



Linnæus University

Sweden

Master's Thesis

**Study of Automated Step Detection
Methods and Dwell Time Analysis of
Single-Molecule Data**

Case of Study: The Turnover of Single Fluorescent ATP on The
Active Site of Myosin



Author: Lathifa Nur Ramdhania
Supervisor: Marko Ušaj, Alf Månsson
Examiner: Carlo Canali
Term: HT22
Subject: Physicc
Level: Advanced
Course code: 4FY52E



Abstract

Single-molecule methods were developed with the objective to obtain kinetic information of biomolecular processes from a single biomolecule. This objective can be achieved by analyzing the binding events in recordings (“trajectories”) of single molecules (e.g., using fluorescence); the so-called dwell times. This study used two automated methods to analyze experimental single-molecule data. First, a method called *AutoStepfinder* is an unsupervised classification and idealization of single-molecule trajectories to find kinetic states. The method iteratively fits steps at locations that yield the biggest reduction in the variance. Second, the variational Bayesian inference method, coupled with a hidden Markov model, vbFRET, was also used. This method applied variational Bayesian approach to estimate the posterior parameters and numbers of hidden state values. We apply those two methods to analyze the single molecule ATPase experiments obtained by total internal reflection fluorescence (TIRF) microscopy on immobilized myosin motor enzymes attached to a coverslip surface. The dwell time analysis results obtained using the *AutoStepfinder* and vbFRET methods were compared to manual analysis in order to elucidate how these methods can facilitate automated analysis of single-molecule binding experiments.



Acknowledgement

Alhamdulillah, all praises to Allah, the Lord of the Universe.

First, I would like to thank my supervisors, Marko Ušaj and Alf Månsson, who have been very kind to give me the opportunity to work on the thesis in your research group. Thank you for all the feedback, guidance, and patience during the process though I am still new to the field.

I also want to express my gratitude to Carlo Canali for his countless help and guidance from the beginning of my study phase until I pursue this Master's thesis.

Finally, to my biggest support system, my husband, thank you for the unlimited support and motivation through the hard days. Along with our daughter, I am very grateful to go through this learning process with you.



Contents

1	Introduction	1
1.1	Single-Molecule Methods.....	1
1.2	Single-Molecule Data Analysis.....	2
1.3	Aims of the Thesis	4
2	Background	5
2.1	Single Molecule Fluorescence Microscopy.....	5
2.2	Total Internal Reflection Fluorescence Microscopy	6
2.3	Analysis of Single Molecule Trajectories	7
3	Methods	10
3.1	Variational Bayesian method	10
3.1.1	Hidden Markov Model	10
3.1.2	Bayes' theorem.....	12
3.1.3	Bayesian inference	12
3.1.4	Variational method.....	13
3.2	<i>AutoStepfinder</i>	16
3.2.1	Basic principle of the algorithm.....	16
3.2.2	Counter fit	16
3.2.3	Dual-pass strategy	17
4	Methods Implementation	20
4.1	Manual Analysis.....	20
4.2	Analysis using <i>AutoStepfinder</i>	20
4.2.1	Software implementation	20
4.2.2	Dwell time data extraction for analysis.....	24
4.3	Analysis using vbFRET	25
4.3.1	Software implementation	25
4.3.2	Dwell time data extraction for analysis.....	27
5	Results and Discussion	30
	Conclusions	36
	Future Work	36
	References	37



1 Introduction

Conventionally, ensembles of molecules have been used in studies to understand some essential molecular properties, particularly the average state of the system. To reveal molecular properties using ensembles, we need many molecules and then average the measured variables of the entire molecular population. Hence, some units usually used in physical chemistry depict a large ensemble of molecules, usually Avogadro's number (mol). For instance, the gas constant unit, R , is equal to Avogadro's number times the Boltzmann constant (k_B). Similarly, the energy per mole unit, RT , is normally given in kilojoule (or kilocalories) per mole [1]. In an experiment where the average ensemble method principle is used, such as fluorescence spectroscopy, we measure the average properties of many molecules even if we only use a sample of a few nmol of e.g., a protein [2]. A volumetrically small sample contains a huge number of individual molecules, e.g., in a mole of a substance, there is $\sim 6 \times 10^{23}$ (Avogadro's constant) molecules.

Despite the benefits of the ensemble method, we cannot know the behavior of individual biomolecules (e.g., protein or enzyme). This behavior can be important since individual molecules may behave differently over time [3]. The single-molecule methods were then introduced to obtain more details about the individual molecule's characteristics and mechanisms. The earliest research that managed to detect a single molecule inside a solid crystal was done by Moerner and Kador in 1989 [4], followed by Orrit and Bernard in 1990 [5], both seen as breakthroughs in this field.

Single-molecule research requires and allows new ways of interdisciplinary studies, including biology, chemistry, and physics, which have developed steadily in recent decades. The research development includes how to improve the single-molecule experiments methodologies and how to use these methodologies to answer important questions [6]. Using the single-molecule method, we can uncover new properties of proteins initially limited by ensemble experiments. Single-molecule methods have recently been applied to living cells [7]–[9], living adult animals [10]–[12], and plant science [13], [14].

1.1 Single-Molecule Methods

Some frequently used methods in single-molecule experiments are force-based and fluorescence-based [15]. Fluorescence-based methods allow high-resolution imaging of single molecules and give information about molecular properties [15], [16]. These methods have high sensitivity and selectivity, which is needed to image labeled individual molecules [17]. Single-molecule detection using the fluorescence method is accomplished by repeating excitation of the desired fluorophore, then detecting and analyzing the emitted fluorescence photons [6].

Most fluorescence-based single-molecule experiments are performed by total internal reflection fluorescence (TIRF) microscopy. The total internal reflection of incoming excitation light from a glass-water interface, such as a glass coverslip and a physiological buffer consisting of mostly water, is the basis for TIRF microscopy [17]. The first imaging of individual labeled myosin molecules using TIRF

microscopy was described by Funatsu et al. in 1995 [18]. They monitored the individual ATP turnover by a single myosin in real-time, utilizing refined epifluorescence and TIRF microscopy. This experiment resulted in high-rate imaging with a reduced background level of more than 2.000-fold lower than in conventional epifluorescence microscopy. Later, many studies that followed contributed to the emergence of single-molecule fluorescence-based experiments [19]–[22].

1.2 Single-Molecule Data Analysis

An appropriate analysis of single-molecule data is required to extract any meaningful information. Unfortunately, analyzing single-molecule data is challenging since it may be difficult to identify the signal from the noise. There are several sources of noise that influence the quality of collected data. One primary source of noise is caused by Brownian motion [23]. Brownian motion is a random motion of tiny particles due to their independent collisions with the thermally excited fluid molecules [24]. Individual molecules whose length range is very small (1\AA – $100\mu\text{m}$ [1]) exhibit Brownian motion in the experiment [23].

The observation of single colloids, e.g., a $2\mu\text{m}$ latex particle, in water with optical microscopy shows Brownian motion (see Figure 1). These motions are usually observed with different timescales and amplitudes. In particular, the same type of motion is also observed for colloids of 1 nm diameter. Such length range also corresponds to the size of single biomolecules; DNA, RNA, and proteins should also exhibit Brownian motion.



Figure 1. Brownian motion of a latex particle, $2\mu\text{m}$ in diameter; the time between images is 4 s ; the movement is in 3-dimensions. Panel 4 and 7-10 show that the bead does not remain in the same focal plane. Picture was taken from [23].

Besides Brownian motion, there are other factors which may also reduce the quality of the single-molecule data. Unwanted photophysical processes, such as reversible blinking and irreversible photo-bleaching, are frequently observed in experiments [25], [26]. Both phenomena can cause an error in the evaluation of single-molecule data [27]. It is also reported that the appearance of unknown fluorescent objects that may contaminate glass surfaces can also decrease the data quality by introducing a false-positive fluorescence dwell-time [28].

In the context of the data types, one commonly acquired single-molecule data is the time trace (or trajectory), as shown in Figure 2. Manual inspection could be employed to select which parts of such a trajectory are chosen as regions of interest. Usually, user considers the trajectory above the specified limit (or “threshold”) to be collected and analyzed further (see dashed line in Figure 2). However, manual analysis is subjected to user bias. Another limitation is the large amount of data produced by the high-speed camera, making manual analysis extremely time-consuming and prone to errors [29].

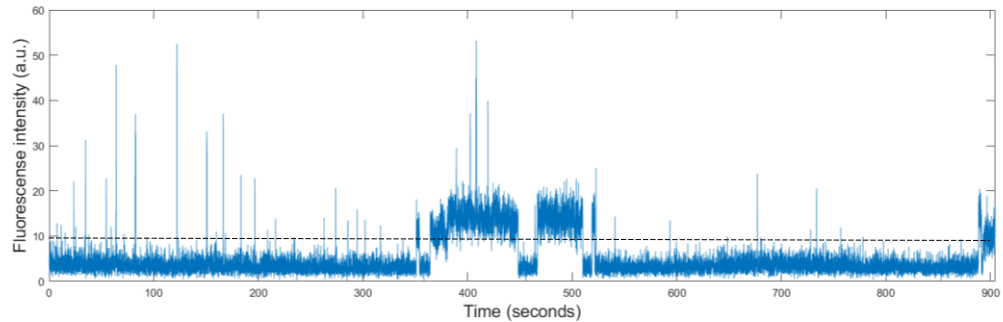


Figure 2. An example of single-molecule data: a time trace of fluorescent ATP molecule interactions with immobilized single myosin molecule. In manual analysis, a trained user determines the “threshold” value. The “threshold” value is indicated with a dashed black line. During the data collection, only the trajectory above the threshold is considered.

There are many studies done in recent years to overcome the challenges in single-molecule data analysis, particularly for trajectories with complex behavior. Simple statistical methods such as thresholding (which is different from setting the ‘threshold’ value in manual analysis) or pairwise distribution analysis manage to analyze clear transitions of state [30]. However, if we wish to uncover a trajectory close to the noise level or the data containing several steps of variable sizes, this method is insufficient [31]. Another method is model-based step detection, which generates various candidate models with steps at different locations, for example, information criterion (IC) [32]. The drawback of using model-based methods is that each model may only fit a trajectory with a specific noise source. Thus, it is not trivial to make a model-based algorithm applicable to various trajectories.

Currently, one of the most used methods is the so-called hidden Markov model (HMM). An advanced statistical method initially used for speech recognition, the HMM has become a reliable method for analyzing single-molecule data due to its ability to discover discrete states within noisy time series and to determine the most likely path between these states [33], [34]. Some researchers have successfully developed software to do single-molecule analysis by applying the HMM, such as McKinney et. al. [35], Bronson et. al. [36], and Meent et al. [37].

A recent study done by Loeff et al. developed another type of software for unsupervised identification and idealization of single-molecule trajectories [31]. This approach enables step fitting of single-molecule trajectories without any preliminary information about the step properties, such as their size, location, and noise impact on the data. The software has advantages of requiring minimal input from the user, a user-friendly interface, and application to any step-like signal.



1.3 Aims of the Thesis

This study aims to develop an automated method to analyze single-molecule data. In particular, this report covers step detection and dwell time analysis of single molecule fluorescence trajectories. The analysis is implemented using several automated methods. The results will be compared with manual analysis to demonstrate how the automated methods can facilitate analysis of single-molecule experimental data.

The report outline is as follows. The background of single-molecule fluorescence-based method and the basics of single-molecule data analysis are provided in chapter 2. Chapter 3 describes an overview of the automated methods used in analysis. In chapter 4, the implementation of the automated methods is described. In chapter 5, the results of the automated analysis of single-molecule data are reported together with discussion.



2 Background

2.1 Single Molecule Fluorescence Microscopy

As mentioned in the Introduction section above, single-molecule experiments are often of two primary types, fluorescence-based or force-based. Fluorescence-based method contrasts with force/displacement experiments, where optical tweezers, magnetic tweezers, and atomic force microscopy are the most common examples. This project covers only fluorescence-based approaches in line with the obtained and available experimental data.

Fluorescence appears when a photon of light is absorbed by a molecule (e.g., a fluorophore) and then re-emitted as a photon with a longer wavelength [38]. Measuring a few photons emitted by a single molecule at a wavelength different from the exciting light is much more convenient than measuring the absorption directly, requiring many incident photons. Hence, it is possible to use the fluorescence-based method to detect individual fluorescent molecules.

In a single-molecule fluorescence-based experiment, the user selects fluorescent spots (also called “hotspots”) on a time projection image obtained from acquired video. The integrated intensity of each fluorescence spot is calculated frame by frame. The time-based changes in fluorescence in a hotspot location are referenced herein as a fluorescence trajectory, stored as time-series data. Most commonly, the trajectory from fluorescence-based single-molecule experiments visualizes the multiple cycles of binding and dissociation of a fluorescent ligand, such as nucleotide (e.g. Cy3-ATP and Cy3-ADP) with a single protein molecule (e.g. heavy meromyosin¹ or myosin subfragment 1 (S1)²) which can also be fluorescently labeled [18]. For example, Funatsu with coworkers firstly visualized fluorescently labelled S1 molecule bound to the surface (Figure 3A) on which fluorescent nucleotide bindings and dissociations

¹ Myosins are motor proteins that have an essential role in muscle contraction, usually associated with actin filaments. The type of myosin present in muscle (myosin II) consists of two heavy chains and two pairs of light chains [51]. Heavy meromyosin is a two-headed soluble motor fragment with the enzymatic activities of myosin [52].

² Subfragment 1 (S1) is the isolated head fragment that contains the ATPase and actin-binding activities of myosin [52].

were recorded (Figure 3B) while data from one such S1 molecule was extracted (Figure 3C).

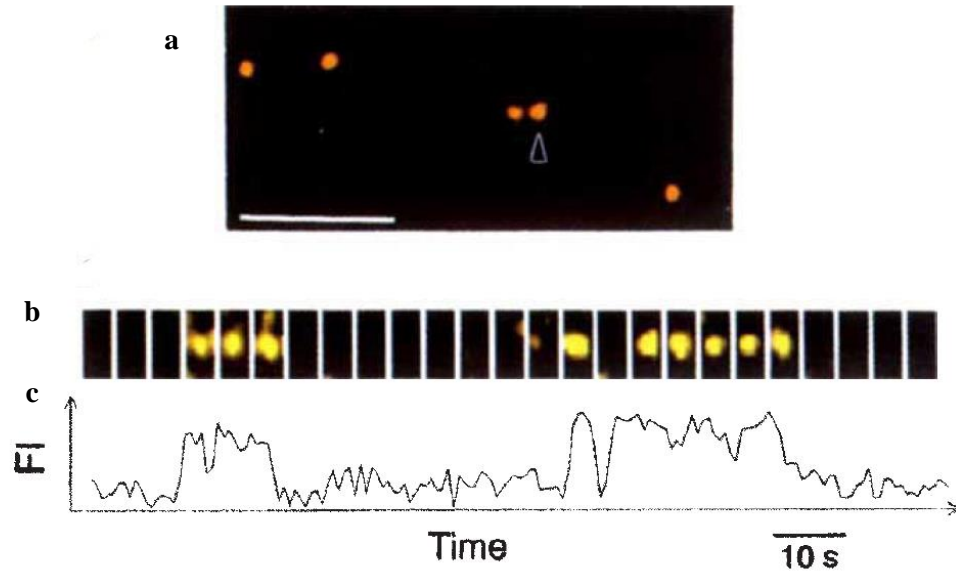


Figure 3 Visualization of individual ATP turnovers by single myosin subfragment 1 (S1) molecules. a) Fluorescence micrograph of individual Cy-5 labelled S1 molecules bound to the surface. b) Typical time-lapse video of Cy-3 nucleotide fluorescence (ATP and ADP) coming in and out of focus by associating and dissociating with a single S-1 molecule indicated by the arrowhead in A. c) Time series data extracted from video presented in B with corresponding fluorescence intensity (FI). Images were taken from [18].

2.2 Total Internal Reflection Fluorescence Microscopy

The typically used microscopy approach to capture a fluorescence-based single-molecule signal is total internal reflection fluorescence (TIRF) microscopy. In TIRF microscopy, the incident laser beam is totally reflected at the interface of two media of different refractive index (usually quartz and water) while an intense field of electromagnetic energy, called an evanescent field, is created [3] (See Figure 4A). The evanescent field is an electromagnetic field with the same frequency as the incident light, whose intensity decays exponentially with distance from the surface. Fluorophore excitations are obtained from this field only at a distance extremely close to a solid surface (within ~100 nm) [39]. This type of excitation produces a low background noise image with almost no out-of-focus fluorescence and prevents exposure of cells or macromolecules to excitation light at any other focal planes in the samples [40].

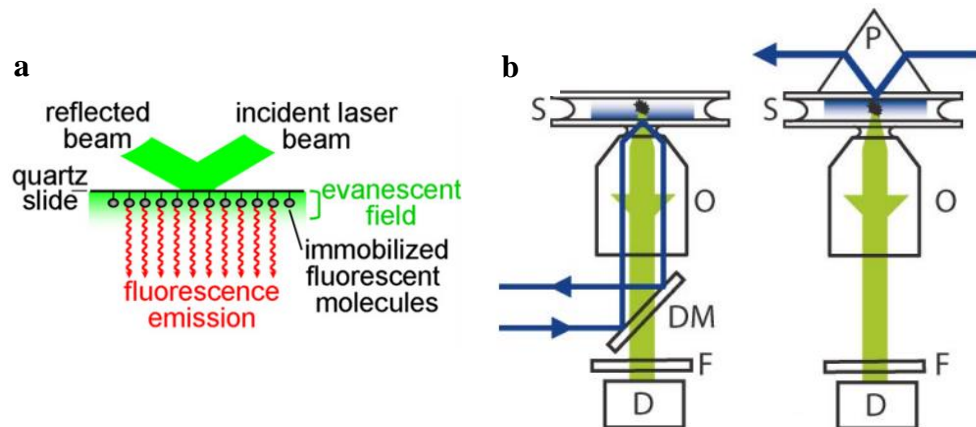


Figure 4. a) Principle of total internal reflection fluorescence microscopy. b) Schematic diagram of experimental setup of objective type total internal reflection fluorescence microscopy. c). Schematic drawing of prism-based total internal reflection fluorescence microscopy. P, prism; S, sample; O, microscopy objective; DM, dichroic mirror; D, detector. Image a) was taken from Zlatanova et al. 2006 [3], while images b) and c) are from Peterman et al. 2004 [41].

Considering the described mode of fluorophore excitations, TIRF microscopy at single-molecule level enables the real-time visualization of the molecular interactions between the surface-tethered species and freely diffusing molecules labeled with a fluorophore [27]. TIRF microscopy is well suited for investigating immobilized and labelled biomolecules because it provides a minimal background and high throughput measurements of conformational dynamics, inter- and intra-molecular interactions, and the stoichiometry of multicomponent complexes [15].

There are two types of TIRF microscopy commonly used: prism type and objective type. In the prism type (Figure 4C), total internal reflection occurs at the interface between the surface of a quartz microscopy slide and the sample (water-based imaging solution). The evanescent field is created on the opposite side of the sample to the objective lens that collects the fluorescence.

Meanwhile, in the objective-type configuration (Figure 4B), emitted fluorescence is collected from the sample on the same side as the excitation light is delivered. The evanescent field is generated in the sample at the coverslip sample interface. Compared with the prism method, the objective-based method is more convenient as the specimen is easily accessible, and the angle of incidence of the laser light can be changed easily [42]. The drawback is that it requires special TIRF grade objective, which is more expensive than a regular objective, which can be used in prism type TIRF microscopy. Data used in this project were generated using objective type TIRF microscopy.

2.3 Analysis of Single Molecule Trajectories

Analysis of the trajectories allows us to identify the kinetic rate constants and interaction affinities, helps to reveal populations of conformers with different binding or catalytic properties, and enable us to distinguish different conformations of



complex assembly [27]. For example, Amrute-Nayak with co-workers [21] have studied ATPase kinetics by myosin molecular motors using TIRF microscopy. They measured the time duration of fluorescently labeled nucleotide molecules bound to a myosin head domain, the so-called *dwell times*, as well as the pauses between high fluorescent states (waiting time). Both dwell times and waiting times yield information about ATPase kinetics and thus, in principle, allow detection of differences in ATPase kinetics among individual myosin molecules [21].

During observation of the binding events in trajectories, the dwell time (high fluorescence state or “on-time”) is associated with the “bound” state with high fluorescence signal. For example, in an experiment where a myosin molecule that is adsorbed to a silanized glass surface interacts with fluorescently labeled nucleotides (e.g., Alexa-647 nucleotide), the resulting dwell times were interpreted as the time spent by Alexa647-ATP or Alexa647-ADP bound to an immobilized myosin molecule [28]. An example of a dwell time from such an experiment is shown in Figure 5.

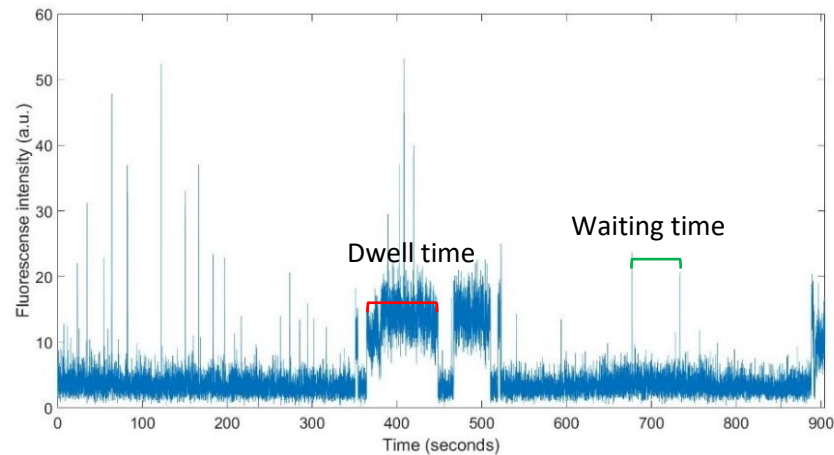


Figure 5 Representative trajectory from an experiment obtained by TIRF microscopy observing fluorescent nucleotide bindings to an immobilized myosin motor protein attached to a coverslip surface. The red-colored bracket represents one of the dwell times, while the green-colored one indicates the waiting time between two short dwell times.

For the dwell time analysis to be carried out, the first step is to create a set of selection rules to classify specific trajectories for analysis from all extracted trajectories. The selection rules, namely for the experimental data of the association of fluorescently labeled macromolecule with the surface-tethered partner, are [27]:

1. At least two true binding events exist. Binding events can be described as the events whose dwell times in more than two frames of the movie exhibit fluorescence above the minimum threshold.
2. The dwell time event should begin after the beginning of the recording and end before the recording is ended.
3. Trajectories experiencing events that caused the decay of intensity, such as focus problems during video acquisition, should be ignored across the total length of recorded video.

4. Trajectories with a very high signal intensity than the majority of signal in most trajectories should be ignored.

Observations of binding events in the trajectories alone is not enough to make a meaningful inference into the mechanism of biomolecular interactions. The overall statistical properties extracted from a large collection of trajectories reflect on macroscopic behavior of the experiment. In this case, the single molecule binding events may be presented as histograms of the observed dwell times. Two types of distributions are most commonly used: ordinary frequency distributions and cumulative frequency distributions.

The ordinary frequency distribution is generated by counting the number of dwell times in each given time duration (bin) (Figure 6A). The resulting distributions can then be compared to a theoretical probability density function that describes the process (e.g., sum of exponential functions). When normalized for the size of the data set, each bin's height defines the probability that a random dwell time falls within a particular time interval [27].

The cumulative frequency distributions reflect the sum of all occurrences of observed dwell times smaller than the given time (Figure 6B). When normalized, the cumulative frequency distribution defines the probability that a random dwell time falls below a given time value. The total number of points in the given cumulative frequency distribution equals the total number of observed dwell times. Similar as above, the cumulative frequency distribution can be fit to the sum of exponential functions to estimate the number of processes with their amplitudes and rate constants.

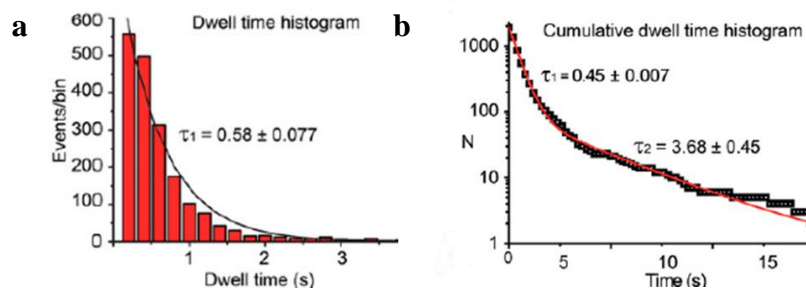


Figure 6. Dwell time distributions of fluorescent nucleotide binding to motor protein myosin-2; a) The ordinary frequency histogram of 3'Cy3-EDA-ATP dwell times recorded from 26 myosin-2 molecules is fitted with a single exponential function (black line). b) Cumulative dwell time distribution using the same data as in a. Experimental data (open squares) were fitted to a double exponential decay function. Images were taken from [21].

3 Methods

3.1 Variational Bayesian method

The variational Bayesian method was used in biophysical time series data analysis in a study by Bronson et al. [36]. The software package called vbFRET was written in MATLAB and is available as open source. vbFRET takes single molecule data, particularly fluorescence resonance energy transfer (FRET) data files, as inputs and fits idealized trajectories to them. It also returns information about the FRET states in the data: the number of states, posterior parameters, and idealized trajectories. In the following section, the basic principle of the algorithm used in vbFRET method will be explained.

3.1.1 Hidden Markov Model

Use of the hidden Markov model (HMM) is a convenient way to analyze single-molecule data because of its ability to find discrete states within noisy time series data and to reliably find the most probable path through these states [33].

In the HMM, the model is assumed to obey a Markov chain or Markov process. A Markov chain is the probabilistic model resting on the assumption that the next event in a series of observations can be predicted upon knowledge of the immediately preceding event, not depending on any knowledge of any prior events [43]. For example, in a Markov chain (see Figure 7), if the system jumps from a state S_1 to another state S_2 after a time step, the state S_2 depends only on S_1 , not any other state visited in the past. What makes a Markov process a hidden Markov model is the states that are not directly determinable because of noise and hence, *hidden* [44].

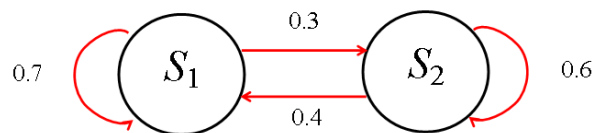


Figure 7. Diagram of a Markov chain with two states labelled S_1 and S_2 . The corresponding transition probabilities represent the probability of the Markov chain changing from one state to another state, with the direction indicated by the arrow. For example, if the Markov chain is in state S_1 , then the probability that it changes to state S_2 is 0.3, while the probability that it remains in state S_1 is 0.7.

A hidden Markov model has three parameters that need to be optimized to find the set of parameters that best describe the data. First, the transition probability matrix describes the probability of any one state changing to any other state or staying in the same state in the subsequent time step. Second, the emission probability distribution contains the probabilities of a specific signal value being emitted by each discrete hidden state. Last, the initiation probability matrix gives the probabilities of starting at each possible discrete state.

Figure 8 below is a simple example of a hidden Markov model with two different FRET states taken from Blanco et al. [43]. The probability of observing state *A* transit to state *B* in a one-time step is governed by single exponential decay independent of how long the molecule has been in state *A*. A noiseless system would generate simple two-state FRET trajectories. However, experimental noise influenced the emission probability distribution of each state (Figure 8B). This makes the process a hidden Markov process since the underlying state (the sequence of true $A \leftrightarrow B$ exchange) is hidden in the data because of noise.

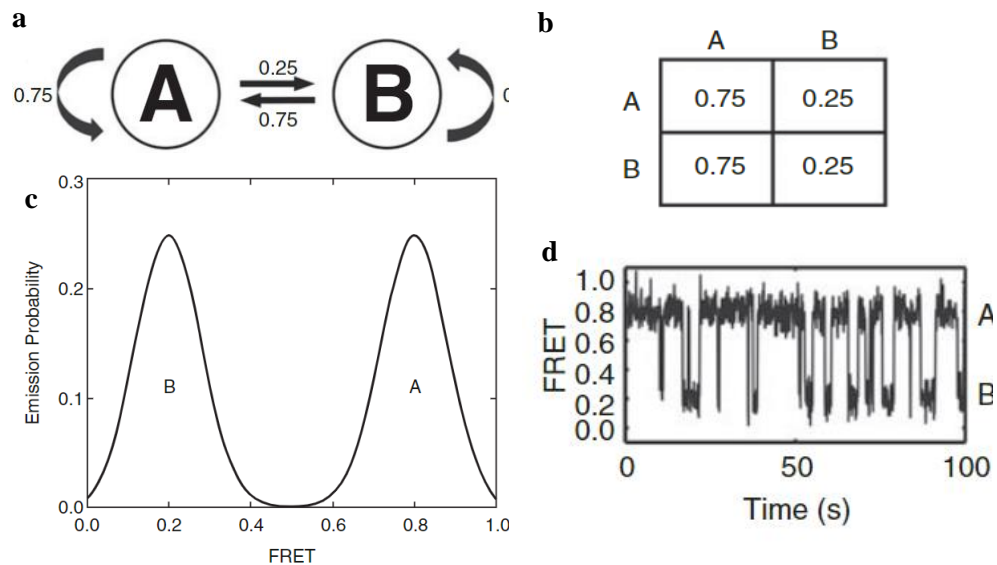


Figure 8. a) An example of a hidden Markov model with two states, *A* and *B*. b) Transition probability matrix of the hidden Markov model from *A*. c) The corresponding emission probability distribution of the two states *A* and *B* from a FRET trajectory calculated by assuming Gaussian noise distribution around the discrete mean FRET values. d) A simulated single-molecule FRET trajectory using two-state model with states *A* (FRET = 0.8) and *B* (FRET = 0.2). Images were taken from [43].

As mentioned, each hidden Markov model has a set of parameters optimized to reconstruct the underlying reality. A solution can be used to find the most probable trajectory of hidden states using the max-sum algorithm, which in the context of HMM is known as the Viterbi algorithm [45].

Using the Viterbi algorithm, we first represent the HMM as a factor graph and consider all paths through the lattice. We evaluate the probability for each path, and then select the one with the highest probability (Figure 9). We compute each path's probability by summing up products of transition and emission probabilities as we work our way forward along each path through the lattice. Consider a particular time step n and a particular state k at that time step. Many possible paths will converge on the corresponding node in the lattice diagram. However, we only need to retain the specific path that so far has the highest probability. Because there are k states at time step n , we need to keep track of k such paths, and there will be k^2 possible paths to

consider. Thus, we need to only retain k of these paths, which corresponds to the best path for each state at time $n+1$. When we reach the final time step N , we will discover which state corresponds to the overall most probable path [46].

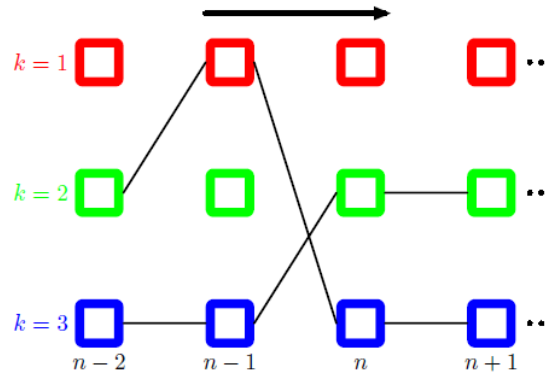


Figure 9. Diagram fragment of the hidden Markov model lattice, with k states and time step n , showing two possible paths. The Viterbi algorithm efficiently determines the most probable path by evaluating the probability for each path and then choosing the highest one. Image was taken from [46].

3.1.2 Bayes' theorem

Bayes' theorem is a simple mathematical formula used for measuring the probability of an event based on prior knowledge of conditions that might be related to the event, also known as conditional probability. Bayes' theorem is denoted as following

$$p(A|B) = \frac{p(B|A)p(A)}{p(B)} \quad (1)$$

where A and B are events and $p(B) \neq 0$.

$p(A|B)$ is conditional probability: the probability of event A occurring given that B is true. It is also called the *posterior* probability of A at given B.

$p(B|A)$ is "inverse" probability of event B given that A is true. It can also be interpreted as *likelihood* of A at given B.

$p(A)$ and $p(B)$ are the unconditional probabilities of observing A and B, respectively, without any given conditions, also known as marginal probability or *prior* probability.

3.1.3 Bayesian inference

Bayesian inference is the application of Bayes' theorem to determine the probability of a set of unknown variables considering a set of observed data. In the context of single-molecule studies, these unknown variables are a vector of parameters of biophysical interest, $\vec{\vartheta}$, and a hidden value of the state variable of interest, z , whereas the observations are an experimental data, y .

The goal of Bayesian inference is now to approximate the so-called posterior probability of model parameters, $\vec{\vartheta}$, considering an experimental data, y , and the number of states, K . Bayes' theorem states that this *posterior probability*, $p(\vec{\vartheta}|y, K)$ is



equal to the product of the likelihood and the prior distributions, normalized by the evidence, $p(y|K)$, or also can be expressed as

$$p(\vec{\vartheta}|y, K) = \frac{p(y|\vec{\vartheta}, K)p(\vec{\vartheta}|K)}{p(y|K)} \quad (2)$$

The likelihood functions

The likelihood function $p(y|\vec{\vartheta}, K)$ is the probability of the vector data y given some unknown vector parameters of interest, $\vec{\vartheta}$, and value of states, K . In other words, the likelihood is the probability of observing a particular y in the experiment, given that $\vec{\vartheta}$ comprises the ‘true’ parameters of model.

The prior distributions

The prior distribution, $p(\vec{\vartheta}|K)$, quantifies the information about the values of a set of model parameters, $\vec{\vartheta}$, for the value of K being tested. A scientist usually has prior knowledge of the natural phenomenon under research, which leads them to generate the hypothesis they are testing before conducting an experiment. After specifying a prior distribution, the likelihood functions redistribute the probability of $\vec{\vartheta}$ specified by the priors to where it is most consistent with the collected data, then yields an updated distribution called the posterior distribution [47].

The posterior distribution

The posterior distribution, $p(\vec{\vartheta}|y, K)$, can be thought of as the quantification of how the experiment updates our certainty of the initial hypothesis (i.e., the prior) in terms of the model parameters. All data analysis methods that are consistent with the scientific method striving to obtain the posterior distribution.

The evidence or marginal likelihood

The evidence, $p(y|K)$, provides the probability that the observed experimental data might have been obtained from the model being tested. It is obtained by marginalizing, which entails integrating all feasible model parameter values.

3.1.4 Variational method

Despite its benefits to the analysis of single-molecule experiments, fully implementing Bayesian inference has been complicated in practice. Specifically, calculating the evidence is analytically and numerically intractable since the observed data are conditionally dependent on an unknown or hidden state that must be inferred. These hidden variables must be marginalized over in calculating the evidence. As a result, calculation of evidence now requires a discrete sum over all states, z , in addition to the integrals over parameter values, $\vec{\vartheta}$:

$$p(y|K) = \sum_z \int d\vartheta p(y, z|\vec{\vartheta}, K)p(\vec{\vartheta}|K) \quad (3)$$

A way to address these challenges is to use a variational method, also known as Variational Bayesian (VB) method. Using this method, we would find the distribution $q(z, \vec{\vartheta})$ that best approximate the posterior, $p(z, \vec{\vartheta}|y, K)$, i.e.,



$$q_*(z, \vec{\vartheta}) = \operatorname{argmin} D_{KL}(q(z, \vec{\vartheta}) \| p(z, \vec{\vartheta} | y, K)), \quad (4)$$

where D_{KL} is the Kullback-Leibler divergence, which quantifies the dissimilarity between $q(z, \vec{\vartheta})$ and $p(z, \vec{\vartheta} | y, K)$ [36]. A simple identity relates this quantity to evidence, $p(y|K)$:

$$\log p(y|K) = -F[q(z, \vec{\vartheta})] + D_{KL}(q(z, \vec{\vartheta}) \| p(z, \vec{\vartheta} | y, K)) \geq -F[q(z, \vec{\vartheta})], \quad (5)$$

where $F[q(z, \vec{\vartheta})]$ is an analytically tractable functional:

$$F[q(z, \vec{\vartheta})] = \sum_z \int d\vartheta q(z, \vec{\vartheta}) \ln \left[\frac{p(y, \vec{\vartheta}, z | K)}{q(z, \vec{\vartheta})} \right] \quad (6)$$

The inequality in Eq. 5 results from the property $D_{KL} \geq 0$, with equality if and only if when $q(z, \vec{\vartheta}) = p(z, \vartheta | y, K)$. Eq. 5 illustrates that minimizing the functional $F[q(z, \vec{\vartheta})]$ at the same time maximizes a lower bound of the evidence and minimizes the dissimilarity between the test distribution, q , and the posterior distribution of parameters [36]. Qualitatively, the best test distribution gives not only the best estimate of the evidence, but also the best estimate of the posterior distribution of the parameters [36].

Here, the experiment is well described by the HMM to model the dynamics of an observed data, y , (e.g., fluorescence intensity, FRET ratio) that is conditionally dependent on hidden process, z [36]. Assuming z is a Markov process, the hidden states can take any one of K possible values at each time t , z_t , conditionally dependent only on its value at the previous time via transition probability matrix. Meanwhile, the observed data depend only on the current-time hidden state via the emission probability. According to the convention of the field, all transition probabilities were modeled as multinomial distributions while all emission probabilities were modeled as Gaussian distributions [36]. In this model, the parameters of interest, $\vec{\vartheta}$, consists of four types of parameters: a K -element vector, $\vec{\pi}$, where the k^{th} component, π_k , holds the probability of starting in the k^{th} state; a transition matrix, A , where a_{ij} is the probability of transitioning from state i to state j ; and two K -element vectors, $\vec{\mu}$ and $\vec{\lambda}$ where μ_k and λ_k are the mean and precision of the Gaussian distribution of the k^{th} state [36].



As in Eq. 3, the evidence follows directly from multiplying the likelihood by priors and marginalizing:

$$p(y|K) = \sum_z \int d\vec{\vartheta} p(\vec{\pi}|K) p(A|K) p(\vec{\mu}, \vec{\lambda}|K) p(z_t|\vec{\pi}, K) \times \left[\prod_{t=2}^T p(z_t|z_{t-1}, A, K) \right] \prod_{t=1}^T p(y_t|z_t, \vec{\mu}, \vec{\lambda}, K). \quad (7)$$

The $p(\vec{\pi}|K)$ and each row of $p(A|K)$ are modeled as Dirichlet distribution while each pair of μ_k and λ_k are modeled jointly as a Gaussian-gamma distribution. These distributions are the standard choice of priors for multinomial and Gaussian distributions. With the same assumptions of that $q(z, \vec{\vartheta})$ factorizes into $q(z)q(\vec{\vartheta})$, this HMM can be solved using variational Bayesian method.

To calculate the model evidence, we treat the components of $\vec{\vartheta}$ as random variables. The vector $\vec{\pi}$ and each row of A are modeled as Dirichlet distributions [36]:

$$p(\vec{\pi}) = \frac{\Gamma(\sum_{k=1}^K u_{\pi}^k)}{\prod_{k=1}^K \Gamma(u_{\pi}^k)} \prod_{k=1}^K \pi_k^{u_{\pi}^k - 1} \quad (8)$$

$$p(a_{j1}, \dots, a_{jK}) = \frac{\Gamma(\sum_{k=1}^K u_a^{jk})}{\prod_{k=1}^K \Gamma(u_a^{jk})} \prod_{k=1}^K a_{jk}^{u_a^{jk} - 1} \quad (9)$$

The probabilities for each pair of μ_k and λ_k are modeled as a Gaussian-Gamma distribution [36]:

$$p(\mu_k, \lambda_k) = \sqrt{\frac{u_{\beta}^k \lambda_k}{2\pi}} e^{-\frac{1}{2} u_{\beta}^k \lambda_k (\mu_k - u_{\mu}^k)^2} \frac{1}{\Gamma(u_{\nu}^k/2)} (2u_{\nu}^k)^{-u_{\nu}^k/2} \lambda_k^{(u_{\nu}^k/2)-1} e^{-\frac{\lambda_k}{2u_{\nu}^k}} \quad (10)$$

The terms u_{π} , u_a , u_{β} , u_{μ} , u_{ν} , and u_w are called *hyperparameters* for the probability distributions over $\vec{\vartheta}$. Hyperparameters in Bayesian statistic is a set of parameters of a prior distribution. It corresponds to probability distributions over the hidden states such that it is most probable that the hidden states are equally likely to be occupied and equally likely to be transitioned to [36].

To sum up, applying this variational method gives the best estimate of the evidence value and yields the best approximation of the posterior distribution of all the parameters. The variational Bayesian method is used in vbFRET method [36] to produce a tractable, analytical form of the posterior for an HMM used to model fluorescence resonance energy transfer (FRET) efficiency trajectories observed in single-molecule FRET (smFRET) experiments.

3.2 *AutoStepfinder*

AutoStepfinder was developed by Loeff et al. as unsupervised classification and idealization of single molecule trajectories that do not require information on the underlying states *a priori* [31]. It is a revised version of previous study by Kerssemakers and co-workers which had successfully developed a first-order approach called *Stepfinder*[48]. This software is written in MATLAB and is available as open source, including its GUI (Graphical User Interface).

3.2.1 Basic principle of the algorithm

Using the *AutoStepfinder* algorithm, we find steps with the assumption that data contain instantaneous steps of interest with variable size (Δ_i) and plateau length (N_i) that are subject to residual noise (σ_R^2) (see Figure 10). The algorithms initiate the fitting procedures by fitting one step to all data points at a location that gives the lowest value of variance (σ^2). This initial partition event generates a fit with two plateaus at a position that represents the average of the data points within the plateau. After the first fit, the plateau that exhibits a step, yielding the largest reduction of σ^2 is selected for the next partition event. The algorithm continues this process of adding a single step to one of the plateaus for each iteration, until it has performed the user-defined maximum iteration number.

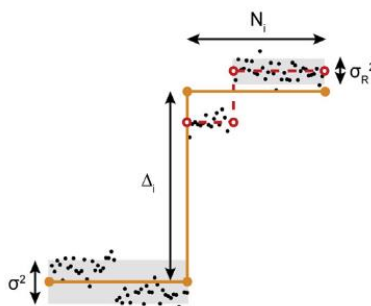


Figure 10. An example of an iterative step fit (orange line) on a single molecule trajectory (black dots)[31].

3.2.2 Counter fit

The quality of the step fits needs to be evaluated by performing an additional fit for each iteration called counter fit. To produce the counter fit, first, *AutoStepfinder* determines the next fit location (i_{next}) within each plateau. After that, the algorithm ignores the current fit and builds a new fit based on the i_{next} locations. *AutoStepfinder* will generate new plateaus with positions that represent the average of the data points within each plateau. The counter fits are all located in between the existing best-fit locations (see Figure 11).

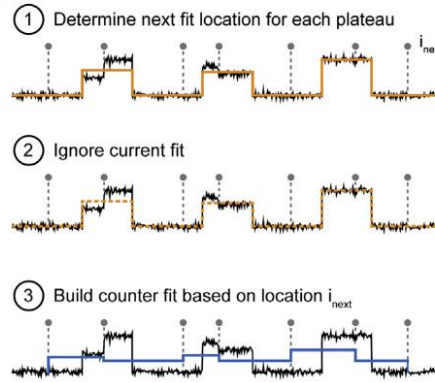


Figure 11. Steps to generating a counter fit [31].

The quality of the fit (S-score) can be quantified by taking ratio of the σ^2 from the existing fit and the counter fit, or also can be written

$$S = \frac{\sigma^2_{counter\ fit}}{\sigma^2_{existing\ fit}} \quad (11)$$

If the existing fit is at the optimal number of iterations, the variance of the existing fit approximates the residual variance in the data (σ^2). In contrast, the counter fit misses all real steps and places step location fits at random plateau positions. Thus, the fit values differ on average $\frac{1}{2}\Delta$ from the data plateaus, yielding a variance of the counter fit that reaches its maximum value ($\sim \frac{\Delta^2}{4\sigma_R^2}$). The maximum S-value (S^{\max}) can be described by $S^{\max} = 1 + P$, where P equals the maximum value of variance for the counter fit ($\frac{\Delta^2}{4\sigma_R^2}$) [31].

3.2.3 Dual-pass strategy

Sometimes the data show steps with wide range of step sizes. In particular, the trajectory may have large and small steps combined (Figure 12C). This type of data can produce S-curves with multiple peaks (S_{p1} and S_{p2}) where its maximum value S_{p2}^{\max} is lower than the global maximum value S_{p1}^{\max} (Figure 12D). To facilitate step detection across a wide variety of scales, the algorithm develops a dual-pass strategy. This procedure aims for optimal fit through the data over two rounds of step fitting as described below:

1. Perform a step fit based on the global maximum of the S-curve (S_{p1}^{\max}) that corresponds to the most prominent features in the data.
2. This step fit is then subtracted from the data and a secondary step fit is performed on the “residual data”.
3. Only if the global maximum of the secondary step fit is above the user-defined threshold — acceptance threshold — will the fit be accepted.

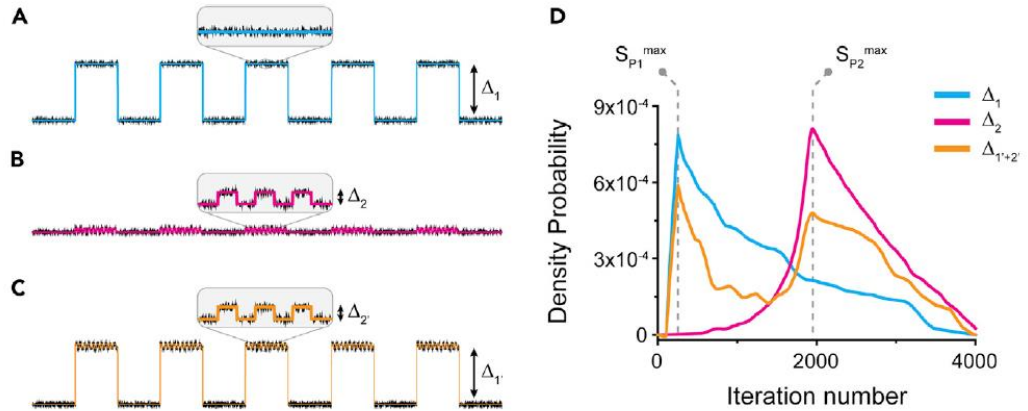


Figure 12. Two rounds (Dual-pass) of fitting to facilitate step detection of a trajectory with a wide variety of scales. A) An example of a trajectory displaying uniform steps with a size of Δ_1 . B) Another example of trajectory displaying uniform steps with a size of Δ_2 . C) A simulated single-molecule trajectory displaying non-uniform steps with step size of Δ_1 and Δ_2 . D) S-curves for the three instances displayed in (A-C). The S-curves for the trajectories that have large and small steps combined (C; orange line) yield two peaks with S_{p1}^{\max} and S_{p2}^{\max} (dashed gray line) being the global maximum of its corresponding peak [31].

Performing this method requires time that increases significantly as the number of data points N_0 increases. To have an efficient number of operations (i) and to reduce the time complexity of the algorithm, we need to re-use the information that is obtained during the localization of the first partition. After the algorithm has determined the average (A_w) value of a plateau (N_w), it determines the location of both N_L and N_R for $x(i)$, using a single operation. The procedure starts with $x(1)$ that is located at the left side of N_w (see Figure 13). The location (A_L) of N_L can be deduced by $A_L(i) = x(i)$, whereas the level of N_R is defined by:

$$A_r(i) = \frac{(N_w \cdot A_w - x(i))}{(N_w - 1)} \quad (12)$$

This procedure is repeated for the next location ($i + 1$) until each location of N_w is calculated. Using this strategy, the algorithm only needs N_0 operations per plateau and $2 \cdot N_0$ for a whole dataset.

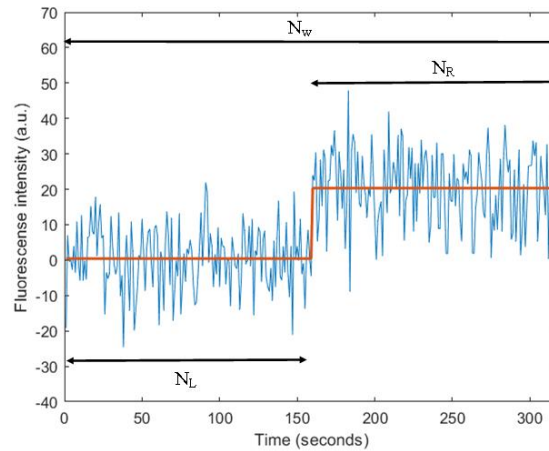


Figure 13. An example of trajectory analysis that utilizes the AutoStepfinder procedure. The existing plateau (N_w) is partitioned into new plateaus (N_L and N_R) at a point with largest reduction in the variance.

The advantage of using AutoStepfinder is its ability to facilitate high-throughput step detection with minimal user input within a user-friendly environment that is both robust and sensitive, allowing users to fit experimental single-molecule trajectories without any prior knowledge of the noise and steps within the data. Some studies have implemented this algorithm recently [49], [50].



4 Methods Implementation

We implemented automated step fitting and dwell time analysis on the experimental data obtained using objective-based TIRF microscopy of fast skeletal muscle myosin II heavy meromyosin fragments adsorbed to silanized glass surface. Single molecule ATPase experiment used Alexa647-ATP as fluorescent substrate. A detailed description of the experimental protocol, i.e., TIRF microscopy settings, procedures for the dwell time collection from experiment, composition of assay buffer, etc. is provided in Ušaj et al. [28].

In this study, 36 trajectories obtained from the experiment were prepared to be analyzed. The trajectories represented diverse behavior profile of the molecules and were selected to get a broader understanding of how well the experimental data can be faithfully analyzed by the selected software (*AutoStepfinder* and *vbFRET*). All fitted trajectories and the extracted dwell time events were compared with manual analysis.

4.1 Manual Analysis

Time series data were manually examined to determine the dwell time durations of fluorescence spots. We selected the time between a one-step increase in intensity above a predetermined threshold as the start of the dwell time and a single-step decrease in intensity below the threshold as the end of the dwell time. Dwell time duration was then calculated as a time difference between these two points. In this experiment, the resulting dwell times were interpreted as the time spent by Alexa647-nucleotide (Alexa647-ATP or Alexa647-ADP) bound to an immobilized myosin molecule.

4.2 Analysis using *AutoStepfinder*

4.2.1 Software implementation

Before performing the fitting using *AutoStepfinder*, several parameters need to be chosen, namely iteration range, time resolution, and acceptance threshold (see *Run Settings* section in Figure 14) [31]. First, the iteration range controls how far *AutoStepfinder* goes with the fitting procedure. After the defined number of iteration ranges is reached, the algorithm ends partitioning the trajectories and discovers the best fit. The iteration range is usually set to $\frac{1}{4}$ of the number of input data points. Second, the time resolution corresponds to the time range between each data point that will be used for time data in the output files of *AutoStepfinder*. Finally, the acceptance threshold establishes a limit for each fitting round and compares it to $S^{\max} - 1$, usually set between 0.1-1. If the set acceptance threshold exceeds the S^{\max} of the first round of fitting, *AutoStepfinder* will not execute the step-finding procedure.

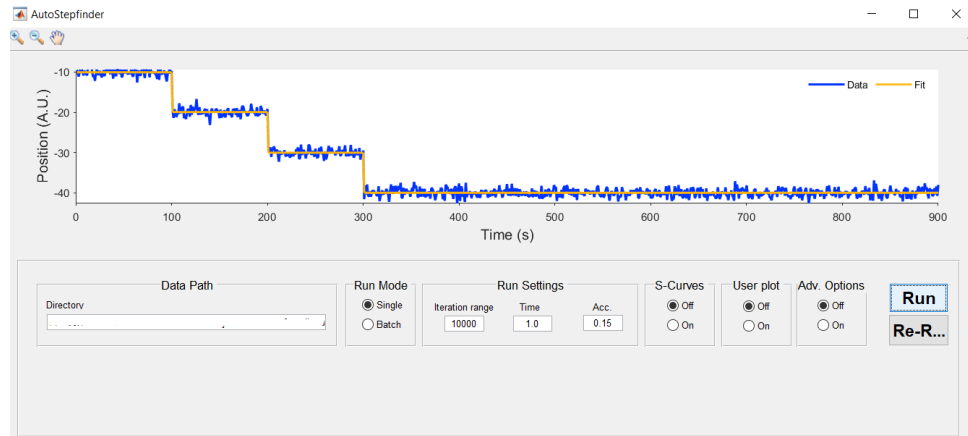


Figure 14. Graphical user interface of *AutoStepfinder*.

AutoStepfinder produced several outputs (available in *.txt or *.mat format) after each run, which contain:

1. *fits*, which represent data that can be used to plot the corresponding fit
2. *properties*, which contain the information required to generate histograms of the step size, step levels of fluorescence intensity (consists of 'Level before' and 'Level after' data), and dwell times (consists of 'Dwell time before' and 'Dwell time after' data)
3. *s_curve*, gives the information required to plot the S-curve,
4. *config*, gives the information consisting of method parameters that were used to generate the fit.

Examples of step fitting results obtained from *AutoStepfinder* and its corresponding distribution of levels is shown in Figure 15A. The blue line represents an experimental trajectory, while the orange line represents a fitted trajectory. Besides that, distribution of step sizes and dwell times can also be plotted from the data results (Figure 15B and 15C).

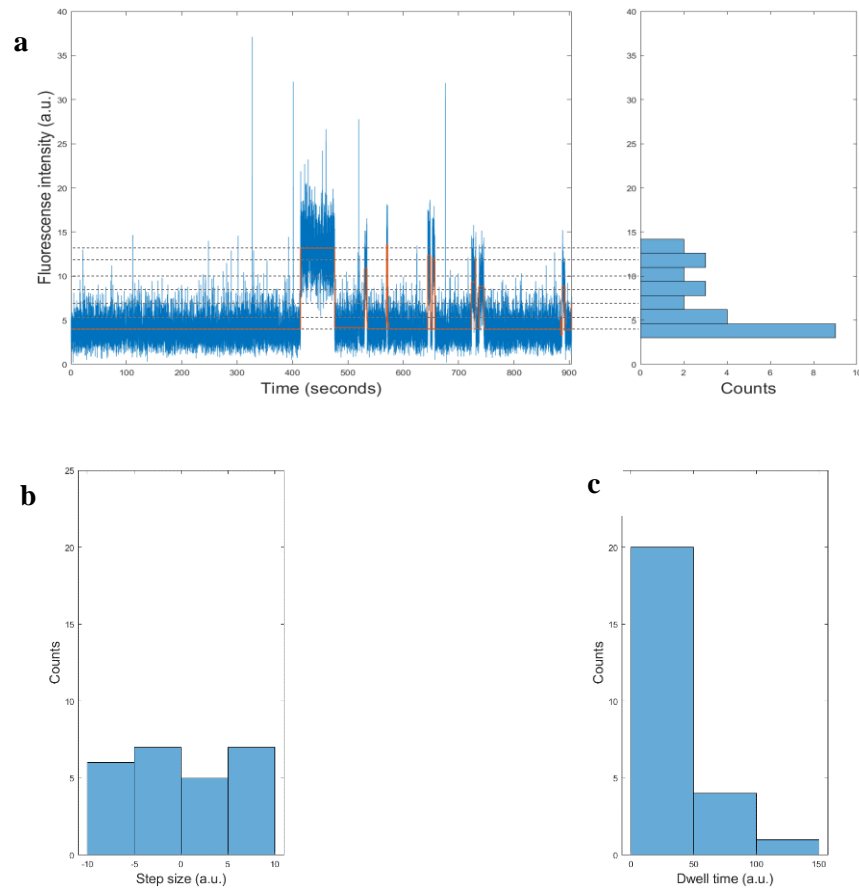


Figure 15. Example of post-processed results obtained from the *AutoStepfinder* algorithm.

a) Left panel shows a time trajectory generated from an experiment (blue line) and its fit result (orange line). Right panel shows distribution of levels of the corresponding trajectory. Histogram was obtained by binning the ‘Levels After’ data on properties.txt output file. b) Distribution of step sizes. Histogram was obtained by binning the ‘Step Size’ data on properties.txt output file. c) Distribution of dwell times. Histogram was obtained by binning the ‘Dwell Time After’ data on properties.txt output file.

Several parameter settings chosen for this implementation using *AutoStepfinder* were set as shown in Table 1. The experiments were recorded for 15 minutes with time resolution 52 ms, so the input data consists of 17,393 data points. Iteration range was set to $\frac{1}{4}$ of the number of data points $\approx 4,500$. For the acceptance threshold value, most of the trajectories were successfully fitted with the default settings value from the software (0.15), except for trajectories number 13, 23, 24, 25, and 29, where use of the default setting resulted in overfitting. In this case, overfitting is seen as the result are fitted to the noise of the data (see the example in Figure 16A). The acceptance thresholds for those trajectories were adjusted (Table 2) to produce a better fitting result as shown in Figure 16B.



AutoStepfinder allows fitting to be done for a single trajectory per run or for a set of trajectories per run in a batch mode. In the present implementation, batch mode was chosen to analyze 36 trajectories with total running time 48.7 seconds for all trajectories. That means it takes roughly 1.3 seconds to analyze each trajectory.

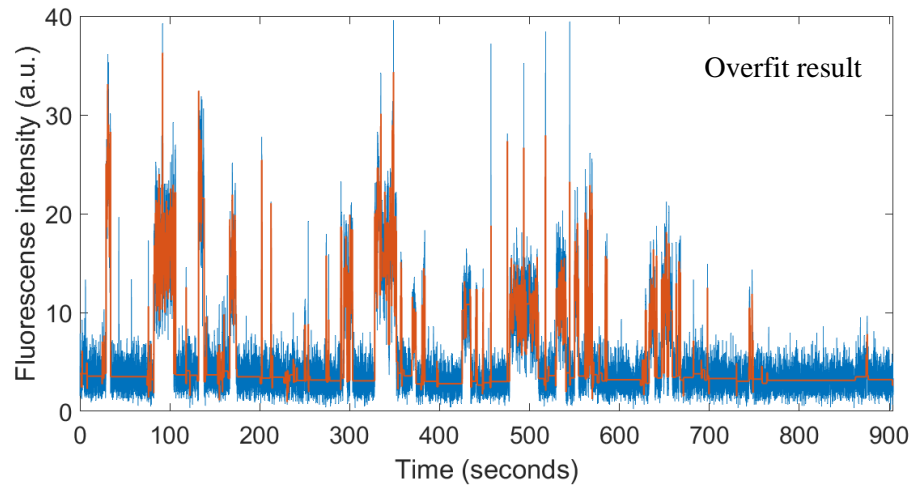
Table 1. Parameter settings for implementation using *AutoStepfinder*

Parameters	Values
Iteration range	4,500
Time resolution	52 ms
Acceptance threshold	0.15, vary for some trajectories (see text below)

Table 2. Adjusted acceptance threshold value used in the process of step fitting of several trajectories. For other trajectories those stated here, the value of acceptance threshold was set to the program default value of 0.15.

Trajectory Number	Acceptance Threshold
13	0.25
23	0.2
24	0.2
25	0.4
29	0.2

a



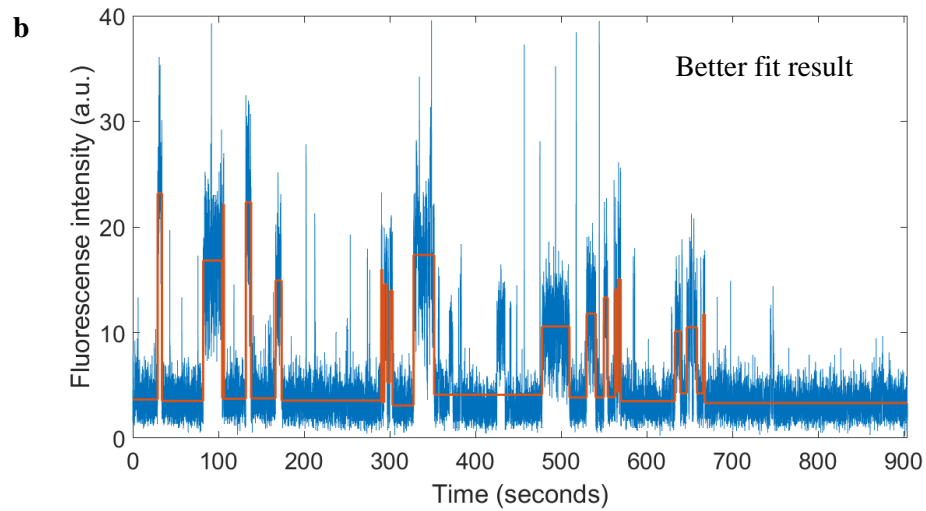


Figure 16. a) Example of trajectory fitted using *AutoStepfinder* with too low acceptance threshold value, produced an overfitting result. b) Adjusted acceptance threshold parameter for trajectory *a* produced a better fit.

4.2.2 Dwell time data extraction for analysis

Both dwell time and waiting time is included in ‘Dwell time’ output data obtained from *AutoStepfinder*. This method interprets dwell time as dwell time ‘on’ and waiting time as dwell time ‘off’. To make it clear, we are going to use the term “dwell time” and “waiting time” in following text. Besides, we were only considering the dwell time data for now and ignoring the waiting data. Hence, the dwell time classification was needed since the waiting time data was not included in the analysis.

Identifying the dwell times was done by comparing the ‘Level before’ and ‘Level after’ data from properties.txt output file using MATLAB. If the ‘Level before’ was larger than the ‘Level after’ data, the corresponding time duration was included as dwell time (see Figure 17). Otherwise, the data was categorized as waiting time and was not included in the dwell time analysis. The collected dwell times were then plotted as cumulative frequency distribution and fitted to sum of exponential functions using non-linear regression in GraphPad Prism software (ver 9.0).

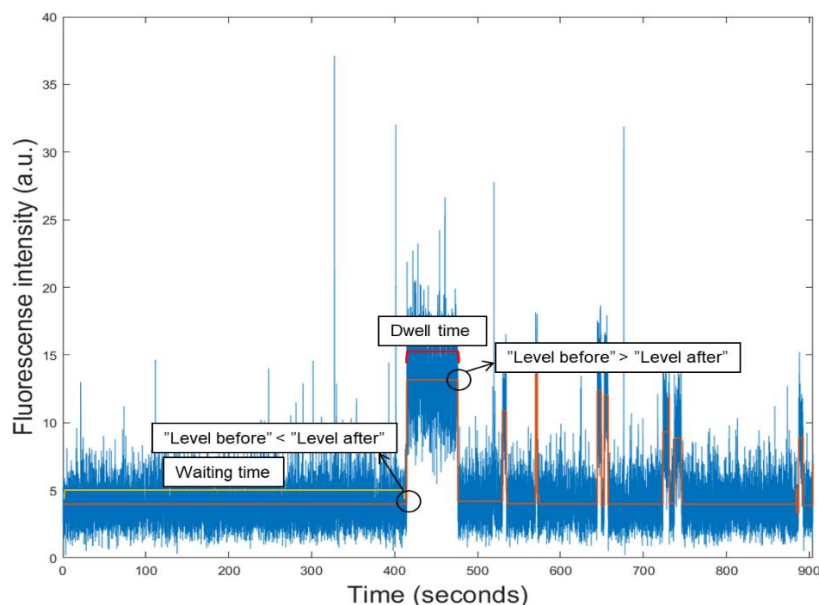


Figure 17. Dwell time identification in a fluorescence trajectory by comparing ‘Level before’ and ‘Level after’ data obtained from *AutoStepfinder* algorithm. The yellow bracket indicates waiting time event meanwhile the red bracket indicates dwell time event.

4.3 Analysis using vbFRET

4.3.1 Software implementation

Although designed for fluorescence resonance energy transfer (FRET) experimental data, the vbFRET software can also be used to analyze one channel single-molecule fluorescence intensity data. Before performing analysis using vbFRET, several parameter values need to be set, i.e., number of possible FRET states and fitting attempts per trace. In this context, the state corresponds to the fluorescence intensity level. The vbFRET program performs the fitting in each state within the given range and generates the result for each state specified. From the fitting result in each state, we can see which state is corresponding to underfit, correct fit (i.e., the best fit to trace), or overfit result. Other data results were stored in *.mat format.

Figure 18 shows an example of fitting results of vbFRET software in which the range of possible states was set between 1 and 5. Also, fitting iteration was set to 10 attempts per trace (see *Analysis settings* indicated with a yellow box in Figure 18). Hyperparameters settings can be found in the *Advanced Analysis Settings* located in the ‘Settings’ menu. Their default settings were chosen to give distributions consistent with experimental data and to influence the inference as weakly as possible, i.e., $u_{\pi} = 1$, $u_{\alpha} = 1$, $u_{\beta} = 0.25$, $u_{\mu} = 0.5$, $u_{\nu} = 5$, and $u_{\omega} = 50$ [36]. Hyperparameters are recommended to remain at the default settings unless we fully understand about Bayesian model of the HMM. The software package also has a simpler version without GUI. The range of states and number of iterations settings can also be done in this version.

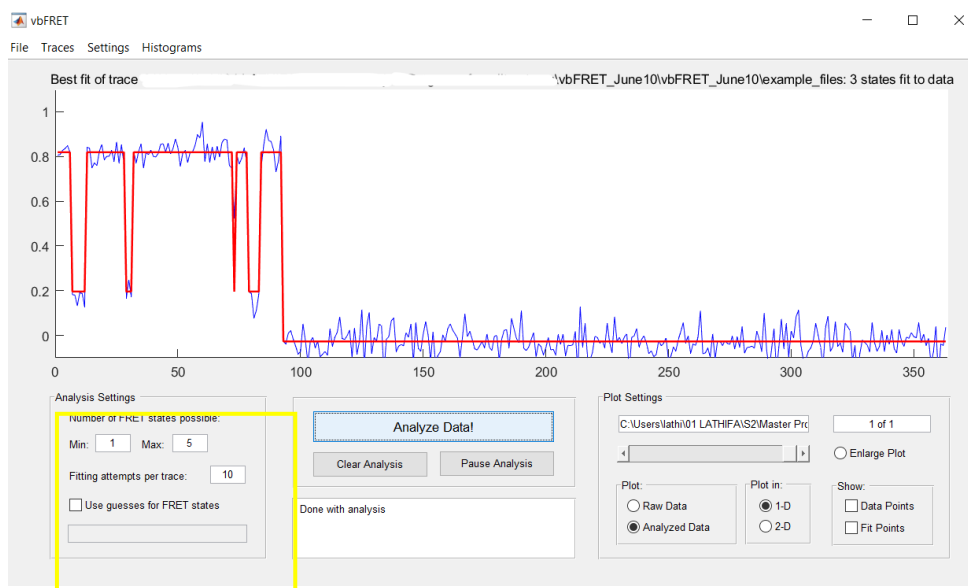


Figure 18. Graphical user interface of vbFRET performing analysis of a trajectory.

As explained above, the range of states may be guessed, corresponding to our knowledge of the data. For the first implementation attempt, the range of state was set between 3 and 5. The number of fitting iterations per trace was set to 10 attempts. Same as the previous method, 36 traces were analyzed using vbFRET. All trace data were collected into one cell array in MATLAB and then analyzed simultaneously. Running time for each trace varied between 20-45 seconds, and the total running time of the fitting process for all traces was ~1000 seconds.

One of the fitted trace results is shown in Figure 19 which was achieved by the range of states set to 3-5. The fitting process was performed for each state setting individually. Thus, three different fitting results (for 3, 4, and 5 states) were generated corresponding to respective state value. The output result of fitted hidden trajectories, fitted hidden states, and posterior parameters were stored after each run to be used for dwell time analysis.

Step fitting aims to identify all correct steps and avoid including any false steps. Unfortunately, along with an increasing number of detected steps, the risk of having a larger count of false-positive steps, i.e., overfitting, is also increasing. Thus, it is important to decide the fluorescence state's minimum and maximum value that adequately describes underlying data. In this case, we found that it is sufficient with 3 states to fit the trajectories, which was inferred from observation of fitted trajectory with 3 states yielding one state corresponding to the dwell times and the remaining states corresponding to the background intensity level (see Figure 19). Using maximum range of 4 or 5 states increase the possibility of any background intensities counted as dwell times.

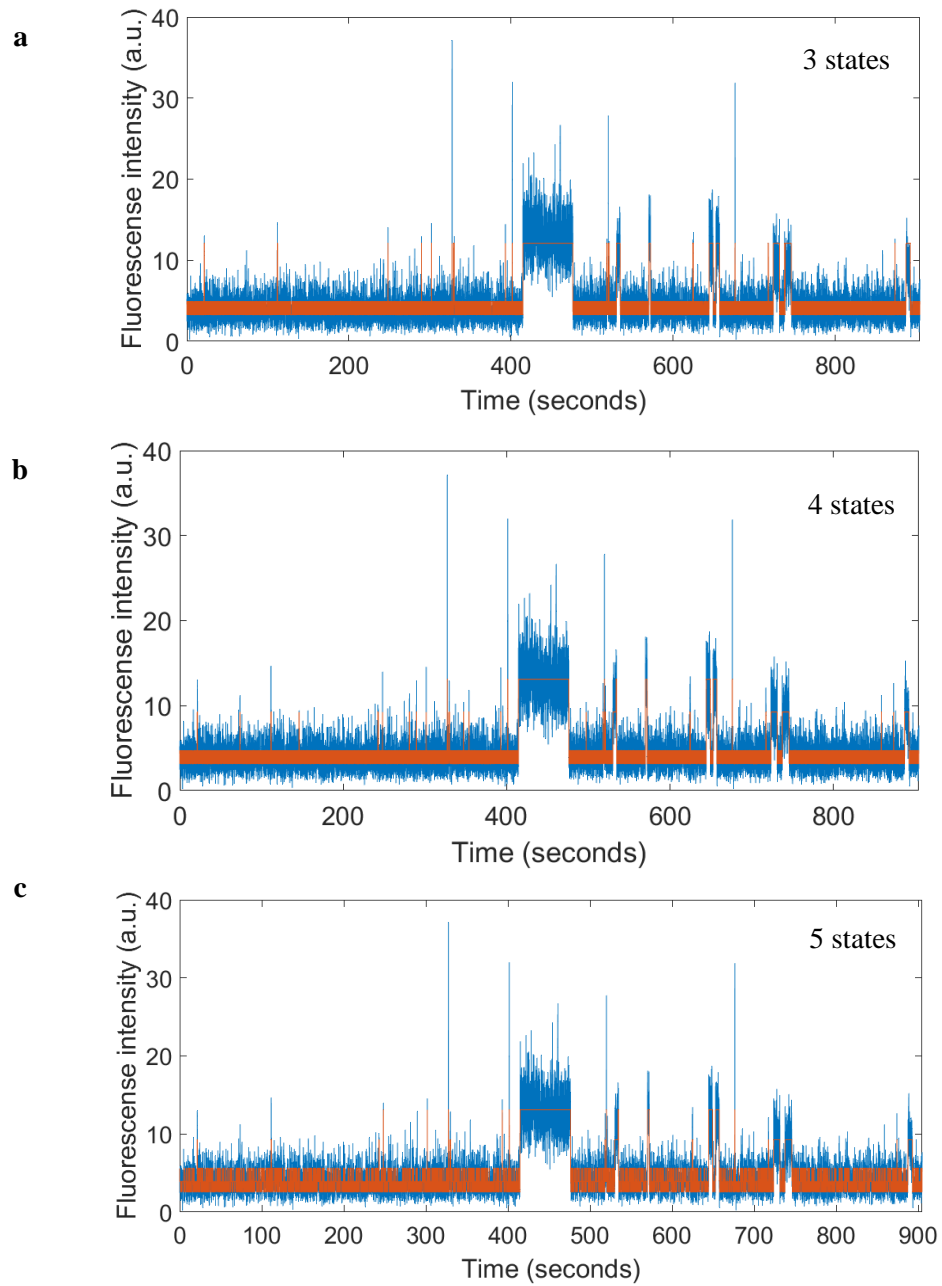


Figure 19. Example of trajectory (blue line) which was fitted by using vbFRET (orange line). Range of states was chosen between 3 and 5, and the fitting was performed for each state value separately: a) 3 states, b) 4 states, and c) 5 states. These results indicate that analysis with 3 states is sufficient. Fitting with 4 and 5 states increases the possibility of any background intensities counted as dwell times.

4.3.2 Dwell time data extraction for analysis

Collecting dwell time data from vbFRET results was achieved using a simple algorithm in MATLAB. This post-process used fitted hidden state and fitted



trajectories data. First, we determined which hidden state corresponds to a dwell time event since the value of hidden states that refer to dwell time varies from trace to trace. It was done with assumption that the states referred to as dwell time events correspond to high fluorescence intensity values of the trajectory. After identifying the 'dwell time state', we collected any time range corresponding to its 'dwell time state'. A simple flowchart of this dwell time extraction is shown in Figure 20. Then, the collected dwell time was fitted using non-linear regression in GraphPad Prism software, the same as done in the analysis based on *AutoStepfinder*.

To summarize this chapter, Table 3 below shows in brief the differences between the two implemented methods. The most significant difference is the basic principle of each method as well as its output result. Since the analysis will be focused on the observed dwell times, the output results need to be processed with distinct algorithms to collect the dwell time data from *AutoStepfinder* and vbFRET outputs.

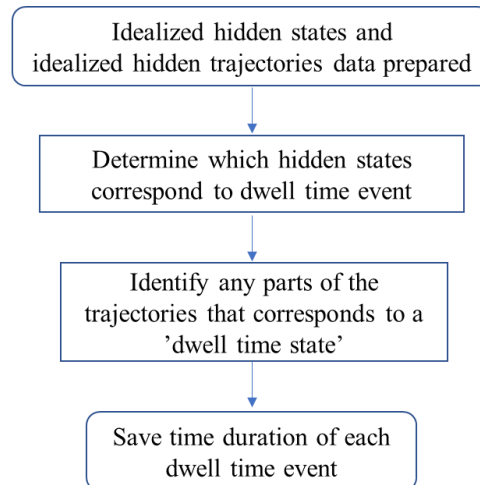


Figure 20. Flowchart of dwell time extraction process obtained from fitting result using vbFRET.



Comparison of two implemented methods

Table 3. Comparison of AutoStepfinder and vbFRET method

	<i>AutoStepfinder</i> (Loeff et al.)	vbFRET (Bronson et al.)
Software objective	Determine the kinetic states within single molecule trajectories without any information on the underlying noise in advance	Revealing the number of conformational states in single molecule FRET data
Algorithm used	Iteratively fits steps at locations that yield the biggest reduction in the variance (σ^2)	Implements variational Bayesian method coupled with hidden Markov model to model the dynamics of an observed time series
Programming language	MATLAB	MATLAB
Parameter settings	<ul style="list-style-type: none">• Iteration range• Acceptance threshold• Time resolution	<ul style="list-style-type: none">• Number of minimum and maximum state• Fitting attempts per trace
Running time	1.3 seconds per trace, 48.7 seconds in total	20-45 seconds per trace, 997.4 seconds in total
Results output	<ul style="list-style-type: none">• Fitted trajectories• Step levels• Dwell time	<ul style="list-style-type: none">• Posterior parameters• Fitted trajectories• Fitted hidden states

5 Results and Discussion

Dwell time analysis results obtained from *AutoStepfinder*

A comparison of the number of dwell time events on each trajectory detected using *AutoStepfinder*, vbFRET, and manual analysis is shown in Figure 21. In general, dwell time results using *AutoStepfinder* are relatively shifted down compared to manual analysis. That is, the number of dwell time events generated is appreciably lower than in the manual analysis with 398 dwell times in total compared to 776 for manual analysis. That could be because the software correctly detects dwell times with long durations but tends to ignore many short dwell times. However, that type of short dwells also needs to be considered as they may represent short unspecific binding of Alexa647-ATP.

It is known that trajectories of single-molecule fluorescence measurement have a unique characteristic of a low number of fluorescence intensity states, which also means the chance of revisiting each state is high [31]. This type of data is more favored to be fitted using hidden Markov model. In contrast, the trajectories with many fluorescence intensity states imply a low chance of revisiting the state, which is more suitable to be analyzed using *AutoStepfinder*, as it detects more states at a higher noise level. Thus, we can say that the *AutoStepfinder* is not well suited to fit our data, at least if we utilize the software without any modification.

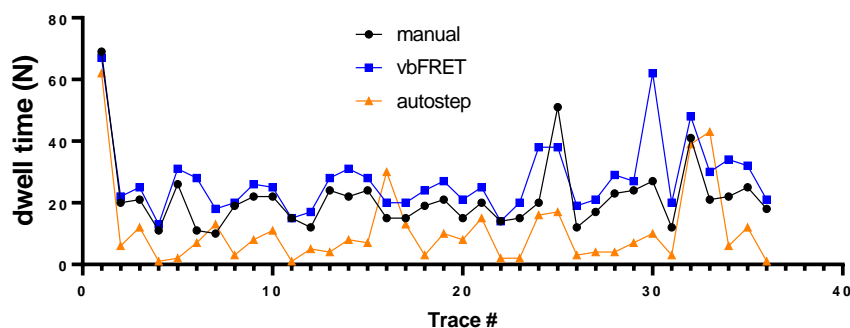


Figure 21. Comparison of number of dwell times (N) generated from manual analysis (black line), vbFRET (blue line), and *AutoStepfinder* (orange line).

Dwell time analysis results obtained from vbFRET

Using the vbFRET method, more dwell times were detected than in manual analysis, except one trace (i.e., trace number 24, see blue colored line in Figure 21) where fewer dwell times were observed. Collectively, more dwell times were detected using vbFRET, 984 events in total, than in the manual analysis where 776 dwell times were detected (Figure 22B and 22C). From the overlaid plot in Figure 22D comparing dwell times from manual analysis and vbFRET, it can be seen more clearly that vbFRET fittings detected more short dwell times with a minor difference in detecting the longer ones (see red circled areas in Figure 22B).



The reasons may be that vbFRET fitting still includes several specific cases that occurred in some trajectories, for example, events with focus problems, events that have not finished yet at the end of the trace, and events that start in a high fluorescence state. The dwell times in such conditions are not included in manual analysis. Note that events with very high amplitude are still included using all these methods. It is also essential to observe that the vbFRET method has a lower fluorescence intensity threshold value for fitting than was used in the manual analysis.

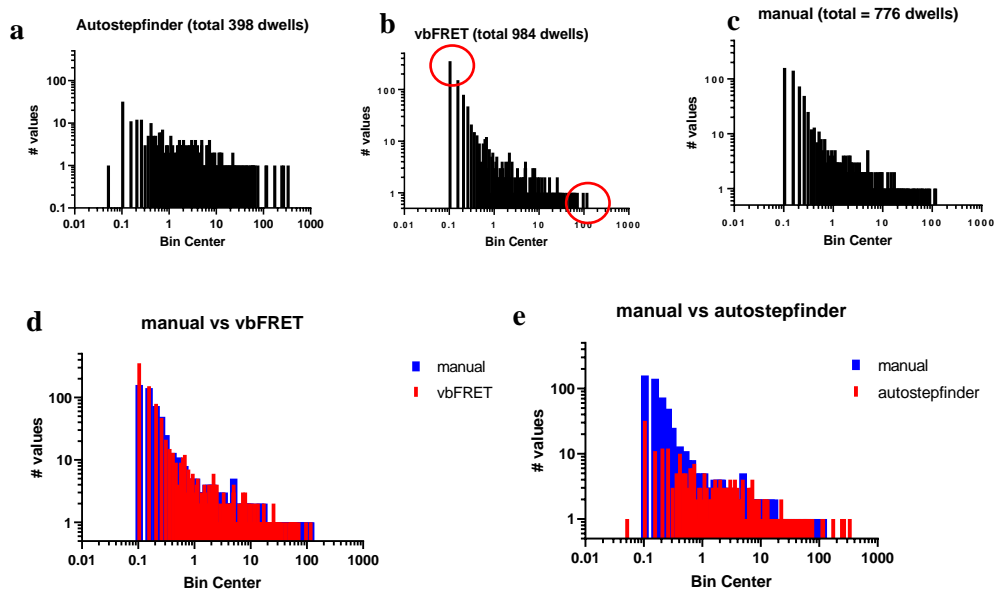


Figure 22. Histograms of dwell time resulting from: a) *Autostepfinder* fitting, b) vbFRET fitting, and c) manual analysis. d) Comparison of dwell time histogram between manual analysis and vbFRET result. e) Comparison of dwell time histogram between manual analysis and vbFRET result.



Cases of a considerably higher number of dwell time

Another notable difference between the manual data and the vbFRET data (see Figure 21) is the fitted result in trace number 30. Here, the method detects a considerably higher number of dwell time events than the manual method. In the trajectory plot shown in Figure 23A, there are some short events on top of long dwell time events, for example, the period 364-448 seconds (the area inside the yellow box in Figure 23A). The method counted these short events as dwell time events. Besides that, many part of traces with low intensities are included as dwell time events. This includes the five first events before the first ‘correct’ event at around 23 seconds in Figure 23A (grey area). With so many false-positive events ‘mistakenly’ counted as dwell times, the number of dwell times in trace number 30 is appreciably higher than the manual count.

To give better fit, change to the parameter “range of the number of states” were made. The modification was the change in the minimum number of fluorescent states from 3 to 1. In other words, the range of states in the fitting procedures was then between 1-5. After adjustment, the 3-state fitting result (Figure 23B) shows that the short events on top of a long dwell time event were being ignored. These results indicate that changing the range of states used in vbFRET procedure affects the fitted results significantly. According to what has been discussed in the ‘Method’ chapter, vbFRET estimates evidence and posterior probability distribution given a *range of possible states*.

The later fit (Figure 23B) still detected short signals with background intensity levels as regular dwell time events; the number of these dwell times was nevertheless reduced. To increase the fitting accuracy, we may have to have specific handling for the signals with a very high fluorescence intensity compared to the overall background fluorescence level.

Another possible improvement is to change the prior hyperparameter settings. However, further analysis is needed to consider for the optimal hyperparameter values. For example, a study has reported a method that outperforms vbFRET by setting the hyperparameter values as part of the inference procedure [37].

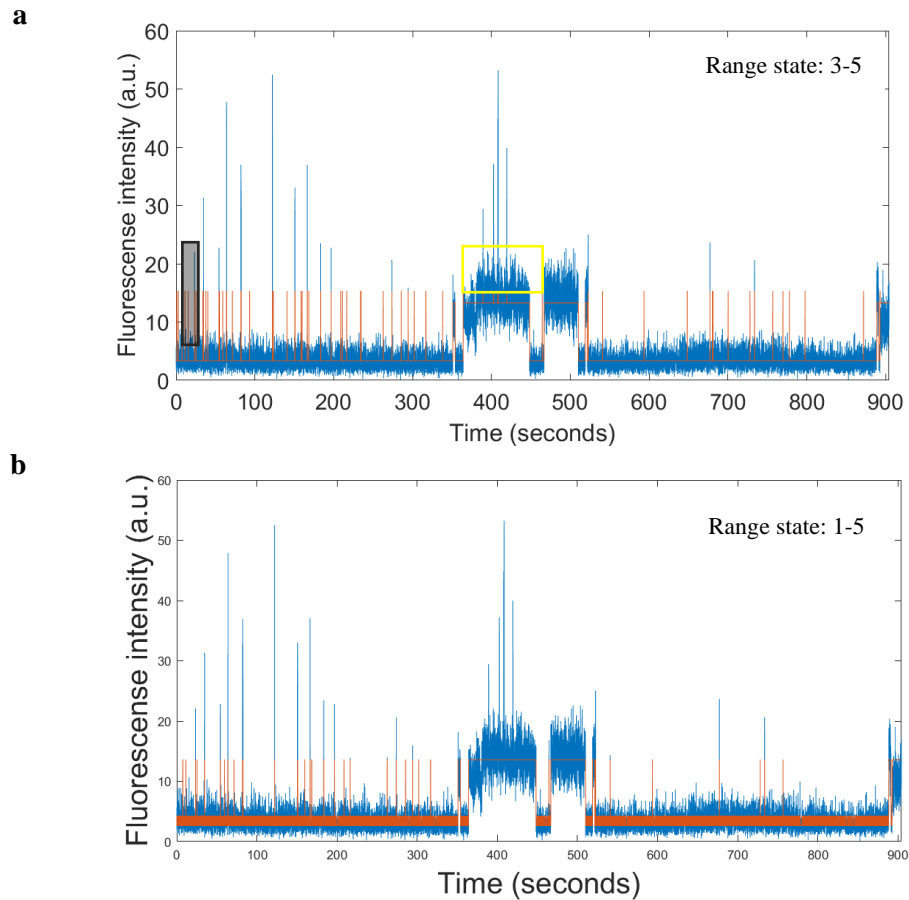


Figure 23. The 3-state fitting results of trace number 30 using vbFRET method with a different range of states: a) the range of states 3-5, b) the range of states changed to 1-5. In the upper panel, several short events indicated with a yellow box are counted as dwell times. Some parts of the trajectory with background intensity levels were also ‘mistakenly’ fitted as dwell times (indicated with a grey area). After the ‘range of states’ modification, the short events in the middle of the long dwell time were ignored (lower panel). Some points with background intensity levels were still counted as dwell times; however, their number was reduced compared to the result under **a**.

Distribution of dwell time durations

From manual analysis and vbFRET, we can examine the distribution of dwell time durations in each trace, as shown in Figure 24. The red dots represent any dwell time events resulting from manual analysis, while the black dots indicate the results obtained by using vbFRET. Denser dot assemblies scattered in some vertical lines represent traces with more dwell time events of specific durations, such as trace number 1, 25, and 32. In general, all traces have relatively large number of short dwell times. Only some of them have medium and long duration dwell time events. From

this plot, the longest durations of the dwell times can be estimated to be around 100 seconds.

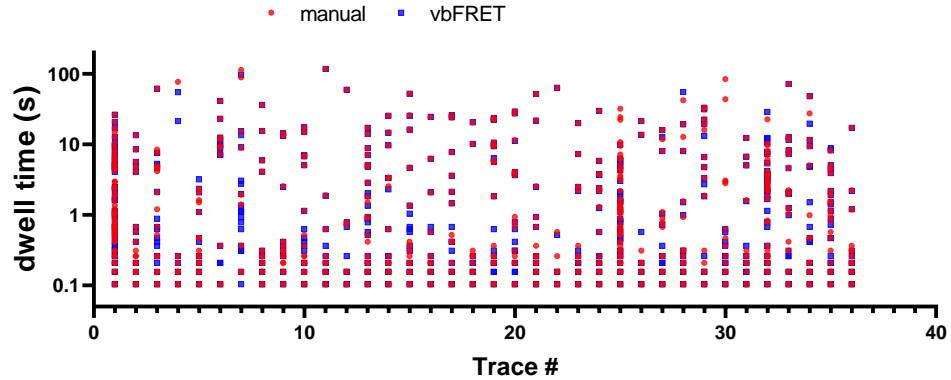


Figure 24. Distribution of dwell time duration in each trace generated from manual analysis and vbFRET fitting.

Rate constant analysis

To estimate the amplitudes and rate constants of underlying biochemical processes which generated the detected dwell times, one can plot them as cumulative frequency distributions (see Amrute-Nayak et al. PNAS 2014 [21]; Ušaj et al. Commun Biol 2021 [28]). The cumulative dwell time distribution describes the decay in number of dwell time events towards the increased duration of the events. Such distribution can then be fitted with sums of exponential functions to extract the relevant biochemical constants. In our case, it was fitted to a sum of three exponential functions.

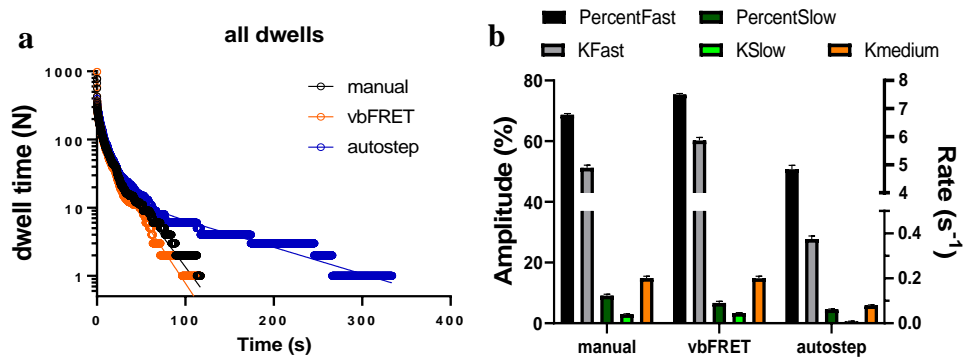


Figure 25. Single molecule ATPase by myosin (heavy meromyosin) using Alexa647-ATP analog. a) Comparison of cumulative dwell time distributions produced from manual analysis, vbFRET, and *AutoStepfinder*. Solid lines represent fitting to sum of triple exponential functions. b) Amplitudes and rate constants of fast, medium, and slowest exponential processes obtained from fittings of distributions in a. Data represent mean \pm 95% CI intervals derived from the fits. Experiments were done at 23 °C.

As it is seen from the cumulative frequency distribution plot in Figure 25A that *Autostepfinder* is more biased toward counting events with long dwell times. The consequence is that some short events were skipped and/or lumped together as longer dwell times even though they should be detected as separate multi dwell events. Moreover, long waiting times with slight change of step level during that time may be misinterpreted as dwell time, as seen in trace in Figure 26. Hence, maximum dwell time durations obtained from *Autostepfinder* were detected to be considerably longer, ~330 seconds. These maximum dwell time durations are much different from those generated by vbFRET and manual analysis, as a maximum dwell time duration of about 120 seconds is obtained in both cases. Correspondingly, the amplitudes and rate constant values derived using manual and vbFRET-based analysis are quite similar (see Figure 25B and Table 4), particularly for the medium and slow phases.

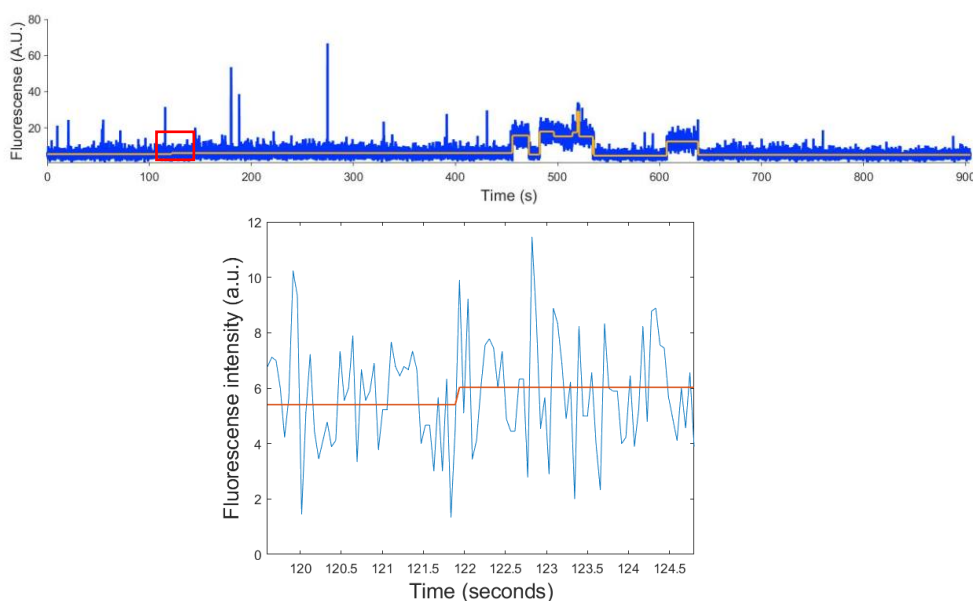


Figure 26. Upper panel shows time trace fitted with *Autostepfinder* which detected the longest dwell time duration between 120 and 450 seconds of the signal, despite having a very small difference in fluorescence intensity signal (indicated with red square) than the background fluorescence signal between 0 and 120 seconds. Lower panel shows a detailed view of time trace (inside red square in upper panel) with slight difference in fluorescence level from the background intensity level which was misinterpreted as the beginning of a dwell time.

Table 4. Rate constants (k) and amplitude (A) values of fast, medium, and slowest exponential processed obtained from fittings in Figure 25. All values are mean $\pm 95\%$ CI.

	Fast		Medium		Slow	
	$k_{\text{fast}} (\text{s}^{-1})$	$A_{\text{fast}} (\%)$	$k_{\text{medium}} (\text{s}^{-1})$	$A_{\text{medium}} (\%)$	$k_{\text{slow}} (\text{s}^{-1})$	$A_{\text{slow}} (\%)$
Manual	4.89	68.7	0.20	22.1	0.040	9.1
vbFRET	5.86	75.3	0.20	17.9	0.045	6.7
<i>AutoStepfinder</i>	0.37	50.7	0.07	44.7	0.009	4.5



Conclusions

The main purpose of this study was to develop an automated method to analyze single-molecule data. Meanwhile, in this report, the objective was to identify and implement the single-molecule analysis using already available methods, i.e., *AutoStepfinder* and vbFRET. The results were compared to manual analysis in order to investigate how the automated methods can facilitate analysis of single-molecule experimental data and determine which method is most suitable for our single-molecule data.

An automatic step detection is aimed to facilitate the selection of relevant events called dwell times. Then, the amplitudes and rate constants of underlying biochemical processes which generated dwell times might be estimated. Namely, an unsupervised method called *AutoStepfinder* could provide the step detections of single-molecule trajectories with minimum information regarding the step size, step location, and noise distributions of the data. Another method, vbFRET, implemented Variational Bayesian inference coupled with a hidden Markov model to identify the states of fluorescence levels also can be used to determine the dwell times.

Methods were implemented to analyze single molecule ATPase experiments using a fluorescent ATP analog. The data were generated using total internal reflection fluorescence (TIRF) microscopy by observing fluorescence nucleotide association and dissociation from immobilized individual myosin motor proteins attached to a coverslip surface. We found that both *AutoStepfinder* and vbFRET can be used for step detection and dwell time analysis. As discussed in Chapter 5, the results confirm that vbFRET performed better in step fitting and ‘correctly’ detects real dwell times. *AutoStepfinder* could be reliable for detecting long dwell times, however, in our case it tended to specifically miss the short dwell times. This could be related to the unique characteristics of single-molecule fluorescence measurements available for this study accompanied with a low number of states and the high chance of revisits per state. We conclude that vbFRET method is more appropriate for analyzing single-molecule ATPase experiments.

Future Work

Some traces from the experiment that exhibit more complex behavior were not selected for this work. Thus, improvements may need to be implemented into post-processing of automatically detected dwell times according to further rules applied to such complex traces. With such modification, it is believed that automated dwell time analysis will be further improved.



References

- [1] T. Ha, 'Single-molecule methods leap ahead', *Nature Methods* 2014 11:10, vol. 11, no. 10, pp. 1015–1018, Sep. 2014, doi: 10.1038/nmeth.3107.
- [2] A. E. Knight, 'Introduction: The "Single Molecule" Paradigm', in *Single Molecule Biology*, A. E. Knight, Ed. New York: Academic Press, 2009, pp. xvii–xxxv. doi: <https://doi.org/10.1016/B978-0-12-374227-8.00011-0>.
- [3] J. Zlatanova and K. van Holde, 'Single-Molecule Biology: What Is It and How Does It Work?', *Mol Cell*, vol. 24, no. 3, pp. 317–329, Nov. 2006, doi: 10.1016/J.MOLCEL.2006.10.017.
- [4] W. E. Moerner and L. Kador, 'Optical detection and spectroscopy of single molecules in a solid', *Phys Rev Lett*, vol. 62, no. 21, pp. 2535–2538, May 1989, doi: 10.1103/PHYSREVLETT.62.2535/FIGURE/1/THUMB.
- [5] M. Orrit and J. Bernard, 'Single pentacene molecules detected by fluorescence excitation in a *p*-terphenyl crystal', *Phys Rev Lett*, vol. 65, no. 21, p. 2716, Nov. 1990, doi: 10.1103/PhysRevLett.65.2716.
- [6] A. A. Deniz, S. Mukhopadhyay, and E. A. Lemke, 'Single-molecule biophysics: At the interface of biology, physics and chemistry', *Journal of the Royal Society Interface*, vol. 5, no. 18. Royal Society, pp. 15–45, Jan. 06, 2008. doi: 10.1098/rsif.2007.1021.
- [7] M. Yasui, M. Hiroshima, J. Kozuka, Y. Sako, and M. Ueda, 'Automated single-molecule imaging in living cells', *Nat Commun*, vol. 9, no. 1, p. 3061, 2018, doi: 10.1038/s41467-018-05524-7.
- [8] F. Persson, I. Barkefors, and J. Elf, 'Single molecule methods with applications in living cells', *Curr Opin Biotechnol*, vol. 24, no. 4, pp. 737–744, Aug. 2013, doi: 10.1016/J.COPBIO.2013.03.013.
- [9] N. Li, R. Zhao, Y. Sun, Z. Ye, K. He, and X. Fang, 'Single-molecule imaging and tracking of molecular dynamics in living cells', *Natl Sci Rev*, vol. 4, no. 5, pp. 739–760, Sep. 2017, doi: 10.1093/nsr/nww055.
- [10] H. Zhan *et al.*, 'In vivo single-molecule imaging identifies altered dynamics of calcium channels in dystrophin-mutant *C. elegans*', *Nature Communications* 2014 5:1, vol. 5, no. 1, pp. 1–12, Sep. 2014, doi: 10.1038/ncomms5974.
- [11] F. B. Robin, W. M. McFadden, B. Yao, and E. M. Munro, 'Single-molecule analysis of cell surface dynamics in *Caenorhabditis elegans* embryos', *Nature Methods* 2014 11:6, vol. 11, no. 6, pp. 677–682, Apr. 2014, doi: 10.1038/nmeth.2928.



- [12] M. J. M. Schaaf *et al.*, ‘Single-Molecule Microscopy Reveals Membrane Microdomain Organization of Cells in a Living Vertebrate’, *Biophys J*, vol. 97, no. 4, pp. 1206–1214, Aug. 2009, doi: 10.1016/J.BPJ.2009.05.044.
- [13] A. Y. Guo, Y. M. Zhang, L. Wang, D. Bai, Y. P. Xu, and W. Q. Wu, ‘Single-Molecule Imaging in Living Plant Cells: A Methodological Review’, *Int J Mol Sci*, vol. 22, no. 10, p. 5071, May 2021, doi: 10.3390/IJMS22105071.
- [14] X. Wang, X. Li, X. Deng, D. T. Luu, C. Maurel, and J. Lin, ‘Single-molecule fluorescence imaging to quantify membrane protein dynamics and oligomerization in living plant cells’, *Nature Protocols 2015 10:12*, vol. 10, no. 12, pp. 2054–2063, Nov. 2015, doi: 10.1038/nprot.2015.132.
- [15] W. Colomb and S. K. Sarkar, ‘Extracting physics of life at the molecular level: A review of single-molecule data analyses’, *Phys Life Rev*, vol. 13, pp. 107–137, Jun. 2015, doi: 10.1016/J.PLREV.2015.01.017.
- [16] K. Blank, G. de Cremer, and J. Hofkens, ‘Fluorescence-based analysis of enzymes at the single-molecule level’, *Biotechnol J*, vol. 4, no. 4, pp. 465–479, Apr. 2009, doi: 10.1002/BIOT.200800262.
- [17] T. Xia, N. Li, and X. Fang, ‘Single-Molecule Fluorescence Imaging in Living Cells’, <http://dx.doi.org.proxy.lnu.se/10.1146/annurev-physchem-040412-110127>, vol. 64, pp. 459–480, Apr. 2013, doi: 10.1146/ANNUREV-PHYSCHEM-040412-110127.
- [18] T. Funatsu, Y. Harada, M. Tokunaga, K. Saito, and T. Yanagida, ‘Imaging of single fluorescent molecules and individual ATP turnovers by single myosin molecules in aqueous solution’, *Nature 1995 374:6522*, vol. 374, no. 6522, pp. 555–559, 1995, doi: 10.1038/374555a0.
- [19] M. Tokunaga, K. Kitamura, K. Saito, A. H. Iwane, and T. Yanagida, ‘Single Molecule Imaging of Fluorophores and Enzymatic Reactions Achieved by Objective-Type Total Internal Reflection Fluorescence Microscopy’, *Biochem Biophys Res Commun*, vol. 235, no. 1, pp. 47–53, Jun. 1997, doi: 10.1006/BBRC.1997.6732.
- [20] A. Ishijima *et al.*, ‘Simultaneous Observation of Individual ATPase and Mechanical Events by a Single Myosin Molecule during Interaction with Actin’, *Cell*, vol. 92, no. 2, pp. 161–171, Jan. 1998, doi: 10.1016/S0092-8674(00)80911-3.
- [21] M. Amrute-Nayak *et al.*, ‘ATP turnover by individual myosin molecules hints at two conformers of the myosin active site’, *Proc Natl Acad Sci U S A*, vol. 111, no. 7, pp. 2536–2541, Feb. 2014, doi: 10.1073/PNAS.1316390111/-/DCSUPPLEMENTAL.
- [22] M. Balaz, M. Sundberg, M. Persson, J. Kvassman, and A. Mansson, ‘Effects of Surface Adsorption on Catalytic Activity of Heavy Meromyosin Studied



- Using a Fluorescent ATP Analogue †', *Biochemistry*, vol. 46, pp. 7233–7251, Jul. 2007, doi: 10.1021/bi700211u.
- [23] D. Riveline, “‘Single molecule’: theory and experiments, an introduction.’, *J Nanobiotechnology*, vol. 11 Suppl 1, no. 1, pp. 1–10, Dec. 2013, doi: 10.1186/1477-3155-11-S1-S1/FIGURES/5.
- [24] M. F. Serag and S. Habuchi, ‘Conserved linear dynamics of single-molecule Brownian motion’, *Nature Communications* 2017 8:1, vol. 8, no. 1, pp. 1–11, Jun. 2017, doi: 10.1038/ncomms15675.
- [25] R. Dave, D. S. Terry, J. B. Munro, and S. C. Blanchard, ‘Mitigating unwanted photophysical processes for improved single-molecule fluorescence imaging’, *Biophys J*, vol. 96, no. 6, pp. 2371–2381, Mar. 2009, doi: 10.1016/J.BPJ.2008.11.061/ATTACHMENT/FD45F890-6BFC-4D5A-BB7C-784FC7C3EFE9/MMC1.PDF.
- [26] T. Ha and P. Tinnefeld, ‘Photophysics of Fluorescent Probes for Single-Molecule Biophysics and Super-Resolution Imaging’, <http://dx.doi.org/10.1146/annurev-physchem-032210-103340>, vol. 63, pp. 595–617, Apr. 2012, doi: 10.1146/ANNUREV-PHYSCHEM-032210-103340.
- [27] E. M. Boehm, S. Subramanyam, M. Ghoneim, M. T. Washington, and M. Spies, ‘Quantifying the assembly of multicomponent molecular machines by single-molecule total internal reflection fluorescence microscopy’, *Methods Enzymol*, vol. 581, p. 105, 2016, doi: 10.1016/BS.MIE.2016.08.019.
- [28] M. Ušaj, L. Moretto, V. Vemula, A. Salhotra, and A. Månsson, ‘Single molecule turnover of fluorescent ATP by myosin and actomyosin unveil elusive enzymatic mechanisms’, *Communications Biology* 2021 4:1, vol. 4, no. 1, pp. 1–12, Jan. 2021, doi: 10.1038/s42003-020-01574-0.
- [29] B. C. Carter, M. Vershinin, and S. P. Gross, ‘A Comparison of Step-Detection Methods: How Well Can You Do?’, *Biophys J*, vol. 94, pp. 306–319, 2008, doi: 10.1529/biophysj.107.110601.
- [30] S. A. McKinney, A. C. Déclais, D. M. J. Lilley, and T. Ha, ‘Structural dynamics of individual Holliday junctions’, *Nature Structural Biology* 2002 10:2, vol. 10, no. 2, pp. 93–97, Dec. 2002, doi: 10.1038/nsb883.
- [31] L. Loeff, J. W. J. Kerssemakers, C. Joo, and C. Dekker, ‘AutoStepfinder: A fast and automated step detection method for single-molecule analysis’, *Patterns*, vol. 2, no. 5, May 2021, doi: 10.1016/j.patter.2021.100256.
- [32] B. Kalafut and K. Visscher, ‘An objective, model-independent method for detection of non-uniform steps in noisy signals’, *Comput Phys Commun*, vol. 179, no. 10, pp. 716–723, Nov. 2008, doi: 10.1016/j.cpc.2008.06.008.



- [33] M. Blanco and N. G. Walter, ‘Analysis of Complex Single Molecule FRET Time Trajectories’, *Methods Enzymol*, vol. 472, p. 153, Jan. 2010, doi: 10.1016/S0076-6879(10)72011-5.
- [34] S. R. Eddy, ‘What is a hidden Markov model?’, *Nature Biotechnology* 2004 22:10, vol. 22, no. 10, pp. 1315–1316, Oct. 2004, doi: 10.1038/nbt1004-1315.
- [35] S. A. McKinney, C. Joo, and T. Ha, ‘Analysis of Single-Molecule FRET Trajectories Using Hidden Markov Modeling’, *Biophys J*, vol. 91, no. 5, pp. 1941–1951, Sep. 2006, doi: 10.1529/BIOPHYSJ.106.082487.
- [36] J. E. Bronson, J. Fei, J. M. Hofman, R. L. Gonzalez, and C. H. Wiggins, ‘Learning Rates and States from Biophysical Time Series: A Bayesian Approach to Model Selection and Single-Molecule FRET Data’, *Biophys J*, vol. 97, no. 12, p. 3196, Dec. 2009, doi: 10.1016/J.BPJ.2009.09.031.
- [37] J. W. van de Meent, J. E. Bronson, C. H. Wiggins, and R. L. Gonzalez, ‘Empirical Bayes methods enable advanced population-level analyses of single-molecule FRET experiments’, *Biophys J*, vol. 106, no. 6, pp. 1327–1337, Mar. 2014, doi: 10.1016/J.BPJ.2013.12.055.
- [38] I. Tinoco, Jr, R. L. Gonzalez, and Jr, ‘Biological mechanisms, one molecule at a time’, *Genes Dev*, vol. 25, no. 12, p. 1205, Jun. 2011, doi: 10.1101/GAD.2050011.
- [39] D. Axelrod, ‘Cell-substrate contacts illuminated by total internal reflection fluorescence.’, *Journal of Cell Biology*, vol. 89, no. 1, pp. 141–145, Apr. 1981, doi: 10.1083/JCB.89.1.141.
- [40] D. Axelrod, ‘Total internal reflection fluorescence microscopy in cell biology’, *Traffic*, vol. 2, no. 11, pp. 764–774, 2001, doi: 10.1034/J.1600-0854.2001.21104.X.
- [41] E. J. G. Peterman, H. Sosa, and W. E. Moerner, ‘SINGLE-MOLECULE FLUORESCENCE SPECTROSCOPY AND MICROSCOPY OF BIOMOLECULAR MOTORS’, *Annu. Rev. Phys. Chem*, vol. 55, pp. 79–96, 2004, doi: 10.1146/annurev.physchem.55.091602.094340.
- [42] M. L. Martin-Fernandez, C. J. Tynan, and S. E. D. Webb, ‘A “pocket guide” to total internal reflection fluorescence’, *J Microsc*, vol. 252, no. 1, pp. 16–22, Oct. 2013, doi: 10.1111/JMI.12070.
- [43] A. E. and W. N. G. Blanco Mario R. and Johnson-Buck, ‘Hidden Markov Modeling in Single-Molecule Biophysics’, in *Encyclopedia of Biophysics*, G. C. K. Roberts, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 971–975. doi: 10.1007/978-3-642-16712-6_493.



- [44] N. Zarrabi, P. Schluesche, M. Meisterernst, M. Börsch, and D. C. Lamb, ‘Analyzing the Dynamics of Single TBP-DNA-NC2 Complexes Using Hidden Markov Models’, *Biophys J*, vol. 115, no. 12, p. 2310, Dec. 2018, doi: 10.1016/J.BPJ.2018.11.015.
- [45] A. J. Viterbi, ‘Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm’, *IEEE Trans Inf Theory*, vol. 13, no. 2, pp. 260–269, 1967, doi: 10.1109/TIT.1967.1054010.
- [46] C. M. Bishop, *Pattern Recognition and Machine Learning*. 2006. Accessed: May 15, 2022. [Online]. Available: <https://link.springer.com/book/9780387310732>
- [47] C. D. Kinz-Thompson, K. K. Ray, and R. L. Gonzalez, ‘Bayesian inference: The comprehensive approach to analyzing single-molecule experiments’, *Annu Rev Biophys*, vol. 50, p. 191, May 2021, doi: 10.1146/ANNUREV-BIOPHYS-082120-103921.
- [48] J. W. J. Kerssemakers, E. Laura Munteanu, L. Laan, T. L. Noetzel, M. E. Janson, and M. Dogterom, ‘Assembly dynamics of microtubules at molecular resolution’, *Nature*, vol. 442, no. 7103, pp. 709–712, Aug. 2006, doi: 10.1038/nature04928.
- [49] A. Bandyopadhyay and M. P. Goldschen-Ohm, ‘Unsupervised selection of optimal single-molecule time series idealization criterion’, *Biophys J*, vol. 120, no. 20, pp. 4472–4483, Oct. 2021, doi: 10.1016/J.BPJ.2021.08.045.
- [50] A. Barulin, P. Roy, J.-B. Claude, and J. Wenger, ‘Ultraviolet optical horn antennas for label-free detection of single proteins’, *Nature Communications* 2022 13:1, vol. 13, no. 1, pp. 1–9, Apr. 2022, doi: 10.1038/s41467-022-29546-4.
- [51] G. M. Cooper, ‘Actin, Myosin, and Cell Movement’, 2000, Accessed: Oct. 21, 2022. [Online]. Available: <https://www.ncbi.nlm.nih.gov/books/NBK9961/>
- [52] Y. Y. Toyoshima, S. J. Kron, E. M. McNally, K. R. Niebling, C. Toyoshima, and J. A. Spudich, ‘Myosin subfragment-1 is sufficient to move actin filaments in vitro’, *Nature*, vol. 328, no. 6130, pp. 536–539, 1987, doi: 10.1038/328536A0.