THE JOURNAL OF PHYSICAL CHEMISTRY B
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCB

Article

# Transition Time Determination of Single-Molecule FRET Trajectories via Wasserstein Distance Analysis in Steady-State Variations in smFRET (WAVE)

*Published as part of The Journal of Physical Chemistry virtual special issue "Hiro-o Hamaguchi Festschrift".*

Ting Chen, Fengnan Gao,* and Yan-Wen Tan*

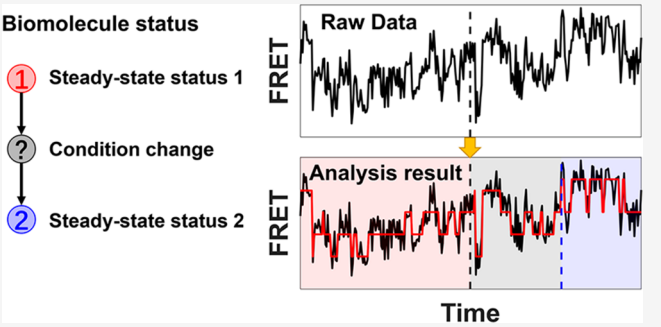
Cite This: *J. Phys. Chem. B* 2023, 127, 7819−7828

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information



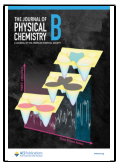
**ABSTRACT:** Many biological molecules respond to external stimuli that can cause their conformational states to shift from one steady state to another. Single-molecule FRET (Fluorescence Resonance Energy Transfer) is of particular interest to not only define the steady-state conformational ensemble usually averaged out in the ensemble of molecules but also characterize the dynamics of biomolecules. To study steady-state transitions, i.e., non-equilibrium transitions, a data analysis methodology is necessary to analyze single-molecule FRET photon trajectories, which contain mixtures of contributions from two steady-state statuses and include non-equilibrium transitions. In this study, we introduce a novel methodology called WAVE (Wasserstein distance Analysis in steady-state Variations in smFRET) to detect and locate non-equilibrium transition positions in FRET trajectories. Our method first utilizes a combined STaSI-HMM (Stepwise Transitions with State Inference Hidden Markov Model) algorithm to convert the original FRET trajectories into discretized trajectories. We then apply Maximum Wasserstein Distance analysis to differentiate the FRET state compositions of the fitting trajectories before and after the non-equilibrium transition. Forward and backward algorithms, based on the Minimum Description Length (MDL) principle, are used to find the refined positions of the non-equilibrium transitions. This methodology allows us to observe changes in experimental conditions in chromophore-tagged biomolecules or vice versa.




## I. INTRODUCTION

Single-molecule spectroscopy has been widely used to study biological macromolecules. One of its most valuable capabilities is probing the dynamic processes of biological molecules, which is typically achieved through measurements of single-molecule Förster-type Resonance Energy Transfer (smFRET).[1] SmFRET has the efficacy of revealing the underlying conformational heterogeneous substate distributions of biological macromolecules. In addition, it can measure the conformational dynamics of biomolecules under an environment that resembles their native physiological condition. To obtain information about the dynamics, several data analytic schemes are popular for analyzing smFRET photon trajectories. The most straightforward method is to summarize the dwell times on the smFRET trajectories. However, this only applies to trajectories with sharp and distinctive state separations. Owing to the fluctuating nature of single-molecule fluorescent signals, the extraction of reliable dynamics usually resorts to sophisticated statistical tools and data analytic methods. Roughly speaking, researchers have studied the issue from two different angles: denoising algorithms to remove intrinsic and observational errors and searching algorithms for change-points reflecting switching of different substates in FRET trajectories. For instance, a deconvolution scheme[2] based on equal information binning[3] has been developed by Haw Yang's group which can eliminate the intrinsic effects of photon-counting noise. Denoising using more complex mathematics such as wavelets has also been proposed.[4] For state-assigning change-point algorithms, Watkins and Yang have devised a change-point algorithm based on Bayesian inference.[5] However, it requires a large number of photons and is, thus, not suitable for fast dynamics. Alternatively, Shuang et al. proposed a simple algorithm called Fast Step Transition and State Identification (STaSI)[6] based on the minimum description length (MDL)[7,8] principle. The MDL principle

Received: April 14, 2023
Revised: August 20, 2023
Published: September 6, 2023







https://doi.org/10.1021/acs.jpcb.3c02498
J. Phys. Chem. B 2023, 127, 7819−7828

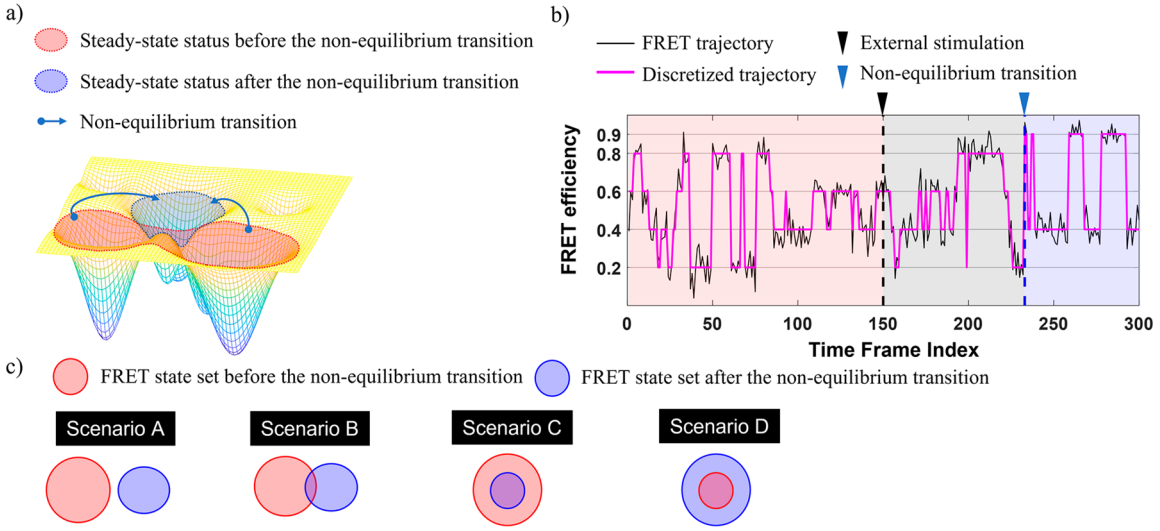



**Figure 1.** Schematic diagram of biomolecules' non-equilibrium transitions. (a) Illustration of the free energy landscape of biomolecules. Following external stimulation, they move from one steady-state status to another through non-equilibrium transition. (b) A typical FRET trajectory that contains a non-equilibrium transition, where the data is a mixture of contributions from two steady-state statuses. The black dashed line represents the position of external stimulation, while the blue dashed line indicates the position of the non-equilibrium transition. (c) Four scenarios of state compositions before and after the non-equilibrium transition, including the state sets before and after the non-equilibrium transition are completely nonconcurrent (scenario A); two state sets are partially overlapped (scenario B); the state set after (before) the non-equilibrium transition is a subset of that before (after) the non-equilibrium transition (scenarios C and D, respectively).

evaluates fitting and penalizes extra fitting parameters,[9,10] and has shown promising results. It can be applied to single-molecule intensity−time trajectories covering the dynamic ranges of common biological macromolecules with reasonable computational costs.

Beyond dwell times, several approaches with different advantages and disadvantages have been utilized to extract the dynamics of biological molecules. The correlation functions of the photon-by-photon trajectories can provide a variety of information, although averaging a pool of trajectories from many molecules is typically required.[11] Motion narrowing theory,[12] borrowed from traditional spectroscopy techniques, has also proven successful in studying protein folding[13] and enzymatic activities.[14] Gopich and Szabo have proposed a method based on maximizing specified likelihood functions,[15] which can be used to probe very fast dynamics[16] or even lifetimes,[17] although it becomes computationally expensive for slow processes. Hidden Markov Models (HMMs)[18−20] have been prominent for single-molecule dynamics analysis. HMM can simultaneously determine the population and transition rates among different substates, where the number of states is usually estimated by optimizing the Bayesian Information Criterion (BIC).[9,10] Recently, single-molecule researchers coming from different laboratories have joined forces to work on a bench-mark analysis of many commonly used single-molecule data analysis algorithms.[21] Applying methods that are appropriate for specific applications is crucial (for more discussions, see the Method section).

The methods mentioned above have been applied to reveal the dynamic nature of biological macromolecules in a variety of ways. Intramolecular conformational dynamics are an essential aspect of the structure−function relationships of these molecules. For some molecules, their conformational states and dynamics may vary due to external stimulation, such as light,[22,23] a sudden change in ionic concentration,[24] temperature,[25] hormones,[26,27] electric potential,[28] or even magnetic fields. Therefore, a data analysis scheme that accommodates non-equilibrium shifts between steady-state statuses is essential to investigate potential shifts in biomolecule conformational free energy landscapes[29] (Figure 1).

We propose a statistical inference scheme to test (statistically) whether a single molecule has transitioned from one steady state to another and, if yes, also to determine the transition time, which is the time lapse between an external stimuli and when the system set into another steady-state status. Our work focuses on the analysis of single-molecule FRET intensity−time trajectories while envisioning that our method can be applied to similar problems. Our approach combines STaSI and HMM[20] to determine the number of (sub)states and exchange rates before and after the external excitation that changes the molecular condition. Since the excitation is often carried out according to preset experimental designs, the precise time when the excitation takes place is considered known, and we can pinpoint the end of free energy landscape migration using a novel maximum Wasserstein distance (MWD) analysis, as proposed below. We showcase the feasibility of our method by applying it to simulated smFRET intensity−time trajectories based on realistic experimental parameters. A proof-of-principle experiment is carried out to study the transition patterns of the denaturing of a TAR−DNA hairpin. The capabilities and limitations of our approach are thoroughly discussed. Given the demonstrated robustness and user-friendliness of our method, we believe it will be more than useful in studying the dynamic processes of biological macromolecules.

## II. METHOD

In the workflow of existing FRET experiments, conditional changes are typically performed before or after data collection. For one thing, it is difficult to change experimental conditions in real-time observations such as altering the salt concentration of the buffer solution in a well-controlled and uniform manner. For the other thing, it is challenging to analyze data that undergo conditional changes during observation time.


https://doi.org/10.1021/acs.jpcb.3c02498
J. Phys. Chem. B 2023, 127, 7819−7828

However, in most FRET experiments, we expect the FRET efficiencies of a biomolecule and the steady-state status in which it resides change alongside condition changes. Thus, biomolecules can transit from one steady-state status to another, which depends on the external excitation, with the process being either reversible or irreversible. For some conditions of interest, the transitions are typically irreversible or require a certain amount of time to relax back to the original state. Consequently, FRET trajectories that encounter condition changes during observation time may contain a mixture of contributions from two steady-state statuses, including a non-equilibrium transition as the change point. Detecting the existence of such transitions and sorting the mixture of different steady-state statuses present significant challenges for researchers.

We introduce a novel method to identify and locate such non-equilibrium transitions in FRET trajectories that result from the changing conditions. The method does not require any prior knowledge of the steady-state models of biomolecules before and/or after the non-equilibrium transitions, but rather it only relies on two assumptions. First, biomolecules are in steady-state statuses or can be fitted by steady-state models before and after the non-equilibrium transition. Second, the duration of the non-equilibrium transition for biomolecules under those given conditions is insignificant in comparison to the overall observation period.

Our method is suitable for working with a collection of FRET trajectories, ranging from dozens to hundreds of homologous biomolecules that undergo the same condition change. The key information we require in advance is the time point of the condition change, which is determined by the experimental design. The method's workflow can be broken down into three steps (Figure 2). First, we employ the STaSI approach, along with a HMM,[20] to fit the entire data set and obtain a discretized fitting trajectory for each FRET trajectory. In step two, a MWD analysis dissects the discrete fitting trajectories at potential change points with the largest Wasserstein distances to identify the state compositions before and after the non-equilibrium transition. In the third and final stage, we use forward and backward algorithms based on the MDL principle to reach a refined localization of the non-equilibrium transition position on each FRET trajectory.

In FRET experiments, photons emitted by labeled biomolecules in multiple channels (usually two different colors) are detected and converted to intensity−time trajectories. These trajectories are contaminated by noise and small errors, which mark the dynamic processes complex and difficult to study. As a result, some data pre-processing is necessary to reduce the complexity of data trajectories for further analysis. Here, we use the STaSI approach, combined with a HMM,[20] to achieve this task. STaSI has proven successful in practice to identify the number and values of states in a single data trajectory.[30] We extended this method to each trajectory of the entire data set, applying the student's $t$ test to detect FRET trajectories' step transitions and break them into segments. These segments with different lengths and from different trajectories are grouped up recursively and form a set of grouping strategies by the proximity principle, in which segments with similar FRET efficiencies from different trajectories are more likely to be merged. We calculate a synthetic objective function for every possible number of states to estimate each grouping strategy's MDL value by

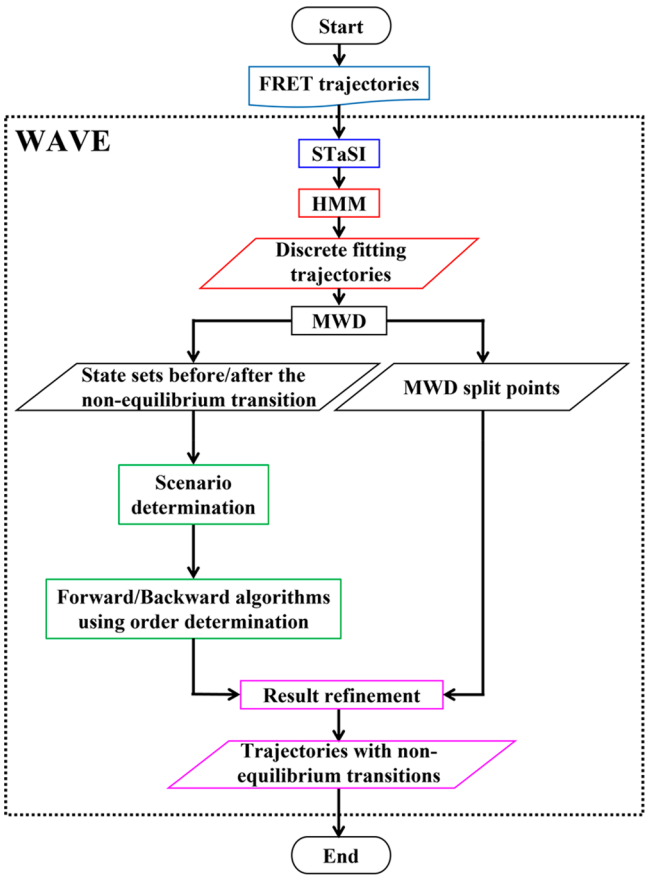



**Figure 2.** Workflow of WAVE. The input data set is first discretized by the STaSI-HMM combined method and is converted into a collection of fitting trajectories. We then conduct a maximum Wasserstein distance (MWD) analysis on each fitting trajectory to identify the state compositions before and after the non-equilibrium transition. For those trajectories that are identified as having undergone non-equilibrium transitions in the hypothesis tests of MWD analysis, we take the point where MWD is attained as a preliminary estimate of the non-equilibrium transition position. We determine the corresponding scenario by comparing state sets before and after the non-equilibrium transition and thus decide on the order of applying the forward/backward algorithm. In the last step, the forward and backward algorithms, based on the minimum description length principle, refine the MWD estimate. A collection of trajectories with non-equilibrium transitions is obtained.

$$\mathrm{MDL}_{\mathrm{total}} = \sum_{i=1}^{N_t} (\gamma F_i + G_i) \tag{1}$$

where $N_t$ is the number of FRET trajectories and $F_i$ and $G_i$ are the goodness of fit and the complexity of the fitting mode corresponding to the $i$th FRET trajectory, respectively. $\gamma$ is a hyper-parameter that controls the ratio between the goodness of fit and the cost of model, and a higher $\gamma$ leads to a model fit that favors more complex models. Note that it is always 1 in the STaSI step. We believe that the optimal number of states corresponding to the global minimum MDL value provides an accurate representation of the biomolecular dynamic processes. Each FRET state, along with its respective FRET efficiency, characterizes a unique conformational state within the molecular system. Precise identification of these quantities is crucial for properly detecting non-equilibrium transition in the


https://doi.org/10.1021/acs.jpcb.3c02498
J. Phys. Chem. B 2023, 127, 7819−7828

given data set and offers a fundamental description of the system under specific time resolution.

Since STaSI has limitations in detecting rapid state transitions, we have combined it with the HMM proposed by B. G. Keller et al.[20] to obtain more precise fitting trajectories. Unlike traditional HMM methods, such as HaMMy[19] and ebFRET,[30] which rely on FRET efficiency−time trajectories, our approach uses a two-channel intensity−time trajectory to fit the model,[31] maximizing the amount of information contained in the FRET intensity−time trajectory while reducing the errors introduced by converting a two-channel FRET intensity−time trajectory into a single-channel FRET efficiency trajectory. The initial number of states and the FRET efficiency of each state provided by STaSI help to overcome the overfitting problem in maximum likelihood estimation in the HMM. Combining the two methods gives us an accurate set of fitting trajectories.

In the next step of MWD analysis, the aim is to dissect fitting trajectories and differentiate the state compositions of the fitting trajectories before and after the non-equilibrium transition. The Wasserstein distance, also known as earth mover's distance or optimal transport distance, is a mathematical measure of (dis-)similarity between two probability distributions.[25] There are other distances available for comparing distributions, such as the Kullback−Leibler distance, which can be used to construct the maximum Kullback−Leibler distance (MKLD) analysis. Additional details of MKLD analysis can be found in Supporting Information Section 2.3 and Figure S1, while we focus solely on the MWD analysis in this manuscript. The Wasserstein distance is defined by calculating the minimum amount of work required to transform one distribution (previous distribution) into the other (target distribution). Specifically, we calculate it as the minimum product of cost function and transfer mass function as follows:

$$\boldsymbol{T}_{\boldsymbol{m}}^{*} = [t_{m,ij}^{*}] = \arg\min\langle \boldsymbol{D}, \boldsymbol{T}_{\boldsymbol{m}}\rangle_{F} = \arg\min \sum_{i=1}^{M}\sum_{j=1}^{N} d_{ij}t_{m,ij} \tag{2}$$

$$\mathrm{WD} = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N} d_{ij}t_{m,ij}^{*}}{\sum_{i=1}^{M}\sum_{j=1}^{N} t_{m,ij}^{*}} \tag{3}$$

The cost function $\boldsymbol{D}$ and the transfer mass function $\boldsymbol{T}_{\boldsymbol{m}}$ can both be described as $M \times N$ matrixes when the previous distribution and target distribution are both discrete distributions, with the numbers of bin $M$, $N$, respectively. The elements $d_{ij}$, $t_{m,ij}$ of $\boldsymbol{D}$, $\boldsymbol{T}_{\boldsymbol{m}}$ represent the distance and the transfer mass between the $i$th bin of the previous distribution and the $j$th bin of the target distribution. The matrix $\boldsymbol{T}_{\boldsymbol{m}}^{*}$, with element $t_{m,ij}^{*}$, denotes the optimal transfer mass function that minimizes the product of $\boldsymbol{D}$ and $\boldsymbol{T}_{\boldsymbol{m}}$. In our context, the cost function models the difficulty of different state transitions of biomolecules, and it is the ratio between state distributions of fitting trajectories after and before condition change positions. The Wasserstein distance naturally takes the lengths forming the two distributions into account.

We consider every end point of segments that follow a change in conditions in each discrete fitting trajectory as a potential "change point" where the steady-state status before the change completes its transition to a second steady-state status. These end points are referred to as Wasserstein distance (WD) split points. We evaluate WDs between the state distributions before and after each split point by moving from each segment after the condition change time point until the end of the trajectory and calculate the WD at each candidate split point. The split point with the largest WD is selected, and we refer to this as the *MWD split point* and the corresponding WD as the MWD. Hypothesis testing is then performed on the obtained MWD value by comparing it to a critical threshold generated by a permutation test. In each test, we generate a total number of $N_{\mathrm{MWD}}$ simulated fitting trajectories with a trajectory length identical to that of the fitting trajectory in question. Their MWD values are sorted from smallest to largest, and the upper $\alpha$ quantile or the $N_{\mathrm{MWD}} \times (1 - \alpha)$-th simulated MWD value will be selected as the critical threshold, with $\alpha$ being the significance level. Trajectories with an MWD value above the critical threshold are identified as having undergone a non-equilibrium transition. The state distribution after the non-equilibrium transition is calculated by accumulating the segments after the MWD split point in trajectories where the non-equilibrium transition has been identified.

There are four scenarios in terms of state compositions before and after the non-equilibrium transition after MWD analysis: In scenario A, completely different state sets appear after the non-equilibrium transition, scenario B has partially overlapped state sets, and, in scenarios C and D, the state set before (after) the non-equilibrium transition as a superset of the state set after (before) the non-equilibrium transition (see Figure 1c). To refine the non-equilibrium transition position estimation in each of these cases, the forward and backward algorithms based on the MDL principle are used. The backward algorithm (see Figure S2 in the Supporting Information) moves progressively back in time from a given starting position in the trajectory and tries to utilize the state set after the non-equilibrium transition to fit the segments before the starting position as much as possible, while the forward algorithm does the same thing by moving in the opposite direction and utilizing the state set before the non-equilibrium transition. The order of the two algorithms depends on the state composition before and after the non-equilibrium transition. If only the state set after (before) the non-equilibrium transition has "non-concurrent" states, which means these states only appear in one state set, we invoke the backward (forward) algorithm first. If both state sets have "non-concurrent" states and have no overlap, which is scenario A, then the order of the algorithms makes no difference. When both state sets have some "common" states, i.e., states appearing in both state sets and "non-concurrent" states, which is scenario B, the order of the algorithms will affect the calculated result. In this case, the backward algorithm is applied first to ensure that the calculated non-equilibrium transition position is probably later than the actual position, making the finding conservative in the sense that the trajectory after the calculated non-equilibrium transition position only includes the contribution from the steady-state status after the non-equilibrium transition.

Following the order of the forward and backward algorithms we established, scenario C involves using the state set that includes all states on the FRET trajectory in the forward algorithm. Consequently, the forward algorithm will be concluded at the end of the trajectory, which fixes the starting position of the backward algorithm. The backward algorithm will begin at the end of the trajectory to locate the non-equilibrium transition position. A similar but reverse situation



https://doi.org/10.1021/acs.jpcb.3c02498
J. Phys. Chem. B 2023, 127, 7819−7828

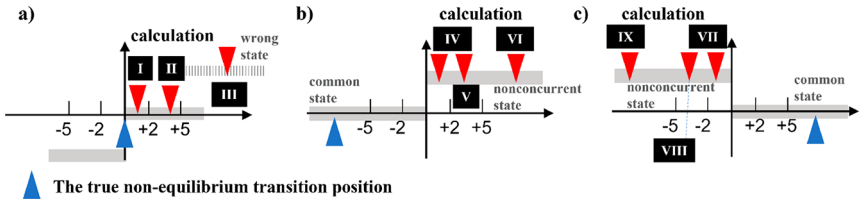



**Figure 3.** Grading criteria in outcomes in different scenarios. The blue arrows below the $x$-axis represent the true non-equilibrium transition positions, while the red arrows above the $x$-axis represent the non-equilibrium transition positions calculated by our method. The $x$-axes are in frames. Heights of the red arrows indicate the FRET efficiencies of the corresponding segments. We always have three categories of outcomes: Pass (P), Acceptable (A), and Failed (F). $\Delta_{abs}$ denotes the absolute difference between the calculated non-equilibrium transition position and the true one. $\Delta_{con}$ ($\Delta_{agg}$) denotes the number of data points that belong to "non-concurrent" states in the state set after (before) the non-equilibrium transition but in the part of the FRET trajectory before (after) the calculated non-equilibrium transition position. (a) Scenario A. Here, our calculated non-equilibrium transition positions can match the true non-equilibrium transition positions, which results in the most stringent criterion, and thus the following: P, $\Delta_{abs} \leq 2$ (I); A, $2 < \Delta_{abs} \leq 5$ (II); F, $5 < \Delta_{abs}$ (III). (b) Scenarios B and D. In these scenarios, our calculated non-equilibrium transition positions were probably later than the true positions, and thus, we set the following: P, $\Delta_{con} \leq 2$ (IV); A, $2 < \Delta_{con} \leq 5$ (V); F, $5 < \Delta_{con}$ (VI). (c) Scenario C. Here, our calculated non-equilibrium transition positions were probably earlier than the true positions, and thus we set the following: P, $\Delta_{agg} \leq 2$ (VII); A, $2 < \Delta_{agg} \leq 5$ (VIII); F, $5 < \Delta_{agg}$ (IX).

arises in scenario D, where the starting position of the forward algorithm is put at the beginning of the trajectory after the backward algorithm is concluded. Because of the inclusion relationship between state sets before and after the non-equilibrium transition in scenarios C and D, the backward/forward algorithm moves back/forward in time and includes more segments belonging to "common" states, which makes the estimated non-equilibrium transition positions in scenario C relatively aggressive, i.e., locating the non-equilibrium transition position earlier than its true position, or conservative in scenario D.

When the state sets before and after the non-equilibrium transition are identical, either the steady-state dynamic models before and after the non-equilibrium transition are practically indistinguishable in our simulation with WAVE or no non-equilibrium transition occurs. In this case, neither forward nor backward algorithms help much, and we resort back to the MWD split point obtained in the MWD analysis. Understanding this scenario is straightforward, and we will not discuss it further in the simulation section.

**Simulation Parameters for Evaluating WAVE.** We generated data sets with varying parameters to study our method's performance in the above-mentioned four scenarios. Each data set contained 100 fluorescence intensity trajectories with the same signal-to-noise ratio (SNR), kinetic rate coefficient, and trajectory length. We generated intensity–time trajectories in an equal-time-binned manner, but our method can be extended to other binning options. The default trajectory length was set to $L = 300$ frames, with the condition change position set at the 150th frame. This length corresponds to the average length of the FRET region obtained by taking a 10 ms frame rate to a photon's arrival time sequence generated by common commercial dyes under laser irradiation in point-scanned confocal microscopy experiments.

For given kinetic rate coefficients, we generate by HMM fluorescence intensity–time trajectories containing intrinsic photon-counting shot noise, which is set at a baseline of $I_{dark} = 5$ photons per frame with crosstalk coefficient $X_d = 0.1$ (a parameter dictating the amount of the signal light leakage from the donor channel to the acceptor channel). We took five levels of total intensity corresponding to different SNRs that start from 3 ($I_{tot,0} = 12.6$, 29.3, 53.6, 85.8, and 125.9 photons per frame that correspond to SNR = 3, 5, 7, 9, and 11, respectively).

To cover the four scenarios of state compositions before and after the non-equilibrium transition, we utilized five sets of kinetic rate coefficients. The seven-state system [A-1] and four-state system [A-2] correspond to scenario A, while the five-state system [B], four-state system [C], and four-state system [D] correspond to scenarios B, C, and D, respectively. The details of these state systems can be found in Supporting Information Section 3. Intermolecular heterogeneity was assumed to be absent in our simulations. For each specific combination of parameters, five data sets were generated.

We set the hyper-parameter of the forward and backward algorithms as follows: $\gamma$ was set to 2 in the backward algorithm and to 4 in the forward algorithm. The level of significance, $\alpha$, i.e., the probability of making type-I errors, for the hypothesis testing in MWD analysis was set at 3%.

State crosstalk occurs when "non-concurrent" states of state set before the non-equilibrium transition appear in fitting trajectories after the non-equilibrium transition positions, or the other way around. This phenomenon arises due to observational noises in the FRET trajectory and brings more difficulty into locating the non-equilibrium transition position. Even if we increase the SNR of the FRET trajectory from 3 to 11, state crosstalk remains prevalent. In our approach, we eliminate any such states from the state set before or after the non-equilibrium transition if their proportion increased by more than 10 times after or before the non-equilibrium transition.

We designed four tests to evaluate the performance of our method. First, using data sets produced by four kinetic rate coefficient sets ([A-1] [B] [C] [D]) with varying SNRs (SNR = 3, 5, 7, 9, and 11), we assessed the accuracy of our method in identifying the correct number of states and finding the non-equilibrium transition positions. In the second test, we used data sets generated by the kinetic rate coefficient set [A-2] to evaluate the effect of trajectory length ($L = 200$, 300, 500, and 700 frames, with the condition change position located at the 100th frame). We searched for the application range of our method by varying the length proportion of the trajectory after the non-equilibrium transition position ($P_L = 40$, 20, 10, and 5%, where $P_L$ is the length proportion of the trajectory after the non-equilibrium transition position) and the SNR (SNR = 5 and 9). In the third test, we used the data sets corresponding to


https://doi.org/10.1021/acs.jpcb.3c02498
J. Phys. Chem. B 2023, 127, 7819−7828

kinetic rate coefficient sets [C] and [D] to study the influence of the hyper-parameter $\gamma$ on the calculation (see Figures S5 and S6 in the Supporting Information). In the fourth test, FRET trajectories generated by the kinetic rate coefficient set [C] with a fixed non-equilibrium transition position ($P_L$ = 40%) and SNR (SNR = 9) were mixed with those of the same SNR but without non-equilibrium transition ($P_L$ = 0%) to form data sets with varying length proportions of the trajectories after the non-equilibrium transition positions. Here, we varied the significance level $\alpha$ to study its effect on the calculated results (see Figure S7 in the Supporting Information).

**Evaluation of Method Accuracy.** To quantify the accuracy of our method in finding the non-equilibrium transition positions, we compared the calculated output to the ground truth (GT). We have different categories of outcomes that consist of Pass (P), Acceptable (A), and Failed (F). The definitions of these three categories vary depending on the scenario, as shown in Figure 3.

## III. RESULTS AND DISCUSSION

**Accuracy of State Identification and Non-Equilibrium Transition Position Localization.** Our method utilized STaSI to pre-process the FRET trajectories and extract the optimal number of FRET states along with their corresponding FRET efficiencies. Subsequently, by utilizing estimates obtained from STaSI, the HMM can effectively identify state transitions in FRET trajectories to obtain the most precise fitting trajectories. Although HMM is originally designed for analyzing data sets with only one steady-state status, it is still feasible to analyze data sets containing non-equilibrium transitions and contributions from two steady-state statuses, since two assumptions of our method preserve the detailed balance of the transition matrix. However, the kinetic heterogeneity of the "common" states present in the state sets before and after non-equilibrium transition cannot be resolved by STaSI.[21] This limitation means that states with the same FRET efficiencies but different dwell times will be considered as a single state in STaSI. Consequently, the transition matrix calculated by HMM will be a mixture of transition matrices from two steady-state statuses, leading to more fitting errors. Nevertheless, for data sets having high SNRs, our method can overcome the difficulty brought in by FRET states with similar FRET efficiencies, and this enables accurate fitting even if the estimated transition matrix is a mixture of two transition matrices.

We assess our method's capability to identify the correct number of states as well as to locate the non-equilibrium transition positions when correct state identification results are found (Figure 4). We employed four distinct groups of data sets, each generated by using a different set of kinetic rate coefficients. These data sets included varying SNRs and were specifically designed to correspond to one of the four scenarios. For data sets where state identification by STaSI yielded a wrong number of states, we also assessed the accuracy of locating the non-equilibrium transition positions (see Figure S3 in the Supporting Information).

The performance of our method in scenario A was evaluated using [A-1]. The results indicate that state identification becomes less accurate at lower SNRs of the data sets. This is expected since the mean standard deviation of FRET trajectories is greater than or close to the minimum interval between states' FRET efficiencies of our seven-state system at SNR = 3, 5, or 7, and the STaSI can identify a four-state

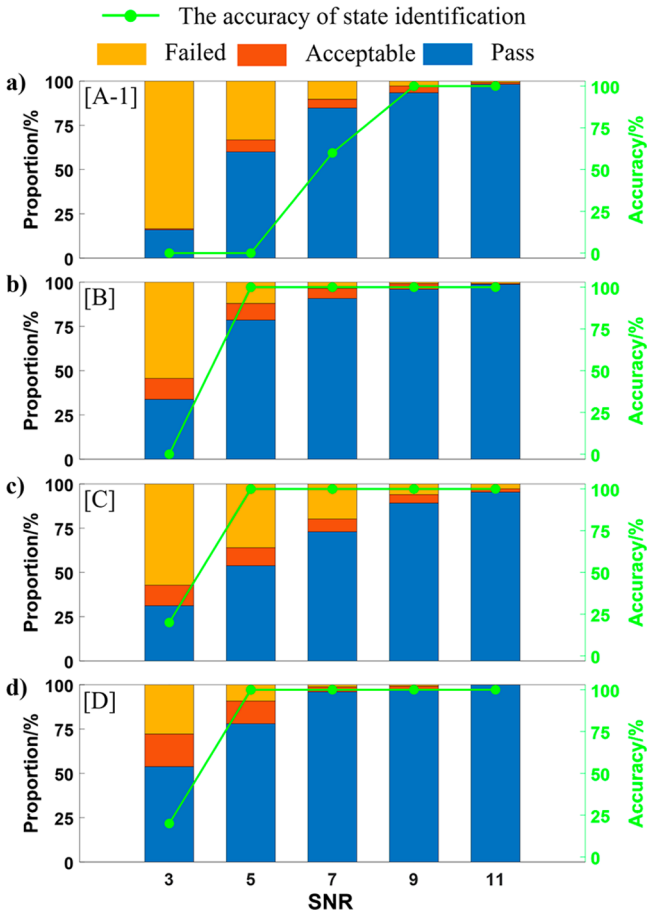



**Figure 4.** Evaluation of our method's abilities to recover the correct number of FRET states and to find the non-equilibrium transition positions when the correct number of states are found. We use data sets produced by four sets of kinetic rate coefficients, corresponding to four scenarios, with varying SNR values of 3, 5, 7, 9, and 11. Default simulation parameters were set to $L$ = 300 frames, with the condition change position located at the 150th frame. These scenarios comprised a non-overlapping seven-state system [A-1] as seen in (a), a partially overlapping five-state system [B] as shown in (b), an inclusion relationship four-state system [C] as depicted in (c), and an inclusion relationship four-state system [D] illustrated in (d). Pass, Acceptable, and Failed proportions were calculated as the average of the results from five data sets with the same experimental parameters.

system accurately[32] when the SNR is larger than or equal to 5. However, if we handpicked the seven-state results from the MDL step for subsequent calculations, our method maintained an acceptable level of accuracy in locating the non-equilibrium transition positions even at SNR = 5. This suggests that our method remains applicable to data sets with low SNRs and multiple states if the number of FRET states and corresponding FRET efficiencies are known in advance. For this reason, we recommend determining the number of states prior to conducting condition change experiments if possible. At high SNRs (SNR = 9 and 11), our method shows satisfactory performance, as the precision of fitting trajectories is extremely high. This is due to the non-overlapping compositions of states before and after the non-equilibrium transition, which results in an approximately block-diagonal transition matrix calculated by HMM and helps overcome the effect of noises on fitting trajectories.


https://doi.org/10.1021/acs.jpcb.3c02498
J. Phys. Chem. B 2023, 127, 7819−7828

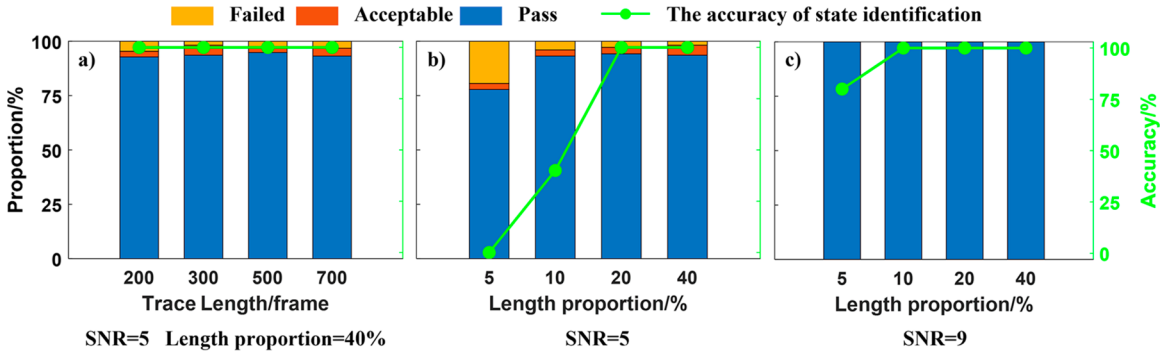



**Figure 5.** Evaluation of our method's ability to recover the correct number of FRET efficiency states and to find the non-equilibrium transition positions when the correct number of states are found. We use data sets produced by the kinetic rate coefficients set [B]. (a) Evaluation with varying trajectory lengths $L$ = 200, 300, 500, and 700 frames (with the condition change position located at the 100th frame), fixed SNR = 5, and fixed length proportion of the trajectory after the non-equilibrium transition position $P_L$ = 40%. (b) Evaluations with varying length proportions of the trajectory after the non-equilibrium transition position ($P_L$ = 40, 20, 10, and 5%), default trajectory length $L$ = 300 frames (with the condition change position set at the 150th frame), and fixed SNR of 5. (c) The same experiment as (b), only with a different SNR of 9. Again, proportions of Pass, Acceptable, and Failed are the averages of results from five data sets with the same experimental parameters.

In scenario B, the performance of our method in dealing with data sets constructed by [B] is much better than that in scenario A, probably due to the reduced model complexity. Interestingly, treating some data sets as six-state systems leads to better results (see Figure S4 in the Supporting Information). This phenomenon is to be expected because now the model identifies and recognizes the kinetic heterogeneity of the "common" states present in the state sets before and after the non-equilibrium transition. As a result, the transition matrix calculated by the HMM becomes decoupled at the "common" state(s), resulting in a closer resemblance to being block-diagonal.

Scenarios C and D are somewhat symmetric in nature, but our method performed differently here because the biomolecule's non-equilibrium transition can only take place after the condition changes. In practice, it is more difficult to identify the non-equilibrium transition in scenario C than in scenario D, since no new state appears after the non-equilibrium transition (or the condition change) in scenario C. The state compositions of FRET trajectories before and after the condition change positions are the same because the FRET trajectories after condition change positions include contribution from steady-state status before the non-equilibrium transition (which typically occurs randomly after the condition change). Only the proportion of each state varies, and such variations are not significant enough for our method to identify a non-equilibrium transition in FRET trajectory, when the length proportion of the FRET trajectory after the non-equilibrium transition position is small.

The accuracy of our method in finding the non-equilibrium transition positions in scenario C is the worst of all four scenarios when the SNR of the processed data set is greater than or equal to 5. Because in four-state system [C], the steady-state statuses before and after the non-equilibrium transition are strongly coupled, the kinetic coefficients in steady-state status before the non-equilibrium transition seriously affect the fitting process of FRET trajectories after the non-equilibrium transition positions, resulting in an increased level of state crosstalk in state distributions after the non-equilibrium transition and a decrease in the precision of the fitting. The accuracy to locate the non-equilibrium transition positions is only improved at increased SNR.

Evaluation using data sets produced by four-state system [D] demonstrates that our method has the best performance in scenario D. We attribute such impressive accuracy in locating the non-equilibrium transition positions to the weak connections between "common" states and "non-concurrent" states in the transition matrix computed by HMM.

Taken together, our results show that our method works well in all four scenarios. The accuracy of state identification depends on the complexities of the FRET trajectories and the SNRs. With decent SNRs, our method finds the true state number and gives good non-equilibrium transition positions in data sets corresponding to scenarios A, B, and D. For scenario C, the accuracy of the results is affected by the steady-state statuses in which biomolecules reside before and after the non-equilibrium transitions, and we recommend increasing the SNRs of the FRET trajectories to have satisfactory results, if the experiment allows. As briefly mentioned in scenario B, the accuracy of our method in locating the non-equilibrium transition positions can be further improved by reflecting in the model the kinetic heterogeneity of the "common" states present in the state sets before and after the non-equilibrium transition (see Figure S4 in the Supporting Information), which can come from expert knowledge. However, achieving this improvement requires a substantial amount of prior knowledge, which our method has been so far successful in avoiding.

**Influence of Trajectory Length and Detection Limit of Our Method.** The number of data points collected in a trajectory depends on two factors. The first is the photobleaching of dyes, which stochastically rely on the molecule's nature, its chemical microenvironment, and the intensity of the excitation. The second is what happens in the acquisition setup, for example, the frame rate for camera-based experiments or the bin time for single-photon-detector-based experiments. Both factors can be manipulated to balance the SNR and the trajectory length.

Both the total trajectory length and the trajectory length corresponding to each steady-state status are relevant in the performance of our method. To study their effects, we generated data sets using the non-overlapping four-state system [A-2], with variations in both the total trajectory length ($L$ = 200, 300, 500, and 700 frames, with the condition change position fixed at the 100th frame) and the proportion


https://doi.org/10.1021/acs.jpcb.3c02498
J. Phys. Chem. B 2023, 127, 7819−7828

of the trajectory after the non-equilibrium transition position ($P_L$ = 40, 20, 10, and 5% with the non-equilibrium transition position fixed in each FRET trajectory). The evaluation results are shown in Figure 5.

When the SNR, the condition change position, and the length proportion of the trajectory after the non-equilibrium transition position were fixed, changing the total trajectory length in data sets generated by four-state system [A-2] had little effect on the performance of our method. It is remarkable that, in this setup (and in scenarios [A-1], [B], and [D]), our method is sufficiently accurate even at low SNRs where STaSI starts to work,[32] which suggests that our method is at least close to being optimal among methods that depend on STaSI.

The accuracy of our method decreases when the SNR decreases and/or the length proportion of the trajectory after the non-equilibrium transition positions decreases. However, this is expected because the fitting trajectories obtained by HMM are less precise when there are fewer data points after the non-equilibrium transition positions and when the noise fluctuation is large. In this situation, merging states to which a small number of data points belong and regarding data sets as states systems of fewer states could lead to better outcomes (see Figure S3 in the Supporting Information).

**Influence of Hyper-Parameters.** Our method contains two essential hyper-parameters that users must specify. The first is $\gamma$, the ratio of the goodness of fit to the cost of the model in the MDL formula, used in STaSI and the forward and backward algorithms. The second is the significance level, $\alpha$, which determines the difficulty for a FRET trajectory to pass our hypothesis testing.

We may specify different $\gamma$ values to reflect different priorities in identifying states in the MDL procedure, where a small $\gamma$ indicates that we care more about reducing the complexity of the fitting trajectory than about reducing the difference between the fitting trajectory and the original FRET trajectory and vice versa. In scenarios A and B, the backward algorithm is applied to find an aggressive estimate of the non-equilibrium transition position with a small $\gamma$, and then we apply the forward algorithm with a large $\gamma$ to refine the estimate. This combination performs well in these two scenarios with rather lax choices of $\gamma$. On the other hand, in scenarios C and D, specifying a good $\gamma$ is important for accurately finding the non-equilibrium transition positions. To experiment with different $\gamma$'s in the backward and forward algorithms, we used the data sets generated by [C] and [D] and examined the calculated results (see Figures S5 and S6 in the Supporting Information). From the perspective of global ensemble fitting, setting $\gamma$ = 2 and 4 appears optimal in terms of performance in scenarios C and D, respectively.

The significance level $\alpha$ reflects how high in proportion false positives are acceptable to the user, where a false positive here is a trajectory that does not exhibit a non-equilibrium transition but is identified as having undergone a non-equilibrium transition. We generally set it to be small at, e.g., 10%, 5%, or 1%. It is worth pointing out that, while having a smaller $\alpha$ may reduce false positive discoveries, it increases potentially more troublesome false negatives, i.e., trajectories that exhibit non-equilibrium transitions but are not identified as having undergone non-equilibrium transitions; the vice versa is also true (see Figure S7 in the Supporting Information).

**Application of WAVE to TAR−DNA Hairpin Folding.** To showcase the practical applications of WAVE, we conducted a new physical experiment to study the transition of the TAR−DNA hairpin system.[33] The folding and unfolding process of the TAR−DNA hairpin can be influenced by the concentration of urea in the buffer solution. In the absence of urea, most of the TAR−DNA hairpins effectively remain folded. However, when exposed to a buffer solution containing 6 M urea, they transition to a partially unfolded or totally unfolded state. The aim of our experiment is to pinpoint the location where the DNA molecules undergo this non-equilibrium transition as we vary the concentration of urea from 0 to 6 M. Our experimental setup involved the use of a total internal reflection fluorescence spectroscopy (TIRF) microscope, and a flow chamber was constructed to immobilize the DNAs and enable manipulation of solution conditions. The fluorescence emitted by individual molecules was recorded by the electron multiplication charge coupled device (EMCCD), capturing frames at a rate of 30 ms per frame.

Due to the 500 ms time required for the solution environment to change in the flow chamber at maximum pump speed, we decided to collect two sets of data for analysis. The first set corresponds to the FRET trajectories of the TAR−DNA hairpin under 0 M urea conditions, providing a baseline measurement of the folded state. The second set corresponds to the FRET trajectories of the TAR−DNA hairpin collected immediately after the pump finished the job, capturing the non-equilibrium transition induced by the presence of 6 M urea concentration. Analyzing these two data sets together using WAVE, we found that the state identification results revealed the presence of five states with FRET efficiencies of 0.24, 0.44, 0.58, 0.75, and 0.97 (see Figure S8 in the Supporting Information). These findings were in good agreement with the results obtained using HaMMy in Jixin Chen's work,[33] where the five states were identified with FRET efficiencies of 0.33, 0.42, 0.57, 0.76, and 0.92. Focusing on the FRET trajectories collected from the moment the solution swapping has finished, we found that, out of 195 trajectories, the non-equilibrium transition positions for 177 of them were located at frame 0. This indicated that most TAR−DNA hairpin molecules underwent the transition from the folded state to the unfolded state during the 500 ms required for the solution change.

During our limited observation duration, we observed that 2 DNA molecules changed their conformation from folded to unfolded, while 3 remained folded and unchanged (see Figure S9 in the Supporting Information). However, upon analysis of the remaining 13 FRET trajectories, we observed a significant deviation between their fitted trajectories and the original FRET trajectories. This discrepancy led to incorrectly identified non-equilibrium transition positions for these trajectories. Overall, WAVE performed well in our experiment, demonstrating a 93% accuracy rate in locating non-equilibrium transition positions. However, the time taken for solution changing poses limitations on precisely measuring the exact position of the non-equilibrium transition. Overcoming such limitations depends more on improvements in experimental technology than on data-analytic schemes. Through this experiment, we have shown the applicability of our method to the FRET intensity−time trajectories collected using EMCCD, which typically have a time resolution of 30 ms per frame and a SNR of about 4. When dealing with such FRET trajectories, our method can capture states with dwell times of as little as approximately one frame and beyond,


https://doi.org/10.1021/acs.jpcb.3c02498
J. Phys. Chem. B 2023, 127, 7819−7828

although better time resolution and more photons per time bin are always desirable.

## IV. CONCLUSION

We present a novel method called WAVE for detecting and locating non-equilibrium transition positions in FRET trajectories that undergo condition changes during observation time. Our method's hypothesis testing part, based on the maximum Wasserstein distance analysis, is capable of identifying whether the trajectory has undergone a non-equilibrium transition. For a collection of FRET trajectories, the steady-state statuses before and after the non-equilibrium transition can be distinguished, and the non-equilibrium transition positions can be accurately identified.

To assess the efficacy of our method, we first conducted tests using simulated data sets. The results confirmed that WAVE performs well in various scenarios with appropriate SNRs, with the best performance observed when new states appear after the non-equilibrium transition. However, its detection power is affected by the SNRs, complexities of FRET trajectories, and the length proportions of the FRET trajectory after the non-equilibrium transition ($P_L$'s). The accuracies increase when the SNR and $P_L$ increase and the complexities of FRET trajectories decrease.

The applicability of our method to real experimental data has been verified through its successful application in studying TAR−DNA hairpin folding. Due to the high accuracy of our method in locating the non-equilibrium transition position, we discovered that the TAR−DNA hairpin underwent a non-equilibrium transition from the folded state to the unfolded state immediately upon exposure to a buffer solution containing 6 M urea concentration. This transition occurred in a shorter time period compared to the 500 ms required for the solution change, so not many transition times were captured.

Overall, WAVE provides a reliable decoupling tool for FRET data obtained from real-time in situ observations under variable conditions, and we believe it would benefit studies focusing on complex and dynamic biological systems.

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpcb.3c02498.

List of abbreviations and variables; additional information on our method, which includes state number determination, trajectory fitting, Wasserstein distance calculation, hypothesis testing, Kullback−Leibler distance calculation, and details of the forward/backward algorithms; five kinetic rate coefficient sets used in our simulation; figures corresponding to the calculated results that are mentioned in this manuscript, including the comparison of MKLD and MWD analysis, the flowchart of the backward algorithm, the accuracy of our method in finding the non-equilibrium transition positions when the kinetic heterogeneity of the "common" states was ignored or considered, and when state recognition yielded an incorrect number of states, the influence of the hyper-parameter $\gamma$ in scenarios C and D, the influence of significance level $\alpha$ on the calculated results, and the FRET trajectories as well as the FRET distributions obtained from our TAR−DNA hairpin experiment. We have provided the detailed and ready-to-use implementation of WAVE in the GitHub repository with the link https://github.com/gaofengnan/WAVE. (PDF)

## ■ AUTHOR INFORMATION


### Corresponding Authors

**Fengnan Gao** − *School of Mathematics and Statistics, University College Dublin, Dublin 4, Ireland; School of Data Science, Fudan University, Shanghai 200433, China;* orcid.org/0000-0002-9300-5729; Email: fngao@fudan.edu.cn

**Yan-Wen Tan** − *State Key Laboratory of Surface Physics, Multiscale Research Institute of Complex Systems, Department of Physics, Fudan University, Shanghai 200433, China;* orcid.org/0000-0002-5522-7996; Email: ywtan@fudan.edu.cn

### Author

**Ting Chen** − *State Key Laboratory of Surface Physics, Multiscale Research Institute of Complex Systems, Department of Physics, Fudan University, Shanghai 200433, China;* orcid.org/0009-0007-4297-470X

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpcb.3c02498


### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

The work of Y.-W.T. was supported by the National Key Research and Development Program of China (No. 2022YFA1404702) and by the National Natural Science Foundation of China (grants 11274076 and 21773039), Shanghai Municipal Science and Technology Major Project (No. 2018SHZDZX01), and ZJLab. The work of F.G. was partially supported by the National Natural Science Foundations of China (grants 11701095 and 11690013).

## ■ REFERENCES


(1) Ha, T.; Ting, A. Y.; Liang, J.; Caldwell, W. B.; Deniz, A. A.; Chemla, D. S.; Schultz, P. G.; Weiss, S. Single-Molecule Fluorescence Spectroscopy of Enzyme Conformational Dynamics and Cleavage Mechanism. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96* (3), 893−898.

(2) Watkins, L. P.; Chang, H.; Yang, H. Quantitative Single-Molecule Conformational Distributions: A Case Study with Poly-(l-Proline). *J. Phys. Chem. A* **2006**, *110* (15), 5191−5203.

(3) Watkins, L. P.; Yang, H. Information Bounds and Optimal Analysis of Dynamic Single Molecule Measurements. *Biophys. J.* **2004**, *86* (6), 4015−4029.

(4) Taylor, J. N.; Makarov, D. E.; Landes, C. F. Denoising Single-Molecule FRET Trajectories with Wavelets and Bayesian Inference. *Biophys. J.* **2010**, *98* (1), 164−173.

(5) Watkins, L. P.; Yang, H. Detection of Intensity Change Points in Time-Resolved Single-Molecule Measurements. *J. Phys. Chem. B* **2005**, *109* (1), 617−628.

(6) Shuang, B.; Cooper, D.; Taylor, J. N.; Kisley, L.; Chen, J.; Wang, W.; Li, C. B.; Komatsuzaki, T.; Landes, C. F. Fast Step Transition and State Identification (STaSI) for Discrete Single-Molecule Data Analysis. *J. Phys. Chem. Lett.* **2014**, *5* (18), 3157−3161.

(7) Hanson, A. J.; Fu, P. C. W. *Applications of MDL to selected families of models*; MIT Press: 2004.

(8) Rissanen, J. A Universal Prior for Integers and Estimation by Minimum Description Length. *Ann. Stat.* **1983**, *11* (2), 416−431.

(9) Schwarz, G. Estimating the Dimension of a Model. *Ann. Stat.* **1978**, *6* (2), 461−464.
(10) Lanterman, A. D. Schwarz, Wallace, and Rissanen: Intertwining Themes in Theories of Model Selection. *Int. Stat. Rev.* **2001**, *69* (2), 185−212.
(11) Yang, H.; Xie, X. S. Statistical Approaches for Probing Single-Molecule Dynamics Photon-by-Photon. *Chem. Phys.* **2002**, *284* (1−2), 423−437.
(12) Anderson, P. W. A Mathematical Model for the Narrowing of Spectral Lines by Exchange or Motion. *J. Phys. Soc. Jpn.* **1954**, *9* (3), 316−339.
(13) Schuler, B.; Lipman, E. A.; Eaton, W. A. Probing the Free-Energy Surface for Protein Folding with Single-Molecule Fluorescence Spectroscopy. *Nature* **2002**, *419* (6908), 743−747.
(14) Hanson, J. A.; Duderstadt, K.; Watkins, L. P.; Bhattacharyya, S.; Brokaw, J.; Chu, J.-W.; Yang, H. Illuminating the Mechanistic Roles of Enzyme Conformational Dynamics. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104* (46), 18055−18060.
(15) Gopich, I. V.; Szabo, A. Decoding the Pattern of Photon Colors in Single-Molecule FRET. *J. Phys. Chem. B* **2009**, *113* (31), 10965−10973.
(16) Chung, H. S.; McHale, K.; Louis, J. M.; Eaton, W. A. Single-Molecule Fluorescence Experiments Determine Protein Folding Transition Path Times. *Science* **2012**, *335* (6071), 981−984.
(17) Gopich, I. V.; Szabo, A. Theory of the Energy Transfer Efficiency and Fluorescence Lifetime Distribution in Single-Molecule FRET. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109* (20), 7747−7752.
(18) Andrec, M.; Levy, R. M.; Talaga, D. S. Direct Determination of Kinetic Rates from Single-Molecule Photon Arrival Trajectories Using Hidden Markov Models. *J. Phys. Chem. A* **2003**, *107* (38), 7454−7464.
(19) McKinney, S. A.; Joo, C.; Ha, T. Analysis of Single-Molecule FRET Trajectories Using Hidden Markov Modeling. *Biophys. J.* **2006**, *91* (5), 1941−1951.
(20) Keller, B. G.; Kobitski, A.; Jäschke, A.; Nienhaus, G. U.; Noé, F. Complex RNA Folding Kinetics Revealed by Single-Molecule FRET and Hidden Markov Models. *J. Am. Chem. Soc.* **2014**, *136* (12), 4534−4543.
(21) Götz, M.; Barth, A.; Bohr, S. S.-R.; Börner, R.; Chen, J.; Cordes, T.; Erie, D. A.; Gebhardt, C.; Hadzic, M. C. A. S.; Hamilton, G. L.; Hatzakis, N. S.; Hugel, T.; Kisley, L.; Lamb, D. C.; De Lannoy, C.; Mahn, C.; Dunukara, D.; De Ridder, D.; Sanabria, H.; Schimpf, J.; Seidel, C. A. M.; Sigel, R. K. O.; Sletfjerding, M. B.; Thomsen, J.; Vollmar, L.; Wanninger, S.; Weninger, K. R.; Xu, P.; Schmid, S. A Blind Benchmark of Analysis Tools to Infer Kinetic Rate Constants from Single-Molecule FRET Trajectories. *Nat. Commun.* **2022**, *13* (1), 5402.
(22) Poddar, H.; Heyes, D. J.; Schirò, G.; Weik, M.; Leys, D.; Scrutton, N. S. A Guide to Time-Resolved Structural Analysis of Light-Activated Proteins. *FEBS J.* **2022**, *289* (3), 576−595.
(23) Li, P.; Cheng, H.; Kumar, V.; Lupala, C. S.; Li, X.; Shi, Y.; Ma, C.; Joo, K.; Lee, J.; Liu, H.; Tan, Y.-W. Direct Experimental Observation of Blue-Light-Induced Conformational Change and Intermolecular Interactions of Cryptochrome. *Commun. Biol.* **2022**, *5* (1), 1103.
(24) Joo, C.; McKinney, S. A.; Lilley, D. M. J; Ha, T. Exploring Rare Conformational Species and Ionic Effects in DNA Holliday Junctions Using Single-Molecule Spectroscopy. *J. Mol. Biol.* **2004**, *341* (3), 739−751.
(25) Zhao, R.; Marshall, M.; Alemán, E. A.; Lamichhane, R.; Feig, A.; Rueda, D. Laser-Assisted Single-Molecule Refolding (LASR). *Biophys. J.* **2010**, *99* (6), 1925−1931.
(26) Shen, C.; Wei, C.; Li, J.; Zhang, X.; Wu, Y. Integrated Single-Molecule Long-Read Sequencing and Illumina Sequencing Reveal the Resistance Mechanism of Psathyrostachys Huashanica in Response to Barley Yellow Dwarf Virus-GAV. *Phytopathol. Res.* **2020**, *2* (1), 19.
(27) Jonas, K. C.; Fanelli, F.; Huhtaniemi, I. T.; Hanyaloglu, A. C. Single Molecule Analysis of Functionally Asymmetric G Protein-Coupled Receptor (GPCR) Oligomers Reveals Diverse Spatial and Structural Assemblies *♦. *J. Biol. Chem.* **2015**, *290* (7), 3875−3892.
(28) Ruggeri, F.; Zosel, F.; Mutter, N.; Różycka, M.; Wojtas, M.; Ożyhar, A.; Schuler, B.; Krishnan, M. Single-Molecule Electrometry. *Nat. Nanotechnol.* **2017**, *12* (5), 488−495.
(29) Hon, J.; Gonzalez, R. L. Bayesian-Estimated Hierarchical HMMs Enable Robust Analysis of Single-Molecule Kinetic Heterogeneity. *Biophys. J.* **2019**, *116* (10), 1790−1802.
(30) van de Meent, J.-W.; Bronson, J. E.; Wiggins, C. H.; Gonzalez, R. L. Empirical Bayes Methods Enable Advanced Population-Level Analyses of Single-Molecule FRET Experiments. *Biophys. J.* **2014**, *106* (6), 1327−1337.
(31) Schmid, S.; Götz, M.; Hugel, T. Single-Molecule Analysis beyond Dwell Times: Demonstration and Assessment in and out of Equilibrium. *Biophys. J.* **2016**, *111* (7), 1375−1384.
(32) Hadzic, M. C. A. S.; Börner, R.; König, S. L. B.; Kowerko, D.; Sigel, R. K. O. Reliable State Identification and State Transition Detection in Fluorescence Intensity-Based Single-Molecule Förster Resonance Energy-Transfer Data. *J. Phys. Chem. B* **2018**, *122* (23), 6134−6147.
(33) Chen, J.; Poddar, N. K.; Tauzin, L. J.; Cooper, D.; Kolomeisky, A. B.; Landes, C. F. Single-Molecule FRET Studies of HIV TAR−DNA Hairpin Unfolding Dynamics. *J. Phys. Chem. B* **2014**, *118* (42), 12130−12139.




https://doi.org/10.1021/acs.jpcb.3c02498
J. Phys. Chem. B 2023, 127, 7819−7828