorrected forward rates, following the procedure used for the experimental estimation of k_f^{rep}.

Assuming steady-state conditions, the probability of being in a state $\sigma=0$ or 1 at any time is given in terms of the rate constants as

$$p_0=k_b/(k_f+k_b),\quad p_1=k_f/(k_f+k_b). \qquad \textbf{[S5]}$$

The survival probabilities in states 0 and 1, denoted by S_0 and S_1, are both exponential, $S_0(t)=\exp(-k_ft)$ and $S_1(t)=\exp(-k_bt)$, respectively. The probability distributions that the states have lived a lifetime between 0 and t, that is, that it decayed before t, are $[1-S_0(t)]$ and $[1-S_1(t)]$. Their time derivatives, namely the associated probability densities of lifetimes, the so-called waiting time distributions $\rho_0(t)$ and $\rho_1(t)$, are given by

$$\rho_0(t)=k_f\exp(-k_ft),\quad \rho_1(t)=k_b\exp(-k_bt). \qquad \textbf{[S6]}$$

The probability densities ρ_0 and ρ_1 describe the statistics of times spent in the 0 and 1 states in the trajectories (Fig. S3). We define the local time t as the time relative to the time when the rotor angle reaches the beginning of an interval j. In experiment, the finite acquisition time of video frames and the shot noise limit the effective time resolution. As a consequence, states that seem to be of $\sigma=1$ but are shorter than $\tau=0.1$ s were discarded in the analysis of Adachi et al. (4). We denote by $p(\sigma_0,n)$ the probability of a succession of n events of changed occupancy occurring at times $\{t_1,t_2,\ldots\}$ during any given 10° interval, given the condition of the system's being in occupancy σ_0 at $t=0$. Similarly, the probability that these n events are missed is denoted by $p_{\text{miss}}(\sigma_0,n)$.

The statistics of binding and release events depend on parameters τ and t. In our calculations, τ and t are formally treated as variables and are subsequently assigned the constant values used by the experimentalists, as indicated earlier, so $p(\sigma_0,n)$ and $p_{\text{miss}}(\sigma_0,n)$ are probabilities that depend on these variables τ and t. These probabilities can be cast as expectation values with respect to τ_n s (defined as t_n-t_{n-1} in Fig. S3). As an example, for $n=2$ and $\sigma_0=0$, $p(0,2)=p_0\langle\Theta(t_2<t<t_3)\rangle$ and $p_{\text{miss}}(0,2)=p_0\langle\Theta(t_2<t<t_3)\Theta(\tau_2<\tau)\rangle$. Here, Θ is a generalized Heaviside function which is 1 if the argument is true and 0 otherwise, and $\langle\cdot\rangle$ is a notation for the integration of the enclosed argument with respect to all τ_n s.

In addition to events of the type $p_{\text{miss}}(\sigma_0,2)$, missed "cross-boundary events," denoted by $p'_{\text{miss}}(\sigma_0,2)$, also have a contribution. Missed cross-boundary events are transitions occurring in a bin with the previous or subsequent transition occurring in another interval, and which transitions are missed due to the time

spent between the two events is too short to be resolved. For example, one such event is when a $0 \to 1$ transition occurs during the jth bin, and the subsequent transition occurs in a subsequent interval, but the time spent in state 1 is too short (i.e., $< \tau$) to be detected. Such $p'_{\text{miss}}(\sigma_0, 1)$, formally written as $p_0 \langle \Theta(t_1 < t < t_2)\Theta(\tau_2 < \tau)\rangle$, leads to a convolution integrals. For example, for $\sigma_0 = 0$, the time-domain expression is $p(0,2|t) = p_0 \int_0^\infty \int_0^\infty \int_0^\infty \rho_0(\tau_1)\rho_1(\tau_2)\rho_0(\tau_3)\delta(t - \tau_1 - \tau_2 - \tau_3)d\tau_3 d\tau_2 d\tau_1$ and $p_{\text{miss}}(0,2|t) = p_0 \int_0^\infty \int_0^\tau \int_0^\infty \rho_0(\tau_1)\rho_1(\tau_2)\rho_0(\tau_3)\delta(t - \tau_1 - \tau_2 - \tau_3)d\tau_3 d\tau_2 d\tau_1$ and a Laplace transform ($t \to s$, $\tau \to u$) formally facilitates their evaluation. For $p(\sigma_0, n)$ the Laplace transform $t \to s$ is a formally simple expression (we use a tilde to denote the transform),

$$\begin{aligned}\tilde{p}(\sigma_0, n|s) &= \int_0^t p(\sigma_0, n|t)e^{-st}dt \\ &= p(\sigma_0)\left\langle \exp(-st_n)\frac{1 - \exp(-s\tau_{n+1})}{s}\right\rangle.\end{aligned} \quad \textbf{[S7]}$$

Using $t_n = \tau_1 + \ldots + \tau_n$, for $\sigma_0 = 0$, an analytic expression follows,

$$\tilde{p}(0, n|s) = p(0) \times \begin{cases} \left[\tilde{\rho}_0(s)\tilde{\rho}_1(s)\right]^{n/2}, & n \text{ even}, \\ \left[\tilde{\rho}_0(s)\right]^{(n+1)/2}\left[\tilde{\rho}_1(s)\right]^{(n-1)/2}, & n \text{ odd}, \end{cases} \quad \textbf{[S8]}$$

where $\tilde{\rho}_0(s) = k_f/(s + k_f)$ and $\tilde{\rho}_1(s) = k_b/(s + k_b)$. An expression for $p(1, n|s)$ results by analogy from Eq. **S8** (i.e., by interchanging states 0 and 1).

In the calculation of the probability of the missed events, for each short event an additional condition applies, which is formally achieved by introducing an additional variable. For the leading terms of $n = 2$, which provide the most contribution, a single additional variable is needed ($\tau \to u$), leading to a double Laplace transform,

$$\tilde{p}_{\text{miss}}(0, 2|u, s) = p_0\tilde{\rho}_0(s)\tilde{\rho}_1(s)u^{-1}(s + k_f)^{-1}. \quad \textbf{[S9]}$$

For the cross-boundary missed events,

$$\tilde{p}'_{\text{miss}}(0, 1|u, s) = p_0(s\,u)^{-1}\tilde{\rho}_0(s)\left[\tilde{\rho}_1(u) - \tilde{\rho}_1(s + u)\right]. \quad \textbf{[S10]}$$

Again, by analogy expressions for $\tilde{p}_{\text{miss}}(1, 2|s)$ and $\tilde{p}'_{\text{miss}}(1, 1|u, s)$ follow from Eqs. **S9** and **S10**.

The above Laplace-space expressions are then inverted by elementary inversion, facilitated by the use of symbolic computation tools (37). Of special importance are the one-event contributions,

$$p(0,1) = p_0 k_f/(k_f - k_b)\left[\exp(-k_b t) - \exp(-k_f t)\right], \quad \textbf{[S11]}$$

$$p(1,1) = p_1 k_b/(k_b - k_f)\left[\exp(-k_f t) - \exp(-k_b t)\right], \quad \textbf{[S12]}$$

and the two-event contributions,

$$p(0,2) = -\frac{k_f k_b^2}{(k_f + k_b)(k_f - k_b)^2} \times \left(e^{-k_f t} - e^{-k_b t} + k_f t e^{-k_f t} - k_b t e^{-k_f t}\right), \quad \textbf{[S13]}$$

$$p(1,2) = \frac{k_f^2 k_b}{(k_f + k_b)(k_f - k_b)^2} \times \left(e^{-k_f t} - e^{-k_b t} + k_f t e^{-k_b t} - k_b t e^{-k_b t}\right). \quad \textbf{[S14]}$$

The leading terms that effectively determine the probability of missed events within an interval are

$$\begin{aligned} p_{\text{miss}}(0,2) &= k_f k_b^2 e^{-k_f t - k_b 3\tau}(\mathrm{k_f} + k_b)^{-1}(k_f - k_b)^{-2} \\ &\left(e^{k_f 3\tau} - e^{k_b 3\tau} + k_f t e^{k_f 3\tau} - k_f t e^{k_b 3\tau}\right. \\ &\left. - k_b t e^{k_f 3\tau} + k_b t e^{k_b 3\tau} - k_f 3\tau\, \mathrm{e}^{k_f 3\tau} + k_b 3\tau\, e^{k_f 3\tau}\right). \end{aligned} \quad \textbf{[S15]}$$

For the cross-boundary missed events, the Laplace inversion yields for the leading terms

$$\begin{aligned} &p'_{\text{miss}}(0,1) + p'_{\text{miss}}(1,1) = \\ &\left[e^{-k_f t}e^{-k_b 3\tau} - e^{-k_b 3\tau} - \frac{k_f e^{-k_f t}}{k_f - k_b} + \frac{k_f e^{-k_b t}}{k_f - k_b}\right. \\ &\left. + k_f k_b e^{-k_b 3\tau}\left(\frac{1}{k_f k_b} + \frac{e^{-k_f(t-3\tau)}}{k_f(k_f - k_b)} - \frac{e^{-k_b(t-3\tau)}}{k_b(k_f - k_b)}\right)\right]. \end{aligned} \quad \textbf{[S16]}$$

To obtain the reported rates from Eqs. **S3** and **S4** we subtract the probability of missed events from the total probability of events,

$$\begin{aligned} k_f^{\text{rep}}(\theta) \cong t^{-1}[&p(0,1) + p(0,2) + p(1,2) \\ &- p_{\text{miss}}(0,2) - p'_{\text{miss}}(0,1) - p'_{\text{miss}}(1,1)], \end{aligned} \quad \textbf{[S17]}$$

$$\begin{aligned} k_b^{\text{rep}}(\theta) &\cong (tp_0)^{-1}\langle N(1 \to 0, \tau, t)\rangle_j \\ &= (tp_0)^{-1}[p(1,1) + p(1,2) + p(0,2) \\ &\quad - p_{\text{miss}}(0,2) - p'_{\text{miss}}(0,1) - p'_{\text{miss}}(1,1)]. \end{aligned} \quad \textbf{[S18]}$$

In the calculations we neglect higher-order contributions (i.e., terms where $n \geq 3$). This approximation is valid if t is smaller or comparable with the inverse of the rates, that is, $t \leq 1/k_f$ and $t \leq 1/k_b$. In our calculations it is always verified that this condition is met. We note that the definitions in Eqs. **S1** and **S2**, for sufficiently slow rotation, yield $k_f^{\text{rep}} < k_b^{\text{rep}}$, even if the actual rate constants have the opposite relation, $k_f > k_b$. Thus, the latter condition could effectively be masked in the experiments, but with the necessary corrections the actual rates can be recovered. As discussed in the text, the number of missed $0 \to 1 \to 0$ events is appreciable at low θ values, that is, at $\theta \sim -50°$, when k_b becomes large compared with $1/t$ and much larger than k_f, resulting in significant corrections for both k_f and k_b, as seen in Fig. S4. The same remark applies to $0 \to 1 \to 0$ events, when $k_f > 1/t$ and $k_f \ll k_b$, but these conditions are not reached in the experiments.

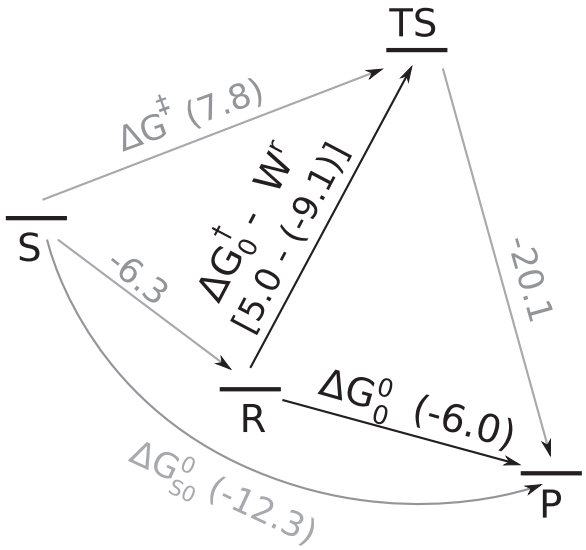

Fig. S1. Diagram of free energy changes (in units of kilocalories per mole) with values provided for the case of free rotation. The three quantities W^r, $\Delta G_0^{\dagger}$, and ΔG_0^0 that yield $\alpha(0)$ and λ are shown in bold. Other free energy terms used to calculate these three quantities are shown in light type.

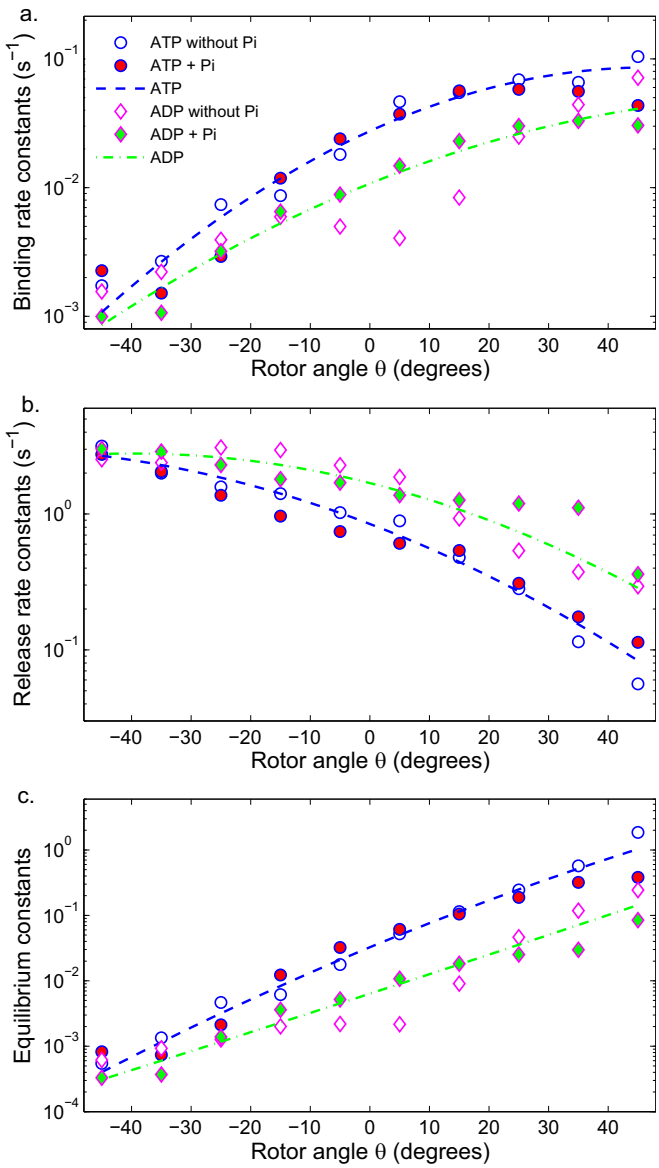

Fig. S2. Reported binding (*A*) and release (*B*) rate constants and equilibrium rate constants (*C*) vs. rotor angle θ in the range −50° to 50°, extracted by Adachi et al. (*4*) from controlled rotation experiments for the fluorescent nucleotides. Open symbols denote the nucleotide (circle for ATP and diamond for ADP) binding and release without P_i in solution, and closed symbols denote the presence of P_i in solution. The lines are a smooth fit to the ATP and ADP data points.

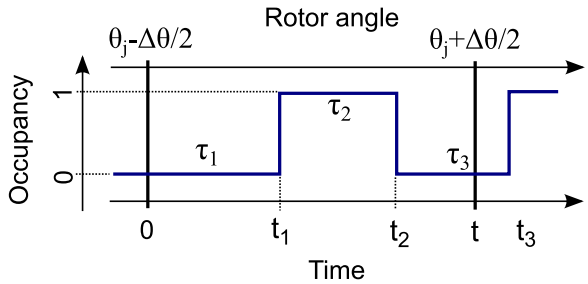

Fig. S3. Example of occupancy change event series described by probability $p(0, 2|t)$ (Eq. **S13**) and related text for definitions) in the jth interval corresponding to a rotation of constant rate from angle $\theta_j - \Delta\theta/2$ to angle $\theta_j + \Delta_\theta/2$. A two-state rate model with rate constants k_f and k_b describes the kinetics of jump events.

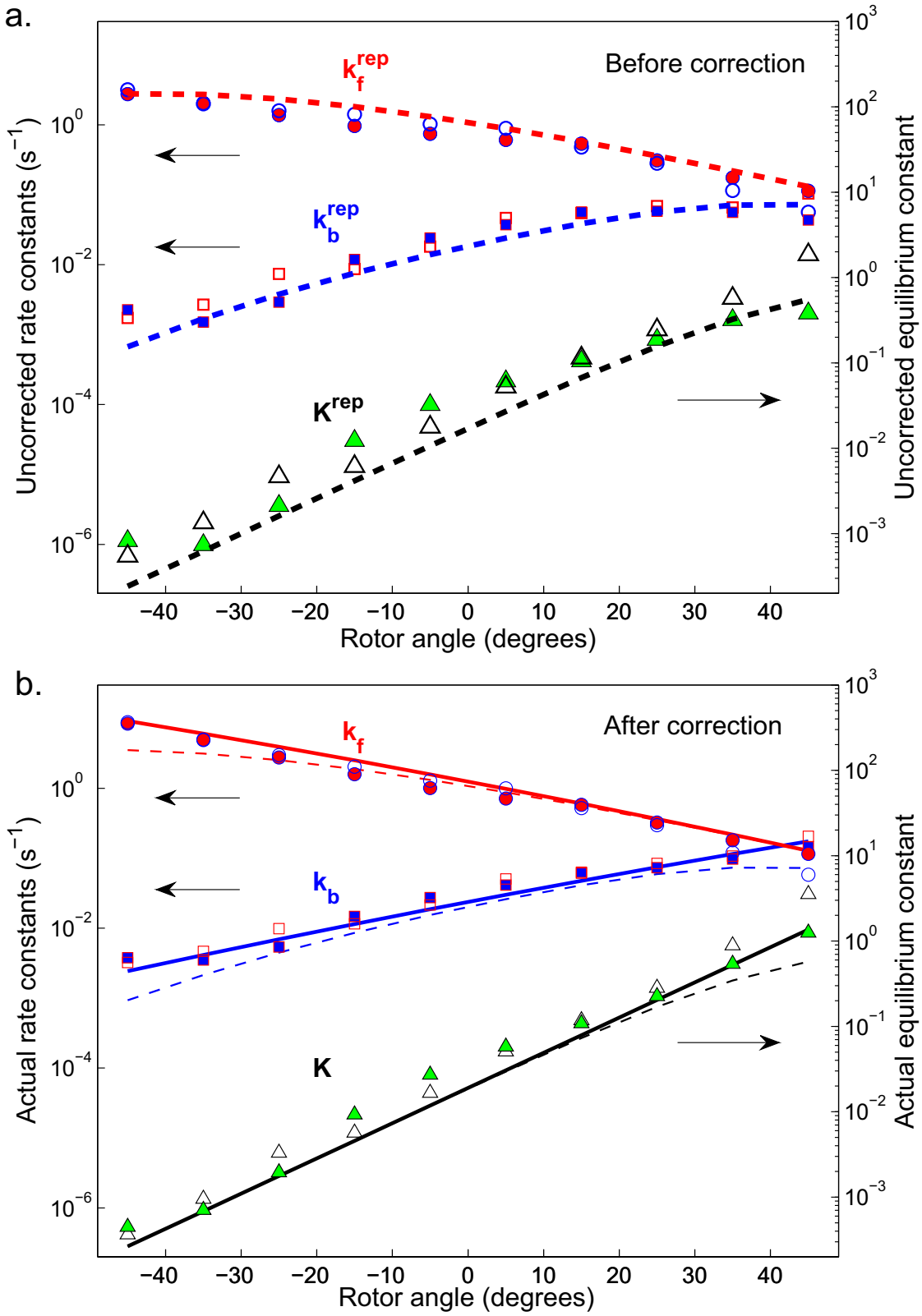

Fig. S4. Reported (*A*) and corrected (*B*) binding and release rate constants and equilibrium rate constants vs. controlled rotation angle for fluorescent ATP in the presence (solid symbols) and absence of P_i (open symbols) in solution. *B* is identical to Fig. 2 in the main text. The reported uncorrected experimental data are compared with their theoretical counterparts by calculating missed events and also correcting for an error due to replacing T_0 by T. Squares, circles, and triangles are experimental points, and solid lines are theoretical values. Dashed lines in *A* and *B* show the theoretical data without corrections.

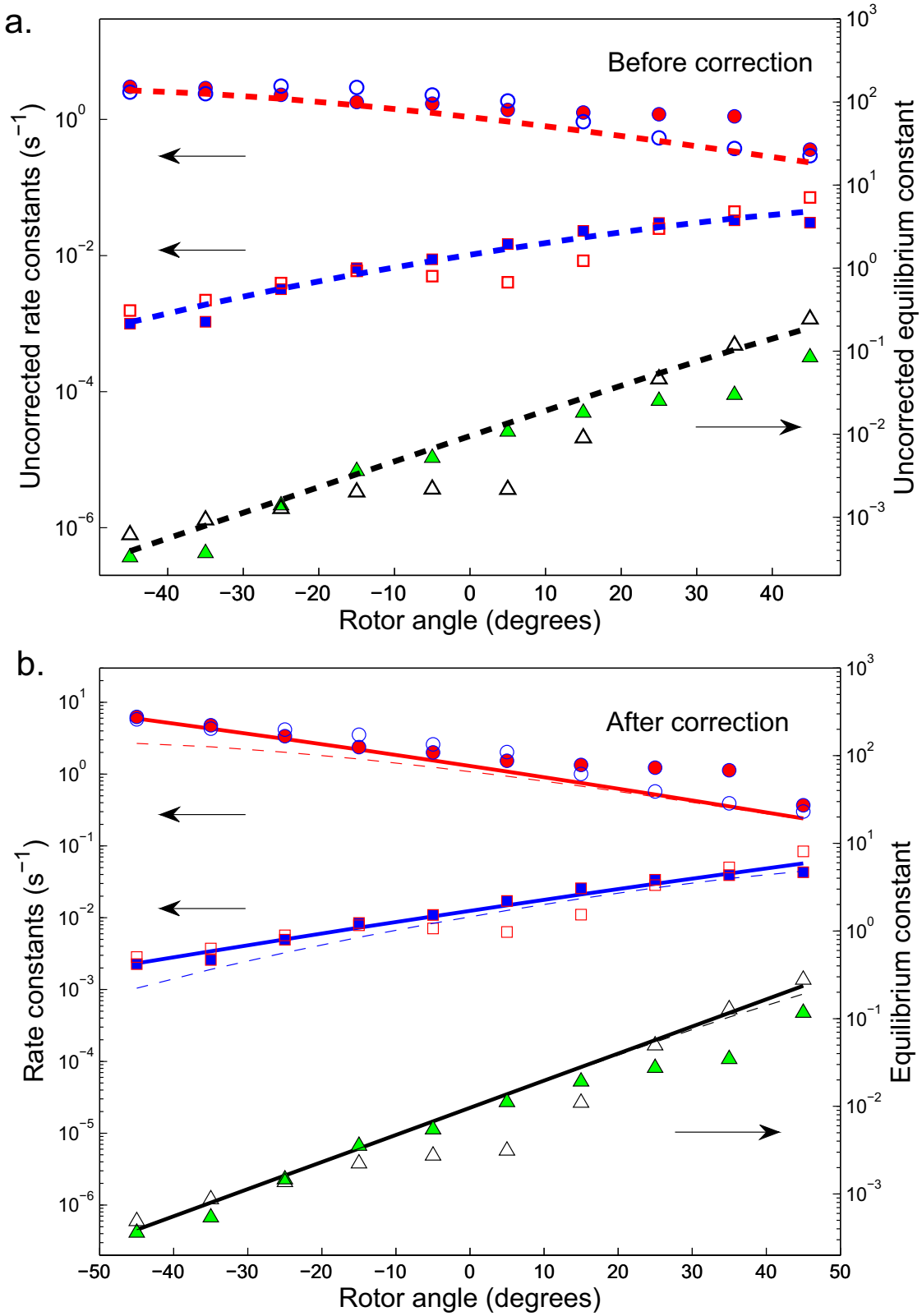

Fig. S5. Reported (*A*) and actual (*B*) binding and release rate constants and equilibrium rate constants for Cy3-ADP in the presence and absence of P_i in solution. *B* is identical to Fig. 3 in the main text. The theoretical curves for $\ln K^{rep}$ vs. θ were fitted to experiment, which yielded $\kappa^{ADP} = 12$ pN·nm/rad^2 and $k_f^{ADP}(0) = 4.6 \times 10^6$ M^{-1}·s^{-1} $= 0.5 \cdot k_f^{ATP}(0)$.