Biophysical Journal, Volume *122*

Supplemental information

Learning continuous potentials from smFRET

J. Shepard Bryan IV and Steve Pressé

SUPPLEMENTARY INFORMATION

S0.5 Derivation of the likelihood

Here we derive the likelihood distribution for photon measurements given particle positions. One important consequence of our Ito approximation, equation 5, is that our likelihood for single photon data will be equivalent to our likelihood for binned data. That is to say that neither photon arrival times nor ordering of photon colors within a time window provide any additional information about the particle position. We will start this section by deriving the likelihood for single photon measurements. After showing that the single photon measurements contain no additional information here as compared to binned photon measurements, we then derive the likelihood for binned photons (equations 6 and 7) that we use throughout this work.

We first write out the probability of collecting J photons with photon arrival times, $\boldsymbol{T}$, and photon colors, $\boldsymbol{\phi}$, within a time window,

$$\mathcal{P}\left(\boldsymbol{T},\boldsymbol{\phi}|x,\lambda_X,\lambda_g,\lambda_r\right)=\mathcal{P}\left(\boldsymbol{T}|x,\lambda_X,\lambda_g,\lambda_r\right)\mathcal{P}\left(\boldsymbol{\phi}|\boldsymbol{T},x,\lambda_X,\lambda_g,\lambda_r\right) \tag{1}$$

where, for simplicity alone, in our derivation here we ignore artifacts induced by crosstalk and detector efficiency. The time between photon arrivals will be exponentially distributed according to the excitation rate, λ_X, and the background rates, λ_g and λ_r. The probability of the photon arrival times, $\boldsymbol{T}$ is the probability of the J inter-photon times multiplied by the probability of no photon following the Jth photon

$$\mathcal{P}\left(\boldsymbol{T}|x,\lambda_X,\lambda_g,\lambda_r\right)\propto\left(1-\int_0^{\Delta t-T_J}dt\;\mathbf{Exp}\left(t;\lambda_X+\lambda_g+\lambda_r\right)\right)\prod_{j=1}^{J}\mathbf{Exp}\left(T_j-T_{j-1};\lambda_X+\lambda_g+\lambda_r\right) \tag{2}$$

$$=e^{-(\lambda_X+\lambda_g+\lambda_r)(\Delta t-T_J)}(\lambda_X+\lambda_g+\lambda_r)^J e^{-(\lambda_X+\lambda_g+\lambda_r)\sum_{j=1}^{J}T_j-T_{j-1})} \tag{3}$$

$$=(\lambda_X+\lambda_g+\lambda_r)^J e^{-(\lambda_X+\lambda_g+\lambda_r)\Delta t} \tag{4}$$

where in our derivation we used $T_0=0$. The probability over the photon colors is the product of the probabilities over each individual photon given by the rates and the FRET efficiency

$$\mathcal{P}\left(\boldsymbol{\phi}|\boldsymbol{T},x,\lambda_X,\lambda_g,\lambda_r\right)\propto\prod_{j=1}^{J}\frac{(\lambda_X f_g(x)+\lambda_g)^{[\phi_j=\text{green}]}(\lambda_X f_r(x)+\lambda_r)^{[\phi_j=\text{red}]}}{\lambda_X+\lambda_g+\lambda_r} \tag{5}$$

$$=\frac{(\lambda_X f_g(x)+\lambda_g)^G(\lambda_X f_r(x)+\lambda_r)^R}{(\lambda_X+\lambda_g+\lambda_r)^J} \tag{6}$$

where $f_g(x)=1-\text{FRET}(x)$, $f_r(x)=\text{FRET}(x)$, $[x=y]$ is the Iverson bracket (which is equal to 1 if $x=y$ and 0 otherwise), and R and G are the total number of observed red and green photons. Putting this all together yields a distribution which has no dependency on individual photon arrival times nor photon color order,

$$\mathcal{P}\left(\boldsymbol{T},\boldsymbol{\phi}|x,\lambda_X,\lambda_g,\lambda_r\right)\propto(\lambda_X f_g(x)+\lambda_g)^G(\lambda_X f_r(x)+\lambda_r)^R e^{-\Delta t(\lambda_X+\lambda_g+\lambda_r)}. \tag{7}$$

Since the likelihood depends neither on individual photon arrival times nor on photon color ordering, we lose no generality by rewriting our likelihood solely in terms of the number of measured photons within a time bin.

We now derive the likelihood for measuring R red photons and G green photons in a time window. The probability of collecting G green photons and R red photons in a time window is the probability of collecting $J=G+R$ photons multiplied by the probability that R of the photons are red

$$\mathcal{P}\left(R,G|\lambda_X,\lambda_g,\lambda_r,x\right)=\mathcal{P}\left(R|J,\lambda_X,\lambda_g,\lambda_r,x\right)\mathcal{P}\left(J|\lambda_X,\lambda_g,\lambda_r\right). \tag{8}$$

The probability of collecting J photons in a time window is Poisson distributed according to the rates,

$$\mathcal{P}\left(J|\lambda_X,\lambda_g,\lambda_r\right)=\mathbf{Poisson}\left(J;\Delta t(\lambda_X+\lambda_g+\lambda_r)\right). \tag{9}$$

The probability that R photons are red is a binomial distribution with weight given by the relative rates of red and green photons

$$\mathcal{P}\left(R|J,\lambda_X,\lambda_g,\lambda_r,x\right)=\mathbf{Binomial}\left(R;\frac{\lambda_X f_r(x)+\lambda_r}{\lambda_X+\lambda_g+\lambda_r},R+G\right). \tag{10}$$

All together this yields

$$\mathcal{P}\left(G,R|\lambda_X,\lambda_r,\lambda_g,x\right) = \textbf{Binomial}\left(R;\frac{\lambda_X f_r(x)+\lambda_r}{\lambda_X+\lambda_g+\lambda_r}\right)\textbf{Poisson}\left(J;\Delta t(\lambda_X+\lambda_g+\lambda_r)\right) \tag{11}$$

$$= \binom{G+R}{R}\left(\frac{\lambda_X f_r(x)+\lambda_r}{\lambda_X+\lambda_g+\lambda_r}\right)^R\left(1-\frac{\lambda_X f_g(x)+\lambda_g}{\lambda_X+\lambda_g+\lambda_r}\right)^G\frac{(\Delta t(\lambda_X+\lambda_r+\lambda_g))^{R+G}}{(R+G)!}e^{-\Delta t(\lambda_X+\lambda_r+\lambda_g)} \tag{12}$$

$$= \frac{(\Delta t(\lambda_X f_r(x)+\lambda_r))^R}{R!}e^{-\Delta t(\lambda_X f_r(x)+\lambda_r)}\frac{(\Delta t(\lambda_X f_g(x)+\lambda_g))^G}{G!}e^{-\Delta t(\lambda_X f_g(x)+\lambda_g)} \tag{13}$$

$$= \textbf{Poisson}\left(R;\Delta t(\lambda_X f_r(x)+\lambda_r)\right)\textbf{Poisson}\left(G;\Delta t(\lambda_X f_g(x)+\lambda_g)\right). \tag{14}$$

This is the likelihood (equations 6 and 7) that we use throughout this work.

S0.6 Conditional probabilities

Here we derive the conditional probabilities used in the Gibbs sampling algorithm of section 0.3. Note that, for clarity, we drop multiplicative terms not directly related to the variable on which we condition in each of the following equations. We do so because these terms are treated as constants during each step of the conditional sampling in the Gibbs sampler.

S0.6.1 Positions

The distribution over positions is the product of the likelihood (equations 6 and 7), the discretized Langevin equation (equation 5), and the prior on the initial position (equation 16)

$$\mathcal{P}\left(\boldsymbol{x}_{1:N}|\boldsymbol{U}^*_{1:M},\zeta,\lambda_X,\lambda_g,\lambda_r,\boldsymbol{g}_{1:N},\boldsymbol{r}_{1:N}\right) \propto \mathcal{P}\left(\boldsymbol{g}_{1:N}|\boldsymbol{x}_{1:N}\right)\mathcal{P}\left(\boldsymbol{r}_{1:N}|\boldsymbol{x}_{1:N}\right)\mathcal{P}\left(\boldsymbol{x}|\boldsymbol{U}^*_{1:M},\zeta,\lambda_X\right) \tag{15}$$

$$\begin{aligned}= &\textbf{Normal}\left(x_1;R0,R0^2\right)\\ &\times\left(\prod_{n=2}^{N}\textbf{Normal}\left(x_n;x_{n-1}+\frac{\Delta t}{\zeta}f(x_{n-1}),\frac{2\Delta tkT}{\zeta}\right)\right)\\ &\times\left(\prod_{n=1}^{N}\textbf{Poisson}\left(g_n;\lambda_X(1-\mathrm{FRET}(x_n))+\lambda_g\right)\right)\\ &\times\left(\prod_{n=1}^{N}\textbf{Poisson}\left(r_n;\lambda_X(\mathrm{FRET}(x_n)+\lambda_r\right)\right).\end{aligned} \tag{16}$$

To sample from this distribution we may sample each x_n individually using a Metropolis Hastings (24) step. Separating equation 16 into conditional distributions at each position yields three equations: a conditional posterior on x_1

$$\begin{aligned}\mathcal{P}\left(x_1|\boldsymbol{x}_{2:N},\boldsymbol{U}^*_{1:M},\zeta,\lambda_X,\lambda_g,\lambda_r,\boldsymbol{g}_{1:N},\boldsymbol{r}_{1:N}\right) \propto &\textbf{Normal}\left(x_1;R0,R0^2\right)\\ &\times\textbf{Normal}\left(x_2;x_1+\frac{\Delta t}{\zeta}f(x_1),\frac{2\Delta tkT}{\zeta}\right)\\ &\times\textbf{Poisson}\left(g_1;\lambda_X(1-\mathrm{FRET}(x_1))+\lambda_g\right)\\ &\times\textbf{Poisson}\left(r_1;\lambda_X(\mathrm{FRET}(x_1)+\lambda_r\right),\end{aligned} \tag{17}$$

an equation for each x_n from time levels 2 to $N-1$

$$\begin{aligned}\mathcal{P}\left(x_n|\boldsymbol{x}_{1:n-1,n+1:N},\boldsymbol{U}^*_{1:M},\zeta,\lambda_X,\lambda_g,\lambda_r,\boldsymbol{g}_{1:N},\boldsymbol{r}_{1:N}\right) \propto &\textbf{Normal}\left(x_n;x_{n-1}+\frac{\Delta t}{\zeta}f(x_{n-1}),\frac{2\Delta tkT}{\zeta}\right)\\ &\times\textbf{Normal}\left(x_{n+1};x_n+\frac{\Delta t}{\zeta}f(x_n),\frac{2\Delta tkT}{\zeta}\right)\\ &\times\textbf{Poisson}\left(g_n;\lambda_X(1-\mathrm{FRET}(x_n))+\lambda_g\right)\\ &\times\textbf{Poisson}\left(r_n;\lambda_X(\mathrm{FRET}(x_n)+\lambda_r\right),\end{aligned} \tag{18}$$

and an equation for the last position, x_N,

$$\begin{aligned}\mathcal{P}\left(x_N|\boldsymbol{x}_{1:N-1},\boldsymbol{U}^*_{1:M},\zeta,\lambda_X,\boldsymbol{g}_{1:N},\boldsymbol{r}_{1:N}\right) \propto &\textbf{Normal}\left(x_N;x_{N-1}+\frac{\Delta t}{\zeta}f(x_{N-1}),\frac{2\Delta t kT}{\zeta}\right)\\ &\times\textbf{Poisson}\left(g_N;\lambda_X(1-\text{FRET}(x_N))+\lambda_g)\right)\\ &\times\textbf{Poisson}\left(r_N;\lambda_X(\text{FRET}(x_N)+\lambda_r)\right).\end{aligned} \tag{19}$$

S0.7 Potential

The conditional distribution for the potential is the product of the discretized Langevin equation (equation 5) and the prior on the potential (equation 17)

$$\mathcal{P}\left(\boldsymbol{U}^*_{1:M}|\boldsymbol{x}_{1:N},\zeta,\lambda_X,\lambda_g,\lambda_r,\boldsymbol{g}_{1:N},\boldsymbol{r}_{1:N}\right) \propto\textbf{Normal}\left(\boldsymbol{U}^*_{1:M};\boldsymbol{0},\boldsymbol{K}\right)\prod_{n=2}^{N}\textbf{Normal}\left(x_n;x_{n-1}+\frac{\Delta t}{\zeta}f(x_{n-1}),\frac{2\Delta t kT}{\zeta}\right) \tag{20}$$

which can be simplified to (34, 35)

$$\mathcal{P}\left(\boldsymbol{U}^*_{1:M}|\boldsymbol{x}_{1:N},\zeta,\lambda_X,\lambda_g,\lambda_r,\boldsymbol{g}_{1:N},\boldsymbol{r}_{1:N}\right) =\textbf{Normal}\left(\boldsymbol{U}^*_{1:M};\ \tilde{\boldsymbol{\mu}},\ \tilde{\boldsymbol{K}}\right) \tag{21}$$

$$\tilde{\boldsymbol{K}} =\left(\boldsymbol{K}^{-1}+\frac{\Delta t}{2\zeta kT}\boldsymbol{K}^{-1}\boldsymbol{K}^{*T}\boldsymbol{K}^{*}\boldsymbol{K}^{-1}\right)^{-1} \tag{22}$$

$$\tilde{\boldsymbol{\mu}} =\frac{\Delta t}{2kT}\tilde{\boldsymbol{K}}\boldsymbol{K}^{-1}\boldsymbol{K}^{*T}\boldsymbol{v}_{1:N-1} \tag{23}$$

where $\boldsymbol{K}$ is the kernel matrix (covariance matrix) between all $\boldsymbol{U}^*_{1:M}$, $\boldsymbol{K}^*$ is the covariance between the potential at $\boldsymbol{x}^*_{1:M}$ and the force at $\boldsymbol{x}_{1:N}$ with elements $K^*_{nm}=-\nabla k(x_n,x^*_m)$, and $\boldsymbol{v}_{1:N-1}$ are the velocities at each time level with elements $v_n=(x_{n+1}-x_n)/\Delta t$. As the final distribution for $\boldsymbol{U}^*_{1:M}$ is Gaussian, we may directly sample $\boldsymbol{U}^*_{1:M}$ from the posterior without invoking Metropolis Hastings (34).

S0.8 Photon rates

The conditional distribution on the excitation rate is the product of the likelihood (equations 7 and 6) and the prior on excitation rate (equation 12)

$$\begin{aligned}\mathcal{P}\left(\lambda_X|\boldsymbol{U}^*_{1:M},\boldsymbol{x}_{1:N},\lambda_g,\lambda_r,\zeta,\boldsymbol{g}_{1:N},\boldsymbol{r}_{1:N}\right) \propto &\textbf{Gamma}\left(\lambda_X;\kappa_{\lambda_X},\theta_{\lambda_X}\right)\\ &\times\left(\prod_{n=1}^{N}\textbf{Poisson}\left(g_N;\Delta t D_g(\lambda_X f_g(x_N)+\lambda_g)\right)\right)\\ &\times\left(\prod_{n=1}^{N}\textbf{Poisson}\left(r_N;\Delta t D_r(\lambda_X f_r(x_n)+\lambda_r)\right)\right).\end{aligned} \tag{24}$$

The distribution for the background rates is constructed in an identical manner except for the prior for which we replace the first term on the right hand side of 24 with equation 13 for λ_r and 14 for λ_g. To sample from either distribution we use Metropolis Hastings by proposing a sample at each iteration of the Gibbs sampler and accepting or rejecting based on the relative probabilities of the proposed variable and the old sample (24).

S0.9 Friction coefficient

The conditional distribution over the friction coefficient is the product of the discretized Langevin equation (equation 5) and the prior on friction (equation 15)

$$\mathcal{P}\left(\zeta|\boldsymbol{U}^*_{1:M},\boldsymbol{x}_{1:N},\lambda_X,\boldsymbol{g}_{1:N},\boldsymbol{r}_{1:N}\right) \propto\textbf{Gamma}\left(\zeta;\kappa_\zeta,\theta_\zeta\right)\prod_{n=2}^{N}\textbf{Normal}\left(x_n;x_{n-1}+\frac{\Delta t}{\zeta}f(x_{n-1}),\frac{2\Delta t kT}{\zeta}\right). \tag{25}$$

To sample from this distribution we use a Metropolis Hastings step by proposing a sample at each iteration of the Gibbs sampler and accepting or rejecting based on the relative probabilities of the proposed variable and the old sample (24).

S0.10 Bayesian Hidden Markov model

We compare the energy landscape learned using our method to an energy landscape learned using a Bayesian Hidden Markov Model (HMM) (24, 50). Here we will briefly describe the structure of our HMM algorithm, then explain how we can use our Bayesian HMM analysis results to infer potential energy landscapes.

Briefly, HMMs work by assuming that the system under consideration has a discrete number of states, $k = 1, 2, ..., K$ governed by a transition matrix, $\boldsymbol{q} = [q_{ij}]_{K\times K}$. At each time level n, the system's state, s_n, is conditioned on the state of the system at the previous time level, s_{n-1}, given the transition matrix, $\boldsymbol{q}$,

$$\mathcal{P}\left(s_n|s_{n-1}, \boldsymbol{q}\right) = \textbf{Categorical}\left(s_n; \boldsymbol{q}_{s_{n-1}}\right) \tag{26}$$

where $\boldsymbol{q}_{s_{n-1}}$ coincides with the row of $\boldsymbol{q}$ corresponding to s_{n-1}. Put differently, the probability that $s_n = j$ given that $s_{n-1} = i$ is equal to q_{ij}.

Each state, k, has its own pair distance, r_k. At each time level, the measured number of photons is conditioned on the pair distance of the system's state at that time level

$$\mathcal{P}\left(g_n\right) = \textbf{Poisson}\left(g_n; \Delta t D_g(\lambda_X f_g(r_{s_n}) + \lambda_g)\right) \tag{27}$$

$$\mathcal{P}\left(r_n\right) = \textbf{Poisson}\left(r_n; \Delta t D_r(\lambda_X f_r(r_{s_n}) + \lambda_r)\right) \tag{28}$$

where $f_r(x)$ and $f_g(x)$ are the FRET rates, including crosstalk terms, defined by 9. Notice that this likelihood is equivalent to the SKIPPER-FRET likelihood, equations 6 and 7.

Working within the Bayesian paradigm, we place priors on all unknowns,

$$\mathcal{P}\left(s_1\right) = \textbf{Categorical}\left(s_1; \boldsymbol{\alpha}_q\right) \tag{29}$$

$$\mathcal{P}\left(r_k\right) = \textbf{Gamma}\left(r_k; \kappa_r, \theta_r\right) \tag{30}$$

$$\mathcal{P}\left(\lambda_X\right) = \textbf{Gamma}\left(\lambda_X; \kappa_{\lambda_X}, \theta_{\lambda_X}\right) \tag{31}$$

$$\mathcal{P}\left(\lambda_g\right) = \textbf{Gamma}\left(\lambda_g; \kappa_{\lambda_g}, \theta_{\lambda_g}\right) \tag{32}$$

$$\mathcal{P}\left(\lambda_r\right) = \textbf{Gamma}\left(\lambda_r; \kappa_{\lambda_r}, \theta_{\lambda_r}\right) \tag{33}$$

$$\mathcal{P}\left(\boldsymbol{q}_k\right) = \textbf{Dirichlet}\left(\boldsymbol{q}_k; \boldsymbol{\alpha}_q\right) \tag{34}$$

where $\textbf{Dirichlet}\left(\boldsymbol{\alpha}_q\right)$ is the Dirichlet distribution (24), conjugate to the Dirichlet dynamics model (equation 26) (24, 50). We choose our hyperparameters to be $\boldsymbol{\alpha}_q = [1/K, 1/K, ..., 1/K]$, $\kappa_r = 2$, and $\theta_r = R_0$.

Equations (26) to (34) form a high dimensional posterior. We sample from our posterior using Gibbs sampling (24) and the forward filter-backward sampling algorithm (24). Once enough samples have been generated, we may choose to use the sample average to provide a point estimate for each variable.

In order to compare the HMM method to SKIPPER-FRET in the Results section, we used our HMM results to estimate the energy of each state. The energy of each state is calculated using the transition probability matrix, $\boldsymbol{q}$. We can find the energies from $\boldsymbol{q}$ by first calculating the equilibrium state probabilities, $\boldsymbol{P}$, defined as

$$\boldsymbol{P} = \boldsymbol{q}\boldsymbol{P} \tag{35}$$

then equating $\boldsymbol{P}$ to the Boltzmann distribution

$$\boldsymbol{P} = \frac{1}{Z}\begin{bmatrix} e^{-\frac{E_1}{kT}} \\ e^{-\frac{E_2}{kT}} \\ \dots \\ e^{-\frac{E_K}{kT}} \end{bmatrix}. \tag{36}$$

Together, equation 35 and 36 allow us to calculate the energy of each state in the HMM model.

S0.10.1 Barrier heights within HMM paradigm

Here we highlight how one would, if required, compute barrier heights within an HMM paradigm under two regimes: 1) features of the barrier are known; or 2) data are collected at different temperatures in addition to features of the barrier being known.

Here we focus on the first as it is of greater interest to experiments on biomolecules operating under one set of physiological temperatures.

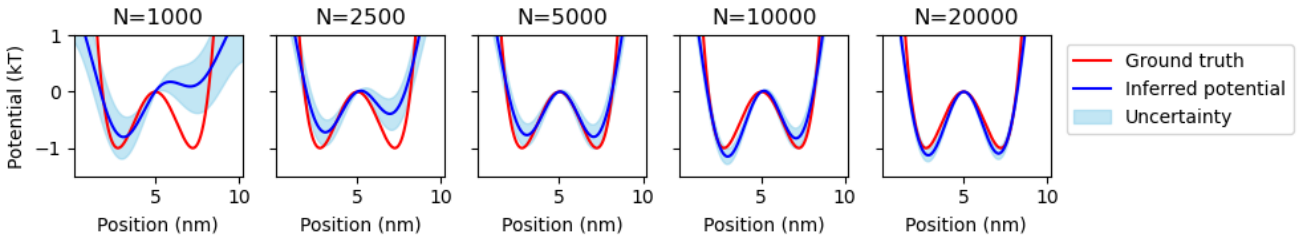

Figure S1: **Robustness test with respect to number of data points**. Each panel plots the potential inferred using SKIPPER-FRET (blue) against the ground truth potential energy landscape (red) for a given number of measurements, N, listed at the top.

In order to demonstrate that one could calculate barrier heights between states in the HMM model, we would first need to assume that the transition probability matrix, $\boldsymbol{q}$, is the solution to a master equation for a rate matrix, $\boldsymbol{\lambda}$,

$$\boldsymbol{q} = \exp(\Delta t \boldsymbol{\lambda}). \tag{37}$$

Solving for $\boldsymbol{\lambda}$ we get

$$\boldsymbol{\lambda} = \text{logm}(\boldsymbol{q})/\Delta t \tag{38}$$

where logm is the matrix logarithm. Assuming that the wells representing each state can be approximated as harmonic oscillators, we can relate $\boldsymbol{\lambda}$ to barrier heights using Kramer's rate equation (2, 22, 31)

$$\lambda_{ij} = \begin{cases} \frac{D\sqrt{c_i c_{ij}}}{2\pi kT} e^{-\frac{E_{ij}-E_i}{kT}} & i \neq j \\ -\sum_{l \neq i} \lambda_{il} & i = j \end{cases} \tag{39}$$

where c_i is the curvature of the well defining state i, c_{ij} is the curvature of the barrier between states i and j, E_{ij} is the energy of the barrier between states i and j, and D is a diffusion parameter dictating the rate of transitions in the absence of a barrier. Solving for the barrier heights we get

$$E_{ij} = E_i - kT \log(2\pi\lambda_{ij} kT) - kT \log(D\sqrt{c_i c_{ij}}). \tag{40}$$

We note that using equation (40), we can only learn the energy of the barrier, E_{ij}, if we know D, c_i, and c_{ij}. However, D, c_i, and c_{ij} are internal parameters of the system which are not otherwise easy to deduce (31). In practice, bounds for barrier height are obtained by using additional approximations and an order of magnitude guess for unknown quantities (2, 31).

Thus we see that the inability to infer accurate potential energy barriers from a single data set without knowledge of hidden internal parameters is a limitation of HMMs applied to smFRET data. By contrast, SKIPPER-FRET can learn barrier heights and friction coefficients from a single data set.

S1 ROBUSTNESS WITH RESPECT TO AMOUNT OF DATA

Here we test SKIPPER-FRET's robustness with respect to the length of the data set. That is, we test how well the inferred potential energy landscape matches the ground truth given different number of time levels, N, available in the data. For our robustness test, we use the same simulated data as the first double well experiment (figures 3 and 4), but truncated at different values of N.

Figure S1 shows the results. As expected, when there are too few time levels for the pair distance to sample both wells, as is the case for $N = 1000$ and $N = 2500$ in figure S1, SKIPPER-FRET cannot infer an accurate potential energy landscape due to missing data on the other well. When there is a sufficient number of time levels for the pair distance trajectory to sample both wells, as is the case for $N \geq 5000$ in figure S1, then the form of the SKIPPER-FRET potential matches the ground truth potential energy landscape more closely. Generally, figure S1 shows that the accuracy of SKIPPER-FRET increases with the number of time levels, and the uncertainty decreases with the number of data points. Note that SKIPPER-FRET's computation time increases linearly with the number of time levels.

Thus it is important to make sure that sufficient number of data supplied before applying SKIPPER-FRET. An ideal data set will have enough time for the pair distance to explore all space. For the purposes of this manuscript, we set $N = 10000$ for all data sets analyzed because this value gave us appropriate balance between accuracy and computation speed.

As it pertains to analysis of real experiments, of course, we can only ascertain the form of the potential for regions visited.

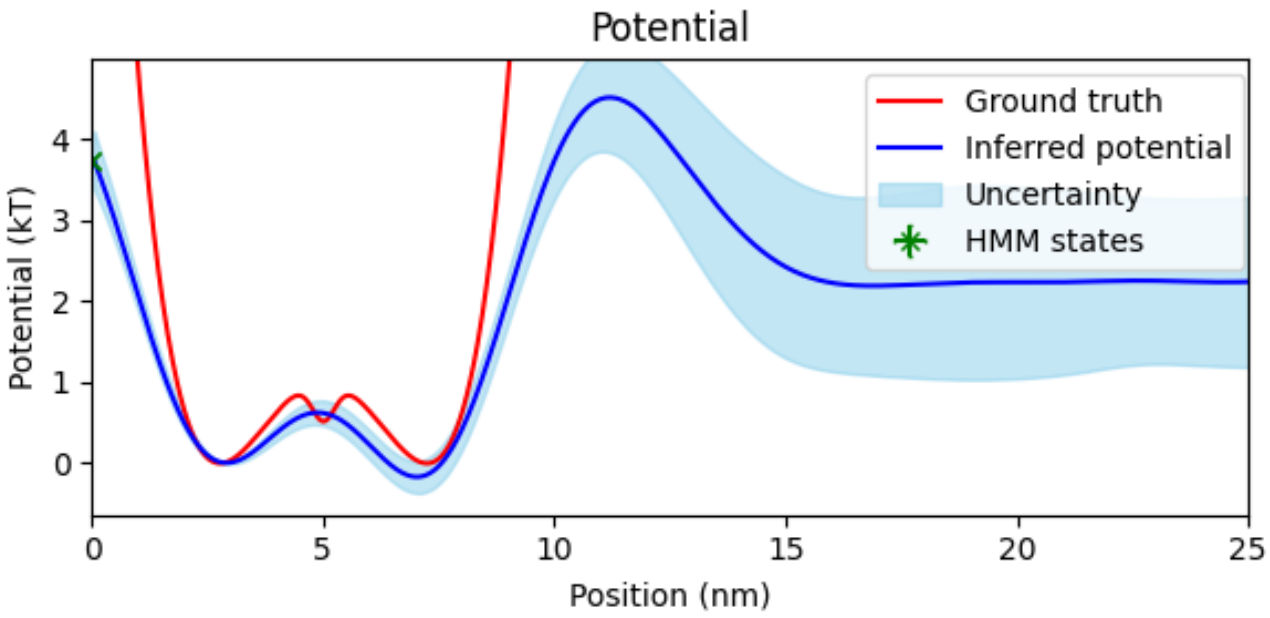

Figure S2: **Simulated potential energy landscape with a sharp dip**. We show our inferred potential energy landscape (blue) with uncertainty (light blue) against the ground truth potential energy landscape used in the simulation (red). We additionally plot markers, with uncertainty, indicating the inferred state energy and pair distance using HMMs (green). The common point of zero potential energy was set at the bottom of the leftmost barrier.

S2 ROBUSTNESS TEST ON POTENTIAL WITH SHARP DIP

Here we demonstrate a failure mode of our method when the potential varies on length scales faster than our length scale hyperparameter, ℓ. In figure S2 we generate simulated data using a potential energy landscape with two large wells and one thin well between them. As seen in figure S2, our method is able to infer the two large wells accurately, but otherwise misses the small middle well. This is because the length scale hyperparameter, set in equation (18), sets the level of detail the SKI-GP method can infer. In particular, in figure S2, we use a length scale of $\ell = 2$nm. However, the width of the well is on the order of 1nm. On the other hand, in figure S3 we see that by setting the length scale hyperparamater to .5nm we pick up the middle well (35). In practice, setting a smaller length scale may result in having noise dictate the shape of the potential wells, such as the well depth underestimation that appears in S3. Thus, our potential should be understood as fundamentally coarse-grained on a length scale set by the length hyperparameter.

S3 EVALUATING THE FRICTION COEFFICIENT

As we do not have a means to estimate the ground truth for the friction coefficient for real data, we need to compare our method against an order of magnitude estimate set by typical scales of the problem. We obtain a rough estimate using dimensional analysis. The units of ζ are mass over time or, equivalently,

$$\zeta \approx \frac{\text{Energy} \times \text{Time}}{\text{Length}^2}. \tag{41}$$

Treating energy scales as kT (with k as Boltzmann's constant and T the temperature, $\approx$ 4pN nm); length scales as the distance between wells $\approx$ 10nm; and time scales as the switching times between wells $\approx$ .1s (see figure 6), we have

$$\zeta \approx \frac{(4\text{pN nm}) \times (.1\text{s})}{(10\text{nm})^2} \tag{42}$$

$$= 4\text{mg/s} \tag{43}$$

consistent with our our estimate of 1.54mg/s.

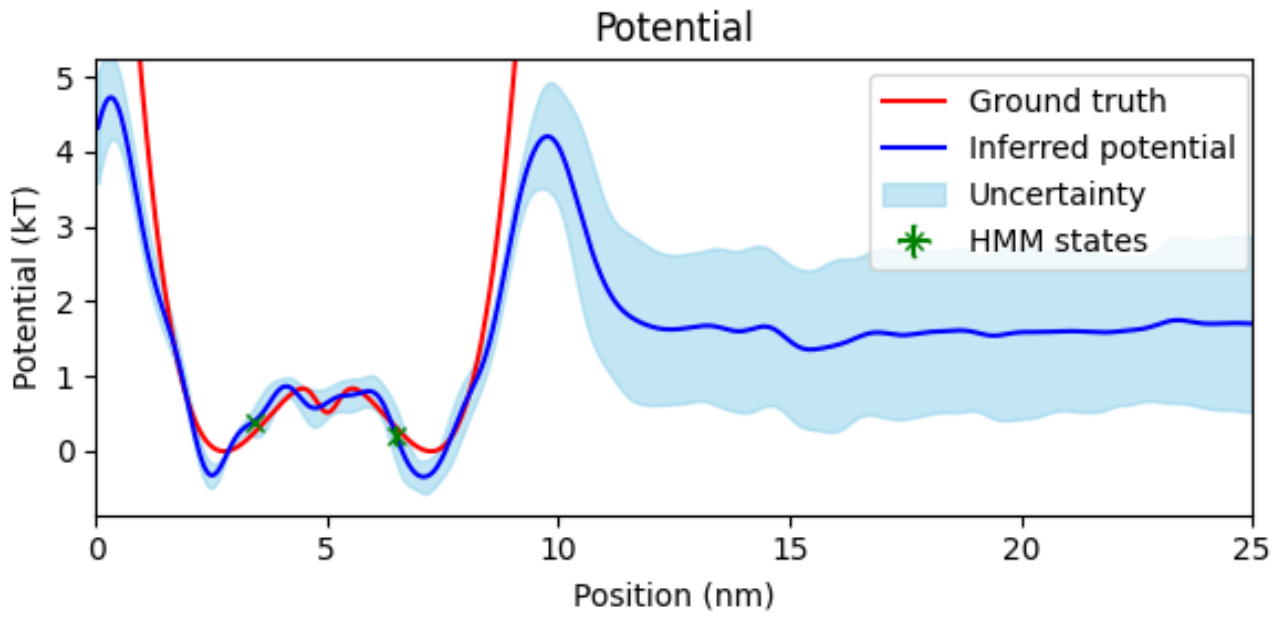

Figure S3: **Sharp dip potential analyzed with small length scale**. We show our inferred potential energy landscape (blue) with uncertainty (light blue) against the ground truth potential energy landscape used in the simulation (red). We additionally plot markers, with uncertainty, indicating the inferred state energy and pair distance using HMMs (green). The common point of zero potential energy was set at the bottom of the leftmost barrier.

S3.1 Parameters Used in the Simulations

We used the following parameters in our simulations.

Δt	1ms
kT	4.114 pNnm
ζ	.03 g/s
λ_X	3 kHz
λ_r	0
λ_g	0
R_0	5 nm