Dynamic free energy landscape of a linear motor chitinase from single-particle tracking trajectories

Running Title: **Dynamic energy landscape of chitinase**

K. Okazaki, A. Nakamura and R. Iino

Abstract

Single-molecule experiments have been used to investigate mechanism of biomolecular motors by visualizing their motions. However, it is challenging to construct an effective physics-based model that consistently explains observed trajectories. In this study, we estimate an underlying diffusion model from observed trajectories. To deal with nonequilibrium trajectories driven by chemical energy consumed by biomolecular motors, we develop a novel framework based on hidden Markov model, which considers switching among multiple energy surfaces depending on chemical states of the motors. The method is tested with simulation trajectories and applied to single-molecule trajectories of processive chitinase, a linear motor that moves with hydrolysis energy of single chitin chain. The dynamic free energy landscape underlying the burnt-bridge Brownian ratchet mechanism of chitinase is revealed. The novel framework allows us to bridge a gap between mechanochemical coupling mechanism and unidirectional motion of biomolecular motors.

Statement of Significance

Biomolecular motors involve in many essential biological processes, from muscle contraction to separating double strands of DNA before its replication. Unlike typical motors that use ATP, processive cellulase and chitinase work with hydrolysis energy of crystalline cellulose and chitin, respectively. The previous single-molecule experiments revealed a unique molecular mechanism of chitinase, called the burnt-bridge Brownian ratchet, for its unidirectional movement. In this study, we obtain chemical-state dependent free energy landscapes and diffusion coefficient by analyzing the single-molecule trajectories of chitinase in more detail and clarify physical principles behind the burnt-bridge Brownian ratchet mechanism. The novel framework developed in this study provides a general approach to understand unidirectional motions of biomolecular motors.

Introduction

Single-molecule experiments are a powerful toolbox to study functional dynamics of biomolecular motors by visualizing their motions. Mechanochemical coupling mechanisms of F_1-ATPase (1-5), myosin (6, 7) and kinesin (8-10), typical motors driven by adenosine triphosphate (ATP), have been revealed by various single-molecule techniques. Recently, processive cellulase and chitinase have been cast new light as a different type of biomolecular motors that use hydrolysis energy of polysaccharides for their unidirectional movements by high-speed atomic force microscopy (11, 12) and high-precision single-particle tracking (13). Together with these experiments, structural information and molecular dynamics simulations reveal how cellulase and chitinase couple elementary steps of processive catalysis to their unidirectional motions (13-16). The emerging mechanism of processive chitinase is that its unidirectional movement is realized by the fast catalytic reaction and biased Brownian motions (13). This type of mechanism, the burnt-bridge Brownian ratchet, was also proposed for other motors such as collagenase that moves processively on the collagen fibril with proteolysis energy (17, 18). The burnt-bridge Brownian ratchet is one of Brownian ratchet model that utilizes thermal fluctuations and switching between different energy surfaces (19-22). Although there have been some studies of simple models based on the burnt-bridge Brownian ratchet (18, 23), not much has been studied about realistic mechanisms by exploiting single-molecule experimental data.

There have been attempts to estimate physicochemical models directly from single-molecule trajectories. Single-molecule fluorescence resonance energy transfer (FRET) trajectories of DNA, RNA and protein conformational dynamics have been analyzed by hidden Markov model (HMM) or maximum likelihood method to estimate kinetic rates and FRET efficiencies associated with different conformational states that are behind FRET signals (24-28) or diffusion model (29). Single-particle tracking trajectories of a cell-surface receptor have been analyzed by HMM to estimate multiple diffusion coefficients that arise from different conformational states of the receptor (30). Energy landscape is another key quantity to understand motions, kinetics and stabilities of biomolecules (31-36), and has been estimated from single-molecule trajectories (37, 38).

Along the lines of these studies, we aim to estimate dynamic energy landscape and diffusion coefficient from single-particle tracking trajectories of biomolecular motors.

Biomolecular motors move along a linear track or rotate relative to a stator. Thus, their motions can be effectively described by one-dimensional (1D) model. This simplification has been successfully adopted for rotary motors (39, 40) as well as linear motors (19). Similar approach has been explored in the field of protein folding. Apparently complicated processes are projected onto a 1D reaction coordinate, and modeled as a diffusion process over a barrier separating unfolded and folded states (27, 41-43). Some methods have been developed to construct the 1D diffusion model from experimental or simulation trajectories (44, 45). Especially, the method developed by Hummer and coworkers has been well tested with extensive applications to simulation trajectories (41, 42, 44, 46). The method can be adapted for analyzing motions of biomolecular motors. However, in contrast to protein folding, unidirectional motions of the motors generally involve switching between different energy surfaces depending on their chemical states.

In this study, we develop a method to estimate 1D diffusion model that involves hopping on multiple energy surfaces from trajectories of biomolecular motors. The novel framework estimates dynamic energy landscape and diffusion coefficient by combining the Hummer's method with HMM. We test the method with trajectories generated from Brownian dynamics simulation with two energy surfaces switching stochastically. Then, we apply the method to single-molecule trajectories of chitinase to obtain its dynamic free energy landscape triggered by the catalytic reaction and the product release.

Methods

Positional transition probability of diffusion model on single energy surface

The present and next sections describe components needed to compute the likelihood function L^{msurf} (see Results). Here we outline the connection of free energy landscape and diffusion coefficient to the positional transition probability, which has been originally developed by Hummer and coworkers with extensive applications to simulation trajectories of protein folding (42, 44, 46). Assuming the 1D model, we aim to calculate

the transition probabilities between discretized positions (bins) from the underlying energy and diffusion coefficient profiles. By spatially discretizing the 1D Smoluchowski equation, the transition rate constant from bin *j* to bin $j \pm 1$ is expressed as (47),

$$k_{j \to j \pm 1} = \frac{D}{\Delta x^2} exp\left(-\frac{F(j \pm 1) - F(j)}{2k_B T}\right)$$

where $F(j)$ is free energy of bin *j*, D is diffusion coefficient, and Δx is bin size. This expression enables us to construct a kinetic model in terms of free energy and diffusion coefficient. Note that in this study we do not consider the position-dependent diffusion coefficient for simplicity. The master equation in matrix form is $\frac{d}{dt}\vec{p} = K\vec{p}$, where K is the rate matrix and $\vec{p} = (p_1 \quad \cdots \quad p_n)^T$ is the state-probability vector for total number of bins *n*. The rate matrix K is constructed from the transition rate constants as,

$$K_{ij} = \begin{cases} k_{j \to i} & i = j \pm 1 \\ -k_{j \to j+1} - k_{j \to j-1} & i = j \\ 0 & otherwise \end{cases}$$

Then, the master equation can be formally solved as,

$$\vec{p}(t) = \exp(Kt)\vec{p}(0) = T(t)\vec{p}(0)$$

Each element of the transition matrix $T(t) = \exp(Kt)$ represents the positional transition probability, $(T(t))_{ij} = p(i, t|j, 0)$, where $p(i, t|j, 0)$ is the conditional probability of being in bin *i* at time t, given it is in bin *j* at time 0. Note that subscript specifying the hidden state in F and $p(i, t|j, 0)$ is omitted here, because we consider only one state (that is, single energy surface).

Eigenvalues of the rate matrix K are related to timescales of the diffusion model (42). The first non-trivial eigenvalue λ_2 represents the slowest timescale of the model, where $\lambda_1 = 0 > \lambda_2 > \lambda_3 \cdots$. The implied timescale is expressed as $\tau_i = -1/\lambda_i$, where $i \geq 2$.

Switching probability between hidden states

The switching probability between the hidden states was calculated from rate constant of the process. For the two-state model (s_i=1 or 2), assuming rate constant from state 1 to 2 (from state 2 to 1) is k_f (k_b),

$$p(s_i \to s_{i+1}) = \begin{cases} 1 - k_f \Delta t & (s_i = 1, s_{i+1} = 1) \\ k_f \Delta t & (s_i = 1, s_{i+1} = 2) \\ 1 - k_b \Delta t & (s_i = 2, s_{i+1} = 2) \\ k_b \Delta t & (s_i = 2, s_{i+1} = 1) \end{cases}$$

This expression is true to the first order of Δt, which is usually a good approximation.

Translational symmetry and prior probability of the model

Because biomolecular motors have the catalytic cycle, their energy profiles are expected to have the translational symmetry with their moving length (or step size) l_{step} after one cycle. Thus, the energy profile of one state is copied and shifted by the step size to represent the next energy profile. Also, the diffusion coefficient of one state is copied to that of the other state. With $n_{shift} = \mathrm{int}(l_{step}/\Delta x)$ and n total number of bins, $F_1(0)$, $\cdots, F_1(n - n_{shift} - 1)$ are copied to $F_2(n_{shift})$, $\cdots, F_2(n-1)$ for the translational symmetry. The regions of the energy profiles that are out of the lower/upper limit of the coordinate after shifting, that is, $F_1(j)$ with $j = n - n_{shift}, \cdots, n-1$ and $F_2(j)$ with $j = 0, \cdots, n_{shift} - 1$, are called the peripheral regions. The peripheral regions, denoted as $j_s \in out$ with subscript s representing the state, are expected to be rarely visited and have high energies. Thus, we used the prior probability that energy of the peripheral regions increases at a constant rate as,

$$p(\mathrm{model}) \propto \prod_{s=1,2;\ j_s \in out} \exp\left[-\frac{(F_s(j_s) - F_s(j_s^0) - \delta|j_s - j_s^0|)^2}{2\sigma^2}\right]$$

where δ is increasing energy per bin, j_s^0 is the last bin before going into the region with $j_1^0 = n - n_{shift} - 1$ and $j_2^0 = n_{shift}$, σ is standard deviation that determines strength of the restraint. Then, according to Bayes' theorem, the posterior probability $p(\mathrm{model}|\mathrm{data})$ is,

$$p(\text{model}|\text{data}) \propto L^{msurf} \times p(\text{model}) = L'$$

The posterior probability L' was maximized during the optimization process (see below).

Monte Carlo optimization

We aim to optimize F and D as well as switching rates to maximize the posterior probability L' with input trajectory data. Explicitly, the model parameters are, $\theta = \left(F_1(0), \cdots, F_1(n-1), F_2(0), \cdots, F_2(n-1), \ln D_1, \ln D_2, k_f, k_b\right)$. We randomly change these parameters as a trial move, compute L' of the trial move, and accept or reject the trial move comparing to the current L' (30, 44). For the translational symmetry described above, some parameters are copied between states. The width of trial moves is adjusted on the fly in the first half of Monte Carlo (MC) steps so that acceptance ratio becomes ~0.3, and fixed in the last half. To compute L^{msurf}, the positional transition probability is calculated for each state as described above for given F and D, then fed into hidden Markov model as described in the main text and Fig. 1B. Metropolis criteria is used to accept or reject the trial move, $p_{acc} = min\left(1, e^{\Delta logL'}\right)$, where $\Delta logL' = logL'_{trial} - logL'_{current}$. In this way, it is ensured that the model parameters are sampled based on the posterior probability L'.

Brownian dynamics simulation combined with kinetic Monte Carlo

Based on the overdamped Langevin dynamics, trajectories were generated numerically by the forward-Euler integrator,

$$x(t+\delta t) = x(t) - \frac{D}{k_B T}\frac{\partial F_{s_t}}{\partial x}\delta t + g(t)$$

where F_{s_t} is potential energy of state s_t, D is diffusion coefficient, $\delta t = 10^{-6}$ is the integration time step, and $g(t)$ are Gaussian random numbers with zero mean and standard deviation $\sqrt{2D\delta t}$. F_{s_t} is switched by the kinetic Monte Carlo with specified rate constants (48). The model potential is $F_{s_t}(x) = a(x-(s_t-1))^2(x-s_t)^2$, where $s_t =$ 1 or 2 and a is a parameter controlling barrier height.

RMSD calculation of energy profile

The RMSD of profile $Y(x)$ to profile $Y_0(x)$ is defined as,

$$RMSD = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y(x_i) - Y_0(x_i))^2}$$

where n is number of bins excluding the peripheral regions. We shifted the whole profile $Y(x_i)$ by a constant C by the least square method before calculating RMSD,

$$C = -\frac{1}{n}\sum_{i=1}^{n}(Y(x_i) - Y_0(x_i))$$

Alignment of chitinase trajectory segments

Each long trajectory was divided into trajectory segments with two forward steps (Fig. 5C). Then, these segments were aligned by a position of high frequency identified from positional histogram of the segments. For each trajectory segment, positional histogram was calculated with bin size 0.15 nm. The position of high frequency was defined as a bin that is more frequent than average frequency, and more frequent than neighboring bins. The minimum value of the identified positions of high frequency was subtracted from the whole trajectory segment, setting origin of the trajectory segments.

Diffusion coefficient of gold nanoparticle

Diffusion coefficient of the 40 nm gold nano-particle was estimated by the Stokes-Einstein relation,

$$D_{AuNP} = \frac{k_B T}{6\pi\eta a} \sim 1.6 \times 10^7\ nm^2\ s^{-1}$$

where $\eta = 7 \times 10^{-4} kg\ m^{-1}\ s^{-1}$ is viscosity of water at 310K and $a = 20\ nm$ is radius of the probe.

Results

Likelihood function of diffusion model on multiple energy surfaces

Here we develop a framework to estimate the diffusion model that involves hopping on multiple energy surfaces. The main idea is that there are hidden states with their own energy surfaces and diffusion coefficients, and one of these states is taken (or more probable) at every snapshot of a trajectory. For example, a trajectory has segments in which they are in state 1 or 2 (Fig. 1A). Thus, in addition to estimating energy profiles and diffusion coefficient, we need to know which state is taken at every snapshot of the trajectory. This type of problem can be solved by hidden Markov model (HMM) (49). The likelihood of a trajectory $X = (x_0, x_1, \cdots, x_N)$ given a model becomes,

$$L^{msurf} = \sum_{S} \prod_{i=0}^{N-2} p_{s_i}(x_{i+1}, t_{i+1}, |x_i, t_i) \cdot p(s_i \to s_{i+1}) \cdot p_{s_{i+1}}(x_{i+2}, t_{i+2} | x_{i+1}, t_{i+1})$$

where $p_{s_i}(x_{i+1}, t_{i+1}, |x_i, t_i) \equiv p_{s_i}(bin_{i+1}, t_{i+1}, |bin_i, t_i) = p_{s_i}(bin_{i+1}, \Delta t | bin_i, 0)$ is the positional transition probability discretized into bins with lag time $\Delta t = t_{i+1} - t_i$ in state s_i and $p(s_i \to s_{i+1})$ is the switching probability of the state from s_i to s_{i+1} in time Δt. All possible paths of hidden states $S = (s_0, s_1, \cdots, s_{N-1})$ are summed over. $p_{s_i}(bin_{i+1}, \Delta t | bin_i, 0)$ was computed in the same manner as the one-energy-surface case developed by Hummer and coworkers (42, 44, 46), and $p(s_i \to s_{i+1})$ was calculated from switching rates (see Methods).

In context of HMM, observed data are positional transitions between bins, $p(s_i \to s_{i+1})$ is the state transition probability, and $p_{s_i}(l, \Delta t | j, 0)$ is proportional to the output probability of the particular transition bin $j \to$ bin l in state s_i (see Fig. 1B). Because sum of the output probabilities has to be one, the positional transition probabilities were divided by number of bins to be used as the output probabilities. Given a model (that is, energy surface and diffusion coefficient for each state and switching rates), the likelihood L^{msurf} and the state probability $p_{t_i}(s) \equiv p(s_i = s | X)$ at any point in the trajectory can be computed by the forward-backward algorithm of HMM. $p_t(s)$ indicates which energy surface is likely taken at time *t*.

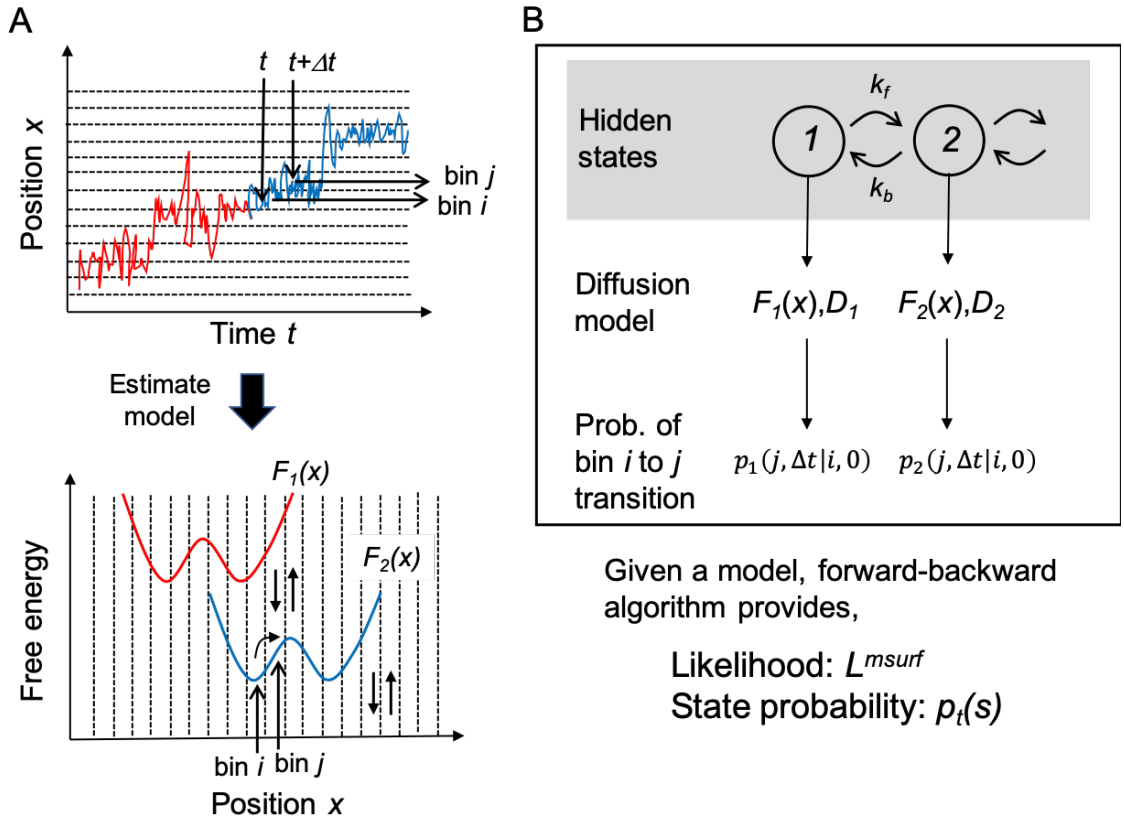

Figure 1. Estimation of dynamic free energy landscape. (A) Estimation of underlying free energy surfaces and diffusion coefficient from one-dimensional trajectories. Red and blue segments of the trajectory shown in top panel correspond to diffusion on red and blue energy surfaces in bottom panel, respectively. (B) Scheme of HMM to estimate diffusion model on multiple energy surfaces.

Reconstruction of diffusion model from simulation trajectories

We tested the method with simulation trajectories for which we know the true model parameters. Fig. 2A shows two energy surfaces used for Brownian dynamics simulation with stochastic switching between the two. Also shown are two out of thirty independent trajectories generated by the simulation. The double-well potential is intended to mimic a situation of chitinase that moves back and forth before catalytic reaction happens. The size and time constant of chitinase stepping motions in mind, units of length and time in Fig. 2 can be interpreted as nanometer and second, respectively. The trajectories were generated by Brownian dynamics simulation for diffusive motions and kinetic Monte Carlo for switching between the two potentials with specified rate constants (see

Methods). The generated trajectories were used to reconstruct underlying energy profile (F) and diffusion coefficient (D). In the estimation, the trajectories were discretized into bins with bin size $\Delta x = 0.1$, number of bins 27, in the range $-0.35 \leq x \leq 2.35$, and transitions between bins after lag step 500 were used as observed data. In the framework of Bayesian inference, the posterior probability L' was maximized in terms of F and D as well as switching rates k_f and k_b by the Monte Carlo optimization (see Methods).

The estimated energy profiles and diffusion coefficient are shown in Fig. 2B and 2C, respectively. The estimated energy profiles averaged over last 20% of the Monte Carlo optimization steps are shown in red and blue lines with the true profile in black line (Fig. 2B). The estimated diffusion coefficient along the Monte Carlo optimization steps is shown with the true value in broken line (Fig. 2C). Both the energy profiles and diffusion coefficient agree quite well with the true profiles. The estimated switching rates agree reasonably well with the true values (Fig. 2D). Furthermore, the estimated state probabilities correctly reproduce timing of the potential switching between the energy surfaces (Fig. 2E). Note that the estimated switching rates were less accurate than the energy profiles and diffusion coefficient, probably because the switching only occurs once or so in each trajectory.

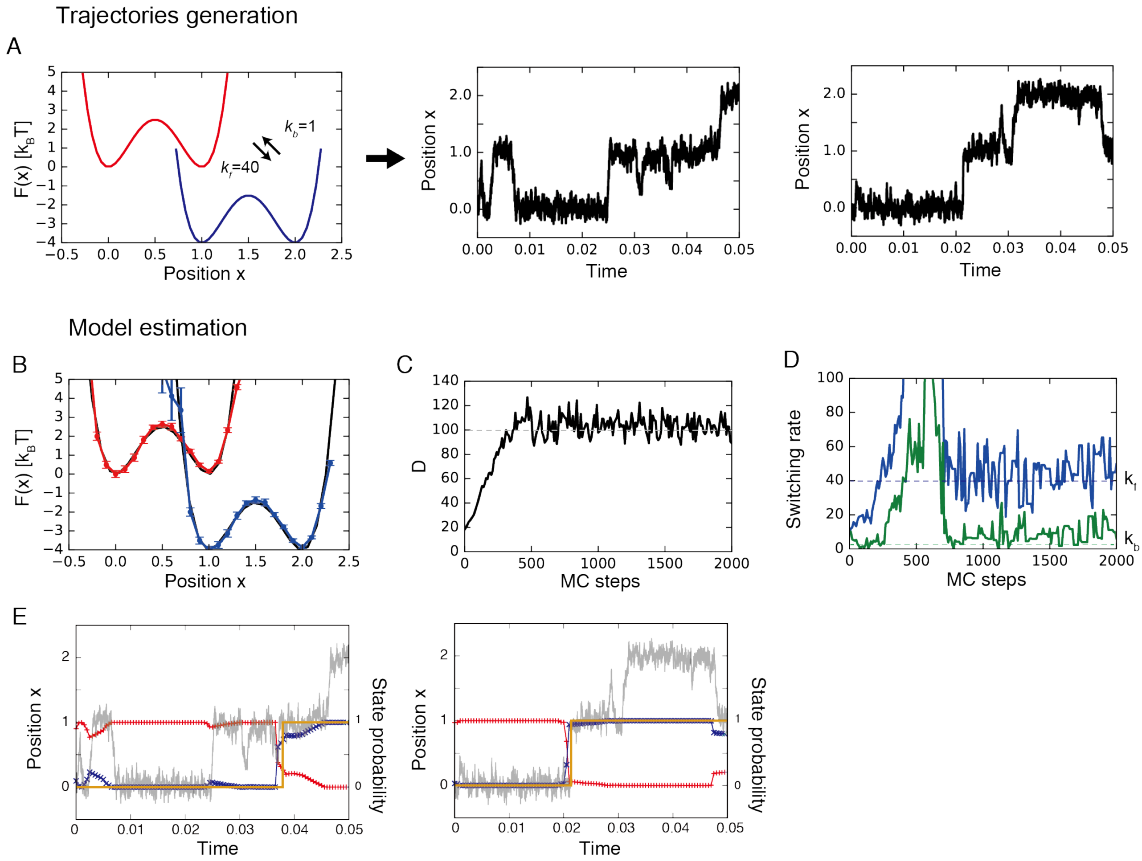

Figure 2. Generation of simulation trajectories and model estimation using the trajectories. (A) Two energy surfaces used for Brownian dynamics simulation with stochastic switching (left panel) and two trajectories generated by the simulation (center and right panels). (B) Estimated energy profiles are shown in red and blue lines with the true energy profile in black lines. (C) Estimated diffusion coefficient during the MC optimization steps (black line) with the true diffusion coefficient (gray broken line). (D) Estimated switching rates in forward and backward direction are shown in blue and green, respectively, with the true values shown in broken lines. (E) The two trajectories with estimated state probability are shown. The orange lines represent the potential switching, whose value 0 (1) means that it is on the red (blue) potential. Estimated state probabilities of the red potential (red line) and blue potential (blue line) are also shown. Length and time units are arbitrary, and can be interpreted as nanometer and second, respectively, in comparison with experiments.

Accuracy of the estimated energy profiles and diffusion coefficient, however, depends on the lag time Δt (or lag step) used to collect positional transition data between bins. In Fig. 3A, Root-Mean-Square-Deviations (RMSDs) of the estimated profiles to the true ones with different lag steps are plotted. We found that RMSD of the energy profile decreases and takes low values around lag step 300 ~ 600 and then increase, while that of diffusion coefficient converges around lag step 300. This result suggests that we have to use the lag step long enough and also not too long to avoid insufficient transition data. In an attempt to decide the optimal lag time without knowing the true profiles, we performed the lag-time test to see how slow timescale converges with increasing lag time (42, 50). The first non-trivial eigenvalue λ_2 of the rate matrix estimated from transition data with the certain lag time was converted into the so-called implied timescale by $\tau_2 = -1/\lambda_2$ (see Methods). The implied timescale along lag steps showed a convergence around lag step 200 (Fig. 3B). We focus on only the first non-trivial mode, because there is a clear timescale gap after this mode. The low RMSDs to the true profiles were achieved after the convergence point of the implied timescale (Fig. 3A and B). Thus, without knowing the true profiles, it is a good idea to use the lag time longer than the convergence point of the implied timescale to achieve high accuracy.

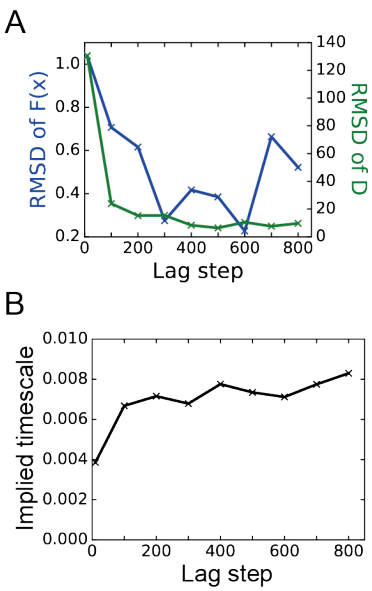

Figure 3. Lag-time dependence of the estimation accuracy. (A) Lag-step dependence of RMSDs of the energy profile and diffusion coefficient to the true profiles. (B) Lag-step dependence of the slowest implied timescale.

Effect of background noise

In the previous section, we showed that our method accurately estimates the underlying dynamic energy landscape and diffusion coefficient from simulation trajectories. However, experimental single-molecule trajectories always contain some background noise. In our case, the 40 nm gold nanoparticle (AuNP), which was used to label chitinase to observe its stepping motions, showed a localization precision of ~0.3 nm at 0.5 ms time resolution with the total internal reflection dark-field microscopy (13, 51). The localization precision is restricted by the detector saturation, and can be improved (51). Still, the limited localization precision as the background noise should be addressed.

By adding background noise to simulation trajectories, we investigated how it affects our estimation. Gaussian noise with mean zero and standard deviation 0.3 nm was added to the position x saved every 500 steps ($\Delta t = 0.5\ \mathrm{ms}$), mimicking the experimental background noise. The median filter with various window sizes was applied to reduce noise level (Fig. 4A). Then, the trajectories were analyzed by our method. To deal with the noise, we used a bigger bin size $\Delta x = 0.2$ with total number of bins 13 in the range $-0.3 \leq x \leq 2.3$. Although it was difficult to fully recover the true energy profiles and diffusion coefficient, they were recovered to some extent in a manner that depends on window size of the median filter. The estimated energy profiles recovered the overall double-well shape with lower energy barriers (Fig. 4B). Using trajectories median-filtered with larger window size resulted in more accurate energy profiles. On contrary, the diffusion coefficients were underestimated further by using larger window size (Fig. 4C). The timing of the potential switching was recovered reasonably well (Fig. 4D), although there were wrongly assigned switching events occasionally. In ~30% of the trajectories, there were some incorrectly assigned transitions. These incorrect assignments mainly arose from altered fluctuations due to the background noise. For example, in the first trajectory in Fig. 4A and 4D, there is an artificial fluctuation in the median-filtered trajectory, which was not present in the original trajectory, to the positive direction from x~1.0 at t~0.025 (indicated by the green arrows). This artificial fluctuation misled our model to assign the incorrect and instantaneous transition at t~0.025 (Fig. 4D). Note that the standard deviation of the background noise used here 0.3 nm is close to the step size 1 nm, and considered as a hard test. Regardless of the limited estimation

accuracy with the background noise, we were able to obtain the essential feature of the energy profiles and rough estimate of the diffusion coefficient.

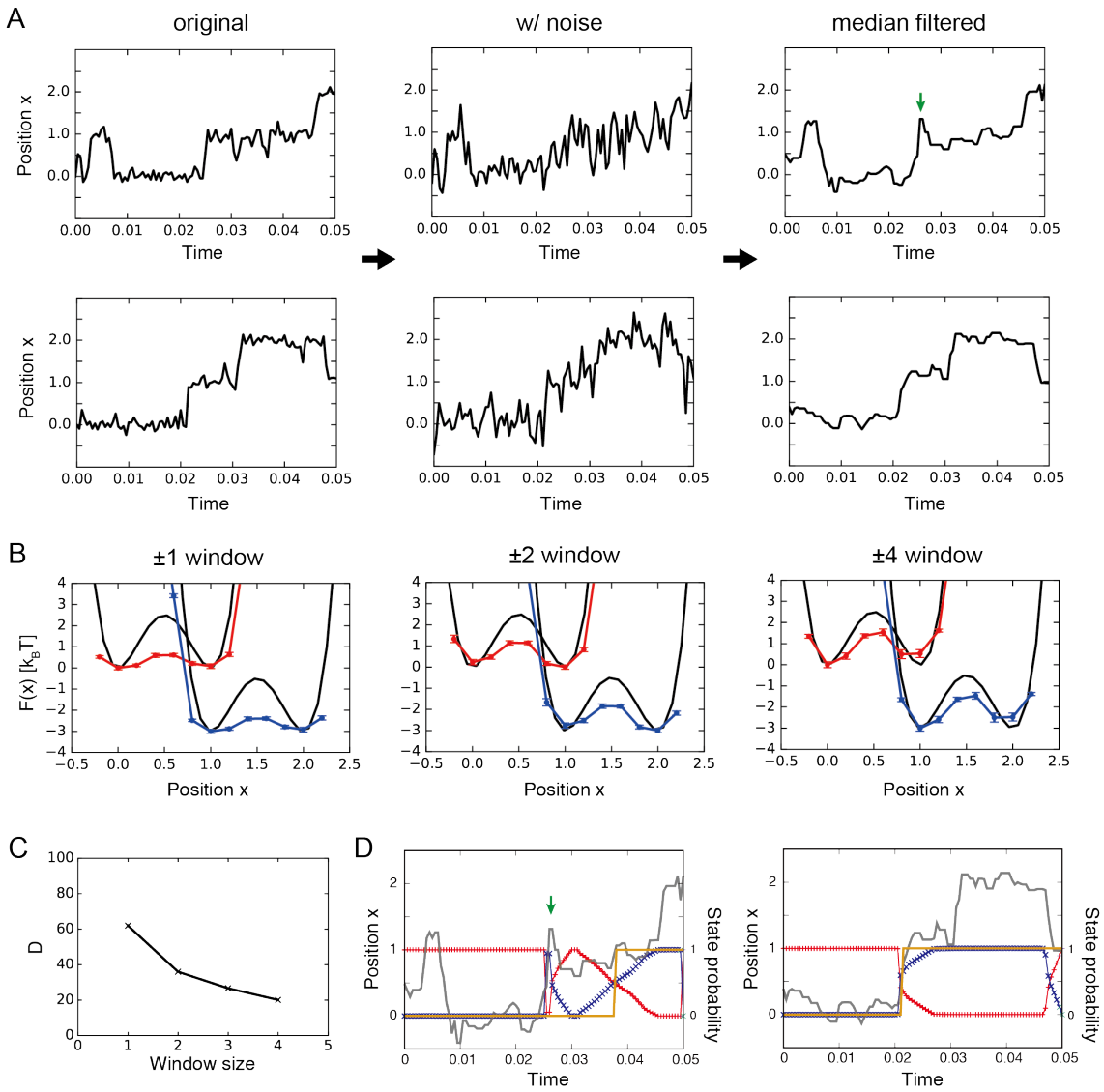

Figure 4. Effect of background noise. (A) Background noise was added to the original simulation trajectory. Then, the noise was reduced by the median filter. The green arrow indicates the artificial fluctuation (see main text). (B) Estimated energy profiles from trajectories processed by the median filter with different window sizes are shown. The window sizes of 1, 2 and 4 are shown in left, center and right panels, respectively. (C) Estimated diffusion coefficients from trajectories processed by the median filter with different window sizes are plotted. (D) Two trajectories median-filtered with window size 2 are shown in gray lines. The orange lines represent the potential switching, whose value

0 (1) means that it is on the red (blue) potential. Estimated state probabilities of state 1 (red energy surface) and 2 (blue energy surface) are shown in red and blue lines, respectively.

The prior probability of the model seems to be important for reliable estimation of the model from trajectories with the background noise. In the analysis of original simulation trajectories without noise (Fig. 2), we used a weak prior probability on the peripheral regions of the energy surfaces ($\sigma = 1.0\ k_BT,\ \delta = 1\ k_BT$ see Methods), which did not affect the estimation. However, using the weak prior probability for trajectories with background noise resulted in mixing of the two energy surfaces (Fig. S1). Consequently, timing of surface switching became obscure. To deal with this situation, we used a stronger prior probability for trajectories with background noise ($\sigma = 0.1\ k_BT,\ \delta = 6\ k_BT$ was used in Fig. 4).

Application to single-molecule trajectories of chitinase

We applied the method to single-molecule trajectories of chitinase. Chitinase is a unique biomolecular motor that moves unidirectionally on crystalline chitin surface with hydrolysis energy of single chitin chain. With single-molecule experiments, we previously showed that it operates by the burnt-bridge Brownian ratchet mechanism, where it is driven by fast catalytic reaction and biased Brownian motions (13). We clarified the molecular mechanism of how chitinase change its free energy landscape so that its motion is biased in forward direction (Fig. 5A). Here, we aim to estimate the dynamic free energy landscape induced by the chemical-state change, directly from single-molecule trajectories.

In the single-molecule experiments, chitinase A from bacteria *Serratia marcescens* was labeled with the 40-nm gold nanoparticle (Fig. 5B), and its unidirectional motion on chitin surface hydrolyzing the chitin chain was observed by the total-internal reflection dark-field microscopy with 0.3 nm localization precision and 0.5 ms temporal resolution (13). We used trajectories that were median-filtered to reduce positional noise. Because the experimental trajectories were rather long in time and a few in number, we divided original trajectories into segments and aligned them to increase sample size. In Fig. 5C,

we show how the trajectory segments were generated. Each trajectory segment has two forward steps that were identified by the step-finding algorithm (13). Then, they were aligned by matching their zero-points (see Methods). All trajectory segments analyzed are shown in Fig. S2.

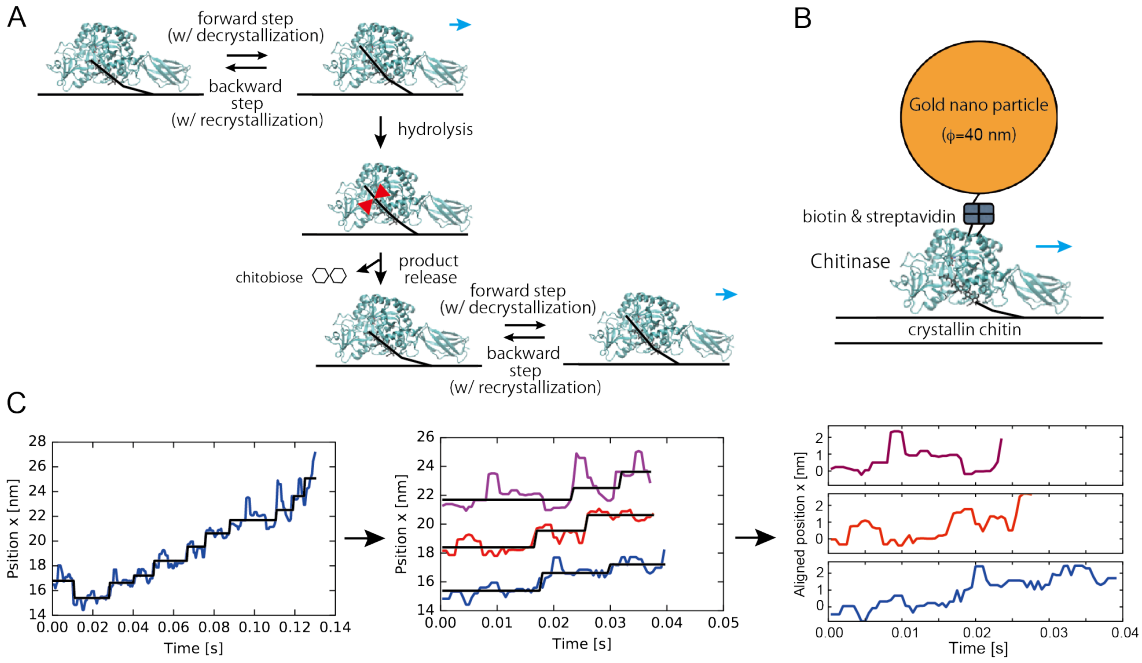

Figure 5. Single-molecule trajectories of chitinase. (A) Mechanochemical coupling scheme of chitinase A from bacteria *Serratia marcescens*. Cyan arrows represent movement direction of chitinase. Red saw teeth represent hydrolysis of single chitin chain shown in black line. (B) Experimental setup of chitinase labeled with gold nanoparticle on crystalline chitin. (C) A single-molecule trajectory of chitinase at 2000 frames per second (0.5 ms temporal resolution, left panel) was divided into three trajectory segments with two forward steps (center panel). The black lines represent steps identified by the step-finding algorithm. The trajectory segments were shifted to start from the origin (right panel).

The trajectory segments were analyzed in the same way as the test cases above. Since trajectory segments have two forward steps, we assume two states (free energy surfaces) that correspond to before and after the chemical-state change (hydrolysis reaction and subsequent product release). This assumption is supported by the molecular insights that

each forward step corresponds to the sliding of chitin chain through the cleft of chitinase and after the step the catalytic reaction and the subsequent product release reset the chitin-chitinase interaction (Fig. 5A). In the estimation, the bin size $\Delta x = 0.2$ nm and number of bins 15 were used in the rage of $-0.5 \leq x \leq 2.5$. For translational symmetry, shifting length 1 nm was used, which was taken from chitinase step size (see Methods). Switching rates between the two free energy surfaces were fixed and determined as follows. The forward switching rate was determined from experimental time constant of the forward steps ~25 ms, which includes diffusive motion over energy barrier (decrystallization and chain sliding), catalytic reaction and product release. After completing these processes, free energy surface is switched. Because the backward switching should not happen in the current experimental setup (this will lead to a synthetic reaction that elongates the chitin chain), its rate was chosen to be sufficiently slow.

The estimated diffusion model is shown in Fig. 6. The estimated free energy surfaces capture essential features of chitinase motions (Fig. 6A). There are minima around 0 and 0.8 nm on the pre-catalytic free energy surface, and around 1 and 1.8 nm on the post-catalytic free energy surface. These minima separated by a low barrier are responsible for stepping motions of chitinase. After the chemical-state change, high energy region $x >$ 1.5 nm on the pre-catalytic surface goes down to make a new minimum around the region on the post-catalytic free energy surface. Additional metastable positions around 1.5 and 2.5 nm on the pre-catalytic and post-catalytic surfaces, respectively, may explain variations in step size. The estimated diffusion coefficient depends on window size of the median filter (Fig. 6B). The larger is the window size, the smaller is the estimated diffusion coefficient. An extrapolation suggests that the diffusion coefficient is 300 ~ 400 $nm^2\ s^{-1}$. This value is five orders of magnitude smaller than the calculated diffusion coefficient of the 40 nm gold probe D_{AuNP} (see Methods). This result explains the experimental observation that the velocity of chitinase is not affected by size of the probes (13). The estimated time series of the state probabilities provide further information on how our model predicts switching of the two free energy surfaces (Fig. 6C). Typically, the switching happens in the range of $0.5 \leq x \leq 1.5$. Note that our model does not have any constraint where the switching happens. This result suggests that a forward motion of chitinase coupled with the sliding motion of the chitin chain into the catalytic site facilitates the catalytic reaction, which results in the forward switching of the free energy surface.

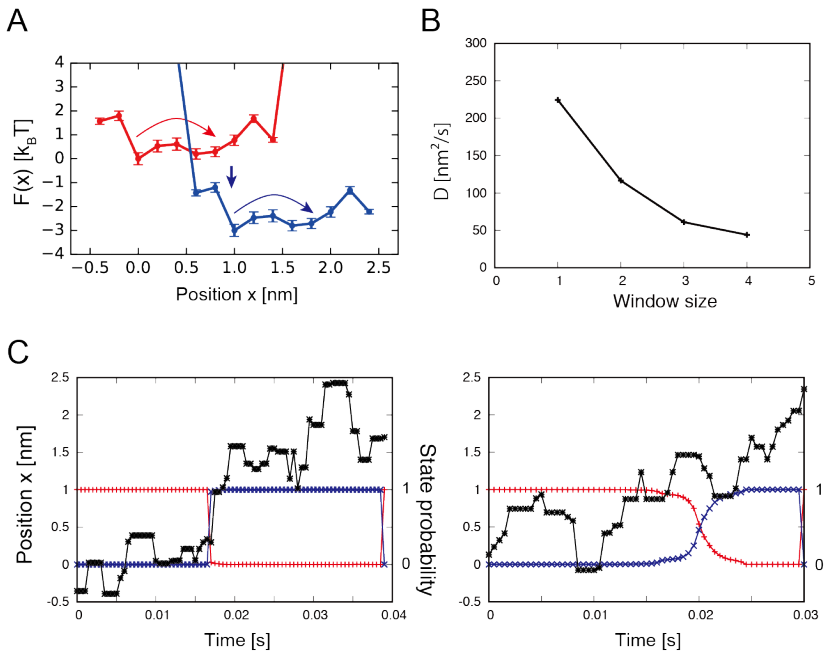

Figure 6. Estimated diffusion model of chitinase. (A) Estimated free energy profiles of before and after the chemical-state change are shown in red and blue, respectively. Arrows represent diffusive motions over barriers and switching of energy surfaces. (B) Estimated diffusion coefficients with different window sizes of the median filter. (C) Examples of trajectory segments (black) and estimated state probability of before (red) and after (blue) the chemical state change are shown.

Discussion

The accuracy of the dynamic free energy profiles and diffusion coefficient depends on the localization precision and temporal resolution of the single-molecule trajectory data. We used the single-molecule trajectories of chitinase observed with 0.3 nm localization precision and 0.5 ms temporal resolution (13). We selected five trajectories that show clear stepping motions for the estimation (Fig. S2). The clear stepping motions were necessary for aligning trajectory segments with two forward steps used in the estimation (Fig. 5C). To tackle the background noise, we used the median filter with various window

sizes. From the test analysis of the simulation trajectories with background noise, it is expected that our method underestimates barrier height and diffusion coefficient. Indeed, the estimated free energy profiles have a low barrier (~1 k_BT), which is likely due to the background noise. Measurements with better localization precision and higher temporal resolution will allow higher quality estimation in future (51).

The estimated friction coefficient with the Einstein-Smoluchowski relation $\xi = k_BT/D$ provides an insight into the energy-conversion mechanism of chitinase. The viscous drag force $\xi\overline{v}$ with the average speed of chitinase $\overline{v}$ = 50 nm s^{-1} (13) through the moving length 1 nm in one cycle results in amount of work ~ 0.1 k_BT. This negligible work against the viscous load, that is, very low Stokes efficiency (52), makes a clear distinction from the rotary motor F_1-ATPase (53-56) and suggests that most of the input hydrolysis energy of chitin is used for something else other than mechanical movement of the motor. The important difference from typical ATPase motors is that the chitin rail becomes one chitobiose-unit shorter every catalytic cycle. Thus, it is essential to take the rail into consideration. When we hypothetically consider only the chitin rail, two reactions happen every cycle: decrystallization of the chitobiose unit and hydrolysis reaction (Fig. 5A). Since the decrystallization is an energy-consuming process, it can be said that, as the net energy balance, the input hydrolysis energy is used for decrystallization of one chitobiose unit from the crystalline β-chitin (the type of chitin used in the experiments). Actually, free energies for these two reactions are comparable with the opposite signs, although only approximate values of the related reactions are available: the hydrolysis free energy of the β-1,4-glycosidic bond in cellulose -16 kJ mol^{-1} $\approx -6.5\,k_BT$ (57) and the decrystallization free energy per chitobiose for α-chitin 5 kcal $mol^{-1} \approx 8\,k_BT$ (58). When we consider the whole chitinase-chitin system, the above simple discussion regarding the net energy balance can be understood in a step-wise manner. In the catalytic cycle, starting from the state before the decrystallization, chitinase decrystallizes the chitin chain and makes a forward step with the chitin chain sliding and binding to the product binding site of chitinase. This binding free energy can compensate the decrystallization energy, which leads to the equally stable state after the forward step (13). Then the hydrolysis reaction of the glycosidic bond and chitobiose release reset the system to the initial state, where the hydrolysis energy is liberated and used for expelling the bound product. In this way, the hydrolysis energy is borrowed in advance as a form of the binding energy of

chitin to the product site to decrystallize the chitobiose unit, and then paid back at the product release after the actual hydrolysis reaction happens. In other words, chitinase develops the mechanism to stabilize the high-energy intermediate due to the decrystallization, which is reset afterwards by the hydrolysis reaction.

Lastly, we mention a previously developed method by Toyabe and colleagues (59) for estimation of chemical-state dependent free energy profiles. Although they used the similar likelihood function (without summing over the hidden state), expression of the positional transition probability and optimization process were quite different from our method. While our method estimates free energy profiles and diffusion coefficient at the same time, their method only optimizes free energy profiles with fixed diffusion coefficient estimated separately. In this regard, our method has an advantage to provide an efficient and consistent estimation. One possible improvement might be explicit consideration of position-dependent switching rate of the energy surfaces (60), which are missing in both methods. However, this feature requires an optimization of a large set of parameters and very precise measurements of positions of the motor. The estimated dynamic free energy landscape by the present method would serve as a basis of developing more detailed models.

Conclusions

Dynamic free energy landscape triggered by the chemical-state changes is the basis of mechanochemical coupling of biomolecular motors. In this study, we introduced a framework to estimate the chemical-state dependent free energy profiles and diffusion coefficient from single-molecule trajectories of biomolecular motors. The method was tested with simulation trajectories, and successfully reproduced energy profiles and diffusion coefficient. Then, the method was applied to single-molecule trajectories of chitinase. The estimated dynamic free energy landscape and diffusion coefficient provide a physical basis to the burnt-bridge Brownian ratchet model previously suggested (13). The double-well free energy profile is shifted forward by the chemical-state change to drive unidirectional motion of chitinase (Fig. 6A). This mechanism makes a clear distinction from the power-stroke mechanism, where it is assumed to have some driving

slope in the free energy profile (20). The estimated diffusion coefficient further provides an insight into the energy-conversion mechanism of chitinase.

Author Contribution

K.O., A.N. and R.I. conceived the study. K.O. designed algorithm and performed research. K.O., A.N. and R.I. discussed results. K.O. wrote manuscript with contributions from A.N. and R.I.

Acknowledgements

K.O. would like to thank Dr. Yohei Nakayama for discussion. This work was supported by Interdisciplinary Research Promotion Project (J281002 to K.O.) of the National Institutes of Natural Sciences (NINS) and by Grant-in-Aid for Scientific Research on Innovative Areas “Molecular Engine” (grant number JP18H05424 to R.I.). K.O. was supported by Building of Consortia for the Development of Human Resources in Science and Technology, MEXT, Japan.

References

1. Noji, H., R. Yasuda, M. Yoshida, and K. Kinosita. 1997. Direct observation of the rotation of F-1-ATPase. Nature 386:299-302.
2. Adachi, K., K. Oiwa, T. Nishizaka, S. Furuike, H. Noji, H. Itoh, M. Yoshida, and K. Kinosita. 2007. Coupling of rotation and catalysis in F-1-ATPase revealed by single-molecule imaging and manipulation. Cell 130:309-321.
3. Watanabe, R., R. Iino, and H. Noji. 2010. Phosphate release in F-1-ATPase catalytic cycle follows ADP release. Nat Chem Biol 6:814-820.
4. Toyabe, S., T. Watanabe-Nakayama, T. Okamoto, S. Kudo, and E. Muneyuki. 2011. Thermodynamic efficiency and mechanochemical coupling of F1-ATPase. Proc Natl Acad Sci U S A 108:17951-17956.

5. Sugawa, M., K. Okazaki, M. Kobayashi, T. Matsui, G. Hummer, T. Masaike, and T. Nishizaka. 2016. F1-ATPase conformational cycle from simultaneous single-molecule FRET and rotation measurements. Proc Natl Acad Sci U S A 113:E2916-2924.
6. Yildiz, A., J. N. Forkey, S. A. McKinney, T. Ha, Y. E. Goldman, and P. R. Selvin. 2003. Myosin V walks hand-over-hand: Single fluorophore imaging with 1.5-nm localization. Science 300:2061-2065.
7. Kodera, N., D. Yamamoto, R. Ishikawa, and T. Ando. 2010. Video imaging of walking myosin V by high-speed atomic force microscopy. Nature 468:72-76.
8. Svoboda, K., C. F. Schmidt, B. J. Schnapp, and S. M. Block. 1993. Direct Observation of Kinesin Stepping by Optical Trapping Interferometry. Nature 365:721-727.
9. Isojima, H., R. Iino, Y. Niitani, H. Noji, and M. Tomishige. 2016. Direct observation of intermediate states during the stepping motion of kinesin-1. Nat Chem Biol 12:290-297.
10. Ariga, T., M. Tomishige, and D. Mizuno. 2018. Nonequilibrium Energetics of Molecular Motor Kinesin. Phys Rev Lett 121.
11. Igarashi, K., T. Uchihashi, A. Koivula, M. Wada, S. Kimura, T. Okamoto, M. Penttila, T. Ando, and M. Samejima. 2011. Traffic jams reduce hydrolytic efficiency of cellulase on cellulose surface. Science 333:1279-1282.
12. Igarashi, K., T. Uchihashi, T. Uchiyama, H. Sugimoto, M. Wada, K. Suzuki, S. Sakuda, T. Ando, T. Watanabe, and M. Samejima. 2014. Two-way traffic of glycoside hydrolase family 18 processive chitinases on crystalline chitin. Nat Commun 5:3975.
13. Nakamura, A., K. I. Okazaki, T. Furuta, M. Sakurai, and R. Iino. 2018. Processive chitinase is Brownian monorail operated by fast catalysis after peeling rail from crystalline chitin. Nat Commun 9:3814.
14. Beckham, G. T., Y. J. Bomble, E. A. Bayer, M. E. Himmel, and M. F. Crowley. 2011. Applications of computational science for understanding enzymatic deconstruction of cellulose. Curr Opin Biotechnol 22:231-238.
15. Knott, B. C., M. F. Crowley, M. E. Himmel, J. Stahlberg, and G. T. Beckham. 2014. Carbohydrate-protein interactions that drive processive polysaccharide translocation in enzymes revealed from a computational study of

cellobiohydrolase processivity. J Am Chem Soc 136:8810-8819.

16. Beckham, G. T., J. Stahlberg, B. C. Knott, M. E. Himmel, M. F. Crowley, M. Sandgren, M. Sorlie, and C. M. Payne. 2014. Towards a molecular-level theory of carbohydrate processivity in glycoside hydrolases. Curr Opin Biotechnol 27:96-106.
17. Saffarian, S., I. E. Collier, B. L. Marmer, E. L. Elson, and G. Goldberg. 2004. Interstitial collagenase is a Brownian ratchet driven by proteolysis of collagen. Science 306:108-111.
18. Saffarian, S., H. Qian, I. Collier, E. Elson, and G. Goldberg. 2006. Powering a burnt bridges Brownian ratchet: a model for an extracellular motor driven by proteolysis of collagen. Phys Rev E Stat Nonlin Soft Matter Phys 73:041909.
19. Julicher, F., A. Ajdari, and J. Prost. 1997. Modeling molecular motors. Rev Mod Phys 69:1269-1281.
20. Wang, H., and G. Oster. 2002. Ratchets, power strokes, and molecular motors. Appl Phys a-Mater 75:315-323.
21. Takano, M., T. P. Terada, and M. Sasai. 2010. Unidirectional Brownian motion observed in an in silico single molecule experiment of an actomyosin motor. Proc Natl Acad Sci U S A 107:7769-7774.
22. Kanada, R., R. Shinagawa, and K. Sasaki. 2018. Diffusion enhancement of Brownian motors revealed by a solvable model. Phys Rev E 98.
23. Mai, J., I. M. Sokolov, and A. Blumen. 2001. Directed particle diffusion under "burnt bridges" conditions. Phys Rev E 64.
24. McKinney, S. A., C. Joo, and T. Ha. 2006. Analysis of single-molecule FRET trajectories using hidden Markov modeling. Biophys J 91:1941-1951.
25. Gopich, I. V., and A. Szabo. 2009. Decoding the pattern of photon colors in single-molecule FRET. J Phys Chem B 113:10965-10973.
26. Keller, B. G., A. Kobitski, A. Jaschke, G. U. Nienhaus, and F. Noe. 2014. Complex RNA folding kinetics revealed by single-molecule FRET and hidden Markov models. J Am Chem Soc 136:4534-4543.
27. Chung, H. S., K. McHale, J. M. Louis, and W. A. Eaton. 2012. Single-Molecule Fluorescence Experiments Determine Protein Folding Transition Path Times. Science 335:981-984.
28. Chung, H. S., and W. A. Eaton. 2018. Protein folding transition path times from

single molecule FRET. Curr Opin Struct Biol 48:30-39.
29. Haas, K. R., H. Yang, and J. W. Chu. 2013. Expectation-maximization of the potential of mean force and diffusion coefficient in Langevin dynamics from single molecule FRET data photon by photon. J Phys Chem B 117:15591-15605.
30. Das, R., C. W. Cairo, and D. Coombs. 2009. A hidden Markov model for single particle tracks quantifies dynamic interactions between LFA-1 and the actin cytoskeleton. PLoS Comput Biol 5:e1000556.
31. Onuchic, J. N., and P. G. Wolynes. 2004. Theory of protein folding. Curr Opin Struc Biol 14:70-75.
32. Okazaki, K., N. Koga, S. Takada, J. N. Onuchic, and P. G. Wolynes. 2006. Multiple-basin energy landscapes for large-amplitude conformational motions of proteins: Structure-based molecular dynamics simulations. P Natl Acad Sci USA 103:11844-11849.
33. Okazaki, K., and S. Takada. 2008. Dynamic energy landscape view of coupled binding and protein conformational change: induced-fit versus population-shift mechanisms. Proc Natl Acad Sci U S A 105:11182-11187.
34. Whitford, P. C., J. N. Onuchic, and K. Y. Sanbonmatsu. 2010. Connecting energy landscapes with experimental rates for aminoacyl-tRNA accommodation in the ribosome. J Am Chem Soc 132:13170-13171.
35. Whitford, P. C., and J. N. Onuchic. 2015. What protein folding teaches us about biological function and molecular machines. Curr Opin Struct Biol 30:57-62.
36. Okazaki, K., and G. Hummer. 2013. Phosphate release coupled to rotary motion of F-1-ATPase. P Natl Acad Sci USA 110:16468-16473.
37. Baba, A., and T. Komatsuzaki. 2007. Construction of effective free energy landscape from single-molecule time series. Proc Natl Acad Sci U S A 104:19297-19302.
38. Okuno, D., R. Iino, and H. Noji. 2010. Stiffness of gamma subunit of F(1)-ATPase. Eur Biophys J 39:1589-1596.
39. Elston, T., H. Y. Wang, and G. Oster. 1998. Energy transduction in ATP synthase. Nature 391:510-513.
40. Wang, H. Y., and G. Oster. 1998. Energy transduction in the F-1 motor of ATP synthase. Nature 396:279-282.
41. Best, R. B., and G. Hummer. 2010. Coordinate-dependent diffusion in protein

folding. P Natl Acad Sci USA 107:1088-1093.

42. Best, R. B., and G. Hummer. 2011. Diffusion models of protein folding. Phys Chem Chem Phys 13:16902-16911.

43. Neupane, K., A. P. Manuel, and M. T. Woodside. 2016. Protein folding trajectories can be described quantitatively by one-dimensional diffusion over measured energy landscapes. Nat Phys 12:700-+.

44. Hummer, G. 2005. Position-dependent diffusion coefficients and free energies from Bayesian analysis of equilibrium and replica molecular dynamics simulations. New J Phys 7.

45. Manuel, A. P., J. Lambert, and M. T. Woodside. 2015. Reconstructing folding energy landscapes from splitting probability analysis of single-molecule trajectories. P Natl Acad Sci USA 112:7183-7188.

46. Buchete, N. V., and G. Hummer. 2008. Coarse master equations for peptide folding dynamics. Journal of Physical Chemistry B 112:6057-6069.

47. Bicout, D. J., and A. Szabo. 1998. Electron transfer reaction dynamics in non-Debye solvents. J Chem Phys 109:2325-2338.

48. Gillespie, D. T. 1977. Exact Stochastic Simulation of Coupled Chemical-Reactions. J Phys Chem-Us 81:2340-2361.

49. Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. 2007. Numerical Recipes 3rd Edition: The Art of Scientific Computing. Cambridge University Press.

50. Chodera, J. D., and F. Noe. 2014. Markov state models of biomolecular conformational dynamics. Curr Opin Struc Biol 25:135-144.

51. Ando, J., A. Nakamura, A. Visootsat, M. Yamamoto, C. Song, K. Murata, and R. Iino. 2018. Single-Nanoparticle Tracking with Angstrom Localization Precision and Microsecond Time Resolution. Biophys J 115:2413-2427.

52. Wang, H. Y., and G. Oster. 2002. The Stokes efficiency for molecular motors and its applications. Europhys Lett 57:134-140.

53. Yasuda, R., H. Noji, K. Kinosita, and M. Yoshida. 1998. F-1-ATPase is a highly efficient molecular motor that rotates with discrete 120 degrees steps. Cell 93:1117-1124.

54. Kinosita, K., K. Adachi, and H. Itoh. 2004. Rotation of F-1-ATPase: How an ATP-driven molecular machine may work. Annu Rev Bioph Biom 33:245-268.

55. Toyabe, S., T. Okamoto, T. Watanabe-Nakayama, H. Taketani, S. Kudo, and E. Muneyuki. 2010. Nonequilibrium energetics of a single F1-ATPase molecule. Phys Rev Lett 104:198103.
56. Okazaki, K., and G. Hummer. 2015. Elasticity, friction, and pathway of gamma-subunit rotation in FoF1-ATP synthase. Proc Natl Acad Sci U S A 112:10720-10725.
57. Sorensen, T. H., N. Cruys-Bagger, K. Borch, and P. Westh. 2015. Free Energy Diagram for the Heterogeneous Enzymatic Hydrolysis of Glycosidic Bonds in Cellulose. J Biol Chem 290:22203-22211.
58. Beckham, G. T., and M. F. Crowley. 2011. Examination of the alpha-Chitin Structure and Decrystallization Thermodynamics at the Nanoscale. Journal of Physical Chemistry B 115:4516-4522.
59. Toyabe, S., H. Ueno, and E. Muneyuki. 2012. Recovery of state-specific potential of molecular motor from single-molecule trajectory. Epl-Europhys Lett 97.
60. Cossio, P., G. Hummer, and A. Szabo. 2018. Transition paths in single-molecule force spectroscopy. J Chem Phys 148.