

RESEARCH ARTICLE

Decoding Single Molecule Time Traces with Dynamic Disorder

Wonseok Hwang¹, Il-Buem Lee², Seok-Cheol Hong^{1,2}, Changbong Hyeon^{1*}

1 Korea Institute for Advanced Study, Seoul, Republic of Korea, **2** Department of Physics, Korea University, Seoul, Republic of Korea

* hyeoncb@kias.re.kr



Abstract

Single molecule time trajectories of biomolecules provide glimpses into complex folding landscapes that are difficult to visualize using conventional ensemble measurements. Recent experiments and theoretical analyses have highlighted dynamic disorder in certain classes of biomolecules, whose dynamic pattern of conformational transitions is affected by slower transition dynamics of internal state hidden in a low dimensional projection. A systematic means to analyze such data is, however, currently not well developed. Here we report a new algorithm—Variational Bayes-double chain Markov model (VB-DCMM)—to analyze single molecule time trajectories that display dynamic disorder. The proposed analysis employing VB-DCMM allows us to detect the presence of dynamic disorder, if any, in each trajectory, identify the number of internal states, and estimate transition rates between the internal states as well as the rates of conformational transition within each internal state. Applying VB-DCMM algorithm to single molecule FRET data of H-DNA in 100 mM-Na⁺ solution, followed by data clustering, we show that at least 6 kinetic paths linking 4 distinct internal states are required to correctly interpret the duplex-triplex transitions of H-DNA.

OPEN ACCESS

Citation: Hwang W, Lee I-B, Hong S-C, Hyeon C (2016) Decoding Single Molecule Time Traces with Dynamic Disorder. *PLoS Comput Biol* 12(12): e1005286. doi:10.1371/journal.pcbi.1005286

Editor: Bert L. de Groot, Max Planck Institute for Biophysical Chemistry, GERMANY

Received: August 11, 2016

Accepted: December 7, 2016

Published: December 27, 2016

Copyright: © 2016 Hwang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Data are available from <https://github.com/TBiophysG/VBDCMM/tree/master/data/HDNA%20KINETICS>.

Funding: This study was partly supported by National Research Foundation of Korea 704 NRF-2015R1D1A1A01060376 ("https://www.nrf.re.kr"). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Author Summary

We have developed a new algorithm to better decode single molecule data with dynamic disorder. Our new algorithm, which represents a substantial improvement over other methodologies, can detect the presence of dynamic disorder in each trajectory and quantify the kinetic characteristics of underlying energy landscape. As a model system, we applied our algorithm to the single molecule FRET time traces of H-DNA. While duplex-triplex transitions of H-DNA are conventionally interpreted in terms of two-state kinetics, slowly varying dynamic patterns corresponding to hidden internal states can also be identified from the individual time traces. Our algorithm reveals that at least 4 distinct internal states are required to correctly interpret the data.

Introduction

Recent technological advances in single molecule experiments on biomolecules have provided an unprecedented chance to investigate dynamics of proteins and nucleic acids at single molecule (SM) level, which has previously been elusive in conventional experiments [1–7]. Folding/unfolding pathways gleaned from individual SM trajectories indicate rugged folding landscapes inherent to biomolecules [4, 8, 9]. Long time trajectories from SM measurements, which now can be extended more than hundreds seconds, allow us to address how a rugged conformational landscape is sampled over time [7, 10, 11]. One of the striking findings from these measurements is that even under the same folding condition, conformational dynamics of individual molecules differ substantially from one another while still maintaining their biological functions. Cofactor-induced conformational transitions of *T. ribozymes* [12], Holliday junctions [13], TPP-riboswitch [14], and preQ₁-riboswitch [15] are the recent seminal examples that exhibit molecular heterogeneity at equilibrium. The variation in the velocities of individual RecBCD helicase motors along the dsDNA [16] is a good example of the molecular heterogeneity out of equilibrium, driven by ATP hydrolysis. Together with other reports [17–28], these could be merely a subset of more widespread, yet unrecognized cases that exhibit dynamical heterogeneity in SM time traces.

The chance of conformational frustration increases with the system size (N_{sys}). For a given N_{sys} , the time for conformational sampling (τ_{sample}) is expected to scale as $\tau_{\text{sample}} \sim e^{N_{\text{sys}}}$ [29]. Suppose that T_{obs} , which is in practice limited by several factors [30–32], is long enough to observe many (more than hundreds) transitions along a trace generated from SM measurement. Two distinct scenarios arise depending on the length of τ_{sample} relative to T_{obs} , (i) If the sampling time is shorter than T_{obs} ($\tau_{\text{sample}} \ll T_{\text{obs}}$), then the conformational space of biomolecule is fully sampled. In this case, the ergodicity of the system is ensured such that for any molecule α (or time trace α) the time average of an observable O_{α} $\langle O \rangle_T = \frac{1}{T_{\text{obs}}} \int_0^{T_{\text{obs}}} O_{\alpha}(\tau) d\tau$, is equivalent to the ensemble average of $O_{\alpha}(t)$ over all α 's ($1 \leq \alpha \leq N_{\text{ens}}$) at any moment t , $\frac{1}{N_{\text{ens}}} \sum_{\alpha=1}^{N_{\text{ens}}} O_{\alpha}(t) = \langle O \rangle_{\text{ens}}$, i.e., $\langle O \rangle_T = \langle O \rangle_{\text{ens}}$; thus thermodynamic properties of the system can be read out by analyzing a single time trace. (ii) In contrast, if $\tau_{\text{sample}} \gg T_{\text{obs}}$ is satisfied due to ruggedness of conformational space characterised with a number of deep local basins of attraction, then each time trace can sample only a local region of the conformational space. In this case, dynamic pattern from each time trace would look different, and a change in the dynamic pattern from one time interval to another would be observed only occasionally.

To be more precise about the second scenario ($\tau_{\text{sample}} \gg T_{\text{obs}}$), suppose that the average time scale for each local basin of attraction to be “sampled” by the conformational dynamics of molecule is τ_{conf} and that the time for the molecule to make transitions between different superbins of attraction is τ_{int} (Fig 1). In principle the relaxation rates and energy barrier heights of biomolecules span continuous spectra. So, the clear time scale separation may not always be warranted. However, to be able to grasp the presence of dynamic disorder, if any, in SM time traces straightforwardly, a separation between two distinct time scales is required such that $\tau_{\text{conf}} \ll \tau_{\text{int}}$ (or $\Delta G_{\text{conf}}^{\ddagger} \ll \Delta G_{\text{int}}^{\ddagger}$). If τ_{conf} and τ_{int} were comparable (or the spectra of relaxation rates were uniform and continuous), an algorithm we will propose here as well as others could hardly be of any help to conceive a concrete landscape model as the one illustrated in Fig 1. Therefore, here we consider τ_{conf} and τ_{int} as two disparate time scales as illustrated in Fig 1. τ_{conf} is the time at which the time average of an observable $\langle O \rangle_{\tau} = \frac{1}{\tau} \int_0^{\tau} O(t) dt$ reaches its steady state value when $\tau > \tau_{\text{conf}}$, corresponding to a time scale in which to fully sample the local basin of attraction. Alternatively, τ_{conf} is limited by a kinetic barrier with the greatest $\Delta G_{\text{conf}}^{\ddagger}$ within the local basin of attraction, so that $\tau_{\text{conf}} \geq e^{\Delta G_{\text{conf}}^{\ddagger}/k_B T}$. On the other hand,

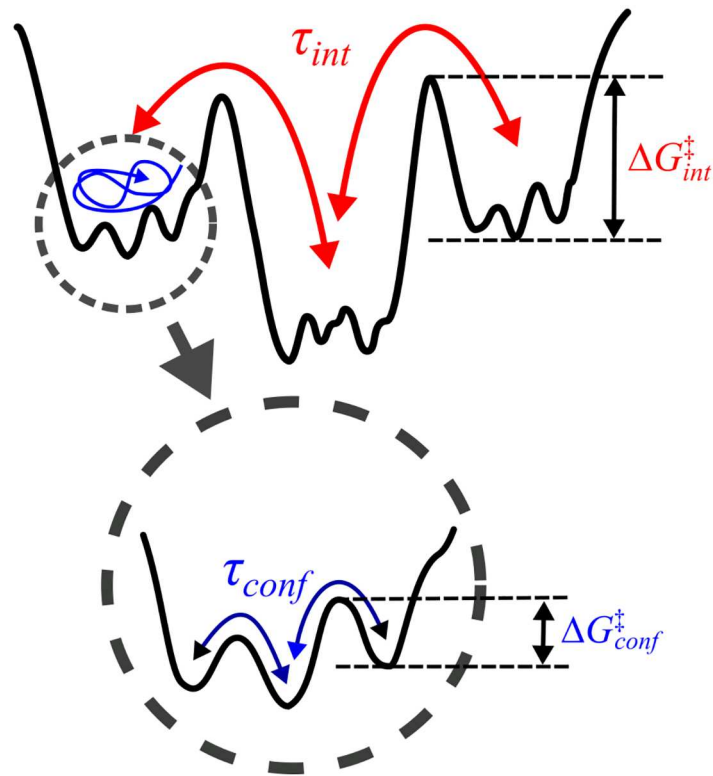


Fig 1. A rugged energy-landscape with hierarchical structure and an emergence of multiple time scales of transitions. τ_{int} is the transition time between different superbasins of attraction whereas τ_{conf} is the time scale of conformational dynamics of molecule *within* each basin. Due to large difference in kinetic barriers ($\Delta G_{int}^{\ddagger} \gg \Delta G_{conf}^{\ddagger}$), $\tau_{int} \gg \tau_{conf}$.

doi:10.1371/journal.pcbi.1005286.g001

τ_{int} is the transition time that is expected to scale with the height of kinetic barriers ($\Delta G_{int}^{\ddagger}$) between the two superbasins as $\tau_{int} \sim e^{\Delta G_{int}^{\ddagger}/k_B T}$. When measurements are conducted with a finite duration of observation time (T_{obs}), we can conceive two entirely different dynamic patterns depending on the relationship between τ_{conf} , τ_{int} , and T_{obs} :

- $\tau_{conf} \ll T_{obs} \ll \tau_{int}$: The interconversion time between distinct basins of attractions is far longer than the observation time. The dynamic patterns from individual trajectories that sample distinct basin of attraction are expected to differ from each other. Since $T_{obs} \ll \tau_{int}$, there is few chance to observe an exchange of dynamic pattern in a single time trace, which corresponds to a case with quenched disorder that each SM time trace looks entirely different. Such cases are reported in Holliday junction [13], *T. ribozyme* [12], and RecBCD [16].
- $\tau_{conf} \ll \tau_{int} \lesssim T_{obs}$: The interconversion time between basins of attraction is shorter than or comparable to the observation time. In this case, it is possible to observe a few rounds ($\sim T_{obs}/\tau_{int}$) of pattern exchanges in a single time trace. Such SM time traces are called to have a dynamic disorder [15, 28, 33–36].

While the most interesting and physically relevant question to ask about the heterogeneity in single molecule time traces is its molecular origin, detection and quantification of such heterogeneity should precede such question for a further analysis. For SM time traces with quenched disorder, it is relatively straightforward to analyze as one can use the criterion of

ergodicity and partition each time trace into its dynamic subensembles [13]. It is, however, more challenging to analyze time traces with dynamic disorder.

In the ion-channel community, ion currents across a single ion-channel measured with patch-clamp technique often demonstrate time series that switch between multiple dynamic patterns, and such a phenomenon is called ‘mode-switching’ [37] or ‘modal gating’ [38]. An algorithm (aggregated Markov model, AMM) developed by ion-channel community to analyze time series exhibiting dynamics disorder is in principle of use, but when applied to our synthetic data, we found that the algorithm tends to overpredict the transitions between hidden states (see S25 Fig and discussion related to it below). Thus, here we have developed a more reliable and systematic algorithm—Variational Bayes-Double Chain Markov Model (VB-DCMM)—which combined variational Bayes method with Double Chain Markov Model (DCMM) [39–43], to analyze SM time traces with dynamic disorder in which dynamic pattern of conformational transition changes at much longer time scale than apparent conformational fluctuations due to a slower transition of a *hidden* variable.

We first explain the algorithm for VB-DCMM, and next apply our VB-DCMM method to synthetic data as a blind test to show that our method can accurately identify the hidden internal states and determine the kinetic rate constants associated with the data. The results from our analysis using VB-DCMM are reliable as long as a clear separation in time scales exists between the apparent conformational transition (τ_{conf}) and the interconversion times (τ_{int}).

As a prototypical example of single molecule time traces with dynamic disorder, data from H-DNA [44, 45] that undergoes duplex-to-triplex conformational transitions (Fig 2A) are analyzed. A kinetic pattern of two-state like conformational transitions between duplex (low FRET ~ 0.1) and triplex form of H-DNA (high FRET ~ 0.9) observed in one time interval changes to another pattern in the next time interval (Fig 2B). DCMM models this peculiar dynamic pattern of H-DNA in Fig 2B by assuming a slowly varying dynamics of a hidden internal state. Fig 2C illustrates how the dynamic pattern of the original time trace of *observable state*, $o_n(t)$ (gray traces in Fig 2C), changes with the *internal state* $x(t)$ at a given time t . The dynamic pattern of $o_n(t)$, displaying multiple transitions, is slave to the slowly changing value of $x(t)$. DCMM implements this idea into an algorithm and allows us to extract the information of $x(t)$ from $o_n(t)$. Finally, we apply VB-DCMM to an ensemble of H-DNA time traces obtained from smFRET experiments and show that the dynamics of H-DNA at $[Na^+] = 100$ mM should be modeled using at least 4 large basins of attraction.

Algorithm

Here, we provide a general overview of the VB-DCMM algorithm, defining terms and parameters. More technical details of derivation and implementation of the algorithm are given in the Supplementary Information.

Modeling time series with dynamic disorder

Markov chain approach is ubiquitously used in modeling biological systems. For example, reversible conformational transitions of biomolecules probed by single molecule fluorescence resonance energy transfer (smFRET) or force spectroscopy are often modeled as a homogeneous Markov process in which the transition rates between experimentally discernible conformational states are uniquely decided. To decipher time series with dynamic disorder that change their dynamic pattern from one time interval to another we assume that there are hidden “internal states”, each of which determines the rate of conformational transitions. A signature of the transition between internal states, which gives rise to dynamic disorder in time series, are difficult to detect using the value of FRET efficiency or end-to-end distance alone

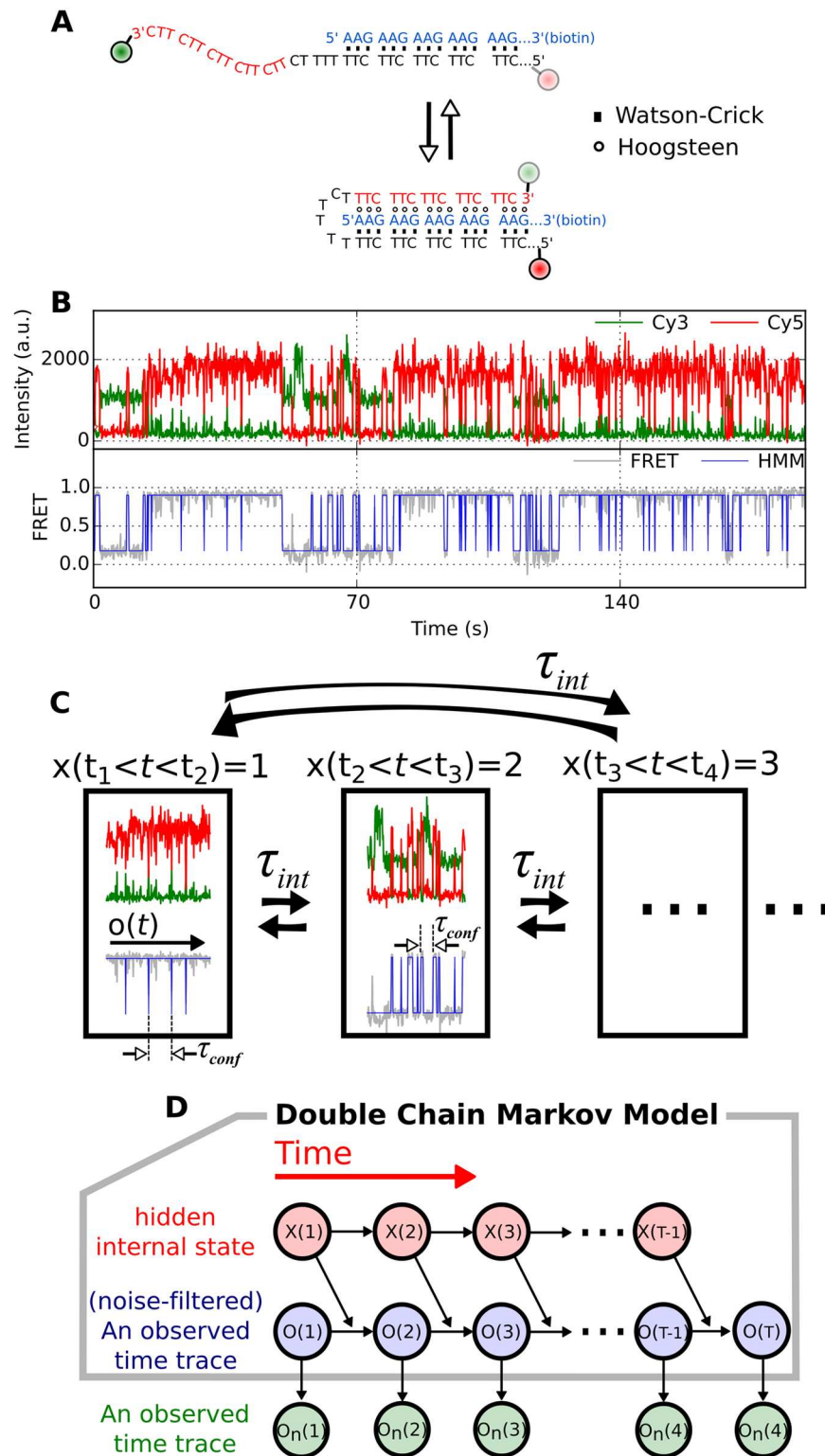


Fig 2. Duplex-triplex transitions of H-DNA with dynamic disorder. (A) Illustration of H-DNA dynamics. The sequences in blue and black form duplex via Watson-Crick base pairing; the sequences in red extended from 3'-end region of the black sequence can pair with the sequences in blue via Hoogsteen base pairs to form the triplex helix. (B) A time trace of H-DNA displaying dynamic disorder. (Top) The fluorescence signals from Cy3 (green) and Cy5 (red) dyes. (Bottom) FRET signal (gray) was calculated using the signals from Cy3

and Cy5. Blue line is the noise-filtered FRET signal obtained using HMM. The low-FRET (~ 0.1) and high-FRET state (~ 0.9) correspond to the duplex and triplex states, respectively. The dynamic pattern of the time trace changes occasionally from one time interval to another. For example, the transitions from low to high FRET state around 70 s are much slower compared with those around 140 s. (C) The model for H-DNA dynamics with dynamic disorder. Hierarchical transitions, (1) transitions within $x(t) = i$, and (2) interconversion between $x(t) = i$ and $x(t') = j (i \neq j)$, can be described using Double Chain Markov Model (DCMM). (D) Graphical representation of DCMM. $x(t)$, $o(t)$, and $o_n(t)$ represent internal state, noise-filtered observable (blue line in (B)), and the original observable at time t (gray line in (B)), respectively. The black arrows signify how each state is determined by others. For example, the state of observable at time t , $o(t)$ is determined by the previous observable state at time $t - 1$, $o(t - 1)$, and the state of the previous internal state, $x(t - 1)$.

doi:10.1371/journal.pcbi.1005286.g002

when the values observed along the time series are indiscernible even if the internal state is altered. By assuming that the transition between internal states is described by a homogeneous Markov process, and that transition between observable (in this study, FRET) follows non-homogeneous Markov process, whose transition rates are slaved to the internal state at each time, we model time trajectories made of these two layers of Markov chains. This algorithm corresponds to the Double Chain Markov Model (DCMM) [39–43] (Fig 2C and 2D).

DCMM is characterized by the following model parameters: (i) Transition matrix A for homogeneous Markov chain, which describes the transition probability between the K -distinct internal states along the time series ($\mathbf{x} = (x(1), x(2), \dots, x(t), \dots, x(T - 1))$). Here K is a total number of internal states in the model, and $x(t)$, specifying internal state at time t , takes one of the values between 1 and K . T is the total observation time. The internal state at time $t+1$ ($x(t + 1)$) is determined by the previous internal state at time t ($x(t)$), whose transition to $x(t + 1)$ is determined by a $K \times K$ Markov transition matrix A as $P(x(t + 1) = \mu) = \sum_{v=1}^K A_{\mu,v} P(x(t) = v)$ where $P(x(t) = v)$ denotes the probability of $x(t)$ being in the v -th internal state; (ii) K transition matrices $\mathbf{B}^{(\mu)}$ with $\mu \in \{1, 2, \dots, K\}$ for non-homogeneous Markov chain describes the transition probability between the observable states along the time series ($\mathbf{o} = (o(1), o(2), \dots, o(T))$). $o(t)$ specifies the state of the observable among N possible states $\{1, 2, \dots, N\}$ at time t . Transition from $o(t)$ to $o(t + 1)$ is determined by an $N \times N$ transition matrix $\mathbf{B}^{x(t)}(t)$, the matrix elements of which are slave to the value of $x(t) (= \mu \in \{1, 2, \dots, K\})$.

For example, if there are two ($K = 2$) internal states, and each internal state has three ($N = 3$) observables in a given time trace recorded with time resolution Δt , then two transition matrices for \mathbf{o} with $\mu = 1, 2$ can be considered (i.e., $\mathbf{B}^{(1)}$ and $\mathbf{B}^{(2)}$):

$$\mathbf{B}^{(\mu)} = \begin{pmatrix} k_{1 \rightarrow 1}^{(\mu)} \Delta t & k_{1 \rightarrow 2}^{(\mu)} \Delta t & k_{1 \rightarrow 3}^{(\mu)} \Delta t \\ k_{2 \rightarrow 1}^{(\mu)} \Delta t & k_{2 \rightarrow 2}^{(\mu)} \Delta t & k_{2 \rightarrow 3}^{(\mu)} \Delta t \\ k_{3 \rightarrow 1}^{(\mu)} \Delta t & k_{3 \rightarrow 2}^{(\mu)} \Delta t & k_{3 \rightarrow 3}^{(\mu)} \Delta t \end{pmatrix}.$$

Next, the transition matrix A for the interconversion between two internal states is:

$$\mathbf{A} = \begin{pmatrix} \gamma^{(1) \rightarrow (1)} \Delta t & \gamma^{(1) \rightarrow (2)} \Delta t \\ \gamma^{(2) \rightarrow (1)} \Delta t & \gamma^{(2) \rightarrow (2)} \Delta t \end{pmatrix}.$$

In the above matrices, the matrix elements must satisfy, $\sum_{j=1}^3 k_{i \rightarrow j}^{(\mu)} \Delta t = 1$ for each $i = 1, 2, 3$ in $\mathbf{B}^{(\mu)}$, and $\gamma^{(1) \rightarrow (1)} \Delta t + \gamma^{(1) \rightarrow (2)} \Delta t = \gamma^{(2) \rightarrow (1)} \Delta t + \gamma^{(2) \rightarrow (2)} \Delta t = 1$ in \mathbf{A} . More detailed descriptions about DCMM are available in the original papers [39–43] particularly in ref. [39] (see also SI). A similar but more general version of DCMM, which can accommodate inputs variables as well as multiple number of internal state sequences, has been suggested by extending the factorial hidden Markov model [46, 47].

Determining the number of internal states

DCMM can estimate the transition matrices \mathbf{A} and $\mathbf{B}^{(\mu)}$ quantitatively, and hence determine the most probable sequence of internal state and associated kinetic rates, $\{k_{a \rightarrow b}^{(\mu)}\}$ and $\{\gamma^{(\mu) \rightarrow (v)}\}$. However, the likelihood (the probability of observing data for given model parameters), maximized by DCMM, $P(\mathbf{o}|\boldsymbol{\pi}, \mathbf{A}, \mathbf{B})$ where $\boldsymbol{\pi} \equiv (\pi_1, \pi_2, \dots, \pi_K)$ with $\pi_\mu = P(x(1) = \mu|\mathbf{o}, \mathbf{A}, \mathbf{B})$, is prone to increase when more number of parameters are used in the model. DCMM can select the best set of parameters for a given model, but not suited to select the best model (i.e., cannot determine the optimal number of internal states K for a given time trace). To overcome this limitation, often used is the maximum evidence method, where the *evidence* ($P(\mathbf{o}|K)$, also called marginal likelihood) is defined as the conditional probability of observing data (\mathbf{o}) for a given model (K), so that

$$P(\mathbf{o}|K) = \int P(\mathbf{o}|\boldsymbol{\lambda})P(\boldsymbol{\lambda}|K)d\boldsymbol{\lambda} \tag{1}$$

where $\boldsymbol{\lambda} \equiv (\boldsymbol{\pi}, \mathbf{A}, \mathbf{B})$ represents the parameter space. In this method, the penalty against model complexity is naturally incorporated during the calculation, allowing to select the best model (see **SI**). By calculating the evidence for each different model (different K , the number of internal states in data), one can select the best model with an optimal number of internal states that maximizes the evidence. The calculation of the evidence, however, involves a massive computational cost to explore the entire parameter space for a given model.

Variational Bayes double chain Markov model

To alleviate the computational cost in employing the maximum evidence method in [Eq 1](#), we employ the Variational Bayes [\[48\]](#), a method that effectively uses a mean-field approximation. The method has previously been used to determine the number of observable states (FRET states) from smFRET data [\[49–51\]](#), the number of diffusive states from single molecule tracking data [\[52\]](#), and the number of DNA-protein conformations from tethered particle motion data [\[53\]](#). It has also been used inside the empirical Bayes method which can analyze several smFRET time series simultaneously [\[54, 55\]](#). In our study, the variational Bayes method combined with DCMM (VB-DCMM) was used to analyze single molecule time traces with dynamic disorder. The analytical expression of the lower bound of the *evidence* (F), offered by VB-DCMM, makes clear where the model penalty comes from, thus providing guidelines to choose the prior parameters to incorporate a prior knowledge of data (see **SI**). Once prior parameters are selected, VB-DCMM iteratively increases the lower bound of $\log(\text{evidence}) (= \log P(\mathbf{o}|K))$ by identifying a better approximation to the true probability distribution.

$$\begin{aligned} \log P(\mathbf{o}|K) &= \int q(\mathbf{Z})\log P(\mathbf{o}|K)d\mathbf{Z} \\ &= F[q] + D_{KL}(q||p) \geq F[q^*]. \end{aligned} \tag{2}$$

where $q(\mathbf{Z})$ is an arbitrary probability distribution of a set of variables, $\mathbf{Z}(\equiv (\mathbf{x}, \boldsymbol{\lambda}))$ consisting of parameters and hidden variables of model,

$$F[q] \equiv \int q(\mathbf{Z})\log (P(\mathbf{o}, \mathbf{Z}|K)/q(\mathbf{Z}))$$

and

$$D_{KL}(q||p) \equiv \int q(\mathbf{Z})\log (q(\mathbf{Z})/P(\mathbf{Z}|\mathbf{o}, K)) \geq 0,$$

where $D_{KL}(q||p)$ is the Kullback-Leibler divergence of $q(\mathbf{Z})$ from $P(\mathbf{Z}|\mathbf{o}, K)$, which we want to minimize. Once the solution from the algorithm converges, the approximate value of $\log P(\mathbf{o}|K^*) (\simeq F[q^*])$ and the (locally) best model parameters (a set of the best kinetic rates), $\boldsymbol{\pi}^*$, A^* and B^* , which determines all the rate constants to describe the given time traces ($\{k_{a \rightarrow b}^{(\mu)}\}$ and $\{\gamma^{(\mu) \rightarrow (\nu)}\}$), are acquired from an approximated probability distribution (See SI for the mathematical details). The performance of VB-DCMM is quite robust over a wide variation of prior parameters (S21 and S22 Figs).

Implementation of the algorithm

The observable sequence \mathbf{o} is obtained by filtering the noise in the experimental data (\mathbf{o}_n) using Hidden Markov Model (HMM) following a similar procedure as the previous studies [49, 56] using a custom code written based on the code from Sagemath software [57]. Next, the \mathbf{o} is analyzed using VB-DCMM to select the best model and to estimate the best model parameters. The optimal sequence of internal states \mathbf{x} is determined by using Viterbi algorithm [39]. All the implementations and data analysis are done by using our custom code. VB-DCMM is freely available at “<https://github.com/TBiophysG/VBDCMM>”

Results and Discussion

Validation of VB-DCMM

To first validate the efficacy of VB-DCMM in identifying internal states in a given SM time trace, we applied VB-DCMM algorithm on synthetic data that mimic a SM time trajectory with dynamic disorder (see Methods). To generate a synthetic SM time trajectory, we first produce a time trajectory specifying the value of internal state from $t = 1$ to $t = T - 1$. The time trajectory of internal state is represented with a symbol $\mathbf{x} \equiv (x(1), x(2), \dots, x(t), \dots, x(T - 1))$. When the total number of distinct internal states in the model is K , one of the values in $\{1, 2, \dots, K\}$ is assigned to $x(t)$. Thus, for $K = 2$ a typical time trajectory of internal state \mathbf{x} looks like $(1, 1, 1, \dots, 1, 1, 2, 2, 2, \dots, 2, 2, 1, 1, \dots, 1, 1), (1, 1, 1, \dots, 1, 2, 2, \dots, 2, 2), (2, 2, \dots, 2, 1, 1, \dots, 1, 1, 2, 2, \dots, 2, 2, 2),$ etc. The time trajectory given in Fig 3A(i) is an example generated with $K = 2$ and $T = 8801$. Next, similar to the structure of \mathbf{x} , the time trajectory of noiseless observables is represented using $\mathbf{o} \equiv (o(1), o(2), \dots, o(t), \dots, o(T))$. In Fig 3A(ii), a trajectory of \mathbf{o} is shown, also demonstrating the influence of \mathbf{x} on \mathbf{o} . Finally, Gaussian noise was added on \mathbf{o} and the range of signal was adjusted to produce the final trajectory $\mathbf{o}_n \equiv (o_n(1), o_n(2), \dots, o_n(T))$ which now resembles a time trajectory of SM FRET signal (Fig 3A(iii)).

Deciphering the information of internal states from an observed time trace involves solving an inverse problem, i.e., decoding \mathbf{o}_n to obtain \mathbf{x} . To decode the trace of internal states from the synthetic data, we follow a 3-step procedure: (1) Filter the noise from \mathbf{o}_n to obtain \mathbf{o} using Hidden Markov Model (HMM) [56] (Fig 3A(iv), blue line); (2) Analyze \mathbf{o} by applying VB-DCMM algorithm with different models $1, 2, \dots, K$ (again, K is the total number of internal states assumed in each model); (3) To select the best model we calculated the conditional probability of observing data for a given model parameter K , $P(\mathbf{o}|K)$, which is often called *evidence* or *marginal likelihood* in machine learning community (Eq 2) [48]. Calculation of $P(\mathbf{o}|K)$ is conducted using the Variational Bayes (VB) method, which gives the lower bound of $\log P(\mathbf{o}|K)$ denoted by $F(K)$. Details of the evidence function $F(K)$ and approximation procedure are provided in the Supplementary Information (SI). Finally, we select the best model K^* which maximizes $F(K)$, i.e., $K^* = \arg \max F(K)$.

To be specific, in order to identify the best model parameter K for the time trace $\mathbf{o}(t)$ given in Fig 3A(iv), we varied K from 1 to 3. The most probable trace of internal states, $\mathbf{x}_{(K)}^{\text{model}}$, was

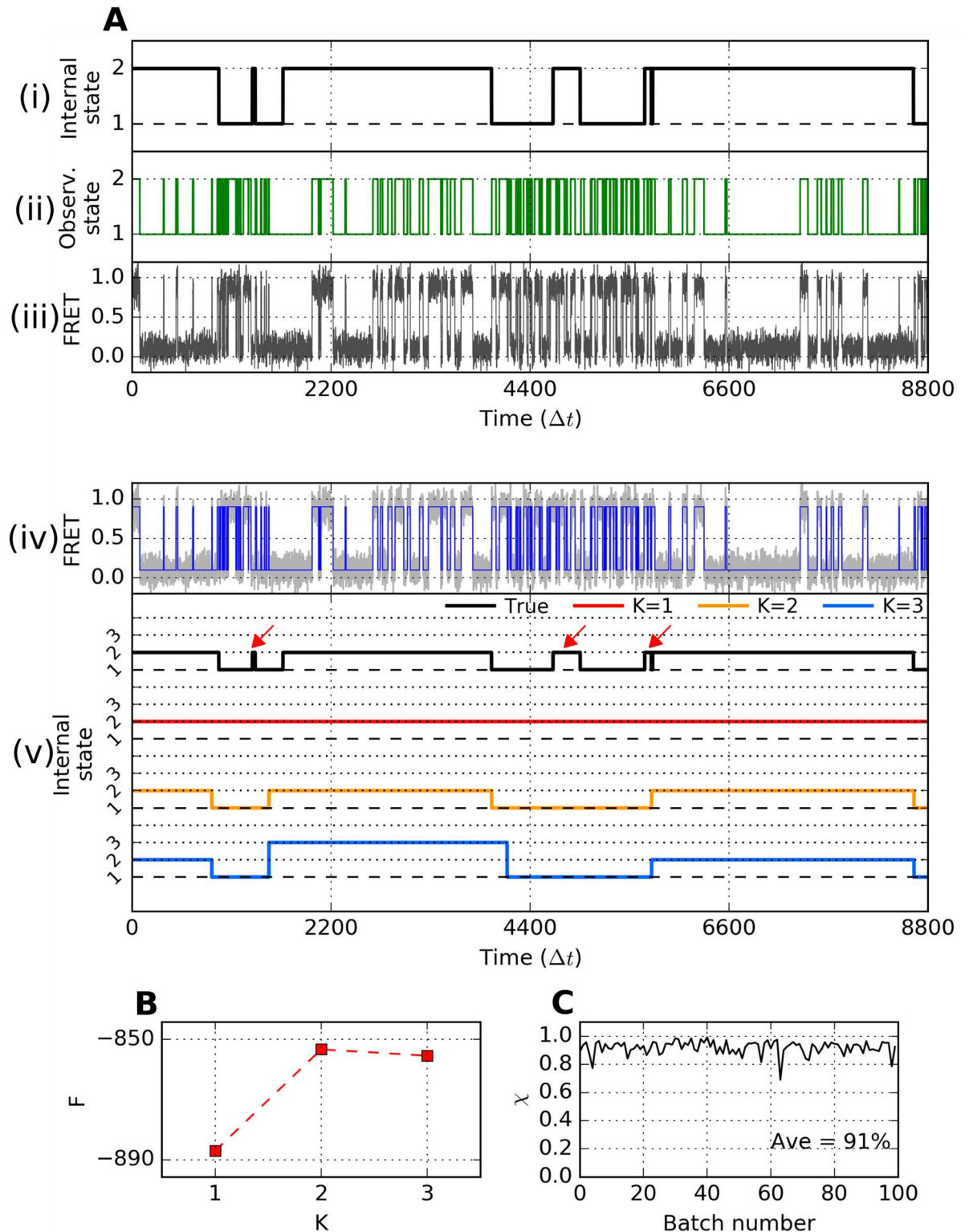


Fig 3. Validation of VB-DCMM on synthetic data. (A) (i) A time trace of internal state generated with $\gamma^{(1) \rightarrow (2)} \Delta t = \gamma^{(2) \rightarrow (1)} \Delta t = 0.001$. (ii) An observable time trace generated based on the trace of internal state in (i) by using internal state-dependent parameters $k_{L \rightarrow H}^{(1)} \Delta t = k_{H \rightarrow L}^{(1)} \Delta t = 0.05$, $k_{L \rightarrow H}^{(2)} \Delta t = 0.00625$, $k_{H \rightarrow L}^{(2)} \Delta t = 0.025$. (iii) An synthetic FRET data with Gaussian noise overlaid on the trace in (ii). (iv) Noised filtered FRET state by HMM (blue line). (v) Traces of internal state with different K , estimated using VB-DCMM on the noise-filtered FRET trace from (iv) (black line is the true internal state trace while red, orange, and blue are internal state estimated

from the model with $K = 1, 2,$ and $3,$ respectively. The indices of internal state were determined by comparing $\mathbf{B}^{(\mu)}$ estimated for each internal state with $\mathbf{B}^{(\mu),\text{true}}$ which is used to generate the synthetic data). (B) Estimated lower bound of the evidence function $F(K)$ of DCMM models with $K = 1, 2,$ and $3.$ (C) Accuracy of detecting internal states. The overlap function χ calculated for 100 synthetic FRET traces generated under the identical condition used for generating the trace of internal state shown in (A).

doi:10.1371/journal.pcbi.1005286.g003

calculated for each model with $K = 1$ (red), $K = 2$ (orange), $K = 3$ (blue) (see Fig 3A(v)). The evidence $F(K)$ calculated using VB method was maximized at $K = K^* = 2,$ and the resulting time trace of the internal states, $\mathbf{x}_{(K^*=2)}^{\text{model}},$ most closely recovers the trajectory of \mathbf{x} (black trace in Fig 3A(v)) except at the time interval where the transitions of $x(t)$ between 1 and 2 occur only transiently or at the boundaries of transitions (red arrows on Fig 3A(v)). This result shows that VB-DCMM can avoid the over-fitting problem that other methods based on maximum likelihood are often fraught with [48].

Conditions required for an accurate recovery of internal states

VB-DCMM detects a signature of change in internal state (x) from a given observable time trace (\mathbf{o}) by evaluating the statistical difference in transition rates. Thus, in the absence of an enough number of transitions in the trace $\mathbf{o},$ the algorithm becomes less reliable. For example, we obtained $F(K = 2) \approx F(K = 3)$ although $F(2) \gg F(3)$ is more desirable (Fig 3B. See another example in S1 Fig). This is due to the lack of statistics in transition events in this particular test trace given in Fig 3A. For example, when only a part of the time trace is selected and analyzed using HMM, the estimated rates of transition from high (H) to low (L) FRET value are $k_{H \rightarrow L}^{\text{est}} \Delta t = 0.016$ in $1500 \lesssim t \lesssim 4000,$ and $k_{H \rightarrow L}^{\text{est}} \Delta t = 0.026$ in $5700 \lesssim t \lesssim 8700.$ Thus, in ($K = 3$)-model the two time intervals, originally generated by using the same kinetic parameter ($k_{H \rightarrow L}^{(2)} \Delta t = 0.025$), are determined to be distinct from each other (blue trace in Fig 3A(v)). By contrast, in ($K = 2$)-model, $k_{H \rightarrow L}^{\text{est}} \Delta t = 0.020$ was estimated over these two time intervals. This type of statistical error is unavoidable for a small $T_{\text{obs}}.$ A more systematic evaluation on the accuracy of the algorithm as a function of T_{obs} and transition rate between distinct internal states will be discussed in the next section.

To assess the accuracy of the best model $\mathbf{x}_{(K^*)}^{\text{model}}$ predicted by VB-DCMM against the solution $\mathbf{x},$ the following overlap function can be used.

$$\chi = \frac{1}{T-1} \sum_{t=1}^{T-1} \delta_{x(t), x_{(K^*)}^{\text{model}}(t)} \quad (3)$$

where δ_{ij} is the Kronecker delta and $T = T_{\text{obs}}/\Delta t$ is the total number of data in the traces (Δt denotes the temporal resolution of the data). For 100 synthetic time traces, generated under the identical parameters used for producing the time trace in Fig 3A, we found that $\chi \approx 0.9$ on average (Fig 3C). Note, however, that $x(t),$ only available for the case of “synthetic data”. Thus, to assess the accuracy of our method against a real time trace from SM experiments, we devised other metrics.

For a given time trace with dynamic disorder, our algorithm quantifies the kinetic features of the time trace in terms of the transition rate between the observable states a and b within the μ -th internal state $k_{a \rightarrow b}^{(\mu)}$ and the transition rate from the μ -th internal state to ν -th internal state $\gamma^{(\mu) \rightarrow (\nu)}$ ($1 \leq \mu, \nu \leq K, 1 \leq a, b \leq N.$ Here, μ is the index for internal state whereas a and b are indices for observable (In FRET displaying low/high two state transitions, these states correspond to the low and high FRET values). K is the total number of hidden internal states, and N denotes the total number of observables). To be able to extract the information of multiple

internal states reliably from a time trace using VB-DCMM, two general conditions are required for the time trace being analyzed.

1. A large time scale separation should be present in the kinetics within each internal state, i.e., $k_{a \rightarrow b}^{(\mu)}$ and $k_{a \rightarrow b}^{(v)}$ ($\mu \neq v$) should be disparate.
2. There should be a clear time scale separation between intra-basins and inter-basin transitions (i.e., τ_{conf} and τ_{int}). More precisely, the intra-basin transition probability $k_{a \rightarrow b}^{(\mu)} \Delta t$ should be much *greater* than the transition probability from the μ -th to any other internal state $\sum_{v \neq \mu} \gamma^{(\mu) \rightarrow (v)} \Delta t (= 1 - \gamma^{(\mu) \rightarrow (\mu)} \Delta t)$.

To substantiate the above-mentioned conditions 1 and 2, we define two metrics D_{conf} and D_{int} which compute the average Hamming-like distances between the distinct rate constants extracted from a given time trace using VB-DCMM analysis:

$$D_{conf} = \frac{2}{K(K-1)} \sum_{\substack{\mu, v=1 \\ \mu > v}}^K \frac{1}{N(N-1)} \sum_{\substack{a, b=1 \\ a \neq b}}^N \left| \log_2 \frac{k_{a \rightarrow b}^{(\mu)}}{k_{a \rightarrow b}^{(v)}} \right| \quad (4)$$

and

$$D_{int} = \frac{1}{K} \sum_{\mu=1}^K \frac{1}{N(N-1)} \sum_{\substack{a, b=1 \\ a \neq b}}^N \left| \log_2 \frac{k_{a \rightarrow b}^{(\mu)}}{\sum_{v \neq \mu} \gamma^{(\mu) \rightarrow (v)}} \right|. \quad (5)$$

D_{conf} measures the dissimilarity between distinct internal states in terms of the intra-basin transition rates. Two distinct internal states (μ, v ($\mu \neq v$)) can be better discerned if the intra-basin transition rate of one internal state (say, $k_{a \rightarrow b}^{(\mu)}$) differs greatly from that of other internal state ($k_{a \rightarrow b}^{(v)}$), so that $|\log_2(k_{a \rightarrow b}^{(\mu)}/k_{a \rightarrow b}^{(v)})|$ is maximized. D_{int} measures the average number of intra-basin transitions in each internal state using the ratio between the transition probabilities, $k_{a \rightarrow b}^{(\mu)} \Delta t$ and $\sum_{v \neq \mu} \gamma^{(\mu) \rightarrow (v)} \Delta t (= 1 - \gamma^{(\mu) \rightarrow (\mu)} \Delta t)$. A greater D_{int} ensures a large time scale separation in dynamics between intra-basin and inter-basin transitions, which improves the reliability of our method to decode the internal state from a given time trace. In general, D_{int} or D_{conf} shows a good correlation with $\langle \chi \rangle$ (see below); thus, one can use (D_{int}, D_{conf}) to assess the accuracy of predicted internal states. Note that the metrics D_{int} and D_{conf} can be estimated for real data, while $\langle \chi \rangle$ can be calculated only against the synthetic data. Since there is a good correlation between (D_{int}, D_{conf}) and χ , one can evaluate (D_{int}, D_{conf}), alternative to χ , to assess the reliability of a predicted result of $x_{(K^*)}^{model}(t)$.

To be more concrete, we applied VB-DCMM algorithm to analyze synthetic data generated with $N = 2$ (transitioning between high and low FRET values) and $K = 2$ (two internal states; $\mu = 1$ and 2) under various scenarios.

- We fixed the transition rates in the state $\mu = 1$ as $k_{L \rightarrow H}^{(1)} \Delta t = k_{H \rightarrow L}^{(1)} \Delta t = 0.05$, and varied the rates associated with the state $\mu = 2$ over the range of $0.125 \leq k_{L \rightarrow H}^{(2)}/k_{L \rightarrow H}^{(1)}, k_{H \rightarrow L}^{(2)}/k_{H \rightarrow L}^{(1)} \leq 8$ (Fig 4A, left). For the interconversion probability between the two internal states we set $\gamma^{(1) \rightarrow (2)} \Delta t = \gamma^{(2) \rightarrow (1)} \Delta t = 0.001$. The accuracy of the model prediction ($\langle \chi \rangle$, Eq (3)) is on average greater than 0.9 as long as the transition rates $k_{L \rightarrow H}^{(\mu)}$ and $k_{L \rightarrow H}^{(v)}$ ($\mu \neq v$) differ more than the factor of 4. Note that in Fig 4A (left), the value of $\langle \chi \rangle$ is greater for $k_{L \rightarrow H}^{(2)}/k_{L \rightarrow H}^{(1)}$, $k_{H \rightarrow L}^{(2)}/k_{H \rightarrow L}^{(1)} \gg 1$ than for $k_{L \rightarrow H}^{(2)}/k_{L \rightarrow H}^{(1)}, k_{H \rightarrow L}^{(2)}/k_{H \rightarrow L}^{(1)} \ll 1$; this is because a statistically sufficient number of transitions make the detection of internal states more reliable. In contrast,

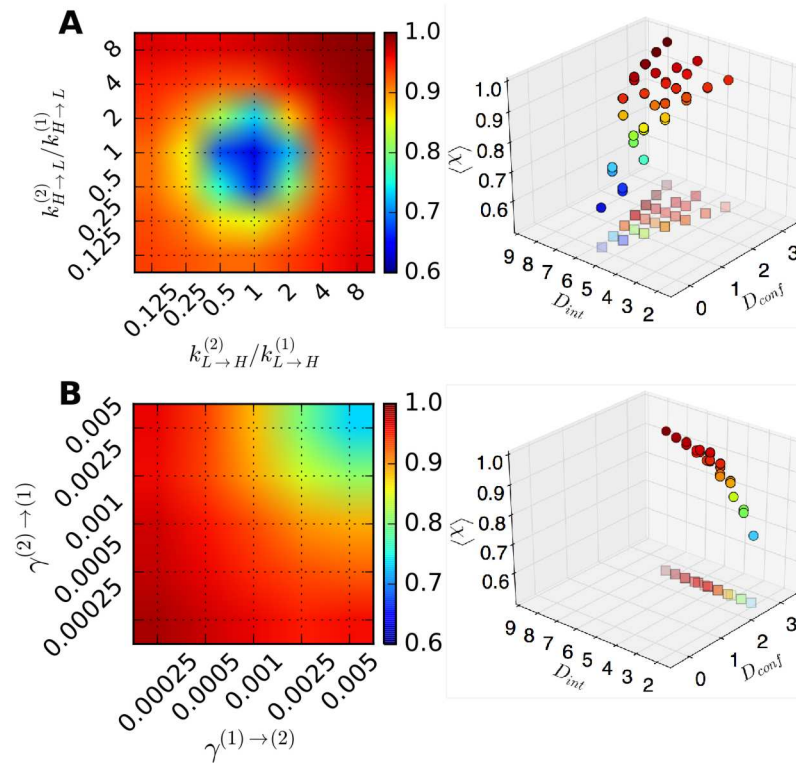


Fig 4. Accuracy of VB-DCMM in detecting internal states under various conditions of $k_{L \rightarrow H}^{(\mu)}$ and $\gamma^{(1) \rightarrow (2)}$ with $T_{obs}/\Delta t = 8800$. (A) The color bar denotes the accuracy of analysis in terms of $\langle \chi \rangle$ under varying $k_{L \rightarrow H}^{(2)}, k_{H \rightarrow L}^{(2)}$ with $K = 2, k_{L \rightarrow H}^{(1)}\Delta t = 0.05, k_{H \rightarrow L}^{(1)}\Delta t = 0.05$, and $\gamma^{(1) \rightarrow (2)}\Delta t = \gamma^{(2) \rightarrow (1)}\Delta t = 0.001$. (B) $\langle \chi \rangle$ under varying $\gamma^{(1) \rightarrow (2)}$ and $\gamma^{(2) \rightarrow (1)}$ with $K = 2, k_{L \rightarrow H}^{(1)}\Delta t = k_{H \rightarrow L}^{(1)}\Delta t = 0.05, k_{L \rightarrow H}^{(2)}\Delta t = 0.00625, k_{H \rightarrow L}^{(2)}\Delta t = 0.0125$. $\langle \chi \rangle$ was calculated by averaging over the results from analysis of 100 traces in each condition. The panels on the right show the relation between the value of $\langle \chi \rangle$ and pairs of D_{int} and D_{conf} values which are evaluated at varying kinetic parameters. Results from the analysis over the data with the same parameters but different length of time trace (or different number of data points $T_{obs}/\Delta t = 2200, 4400$) are provided in S4 Fig.

doi:10.1371/journal.pcbi.1005286.g004

when $k_{L \rightarrow H}^{(2)}/k_{L \rightarrow H}^{(1)}, k_{H \rightarrow L}^{(2)}/k_{H \rightarrow L}^{(1)} \simeq 1$, i.e. when the kinetics inside the two internal states are essentially identical, it is difficult to discern the two internal states. In this case, $K = 1$ instead of $K = 2$ is effectively the correct number of internal states. Indeed, when $K = 1$ is assumed (i.e., assuming true internal state $x(t) = 1$ for all t in Eq (3)), the re-calculated $\langle \chi \rangle$ is close to 1 (see S2 Fig).

- To explore the effect of interconversion between distinct internal states on the performance of algorithm, we generated synthetic data with $k_{L \rightarrow H}^{(1)}\Delta t = k_{H \rightarrow L}^{(1)}\Delta t = 0.05$, $k_{L \rightarrow H}^{(2)}/k_{L \rightarrow H}^{(1)} = 0.125$, and $k_{H \rightarrow L}^{(2)}/k_{H \rightarrow L}^{(1)} = 0.25$ by, this time, varying $\gamma^{(1) \rightarrow (2)}\Delta t$ and $\gamma^{(2) \rightarrow (1)}\Delta t = 0.00025 \sim 0.005$ (Fig 4B, left). The results clearly show that the case with smaller $\gamma^{(\mu) \rightarrow (v)}$ results in a higher $\langle \chi \rangle$, which is expected because each internal state can have more number of transitions in the traces σ when the interconversion is slower (Fig 4B). Re-plotting $\langle \chi \rangle$ as a function of D_{conf} and D_{int} reveals clear dependence of the accuracy on D_{int} (Fig 4B, right). Similar trends are observed for other conditions of $k_{L \rightarrow H}^{(2)}/k_{L \rightarrow H}^{(1)}$ and $k_{H \rightarrow L}^{(2)}/k_{H \rightarrow L}^{(1)}$ (S3 Fig).
- Analyses on synthetic data generated using the same input parameters with those in Fig 4, but with a different number of data points in each trace, $T_{obs}/\Delta t = 4400$, and 2200 (S4 Fig)

show a similar trend as observed in Fig 4 with $T_{obs}/\Delta t = 8800$ but with slightly smaller $\langle \chi \rangle$ values.

- Extension of VB-DCMM algorithm to a more complicated case for $K > 2$ (S5 and S14 Figs) or $N > 2$ (S6 and S14 Figs) is straightforward. Application of VB-DCMM to a trajectory in which each internal state trajectory has different N is also straightforward (S7 Fig). In the latter case, the data is analyzed by assuming that all internal states have the same number of possible observables, N ; but the analysis would indicate that transition associated with a small transition rate is essentially disallowed. In all situations considered for various K and N , VB-DCMM can be used for the reliable recovery of the sequence of true internal states.
- Analyses of synthetic traces show that the accuracy of the algorithm improves with both D_{conf} and D_{int} (Fig 4 right panels and Fig 5). Thus, these two metrics allow one to judge the reliability of the information on internal states extracted from a given time trace. Alternatively, a single parameter $D_{tot}(= D_{conf} + \alpha D_{int})$ with an empirically acquired coefficient $\alpha \approx 0.8$ can be used to judge the reliability of the extracted information. Note that $\langle \chi \rangle$ remains similar as long as D_{tot} remains constant (Fig 5B). Hence, when $\langle \chi \rangle$ is plotted against D_{tot} , all synthetic data generated using different parameters approximately collapse onto a single universal curve (S8 Fig).
- There are multiple ways of assessing the efficacy of VB-DCMM in decoding the internal states. In addition to $\langle \chi \rangle$, D_{conf} , D_{int} , and D_{tot} as the possible measures for the assessment, one can also use the statistical property that the dwell times of homogeneous Markov process satisfies $\sqrt{\langle \tau^2 \rangle - \langle \tau \rangle^2} / \langle \tau \rangle \sim 1$ (see SI for details).

Application of VB-DCMM on H-DNA data

Now, to analyze the duplex-triplex transitions of H-DNA (Fig 6), we obtain \mathbf{o} by filtering the noise from FRET signal (Fig 6(ii), blue line) and apply the VB-DCMM algorithm to decode the hidden internal state in the signals. Fig 6(iii) shows time series of internal state, $\mathbf{x}_{(K)}^{model}$, calculated from the VB-DCMM by varying K from 1 to 5. It is of note that the number of actually observed internal states in the $\mathbf{x}_{(K)}^{model}$ for a given input parameter K does not change after some $K_{obs}(\leq K)$ ($K_{obs} = 2$ (Fig 6A), 2 (Fig 6B), 2 (Fig 6C), and 1 (Fig 6D)). (See also other time traces of synthetic data and H-DNA analyzed in SI: S1A Fig ($K_{obs} = 1$), S9A Fig ($K_{obs} = 3$), S9B Fig ($K_{obs} = 2$), S9C Fig ($K_{obs} = 3$), S9D Fig ($K_{obs} = 3$), S10A Fig ($K_{obs} = 4$), S10B Fig ($K_{obs} = 2$), S10C Fig ($K_{obs} = 2$), S10D Fig ($K_{obs} = 3$), and S12 Fig ($K_{obs} = 3$)). A similar behavior is also observed when analyzing data using the variational Bayes Gaussian mixture model [48].

To account for the contribution due to degeneracy in labeling the internal states, $\log K!$ term is conventionally considered in formulating the evidence function $F(K)$ (See SI for the details); however, in our problems, the actual number of degeneracy in labeling internal states should be ${}_K C_{K_{obs}} \times K_{obs}!$ instead of $K!$. Therefore, we replace the $\log K!$ term in $F(K)$ with $\log [K!/(K - K_{obs})!]$, and considered a modified evidence function, $G(K)$, to identify an optimal K for a given time trace:

$$G(K) \equiv F(K) - \log (K - K_{obs})! \tag{6}$$

$G(K)$ shows a clear peak, allowing us to identify the optimal $K(= K^*)$ with ease (blue circles on the right side of Figs 6, S9 and S10). Use of $G(K)$ instead of $F(K)$ in analyzing synthetic data does not alter K^* (S12 and S14B Figs).

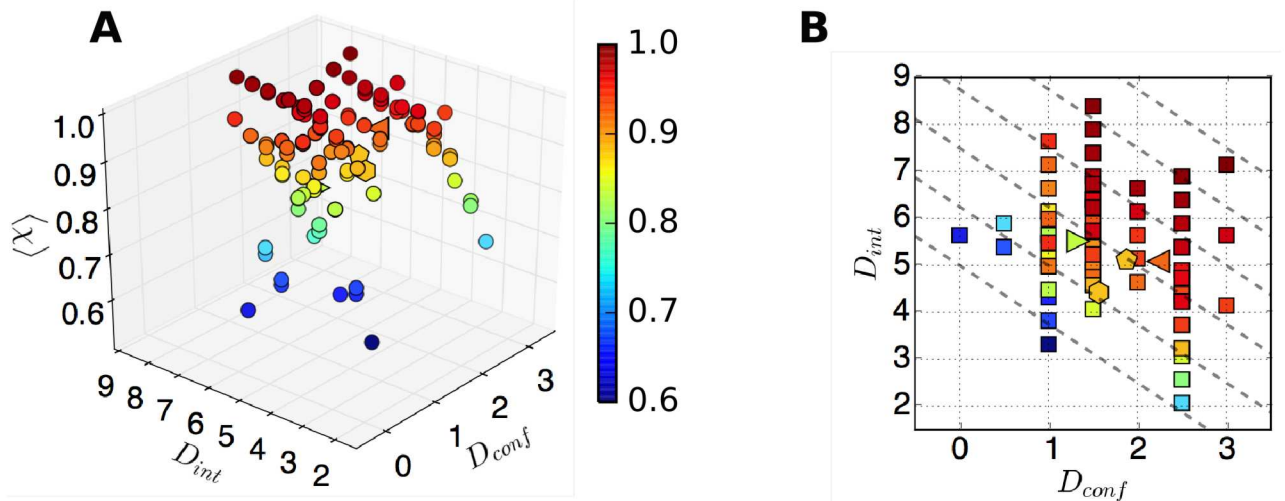


Fig 5. Average accuracy of internal state detection as a function of D_{conf} , and D_{int} . To construct this diagram, we employed various synthetic data in Fig 4 (circle, two internal states ($K=2$), two FRET states ($N=2$)), S5 Fig (left triangle, $K=3$, $N=2$), and S6 Fig (hexagon, $K=2$, $N=3$). The right triangle symbol denotes the result from the similar analysis shown in S5 Fig with $K=3$ but with smaller relative differences in the transition rates, K 's. Pentagon represents the result obtained with $K=4$ and $N=2$. (A) Color code denotes the accuracy of internal states predictions in terms of $\langle \chi \rangle$, averaged over 100 traces for each condition. (B) The dashed lines corresponding to $\Delta = D_{conf} + 0.8D_{int} = 4, 5, \dots, 9$ are overlaid on the 2-D scatter plot of $\langle \chi \rangle(D_{conf}, D_{int})$ calculated in (A).

doi:10.1371/journal.pcbi.1005286.g005

Among the time traces of H-DNA, traces with more than 3 interconversions between distinct internal states, which enables us to estimate $\gamma^{(\mu \rightarrow \nu)}$, are rare, especially when $[Na^+] = 100$ mM; thus it is not feasible to get a statistically meaningful scatter plot of (D_{conf}, D_{int}) (see S23D, S23E and S23F Fig); however, for those displayed in S23D, S23E and S23F Fig, $\langle D_{tot} \rangle \approx 7$ suggests that $\chi \gtrsim 0.9$ (from Fig 5). Therefore, at least the intra-basin rate constants extracted from H-DNA data using VB-DCMM are reliable.

Time traces that have τ_{int} comparable to experimental observation time ($\tau_{int} \approx \mathcal{T}_{obs}$) would exhibit on average no or only a single transition event between distinct internal states. Indeed, we find that only a subset of total number of internal states is sampled by individual time traces due to the limited observation time. For instance, at $[NaCl] = 100$ mM, our analysis identified $K^* \leq 2$ in 265 out of 269 traces, and that only 4 time traces display $K^* > 2$ (S11 Fig). Therefore, in order to identify the internal states present in the transition dynamics of H-DNA, clustering analysis is required against the whole ensemble of time trajectories. We provide the procedure of clustering analysis and results in details in the following section.

Clustering H-DNA data

VB-DCMM algorithm allows us to decompose individual H-DNA time traces with dynamic disorder into multiple “components”, each of which should satisfies the property of *homogeneous Markov chain*. In order to understand the structure of conformational space of H-DNA, the ensemble of components acquired from the VB-DCMM analysis should be clustered into the same kind. To this end, we produce scatter plots of $(k_{L \rightarrow H}, k_{H \rightarrow L})$, representing the kinetic property of the ensemble of time traces, using the transition rates estimated for individual time traces. The scatter plots of $(k_{L \rightarrow H}, k_{H \rightarrow L})$ were calculated for the ensemble of H-DNA time traces (i) before (Fig 7A, left) and (ii) after decomposing the individual heterogeneous time traces retaining multiple components into the homogeneous ones (Fig 7A, right). The scatter plot of $(k_{L \rightarrow H}, k_{H \rightarrow L})$ after the decomposition has a greater dispersion, which is expected since

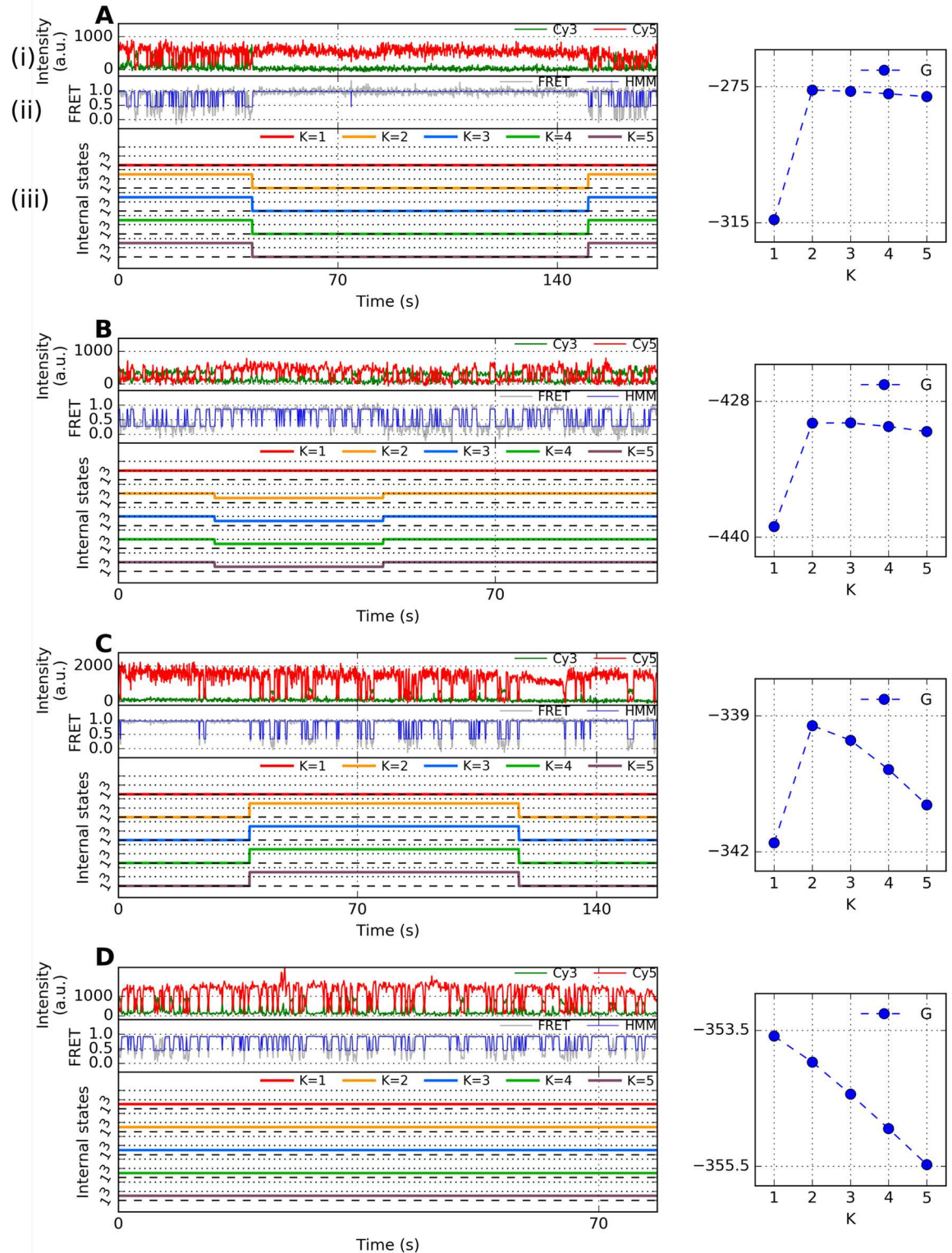


Fig 6. Representative time traces of H-DNA at $[Na^+] = 100$ mM and their analysis. (A) (i) Fluorescence signal and (ii) their FRET state. (iii) Internal states estimated for $K = 1, 2, \dots, 5$. Right panel shows $G(K)$ (blue circle) where K_{obs} specifies the number of detected internal states in individual traces (blue). (B, C, D) Other representative time traces and their $G(K)$ obtained under the same experimental condition.

doi:10.1371/journal.pcbi.1005286.g006

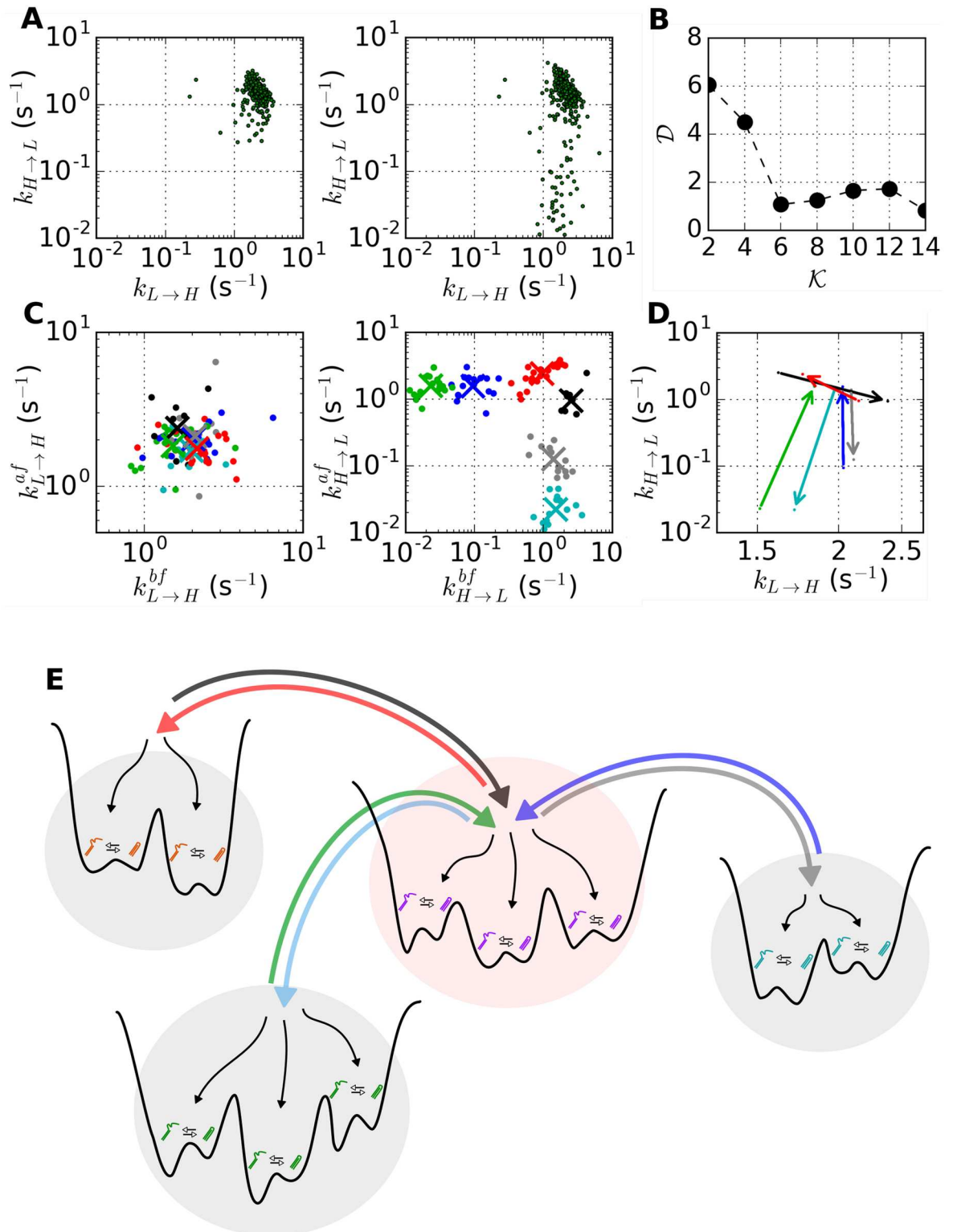


Fig 7. Clustering H-DNA data at $[Na^+] = 100$ mM condition. (A) The scatter plots of $(k_{L \rightarrow H}, k_{H \rightarrow L})$ before (left) and after (right) applying VB-DCMM from $[Na^+] = 100$ mM data. (B) The “average pairing distance” $D(\kappa)$ as a function of the number of clusters (κ) (see [Methods](#)). The minimum value of $D(\kappa)$ is found at $\kappa = 6$. (C) Left: the scatter plot of clustered data projected on $(k_{L \rightarrow H}^{bf}, k_{L \rightarrow H}^{af})$ plane. Right: scatter plot of clustered data projected on $(k_{H \rightarrow L}^{bf}, k_{H \rightarrow L}^{af})$ plane. The data belonging to different clusters are depicted in different colors, and the centroid of each cluster is marked with the \times symbol. Total 98 data points were used for

analysis. (D) The result of the above clustering is represented using 6 kinetic arrows which represent the centroids of each cluster represented in $(k_{L \rightarrow H}, k_{H \rightarrow L})$ plane. The starting point of the arrow is $(\langle k_{L \rightarrow H}^{(\mu), \text{bf}} \rangle, \langle k_{H \rightarrow L}^{(\mu), \text{bf}} \rangle)$ whereas the ending point of the arrow is $(\langle k_{L \rightarrow H}^{(\mu), \text{af}} \rangle, \langle k_{H \rightarrow L}^{(\mu), \text{af}} \rangle)$, where the superscript μ represents the index of each cluster $\mu = 1, 2, \dots, 6$. The colors used for depicting kinetic arrows are consistent with the data points in (C). (E) A schematic of the conformational landscape of H-DNA.

doi:10.1371/journal.pcbi.1005286.g007

a data point $(k_{L \rightarrow H}, k_{H \rightarrow L})$ for a time trace with dynamic disorder is a mixture of $(k_{L \rightarrow H}^{(\mu)}, k_{H \rightarrow L}^{(\mu)})$ with $\mu = 1, 2, \dots, K$. In the presence of clear distinction between internal states ($\mu \neq \nu$), the clustering of $(k_{L \rightarrow H}^{(\mu)}, k_{H \rightarrow L}^{(\mu)})$ would be straightforward, which is indeed the case for the synthetic data (S13A Fig). However, for the H-DNA data, even after the decomposition, the clustering of data on $(k_{L \rightarrow H}, k_{H \rightarrow L})$ plane (Fig 7A) is not that clear.

To improve the quality of clustering, we extended the clustering of the kinetic data to a higher dimension by considering the kinetic information of internal states that are contiguous (kinetically linked) along time traces. To be specific, for a time trace exhibiting a transition from the μ -th to ν -th internal state ($\mu \neq \nu$), one can consider that the inter-basin transition has occurred from the time interval represented by its pair of kinetic rate $(k_{L \rightarrow H}^{\text{bf}}, k_{H \rightarrow L}^{\text{bf}}) [= (k_{L \rightarrow H}^{(\mu)}, k_{H \rightarrow L}^{(\mu)})]$ to the next time interval represented by $(k_{L \rightarrow H}^{\text{af}}, k_{H \rightarrow L}^{\text{af}}) [= (k_{L \rightarrow H}^{(\nu)}, k_{H \rightarrow L}^{(\nu)})]$, where the superscripts, ‘bf’ and ‘af’ denotes ‘before’ and ‘after’ the transition, respectively. Thus, instead of $(k_{L \rightarrow H}, k_{H \rightarrow L})$, a clustering at a higher dimension can be carried out by measuring the Euclidean distance between a pair of the four-dimensional (4-dim) arrays, $(\log k_{L \rightarrow H}^{\text{bf}}, \log k_{H \rightarrow L}^{\text{bf}}, \log k_{L \rightarrow H}^{\text{af}}, \log k_{H \rightarrow L}^{\text{af}})$.

In order to cluster the 4-dim arrays we used the k-means clustering algorithm. Application of the algorithm to the H-DNA data at $[\text{Na}^+] = 100 \text{ mM}$ reveals that the average pairing distance, $\mathcal{D}(\mathcal{K})$ (see Methods), is minimized when the number of clusters is 6 ($\mathcal{K} = 6$), namely, the model with 6 clusters provides the best interpretation of the data (Fig 7B). Although the model with 14 clusters shows a smaller \mathcal{D} , we selected $\mathcal{K} = 6$ as the best solution, since for $\mathcal{K} = 14$ each of 12 clusters out of 14 has less than 10 data points, which makes the result of clustering statistically less significant (S16 Fig). This results remain qualitatively identical when L1 distance (so called “city block” distance) was used instead of “square-euclidean” distance (S19 Fig). Furthermore, the clustering algorithm using “affinity propagation” [58], which considers all the data points as possible exemplars (analogous to centroids in k-means clustering method) and iteratively exchanges messages between them, also gives qualitatively identical results, confirming the robustness of the conclusion on H-DNA dynamics obtained from VB-DCMM and k-mean clustering (see S20 Fig).

We present the result of clustering either (i) by projecting it on the two separate kinetic planes, $(k_{L \rightarrow H}^{\text{bf}}, k_{L \rightarrow H}^{\text{af}})$ and $(k_{H \rightarrow L}^{\text{bf}}, k_{H \rightarrow L}^{\text{af}})$, which visualize the inter-basin transitions in terms of variable $L \rightarrow H$ and $H \rightarrow L$ transition rates (see Fig 7C), or (ii) by using “interconversion arrows” linking the kinetic rates of two internal states, $[(\log k_{L \rightarrow H}^{\text{bf}}, \log k_{H \rightarrow L}^{\text{bf}}) \rightarrow (\log k_{L \rightarrow H}^{\text{af}}, \log k_{H \rightarrow L}^{\text{af}})]$ on the $(k_{L \rightarrow H}, k_{H \rightarrow L})$ plane (Fig 7D). Note that in the scatter plot visualized with $(k_{H \rightarrow L}^{\text{bf}}, k_{H \rightarrow L}^{\text{af}})$, the distinction between different clusters is clear (the right panel of Fig 7C). Furthermore, for a system in equilibrium or at least near equilibrium, the interconversion between two internal states, say μ and ν , should occur in both directions, i.e., $\mu \rightarrow \nu$ and $\nu \rightarrow \mu$. In the representation (i), a symmetry of $(k_{a \rightarrow b}^{(\mu)}, k_{a \rightarrow b}^{(\nu)}) = (k_{a \rightarrow b}^{(\nu)}, k_{a \rightarrow b}^{(\mu)})$ is expected in the both panels of Fig 7C; and in the representation (ii), the “arrows”, amounting to the kinetic connectivity between distinct internal states, should be bi-directional. The symmetry of the data plotted in Fig 7C or the bidirectionality of the kinetic arrows confirms the condition of detailed balance being satisfied in the system in equilibrium. Fig 7D depicts 6 kinetic arrows (3 pairs of reversible kinetic arrows) connecting the centroids of $(\log k_{L \rightarrow H}^{\text{bf}}, \log k_{L \rightarrow H}^{\text{af}})$ or $(\log k_{H \rightarrow L}^{\text{bf}}, \log k_{H \rightarrow L}^{\text{af}})$ data.

Application of the above clustering method to synthetic data with $K = 3$, $N = 2$ (S13 Fig) is straightforward. To check the efficacy of clustering method for a more complicated case, we have tested with synthetic data generated with $K = 4$, $N = 4$, i.e. when there are as many as 4 observable states in each internal state (S14 and S15 Figs). In the case with 4 observable states, total 12 possible intrabasin transitions are conceivable. Thus, the dimension of the array associated with interbasin transition is 24. As long as there is a clear time scale separation, it is expected that the pairing distance $\mathcal{D}(\mathcal{K} = 12)$ shows minimum as there are 12 connection paths between 4 internal states. Indeed, $\mathcal{D}(\mathcal{K})$ is minimized at $\mathcal{K} = 12$ (S15A Fig).

Lastly, it is noteworthy that the clustering method presented here is not limited to data analysis for systems in equilibrium, but can be extended to systems in nonequilibrium steady state [59] where the individual state-to-state kinetic transition rate is well defined using the reversible Markov process although the condition of detailed balance is no longer anticipated [60, 61]. The symmetry of data point and bidirectionality of kinetic arrows as in Fig 7C and 7D are still of use to cluster the kinetic information generated from a system in nonequilibrium steady states.

Folding energy landscape of H-DNA

We classified the “components” of a similar kinetic pattern ($k_{L \rightarrow H}$, $k_{H \rightarrow L}$) obtained from VB-DCMM into a single cluster which represents a kinetic path linking two independent basins of attraction (or internal states). For example, the kinetic paths in Fig 7D can be best understood by hypothesizing 4 internal states (four basins) linked by 6 kinetic paths. Thus, the conformational transition landscape of H-DNA at $[\text{Na}^+] = 100$ mM condition consists of 4 internal states with 3 reversible kinetic paths being established as illustrated in Fig 7E. At lower salt concentrations ($[\text{Na}^+] = 50$ mM (S17 Fig) and $[\text{Na}^+] = 26$ mM (S18 Fig)), H-DNA transitions slow down and the dispersion of data also increases; however, the overall structure of conformational landscape of H-DNA remains unchanged from the picture suggested in Fig 7E; thus, there is a central superbasin to which three other superbasins are kinetically connected (S17 and S18 Figs).

Contributions of our work

In comparison to other pre-existing methods, the advantage of our VB-DCMM in decoding dynamic disorder from a given trajectory is highlighted as follows:

- (1) Dynamic disorders in single molecule time trajectories are modeled using DCMM by assuming the presence of hidden internal states. While Aggregated Markov Model (AMM), which has been adopted in ion-channel community for time trace analysis of varying current [62–76], can be employed to analyze our data with dynamic disorder, DCMM is better in correctly decoding dynamic disorder than AMM. We found that AMM is prone to overpredict the transition between kinetic patterns (S25 Fig). Our method is more suitable to the data showing persistent dynamic patterns by suppressing unwanted frequent transition between kinetic patterns. Detailed explanations of connection and quantitative comparison between DCMM and AMM are provided in SI and S25 Fig.
- (2) In this paper, Bayesian version of DCMM was developed by using variational Bayes (VB) method, which enabled us to determine the number of internal states straightforwardly. Although Bayesian version of DCMM using Markov chain Monte Carlo (MCMC) method has previously been developed for the credit portfolio modeling [77], the idea of Bayesian inference in ref. [77] was used only for the purpose of calculating a posterior distribution of model parameters. To determine the number of hidden states corresponding to the internal states in

this study, the authors in ref. [77] used the economic cycle fluctuation model, instead. Our study combining VB with DCMM (i) can determine the number of internal states in a more objective fashion, (ii) offers intuitive way to incorporate prior knowledge, and (iii) is computationally more efficient than MCMC (See SI for details).

(3) We tested VB-DCMM under various conditions, by varying the kinetic rates, the number of observables, the number of hidden states, and prior parameters. New metrics were also devised to quantify the performance of algorithm systematically.

(4) Finally the connection paths (kinetic arrows) between internal states of H-DNA are clustered by using the kinetic components extracted from VB-DCMM and by applying k-means clustering algorithm to high dimensional arrays.

To recapitulate, our entire process of analyzing single molecule data is composed of three stages: (i) noise-filtering using HMM; (ii) decomposition of heterogeneous time traces into the homogeneous components using VB-DCMM; (iii) clustering the decomposed components into the same cluster.

In principle, this three-stage analysis can be made more systematic by combining the noise-filtering and clustering procedure with VB-DCMM. To be more specific, (1) The noise-filtering of observable trace (o_n) is processed, independently from the main VB-DCMM algorithm, by using HMM, which has been proved to be reliable in noise-filtering [56], and the maximum number of observables (N) are predetermined as an input parameter. Current version of algorithm can be further automated by combining with the Bayesian version of HMM [49], which can determine the number of observables while filtering the noise in data (See Fig 2D). The resulting model will have a similar structure with the modified factorial HMM [46, 47]. (2) The heterogeneous components identified from individual time traces are clustered separately from our main algorithm. It would be also desirable to unify the post-processing step (clustering) with VB-DCMM using empirical Bayes method which has been applied recently to analyze single molecule data [54, 55].

However, it should also be noted that a blind integration of noise-filtering and clustering steps inevitably complicates the implementation of VB-DCMM, as more number of prior parameters are ought to be decided by users. For example, Bayesian implementation of HMM for noise filtering demands manual determination of additional $N(N + 5)$ prior-parameters [49]. Compared to this, currently VB-DCMM requires users to pre-determine only one prior parameter which characterizes the final transition rate matrix, A (see the subsection: *Selection of prior parameters* in SI). Moreover, the integration of other methods will obscure the flow of analysis, making it difficult to identify an error-causing step. Keeping each step in the algorithm separate makes the integration of VB-DCMM to other applications more transparent (for example, if noise-filtering by HMM is unsuccessful, other advanced method can be employed [49]). We leave it as our future work to develop an algorithm that integrates the above-mentioned three procedures (noise-filtering, VB-DCMM, and clustering) without increasing complexity or obscuring the flow of analysis.

In decoding SM FRET data, the most notable difference of our VB-DCMM from the previous studies employing the probabilistic models such as maximum likelihood and Bayesian statistics is that VB-DCMM explicitly considers the situation that transition rates can change from one time interval to another within individual time traces. The previous studies [49–51, 56, 78–80] assumed that the transition rates were constant within individual time traces. Also, currently, VB-DCMM is applicable to window-averaging FRET trajectories. It will be of great interest to extend VB-DCMM to analyzing time trajectories in which arrival times for individual photons are available. VB-DCMM is particularly powerful when there is a separation in time scales between τ_{int} and τ_{conf} .

Concluding remarks

While the notion of dynamical heterogeneity or broken ergodicity seems better recognized in the research field of nucleic acids [81] than in proteins, which likely arises from more homopolymer-like nature of building block of nucleotides [82], biomolecules in general can have a rugged folding landscape with many local basins of attraction and kinetic barriers with varying heights [83]. Conformational dynamics of biomolecules on rugged landscapes can be heterogeneous, which gives rise to static or dynamic disorder depending on the time scale of observation or the height distribution of kinetic barrier. The presence of heterogeneity or disorder among individual molecules, unveiled by *in vitro* SM experiments could be surprising at first sight; however, it is also important to note that the general hypotheses in the conventional molecular biology towards a single native state have been put forward based on the observations from ensemble experiments where the heterogeneity, if any, is usually masked by the process of ensemble averaging. Given that the complexity of a molecular system increases with the system size (N_{sys}) as $\sim e^{N_{\text{sys}}}$ [29], it should not be too surprising to find such disorder in biomolecules in itself. Cells are equipped with molecular chaperones that can tame misfolding-prone biomolecules with rugged landscapes [84–87]; thus the principle of optimization in biology, if it fails at the level of a molecule in isolation, can be extended further to the molecular system including its environmental factors.

It is not easy to elucidate the molecular origin of disorder in a conclusive manner; yet, it has recently been suspected that interactions of biomolecules with cofactor such as ATP and multivalent metal-ions could be the microscopic causes for those molecules exhibiting dynamical heterogeneity [12, 13, 16, 88, 89]. Modulating the concentration of Mg^{2+} ions from high to low and again to high induced inter-conversions of dynamic patterns in equilibrium conformational fluctuations of *T. ribozyme* [12] and Holliday junctions [13]. Distinct velocities of ATP-empowered individual RecBED helicase motors, which can move progressively along dsDNA by unwinding it into two separate strands, can be reset by introducing a long pause by halting the supply of ATP. For the time trajectories of biomolecules displaying quenched disorder, a method to analyze such data was proposed using a concept from glass physics [13]. Here, to deal with more general scenarios, we have developed a method to analyze single molecule time traces with dynamic disorder.

As demonstrated by testing the VB-DCMM algorithm on synthetic data, the algorithm is quite accurate in decoding dynamic disorder as long as a time trajectory of interest contains multiple time intervals, each of which display kinetic pattern distinct from others. When a clear separation in timescale is present between two distinct kinetic patterns, large value of D_{conf} , D_{int} , and D_{tot} would be acquired.

While we developed the VB-DCMM algorithm primarily to analyze dynamic disorder in duplex-triplex transitions of H-DNA, the method is applicable to any data in the form of one-dimensional time series with multiple transitions. Together with a further technical advance in SM, which eliminates experimental artifacts as well as extends the measurement time, our algorithm developed here will contribute to better understanding of biomolecules that display heterogeneous dynamics.

Methods

Generation of synthetic data

Internal state sequence \mathbf{x} was generated by using Monte Carlo method with a constant transition matrix (homogeneous Markov chain model). The observable sequence \mathbf{o} was generated

by using the same method but with the transition matrix that was defined at each time t based on the internal state $x(t)$. Finally, Gaussian noise was added on \mathbf{o} to produce \mathbf{o}_n .

Single-molecule FRET measurements to monitor duplex-triplex transitions of H-DNA

We purchased triplex forming oligonucleotides from Integrated DNA Technologies (Coralville, IA, USA). The oligonucleotides were dissolved in T50 buffer solution (10 mM Tris-HCl, 50 mM NaCl, pH = 7.5) and were heated beyond the melting temperature of DNA duplex ($\sim 90^\circ\text{C}$), and slowly cooled down on a heat block to room temperature over 8 hour to properly hybridize them. The DNA prepared as such is called ‘‘H-DNA’’ here. The sequences of the triplex forming strands (purine-rich and pyrimidine-rich) are: Purine-rich strand: 5’ AAG AAG AAG AAG AAG (Cy5) TGG CGA CGG CAG CGA (Biotin) 3’, Pyrimidine-rich strand: 5’ TCG CTG CCG TCG CCA CTT CTT CTT CTT CTT TTT TCT TCT TCT TCT TCT TC (Cy3) 3’. In the purine-rich strand, the biotin at 3’ terminus is used to attach the H-DNA molecule to a neutravidin-coated cover-glass. The Cy3 and Cy5 dyes in the H-DNA molecule correspond to a donor and an acceptor for FRET measurements, respectively. In order to observe the transition between folded triplex and unfolded DNA, we used the reaction buffer containing 50 mM HEPES(Sigma-Aldrich) and various concentrations of Na^+ (26, 50, 100 mM). These buffer solutions also contained 2 mM trolox, 10% glucose and glyoxy for single-molecule fluorescence experiments. We utilized a home-made TIRF (Total Internal Reflection Fluorescence) microscope to measure the FRET efficiency between donor and acceptor dyes, which reveals the conformational state of the H-DNA molecule. A 532-nm laser (CrystaLaser DPSS, 10 mW) was used to excite donor molecules and fluorescence intensities of both dyes were measured by an EMCCD (Andor iXon DV887, Andor technology). To observe the change of FRET efficiency in real time, we measured the time-lapse FRET traces with the repetition rate of 10 Hz. To study kinetic features of the conformational transition with dynamic disorder, we acquired the FRET time traces for a long period (>100 sec).

Clustering at a higher dimension

For given N and K , total $N(N - 1)$ intra-basin transition rates $k_{a,b}^{(\mu)}$ ($a, b \in \{1, 2, \dots, N\}, a \neq b$) are defined in the μ -th basin (or μ -th internal state) and total $K(K - 1)$ inter-basin transitions are conceivable. To cluster the kinetic information of H-DNA data obtained from VB-DCMM, we consider the *kinetic arrow*, $2N(N - 1)$ -dimensional array of data, which has the structure of $C_i \equiv (\{\log k_{a,b}^{i,\text{bf}}\}, \{\log k_{a,b}^{i,\text{af}}\})$ where the subscript i denotes an index referring to one of $K(K - 1)$ possible inter-basin transitions linking two internal states ($\mu \neq \nu$). For a kinetic scheme made of a network of reversible transitions between K internal states, the transition between two internal states should be bidirectional; thus for a given inter-basin transition path i , there should be a kinetic path j antiparallel to the path i , satisfying $\|C_i - \tilde{C}_j\| \approx 0$, where $\tilde{C}_j \equiv (\{\log k_{a,b}^{j,\text{af}}\}, \{\log k_{a,b}^{j,\text{bf}}\})$. In our problem, the set of all the data generated as an outcome of VB-DCMM can in principle be clustered into the disjoint subsets of size 2 partitioning the \mathcal{K} transition paths, $\{\mathcal{K} | 1 \leq \mathcal{K} \leq K(K - 1)\}$, and one realization of such disjoint subsets will minimize the pairwise sum of Euclidean distances $\|C_\alpha - \tilde{C}_\beta\|$ for all α and β ; however, the method suffers from high computational cost as the possible number of clusters increases rapidly with N and K .

To alleviate the computational cost for large N and K , we modified the original method. We first searched the best partitioning set of data $S^*(\mathcal{K})$ for a given \mathcal{K} that minimizes the

Euclidean distance between all the pairs of centroids,

$$\mathcal{D}^c(\mathcal{K}) = \frac{2}{\mathcal{K}} \sum_{(i,j)} (d_{ij}^c)^2 \tag{7}$$

where $d_{ij}^c = \|\mathbf{C}_i^c - \tilde{\mathbf{C}}_j^c\|$ with $\mathbf{C}_i^c \equiv (\{\langle \log k_{a,b}^{i,bf} \rangle\}, \{\langle \log k_{a,b}^{i,af} \rangle\})$, $\tilde{\mathbf{C}}_j^c \equiv (\{\langle \log k_{a,b}^{j,af} \rangle\}, \{\langle \log k_{a,b}^{j,bf} \rangle\})$, and $\langle \dots \rangle$ denotes the centroid of clustered data. To obtain the best clustering result for a given \mathcal{K} , we conducted k-means clustering using *k_means* function from scikit-learn libraries [90] with 20,000 different random initial conditions in each analysis. It is expected that $\mathcal{D}^c(\mathcal{K}) = \frac{2}{\mathcal{K}} \sum_{(i,j)} (d_{ij}^c)^2 \geq \frac{2}{\mathcal{K}} \sum_{S^*(\mathcal{K})} (d_{ij}^c)^2$. The summation, $\sum_{(i,j)}$, signifies that the sum is taken over the disjoint subsets of size 2 partitioning a set $\{1, \dots, \mathcal{K}\}$ with \mathcal{K} being an even number) and $S^*(\mathcal{K})$ is the best partitioning set that minimizes the value of $\mathcal{D}^c(\mathcal{K})$ for a given \mathcal{K} . For example, provided that there are 4 kinetic arrows made of centroids ($i = 1, 2, 3, 4$), which minimizes \mathcal{D}^c at $\mathcal{K} = 2$ when $i = 1$ is paired with $i = 3$ and $i = 2$ with $i = 4$, then $S^*(2) = \{\{1, 3\}, \{2, 4\}\}$ and $\mathcal{D}^c(2) = d_{13}^c + d_{24}^c$.

Next, in order to decide the optimal \mathcal{K} , we calculated pairing distance between paired clusters in $S^*(\mathcal{K})$ again, but this time using all the elements in each cluster. The total pairing score

$$\mathcal{D}(\mathcal{K}) \equiv \frac{2}{\mathcal{K}} \sum_{S^*(\mathcal{K})} \langle d_{ij} \rangle, \tag{8}$$

where the average pairing distance between two clusters i and j is defined as

$\langle d_{ij} \rangle \equiv \frac{1}{M_i M_j} \sum_n^{M_i} \sum_m^{M_j} \|\mathbf{C}_{i_n} - \tilde{\mathbf{C}}_{j_m}\|$ where $\mathbf{C}_{i_n} = (\{\log k_{a,b}^{i_n,bf}\}, \{\log k_{a,b}^{i_n,af}\})$, $\tilde{\mathbf{C}}_{j_m} = (\{\log k_{a,b}^{j_m,af}\}, \{\log k_{a,b}^{j_m,bf}\})$, and n refers to an index for the element in the i -th cluster and m to an index for the elements in the j -th cluster. M_i is the total number of the elements in the i -th cluster. Finally, the optimal \mathcal{K}^* , minimizing $\mathcal{D}(\mathcal{K})$, is selected, i.e., $\mathcal{K}^* = \arg \min \mathcal{D}(\mathcal{K})$, and the interpretation of data is conducted for the best partitioning set $S^*(\mathcal{K} = \mathcal{K}^*)$.

For H-DNA data at three different Na^+ concentrations, the optimal \mathcal{K}^* are determined at $\mathcal{K}^* = 6$ for $[\text{Na}^+] = 100 \text{ mM}$ (Fig 7A), $\mathcal{K}^* = 10$ for $[\text{Na}^+] = 50 \text{ mM}$ (S17B Fig), $\mathcal{K}^* = 12$ for $[\text{Na}^+] = 26 \text{ mM}$ (S18B Fig). This implies that the complexity of conformational space of H-DNA increases at low salt condition (also see the scatter plot of $(k_{L \rightarrow H}, k_{H \rightarrow L})$ in Figs 7A, S17A and S18A).

The clustering results presented in this study remain robust regardless of the choice of distance metric. K-means clustering using *L1* distance (“city block”) measure with 20,000 different random initial conditions also was led to qualitatively similar results (S19 Fig). Furthermore, as an alternative clustering algorithm, we also tested “affinity propagation” [58] on our data, and the results remain qualitatively identical (see S20 Fig). In the affinity propagation method, negative square-euclidean distance was employed as a similarity metric ($s(i, j) = -\|\mathbf{x}_i - \mathbf{x}_j\|^2$) where \mathbf{x}_i denotes the coordinate of the i -th data point. The objective of the algorithm is to optimize the factorized probability distribution which approximates the net similarity \mathcal{S} , defined as $\mathcal{S} \sim \prod_{i=1}^N e^{s(i, c_i)}$. Here, c_i is the index of the exemplar of i -th data point \mathbf{x}_i . For example, if $c_i = k$, \mathbf{x}_k is an exemplar of \mathbf{x}_i and \mathbf{x}_i belongs to the cluster represented by \mathbf{x}_k . Multiple iterations of message passing are carried out until convergence is achieved in the result and the best result of clustering is acquired. For implementation, we used *AffinityPropagation* class from *scikit* [90] library with varying “preference” as an input parameter, where the preference denotes the logarithm of probability that i -th data point \mathbf{x}_i selects itself as an exemplar. Further details of the algorithm are available in Ref. [58].

Supporting Information

S1 Text. Supplementary Information.

(PDF)

S1 Fig. VB-DCMM analysis on synthetic data generated with the following parameters:

$$K^{\text{true}} = 2, \gamma^{(1) \rightarrow (2)} \Delta t = \gamma^{(2) \rightarrow (1)} \Delta t = 0.001, k_{L \rightarrow H}^{(1)} \Delta t = k_{H \rightarrow L}^{(1)} \Delta t = 0.05,$$

$$k_{L \rightarrow H}^{(2)} \Delta t = k_{H \rightarrow L}^{(2)} \Delta t = 0.0025.$$

(TIF)

S2 Fig. Re-calculated accuracy of model prediction ($\langle \chi \rangle$) by assuming “ $K = 1$ ”.

(TIF)

S3 Fig. Systematic validation of VB-DCMM on synthetic data generated under various conditions with $T_{\text{obs}}/\Delta t = 8800$.

(TIF)

S4 Fig. Accuracy of VB-DCMM on synthetic data with $T_{\text{obs}}/\Delta t = 4400$ and 2200.

(TIF)

S5 Fig. VB-DCMM analysis on synthetic data having three internal states.

(TIF)

S6 Fig. VB-DCMM analysis on synthetic data having 3 observables.

(TIF)

S7 Fig. VB-DCMM analysis on synthetic data having 4 observables ($o = 1, 2, 3,$ and 4) when internal state $x = 1$, or having 2 ($o = 1, 3$) observables when $x = 2$.

(TIF)

S8 Fig. Average accuracy of the model prediction ($\langle \chi \rangle$) versus $D_{\text{tot}} = D_{\text{conf}} + 0.8D_{\text{int}}$ from various synthetic data.

(TIF)

S9 Fig. Representative time traces of H-DNA dynamics and their analysis using VB-DCMM at $[\text{Na}^+] = 50$ mM.

(TIF)

S10 Fig. Representative time traces of H-DNA dynamics and their analysis using VB-DCMM at $[\text{Na}^+] = 26$ mM.

(TIF)

S11 Fig. Representative time traces of H-DNA dynamics which display more than two internal states within the trace at $[\text{Na}^+] = 26$ mM.

(TIF)

S12 Fig. VB-DCMM analysis on synthetic data generated with the following parameters:

$$T_{\text{obs}}/\Delta t = 4400, K^{\text{true}} = 2, \gamma^{(1) \rightarrow (2)} \Delta t = \gamma^{(2) \rightarrow (1)} \Delta t = 0.001, k_{L \rightarrow H}^{(1)} \Delta t = k_{H \rightarrow L}^{(1)} \Delta t = 0.05,$$

$$k_{L \rightarrow H}^{(2)} \Delta t = 0.00625, k_{H \rightarrow L}^{(2)} \Delta t = 0.0125.$$

(TIF)

S13 Fig. Clustering synthetic data with three internal states ($K = 3$) with two observable states ($N = 2$).

(TIF)

S14 Fig. VB-DCMM analysis on synthetic data having four internal states ($K = 4$) with four observable states ($N = 4$).

(TIF)

S15 Fig. Clustering synthetic data with three internal states ($K = 4$) with four observable states ($N = 4$).

(TIF)

S16 Fig. Clustering H-DNA data ($[Na^+] = 100$ mM).

(TIF)

S17 Fig. Clustering H-DNA data ($[Na^+] = 50$ mM).

(TIF)

S18 Fig. Clustering H-DNA data ($[Na^+] = 26$ mM).

(TIF)

S19 Fig. H-DNA data ($[Na^+] = 100$ mM) analyzed with k-means clustering algorithm using “city block” distance ($L1$ -distance).

(TIF)

S20 Fig. Clustering results of H-DNA data ($[Na^+] = 100$ mM) from “affinity propagation” method.

(TIF)

S21 Fig. Accuracy of the model prediction in terms of $\langle \chi \rangle$ under varying prior parameters u_{bs} and u_{bd} .

(TIF)

S22 Fig. Accuracy of the model prediction in terms of $\langle \chi \rangle$ under varying prior parameters u_{as} and u_{ad} .

(TIF)

S23 Fig. Effect of decomposing the original H-DNA time traces into its homogeneous Markov components.

(TIF)

S24 Fig. Comparison of $\varphi_{20} = \sigma_{20}/\mu_{20}$ histograms on synthetic data before and after removing dynamic heterogeneity (red) by decomposing the original traces into the pieces according to estimated internal state trace.

(TIF)

S25 Fig. Comparison between VB-DCMM and sticky-iAMM.

(TIF)

Acknowledgments

We thank the KIAS Center for Advanced Computation for providing computing resources.

Author Contributions

Conceptualization: WH CH.

Data curation: IBL SCH WH.

Formal analysis: WH.

Funding acquisition: SCH CH.

Investigation: IBL SCH.

Methodology: WH CH.

Project administration: SCH CH.

Resources: IBL SCH.

Software: WH.

Supervision: SCH CH.

Validation: WH.

Visualization: WH.

Writing – original draft: WH CH.

Writing – review & editing: WH CH.

References

1. Liphardt J, Onoa B, Smith SB, Tinoco I, Bustamante C. Reversible Unfolding of Single RNA Molecules by Mechanical Force. *Science*. 2001; 292(5517):733–737. doi: [10.1126/science.1058498](https://doi.org/10.1126/science.1058498) PMID: [11326101](https://pubmed.ncbi.nlm.nih.gov/11326101/)
2. Yang H, Luo G, Karnchanaphanurach P, Louie T, Rech I, Cova S, et al. Protein conformational dynamics probed by single-molecule electron transfer. *Science*. 2003; 302:262–266. doi: [10.1126/science.1086911](https://doi.org/10.1126/science.1086911) PMID: [14551431](https://pubmed.ncbi.nlm.nih.gov/14551431/)
3. Greenleaf WJ, Frieda KL, Foster DAN, Woodside MT, Block SM. Direct Observation of Hierarchical Folding in Single Riboswitch Aptamers. *Science*. 2008; 319(5863):630–633. doi: [10.1126/science.1151298](https://doi.org/10.1126/science.1151298) PMID: [18174398](https://pubmed.ncbi.nlm.nih.gov/18174398/)
4. Zhuang X, Kim H, Pereira MJB, Babcock HP, Walter NG, Chu S. Correlating Structural Dynamics and Function in Single Ribozyme Molecules. *Science*. 2002; 296(5572):1473–1476. doi: [10.1126/science.1069013](https://doi.org/10.1126/science.1069013) PMID: [12029135](https://pubmed.ncbi.nlm.nih.gov/12029135/)
5. Rhoades E, Gussakovskiy E, Haran G. Watching proteins fold one molecule at a time. *Proc Natl Acad Sci*. 2003; 100(6):3197–3202. doi: [10.1073/pnas.2628068100](https://doi.org/10.1073/pnas.2628068100) PMID: [12612345](https://pubmed.ncbi.nlm.nih.gov/12612345/)
6. Zhuang X, Rief M. Single-molecule folding. *Curr Opin Struct Biol*. 2003; 13(1):88–97. doi: [10.1016/S0959-440X\(03\)00011-3](https://doi.org/10.1016/S0959-440X(03)00011-3) PMID: [12581665](https://pubmed.ncbi.nlm.nih.gov/12581665/)
7. Stigler J, Ziegler F, Gieseke A, Gebhardt JCM, Rief M. The Complex Folding Network of Single Calmodulin Molecules. *Science*. 2011; 334(6055):512–516. doi: [10.1126/science.1207598](https://doi.org/10.1126/science.1207598) PMID: [22034433](https://pubmed.ncbi.nlm.nih.gov/22034433/)
8. Thirumalai D, Hyeon C. RNA and Protein Folding: Common Themes and Variations†. *Biochemistry*. 2005; 44(13):4957–4970. doi: [10.1021/bi047314+](https://doi.org/10.1021/bi047314+) PMID: [15794634](https://pubmed.ncbi.nlm.nih.gov/15794634/)
9. Mickler M, Dima RI, Dietz H, Hyeon C, Thirumalai D, Rief M. Revealing the bifurcation in the unfolding pathways of GFP by using single-molecule experiments and simulations. *Proc Natl Acad Sci*. 2007; 104(51):20268–20273. doi: [10.1073/pnas.0705458104](https://doi.org/10.1073/pnas.0705458104) PMID: [18079292](https://pubmed.ncbi.nlm.nih.gov/18079292/)
10. Rognoni L, Möst T, Žoldák G, Rief M. Force-dependent isomerization kinetics of a highly conserved proline switch modulates the mechanosensing region of filamin. *Proc Natl Acad Sci U S A*. 2014; 111(15):5568–5573. doi: [10.1073/pnas.1319448111](https://doi.org/10.1073/pnas.1319448111) PMID: [24706888](https://pubmed.ncbi.nlm.nih.gov/24706888/)
11. Qu X, Smith GJ, Lee KT, Sosnick TR, Pan T, Scherer NF. Single-molecule nonequilibrium periodic Mg²⁺-concentration jump experiments reveal details of the early folding pathways of a large RNA. *Proc Natl Acad Sci*. 2008; 105(18):6602–6607. doi: [10.1073/pnas.0801436105](https://doi.org/10.1073/pnas.0801436105) PMID: [18448679](https://pubmed.ncbi.nlm.nih.gov/18448679/)
12. Solomatin SV, Greenfield M, Chu S, Herschlag D. Multiple native states reveal persistent ruggedness of an RNA folding landscape. *Nature*. 2010; 463(7281):681–684. doi: [10.1038/nature08717](https://doi.org/10.1038/nature08717) PMID: [20130651](https://pubmed.ncbi.nlm.nih.gov/20130651/)
13. Hyeon C, Lee J, Yoon J, Hohng S, Thirumalai D. Hidden complexity in the isomerization dynamics of Holliday junctions. *Nat Chem*. 2012; 4(11):907–914. doi: [10.1038/nchem.1463](https://doi.org/10.1038/nchem.1463) PMID: [23089865](https://pubmed.ncbi.nlm.nih.gov/23089865/)
14. Haller A, Altman RB, Soulière MF, Blanchard SC, Micura R. Folding and ligand recognition of the TPP riboswitch aptamer at single-molecule resolution. *Proc Natl Acad Sci*. 2013; 110(11):4188–4193. doi: [10.1073/pnas.1218062110](https://doi.org/10.1073/pnas.1218062110) PMID: [23440214](https://pubmed.ncbi.nlm.nih.gov/23440214/)

15. Rinaldi AJ, Lund PE, Blanco MR, Walter NG. The Shine-Dalgarno sequence of riboswitch-regulated single mRNAs shows ligand-dependent accessibility bursts. *Nat Commun.* 2016; 7:1–10. doi: [10.1038/ncomms9976](https://doi.org/10.1038/ncomms9976) PMID: [26781350](https://pubmed.ncbi.nlm.nih.gov/26781350/)
16. Liu B, Baskin RJ, Kowalczykowski SC. DNA unwinding heterogeneity by RecBCD results from static molecules able to equilibrate. *Nature.* 2013; 500(7463):482–485. doi: [10.1038/nature12333](https://doi.org/10.1038/nature12333) PMID: [23851395](https://pubmed.ncbi.nlm.nih.gov/23851395/)
17. Lu HP, Xun L, Xie XS. Single-Molecule Enzymatic Dynamics. *Science.* 1998; 282(5395):1877–1882. doi: [10.1126/science.282.5395.1877](https://doi.org/10.1126/science.282.5395.1877) PMID: [9836635](https://pubmed.ncbi.nlm.nih.gov/9836635/)
18. English BP, Min W, van Oijen AM, Lee KT, Luo G, Sun H, et al. Ever-fluctuating single enzyme molecules: Michaelis-Menten equation revisited. *Nat Chem Biol.* 2005; 2(2):87–94. doi: [10.1038/nchembio759](https://doi.org/10.1038/nchembio759) PMID: [16415859](https://pubmed.ncbi.nlm.nih.gov/16415859/)
19. Zhuang X, Kim H, Pereira MJB, Babcock HP, Walter NG, Chu S. Correlating Structural Dynamics and Function in Single Ribozyme Molecules. *Science.* 2002; 296(5572):1473–1476. doi: [10.1126/science.1069013](https://doi.org/10.1126/science.1069013) PMID: [12029135](https://pubmed.ncbi.nlm.nih.gov/12029135/)
20. van Oijen AM, Blainey PC, Crampton DJ, Richardson CC, Ellenberger T, Xie XS. Single-Molecule Kinetics of λ Exonuclease Reveal Base Dependence and Dynamic Disorder. *Science.* 2003; 301(5637):1235–1238. doi: [10.1126/science.1084387](https://doi.org/10.1126/science.1084387) PMID: [12947199](https://pubmed.ncbi.nlm.nih.gov/12947199/)
21. Velonia K, Flomenbom O, Loos D, Masuo S, Cotlet M, Engelborghs Y, et al. Single-Enzyme Kinetics of CALB-Catalyzed Hydrolysis. *Angew Chem Int Ed.* 2005; 44(4):560–564. doi: [10.1002/anie.200460625](https://doi.org/10.1002/anie.200460625) PMID: [15619259](https://pubmed.ncbi.nlm.nih.gov/15619259/)
22. Flomenbom O, Velonia K, Loos D, Masuo S, Cotlet M, Engelborghs Y, et al. Stretched exponential decay and correlations in the catalytic activity of fluctuating single lipase molecules. *Proc Natl Acad Sci.* 2005; 102(7):2368–2372. doi: [10.1073/pnas.0409039102](https://doi.org/10.1073/pnas.0409039102) PMID: [15695587](https://pubmed.ncbi.nlm.nih.gov/15695587/)
23. Yang H, Luo G, Karnchanaphanurach P, Louie TM, Rech I, Cova S, et al. Protein Conformational Dynamics Probed by Single-Molecule Electron Transfer. *Science.* 2003; 302(5643):262–266. doi: [10.1126/science.1086911](https://doi.org/10.1126/science.1086911) PMID: [14551431](https://pubmed.ncbi.nlm.nih.gov/14551431/)
24. Min W, Luo G, Cherayil BJ, Kou SC, Xie XS. Observation of a Power-Law Memory Kernel for Fluctuations within a Single Protein Molecule. *Phys Rev Lett.* 2005; 94:198302. doi: [10.1103/PhysRevLett.94.198302](https://doi.org/10.1103/PhysRevLett.94.198302) PMID: [16090221](https://pubmed.ncbi.nlm.nih.gov/16090221/)
25. Cremer GD, Roeffaers MBJ, Baruah M, Sliwa M, Sels BF, Hofkens J, et al. Dynamic Disorder and Stepwise Deactivation in a Chymotrypsin Catalyzed Hydrolysis Reaction. *J Am Chem Soc.* 2007; 129(50):15458–15459. doi: [10.1021/ja077621d](https://doi.org/10.1021/ja077621d) PMID: [18031046](https://pubmed.ncbi.nlm.nih.gov/18031046/)
26. Rissin DM, Gorris HH, Walt DR. Distinct and Long-Lived Activity States of Single Enzyme Molecules. *J Am Chem Soc.* 2008; 130(15):5349–5353. doi: [10.1021/ja711414f](https://doi.org/10.1021/ja711414f) PMID: [18318491](https://pubmed.ncbi.nlm.nih.gov/18318491/)
27. Piwonski HM, Gomanovsky M, Bensimon D, Horovitz A, Haran G. Allosteric inhibition of individual enzyme molecules trapped in lipid vesicles. *Proc Natl Acad Sci U S A.* 2012; 109(22):E1437–E1443. doi: [10.1073/pnas.1116670109](https://doi.org/10.1073/pnas.1116670109) PMID: [22562794](https://pubmed.ncbi.nlm.nih.gov/22562794/)
28. Wu Z, Xing J. Functional Roles of Slow Enzyme Conformational Changes in Network Dynamics. *Biophys J.* 2012; 103(5):1052–1059. doi: [10.1016/j.bpj.2012.08.008](https://doi.org/10.1016/j.bpj.2012.08.008) PMID: [23009855](https://pubmed.ncbi.nlm.nih.gov/23009855/)
29. Palmer RG. Broken ergodicity. *Adv Phys.* 1982; 31(6):669–735. doi: [10.1080/00018738200101438](https://doi.org/10.1080/00018738200101438)
30. Hübner CG, Renn A, Renge I, Wild UP. Direct observation of the triplet lifetime quenching of single dye molecules by molecular oxygen. *J Chem Phys.* 2001; 115(21):9619–9622. doi: [10.1063/1.1421382](https://doi.org/10.1063/1.1421382)
31. Elenko MP, Szostak JW, van Oijen AM. Single-molecule binding experiments on long time scales. *Rev Sci Instrum.* 2010; 81(8). doi: [10.1063/1.3473936](https://doi.org/10.1063/1.3473936) PMID: [20815611](https://pubmed.ncbi.nlm.nih.gov/20815611/)
32. Hwang W, Bae S, Hohng S. Autofocusing system based on optical astigmatism analysis of single-molecule images. *Opt Express.* 2012; 20(28):29353–29360. doi: [10.1364/OE.20.029353](https://doi.org/10.1364/OE.20.029353) PMID: [23388762](https://pubmed.ncbi.nlm.nih.gov/23388762/)
33. Hyeon C, Hinczewski M, Thirumalai D. Evidence of Disorder in Biological Molecules from Single Molecule Pulling Experiments. *Phys Rev Lett.* 2014; 112:138101. doi: [10.1103/PhysRevLett.112.138101](https://doi.org/10.1103/PhysRevLett.112.138101) PMID: [24745458](https://pubmed.ncbi.nlm.nih.gov/24745458/)
34. Hinczewski M, Hyeon C, Thirumalai D. Directly measuring single molecule heterogeneity using force spectroscopy. *Proc Natl Acad Sci U S A.* 2016; 113:E3852–E3861. doi: [10.1073/pnas.1518389113](https://doi.org/10.1073/pnas.1518389113) PMID: [27317744](https://pubmed.ncbi.nlm.nih.gov/27317744/)
35. Zwanzig R. Dynamical disorder: Passage through a fluctuating bottleneck. *The J Chem Phys.* 1992; 97(5):3587–3589. doi: [10.1063/1.462993](https://doi.org/10.1063/1.462993)
36. Zwanzig R. Rate processes with dynamical disorder. *Acc Chem Res.* 1990; 23(5):148–152. doi: [10.1021/ar00173a005](https://doi.org/10.1021/ar00173a005)

37. Wilson GF, Kaczmarek LK. Mode-switching of a voltage-gated cation channel is mediated by a protein kinase A-regulated tyrosine phosphatase. *Nature*. 1993; 366(6454):433–438. doi: [10.1038/366433a0](https://doi.org/10.1038/366433a0) PMID: [8247151](https://pubmed.ncbi.nlm.nih.gov/8247151/)
38. Siekmann I, Sneyd J, Crampin EJ. Statistical analysis of modal gating in ion channels. *Proc R Soc A*. 2014; 470(2166):20140030–20140030. doi: [10.1098/rspa.2014.0030](https://doi.org/10.1098/rspa.2014.0030)
39. Berchtold A. The Double Chain Markov Model. Technical Report, Washington Univ. 1999;(348).
40. Poritz A. Linear predictive hidden Markov models and the speech signal. In: *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'82*. vol. 7; 1982. p. 1291–1294.
41. Kenny P, Lennig M, Mermelstein P. A linear predictive HMM for vector-valued observations with applications to speech recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*. 1990; 38(2):220–225. doi: [10.1109/29.103057](https://doi.org/10.1109/29.103057)
42. Wellekens C. Explicit time correlation in hidden Markov models for speech recognition. In: *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'87*. vol. 12; 1987. p. 384–386.
43. Paliwal KK. Use of temporal correlation between successive frames in a hidden Markov model based speech recognizer. In: *Acoustics, Speech, and Signal Processing, 1993. ICASSP-93., 1993 IEEE International Conference on*. vol. 2; 1993. p. 215–218 vol.2.
44. Lee IB, Lee JY, Lee NK, Hong SC. Direct observation of the formation of DNA triplexes by single-molecule FRET measurements. *Curr Appl Phys*. 2012; 12(4):1027–1032. doi: [10.1016/j.cap.2011.12.026](https://doi.org/10.1016/j.cap.2011.12.026)
45. Lee IB, Hong SC, Lee NK, Johner A. Kinetics of the Triplex-Duplex Transition in DNA. *Biophys J*. 2012; 103(12):2492–2501. doi: [10.1016/j.bpj.2012.10.029](https://doi.org/10.1016/j.bpj.2012.10.029) PMID: [23260051](https://pubmed.ncbi.nlm.nih.gov/23260051/)
46. Ghahramani Z, Jordan MI. Factorial Hidden Markov Models. *Mach Learn*. 1997; 29(2):245–273. doi: [10.1023/A:1007425814087](https://doi.org/10.1023/A:1007425814087)
47. Jordan MI, Ghahramani Z, Saul LK. Hidden Markov decision trees. *Adv Neural Inf Process Syst*. 1997;.
48. Bishop CM. *Pattern Recognition and Machine Learning*. Springer; 2006.
49. Bronson JE, Fei J, Hofman JM, Jr RLG, Wiggins CH. Learning Rates and States from Biophysical Time Series: A Bayesian Approach to Model Selection and Single-Molecule FRET Data. *Biophys J*. 2009; 97(12):3196–3205. doi: [10.1016/j.bpj.2009.09.031](https://doi.org/10.1016/j.bpj.2009.09.031) PMID: [20006957](https://pubmed.ncbi.nlm.nih.gov/20006957/)
50. Bronson J, Hofman J, Fei J, Gonzalez R, Wiggins C. Graphical models for inferring single molecule dynamics. *BMC Bioinformatics*. 2010; 11(Suppl 8):S2. doi: [10.1186/1471-2105-11-S8-S2](https://doi.org/10.1186/1471-2105-11-S8-S2) PMID: [21034427](https://pubmed.ncbi.nlm.nih.gov/21034427/)
51. Okamoto K, Sako Y. Variational Bayes Analysis of a Photon-Based Hidden Markov Model for Single-Molecule FRET Trajectories. *Biophys J*. 2012; 103(6):1315–1324. doi: [10.1016/j.bpj.2012.07.047](https://doi.org/10.1016/j.bpj.2012.07.047) PMID: [22995504](https://pubmed.ncbi.nlm.nih.gov/22995504/)
52. Persson F, Linden M, Unoson C, Elf J. Extracting intracellular diffusive states and transition rates from single-molecule tracking data. *Nat Meth*. 2013; 10(3):265–269. doi: [10.1038/nmeth.2367](https://doi.org/10.1038/nmeth.2367) PMID: [23396281](https://pubmed.ncbi.nlm.nih.gov/23396281/)
53. Johnson S, van de Meent JW, Phillips R, Wiggins CH, Lindén M. Multiple LacI-mediated loops revealed by Bayesian statistics and tethered particle motion. *Nucleic Acids Res*. 2014; 42(16):10265–10277. doi: [10.1093/nar/gku563](https://doi.org/10.1093/nar/gku563) PMID: [25120267](https://pubmed.ncbi.nlm.nih.gov/25120267/)
54. van de Meent JW, Bronson JE, Wood F, Gonzalez RL Jr, Wiggins CH. Hierarchically-coupled hidden Markov models for learning kinetic rates from single-molecule data. *Proc Int Conf Machine Learn*. 2013; 28:361–369. PMID: [26985282](https://pubmed.ncbi.nlm.nih.gov/26985282/)
55. van de Meent JW, Bronson JE, Wiggins CH, Jr RLG. Empirical Bayes Methods Enable Advanced Population-Level Analyses of Single-Molecule FRET Experiments. *Biophys J*. 2014; 106(6):1327–1337. doi: [10.1016/j.bpj.2013.12.055](https://doi.org/10.1016/j.bpj.2013.12.055) PMID: [24655508](https://pubmed.ncbi.nlm.nih.gov/24655508/)
56. McKinney SA, Joo C, Ha T. Analysis of single-molecule FRET trajectories using hidden Markov modeling. *Biophys J*. 2006; 91(5):1941–1951. doi: [10.1529/biophysj.106.082487](https://doi.org/10.1529/biophysj.106.082487) PMID: [16766620](https://pubmed.ncbi.nlm.nih.gov/16766620/)
57. A Stein W, et al. *Sage Mathematics Software (Version 6.3.)*; 2014.
58. Frey BJ, Dueck D. Clustering by Passing Messages Between Data Points. *Science*. 2007; 315(5814):972–976. doi: [10.1126/science.1136800](https://doi.org/10.1126/science.1136800) PMID: [17218491](https://pubmed.ncbi.nlm.nih.gov/17218491/)
59. English BP, Min W, Van Oijen AM, Lee KT, Luo G, Sun H, et al. Ever-fluctuating single enzyme molecules: Michaelis-Menten equation revisited. *Nature Chem Biol*. 2006; 2(2):87–94. doi: [10.1038/nchembio759](https://doi.org/10.1038/nchembio759) PMID: [16415859](https://pubmed.ncbi.nlm.nih.gov/16415859/)
60. Beard DA, Qian H. *Chemical Biophysics: Quantitative Analysis of Cellular Systems*. Cambridge University Press; 2008.
61. Hyeon C, Klumpp S, Onuchic JN. Kinesin's backsteps under mechanical load. *Phys Chem Chem Phys*. 2009; 11:4899–4910. doi: [10.1039/b903536b](https://doi.org/10.1039/b903536b) PMID: [19506765](https://pubmed.ncbi.nlm.nih.gov/19506765/)

62. Donald R Fredkin JAR. On Aggregated Markov Processes. *J Appl Probab.* 1986; 23(1):208–214. doi: [10.2307/3214130](https://doi.org/10.2307/3214130)
63. Kienker P. Equivalence of Aggregated Markov Models of Ion-Channel Gating. *Proc R Soc Lond B.* 1989; 236(1284):269–309. doi: [10.1098/rspb.1989.0024](https://doi.org/10.1098/rspb.1989.0024) PMID: [2471201](https://pubmed.ncbi.nlm.nih.gov/2471201/)
64. Colquhoun D H AG. On the Stochastic Properties of Single Ion Channels. *Proc R Soc Lond B.* 1981; 211(1183):205–235. doi: [10.1098/rspb.1981.0003](https://doi.org/10.1098/rspb.1981.0003) PMID: [6111797](https://pubmed.ncbi.nlm.nih.gov/6111797/)
65. Horn R, Lange K. Estimating kinetic constants from single channel data. *Biophys J.* 1983; 43(2):207–223. doi: [10.1016/S0006-3495\(83\)84341-0](https://doi.org/10.1016/S0006-3495(83)84341-0) PMID: [6311301](https://pubmed.ncbi.nlm.nih.gov/6311301/)
66. Qin F. Restoration of Single-Channel Currents Using the Segmental k-Means Method Based on Hidden Markov Modeling. *Biophys J.* 2004; 86(3):1488–1501. doi: [10.1016/S0006-3495\(04\)74217-4](https://doi.org/10.1016/S0006-3495(04)74217-4) PMID: [14990476](https://pubmed.ncbi.nlm.nih.gov/14990476/)
67. Qin F, Auerbach A, Sachs F. Maximum likelihood estimation of aggregated Markov processes. *Proc R Soc Lond B.* 1997; 264(1380):375–383. doi: [10.1098/rspb.1997.0054](https://doi.org/10.1098/rspb.1997.0054) PMID: [9107053](https://pubmed.ncbi.nlm.nih.gov/9107053/)
68. Wagner M, Michalek S, Timmer J. Estimating transition rates in aggregated Markov models of ion channel gating with loops and with nearly equal dwell times. *Proc R Soc Lond B.* 1999; 266(1431):1919–1926. doi: [10.1098/rspb.1999.0867](https://doi.org/10.1098/rspb.1999.0867)
69. Ball FG, Cai Y, Kadane JB, O'Hagan A. Bayesian inference for ion-channel gating mechanisms directly from single-channel recordings, using Markov chain Monte Carlo. *Proc R Soc Lond A.* 1999; 455(1988):2879–2932. doi: [10.1098/rspa.1999.0432](https://doi.org/10.1098/rspa.1999.0432)
70. Rosales RA. MCMC for hidden Markov models incorporating aggregation of states and filtering. *Bull Math Biol.* 2004; 66(5):1173–1199. doi: [10.1016/j.bulm.2003.12.001](https://doi.org/10.1016/j.bulm.2003.12.001) PMID: [15294422](https://pubmed.ncbi.nlm.nih.gov/15294422/)
71. Rosales R, Stark JA, Fitzgerald WJ, Hladky SB. Bayesian Restoration of Ion Channel Records using Hidden Markov Models. *Biophys J.* 2001; 80(3):1088–1103. doi: [10.1016/S0006-3495\(01\)76087-0](https://doi.org/10.1016/S0006-3495(01)76087-0) PMID: [11222275](https://pubmed.ncbi.nlm.nih.gov/11222275/)
72. Gin E, Falcke M, Wagner LE, Yule DI, Sneyd J. Markov chain Monte Carlo fitting of single-channel data from inositol trisphosphate receptors. *J Theor Biol.* 2009; 257(3):460–474. doi: [10.1016/j.jtbi.2008.12.020](https://doi.org/10.1016/j.jtbi.2008.12.020) PMID: [19168073](https://pubmed.ncbi.nlm.nih.gov/19168073/)
73. Siekmann I, L W II, Yule D, Fox C, Bryant D, Crampin E, et al. MCMC Estimation of Markov Models for Ion Channels. *Biophys J.* 2011; 100(8):1919–1929. doi: [10.1016/j.bpj.2011.02.059](https://doi.org/10.1016/j.bpj.2011.02.059) PMID: [21504728](https://pubmed.ncbi.nlm.nih.gov/21504728/)
74. Siekmann I, Sneyd J, Crampin E. MCMC Can Detect Nonidentifiable Models. *Biophys J.* 2012; 103(11):2275–2286. doi: [10.1016/j.bpj.2012.10.024](https://doi.org/10.1016/j.bpj.2012.10.024) PMID: [23283226](https://pubmed.ncbi.nlm.nih.gov/23283226/)
75. Hodgson MEA, Green PJ. Bayesian choice among Markov models of ion channels using Markov chain Monte Carlo. *Proc R Soc Lond A.* 1999; 455(1989):3425–3448. doi: [10.1098/rspa.1999.0459](https://doi.org/10.1098/rspa.1999.0459)
76. Hines K, Bankston J, Aldrich R. Analyzing Single-Molecule Time Series via Nonparametric Bayesian Inference. *Biophys J.* 2015; 108(3):540–556. doi: [10.1016/j.bpj.2014.12.016](https://doi.org/10.1016/j.bpj.2014.12.016) PMID: [25650922](https://pubmed.ncbi.nlm.nih.gov/25650922/)
77. Fitzpatrick M, Marchev D. Efficient Bayesian estimation of the multivariate Double Chain Markov Model. *Stat and Comput.* 2013; 23(4):467–480. doi: [10.1007/s11222-012-9323-y](https://doi.org/10.1007/s11222-012-9323-y)
78. Schröder GF, Grubmüller H. Maximum likelihood trajectories from single molecule fluorescence resonance energy transfer experiments. *J Chem Phys.* 2003; 119(18):9920–6. doi: [10.1063/1.1616511](https://doi.org/10.1063/1.1616511)
79. Antonik M, Felekyan S, Gaiduk A, Seidel CAM. Separating Structural Heterogeneities from Stochastic Variations in Fluorescence Resonance Energy Transfer Distributions via Photon Distribution Analysis. *J Phys Chem B.* 2006; 110(13):6970–6978. doi: [10.1021/jp057257+](https://doi.org/10.1021/jp057257+) PMID: [16571010](https://pubmed.ncbi.nlm.nih.gov/16571010/)
80. Gopich IV, Szabo A. Decoding the Pattern of Photon Colors in Single-Molecule FRET. *J Phys Chem B.* 2009; 113(31):10965–10973. doi: [10.1021/jp903671p](https://doi.org/10.1021/jp903671p) PMID: [19588948](https://pubmed.ncbi.nlm.nih.gov/19588948/)
81. Al-Hashimi HM, Walter NG. RNA dynamics: it is about time. *Curr Opin Struct Biol.* 2008; 18(3):321–329. doi: [10.1016/j.sbi.2008.04.004](https://doi.org/10.1016/j.sbi.2008.04.004) PMID: [18547802](https://pubmed.ncbi.nlm.nih.gov/18547802/)
82. Thirumalai D, Hyeon C. RNA and Protein folding: Common Themes and Variations. *Biochemistry.* 2005; 44(13):4957–4970. doi: [10.1021/bi047314+](https://doi.org/10.1021/bi047314+) PMID: [15794634](https://pubmed.ncbi.nlm.nih.gov/15794634/)
83. Thirumalai D, O'Brien EP, Morrison G, Hyeon C. Theoretical Perspectives on Protein Folding. *Annu Rev Biophys.* 2010; 39:159–183. doi: [10.1146/annurev-biophys-051309-103835](https://doi.org/10.1146/annurev-biophys-051309-103835) PMID: [20192765](https://pubmed.ncbi.nlm.nih.gov/20192765/)
84. Bhaskaran H, Russell R. Kinetic redistribution of native and misfolded RNAs by DEAD-box chaperone. *Nature.* 2007; 449:1014–1018. doi: [10.1038/nature06235](https://doi.org/10.1038/nature06235) PMID: [17960235](https://pubmed.ncbi.nlm.nih.gov/17960235/)
85. Woodson SA. Taming free energy landscapes with RNA chaperones. *RNA biology.* 2010; 7(6):677–686. doi: [10.4161/ma.7.6.13615](https://doi.org/10.4161/ma.7.6.13615) PMID: [21045544](https://pubmed.ncbi.nlm.nih.gov/21045544/)
86. Thirumalai D, Lorimer GH. Chaperonin-mediated protein folding. *Ann Rev Biophys Biomol Struct.* 2001; 30:245–269. doi: [10.1146/annurev.biophys.30.1.245](https://doi.org/10.1146/annurev.biophys.30.1.245)

87. Hyeon C, Thirumalai D. Generalized iterative annealing model for the action of RNA chaperones. *J Chem Phys*. 2013; 139(12):121924. doi: [10.1063/1.4818594](https://doi.org/10.1063/1.4818594) PMID: [24089736](https://pubmed.ncbi.nlm.nih.gov/24089736/)
88. Kowanko D, König SLB, Skilandat M, Kruschel D, Hadzic MCAS, Cardo L, et al. Cation-induced kinetic heterogeneity of the intron–exon recognition in single group II introns. *Proc Natl Acad Sci*. 2015; 112(11):3403–3408. doi: [10.1073/pnas.1322759112](https://doi.org/10.1073/pnas.1322759112) PMID: [25737541](https://pubmed.ncbi.nlm.nih.gov/25737541/)
89. Segev E, Smith Y, Ben-Yehuda S. RNA Dynamics in Aging Bacterial Spores. *Cell*. 2012; 148(1–2):139–149. PMID: [22209493](https://pubmed.ncbi.nlm.nih.gov/22209493/)
90. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. *J Mach Learn Res*. 2011; 12:2825–2830.