# Supporting Information (SI) for

# A Two-Step Method for smFRET Data Analysis

Jixin Chen,*[,†] Joseph R. Pyle,[†] Kurt Waldo Sy Piecco,[†] Anatoly B. Kolomeisky,[‡] Christy F. Landes[‡,§]

[†]*Department of Chemistry and Biochemistry, Ohio University, Athens, OH 45701*

[‡]*Department of Chemistry, Rice University, Houston, TX 77251*

[§]*Department of Electrical and Computer Engineering, Rice University, Houston, TX 77251*

**Corresponding Author**

* Email: chenj@ohio.edu

Contents

1. Trajectory simulation ..... S2
2. Data analysis ..... S3
3. Comparing our rate-constant analysis code with established software packages ..... S6
4. Mimicking the camera blurring effect ..... S7
5. Estimate the desired simulation length in postFRET ..... S7
6. The converging rate of postFRET simulations ..... S8
7. Estimate the errors in postFRET simulations ..... S8
8. Sequential-reaction model ..... S9
9. A four-state trajectory ..... S10
10. A "whack-a-mole" like searching algorithm ..... S10
11. A five-state trajectory ..... S12
12. A six-state trajectory ..... S13
13. A two-state trajectory ..... S14

References ..... S14

## 1. Trajectory simulation

**Monte Carlo transition simulation.** All simulations are carried out with a Dell Optiplex 9020 desktop computer equipped with an Intel i7 3.4 GHz CPU, 8 GB memory, and MATLAB 2014b with the Curve Fitting Toolbox. The MATLAB codes for the simulations have been included in a separate file.

The FRET trajectories are simulated with Monte Carlo method using several parameters. The key parameters include the number of states, FRET values of the states, transition rate constants between any states (a matrix), emission intensities, number of molecules, dye bleaching lifetime, simulation step time, binning time, *etc*. A few of these are listed in **Figure S1**, including the number of states, $N$, the transition rate constant between the state $i$ and $j$, $k_{ij}$, and the FRET values. The code is written in such a way that one can easily change any of the parameters to generate time trajectories for any number of states and any rate constant matrix. If a transition is irreversible, its back-transition rate can be set to 0. The molecule is assigned to have an initial state, e.g. state 1, and its statistical transition probability is calculated based on the sum of all rate constants from state 1:[1]

$$\boldsymbol{p_1 = 1 - e^{-\sum_{i=2}^{N} k_{1i}\Delta t}} \qquad (1)$$

where $\Delta t$ is the simulation time resolution in second. If the molecule is confirmed to change states at a certain step, the transition probabilities to each state $p_2$, $p_3$, ... $p_N$ are set to be proportional to the corresponding rate constant.[1]

$$\boldsymbol{p_{1i} = \frac{k_{1i}}{\sum_{j=1}^{j=N} k_{1j}}} \qquad (2)$$

where $p_{1i}$ is the probability of transferring to the $i^{th}$ state, $k_{1i}$ and $k_{1j}$ is the rate constant of state 1 to state $i$ and $j$ respectively, $k_{11}$ and $p_{11}$ is set to 0 for convenience ($\boldsymbol{k_{ii} \equiv 0}$ and $\boldsymbol{p_{ii} \equiv 0}$ do not have physical meanings).

Thus, one or two random numbers are used to determine whether each data point stays with its original value or changes to one of the other values in the next data point. For example, if the current data point is at state 1, and if the first random number $r_1$ (between 0 to 1) is $\boldsymbol{r_1 \leq p_1}$, then the next data point will be a state different than state 1. Another random number $r_2$ (between 0 to 1) is generate and is compared to an $N$-element vector $\boldsymbol{Vp(j)} = \boldsymbol{\sum_{i=1}^{i=j} p_{1i}}$, where $j$ is a vector $\{1,2,3...N\}$. Thus, $N$ state gives an $N\times N$ transition probability matrix. The range where $r_2$ locate in this vector determines its final state, e.g. if $\boldsymbol{Vp(2) \leq r_2 \leq Vp}(3)$, then the new state will be state 3. The simulation stops when the trajectory reaches the desired simulation time (dye bleaching time). After the simulation, the states are assigned to the FRET values. Then donor and acceptor photon counts are calculated for each FRET data point and binned into new trajectories. Redefine the probability matrix to add the probabilities of staying unchanged (set to $\boldsymbol{p_{ii} \equiv 0}$ now) will allow determining the states using one random number.

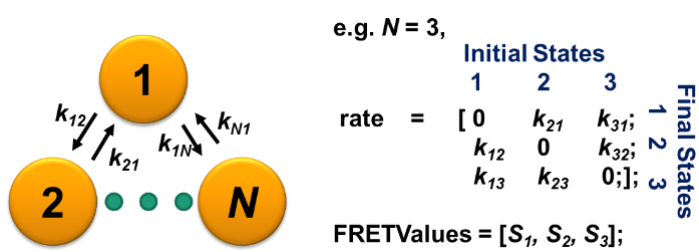


**Figure S1.** Input parameters for the Monte Carlo simulation. Assuming transitions happen between any two of the $N$ states, all the transition rate constants $k_{ij}$ >= 0 ($s^{-1}$), with 0 equivalent to transition forbidden. An $N\times N$ rate constant matrix is initialized. State FRET values $S_i$ are assigned to the $i^{th}$ state after simulation.

**Molecules and photobleaching.** Multiple molecules (e.g. 100) are simulated to mimic common experiments, with each dye-pair having a bleaching time that is exponentially distributed around the average value of the bleaching lifetime (typically set to 5 s). The initial state of each molecule is randomly distributed based on the ratio of the state lifetime over the sum of the lifetimes of all states, i.e. the equilibrium concentrations of the states. These 100 molecules are then combined into a single trajectory. The trajectory length of each molecule is stored so that the information can be used to exclude the first and the last state-transitions of every molecule. This exclusion is to avoid the edge effect of each trajectory (starting and stopping observation during the lifetime of a

state) and is performed after the state identification. Trajectories that have fewer than 2 transitions will thus be excluded from transition analysis, but they will still contribute to the state identification.

**Binning the data.** Each state is then split into donor and acceptor channels, with total photon counts set to be 20 at step time 1 ms. Noise is not considered in this work. The photon counts of the two channels are then binned. The binned FRET value is calculated from the binned photon counts of the two channels:[2]

$$FRET = I_A/(I_A+I_D) \qquad (3)$$

where $I_D$ and $I_A$ are the respective fluorescence intensities of the donor and acceptor molecules respectively.

## 2. Data analysis

Any of the established packages can be used to analyze the data. But for a fast demonstration, we wrote our own code to analyze the simulated noise-free smFRET trajectories. The accuracy of this rate constant analysis is the same as the established software package when dealing with noise free or low noise (signal-to-noise ratio >10) trajectories, with <5% deviation from HaMMy or SMART for non-binned or binned data (see below for details). Because our codes make use of the known FRET values, it is ultrafast for rate constant analysis. The detailed method is explained in the next few parts.

**State identification.** The nearest state assignment via a thresholding algorithm is adapted in this study. Briefly, if the number of states and the FRET values of the states are known, while the dwell times are not known, the data on the time trajectory can be assigned simply to its nearest state. For example, for 2 states at *FRET* values 0.9 and 0.1 with a 0.8 gap, 0.4 will be assigned to state 0.1, and 0.6 will be assigned to state 0.9. This state assignment method is extremely fast, and it is suitable to analyze data with no or relatively small noise to gap ratio.

**One-point-data state reassignment.** The method has been explained in vbFRET.[3] This option is switchable in postFRET but it should be consistent in both experimental data analysis and postFRET analysis. For a trajectory with slow transitions, the camera blurring effect can be corrected by reassigning any state that lasts only one data point,[3] e.g. if state A jumps to state B then to state C, and state B lasts only one data point, then merge state B into state A or C. Thus the questionable transition AB is now relatively more reliably reassigned to AC and questionable transition BC is removed.

**Rate constants analysis.** Once the states are identified, the next step is to determine the rate constant for each transition. The dwell-times of transitions from one state to other states are summarized in histograms (**Figure S2**). For each transition, the lifetime of the initial state (before transfer to another state) can be easily calculated by the average dwell-time (model-free) or fit the histograms with a first order exponential decay function (first order reactions as simulated). However, it is not trivial to obtain rate constants from these lifetimes, because $k_{nf} \neq 1/\tau_{nf}$, where $k_{nf}$ is the rate constant from state *n* to state *f*. The fitted decay time constants in **Figure S2**, for well-behaved distributions of single exponential decays, are rather the combination of multiple rate constants from each state $S_n$ to all other states $S_j$ based on the nature of parallel reactions:[1,4–6]

$$\boldsymbol{\tau_{nf}} = \frac{\mathbf{1}}{\sum_{j=1}^{j=N} \boldsymbol{k_{nj}}} \qquad (4)$$

where $\tau_{nf}$ is the observed lifetime of state $S_n$ that is transferring to state $S_f$, $k_{nj}$ is the true rate constant of the transfer from state $S_n$ to $S_j$, and *N* is the total number of states. This equation means that the transitions from a given state to all other states have the same time decay constant, even though the rate constants and the probabilities of these transitions are very different. For example in **Figure S2**, lifetimes from State 1 to 2 and state 1 to 3 are 0.64 s and 0.71 s respectively. Both are close to 1/(0.5+1) = 0.67 s. This confirms that our Monte Carlo simulation function works as expected. Only for a one-in and one-out situation such as two-state transition, this equation can be simplified to $k_{12} = 1/\tau_{12}$, and $k_{21} = 1/\tau_{21}$.

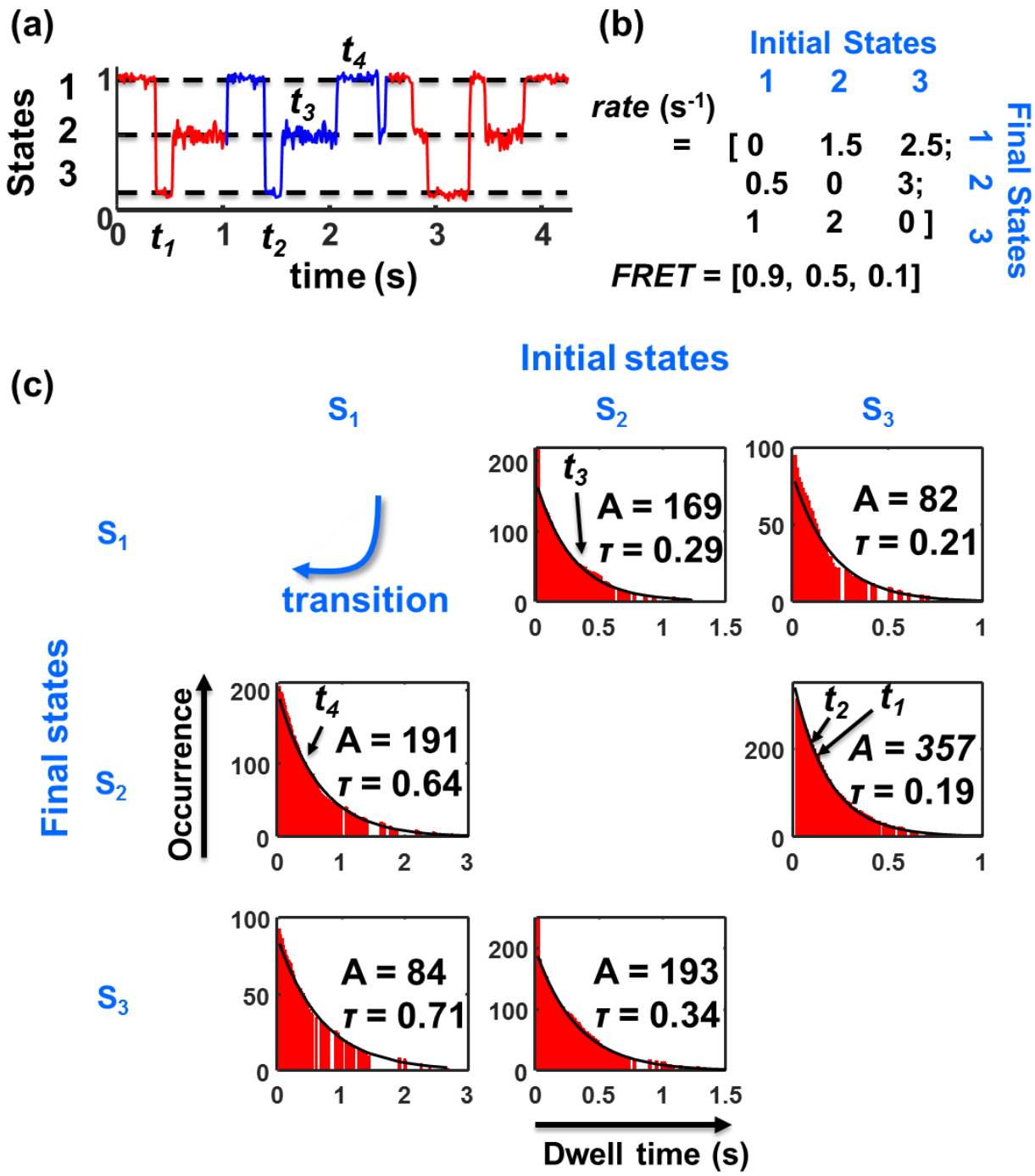


**Figure S2.** Scheme of summarizing state transitions. (A) The dwell times of the transitions are recorded (molecules are combined with alternative blue and red colors, a small Poisson noise added), e.g. $t_1$ is a transition from state $S_3$ to state $S_2$ with dwell-time $t_1$ for molecule #1; $t_3$ is a transition from state 2 to state 1 for molecule #2. (B) Simulation conditions with rate constants showed, 100 molecules, bleaching lifetime 5 s; total time ~460 s; time step 1 ms; binning time 10 ms; and FRET values 0.9, 0.5, and 0.1. (C) The dwell times are then summarized as a modified reverse cumulative distribution histograms of occurrence counts, e.g. $t_1$ and $t_2$ contribute to two occurrence counts of the transition from state $S_3$ to $S_2$ in their dwell time ranges. *The modification is made to remove the constant cumulative probability between two adjacent discrete dwell times as shown by the gaps between the bars.* This modification reduces the importance of the long dwell time data points during fitting comparing to regular cumulative distribution. Because first-order reactions are simulated, the distribution of these histograms exponentially decays over the dwell time, $y = Ae^{-x/\tau}$. The lifetime of each transition is the average dwell time of the histogram, which is represented by the decay constant $\tau$; and the total occurrence is the $y$-intercept which is represented by the pre-exponential factor $A$. Note: the last 2% counts on the longer dwell time side (tail) are excluded from the fitting because of its statistical insignificance (rare events, represented by the lower density bars near the tails).

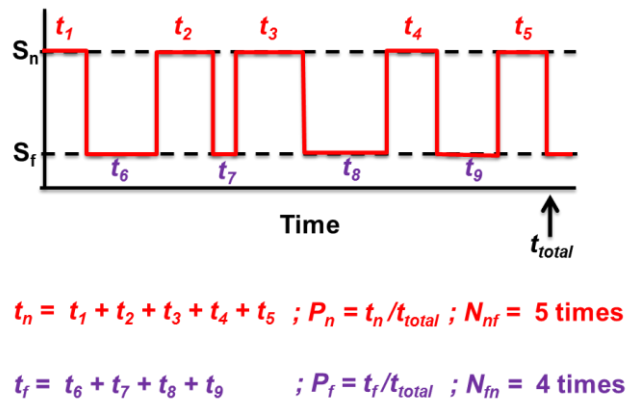


**Figure S3.** Scheme of counting the transition number for two-state transitions. The dwell times are summed to 2 subtotal of times ($N_{nf}$ and $N_{fn}$) for the two states $S_n$ and $S_f$. The probability of each state is the subtotal time over total experimental time $t_{total}$. The number of transitions is then counted.

We then connect the smFRET trajectories to macroscopic ensemble results using simple chemical-kinetic arguments. The fraction of each state in the ensemble is linearly proportional to the probability of observing this state in single-molecule experiments, which is given by the ratio of the dwell time of this state over the sum of the dwell times of all states, i.e., the effective total experimental time (**Figure S3**).

$$\boldsymbol{P_n} = \frac{t_n}{t_{total}} = \frac{\sum_{f=1}^{f=N} t_{nf}}{t_{total}} \qquad (5)$$

where $t_{nf}$ is the dwell-time for transition events from the state $S_n$ to $S_f$. Because these parallel transitions have been assumed to be first order reactions, the overall rate of each state is equal to

$$\boldsymbol{r_n} = \frac{dP_n}{dt} = -\sum_{f=1}^{f=N} \boldsymbol{k_{nf}}\, \boldsymbol{P_n} \qquad (6)$$

where $k_{nf}$ is the rate constant of the transition from state $S_n$ to one of the final states $S_f$. By the rate constant definition, the overall rate of each state is the number of transitions per time scale:

$$\boldsymbol{r_n} = \frac{dP_n}{dt} = -\frac{N_n}{t_{total}} = -\frac{\sum_{f=1}^{f=N} N_{nf}}{t_{total}} \qquad (7)$$

where $N_n$ is the number of transition events from state $S_n$ to all other four states, $N_{nf}$ is the number of transition events from state $S_n$ to $S_f$.

From Equation 4 to Equation 7, one obtains

$$\sum_{f=1}^{f=N} \boldsymbol{k_{nf}} \equiv \frac{1}{\tau_{ni}} = \frac{N_n}{t_n} = \frac{\sum_{f=1}^{f=N} N_{nf}}{\sum_{f=1}^{f=N} t_{nf}} \qquad (8)$$

For parallel first order reactions, the probability of the transitions always obeys

$$\frac{k_{ni}}{k_{nj}} = \frac{N_{ni}}{N_{nj}} \qquad (9)$$

where $n$ represents the initial state and $i$ or $j$ represents one of the final states. From Equation 8 and 9, the rate constant from state $S_n$ to any other state $S_f$ is

$$\boldsymbol{k_{nf}} = \frac{N_{nf}}{\sum_{f=1}^{f=n} t_{nf}} = \frac{N_{nf}}{t_n} \qquad (10)$$

Equation 10 is a reasonable result because the rate constant of each transition $S_n$ to $S_f$ is just the frequency of the transitions within the effective subtotal observation time for the state of $S_n$. Using Equation 10, any individual transition rate constant $k_{nf}$ can be evaluated from the simulated data.

The same Equation 10 can be obtained with another simple argument based on the ergodic principle. It is worth noting that the ergodicity of an individual molecule is broken due to the limited time of each measurement.[7] Because a given single molecule is statistically the same as all other molecules in the simulation, we can assume

that the smFRET trajectory is a random combination of the steady-state FRET efficiencies of multiple molecules with each molecule occupying a very short duration time $\delta t'$. Thus, the time trajectory of one molecule is converted to many molecules that are measured at the same time. So the dwell time $t_n$ represents the total number of molecules $N_n' = t_n/\delta t'$ at state $S_n$. So by the ensemble definition, for a first order reaction, the transition rate from the state $S_n$ to another state $S_f$ is the number of molecules that transfer during the time $\delta t'$:

$$\boldsymbol{r_{nf}} = -\frac{\boldsymbol{N_{nf}}}{\boldsymbol{\delta t'}} = -\boldsymbol{k_{nf}}\boldsymbol{N_n}' \qquad (11)$$

Substituting $N_n' = t_n/\delta t'$ into Equation 11 leads to Equation 10. The value $t_n/t_{total}$ represents the equilibrium concentration of the state $S_n$ in all molecules.

The fitting in **Figure S2** is a method to obtain the average lifetimes of first order transitions. The concept in Equation 10 should be general to other conditions. Thus, the average dwell time is still provided in the analysis for less statistically significant data when the number of transitions is small and fitting is less reliable than averaging. The error of such a treatment, however, has to be statistically evaluated with multiple simulations. With Equation 10, it becomes simple to calculate the rate constants of the transitions in **Figure S2**:

$$\boldsymbol{k_{nf}} = \frac{\boldsymbol{A_{nf}}}{\sum_{f=1}^{f=N} \boldsymbol{A_{nf}\tau_{nf}}} \qquad (12)$$

where $A$ and $\tau$ are the fitting parameters shown in **Figure S2**, representing the number and the lifetime of the transitions respectively.

## 3. Comparing our rate-constant analysis code with established software packages

The performance of the MATLAB code for rate constant analysis is compared to HaMMy and vbFRET for both non-binned trajectories and binned trajectories. The average error on rate constants (the weighted L1 norm explained in the main text) $wL1_{AT}$ scores have been listed in **Table S1**. The relatively small values and variations (<5%, highlighted in red) between the different methods confirm that both our trajectory simulation codes and the rate constant analysis codes work as designed. For the binned trajectory, larger variation is observed for different method due to camera blurring and the different state identification method employed.

The exponential decay fitting used in our method does correct some of the false transitions with very short dwell time. For example, in **Figure S2C**, the distributions from state 2 to state 1 and 3 have many false transitions that have a dwell time of just one data point. These transitions will be excluded by the fitting to bring down the number of transitions. However, for a false transition that has a longer dwell time and at the end of which a false state is assigned, this method is not able to correct it. Thus, the overall performance of our method is at the same level of those used in the other software packages that take the average dwell time of all transitions as the lifetime with no corrections on the number of transitions.

For a smFRET trajectory with relatively slow transition rate, the one-point-data reassigned method of our code works as well as that in vbFRET-C (**Table S2**). Both methods are able to reduce the average error on rate constants $wL1_{AT}$ <10%.

**Table S1**. An example of simulated smFRET trajectory analyzed by different methods for rate constants.

| | | **Preset values** | | | **HaMMy** | | | **vbFRET** | | | **Our codes** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | **1** | **2** | **3** | **1** | **2** | **3** | **1** | **2** | **3** | **1** | **2** | **3** |
| **Not binned (true)** | 1 | | 15[a] | 25 | | 14 | 26 | | 14 | 26 | | 15 | 25 |
| | 2 | 5 | | 30 | 5 | | 32 | 5 | | 29 | 5 | | 31 |
| | 3 | 10 | 20 | | 10 | 21 | | 10 | 20 | | 10 | 19 | |
| | ***wL1*$_{AT}$** | | **N/A** | | | **3.7%** | | | **2.3%** | | | **1.3%** | |
| **Binned[b]** | 1 | | | | | 18 | 11 | | 17 | 2 | | 18 | 14 |
| | 2 | | | | 9 | | 27 | 13 | | 41 | 11 | | 37 |
| | 3 | | | | 4 | 15 | | 1 | 12 | | 3 | 13 | |
| | ***wL1*$_{AT}$** | | | | | **42%** | | | **72%** | | | **52%** | |

[a]Rate constants unit in $s^{-1}$. True trajectory simulated at 1 ms time resolution. [b]Ten data points binned into one, resulting a time resolution 10 ms.

**Table S2**. A simulated trajectory analyzed with different software packages.

| | **True values**[a] | | | **HaMMy** | | | **SMART** | | | **vbFRET** | | | **vbFRET-C** | | | **Our codes with one-point-data reassigned** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 |
| 1 | | 1.63 | 2.64 | | 2.6 | 1.2 | | 2.6 | 1.2 | | 0.3 | 3.8 | | 1.7 | 2.4 | | 1.6 | 2.8 |
| 2 | 0.46 | | 2.96 | 0.9 | | 4.3 | 0.9 | | 4.1 | 0.1 | | 4.0 | 0.5 | | 2.9 | 0.5 | | 2.9 |
| 3 | 0.95 | 1.99 | | 0.5 | 2.9 | | 0.5 | 3.0 | | 1.3 | 3.7 | | 0.9 | 1.8 | | 0.9 | 1.9 | |
| ***wL*1**$_{AT}$ | **N/A** | | | **61%** | | | **60%** | | | **60%** | | | **8%** | | | **5%** | | |

[a]Results of non-binned data analyzed by vbFRET. Ten 1 ms data points binned into one 10 ms data point.

## 4. Mimicking the camera blurring effect

Selecting the simulation time resolution. How many data points should be binned to simulate the camera blurring effect? A few simulations (**Table S3**) indicate that two data point binned into one is enough to simulate the camera blurring effect for this set of rate constants. We choose 1 ms time resolution for all the rest simulations, which simulates the true transition point within each data point at 1/10 accuracy of the experimental time resolution, which is set to 10 ms. We speculate that transitions faster than this resolution will be unobservable during an experiment with a typical signal-to-noise ratio.

**Table S3**.The effect of simulation step resolution on rate constant analysis (binning time 10 ms).

| | **Preset values** | | | **10 ms (no binning)**[a] | | | **5 ms (2 in 1)** | | | **1 ms (10 in 1)** | | | **0.5 ms (20 in 1)** | | | **0.05 ms (200 in 1)** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 |
| 1 | | 15 | 25 | | 15 | 25 | | 16 | 13 | | 17 | 13 | | 18 | 12 | | 18[c] | 12 |
| 2 | 5 | | 30 | 5 | | 30 | 11 | | 42 | 11 | | 39 | 11 | | 38 | 10 | | 38 |
| 3 | 10 | 20 | | 10 | 20 | | 4 | 14 | | 4 | 15 | | 3 | 15 | | 4 | 15 | |
| ***wL*1**$_{AT}$ | **N/A** | | | **0%**[b] | | | **52%** | | | **48%** | | | **51%** | | | **48%** | | |

[a]The smFRET trajectories are simulated at given step resolution and then binned into 10 ms for each data point; total length 500 s; FRET values 0.9, 0.5, and 0.1; results are averages of 5 simulations. [b]The score is the rate constants analyzed from the binned trajectories referenced to the results from the non-binned trajectories which are ±1% of the preset values. [c]The standard deviation of each rate constant is <6% of its value.

## 5. Estimate the desired simulation length in postFRET

Use a given experimental rate constants as the initial guess to generate simulated smFRET trajectories at a different length (**Figure S4**). For each length, five simulations are carried out to obtain the standard deviations on the final rate constants from this set of initial guess. The fractions of the standard deviations to the resulting rate constants are taken as an estimation of the variation of the simulations.

The longer the trajectory length, the better the confidence level for the obtained rate constant with a computational time proportional to the trajectory length. A length that gives a variation within 3% to 10% is taken in this report. Thus, when using the simulated results to compare with the experimental results, the confidence on the simulation is roughly expected at the same level.

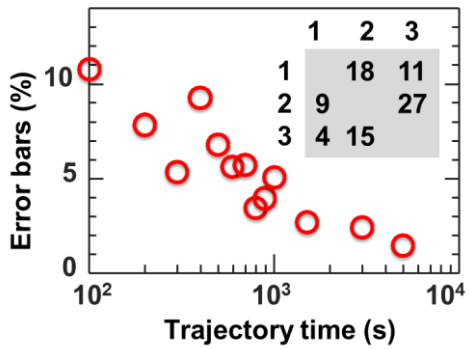


**Figure S4**. The standard deviation of five simulations over the resulting rate constants with the initial guess shown in the inset. The y-axis is the average values of the error bars for the six rate constants. The x-axis is the length of the smFRET trajectories for the simulations. FRET states 0.9, 0.4, and 0.1.

## 6. The converging rate of postFRET simulations

The curve of $wL1_{SE}$ scores vs the iteration numbers reflects the converging rate of the example postFRET simulations. Note that this $wL1_{AT}$ score is the guessed rate constants references to the true rate constant values. While in the postFRET simulation the $wL1_{SE}$ score is the analyzed results of simulated trajectories vs the analyzed results of the experimental trajectory.

The results (**Figure S5**) indicates that postFRET quickly reached a $wL1$ score of <5% within 3 iterations (red curve), which gives guesses of ~10% off the true values (blue curve). Then the $wL1$ scores vary around 4±1% (red curve). Post-analysis shows that the guesses vary around the true values with a standard deviation of ~2% (blue curve). Thus, we conclude that the converging rate of this example postFRET simulation is 3 iterations.

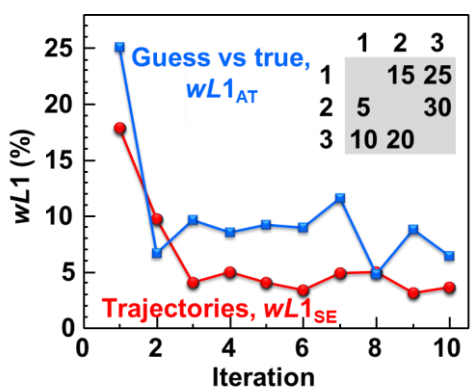


**Figure S5**. The converging rate of an example postFRET simulation. Red data show the $wL1_{SE}$ scores between the simulated postFRET trajectories and the simulated experimental trajectories. Blue data show the $wL1_{AT}$ scores between the best guess and the true values. The inset shows the true values in the $wL1_{AT}$ calculation. Simulation length of each trajectory is 1500 s. FRET states 0.9, 0.4, and 0.1.

## 7. Estimate the errors in postFRET simulations

We hypothesize that the average error bar of the guesses will estimate the error of postFRET analysis. The actual relationship, however, requires comprehensive study in the future. This hypothesis is consistent with the

simulation results (**Table S4**). For Monte Carlo simulated experimental data whose true rate constants are known, it is obvious that the uncertainty level of the results is 7% (**Table S4**). Assuming the true values are unknown, **Table S4** shows an estimated error ~6% from the standard deviations of several iterations. This value is consistent with our hypothesis. However, the accuracy of this estimation is subjected to the number of simulations and the accuracy of each Monte Carlo simulation.

**Table S4**. Estimated errors in postFRET of main text **table 1**.

| | **Preset true values** | | | **Our codes postFRET** | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 1 | 2 | 3 |
| 1 | | 15 | 25 | | 14.9±1.3[a] | 27.3±1.0 |
| 2 | 5 | | 30 | 5.8±0.4 | | 28.1±2.0 |
| 3 | 10 | 20 | | 9.5±0.2 | 19.1±1.2 | |
| **Average error bar** | | **N/A** | | | **6%** | |
| ***wL*1$_{AT}$** | | **N/A** | | | **7%** | |

[a]The error bar is the standard deviation of the guessing rate constant of the 5-10 iterations of the simulation (**Figure S5**), which gives an error bar 1.3/14.9 = 8.7%.

## 8. Sequential-reaction model

For a sequential reaction model instead of a network transition model (**Figure S6** inset), when the kinetics is fast enough, the existing methods have difficulties in analyzing the true rate constants. At simulation length 1000 s, ~5% standard deviations for 5 simulations (**Figure S6**), postFRET is able to mimic the trajectories and find acceptable matches at $wL1_{AT}$ ~ 2% (**Table S5**).

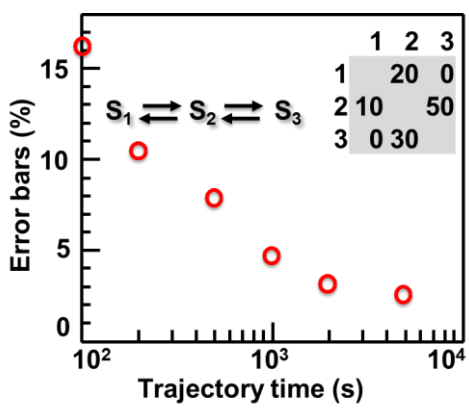


**Figure S6.** The standard deviation of ten simulations over the resulting rate constants with the preset values shown in the inset. Trajectories simulated at 1 ms time resolution and then binned at 10 ms. Transitions between state $S_1$ and $S_3$ are forbidden and the rate constants are set to 0 in the matrix. FRET states 0.9, 0.5, and 0.1.

**Table S5**. Data analysis results for a three-state smFRET trajectory (each trajectory length 1000 s).

| | Preset values | | | Our code one-step | | | Our code one-data-point removed one-step | | |
|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 |
| 1 | | 20 | 0[a] | | 18.7 | 2.7 | | 18.1 | 9.6 |
| 2 | 10 | | 50 | 8.4 | | 37.1 | 7.5 | | 30.4 |
| 3 | 0[a] | 30 | | 0.7 | 22.2 | | 2.0 | 21.2 | |
| ***wL*1$_{AT}$** | **N/A** | | | **19%** | | | **29%** | | |
| | **vbFRET** | | | **vbFRET-C** | | | **Our code two-step postFRET** | | |
| 1 | | 12.4 | 0.05 | | 11.7 | 3.3 | | 20.6 | 0.01 |
| 2 | 9.4 | | 39.2 | 8.6 | | 34.6 | 9.7 | | 50.0 |
| 3 | 0.01 | 10.7 | | 0.7 | 9.9 | | 0.03 | 29.9 | |
| ***wL*1$_{AT}$** | **32%** | | | **38%** | | | **2%** | | |

[a]The highlighted transitions are forbidden and the rate constant is set to zero. The *wL*1$_{AT}$ calculation excludes these rate constants.

## 9. A four-state trajectory

An experimental trajectory with preset rate constants shown in **Table S6** is simulated and analyzed by vbFRET and postFRET. A trajectory length of 1000 s is adapted during postFRET simulation. The standard deviation of the rate constants obtained with the initial guess is ~5% in this case (**Figure S7**). The *wL*1$_{AT}$ score for postFRET is 8%.

**Table S6**. Data analysis results for a four-state smFRET trajectory (1000 s).

| | Preset | | | | vbFRET | | | | vbFRET-C | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| 1 | | 15 | 15 | 15[a] | | 11.5 | 0.9 | 0.8 | | 7.6 | 8.3 | 7.2 |
| 2 | 15 | | 15 | 15 | 34.7 | | 34.7 | 33.1 | 11.1 | | 12.8 | 12.1 |
| 3 | 15 | 15 | | 15 | 0.9 | 11.9 | | 2.1 | 11.8 | 12.6 | | 11.2 |
| 4 | 15 | 15 | 15 | | 0.9 | 11.3 | 2.0 | | 7.1 | 8.1 | 7.4 | |
| ***wL*1$_{AT}$** | **N/A** | | | | **83%** | | | | **35%** | | | |
| | | | | | **Our codes[b]** | | | | **postFRET[c]** | | | |
| | | | | | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| 1 | | | | | | 16.4 | 7.7 | 3.6 | | 17.5 | 14.0 | 15.1 |
| 2 | | | | | 26.7 | | 22.7 | 13.9 | 14.4 | | 15.6 | 14.2 |
| 3 | | | | | 13.3 | 21.3 | | 24.3 | 15.5 | 11.4 | | 14.8 |
| 4 | | | | | 4.1 | 8.5 | 14.2 | | 13.1 | 16.9 | 16.1 | |
| ***wL*1$_{AT}$** | | | | | **42%** | | | | **8%** | | | |

[a]Highlighting one rate constant in different methods. [b]One-point-data reassignment option OFF. [c]Results are the averages of 5 simulations, each length is 1000 s.

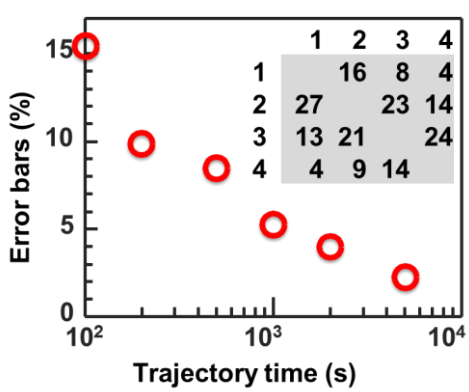


**Figure S7**. The standard deviation of five simulations over the resulting rate constants with the initial guess shown in the inset. FRET states are set to 0.8, 0.6, 0.4, and 0.2.

## 10. A "whack-a-mole" like searching algorithm

An experimental trajectory with four random FRET states and random rate constants is generated to further test the postFRET method (**Table S7** True).

A "whack-a-mole" like optimization algorithm (WAM) is adapted for this search, which is also suitable for other searches of more complicated systems shown later. The transition contributing the biggest to the $wL1_{SE}$ are identified, and an optimization of this rate constant is carried out to find the minimum $wL1_{SE}$ score. Because this algorithm alone may have difficulty to overcome local minimums and be locked in a few transitions during searching for a complicated model, an addition selection rule of simulated annealing is applied to the scanning. Because the outer transitions e.g. 1 to 4 or 4 to 1 are affected more by the camera blurring effect, the searching frequency is favorably applied to these transitions. The weight is set to abs($i$-$j$) for a transition i to j, meaning transition 1 to 4 is three times more likely to be searched than transition 1 to 2. First, the transition giving the biggest error is identified. Then the searching engine will decide that there is an $e^{-E/\Delta n}$ probability of adding an addition scan, where $E$ is set to 4 and $\Delta n$ is the normalized difference between the scanning history of the most frequently scanned and the least scanned transitions. If a random number generated is less or equal to the probability, a scan is carried out for the least scanned transition.

The experimental results are still used as the initial guess of the search which gives a set of rate constants that is ~35% off the experimental results ($wL1_{SE}$ score).The WAM searching procedure going on and $wL1_{SE}$ dropped to ~6% after 100 scans (**Figure S8**). The averages and standard deviations of this 100 scans are listed in **Table S7**. All averages are within the true values plus or minus one to three standard deviations except for the two transitions with substantially small rate constants than the others (highlighted), which is within 3.5 times the standard deviation. However, the right equilibrium constant between these two states is guessed.

**Table S7**. Data analysis results for a four-state smFRET trajectory (1000 s).

| | **True[a]** | | | | **vbFRET** | | | |
|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| 1 | | 1.3[b] | 8.5 | 15.2 | | 0.1 | 11.1 | 1.2 |
| 2 | 0.4[b] | | 14.1 | 6.8 | 0.0 | | 7.8 | 0.5 |
| 3 | 13.4 | 19.3 | | 2.6 | 22.9 | 30.9 | | 20.4 |
| 4 | 12.6 | 18.3 | 7.7 | | 1.1 | 1.2 | 13.1 | |
| ***wL*1$_{AT}$** | **N/A** | | | | **112%[c]** | | | |
| | **vbFRET-C** | | | | **postFRET** | | | |
| | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| 1 | | 7.7 | 0.6 | 8.8 | | 4.8±1.0 | 7.7±0.9 | 14.1±0.8 |
| 2 | 12.8 | | 24.8 | 11.1 | 1.5±0.4 | | 14.9±1.0 | 8.0±0.8 |
| 3 | 0.9 | 16.0 | | 0.2 | 11.6±0.9 | 16.9±1.2 | | 2. 5±0.4 |
| 4 | 7.8 | 13.7 | 0.6 | | 12.7±0.6 | 18.7±0.9 | 7.1±0.4 | |
| ***wL*1$_{AT}$** | **53%[c]** | | | | **6%[c]** | | | |

[a]FRET state 0.98, 0.88, 0.60, and 0.43. Highlighting one rate constant in different methods. [b]These two rate constants are excluded from the final *wL*1AT calculation because they are statistically rare during the 1000 s simulation. They contain larger errors than the others. [c]The value does not include the contributions of the highlighted transitions.

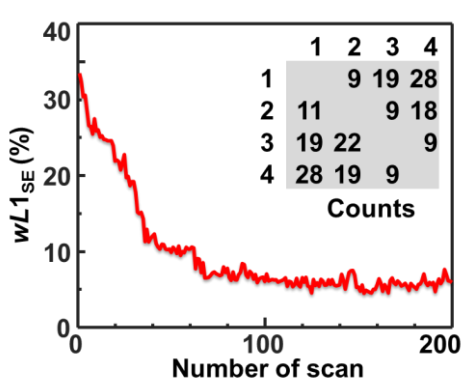


**Figure S8**. The $wL1_{SE}$ score history of a postFRET-WAM searching. The inset shows the times of the rate constants being searched.

## 11. A five-state trajectory

The preset values of the simulated experimental trajectory are listed in **Table S8**. vbFRET and vbFRET-C give analysis results contain 104% and 47% average errors. postFRET is able to reduce the error to 9%. The simulation length of postFRET is 500 s for each trajectory which is estimated to give an average variation ~10% for the rate constant analysis (**Figure S9**).

The sequential scanning method used in the three-state and four-state system requires very long time computation to work with the five-state system (a few days). While the WAM scanning algorithm reduces the searching to a few hours. The search optimizes the rate constants to acceptable values (**Figure S10**). Note that for such a complicated system, the one-point-data reassigned method is useful to reduce the time cost during searching.

**Table S8**. Data analysis results for a five-state smFRET trajectory (each trajectory length 500 s).

| | **Preset** | | | | | **vbFRET** | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| 1 | | 10 | 10 | 10 | 10 | | 14.0 | 1.4 | 3.8 | 0.8 |
| 2 | 10 | | 2 | 10 | 10 | 21.7 | | 10.9 | 14.3 | 7.1 |
| 3 | 10 | 10 | | 10 | 10 | 1.6 | 10.3 | | 9.8 | 1.6 |
| 4 | 5 | 10 | 10 | | 1 | 5.5 | 13.2 | 16.0 | | 17.3 |
| 5 | 10 | 10 | 10 | 0[a] | | 0.8 | 4.2 | 1.6 | 10.1 | |
| ***wL*1$_{AT}$** | | | **N/A** | | | | | **104%** | | |
| | **vbFRET-C** | | | | | **postFRET** | | | | |
| 1 | | 5.9 | 5.4 | 6.8 | 5.4 | | 9.1 | 10.0 | 9.4 | 10.7 |
| 2 | 7.8 | | 4.3 | 10.1 | 8.2 | 10.4 | | 3.4 | 8.8 | 8.6 |
| 3 | 4.8 | 6.1 | | 6.6 | 5.1 | 9.2 | 9.2 | | 11.4 | 9.8 |
| 4 | 6.9 | 10.3 | 9.4 | | 3.8 | 5.4 | 10.8 | 9.1 | | 0.6 |
| 5 | 5.5 | 6.7 | 5.1 | 2.6 | | 10.7 | 12.4 | 7.9 | 0.02 | |
| ***wL*1$_{AT}$** | | | **47%** | | | | | **9%** | | |

[a]The transition from State 4 to 5 is forbidden ($k_{45}$ = 0 s$^{-1}$ and the lifetime is infinity) and its rate constant is excluded from the *wL*1$_{AT}$ calculation. However since the analyzed rate constant is not zero for this transition (highlighted), it is still included in the *wL*1$_{SE}$ calculation of postFRET and the final guess is highlighted to approaching the true value zero. Rate constants have relatively larger deviation than others to the true values are highlighted (>20%).

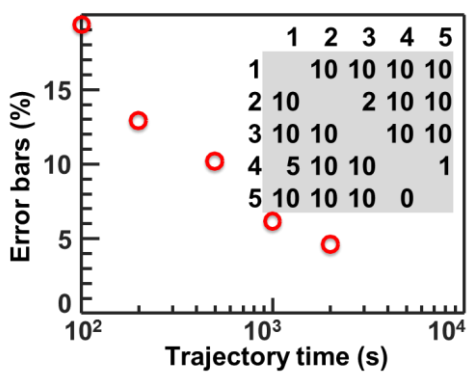


**Figure S9**. The standard deviation of five simulations over the resulting rate constants with the preset values in the inset. FRET states are set to 0.9, 0.7, 0.5, 0.3, and 0.1.

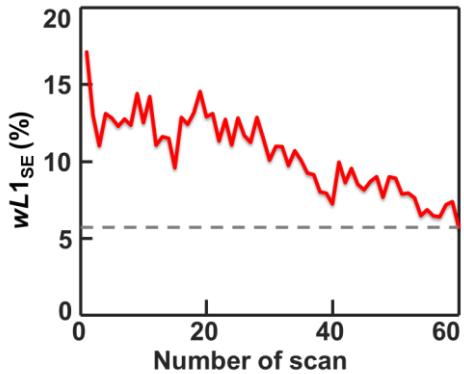


**Figure S10**. The wL1 score history of a "whack-a-mole" like search. The threshold is set to 6%. One-point-data-reassignment option is ON for experimental and simulated trajectories.

## 12. A six-state trajectory

For a system such complicated and with camera blurring effect (**Figure S11**), the best fitting results from vbFRET are not consistent with the real FRET states. HaMMy is able to fit the trajectory in six states relatively close (**Table S9**). This fitting yields a $wL1_{AT}$ score 41%. Our code with the one-point-data reassigned option turned ON yields a set of rate constants with the $wL1_{AT}$ score 23% (**Table S9**). The postFRET method with WAM searching mode and a searching trajectory length of 10000 s (~3% confidence, **Figure S11**), is able to correct the rate constants to $wL1_{AT}$ score 9% (**Table S9**). Note that for such a complicated system, the one-point-data reassigned method is critical during the search, and searching one by one works faster than WAM search.

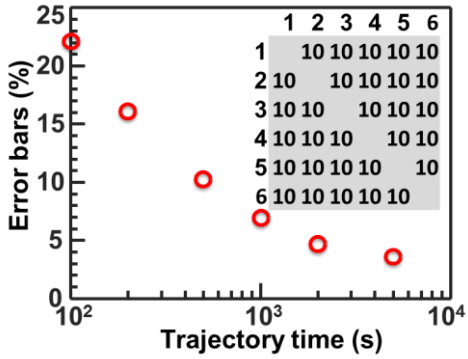


**Figure S11**. The standard deviation of five simulations over the resulting rate constants with the preset values in the inset. FRET states are set to 0.95, 0.8, 0.65, 0.5, 0.35, and 0.2.

**Table S9**. Data analysis results for a six-state smFRET trajectory.

| | Preset | | | | | | HaMMY | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 1 | 2 | 3 | 4 | 5 | 6 |
| 1 | | 10 | 10 | 10 | 10 | 10 | | 11.1 | 6.6 | 3.9 | 3.1 | 1.9 |
| 2 | 10 | | 10 | 10 | 10 | 10 | 15.3 | | 13.6 | 9.6 | 5.9 | 3.9 |
| 3 | 10 | 10 | | 10 | 10 | 10 | 11.0 | 15.1 | | 15.4 | 10.7 | 6.0 |
| 4 | 10 | 10 | 10 | | 10 | 10 | 5.6 | 11.1 | 15.5 | | 15.5 | 10.8 |
| 5 | 10 | 10 | 10 | 10 | | 10 | 3.7 | 5.8 | 9.4 | 13.8 | | 14.6 |
| 6 | 10 | 10 | 10 | 10 | 10 | | 1.9 | 2.8 | 4.1 | 7.1 | 11.3 | |
| ***wL*1$_{AT}$** | **N/A** | | | | | | **41%** | | | | | |
| | **One-point-data reassigned (our codes)** | | | | | | **postFRET[a]** | | | | | |
| 1 | | 6.2 | 6.9 | 6.1 | 7.3 | 5.4 | | 10.1 | 7.2 | 9.0 | 10.8 | 9.9 |
| 2 | 8.4 | | 10.3 | 7.3 | 7.2 | 8.1 | 10.8 | | 11.7 | 8.5 | 11.1 | 7.7 |
| 3 | 8.4 | 11.3 | | 11.7 | 9.9 | 8.3 | 9.5 | 10.8 | | 10.8 | 10.6 | 8.7 |
| 4 | 9.0 | 8.9 | 11.5 | | 9.6 | 9.5 | 10.7 | 10.1 | 9.7 | | 9.8 | 10.4 |
| 5 | 7.2 | 7.1 | 8.0 | 9.1 | | 8.0 | 8.9 | 9.7 | 10.6 | 12.7 | | 10.4 |
| 6 | 5.7 | 6.5 | 5.9 | 6.9 | 6.2 | | 9.5 | 8.2 | 10.7 | 8.9 | 9.5 | |
| ***wL*1$_{AT}$** | **23%** | | | | | | **9%** | | | | | |

[a]Each trajectory length 10000 s, one-point-data reassigned option is ON, WAM searching is OFF, and searching goal $wL1_{SE}$ is set to 3.5%. Rate constants with relative large deviation from true values (>20%) are highlighted.

## 13. A two-state trajectory

For a two-state system, if one transition has the lifetime equals to the experimental time resolution, 63% of its dwell times will last shorter than one-data-point. These dwell times lose their information during the measurements, making the rate constant very difficult to be extracted from such a trajectory. It reaches the limitation of the information one has obtained. An example trajectory is shown in **Figure 3** in the main text with the lifetime of one transition the same as and the other twice the time resolution 10 ms.

Indeed, when forcing the HaMMy to a 2-state model, the software identified two states with FRET values averaged by the blurring to 0.82 and 0.24 instead of the true values 0.9 and 0.1 respectively. The $wL1_{AT}$ score is 56% (**Table S10**). Other one-step analysis methods yield different results, but all have relatively large errors. The 2-step method postFRET is able to reduce the score to 1%.

**Table S10**. Data analysis results for the two-state smFRET trajectory (each trajectory length 500 s).

| | Preset values | | HaMMy[a] | | vbFRET-C[b] | | Our code one-step[c] | | postFRET[c] | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 |
| 1 | | 100 | | 46 | | 38 | | 73.2 | | 99.7±1.8 |
| 2 | 50 | | 21 | | 28 | | 25.0 | | 49.4±1.3 | |
| ***wL*1$_{AT}$** | **N/A** | | **56%** | | **53%** | | **38%** | | **1%** | |

[a]Force 2-state. States are identified to be *FRET* = 0.82 and 0.24. [b]Force 2-state. States are identified to be *FRET* = 0.90 and 0.45. [c]Force 2-state, and *FRET* = 0.9 and 0.1.

## References


(1) Raff, L. M. *Principles of Physical Chemistry*; Prentice Hall Upper Saddle River, NJ, 2001.

(2) Forster, T. Energiewanderung und Fluoreszenz. *Naturwissenschaften* **1946**, *33*, 166–175.

(3) Bronson, J. E.; Fei, J.; Hofman, J. M.; Gonzalez Jr., R. L.; Wiggins, C. H.; Gonzalez Jr, R. L.; Wiggins, C. H. Learning Rates and States from Biophysical Time Series: A Bayesian Approach to Model Selection and Single-Molecule FRET Data. *Biophys. J.* **2009**, *97*, 3196–3205.

(4) Gillespie, D. T. Exact Stochastic Simulation of Coupled Chemical Reactions. *J. Phys. Chem.* **1977**, *81*, 2340–2361.

(5) Benitez, J. J.; Keller, A. M.; Ochieng, P.; Yatsunyk, L. A.; Huffman, D. L.; Rosenzweig, A. C.; Chen, P. Probing Transient Copper Chaperone-Wilson Disease Protein Interactions at the Single-Molecule Level with Nanovesicle Trapping. *J. Am. Chem. Soc.* **2008**, *130*, 2446–2447.

(6) Landes, C. F.; Rambhadran, A.; Taylor, J. N.; Salatan, F.; Jayaraman, V. Structural Landscape of Isolated Agonist-Brinding Domains Form Single AMPA Receptors. *Nat. Chem. Biol.* **2011**, *7*, 168–173.

(7) Hyeon, C.; Lee, J.; Yoon, J.; Hohng, S.; Thirumalai, D. Hidden Complexity in the Isomerization Dynamics of Holliday Junctions. *Nat. Chem.* **2012**, *4*, 907–914.