2019

Identification of Multiple Kinetic Populations of DNA-Binding Proteins in Live Cells

Han Ngoc Ho
University of Wollongong, hnh725@uowmail.edu.au

Daniel Zalami
Universitat Bayreuth

Jurgen Kohler
Universitat Bayreuth

Antoine M. van Oijen
University of Wollongong, vanoijen@uow.edu.au

Harshad Ghodke
University of Wollongong, harshad@uow.edu.au

Identification of Multiple Kinetic Populations of DNA-Binding Proteins in Live Cells

Abstract

Understanding how multiprotein complexes function in cells requires detailed quantitative understanding of their association and dissociation kinetics. Analysis of the heterogeneity of binding lifetimes enables the interrogation of the various intermediate states formed during the reaction. Single-molecule fluorescence imaging permits the measurement of reaction kinetics inside living organisms with minimal perturbation. However, poor photophysical properties of fluorescent probes limit the dynamic range and accuracy of measurements of off rates in live cells. Time-lapse single-molecule fluorescence imaging can partially overcome the limits of photobleaching; however, limitations of this technique remain uncharacterized. Here, we present a structured analysis of which timescales are most accessible using the time-lapse imaging approach and explore uncertainties in determining kinetic subpopulations. We demonstrate the effect of shot noise on the precision of the measurements as well as the resolution and dynamic range limits that are inherent to the method. Our work provides a convenient implementation to determine theoretical errors from measurements and to support interpretation of experimental data.

Disciplines

Medicine and Health Sciences

Publication Details

Ho, H. N., Zalami, D., Kohler, J., van Oijen, A. M. & Ghodke, H. (2019). Identification of Multiple Kinetic Populations of DNA-Binding Proteins in Live Cells. Biophysical Journal, 117 (5), 950-961.

Identification of multiple kinetic populations of DNA-binding proteins in live cells

Running title: Measuring binding lifetimes in cells

Han N. Ho[1,2,6], Daniel Zalami[3], Jürgen Köhler[3,4,5], Antoine M. van Oijen[1,2,*] and Harshad Ghodke[1,2,*]

[1]Molecular Horizons and School of Chemistry and Molecular Bioscience, University of Wollongong, Wollongong, Australia

[2]Illawarra Health and Medical Research Institute, Wollongong, Australia

[3]Spectroscopy of Soft Matter, University of Bayreuth, Bayreuth, Germany

[4]Bavarian Polymer Institute, Bayreuth, Germany

[5]Bayreuth Institute of Macromolecular Research (BIMF), University of Bayreuth, Bayreuth, Germany

[6]Present address: The Francis Crick Institute, London, UK

* Correspondence: harshad@uow.edu.au or vanoijen@uow.edu.au

ABSTRACT

Understanding how multi-protein complexes function in cells requires detailed quantitative understanding of their association and dissociation kinetics. Analysis of the heterogeneity of binding lifetimes enables interrogation of the various intermediate states formed during the reaction. Single-molecule fluorescence imaging permits the measurement of reaction kinetics inside living organisms with minimal perturbation. However, poor photo-physical properties of fluorescent probes limit the dynamic range and accuracy of measurements of off rates in live cells. Time-lapse single-molecule fluorescence imaging can partially overcome the limits of photobleaching, however, limitations of this technique remain uncharacterized. Here, we present a structured analysis of which timescales are most accessible using the time-lapse

imaging approach and explore uncertainties in determining kinetic sub-populations. We demonstrate the effect of shot noise on the precision of the measurements, as well as the resolution and dynamic range limits that are inherent to the method. Our work provides a convenient implementation to determine theoretical errors from measurements and to support interpretation of experimental data.

STATEMENT OF SIGNIFICANCE

Measuring lifetimes of interactions between DNA-binding proteins and their substrates is important for understanding how they function in cells. In principle, time-lapse imaging of fluorescently-tagged proteins using single-molecule methods can be used to identify multiple sub-populations of DNA-binding proteins and determine binding lifetimes lasting for several tens of minutes. Despite this potential, currently available guidelines for the selection of binding models are unreliable, and the practical implementation of this approach is limited. Here, using experimental and simulated data we identify the minimum size of the dataset required to resolve multiple populations reliably and measure binding lifetimes with desired accuracy. This work serves to provide a guide to data collection, and measurement of DNA-binding lifetimes from single-molecule time-lapse imaging data.

INTRODUCTION

Understanding fundamental processes of life requires characterization of the kinetics of interactions between biological molecules. At single-molecule levels, these systems often exhibit kinetic heterogeneity that is inherent to the presence of multiple intermediate states (1-17). Advances in single-molecule imaging have enabled the detection and characterization of heterogeneous sub-populations in reactions conducted *in vitro* as well as, *in vivo*. Ultimately, these investigations enable the construction of detailed molecular mechanisms to explain how various biomolecular interactions proceed.

Compared to *in vitro* studies, live-cell investigations offer the key advantage of studying biochemical reactions at physiological conditions that can be difficult to reconstitute. Single-molecule live-cell imaging commonly relies on fluorescent proteins that are genetically fused to the protein of interest (Fig. 1A) (18-22). Tracking the fluorescence signal of thousands of molecules, one molecule at a time, enables the building of physical models, from which physical parameters such as diffusion constants and detachment rates from DNA can be determined. Where detachment rates are concerned, the trajectory lengths of thousands of molecules are aligned to obtain a cumulative residence time distribution (CRTD). At the single-molecule level, the dissociation of a protein from its substrate is a stochastic process. This phenomenon can be adequately described as a two-state kinetic model with the interconversion of populations being modelled as a Poisson process. The resulting CRTD can be fit to exponential functions to obtain decay rates. In the case of a fluorescently tagged protein where loss of fluorescence is attributable to either dissociation, or photobleaching of the chromophore, the decay rate represents a combination of dissociation rates and photobleaching rate (Fig. 1B-C) (23).

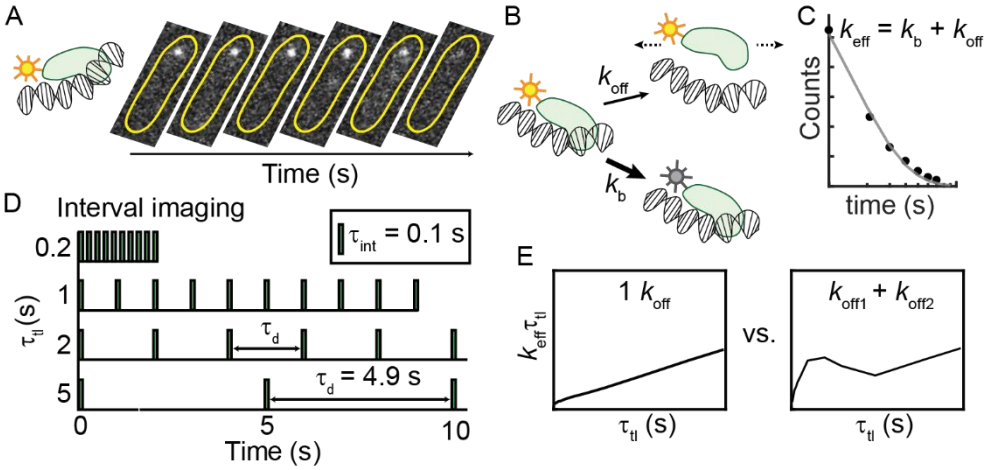


FIGURE 1 Experimental approach for characterizing kinetic heterogeneity of protein binding in live cells using single-molecule fluorescence imaging. (A) The protein of interest is tagged with a fluorescent protein. When the protein binds to DNA substrate, its fluorescence signal appears as a diffraction-limited focus that can be tracked in real time. Subsequent dissociation results in the disappearance of the focus and a redistribution of fluorescence signal throughout the cell. Yellow outlines illustrate the bacterial cell membrane. (B) The loss of fluorescence is attributable to either dissociation, or photobleaching of the chromophore. (C) Cumulative residence time distribution (CRTD) constructed from binding durations of thousands of molecules. Fitting the exponential function (Eq. 1) to CRTD yields an effective rate k_{eff}, which is the sum of off rate (k_{off}) of the protein of interest and photobleaching rate (k_b) of the fluorescent probe (23). (D) To deconvolute k_b and k_{off}, excitation and integration durations (τ_{int}) can be spaced with various dark intervals (τ_d). (E) Through exponential analyses, CRTDs obtained at various intervals result in $k_{eff}\tau_{tl}$ plots which are indicators of kinetic heterogeneity (23). A single kinetic population yields a straight line whereas deviations from linear fits indicate the presence of a second kinetic sub-population. For a single kinetic population, the slope is the off rate and y-intercept is proportional to the photobleaching rate.

Photobleaching, a result of fluorescent proteins being damaged upon exposure to excitation sources, leads to the loss of fluorescence signal (24). Under excitation conditions that guarantee good signal-to-background ratios, fluorescent proteins can only stay 'on' for a few frames during continuous acquisitions. This limited visualization window reflects the 'photon budget' (25). Thus, when photobleaching occurs faster than the dissociation process, lifetime measurements are limited by the photobleaching rate. To overcome this problem and extend the observation time, the observation time window can be expanded by temporally spacing the photon budget using stroboscopic imaging (26). In this method, a dark interval (τ_d) is inserted

between integration time (τ_{int}), effectively scaling the observation time with a factor of τ_{tl}/τ_{int} ($\tau_{tl} = \tau_{int} + \tau_d$). Instead of using one dark interval, Gebhardt and co-workers (2013) developed an approach involving 'time-lapse illumination with a fixed integration time, interspersed with dark periods of varying duration' in which fluorescence acquisitions are collected at a series of time-lapse intervals (Fig. 1D) (23, 27). This method has also been variously referred to as 'time-lapse imaging' (28), 'time-lapse illumination with different dark times' (29), 'time-lapse imaging at multiple timescales' (30) and 'stroboscopic single particle tracking PALM' (31). For the purpose of brevity, and to distinguish from a time-lapse imaging mode with a single dark interval, we have adopted the term 'interval imaging' in our lab (32). Briefly, the approach works as follows: First, several movies (each with a unique dark interval) are collected while keeping the photon budget constant (in practice this is achieved by keeping the number of frames constant across all the movies). In cases where the copy number of the tagged protein is high and single-molecule imaging conditions may be difficult to attain, the cellular fluorescence is first photobleached such that only single-molecule fluorescence is observable. Subsequently, using particle tracking algorithms that enable measurements of lifetimes of bound molecules within a specified localization radius, a CRTD can be compiled. Fitting the CRTDs to effective rates (k_{eff}), one can obtain the so-called $k_{eff}\tau_{tl}$ plot which is linear for mono-exponential distributions (Fig. 1E) (23). In this case, since the photobleaching rate is maintained constant across all conditions, it can be read off from the intercept on the Y-axis. A population of molecules dissociating with a finite and measurable off rate manifests as a straight line, where the slope reports on the off rate of the dissociation kinetics. A mixed population composed of species dissociating with multiple lifetimes manifests as a deviation from the linear fit (Fig. 1E) (23). Fitting the experimental data to a model describing mixed populations can then be used to extract the relative amplitudes and rates of the various populations. This power to deconvolute the photobleaching rate from multiple off rates has

been successfully harnessed to dissect the kinetic heterogeneity of various DNA binding proteins including transcription factors and DNA replication and repair proteins in live cells (23, 27-33).

However, limitations arising from the practical implementation of this elegant method remain uncharacterized. In particular, we address the following questions: 1) What is the minimum number of observations needed to determine the binding lifetime of a species within a specified confidence? 2) For a given experimental setup, what is the dynamic range in binding lifetimes that can be detected? 3) How many populations can be resolved? and 4) What limits the ability to reliably resolve multiple populations? We consider four cases below to answer these questions. This study serves to provide a practical guide to realize the power as well as limitations of practical implementations of the interval imaging approach to measure intracellular binding kinetics of fluorescently tagged proteins.

METHODS

Rationale and model

For an introduction to the method, we direct the reader to seminal work by Gebhardt and co-workers who have developed and demonstrated the time-lapse imaging approach discussed here (23). Here, we first summarize the theoretical development to establish the context of the problem for this report. Consider a system containing '*A*' number of fluorescently tagged DNA-bound proteins, wherein the proteins dissociate from DNA with a single off rate (k_{off}). Upon exposure to excitation photon sources, the fluorescent proteins exhibit photobleaching with a rate k_b, resulting in the loss of fluorescence signal. Additionally, dissociation contributes to the loss of fluorescent foci as protein molecules move out of the localization radius. Since dissociation and photobleaching are independent, and both are Poisson processes, the loss of observations as a function of time *t* can be described as:

$$f_1(t) = A\exp(-(k_\mathrm{b} + k_\mathrm{off})t) \qquad \text{(1, ref. (23))}$$

Observation times of genetically expressible fluorescent proteins are severely limited to the duration of a few acquisition frames due to photobleaching, limiting measurements of long-lived binding events (34). To extend observation times, the frame rate can be reduced by inserting a dark interval (τ_d) after a short integration time (τ_int). Scaling the photobleaching rate appropriately, Eq. 1 then becomes:

$$f_2(t) = A\exp(-(k_\mathrm{b}\tau_\mathrm{int}/\tau_\mathrm{tl} + k_\mathrm{off})t) \qquad \text{(2, ref. (23))}$$

where the time-lapse time τ_tl is the sum of τ_int and τ_d. The sum of two decay rates k_b and k_off can be approximated with an effective decay rate (k_eff):

$$k_\mathrm{eff} = k_\mathrm{b}\tau_\mathrm{int}/\tau_\mathrm{tl} + k_\mathrm{off} \qquad \text{(3, ref. (23))}$$

Rearrangement of Eq. 3 yields:

$$k_\mathrm{eff}\tau_\mathrm{tl} = k_\mathrm{b}\tau_\mathrm{int} + k_\mathrm{off}\,\tau_\mathrm{tl} \qquad \text{(4, ref. (23))}$$

As $k_\mathrm{b}\tau_\mathrm{int}$ is maintained constant at a certain imaging condition, $k_\mathrm{eff}\tau_\mathrm{tl}$ increases linearly with τ_tl, with the coefficient (slope) k_off.

In systems with two sub-populations each dissociating at different rates k_off1 and k_off2, Eq. 2 then becomes:

$$f_3(t) = A\big(B\exp(-(k_\mathrm{b}\tau_\mathrm{int}/\tau_\mathrm{tl} + k_\mathrm{off1})t) + (1-B)\exp(-(k_\mathrm{b}\tau_\mathrm{int}/\tau_\mathrm{tl} + k_\mathrm{off2})t)\big) \qquad \text{(5, ref. (23))}$$

151 where B ($0 < B < 1$) and $(1 - B)$ are the amplitudes of k_{off1} and k_{off2} sub-populations respectively.

152 Similarly, a system with three kinetic sub-populations can be described by:

$$f_4(t) = A\big(B_1\exp(-(k_b\tau_{int}/\tau_{tl} + k_{off1})t) + B_2\exp(-(k_b\tau_{int}/\tau_{tl} + k_{off2})t) + (1 - B_1 - B_2)\exp(-(k_b\tau_{int}/\tau_{tl} + k_{off3})t)\big) \quad \text{(6, ref. (29))}$$

153 where B_1, B_2 ($0 < B_1, B_2 < 1$ and $B_1 + B_2 < 1$) and $(1 - B_1 - B_2)$ represent the amplitudes of k_{off1},

154 k_{off2} and k_{off3} sub-populations respectively.

155

Experimental considerations

The specifics of the experimental setup for different model organisms should be tailored to requirements for the respective system. However, to provide the reader with a starting point, we describe the experimental configuration used in our lab to measure binding lifetimes of DNA-repair proteins labelled with the fluorescent protein, YPet in the model organism *Escherichia coli* (*E. coli*) (see Fig. S1 and ref. (32)). Bacterial cells in early exponential phase are loaded into a custom-built flow cell made up a glass coverslip and a quartz top. The bottom coverslip is functionalized with (3-Aminopropyl)triethoxysilane (APTES, Alfa Aesar, USA) to facilitate cell adhesion to the surface of the coverslip. The temperature of the flow cell is kept constant at 30 °C. Cells are supplied with aerated rich defined media (EZ rich defined medium supplemented with glucose, Teknova) to maintain fast growth. YPet is excited with 514-nm laser (Sapphire LP laser, Coherent, USA) in near-TIRF configuration (35) at a power density of 71 W/cm^2 (measured directly above the inverted objective). Fluorescent signal is recorded using an electron-multiplying (EM)-CCD camera (Photometrics Evolve, Photometrics, USA), with an EM gain of 1,000. The camera exposure time is 0.1 s and time-lapse imaging is acquired with a 10-s τ_{tl} set (Table S2). Typically, a time-lapse imaging experiment lasts three to five hours and in generally four to ten experiments are required to obtain more than a thousand binding events at each τ_{tl}.

Resolution of binding events in bacterial cells expressing copy numbers of fluorescent proteins in excess of ~ 20 copies per cell is challenging due to the limitations of particle tracking algorithms to resolve closely spaced foci. Further, distinguishing bound molecules from freely diffusive molecules in the cytosol is also challenging when copy numbers are high. In this case, to enable reliable observation of single- molecule cells are exposed to continuous illumination such that the majority of the emitters are darkened or photo-bleached, and only stochastically

reactivated emitters are observed in single-molecule imaging conditions (36).

This setup allows us to unambiguously detect single-molecule foci using a relative signal-to-background ratio between six and eight. Foci detected in at least two consecutive frames within a 300-nm (3 pixels) radius are defined as a binding event. For each τ_{tl}, all binding events are combined, and bootstrapping analysis is performed by randomly selecting with replacements 80% of all binding events. CRTDs are constructed from bootstrapped samples and are fit to exponential models to obtain $k_{eff}\tau_{tl}$ plots, as well as k_b and τ.

Simulating concurrent dissociation and photobleaching

In order to maintain full control of the kinetic variables, we chose to perform simulations of the experiment. Simulations of exponential distributions and curve-fitting were performed with custom-written program in MATLAB (The MathWorks, Natick, MA). We simulated exponential distributions (Eq. 2,5,6) using the *exprnd* function in MATLAB (Supplementary Notes). This function generates exponentially distributed random numbers with a specified decay constant. Here, each number returned by *exprnd* function represents the lifetime of a simulated 'trajectory'. For the purposes of this work, we have not accounted for blinking of bound molecules that may yield prematurely truncated binding events. Accommodation of such a feature will require reasonable estimates of FP blinking under the conditions of the experiment that will be unique to the fluorescent probe used. To simulate a sub-population of molecules dissociating with a specified off rate, a set of trajectories was generated and binned to produce histograms with ten bins, whose edges correspond to frame times (integer multiples of τ_{tl}). The *exprnd* function was iterated until the counts of the first bin exceeded the number of binding events in that sub-population (typically between three and six iterations, see Fig. S2). To simulate experiments where multiple sub-populations are present, each sub-population was simulated in defined proportions and all trajectories were pooled together. Finally, to generate the CRTDs, we rejected molecules in the first bin (0 to τ_{tl}) and only carried forward observations from τ_{tl} to $10\tau_{tl}$ to the next step in accordance with our definition of a binding event (or trajectories), *i.e*, the observation must be present in two consecutive frames.

To simulate uncertainty in each simulation sample, ten rounds of bootstrapping were performed, each involved randomly sampling 80% of the simulated population. Next, fitting was performed on each bootstrapped CRTDs (henceforth referred simply as CRTDs). First, the CRTD at each τ_{tl} was fit to a mono-exponential model to obtain k_{eff} (Eq. 2, 3 and Fig. 1C).

These values for k_{eff}, corresponding to the number of τ_{tl}, were then used to construct the $k_{eff}\tau_{tl}$ plot. Error bands in these plots represent standard deviations from ten bootstrapped samples.

Second, the CRTDs for all τ_{tl} were fit to objective functions based on Eq. 2, 5 and 6 (global fitting, see Supplementary Notes). The list of parameters, initial conditions, bound constraints, termination criteria and algorithm is presented in Table S1. Throughout the paper, A was set as a local parameter to mimic experimental conditions where counts may be different across τ_{tl}, even though this often leads to less accurate results compared to when A was set as a global parameter (Fig. S3B-C).

For each simulation, outcomes from globally fitting the ten bootstrapped CRTDs were averaged and reported. To determine uncertainty in the estimate, we repeated the simulation a hundred times. The standard deviations of the binding lifetime (σ_τ) from a hundred simulations using the same conditions was calculated according to Eq. 7.

$$\sigma_\tau = \left(\sum_{i=1}^{100} (\tau_i - \langle\tau\rangle)^2 \Big/ 100 \right)^{1/2} \quad (7)$$

where $<\tau>$ denotes the true binding lifetime, which is calculated by $1/< k_{off}>$.

Unless otherwise stated, $k_b\tau_{int}$ was fixed at 0.7 to mimic experimental values obtained in our published work (32). Four sets of τ_{tl} were used: 10-s τ_{tl}, 100-s τ_{tl}, the three- and five- τ_{tl} sets (Table S2).

RESULTS

Influence of experimental sample size on uncertainty of the estimate of the binding lifetime

First, we set out to investigate whether the size of the experimental data set influences the uncertainty in the error estimate of the outcomes from global fitting, such as the binding lifetime τ and photobleaching rate k_b. This can be achieved by randomly selecting a fraction of experimental data (3%-30%) at each τ_{tl}, following by bootstrapping and global fitting. Toward this goal, we revisited published data from our laboratory where interval imaging was used to determine dissociation kinetics of the transcription-repair coupling factor Mfd from DNA in live *E. coli* (32). The entire dataset (100%) contains between 1,000 to 2,000 trajectories (counts lasting at least two frames) at each τ_{tl} (Fig. 2A, right-most panel). Representative CRTDs following sub-sampling the experimental dataset (3%, 10% and 30%) at each τ_{tl} are shown in (Fig. 2A). While the $k_{eff}\tau_{tl}$ plot derived from the whole dataset resembles a straight line, deviations from linear fits in $k_{eff}\tau_{tl}$ plots can be seen when only a sub-set of experimental data was used (Fig. 2B).

To determine the uncertainties in k_b and τ as a result of under-sampling, we repeated the sub-sampling a hundred times and k_b and τ values were obtained from global fitting using Eq. 2 (Fig. 2C-D). Here, uncertainties in the estimates of k_b and τ are smallest when the entire data set is used (2% and 5% respectively, Fig. 2C-D), and as expected, increase with decreasing number of counts (Fig. 2C-D). For k_b, uncertainties increase from 3% to 10% as the percentage of experimental data drop from 30% to 3% while uncertainties in determining binding lifetimes increase from 8% to 35% (Fig. 2C-D).

Fitting individual CRTD to mono-exponential model to obtain $k_{eff}\tau_{tl}$ plots has been suggested

to be used as a guide to determine kinetic heterogeneity (23). Our analysis demonstrates that deviation from linear fits in the $k_{eff}\tau_{tl}$ plots can potentially simply reflect under-sampling. Since deviations from linear fits in $k_{eff}\tau_{tl}$ plots can also be used to guide the choice of bi- and tri-exponential models (23), a fundamental question that faces users is, what governs the choice of exponential model? What is the minimum size of data, for which a multi-exponential model is appropriate for consideration? Are deviations in the $k_{eff}\tau_{tl}$ plots reliable indicators for the choice of model? To explore these questions in greater detail, we chose to perform simulations that permit us to retain full control of the model parameters, and overcome practical limitations of generating large data sets from microscopy experiments.

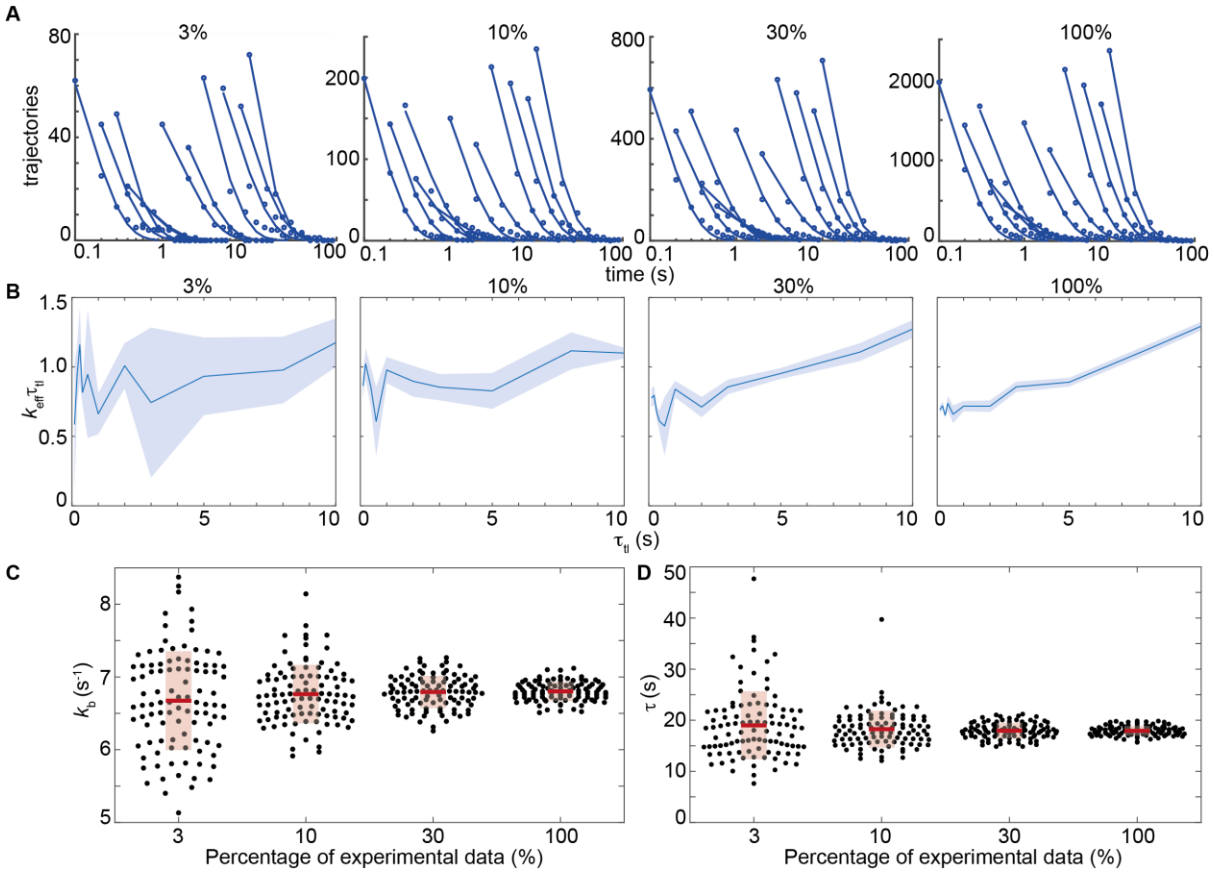
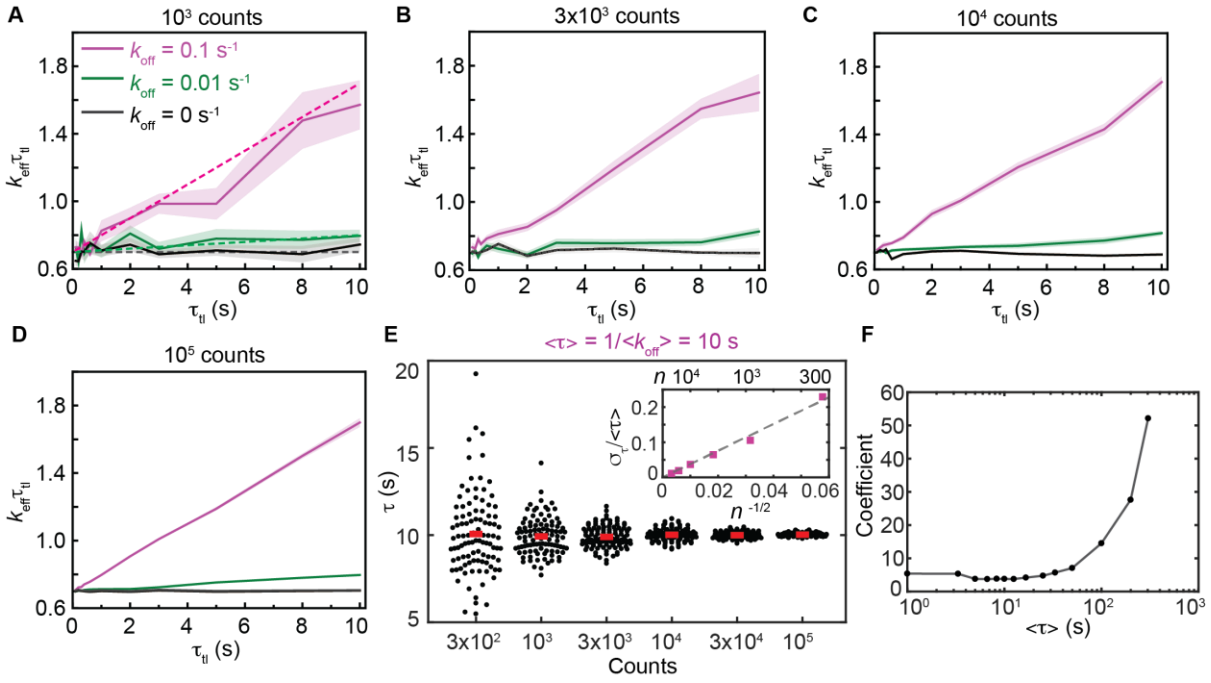


FIGURE 2 Determination of the photobleaching rate and binding lifetime from sub-sampling experimental data presented in ref (32). (A) Representative CRTDs when only 3%, 10% or 30% of experimental trajectories were randomly selected. Counts can be approximated as y-intercepts of exponential fits of CRTDs. The CRTD from the full dataset (right most panel) is

reproduced from ref. (32). (B) $k_{eff}\tau_{tl}$ plots of the corresponding CRTDs (above). Shaded error bands are standard deviations from ten bootstrapped samples. (C) Scatter plots show distributions of k_b obtained using global fitting 100 subsets of the experimental data at the indicated fraction. Each point represents the average of results from ten bootstrapped samples. (D) Scatter plots show distributions of τ obtained using global fitting 100 subsets of the experimental data at the indicated fraction. Similarly, each point represents the average of results from ten bootstrapped samples. Red bars and boxes represent means and standard deviations of the fitting outcomes of 100 subsets of the dataset respectively. The experimentally measured value of τ = 17.9 ± 0.9 s for the entire data set is reproduced from ref (32).

Case I: Influence of the size of the data set on the measured lifetime for a single dissociating species

We first explored the relationship between the number of counts (n) at each τ_{tl} and uncertainties in estimates of binding lifetimes from mono-exponential distributions. To this end, we simulated a population of molecules dissociating with k_{off} of 0.1 s^{-1}, corresponding to a binding lifetime $<\tau>$ of 10 s, and photobleaching rate k_b of 7 s^{-1} (see Methods). While τ_{int} was constant at 0.1 s, τ_{tl} was varied from 0.1 s to 10 s (Table S2). These values of k_b, $<\tau>$, τ_{int} and τ_{tl} were initially chosen to closely match experimental values used in our published work (see Fig. 2 and ref. (32)). The theoretical $k_{eff}\tau_{tl}$ plot is shown as the dashed line (Fig. 3A). At n = $1x10^3$ observations (Fig. 3A), the $k_{eff}\tau_{tl}$ plot deviates noticeably from the theoretical line (purple curve). However, as n increases, the error bands reduce, and the plots closely resemble straight lines (purple curves, Fig. 3B-D). At $1x10^5$ observations, linearly fitting the $k_{eff}\tau_{tl}$ plot (Fig. 3D) yielded a slope of 0.1 and y-intercept of 0.6992, reflecting the specified k_{off} (0.1 s^{-1}) and $k_b\tau_{int}$ (0.7). As expected, mono-exponential distributions with the same $k_b\tau_{int}$ but smaller off rate (k_{off} = 0.01 s^{-1}) or without off rate (k_{off} = 0 s^{-1}) yielded lines with smaller slope (Fig. 3A-D, green curves) or essentially flat lines (Fig. 3A-D, black curves).



290

291 FIGURE 3 Determination of binding lifetimes from mono-exponential distributions. (A-D)
292 $k_{eff}\tau_{tl}$ plots of mono-exponential distributions with $k_b\tau_{int}$ of 0.7 and k_{off} of 0.1 s^{-1} (purple curves),
293 0.01 s^{-1} (green curves) or 0 s^{-1} (black curves). Panels A-D reflect $k_{eff}\tau_{tl}$ plots obtained from
294 simulations containing number of observations (n) equaling (A) $1x10^3$, (B) $3x10^3$, (C) $1x10^4$
295 or (D) $1x10^5$ counts in the first bin (see Methods). (A) Dashed lines correspond to theoretical
296 $k_{eff}\tau_{tl}$ plots at the specified k_{off} values. Shaded error bands are standard deviations from ten
297 bootstrapped samples. (E) Scatter plots show distributions of τ obtained using global fitting
298 from 100 simulated samples for each n value. Red bars represent the mean values. (Inset) The
299 relative error in determining τ (σ_τ /$<\tau>$) reduces with $n^{-1/2}$ for increasing n. Dashed line is the
300 linear fit to six data points. (F) Coefficient in function of $\sigma_\tau/<\tau>$ versus $n^{-1/2}$ at various $<\tau>$.
301 The sharp increase in coefficients for $<\tau>$ larger than 50 s indicates larger uncertainties in
302 measuring slow processes when the maximum τ_{tl} is limited to 10 s.

303

To characterize the uncertainty (standard deviation, σ_τ) in the estimate of the binding lifetime, we repeated the simulation a hundred times for each value of *n* and determined τ using global fitting (Fig. 3E). As expected for shot noise (37), the relative error $\sigma_\tau/\langle\tau\rangle$ is proportional to the inverse of the square root of *n* with a coefficient of 3.8 (Fig. 3E, inset). Importantly, the coefficient fluctuates between 3.7 and 5.7 for $\langle\tau\rangle \leq 50$ s, but rises sharply for $\langle\tau\rangle$ greater than 50 s (Fig. 3F). This result demonstrates that the uncertainty in estimating lifetime of long-lived binding events becomes arbitrarily large when the extended lifetime of the fluorophore (by introduction of τ_d) becomes comparable to the binding lifetime. In principle, this limit can be readily overcome by simply selecting larger τ_{tl} values; indeed, simulations of mono-exponential distributions of long-lived binding events (τ = 100 s) indicated that σ_τ is lower at lower values of *n*, when τ_{tl} is extended to 100 s, compared to 10 s (Fig. S3).

Therefore, we propose that accurate measurements of lifetime of long-lived binding events require significant increases in either the number of observations (*n*) or the length of τ_{tl} for a fixed photobleaching rate. However, it should be noted that extension of τ_{tl} up to 100 s may not be experimentally feasible for all systems. In our work involving bacterial live-cell imaging in rich media, cell growth and division on the timescale of imaging limit the tracking binding events lasting on the timescale of tens of minutes. Practical limitations imposed by the model organism, growth conditions and choice of fluorescent protein dictate optimal experimental design.

Further, we anticipated that photobleaching rate also contributes to σ_τ as faster photobleaching reduces observation times. To examine the effect of $k_b\tau_{int}$, we performed a comprehensive set of simulations with the 10-s τ_{tl} set (Table S2) and $k_b\tau_{int}$ varying from 0.007 to 2.1 (k_b from 0.07 to 70 s^{-1} and τ_{int} from 0.01 to 0.1 s). We obtained the relationship between σ_τ/τ, *n* and $k_b\tau_{int}$ as

in Eq. 8.

$$\frac{\sigma_\tau}{\tau} = \frac{(2.7379 k_b \tau_{int})^2}{n^{1/2}} \quad (8)$$

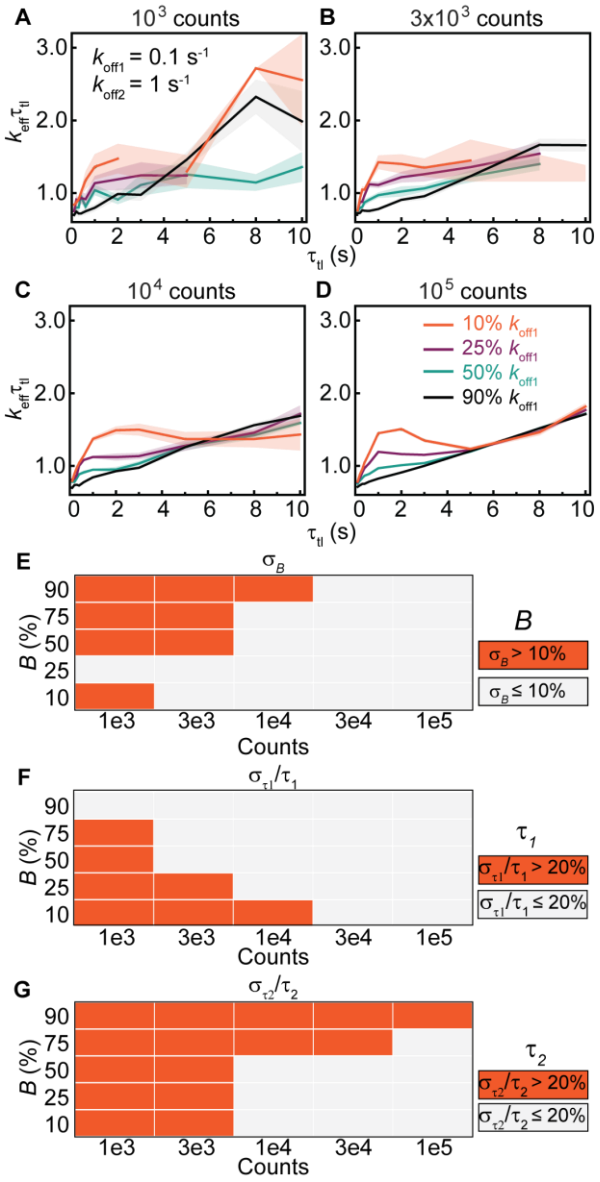
This formula describes the lower bound of errors as other sources of practical errors, such as localization uncertainties and experimental variations, have not been considered. The minimum number of observations required to determine τ ($<\tau> \leq 50$ s) with a given uncertainty is therefore:

$$n = (\tau/\sigma_\tau)^2 \times (2.7379 k_b \tau_{int})^4 \quad (9)$$

For example, when $k_b\tau_{int}$ is 0.7, the number of observations required to achieve relative error of 10% in the estimate of τ (where $<\tau> \leq 50$ s) is about 1350 (see Fig. 3E). This equation also highlights the importance of using fluorophores with high photo-stability: a two-fold increase in k_b needs to be compensated by a 16-fold increase in n.

Case II: Detection of two species with resolvable lifetimes

Next, we examined the situation where a second kinetic sub-population is present in the system. A second population with a faster off rate yields $k_{eff}\tau_{tl}$ plots that deviate from straight lines (23). However, as we demonstrate deviations can also be a result of shot noise at low n (see Fig. 2B and Fig. 3A-D). To identify the minimum n at which one can determine with a specified confidence that a bi-exponential model is appropriate, we simulated CRTDs using Eq. 5. First, we performed simulations with off rates that are an order of magnitude apart: k_{off1} = 0.1 s^{-1} (intermediate rate) and k_{off2} = 1 s^{-1} (fast rate). The amplitude B of the intermediate dissociating population was varied from 10% to 90% (Fig. 4).



346

347 FIGURE 4 Determination of binding lifetimes and amplitudes from bi-exponential
348 distributions with an intermediate rate (k_{off1}) and a fast rate (k_{off2} = 10 k_{off1}). (A-D) $k_{eff}\tau_{tl}$ plots
349 of bi-exponential distributions with $k_b\tau_{int}$ of 0.7, k_{off1} and k_{off2} of 0.1 and 1.0 s^{-1} respectively,
350 with (A) 10^3, (B) $3x10^3$, (C) 10^4 or (D) 10^5 observations. The amplitude of k_{off1} (B) is 10%
351 (orange), 25% (purple), 50% (green) or 90% (black). Shaded error bars are standard deviations
352 from ten bootstrapped samples. (E-G) Heatmaps show errors in estimates of *B*, τ_1 and τ_2
353 obtained using global fitting of 100 simulated distributions for each *n* value (see Fig. S4 for
354 distributions).

355 When the majority of the population dissociates with the intermediate rate k_{off1} (*B* = 90%), the
356 $k_{eff}\tau_{tl}$ plots resemble those of mono-exponential distribution with the single k_{off} of 0.1 s^{-1}
357 (compare Fig. 4A-D, black curves and Fig. 3A-D). As before, increasing the number of

observations significantly improved the quality of the $k_{eff}\tau_{tl}$ plots (Fig. 4A-D). These simulations reveal that a short-lived second sub-population does not manifest as a visible feature in the $k_{eff}\tau_{tl}$ plots when it is present only to the extent of 10% in the observations. To examine if the two populations could be resolved with global fitting using the bi-exponential model, we determined binding lifetimes and amplitudes from 100 simulations (Fig. S4). Unsurprisingly, we found that the accuracies and precisions of determining B, τ_1 and τ_2 increase with n. While estimation of τ_1 is robust (Fig. 4F, Fig. S4B), global fitting of CRTDs to the bi-exponential model at low counts suffers from a bias towards the fast dissociating sub-population, with its amplitude being overestimated and τ_2 being underestimated (Fig. 4E, G, Fig. S4A, C). This bias is observed to a lesser extent when k_{off1} is present at 75% or 50% (Fig. 4E-G, Fig. S4).

As the amplitude of the fast dissociating sub-population increased (B equal to 25% or 10%), fewer observations were found at long intervals. Insufficient counts resulted in missing data points at these τ_{tl} ($\tau_{tl} \geq 5$ s) in $k_{eff}\tau_{tl}$ plots at low counts (10^3 and $3x10^3$, Fig. 4A-B). However, the $k_{eff}\tau_{tl}$ plots extended to the full τ_{tl} range of 10 s when n increases to 10^4 and 10^5 (Fig. 4C-D). As expected, deviations from straight lines were found in the 0-5 s regime, reflecting the presence of the fast dissociating sub-population. Since contributions from the fast dissociating sub-population drop sharply at long timescales, the $k_{eff}\tau_{tl}$ plots converge to the straight line exhibited by mono-exponential distributions with k_{off1} (Fig. 4C-D). Further analysis by integrating the area under the peaks in the 0-to-5 s region shows the area increases exponentially with the amplitude of the fast dissociating sub-population (Fig. S5). When the fast dissociating sub-population represents the majority, the accuracy and precision in determining B, τ_1 and τ_2 also increase with n (Fig. 4E-G, Fig. S4).

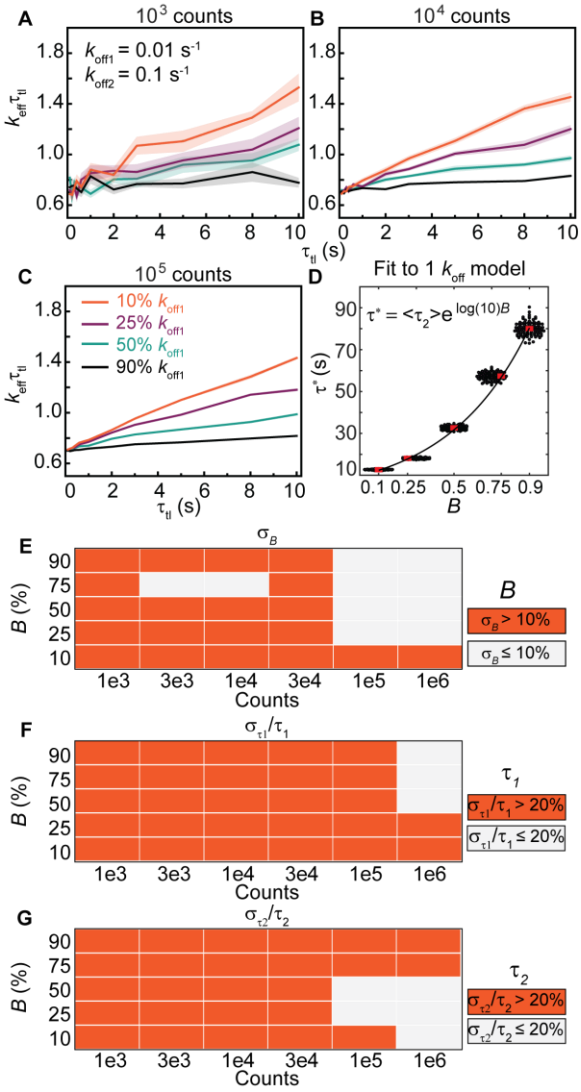
Based on the observation that accurate measurements of long-lived binding events require the

extension of τ_{tl} to greater than 10 s, we anticipated that resolving two kinetic sub-populations [one with a slow rate (k_{off1} of 0.01 s^{-1}; $<\tau_1>$ = 100 s) and an intermediate rate (k_{off2} of 0.1 s^{-1}; $<\tau_2>$ = 10 s)] is challenging when the largest τ_{tl} is 10-s. Consistent with this, the $k_{eff}\tau_{tl}$ plots in the 0-10s range appear linear (Fig. 5A-C), resembling those of mono-exponential distributions. Hence, we attempted to fit the CRTDs at 10^5 counts to mono-exponential model (Eq. 2), yielding apparent binding lifetimes (τ^*) that lie between $<\tau_1>$ and $<\tau_2>$ (Fig. 5D). Fitting mean τ^* vs. *B* to exponential function results in Equation 10:

$$\tau^* = \langle\tau_2\rangle e^{log(\langle\tau_1\rangle/\langle\tau_2\rangle)B} \quad (10)$$

Thus, *B* can be derived from τ^* where $<\tau_1>$ and $<\tau_2>$ are known.

From the simulations, fitting the CRTDs with *n* less than $3x10^4$ to the bi-exponential model yields unreliable results (Fig. 5E-G, Fig. S6). Across various amplitudes of k_{off1}, the species with lifetime τ_1, is often underestimated and corresponds to τ^* at that amplitude (compare Fig. 5F to Fig. 5D). Similarly, τ_2 is also underestimated, but eventually approaches $<\tau_2>$ of 10 s when *n* reached 10^6 counts and the amplitude of k_{off2} sub-population is more than 25% (Fig. 5G).



396

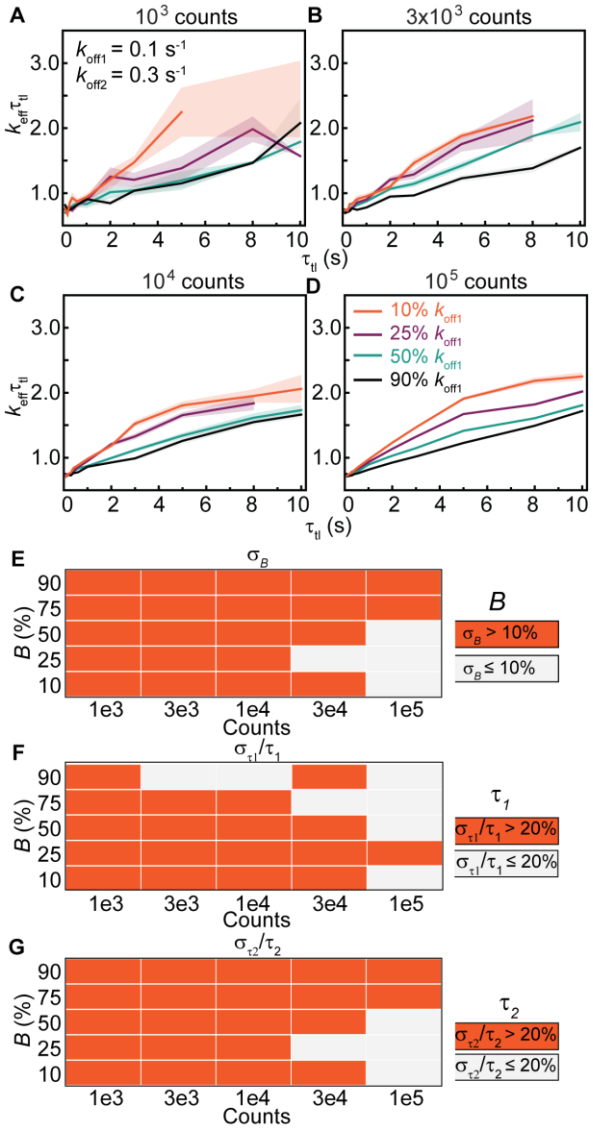
397 FIGURE 5 Determination of binding lifetimes and amplitudes from bi-exponential
398 distributions with a slow rate (k_{off1}) and an intermediate rate (k_{off2} = 10k_{off1}). (A-C) $k_{eff}\tau_{tl}$ plots
399 of bi-exponential distributions with $k_b\tau_{int}$ of 0.7, k_{off1} and k_{off2} of 0.01 and 0.1 s^{-1} respectively,
400 with (A) 10^3, (B) 10^4 or (C) 10^5 observations. The amplitude of k_{off1} (B) is 10% (orange), 25%
401 (purple), 50% (green) or 90% (black). Shaded error bars are standard deviations from ten
402 bootstrapped samples. (D) Scatter plots show distribution of apparent τ (τ^*) obtained from
403 fitting of 100 simulated bi-exponential distributions at a specified *B* and 10^5 counts to mono-
404 exponential model. Line is exponential fit between the average of τ^* (red bars) and *B*. (E-G)
405 Heatmaps show errors in estimates of *B*, τ_1 and τ_2 obtained using global fitting of 100 simulated
406 distributions for each *n* value (see Fig. S6 for distributions).

407

On the other hand, when the above distributions were simulated using the 100-s τ_{tl} set, deviations from straight lines in $k_{eff}\tau_{tl}$ plots were observed in the 0-30s regime and when B is smaller than 75% (Fig. S7A). In this case, as expected, accuracies in determining B, τ_1 and τ_2 follow the same trends as discussed in Fig. 4 (Fig. S7B-D).

Case III: Detection of two species with closely matched lifetimes

Due to the resolution limit that is inherent to exponential analysis (34), we anticipated the ability to resolve rates that are closely spaced would reduce. To test this hypothesis, we simulated bi-exponential distributions with rates that are only three-fold apart: an intermediate rate k_{off1} of 0.1 s^{-1} and a fast rate k_{off2} of 0.3 s^{-1}. Under conditions that yield sufficient observations at long intervals ($n \geq 10^4$), examination of the $k_{eff}\tau_{tl}$ plots often fails to identify the presence of multiple sub-populations in the form of deviation from straight lines (Fig. 6A-D). Only when the fast rate is present at 90%, can deviations be observed in the form of a broad convex spanning from 0 to 10 s (orange curves, Fig. 6C-D). Fitting to Eq. 5 yields unreliable results for B and τ_2 for $n \leq 10^4$ (Fig. 6E, G, Fig. S8A, C) whereas the accuracy in determining τ_1 requires 3 x10^3 observations or $<B>$ to be larger than 25% (Fig. 6F, Fig. S8B). Fitting CRTDs at low counts ($n \leq 10^4$) to the bi-exponential model should be avoided as one often obtains two kinetics sub-populations with artificially enhanced rate separation and substantial amplitudes, regardless of the true amplitudes (Fig. 6E-G, Fig. S8).



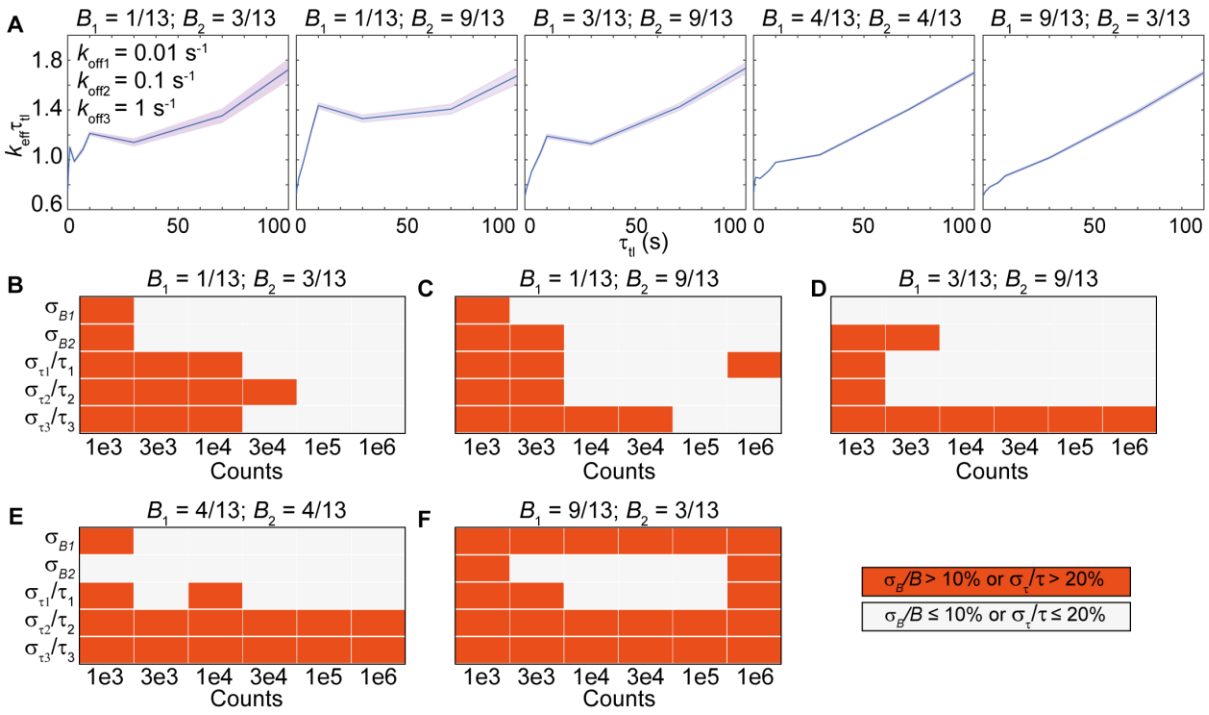
426

427 FIGURE 6 Determination of binding lifetimes and amplitudes from bi-exponential
428 distributions with closely spaced rates (k_{off2} = 3k_{off1}). (A-D) $k_{eff}\tau_{tl}$ plots of bi-exponential
429 distributions with $k_b\tau_{int}$ of 0.7, k_{off1} and k_{off2} of 0.1 and 0.3 s^{-1} respectively, with (A) 10^3, (B)
430 $3x10^3$, (C) 10^4 or (D) 10^5 observations. The amplitude of k_{off1} (*B*) is 10% (orange), 25%
431 (purple), 50% (green) or 90% (black). Shaded error bars are standard deviations from ten
432 bootstrapped samples. (E-G) Heatmaps show errors in estimates of *B*, τ_1 and τ_2 obtained using
433 global fitting of 100 simulated distributions for each *n* value (see Fig. S8 for distributions).

434

Case IV: Detection of three species

The resolution limit as well as dynamic range limit that we demonstrated above raise the question if tri-exponential distributions can be faithfully resolved under the specified experimental condition (ranges of τ_{tl} and n). To address this issue, we simulated tri-exponential distributions (Eq. 6), with off rates spanning two orders of magnitude (0.01, 0.1 and 1 s^{-1}), using the 100-s τ_{tl} set. The diversity in $k_{eff}\tau_{tl}$ plots obtained by varying B_1 and B_2 is illustrated in Fig. 7A. Three kinetic sub-populations are apparent when B_1 is a third of B_2 and B_2 in turn is a third of B_3 ($1 - B_1 - B_2$). We further characterized uncertainties in amplitudes and binding lifetimes obtained using global fitting to the tri-exponential model (Methods). In general, accuracy in determining the amplitudes and lifetimes improves with increasing n (Fig. 7B-F, Fig. S9). However, when the slowly dissociating sub-population dominates (B_1 = 9/13), increasing n does not yield more accurate estimates. As in the case of the bi-exponential simulations, we observed consistent biases towards faster binding lifetimes (Fig. S9).



449

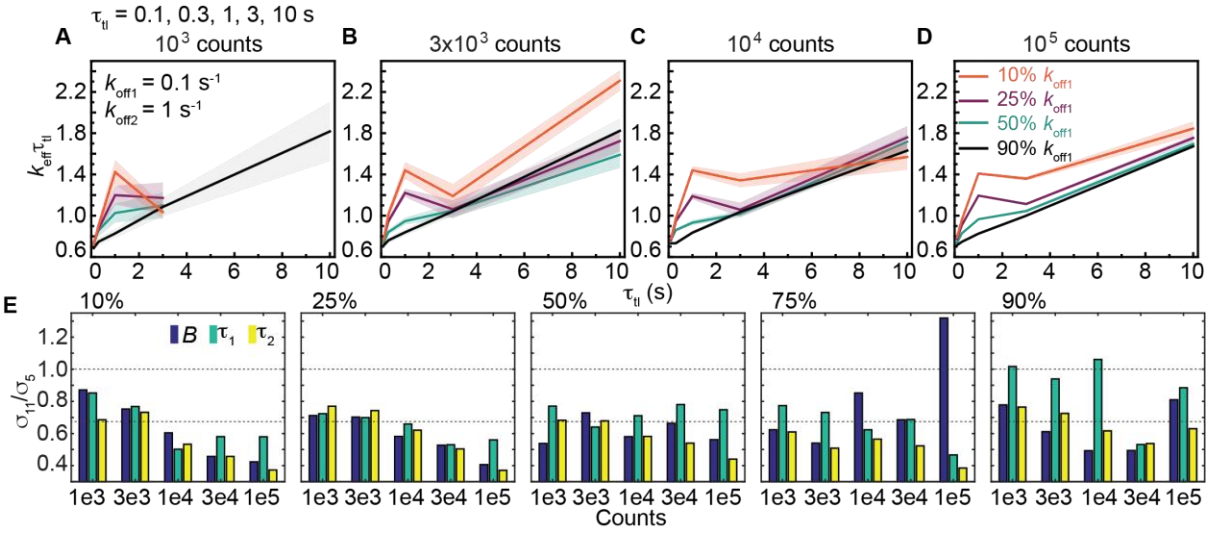
450 FIGURE 7 Determination of binding lifetimes and amplitudes from tri-exponential
451 distributions with a slow rate (k_{off1}), an intermediate rate (k_{off2} = 10k_{off1}) and a fast rate (k_{off3} =
452 10k_{off2}), using the 100-s τ_{tl} set. From left to right, five panels in each row correspond to different
453 amplitudes of each sub-population (displayed on top). (A) $k_{eff}\tau_{tl}$ plots of tri-exponential
454 distributions with $k_b\tau_{int}$ of 0.7, k_{off1}, k_{off2} and k_{off3} of 0.01, 0.1 and 1 s^{-1} respectively, with 10^6
455 observations. Shaded error bands are standard deviations from ten bootstrapped samples. (B-
456 F) Heatmaps show errors in estimates of B_1, B_2, τ_1, τ_2 and τ_3 obtained using global fitting of
457 100 simulated distributions at various pre-set values of B_1 and B_2 (see Fig. S9 for distributions).

458

The choice of τ_{tl}

Given a finite amount of experimental time, should one sample with more τ_{tl} values (increase $N_{interval}$) or obtain more observations (increase n) with a set containing fewer τ_{tl} values? To identify the optimum choice of τ_{tl}, we simulated bi-exponential distributions with an intermediate rate (k_{off1} = 0.1 s^{-1}) and a fast rate (k_{off2} = 1 s^{-1}) using a τ_{tl} set containing either three (N_3) or five (N_5) τ_{tl} values, ranging from 0.1 to 10 s (Table S2). Since fitting outcomes are unreliable in the three τ_{tl} set (compare Fig. S10 to Fig. S11), we decided to examine the simulations with the five τ_{tl} set further. These simulations yielded $k_{eff}\tau_{tl}$ plots that closely resemble those in Fig. 4 (see Fig. 8A-D) and similarly, deviations from straight lines are also reliable indicators of kinetic heterogeneity when B is less than 90%. As expected, estimates of B, τ_1 and τ_2 are more accurate with larger n (Fig. S10).

Compared the simulations using the five τ_{tl} and the 10-s τ_{tl} (11 τ_{tl} values) sets for the same n, errors of estimates are almost always smaller in simulated distributions with the 10-s τ_{tl} set ($\sigma_{11}/\sigma_5 < 1$, see Fig. 8E). By extension of Eq. 8, error ratios (σ_{11}/σ_5) smaller than $1/\sqrt{(11/5)}$ or 0.67 indicate the benefit of increasing $N_{interval}$ outweighs the benefit of increasing n with the five τ_{tl} set whereas error ratios larger than 0.67 represent redundancy in τ_{tl}. Redundancy in τ_{tl} was observed in some cases when the intermediate dissociating sub-population is the majority (B between 75% and 90%) (Fig. 8E). However, when the majority dissociates with the fast rate (B between 10% and 50%), the benefit of sampling with more τ_{tl} is clear ($\sigma_{11}/\sigma_5 < 0.67$), especially with $n \geq 10^4$. Thus, we concluded the net benefit of increasing $N_{interval}$ is greater than increasing the number of counts with a set of fewer τ_{tl} values.



480

481 FIGURE 8 Determination of binding lifetimes and amplitudes from bi-exponential
482 distributions using a τ_{tl} set containing five τ_{tl} values (Table S2), an intermediate rate (k_{off1} = 0.1
483 s^{-1}) and a fast rate (k_{off2} = 1 s^{-1}). (A-D) $k_{eff}\tau_{tl}$ plots of bi-exponential distributions with (A) 10^3,
484 (B) $3x10^3$, (C) 10^4 or (D) 10^5 observations. The amplitude of k_{off1} (*B*) is 10% (orange), 25%
485 (purple), 50% (green) or 90% (black). Shaded error bands are standard deviations from ten
486 bootstrapped samples. (E) Bar plots show ratios of error estimates obtained from simulations
487 with eleven and five τ_{tl} values at the same *n*. Blue: *B*, green: τ_1, yellow: τ_2.

488

DISCUSSION

In this work, we used experimental and simulated data to explore the influence of shot noise, resolution limit and dynamic range limit on resolving multiple kinetic sub-populations in single-molecule time-lapse imaging experiments (Fig. 9). Within the dynamic range and resolution limit, determination of binding lifetimes and amplitudes in mono-exponential and multi-exponential distributions are reliable in general, especially with at least 10^4 counts.

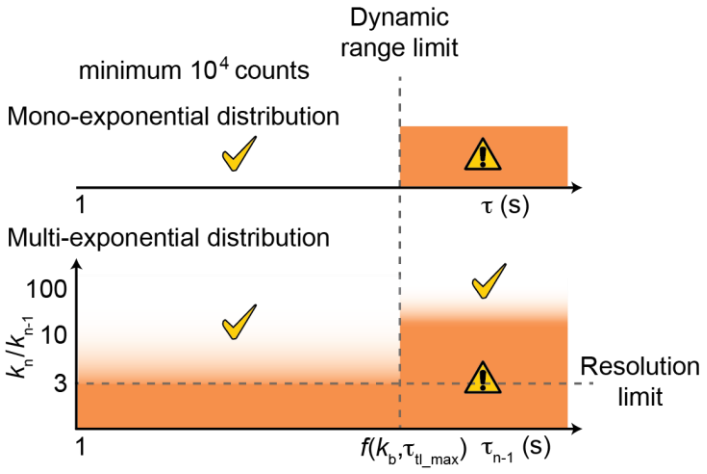


FIGURE 9 Dynamic range and resolution limits in resolving multiple populations using the time-lapse imaging technique with photobleaching-prone fluorescent probes. Dynamic range limit is a function of photobleaching rate (k_b) and the maximum τ_{tl} (τ_{tl_max}) used in experimental conditions. τ_{n-1} is the longest binding lifetime in a multi-exponential distribution within pairs of off rates k_n and k_{n-1}. Orange zones indicate conditions where errors in estimates of τ and the amplitude are high.

As showed in Eq. 8, the relative error in τ determination scales with the square of $k_b\tau_{int}$ and the inverse square root of n. This emphasizes the importance of choosing imaging conditions to minimize $k_b\tau_{int}$ as a two-fold increase in $k_b\tau_{int}$ needs to be compensated by a 16-fold increase in n. A balance has to be struck here to ensure good signal-to-background ratio, a prerequisite for reliable particle tracking. These findings also highlight the importance of developing and using fluorophores with higher photo-stability and brightness for live-cell applications as these would greatly reduce uncertainties in measurements. In practice, the choice of fluorescent

509 protein should be made with great care, as fluorescent proteins often exhibit undesirable
510 properties that limit their utility (38-42).

511 Errors obtained from repeating the experiments can be an underestimation compared to
512 inherent errors conferred by shot noise when fitting is ill-conditioned (43), which is often the
513 case when minimizing using multi-objective functions (44). Therefore, reports of binding
514 lifetimes measurements using these time-lapse imaging approaches should clearly state $k_b\tau_{int}$
515 from fitting and n from experimental data. This would enable a theoretical error estimation of
516 τ and avoid over-interpretation of experimental results.

517 We found $k_{eff}\tau_{tl}$ plots useful for guiding the fitting model when the number of counts is
518 sufficiently large (more than 10^4) as deviations from straight lines faithfully reflect
519 heterogeneity in binding kinetics. The reverse is not necessarily true. Good linear fits, seen at
520 large n values, can reflect one of the following three scenarios: (i) the absence of multiple
521 populations, (ii) sub-populations with off rates that are within the resolution limit, or (iii) sub-
522 populations where the off rate of one population lies beyond the dynamic range. This dynamic
523 range is determined by the photobleaching rate and the maximum τ_{tl} used in the experiment.
524 When the mono-exponential model is used to fit those data, an apparent binding lifetime τ^*,
525 whose value lies between the two true binding lifetimes, is obtained. While sub-optimal, τ^*
526 depends on the proportion of molecules in each kinetic sub-population: a larger presence of the
527 fast dissociating sub-population yields smaller τ^*. This in turn can report on change in binding
528 kinetics when the biology is manipulable – for instance with binding partners or drugs.

529 Can statistical information such as reduced χ^2 be used to decide the model that best describes
530 the data? Computing these criteria requires the determination of the degree of freedom, which
531 still needs to be analytically derived for the non-linear models used in this method (45-48).

Instead of using statistical criteria, the selection of the fitting model using $k_{eff}\tau_{tl}$ plot can be complemented with experimental design. For example, in case where a bi-exponential model is invoked, it might be tempting to attribute sub-populations to molecules performing certain activities such as binding of DNA repair proteins to a damaged or non-damaged substrate. These hypotheses can be tested using structure-function mutants in which one or few catalytic activities are inhibited, hence, yielding predictable changes in $k_{eff}\tau_{tl}$ plots and fitting results. Finally, where possible, we recommend approaches that utilize multiple experimental designs to reproducibly observe or enrich the hypothesized populations.

SOFTWARE

Our algorithms are freely available as open source MATLAB codes from https://github.com/hanngocho/off-rate-simulation.

SUPPORTING INFORMATION

The supplementary information is available following publication.

AUTHOR CONTRIBUTIONS

H.N.H., H.G. and A.M.v.O. designed research. H.N.H. wrote the codes, conducted the simulations and wrote the first draft. D.Z. contributed to the codes, with support from J.K. H.G. and A.M.v.O revised the manuscript and supervised research.

ACKOWLEDGEMENTS

This work was supported by Australian Research Council Grant DP180100858, Australian Laureate Fellowship FL140100027 (to A.M.v.O.). We thank Dr. Joris M. H. Goudsmits for assisting with the simulation codes in MATLAB. D.Z. gratefully acknowledges support from the Elite Network of Bavaria (ENB) program “Macromolecular Science”.

REFERENCES

1. Lu, H. P., L. Xun, and X. S. Xie. 1998. Single-molecule enzymatic dynamics. Science 282:1877-1882.

2. Xie, X. S. 2002. Single-molecule approach to dispersed kinetics and dynamic disorder: Probing conformational fluctuation and enzymatic dynamics. J Chem Phys 117:11024-11032.

3. Larson, M. H., R. Landick, and S. M. Block. 2011. Single-molecule studies of RNA polymerase: one singular sensation, every little step it takes. Mol Cell 41:249-262.

4. Stratmann, S. A., and A. M. van Oijen. 2014. DNA replication at the single-molecule level. Chem Soc Rev 43:1201-1220.

5. Gahlmann, A., and W. E. Moerner. 2014. Exploring bacterial cell biology with single-molecule tracking and super-resolution imaging. Nat Rev Microbiol 12:9-22.

6. Bell, J. C., and S. C. Kowalczykowski. 2016. Mechanics and Single-Molecule Interrogation of DNA Recombination. Annu Rev Biochem 85:193-226.

7. Kaniecki, K., L. De Tullio, and E. C. Greene. 2018. A change of view: homologous recombination at single-molecule resolution. Nat Rev Genet 19:191-+.

8. Kong, M., E. C. Beckwitt, L. Springall, N. M. Kad, and B. Van Houten. 2017. Single-Molecule Methods for Nucleotide Excision Repair: Building a System to Watch Repair in Real Time. Methods Enzymol 592:213-257.

9. Floyd, D. L., J. R. Ragains, J. J. Skehel, S. C. Harrison, and A. M. van Oijen. 2008. Single-particle kinetics of influenza virus membrane fusion. Proc Natl Acad Sci U S A 105:15382-15387.

10. Yan, X., T. A. Hoek, R. D. Vale, and M. E. Tanenbaum. 2016. Dynamics of Translation of Single mRNA Molecules In Vivo. Cell 165:976-989.

11. Ghodke, H., H. Wang, C. L. Hsieh, S. Woldemeskel, S. C. Watkins, V. Rapic-Otrin, and B. Van Houten. 2014. Single-molecule analysis reveals human UV-damaged DNA-binding protein (UV-DDB) dimerizes on DNA via multiple kinetic intermediates. Proc Natl Acad Sci U S A 111:E1862-1871.

12. Volkov, I. L., M. Linden, J. Aguirre Rivera, K. W. Ieong, M. Metelev, J. Elf, and M. Johansson. 2018. tRNA tracking for direct measurements of protein synthesis kinetics in live cells. Nat Chem Biol 14:618-626.

13. Leake, M. C., N. P. Greene, R. M. Godun, T. Granjon, G. Buchanan, S. Chen, R. M. Berry, T. Palmer, and B. C. Berks. 2008. Variable stoichiometry of the TatA component of the twin-arginine protein transport system observed by in vivo single-molecule imaging. Proc Natl Acad Sci U S A 105:15376-15381.

14. Sako, Y., S. Minoghchi, and T. Yanagida. 2000. Single-molecule imaging of EGFR signalling on the surface of living cells. Nat Cell Biol 2:168-172.

15. Dixit, R., J. L. Ross, Y. E. Goldman, and E. L. Holzbaur. 2008. Differential regulation of dynein and kinesin motor proteins by tau. Science 319:1086-1089.

16. Liu, J., J. Hanne, B. M. Britton, J. Bennett, D. Kim, J. B. Lee, and R. Fishel. 2016.

Cascading MutS and MutL sliding clamps control DNA diffusion to activate mismatch repair. Nature 539:583-587.

17. Presman, D. M., D. A. Ball, V. Paakinaho, J. B. Grimm, L. D. Lavis, T. S. Karpova, and G. L. Hager. 2017. Quantifying transcription factor binding dynamics at the single-molecule level in live cells. Methods 123:76-88.

18. Li, Y., J. W. Schroeder, L. A. Simmons, and J. S. Biteen. 2018. Visualizing bacterial DNA replication and repair with molecular resolution. Curr Opin Microbiol 43:38-45.

19. Ghodke, H., H. Ho, and A. M. van Oijen. 2018. Single-molecule live-cell imaging of bacterial DNA repair and damage tolerance. Biochem Soc Trans 46:23-35.

20. Kapanidis, A. N., S. Uphoff, and M. Stracy. 2018. Understanding Protein Mobility in Bacteria by Tracking Single Molecules. J Mol Biol 430:4443-4455.

21. Elf, J., and I. Barkefors. 2018. Single-Molecule Kinetics in Living Cells. Annu Rev Biochem.

22. Mazza, D., A. Abernathy, N. Golob, T. Morisaki, and J. G. McNally. 2012. A benchmark for chromatin binding measurements in live cells. Nucleic Acids Res 40:e119.

23. Gebhardt, J. C., D. M. Suter, R. Roy, Z. W. Zhao, A. R. Chapman, S. Basu, T. Maniatis, and X. S. Xie. 2013. Single-molecule imaging of transcription factor binding to DNA in live mammalian cells. Nat Methods 10:421-426.

24. Shaner, N. C., P. A. Steinbach, and R. Y. Tsien. 2005. A guide to choosing fluorescent proteins. Nat Meth 2:905-909.

25. Liu, Z., L. D. Lavis, and E. Betzig. 2015. Imaging live-cell dynamics and structure at the single-molecule level. Mol Cell 58:644-659.

26. Elf, J., G. W. Li, and X. S. Xie. 2007. Probing transcription factor dynamics at the single-molecule level in a living cell. Science 316:1191-1194.

27. Hipp, L., J. Beer, O. Kuchler, M. Reisser, D. Sinske, J. Michaelis, J. C. M. Gebhardt, and B. Knoll. 2019. Single-molecule imaging of the transcription factor SRF reveals prolonged chromatin-binding kinetics upon cell stimulation. Proc Natl Acad Sci U S A 116:880-889.

28. Reisser, M., A. Palmer, A. P. Popp, C. Jahn, G. Weidinger, and J. C. M. Gebhardt. 2018. Single-molecule imaging correlates decreasing nuclear volume with increasing TF-chromatin associations during zebrafish development. Nat Commun 9:5218.

29. Agarwal, H., M. Reisser, C. Wortmann, and J. C. M. Gebhardt. 2017. Direct Observation of Cell-Cycle-Dependent Interactions between CTCF and Chromatin. Biophys J.

30. Liao, Y., Y. Li, J. W. Schroeder, L. A. Simmons, and J. S. Biteen. 2016. Single-Molecule DNA Polymerase Dynamics at a Bacterial Replisome in Live Cells. Biophys J 111:2562-2569.

31. Mignolet, J., S. Holden, M. Berge, G. Panis, E. Eroglu, L. Theraulaz, S. Manley, and P. H. Viollier. 2016. Functional dichotomy and distinct nanoscale assemblies of a cell cycle-controlled bipolar zinc-finger regulator. Elife 5.

32. Ho, H. N., A. M. van Oijen, and H. Ghodke. 2018. The transcription-repair coupling

factor Mfd associates with RNA polymerase in the absence of exogenous damage. Nat Commun 9:1570.

33. Clauss, K., A. P. Popp, L. Schulze, J. Hettich, M. Reisser, L. Escoter Torres, N. H. Uhlenhaut, and J. C. M. Gebhardt. 2017. DNA residence time is a regulatory factor of transcription repression. Nucleic Acids Res 45:11121-11130.

34. Istratov, A. A., and O. F. Vyvenko. 1999. Exponential analysis in physical phenomena. Rev Sci Instrum 70:1233-1257.

35. Tokunaga, M., N. Imamoto, and K. Sakata-Sogawa. 2008. Highly inclined thin illumination enables clear single-molecule imaging in cells. Nat Methods 5:159-161.

36. Dickson, R. M., A. B. Cubitt, R. Y. Tsien, and W. E. Moerner. 1997. On/off blinking and switching behaviour of single molecules of green fluorescent protein. Nature 388:355-358.

37. Schottky, W. 1918. Über spontane Stromschwankungen in verschiedenen Elektrizitätsleitern. Annalen der Physik 362:541-567.

38. Swulius, M. T., and G. J. Jensen. 2012. The helical MreB cytoskeleton in Escherichia coli MC1000/pLE7 is an artifact of the N-Terminal yellow fluorescent protein tag. J Bacteriol 194:6382-6386.

39. Landgraf, D., B. Okumus, P. Chien, T. A. Baker, and J. Paulsson. 2012. Segregation of molecules at cell division reveals native protein localization. Nature Methods 9:480-U498.

40. Wang, S., J. R. Moffitt, G. T. Dempsey, X. S. Xie, and X. Zhuang. 2014. Characterization and development of photoactivatable fluorescent proteins for single-molecule-based superresolution imaging. Proc Natl Acad Sci U S A 111:8452-8457.

41. Tuson, H. H., A. Aliaj, E. R. Brandes, L. A. Simmons, and J. S. Biteen. 2016. Addressing the Requirements of High-Sensitivity Single-Molecule Imaging of Low-Copy-Number Proteins in Bacteria. Chemphyschem 17:1435-1440.

42. Ghodke, H., V. E. Caldas, C. M. Punter, A. M. van Oijen, and A. Robinson. 2016. Single-Molecule Specific Mislocalization of Red Fluorescent Proteins in Live Escherichia coli. Biophys J 111:25-27.

43. Shrager, R. I., and R. W. Hendler. 1998. Some pitfalls in curve-fitting and how to avoid them: a case in point. J Biochem Biophys Methods 36:157-173.

44. Kaiser, M. J. 2001. Nonlinear multiobjective optimization. Interfaces 31:122-123.

45. Andrae, R., Schulze-Hartung, T. & Melchior, P. 2010. Dos and don'ts of reduced chi-squared. arXiv: 1012.3754.

46. Zou, H., Hastie, T., Tibshirani, R. 2007. ON THE “DEGREES OF FREEDOM” OF THE LASSO. The Annals of Statistics 35:2173-2192.

47. Hansen, N. R., Sokol, A. 2014. Degrees of freedom for nonlinear least squares estimation. arXiv:1402.2997.

48. Tibshirani, R. 2014. Degrees of Freedom and Model Search. arXiv:1402.1920.

Supplementary Information for

Identification of multiple kinetic populations of DNA-binding proteins in live cells

Running title: Measuring binding lifetimes in cells

Han N. Ho[1,2,6], Daniel Zalami[3], Jürgen Köhler[3,4,5], Antoine M. van Oijen[1,2,*] and Harshad Ghodke[1,2,*]

[1]Molecular Horizons and School of Chemistry and Molecular Bioscience, University of Wollongong, Wollongong, Australia

[2]Illawarra Health and Medical Research Institute, Wollongong, Australia

[3]Spectroscopy of soft Matter, University of Bayreuth, Bayreuth, Germany

[4]Bavarian Polymer Institute, Bayreuth, Germany

[5]Bayreuth Institute of Macromolecular Research (BIMF), University of Bayreuth, Bayreuth, Germany

[6]Present address: The Francis Crick Institute, London, UK

* Correspondence: harshad@uow.edu.au or vanoijen@uow.edu.au

Supplementary table

Table S1. Initial conditions, constraints and termination tolerance used in global fitting. n_0 is the minimum number of counts in the second bin across τ_{tl}.

Model	Initial conditions	Bound constraints	Termination tolerance	Algorithm	MATLAB function
Mono (Eq. 2)	$k_b = 1\ s^{-1}$ $k_{off} = 1\ s^{-1}$	$k_b > 0\ s^{-1}$ $0\ s^{-1} < k_{off} < 1/\tau_{int}\ s^{-1}$	10^{-6}	trust-region-reflective	*lsqnonlin*
Bi (Eq. 5)	$k_b = 1\ s^{-1}$ $k_{off1} = 1\ s^{-1}$ $B = 0.5$ $k_{off2} = 2\ s^{-1}$	$k_b > 0$ $10^{-3}\ s^{-1} < k_{off1} < 1/\tau_{int}\ s^{-1}$ $1/n_0 < B < 1 - 1/n_0$ $10^{-3}\ s^{-1} < k_{off2} < 1/\tau_{int}\ s^{-1}$	10^{-6}	trust-region-reflective	*lsqnonlin*
Tri (Eq. 6)	$k_b = 1\ s^{-1}$ $k_{off1} = 0.05\ s^{-1}$ $B_1 = 0.3$ $k_{off2} = 0.5\ s^{-1}$ $B_2 = 0.3$ $k_{off2} = 5\ s^{-1}$	$k_b > 0\ s^{-1}$ $10^{-3}\ s^{-1} < k_{off1} < 1/\tau_{int}\ s^{-1}$ $1/n_0 < B_1 < 1 - 1/n_0$ $10^{-3}\ s^{-1} < k_{off2} < 1/\tau_{int}\ s^{-1}$ $1/n_0 < B_2 < 1 - 1/n_0$ $10^{-3}\ s^{-1} < k_{off3} < 1/\tau_{int}\ s^{-1}$ $B_1 + B_2 < 1 - 2/n_0$	10^{-9}	trust-region-reflective	*fmincon*

Table S2. The τ_{tl} sets used in the study.

τ_{tl} sets	**τ_{tl} values (s)**
10-s	0.1, 0.2, 0.3, 0.4, 0.6, 1, 2, 3, 5, 8, 10
100-s	0.1, 0.3, 0.7, 1, 3, 7, 10, 30, 70, 100
Three-	0.1, 1, 10
Five-	0.1, 0.3, 1, 3, 10

Supplementary figures

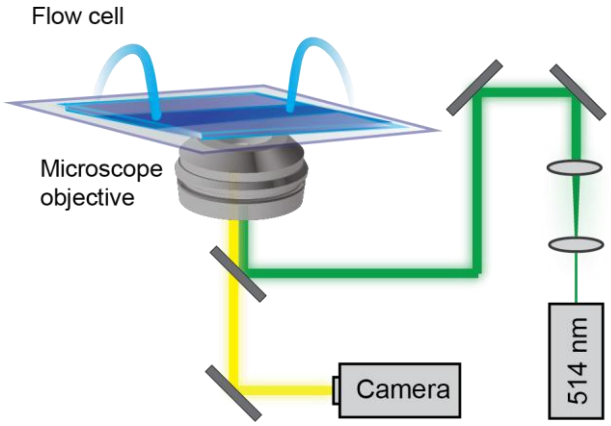


Figure S1. Schematic of experimental setups in single-molecule live-cell imaging. Bacteria expressing fluorescently labelled proteins are loaded in a flow cell with a constant supply of media at 30 ºC. The fluorescent label (YPet) is excited with 514-nm light and fluorescence signal is recorded with an electron-multiplying CCD camera.

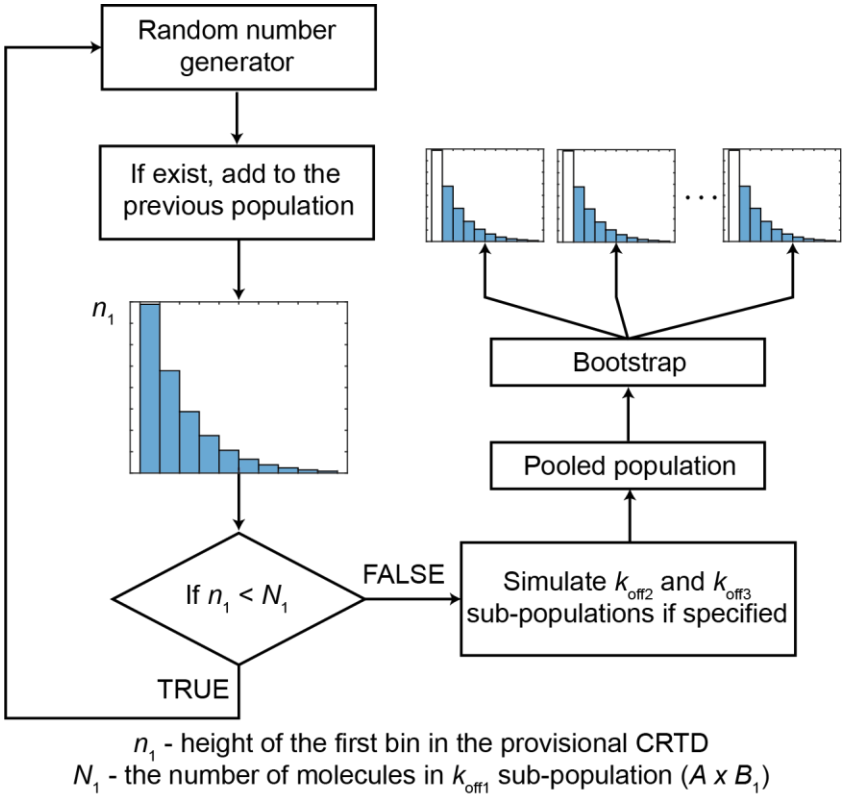


Figure S2. Schematic of the simulation of the cumulative residence time distribution (CRTD) at a specified τ_{tl}. The molecules were generated by a random number generator to produce a group of numbers following an exponential distribution (defined by k_{off1}, k_b, τ_{int} and τ_{tl}) (see Eq. 4-6 in main text). The number generator function was called a few times (typically 3-6) until the number of molecules in the first bin (n_1) of the histogram exceeded the user-specified number of molecules (N_1, $N_1 = A \times B$ in mono-exponential distribution, or $N_1 = A \times B_1$ in multiple-exponential distribution). The k_{off2} and k_{off3} sub-populations were simulated in the same manner. Then, molecules from all simulated sub-populations were pooled and subject to bootstrapping analysis to construct the bootstrapped CRTDs (referred simply as CRTDs). This procedure was repeated for all specified values of τ_{tl}. The global fitting was performed on CRTDs from all τ_{tl}, using a CRTD for each τ_{tl}.

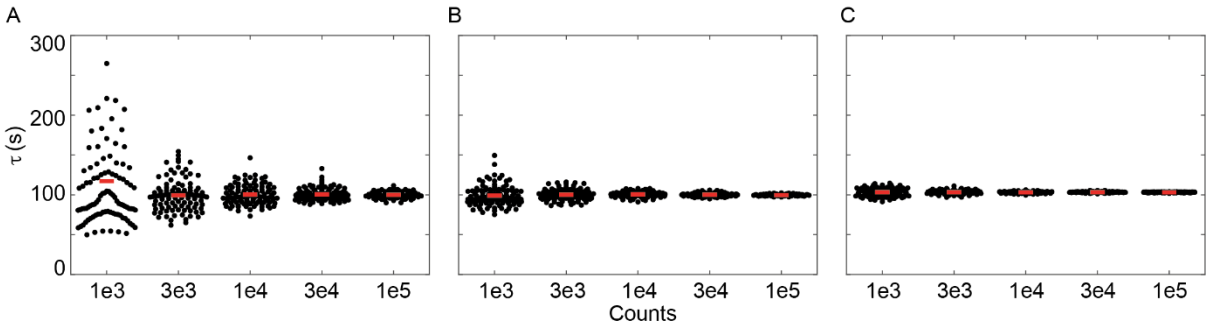


Figure S3. Scatter plots show distributions of τ obtained using global fitting on 100 simulated mono-exponential ($<\tau>$ = 100 s) for each *n* value. (A) Simulation using the 10-s τ_{tl} set. (B) Simulation using the 100-s τ_{tl} set. (C) Simulated data from (B) were globally fitted with the amplitude as the global parameter. Apart from this panel, all global fittings in this study were performed with *A* as the local parameter. Red bars represent the averages.

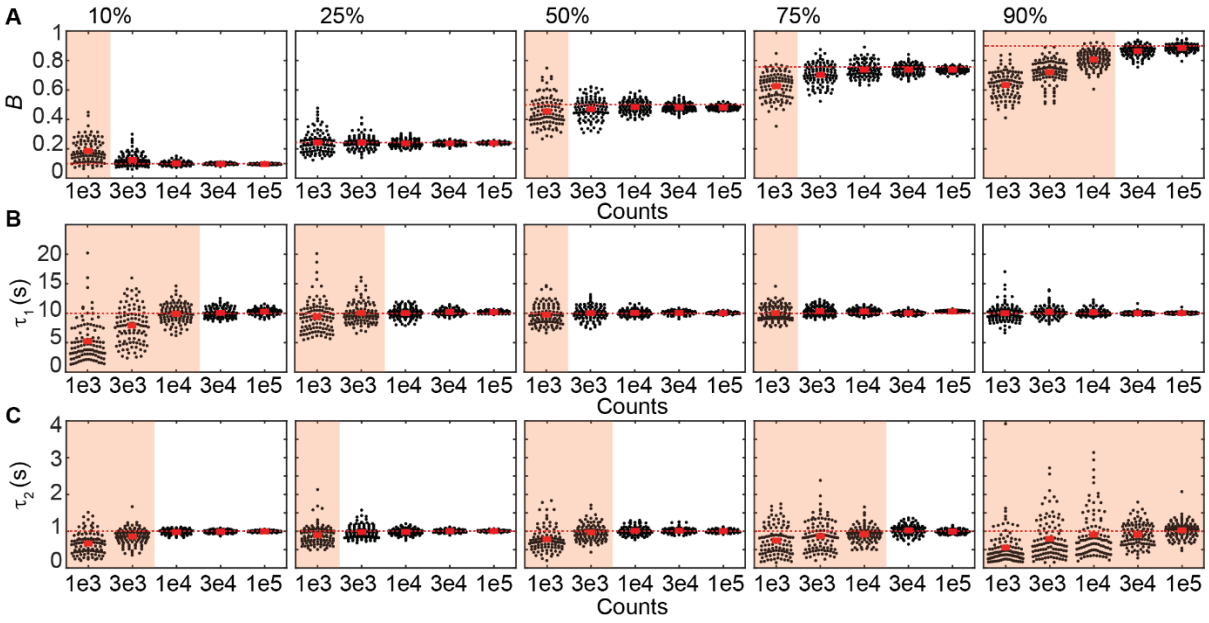


Figure S4. Determination of time constants and amplitudes from bi-exponential distributions with an intermediate rate (k_{off1}) and a fast rate ($k_{off2} = 10k_{off1}$). (A-C) Scatter plots show distributions of B, τ_1 and τ_2 obtained using global fitting from 100 simulated distributions for each n value. Each panel corresponds to a pre-set B, which increases from 10%, 25%, 50%, 75% to 90% from left to right. In each panel, n increases from 10^3 (1e3) to 10^5 (1e5). Dashed lines and red bars represent the true values and the average respectively. Orange shades represent distributions where σ_B is larger than 0.1 or σ_τ/τ is larger than 20%. To enhance visibility, outliers (less than 5% when present) were omitted from scatter plots.

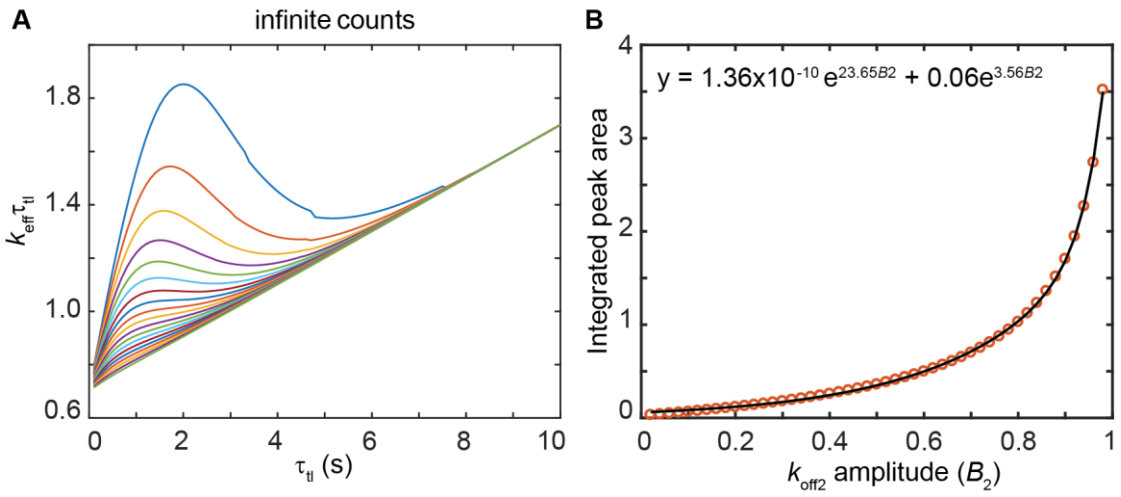


Figure S5. Bi-exponential distributions with an intermediate rate (k_{off1} = 0.1 s^{-1}) and a fast rate (k_{off2} = 1 s^{-1}) with infinite counts. (A) Representative $k_{eff}\tau_{tl}$ plots at 20 amplitudes of k_{off2}. From top to bottom, the amplitude reduces from 95% to 5%. (B) Integrated peak areas as a function of k_{off2} amplitudes (open circles). Line is the exponential fit to data points (R^2: 0.9996). The peak area is calculated as the difference between areas under the $k_{eff}\tau_{tl}$ plots and the area under the line $y = 0.7 + 0.1\tau_{tl}$.

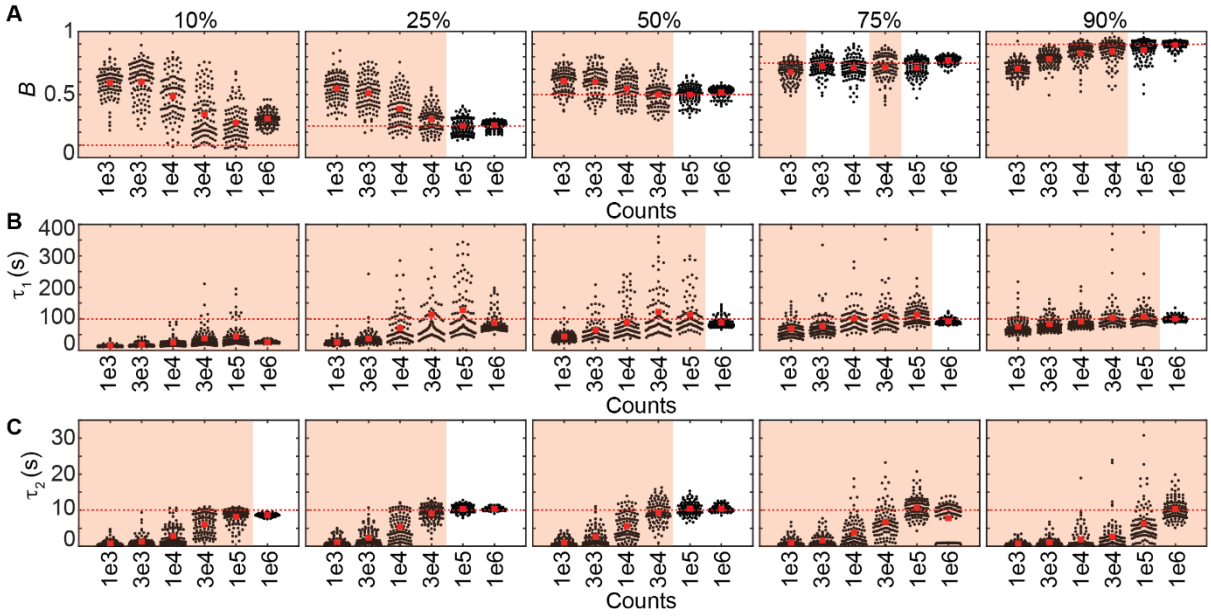


Figure S6. Determination of time constants and amplitudes from bi-exponential distributions with a slow rate (k_{off1} = 0.01 s^{-1}) and an intermediate rate (k_{off2} = 0.1 s^{-1}). (A-C) Scatter plots show distributions of B, τ_1 and τ_2 obtained from fitting of 100 simulated distributions to bi-exponential model. Each panel corresponds to a pre-set amplitude of B, which increases from 10%, 25%, 50%, 75% to 90% from left to right. In each panel, n increases from 10^3 (1e3) to 10^6 (1e6). Dashed lines and red bars represent the true values and the average respectively. Orange shades represent distributions where σ_B is larger than 0.1 or σ_τ/τ is larger than 20%. To enhance visibility, outliers (less than 5% when present) were omitted from scatter plots.

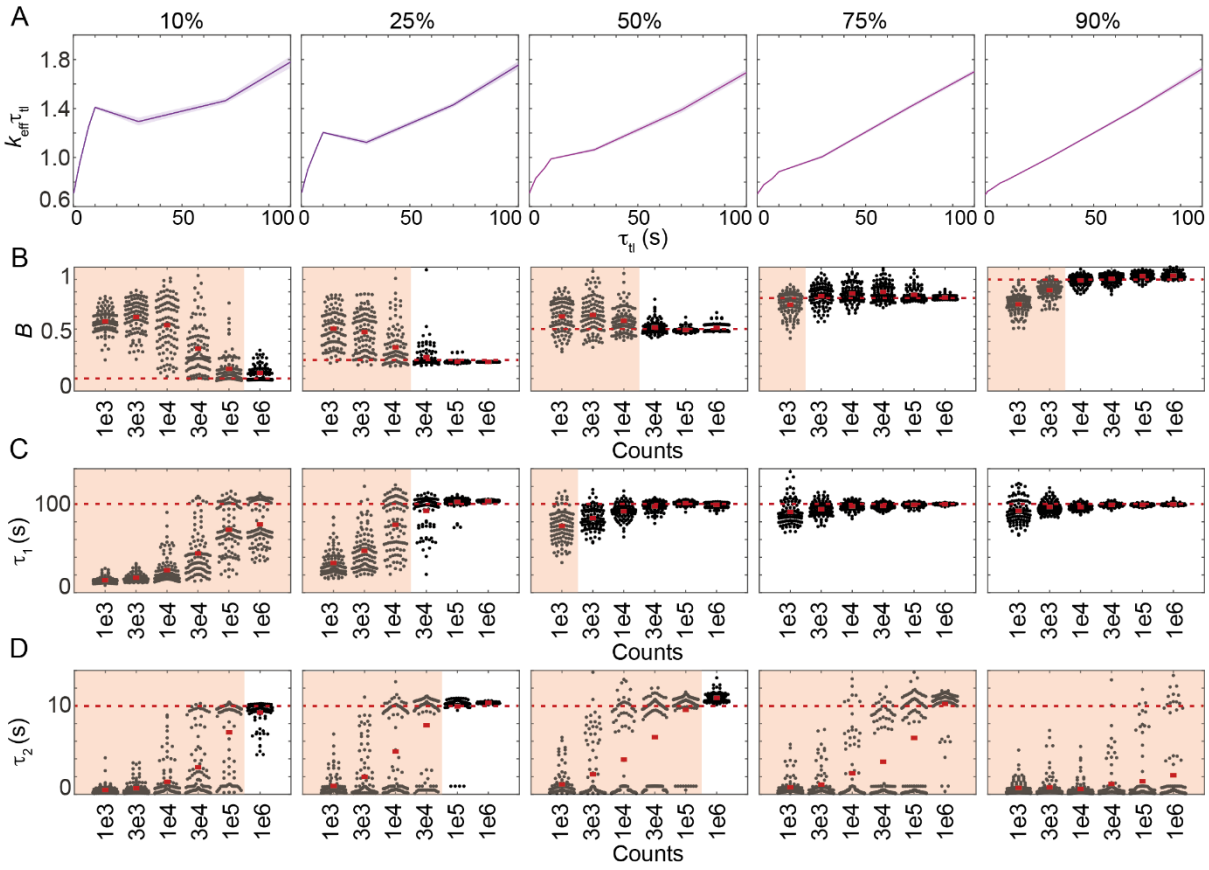


Figure S7. Determination of time constants and amplitudes from bi-exponential distributions with a slow rate (k_{off1}) and an intermediate rate (k_{off2} = 10k_{off1}), simulated using the 100-s τ_{tl} set. (A) $k_{eff}\tau_{tl}$ plots of bi-exponential distributions with $k_b\tau_{int}$ of 0.7, k_{off1} and k_{off2} of 0.01 and 0.1 s^{-1} respectively, with 10^5 observations. The amplitude of k_{off1} (B, shown on top) increases from left to right (10% to 90%). Shaded error bands are standard deviations from ten bootstrapped samples. (B-D) Scatter plots show distributions of B, τ_1 and τ_2 obtained from fitting of 100 simulated distributions to bi-exponential model. Each panel corresponds to a pre-set amplitude of B, which increases from 10%, 25%, 50%, 75% to 90% from left to right. In each panel, n increases from 10^3 (1e3) to 10^6 (1e6). Dashed lines and red bars represent the true values and the average respectively. Orange shades represent distributions where σ_B is larger than 0.1 or σ_τ/τ is larger than 20%. To enhance visibility, outliers (less than 5% when present) were omitted from scatter plots.

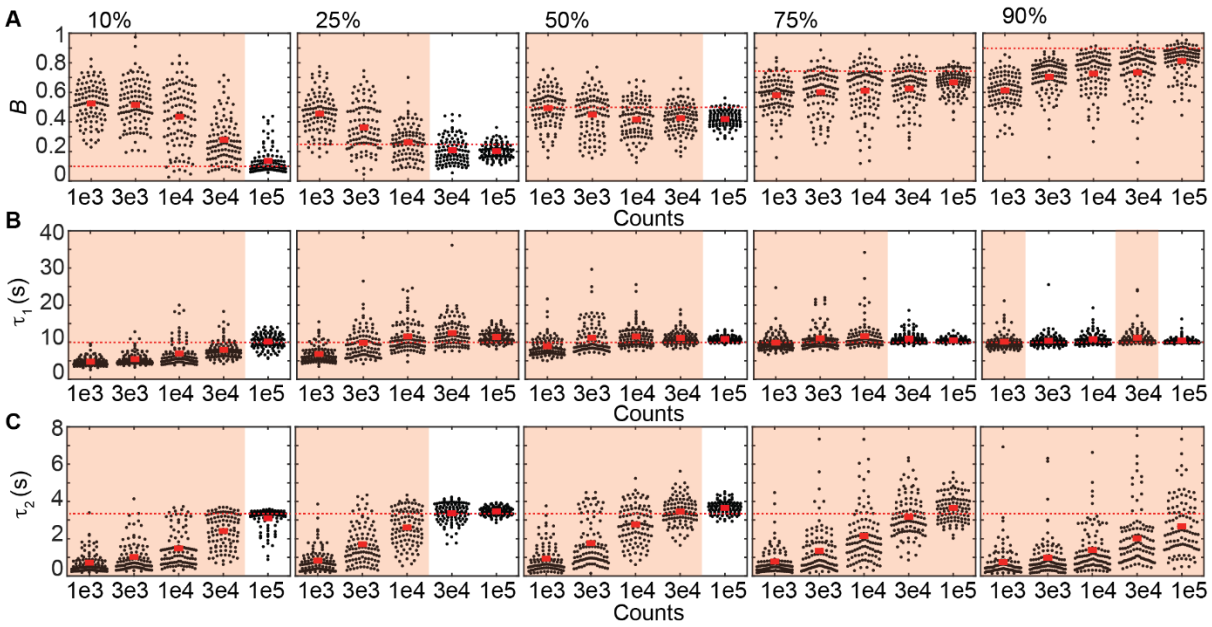


Figure S8. Determination of binding lifetimes and amplitudes from bi-exponential distributions with closely spaced rates ($k_{off2} = 3k_{off1}$). (A-C) Scatter plots show distributions of B, τ_1 and τ_2 obtained from fitting of 100 simulated distributions for each n value. Each panel corresponds to a pre-set B, which increases from 10%, 25%, 50%, 75% to 90% from left to right. In each panel, n increases from 10^3 (1e3) to 10^5 (1e5). Dashed lines and red bars represent the true values and the average respectively. Orange shades represent distributions where σ_B is larger than 0.1 or σ_τ/τ is larger than 20%. To enhance visibility, outliers (less than 5% when present) were omitted from scatter plots.

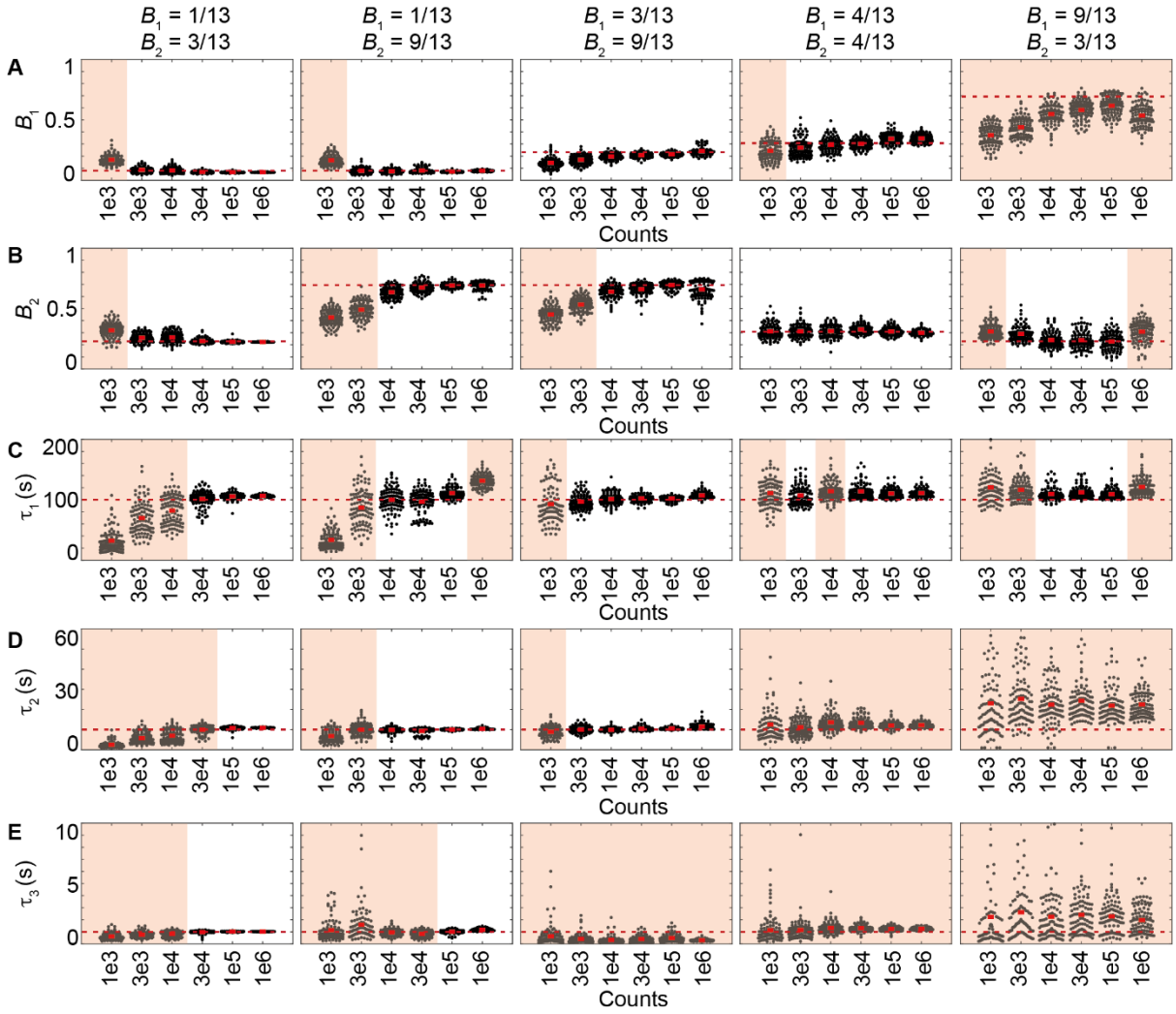


Figure S9. Determination of binding lifetimes and amplitudes from tri-exponential distributions with a slow rate (k_{off1}), an intermediate rate ($k_{off2} = 10k_{off1}$) and a fast rate ($k_{off3} = 10k_{off2}$), using the 100-s τ_{tl} set. From left to right, five panels in each row correspond to different amplitudes of each sub-population (displayed on top). (A-E) Scatter plots show distributions of amplitudes (B_1 and B_2), τ_1, τ_2 and τ_3 obtained using global fitting 100 simulated samples. In each panel, n increases from 10^3 (1e3) to 10^6 (1e6). Dashed lines and red bars represent the true values and the averages respectively. Orange shades represent distributions where σ_B is larger than 0.1 or σ_τ/τ is larger than 20%. To enhance visibility, outliers (less than 5% when present) were omitted from scatter plots.

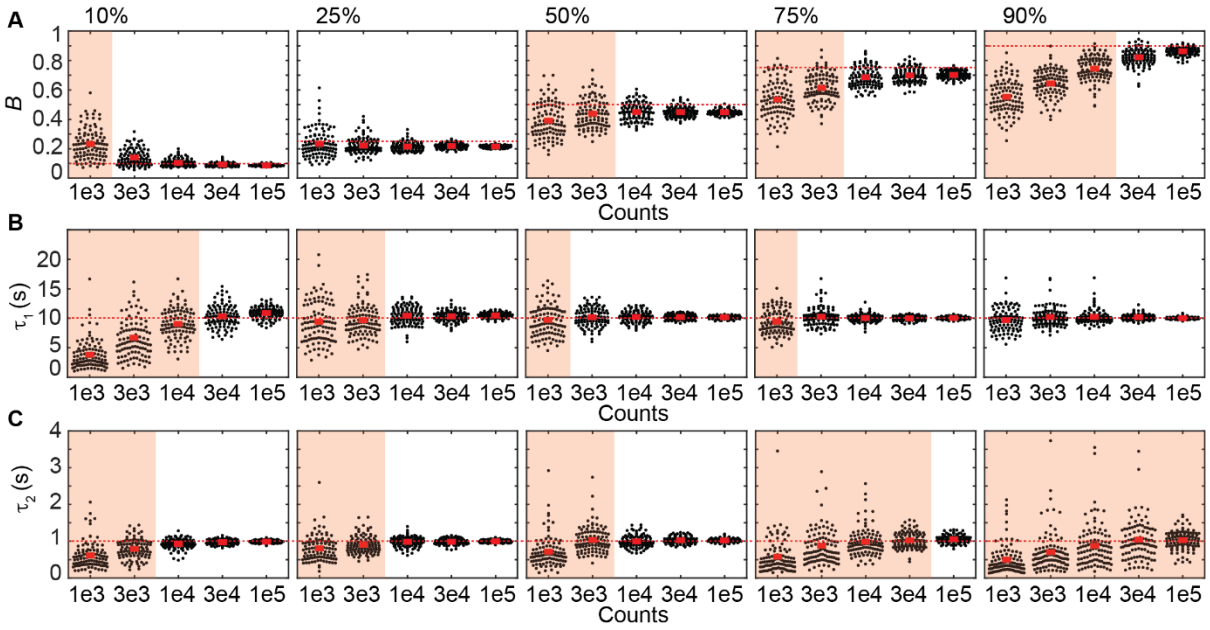


Figure S10. Determination of time constants and amplitudes from bi-exponential distributions simulated with the five τ_{tl} set, and an intermediate rate (k_{off1} = 0.1 s^{-1}) and a fast rate (k_{off2} = 1 s^{-1}). (A-C) Scatter plots show distributions of B, τ_1 and τ_2 obtained from fitting of 100 simulated distributions to bi-exponential model. Each panel corresponds to a pre-set amplitude of B, which increases from 10%, 25%, 50%, 75% to 90% from left to right. In each panel, n increases from 10^3 (1e3) to 10^5 (1e5). Dashed lines and red bars represent the true values and the average respectively. Orange shades represent distributions where σ_B is larger than 0.1 or σ_τ/τ is larger than 20%. To enhance visibility, outliers (less than 5% when present) were omitted from scatter plots.

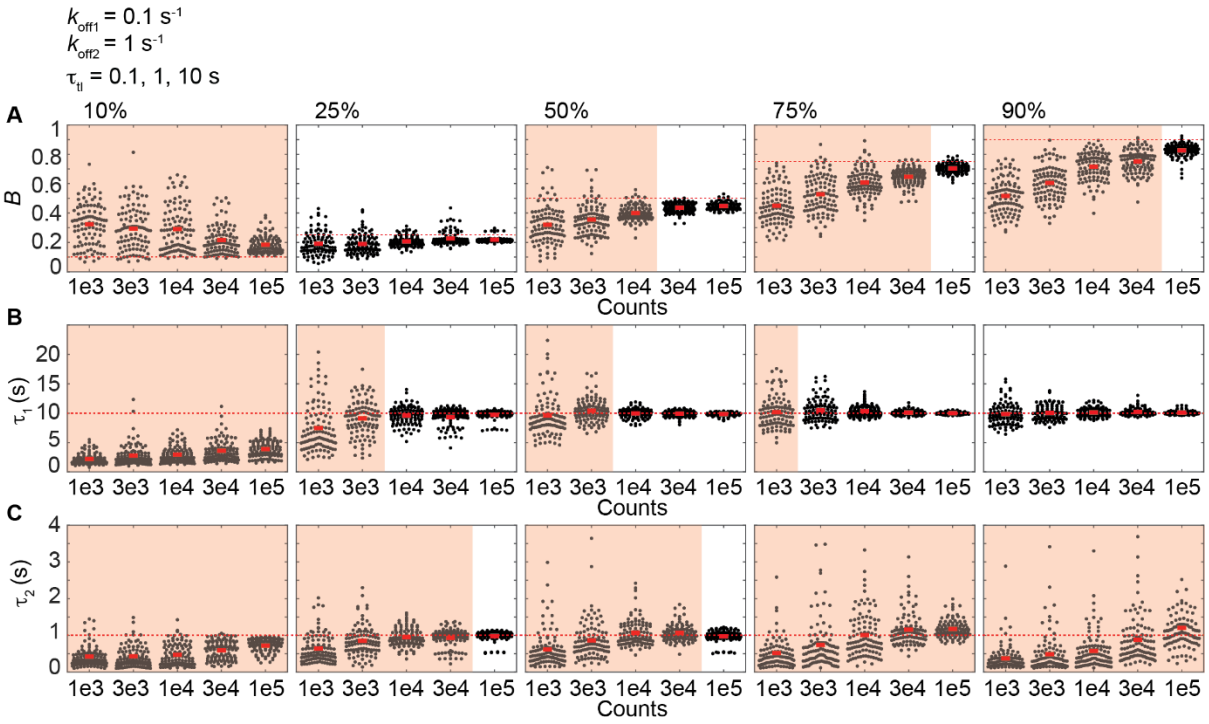


Figure S11. Determination of time constants and amplitudes from bi-exponential distributions simulated with the three τ_{tl} set, and an intermediate rate (k_{off1} = 0.1 s^{-1}) and a fast rate (k_{off2} = 1 s^{-1}). (A-C) Scatter plots show distributions of B, τ_1 and τ_2 obtained from fitting of 100 simulated distributions to bi-exponential model. Each panel corresponds to a pre-set amplitude of B, which increases from 10%, 25%, 50%, 75% to 90% from left to right. In each panel, n increases from 10^3 (1e3) to 10^5 (1e5). Dashed lines and red bars represent the true values and the average respectively. Orange shades represent distributions where σ_B is larger than 0.1 or σ_τ/τ is larger than 20%. To enhance visibility, outliers (less than 5% when present) were omitted from scatter plots.

Supplementary Notes

1. Simulation of a set of binding events whose lifetimes follow an exponential distribution with user-defined mean

```
function [counts, each_molecule] = simulate_res_time(mu,edges,n_count)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%% Inputs:
%%    mu: mean of exponential distribution for a particular τt1
%%    edges: bin edges of histograms
%%    n_count: the number of counts for a particular τt1
%% Outputs:
%%    counts: vector describing CRTD
%%    each_molecule: vector containing all random number corresponding to
%%                   lifetimes of binding events
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
each_molecule = [];
counts = zeros(10,1);
%% generate a set of random numbers corresponding to lifetimes of binding
%% events until counts in the first bin exceed user-defined counts
while counts(1) < n_count
% single iteration of the exprnd function
    sim = exprnd(mu,round(n_count/2.71),1);
% construct the histogram with edges corresponding to frame times
% N is a vector containing counts in all bins [from the latest iteration]
    [N,~] = histcounts(sim,edges);
    counts = counts + N'; % add counts to the previous iterations
% combine lifetimes of binding events to existing population from previous
% iteration of the exprnd function
    each_molecule = [each_molecule; sim];
end
end % end of the function
```

2. Simulation of mono-, bi- or tri-exponential distribution across all τ_{tl}

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%% Inputs:
%%    ttl:   vector containing the set of time-lapse intervals
%%    kb:    photobleaching rate (unit: s^-1)
%%    tint:  camera integration time
%%    koff1: user-defined off rate 1
%%    koff2: user-defined off rate 2
%%    koff3: user-defined off rate 3
%%    B(1):  amplitude of the first kinetic sub-population
%%    B(2):  amplitude of the second kinetic sub-population
%%    n_count_total: user-defined counts for each simulation
%% Outputs:
%%    bin: matrix containing CRTDs for all time-lapse intervals
%%    d.data: contains the simulated population at a particular time-
%%    lapse interval
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
for i = 1:length(ttl)    % simulate CRTD for each time-lapse interval
     time = ttl(i)*(0:10)';  % determine frame times for binning
     %% define exponential distribution for each sub-population
     keff1 = (kb*tint/ttl(i) + koff1); % effective rate 1
     % mean of the exponential distribution of the first sub-population
     mu1 = 1/keff1;
     keff2 = (kb*tint/ttl(i) + koff2); % effective rate 2
     % mean of the exponential distribution of the second sub-population
     mu2 = 1/keff2;
     keff3 = (kb*tint/ttl(i) + koff3); % effective rate 3
     % mean of the exponential distribution of the third sub-population
     mu3 = 1/keff3;
     %% determine the number of counts for each sub-population based
     %% on the amplitudes B1 and B2
     % counts of the first kinetic sub-population
     n_count1 = round(B(1)*n_count_total);
     % counts of the second kinetic sub-population
     n_count2 = round(B(2)*n_count_total);
     % counts of the third kinetic sub-population
     n_count3 = n_count_total - n_count1 - n_count2;
     % bin1, bin2 and bin3 are vectors containing CRTDs of koff1, koff2 and
     % koff3 sub-population respectively
     % population1, population2 and population3 are vectors containing
     % simulated koff1, koff2 and koff3 sub-population respectively.
     bin2 = zeros(10,1); population2 = [];
     bin3 = zeros(10,1); population3 = [];
     % simulate koff1 sub-population
     [bin1, population1] = simulate_res_time(mu1,time,n_count1);
     % simulate koff2 sub-population
     if n_count2 > 1
          [bin2, population2] = simulate_res_time(mu2,time,n_count2);
     end
     % simulate koff3 sub-population
     if n_count3 > 1
          [bin3, population3] = simulate_res_time(mu3,time,n_count3);
     end
     % combine CRTDs from sub-population CRTDs
            bin(:,i) = bin1 + bin2 + bin3;
     % combine simulated population from simulated sub-populations
     d(i).data = [population1; population2; population3];
end
```

3. Global fitting

```
function [p_out] = globalFit(i_model, X, Y, tint)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%% Inputs:
%%    i_model = 1 - mono-exponential model
%%    i_model = 2 - bi-exponential model
%%    i_model = 3 - tri-exponential model
%%    X: matrix containing frame times of all time-lapse intervals
%%          - row: frame times corresponding to one time-lapse interval
%%          - column: increase in frame times
%%    Y: matrix containing simulated CRTDs of all time-lapse intervals
%%    tint: camera integration time
%%    para: initial conditions
%%      - mono-exponential model: [kb, koff1, counts]
%%      - bi-exponential model: [kb, koff1, B1, koff2, counts]
%%      - tri-exponential model: [kb, koff1, B1, koff2, B2, koff3, counts]
%%    lb: lower constraints
%%      - mono-exponential model: [kb, koff1, counts]
%%      - bi-exponential model: [kb, koff1, B1, koff2, counts]
%%      - tri-exponential model: [kb, koff1, B1, koff2, B2, koff3, counts]
%%    ub: upper constraints
%%      - mono-exponential model: [kb, koff1, counts]
%%      - bi-exponential model: [kb, koff1, B1, koff2, counts]
%%      - tri-exponential model: [kb, koff1, B1, koff2, B2, koff3, counts]
%% Outputs:
%%    p_out: vector containing outcomes of global fitting
%%          - p(1): kb
%%          - p(2): koff1
%%          - p(3): B1
%%          - p(4): koff2
%%          - p(5): B2
%%          - p(6): koff3
%%          - p(7): 1 - B1 - B2
%%          - p(8)-p(end): counts at time 0 for all time-lapse intervals
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%% Known Parameters
ttl = X(:,1); % vector containing all time-lapse intervals
a_para = Y(:,1); % Initialize the vector for counts at time 0
weights = ones(size(X)); % fitting weights
lower_B = 1/min(a_para(a_para>0)); % the lower bound for the amplitudes
upper_koff = 1/tint;     % the upper bound for off rates
if i_model == 1   % fitting to mono-exponential function
     para = [1,  1, a_para']; % initial conditions: kb, koff1, counts
     lb   = [0,  0, zeros(size(ttl))']; % lower bounds: kb, koff1, counts
     % upper bounds: kb, koff1, counts
     ub   = [Inf,upper_koff, Inf*ones(size(ttl))'];
     % define function to minimize
     f1 = @(p)(   (model(i_model,p,X,tint,ttl)-Y).*weights );
     opts = optimset('Display','off');
     % Global fitting using the lsqnonlin function
     [p] = lsqnonlin(f1,para,lb,ub,opts);
     p_out = [p(1:2),1,zeros(1,4),p(3:end)];
elseif i_model == 2 % fitting to bi-exponential function
     para = [1, 1, 0.5, 2, a_para'];
     lb   = [0, 1e-3, lower_B, 1e-3, zeros(size(ttl))'];
     ub   = [Inf, upper_koff, 1-lower_B, upper_koff, Inf*ones(size(ttl))'];
     % define function to minimize
     f1 = @(p)(   (model(i_model,p,X,tint,ttl)-Y).*weights );
     opts = optimset('Display','off');
```

```
     % Global fitting using the lsqnonlin function
     [p] = lsqnonlin(f1,para,lb,ub,opts);
      % assign the smaller off rate to be koff1
      p_temp = sortrows([p(2) p(3); p(4) (1 - p(3))]);
      p_temp = p_temp';
      p_out  = [p(1), p_temp(:)', zeros(1,2), p(5:end)];
elseif i_model == 3
      para = [1, 0.05, 0.3, 0.5, 0.3, 5, a_para'];
      lb   = [0, 1e-3, lower_B, 1e-3, lower_B, 1e-3, zeros(size(ttl))'];
      ub   = [Inf, upper_koff, 1-lower_B, upper_koff, 1-lower_B,upper_koff,
              Inf*ones(size(ttl))'];
      % define function to minimize
      f1 = @(p)( sum(sum((model(i_model,p,X,tint,ttl)-Y).^2.*weights,2 )));
      opts = optimoptions('fmincon', 'MaxFunctionEvaluations',10000,...
            'MaxIter',3000,'Algorithm','interior-point','StepTolerance',
            1.0000e-9);
      b = 1-2*lower_B;
      A = [0,0,1,0,1,0,zeros(1,size(a_para,1))];
      % Global fitting using the fmincon function
      [p] = fmincon(f1,para,A,b,[],[],lb,ub,[],opts);
      % assign the smallest off rate to be koff1 and the second smallest to
      % be koff2
      p_temp = sortrows([p(2) p(3); p(4) p(5); p(6) (1-p(3)-p(5))]);
      p_temp = p_temp';
      p_out = [p(1),p_temp(:)',p(7:end)];
end
end % end of function
```

4. Define fitting models

```
function f = model(i_model,para,X,tint,ttl)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%% Inputs:
%%    i_model = 1 - mono-exponential model
%%    i_model = 2 - bi-exponential model
%%    i_model = 3 - tri-exponential model
%%    para: global parameters
%%    X:    frame times
%%    tint: camera integration times
%%    ttl: time-lapse time
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% ampl: vector containing counts for all time-lapse intervals
p = tint./ttl; p = p(:);
if i_model == 1
    kb = para(1);
    koff1 = para(2);
    ampl = para(3:end);
    % mono-exponential model
    f = (ampl'*ones(1,size(X,2))).*
        (exp(-((kb.*p + koff1)*ones(1,size(X,2))).*X));
elseif i_model == 2 % bi-exponential model
    kb = para(1);
    koff1 = para(2);
    B1 = para(3);
    koff2 = para(4);
    ampl = para(5:end);
    % bi-exponential model
    f = (ampl'*ones(1,size(X,2))).*(B1.*exp(-((kb.*p + koff1)*
      ones(1,size(X,2))).*X)+(1-B1).*exp(-(kb.*p + koff2)*
      ones(1,size(X,2)).*X));
```

```
elseif i_model == 3
    kb = para(1);
    koff1 = para(2);    B1 = para(3);
    koff2 = para(4);    B2 = para(5);
    koff3 = para(6);
    ampl = para(7:end);
    % tri-exponential model
    f = (ampl'*ones(1,size(X,2))).*
        (B1.*exp(-((kb.*p + koff1) * ones(1,size(X,2))).*X)
        + B2.* exp( -(kb.*p + koff2)*ones(1,size(X,2)).*X )+
        (1-B1-B2).* exp( -(kb.*p + koff3)*ones(1,size(X,2)).*X ));
end
end % end of function
```