

SCIENTIFIC REPORTS



OPEN

SMTracker: a tool for quantitative analysis, exploration and visualization of single-molecule tracking data reveals highly dynamic binding of *B. subtilis* global repressor AbrB throughout the genome

Thomas C. Rösch^{1,2}, Luis M. Oviedo-Bocanegra^{1,2}, Georg Fritz^{1,3} & Peter L. Graumann^{1,2}

Single-particle (molecule) tracking (SPT/SMT) is a powerful method to study dynamic processes in living cells at high spatial and temporal resolution. Even though SMT is becoming a widely used method in bacterial cell biology, there is no program employing different analytical tools for the quantitative evaluation of tracking data. We developed SMTracker, a MATLAB-based graphical user interface (GUI) for automatically quantifying, visualizing and managing SMT data via five interactive panels, allowing the user to interactively explore tracking data from several conditions, movies and cells on a track-by-track basis. Diffusion constants are calculated a) by a Gaussian mixture model (GMM) panel, analyzing the distribution of positional displacements in *x*- and *y*-direction using a multi-state diffusion model (e.g. DNA-bound vs. freely diffusing molecules), and inferring the diffusion constants and relative fraction of molecules in each state, or b) by square displacement analysis (SQD), using the cumulative probability distribution of square displacements to estimate the diffusion constants and relative fractions of up to three diffusive states, or c) through mean-squared displacement (MSD) analyses, allowing the discrimination between Brownian, sub- or superdiffusive behavior. A spatial distribution analysis (SDA) panel analyzes the subcellular localization of molecules, summarizing the localization of trajectories in 2D- heat maps. Using SMTracker, we show that the global transcriptional repressor AbrB performs highly dynamic binding throughout the *Bacillus subtilis* genome, with short dwell times that indicate high on/off rates *in vivo*. While about a third of AbrB molecules are in a DNA-bound state, 40% diffuse through the chromosome, and the remaining molecules freely diffuse through the cells. AbrB also forms one or two regions of high intensity binding on the nucleoids, similar to the global gene silencer H-NS in *Escherichia coli*, indicating that AbrB may also confer a structural function in genome organization.

With the advent of single-molecule localization microscopy, there is an increasing number of studies reporting the dynamics of single molecules at the millisecond range and at high optical resolution¹. Contrarily to ensemble measurements, such as fluorescence correlation spectroscopy (FCS) and fluorescence recovery after photobleaching (FRAP) experiments, single molecule tracking (SMT) provides details about individual molecules

¹SYNMIKRO, LOEWE Center for Synthetic Microbiology, Marburg, Germany. ²Department of Chemistry, Philipps Universität Marburg, Marburg, Germany. ³Department of Physics, Philipps Universität Marburg, Marburg, Germany. Correspondence and requests for materials should be addressed to G.F. (email: georg.fritz@synmikro.uni-marburg.de) or P.L.G. (email: peter.graumann@synmikro.uni-marburg.de)

and their diffusive behavior² and has revealed unprecedented insights into the mechanism of diverse cellular processes such as signal transduction³, chromosome segregation^{4,5}, transcription⁶, translation^{7,8}, replication⁹ and DNA-repair^{10–12}.

Despite several established software solutions to detect and connect sequentially acquired signals of fluorescent fusion proteins¹³, we are aware of only few GUI-based tools to easily assess, visualize and further analyze diffusion data without advanced computational skills. The softwares InferenceMAP and Diatrack^{14,15} are able to infer diffusion constants from tracking data and allows the user to visualize spatial maps of protein mobilities, primarily designed for eukaryotic cells. We therefore devised a GUI-based solution to easily quantify, explore and visualize the molecular kinetics and dynamics of single particles in bacterial cells. This program fills a void for tools that provide access to different diffusive states of individual molecules, as often incurred by binding and unbinding to other subcellular structures. We therefore developed SMTracker, a versatile and user-friendly program operating in MATLAB, which can be run without prior programming knowledge. The software allows detailed access to single trajectories, to analyze the mode of diffusion via different approaches, the determination of binding kinetics (e.g. of protein-DNA interactions) and the 2-dimensional visualization of subcellular distributions of molecules and their tracks in a normalised cell.

SMTracker contains 5 different panels (Fig. 1): (i) The import/exploration panel allows the user to interactively explore tracking data from several conditions, movies and cells on a track-by-track basis. (ii) The Gaussian mixture model (GMM) panel analyzes the distribution of positional displacements in x - and y -direction using a multi-state diffusion model (e.g. DNA-bound vs. freely diffusing molecules), and infers the diffusion constants and relative fraction of molecules in each state. (iii) The squared displacement analysis (SQD) panel uses the cumulative probability distribution of squared displacements to estimate the diffusion constants and relative fractions of up to three diffusive states. Although the GMM and SQD analyses are closely related, each method has strengths and weaknesses in accurately inferring molecular properties under different experimental conditions, e.g. for different magnitudes of the diffusion constant, as shown by benchmarking both methods on computer-simulated SMT data. (iv) The mean-square displacement (MSD) panel determines the mode of diffusion from a fit of the MSD plotted against the time lag between image frames, allowing the discrimination between Brownian, sub- or superdiffusive behavior. (v) The spatial distribution analysis (SDA) panel analyzes the subcellular localization of molecules, summarizing the localization of trajectories in 2D- and 3D-heat maps and in normalized distributions projected in x - and y -orientation. This visualization allows distinguishing different localization patterns (cytoplasmic, nucleoid- or membrane-bound) arising from different biological conditions.

AbrB is a tetrameric DNA binding protein that generally represses many genes during exponential growth, which are released from repression at the transition to stationary phase, rendering AbrB an important global regulator for stationary phase gene expression (including biofilm formation) in *Bacillus subtilis* and other Gram positive bacteria^{16–18}. In many different bacterial species, including Cyanobacteria, AbrB confers an important function in gene regulation for e.g. metabolic processes, extending its activity into the exponential growth phase^{19,20}. In *B. subtilis*, AbrB also binds to DNA as a heterotetramer, together with its paralog Abh, to a total of over 600 binding sites throughout the genome²¹. Of these, about 100 have been implicated in regulating promoter activities in a direct manner, which influences the expression of almost 200 genes. In spite of its genome-wide importance, and role as transition-state regulator, AbrB has never been characterized in a cell biological approach. Using SMTracker, we determined single molecule dynamics of AbrB, and also found a surprising clustering of AbrB molecules on the nucleoids that may confer a more general role as chromosome-structuring element.

Results and Discussion

Import/exploratory panel. As an input, SMTracker requires single-molecule tracking data, cell contour data and original fluorescence movies (Figs. 1a and S1). SMTracker supports data input from two of the most commonly used particle tracking tools U-track²² and TrackMate²³ and computes the x - and y -coordinates of molecular trajectories relative to the geometry of each cell, as obtained by the cell segmentation tools MicrobeTracker²⁴ or Oufti²⁵. After automatic data import – only requiring the pixel size of the microscopy system and the time interval Δt between image frames as user input – each track is associated to a corresponding cell and the user can interactively explore tracking data from several conditions, movies and cells on a track-by-track basis. Cells and tracks can be manually removed, if necessary, and overlaid in up to three different channels acquired during imaging (e.g. DAPI, membrane or phase contrast), allowing for a manual analysis of protein localization relative to cellular markers. Besides choosing a specific trajectory or cell from a pop-up menu, the user can drag each cell or trajectory on the screen (Fig. S4) and thereby access statistical details about the track (Fig. S5) and visualize the track as a movie. Additionally, this panel provides general statistical information about the experiment (e.g. the number of movies, cells, trajectories, etc.) and the conditions studied.

Gaussian mixture model (GMM) analysis panel. To directly compare the change in intracellular protein mobility upon changing conditions of the cell (e.g. drug treatment), we developed a method based on Gaussian mixture models (GMM). The method assumes that molecules may exist in multiple states, e.g. DNA-bound vs. free, or septum-bound and freely diffusive, each of which is associated with a different diffusion constant D_1 and D_2 , respectively, and that the fraction α of molecules in the slow diffusive state (or $1 - \alpha$ in the fast diffusive state) may change between experimental conditions (Fig. 1b). For each subpopulation the random motion between consecutive image frames leads to positional displacements in x - and y -direction, Δx and Δy , following a Gaussian distribution with zero mean and standard deviation $\sigma = \sqrt{2D\Delta t}$, where D is the respective (1-dimensional) diffusion constant and Δt the time interval between image frames. By fitting the superposition of one, two or optionally three Gaussians to the experimental displacement distributions the algorithm infers the diffusion constants D_1 , D_2 (and D_3) as well as the fractions α_1 , α_2 and α_3 of the respective molecules in the different diffusive states. In particular, if we let

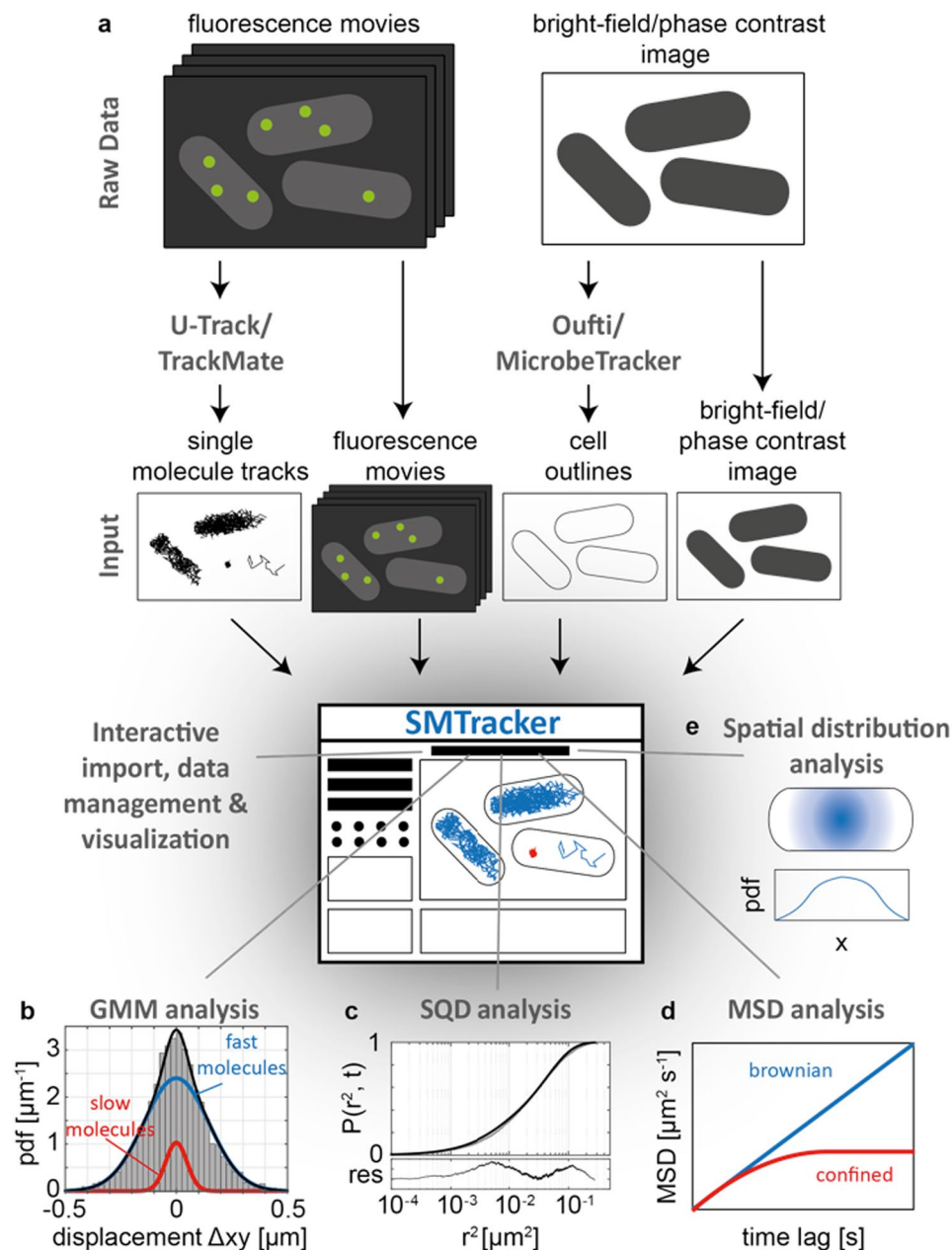


Figure 1. The SMTracker software suite. (a) After importing raw data of tracks and cell outlines, the import panel allows the user to interactively explore the data, identify characteristic patterns and potential experimental artifacts. (b) The Gaussian mixture model (GMM) analysis runs a fit on the probability density function (pdf) of positional 1D displacements, assuming a two- (or three-) state model with a mobile (blue line) and immobile (red line) subpopulation of molecules, thereby extracting the diffusion coefficient and fraction size of each subpopulation. (c) Squared displacement (SQD) analysis considers the cumulative probability function of squared displacements and provides an alternative way to analyze up to 3 different subpopulations in the biological sample. (d) The mean-squared displacement (MSD) analysis plots the time-ensemble averaged MSD vs. the time-lag, revealing the type of motion exhibited by the molecules. A linear dependency of the MSD on the time-lag indicates Brownian diffusion (blue line), whereas an asymptotic curve is indicative of confined (or sub-diffusive) motion. (e) In the spatial distribution analysis (SDA) panel the software generates 2D- and 3D- heat maps and distributions of trajectories along the x- and y-axis of the cell, allowing the identification of specific localization patterns of molecules, e.g., cytoplasmic, membrane- or nucleoid-bound.

$$F(\Delta x | \mu, \sigma) = \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{\Delta x - \mu}{\sqrt{2} \sigma} \right) \right]$$

be the cumulative distribution function (CDF) of the Gaussian distribution, we performed a least square fit of the function

$$G(\Delta x|\alpha, D_1, D_2) = \sum_{i=1}^q \alpha_i F(\Delta x|0, \sqrt{2D_i\Delta t}), \text{ with } \sum_{i=1}^q \alpha_i = 1,$$

to the experimental CDF, $H(\Delta x)$, by using a trust-region reflective Newton method (MATLAB, The MathWorks, Inc.). Note that the algorithm allows the user to either pool x - and y -displacements into a single distribution H (which is the preferred choice if x - and y -displacements are given in an image-centric coordinate system and cells are randomly oriented) or to treat x - and y -displacements independently (which is the preferred choice if the analysis is performed in a cell-centric coordinate system, with the x - and y -axis corresponding to the long and the short cell axis, respectively). To evaluate the goodness of fit, a Kolmogorov-Smirnov test has been applied.

An example of a protein investigated by GMM can be seen in suppl. Fig. S6: in the left panel, many small steps (around “0”) can be seen, and fewer larger steps (these are shown by the dotted line), while in the right panel, the protein has become much more mobile; now, the static fraction (dotted line) is much smaller than the one in the left panel. The red curves show two population fits, which explain the data well, while a single fit (assuming a single population), shown in green, could not explain the observed distribution of steps. The bubble plot GMM (Fig. S6) shows the size of the two populations, and their average diffusion constant on the y -axis. It can be seen that the right condition has an inverse relation between static and mobile molecules, revealing that the protein markedly changes its dynamics between the two conditions.

To better compare changes in molecular behavior between different treatments of cells, the algorithm optionally allows the user to keep the diffusion constants D_1 , D_2 (and D_3) identical between conditions and to attribute changes in the displacement distributions only to changes in the fractions α_i . A two-sample Kolmogorov-Smirnov test has been implemented in order to evaluate significant differences between the experimental cumulative distributions of displacements. The software also runs a model comparison test based on the Bayesian Information Criterion (BIC)²⁶, and displays which model (i.e. the number of diffusive states), explains the experimental data best, i.e., the model for which

$$BIC = \chi^2 + k \cdot \ln(n)$$

is minimal. Here, χ^2 denotes the residual between model and experimental CDF for the optimal parameter set, k the number of model parameters and n the number of experimental data points.

As an additional output the software also calculates the residence times for the molecules of each condition, e.g. how long a protein binds to DNA, by counting the time a molecule stays within a predefined radius R . This radius is based on the standard deviation of the immobile fraction of molecules, $\sigma_1 = \sqrt{2D_1\Delta t}$, such that it includes 99.7% of the smaller displacements ($R = 3 \cdot \sigma_1$), thereby also allowing the classification into confined and non-confined trajectories.

Mean-square displacement (MSD) analysis panel. The most common way of analyzing SMT data is the extraction of the time-ensemble averaged MSD, which is plotted as a function of the time lag (t_{lag})²⁷. While normal (Brownian) diffusion shows linear relation between the MSD and t_{lag} , confined motion is represented by a MSD that reaches a plateau for high t_{lag} and directed motion is characterized by parabolic MSD curve (Fig. 1d). By fitting the MSD plot with a linear function at short time-lags (Fig. S7), the diffusion coefficient can be estimated²⁸. To this end the user can interactively define the number of t_{lag} taken into account to linearly fit the MSD function, which condition and which direction (x -, y - or pooled xy -orientation) should be plotted. Additionally, the panel provides a measure for the localization error ξ of the microscopy system, which is derived from the y -axis intercept of the time-averaged MSD curves generated for single trajectories²⁸. To improve precision in the estimation of ξ the software limits this analysis to tracks with a high goodness of fit (R^2) value and a low diffusion coefficient.

Here, it should be noted that the MSD analysis is most appropriate if individual molecules are in one diffusive state characterized by a single diffusion constant. However, if individual molecules switch between diffusive states (e.g. by DNA binding and unbinding), the MSD analysis will only report an average diffusion constant for each molecule. Thus, even if molecules are in clearly distinct diffusive states, but if they switch between them, the time-averaging performed on each trajectory can lead to a whole spectrum of apparent diffusion constants between D_1 and D_2 .

Squared displacement (SQD) analysis panel. The SQD panel integrates an alternative way of analyzing the diffusive behavior of the molecules at the sub-population level (Fig. 1c), which was first described by Schütz *et al.*²⁹ and has been applied in a variety of studies^{5,30,31}. The basic idea is similar to the GMM analysis, but the SQD analysis considers the solution of the diffusion equation in terms of the 2-dimensional radial displacement $r(t)^2 = (x(t) - x(0))^2 + (y(t) - y(0))^2$, which leads for one diffusive molecule species to the cumulative distribution function

$$P(r^2, t) = 1 - \exp\left(-\frac{r^2}{\sigma(t)^2}\right),$$

with

$$\sigma(t) = \sqrt{4Dt}.$$

Accordingly, for multiple diffusive species with diffusion constants D_i ($i = 1, \dots, q$) and relative fractions α_i , the CDF reads

$$P(r^2, t) = 1 - \sum_{i=1}^q \alpha_i \cdot \left(\exp\left(-\frac{r^2}{4D_i t}\right) \right), \quad (1)$$

where $\sum_{i=1}^q \alpha_i = 1$. For a given number of diffusive states q , the algorithm implemented in the SQD panel performs a simultaneous nonlinear least square fit of Eq. (1) to the experimental CDFs for $t = 1\Delta t, 2\Delta t, 3\Delta t, 4\Delta t$ - thereby estimating the α_i and D_i ³². Here, the user can either choose manually between a one-, two- or three-state model to estimate the α_i and D_i of a given dataset, or to automatically choose the best-fitting model according to the BIC criterion (Fig. S8). Please note that in contrast to the GMM method, the CDF of squared displacements are fitted independently for each condition or treatment. For the evaluation of the difference between the CDFs of the modelled distributions of each pair of conditions, a two-sample Kolmogorov-Smirnov is performed.

Spatial distribution (SDA) panel. The spatial distribution panel summarizes the localization of the trajectories in 2D- and 3D-heat maps and provides normalized distributions for the localization in x - and y -orientation. To this end all trajectories are transformed into a cell-centric coordinate system and then scaled to a unit cell of $1 \mu m \times 1 \mu m$ (Fig. S9). This visualization allows distinguishing different localization patterns (cytoplasmic, nucleoid and membrane) (Fig. 1e), and the visualization of changes in localization patterns of a given protein in response to e.g. changes in environmental conditions, or by the absence of another protein thought to interact with the protein of interest.

Benchmarking SMTracker performance with synthetic SMT data. In order to validate the performance of SMTracker, we used computational simulations to generate sets of synthetic SMT data with known properties, i.e., diffusion constants and fraction sizes of molecules in the different diffusive states. Applying the GMM and SQD methods to the synthetic data then allows benchmarking the performance of each method as a function of simulation parameters, as detailed in the following.

Simulation of synthetic single molecule trajectories. Simulations of single molecule trajectories were performed on a total of $n_{tot} = n_1 + n_2$ molecules, where n_1 and n_2 are the number of molecules in a slow and fast diffusive state (with diffusion constants D_1 and D_2), respectively. For each subpopulation, Brownian motion of molecules was modeled by a 3-dimensional discrete-time random walk in x -, y -, and z -direction according to

$$x_{i+1} = x_i + \delta_x; \delta_x \sim N(0, \sqrt{2D\tau})$$

$$y_{i+1} = y_i + \delta_y; \delta_y \sim N(0, \sqrt{2D\tau})$$

$$z_{i+1} = z_i + \delta_z; \delta_z \sim N(0, \sqrt{2D\tau}).$$

Here, δ_x , δ_y and δ_z are Gaussian-distributed random variables with zero mean and standard deviation $\sigma = \sqrt{2D\tau}$, with $D \in \{D_1, D_2\}$ being the diffusion constant of the respective subpopulation and τ the time interval between simulation steps $i = 1, 2, \dots, N$. The initial particle positions (x_0, y_0, z_0) were randomly sampled according to a uniform distribution within a 3-dimensional, rod-shaped cell, formed by a cylinder and two hemispherical end caps with a radius of $0.5 \mu m$ and a total length of $3 \mu m$. From these starting points tridimensional random trajectories were simulated with a time lag of $\tau = 0.02$ ms. Reflecting boundary conditions at the cell membrane were implemented by rejecting (and re-drawing) all random moves ($\delta_x, \delta_y, \delta_z$) leading to positional vectors outside the rod-shaped cell. Additionally, the simulation accounts for a finite localization precision in typical experimental setups by adding a random vector ($\varepsilon_x, \varepsilon_y, \varepsilon_z$) to the “true” positions in the simulation, leading to the “observed” positions (X_i, Y_i, Z_i) at each point in time i

$$X_i = x_i + \varepsilon_x; \varepsilon_x \sim N(0, \sigma_{x,loc})$$

$$Y_i = y_i + \varepsilon_y; \varepsilon_y \sim N(0, \sigma_{y,loc})$$

$$Z_i = z_i + \varepsilon_z; \varepsilon_z \sim N(0, \sigma_{z,loc}).$$

here the ($\varepsilon_x, \varepsilon_y, \varepsilon_z$) are again normally distributed random numbers with zero mean and standard deviation $\sigma_{x,loc} = \sigma_{y,loc} = \sigma_{z,loc} = 30$ nm. For the generation of synthetic single molecule trajectories, we used a sampling interval of $\Delta t = 20$ ms, which is often used in experimental setups (see Tables 1 and 2 for a summary of all simulation parameters).

Benchmarking of GMM and SQD method performance. For benchmarking the performance of the GMM and SQD methods, we simulated a set of synthetic datasets varying in the diffusion constants D_1 and D_2 , as

Number of tracks	2000
Max number of points/track	20
Simulation time lag τ (ms)	0.02
Observation time lag (ms)	20
Number of simulation runs	30
Diffusion constant D_1 ($\mu\text{m}^2 \text{s}^{-1}$)	[0.01 0.05 0.1]
Diffusion constant D_2 ($\mu\text{m}^2 \text{s}^{-1}$)	[0.1 0.5 1]
Fraction size α (% molecules at diffusion rate D_1)	[20 40 60 80]
Localization error ξ (nm)	30

Table 1. Configuration of the synthetic trajectories.

Fitting upper boundaries [D_1, D_2] ($\mu\text{m}^2 \text{s}^{-1}$)	[1, 10]
Fitting lower boundaries [D_1, D_2] ($\mu\text{m}^2 \text{s}^{-1}$)	[0.0225, 0.0225]
Fit method	Nonlinear least-squares
Tolerance	10^{-8}

Table 2. Configuration of the fitting procedure.

well as in the fraction α of the slow-diffusive subpopulation of molecules (Figs 2–4). Each dataset consists of 30 simulation runs with 2000 single molecule tracks with a length of 20 frames. The tracks of each synthetic dataset were analyzed via the GMM and SQD methods and the resulting estimated parameters were compared with the true values entering the simulation (Fig. 2 – D_1 ; Fig. 3 – D_2 ; Fig. 4 – α). Overall, we found that both methods capture the diffusion constants and the fraction sizes of the two subpopulations with reasonable accuracy, provided that the diffusion constants are significantly different from each other and both range in a regime amenable to experimental resolution, as detailed further below.

When comparing both inference methods, it became apparent that the SQD method is more accurate in inferring the diffusion constant of the slower subpopulation (Fig. 2) while the GMM method is more precise in estimating the diffusion constant of the fast subpopulation (Fig. 3). Intuitively, this performance difference is related to the fact that the GMM method only takes into account the displacements between single frames, while the SQD method considers the displacements between multiple (up to 4) frames simultaneously. Accordingly, if the displacements are small (as for the case of the slowly diffusing fraction of molecules) the SQD method averages over more displacements and thereby achieves a higher accuracy in estimating the diffusion constant than the GMM method. Note, however, that both methods can only resolve diffusion constants larger than a lower limit determined by the finite localization precision σ_{loc} of the experimental setup, as given by $D_{loc} = \frac{\sigma_{loc}^2}{2\Delta t}$ (Fig. 2; *dashed lines*). On the other hand, if displacements are large (as for the case of the rapidly diffusing fraction of molecules) averaging the displacements over multiple frames leads to a slight underestimation of the diffusion constant by the SQD method (Fig. 3), because the longer the observation timescale, the more likely the trajectory is restricted by the confinement to the cell volume, leading to an apparent reduction in the observed displacements. To provide the user with the ability to flexibly adjust the accuracy of the SQD method to either slow or fast diffusive molecules we further implemented the option to choose between fits considering $t = 1\Delta t, 2\Delta t, 3\Delta t, 4\Delta t$ time frames or all of them simultaneously. Thus, in order to obtain high accuracy in the estimation of low diffusion constants the user should choose to take all time frames into account, whereas for accuracy in the estimation of high diffusion constants the best choice will be $t = 1\Delta t$. Under the latter conditions the GMM and SQD methods are in fact equivalent and yield identical results (Figs S10–S12). Comparing the estimated fraction sizes α shows that for most parameter combinations both methods show a comparable accuracy, with estimated fraction sizes deviating from the true fraction sizes by less than 10% (Fig. 4). Only if the diffusion constant of the slow and the rapid subpopulation are (almost) identical (Fig. 4D,G; $D_1 = 0.05, 0.1 \mu\text{m}^2/\text{s}$; $D_2 = 0.1 \mu\text{m}^2/\text{s}$), the variance in the estimated values of α increases, indicating that more than a 2-fold difference in diffusion constants is required to reliably infer the fraction sizes of two subpopulations.

Transition state regulator AbrB shows highly dynamic binding throughout the *Bacillus subtilis* nucleoids, and forms one or two subcellular clusters on the chromosomes. Next, we applied the SMTracker software to analyze experimentally acquired single molecule tracking data by focussing on the subcellular dynamics of the transcription factor AbrB in the Gram-positive model organism *Bacillus subtilis*. AbrB has long been known as a global regulator of genes whose expression is de-repressed as cells transition from exponential into stationary growth. Due to its structural similarity to small nucleoid-associated proteins in bacteria (often also called histone-like proteins), it has been speculated that AbrB may also confer a structural role in chromosome organization, in addition to its specific role in gene regulation. To investigate this point from a cell biological perspective, we generated a C-terminal fusion of YFP to AbrB in *B. subtilis*, which was still able to support biofilm formation (data not shown), indicating that it functions similar to wild type AbrB. The fusion

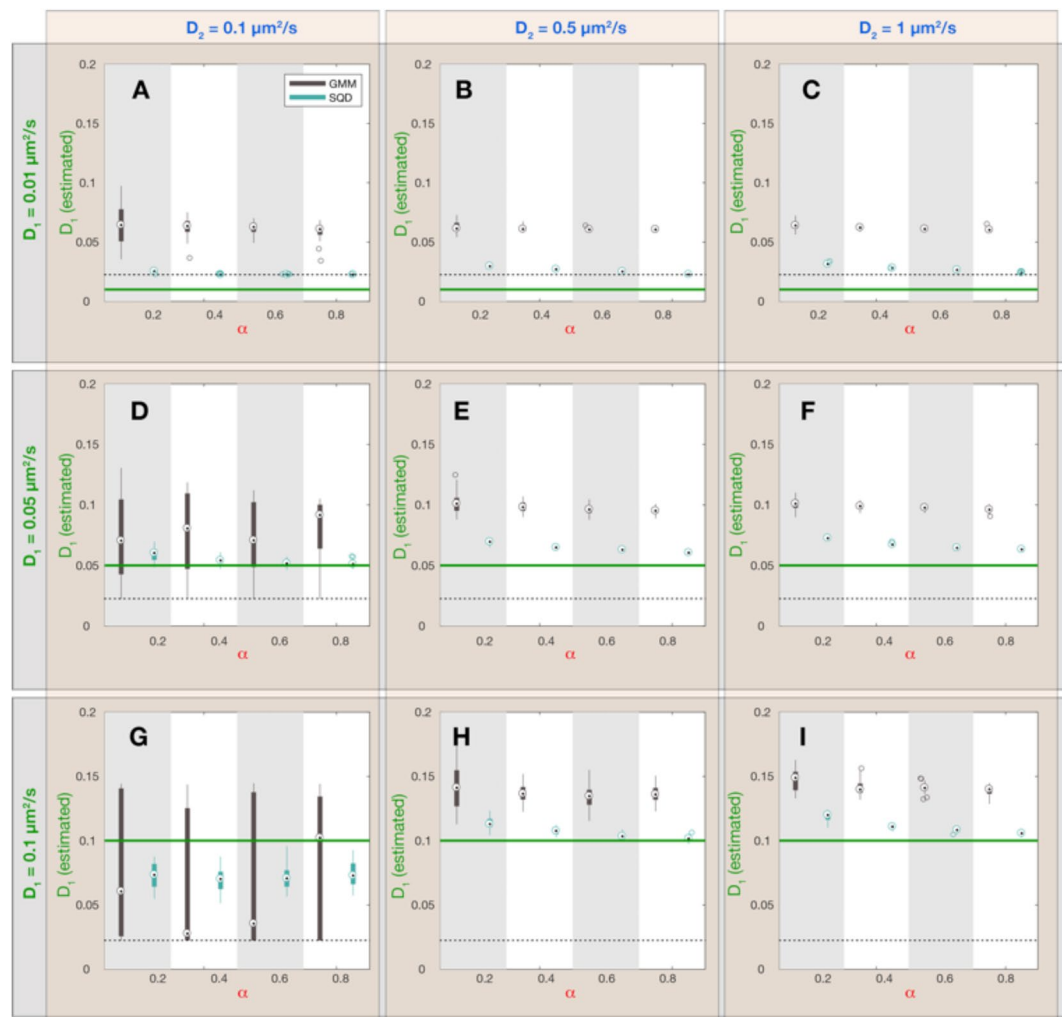


Figure 2. Dependence of estimated diffusion constant D_1 (slow subpopulation) on simulation parameters and inference method. Each panel (A–I) corresponds to a combination of simulation parameters D_1 and D_2 , as indicated by the green and blue labels outside the plot panels, respectively. Within each panel 4 values of the fraction sizes α were chosen and synthetic simulations were performed on all parameter combinations. For each parameter combination (D_1 , D_2 and α) 30 simulations with 2000 tracks of 20 frames duration (see Table 1 for all simulation parameters) were executed and the resulting data was analyzed with the GMM and SQD method, yielding two different estimates for the value of D_1 (GMM: black boxplots; SQD: cyan boxplots). On each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points not considered outliers and the outliers are plotted as individual circles. The green horizontal lines indicate the true value for D_1 in the simulation. The black dashed lines indicate the lower limit $D_{loc} = \frac{\sigma_{loc}^2}{2\Delta t} = 0.0225 \frac{\mu m^2}{s}$ for an estimate of the diffusion constant, given a finite localization precision $\sigma_{loc} = 30 \text{ nm}$ and a frame rate of $\Delta t = 20 \text{ ms}$ used in our simulations.

was expressed from the original gene locus as sole source of the protein in the cell, under control of the original promoter, ensuring that wild type levels are expressed.

Using 15 ms stream acquisitions, we obtained 1900 tracks from 66 cells, having an average duration of 180 ms. Using SMTracker, we characterized AbrB dynamics using above-mentioned features. Figure 5A shows a projection of all fluorescence intensities obtained in a chain of cells over time, revealing that AbrB is mainly restricted to the centrally located chromosome(s). In Fig. 5C, the positions of AbrB-YFP tracks in a slow/static state (red) and mobile state (blue) are shown within a standardized cell of $3 \times 1 \mu m$, revealing that AbrB binds throughout the entire nucleoid, and thus throughout the chromosome. This is in agreement with its more than 600 binding sites distributed across the genome²¹. GMM analyses show that 68% of the time AbrB molecules move with a diffusion constant of $0.084 \mu m^2/s$, whereas in 32% of the time they move with a diffusion constant of $0.8 \mu m^2/s$ (Fig. 5E). Interestingly, the diffusion constant for the fast subpopulation of AbrB (total mass of 156 kDa including YFP) is very similar to the diffusion constant of $0.53 \mu m^2/s$ for the freely diffusive fraction of the Smc dimer³³ (total mass of 330 kDa including YFPs), and $0.51 \mu m^2/s$ for the freely diffusive fraction of the DnaA dimer⁹ (total mass of 156 kDa including YFPs), suggesting that this corresponds to the freely diffusive state of AbrB. In contrast, the previously reported diffusion constant for the DNA-bound state of Spo0J ($0.02 \mu m^2/s$)⁹ is about 4-fold lower than

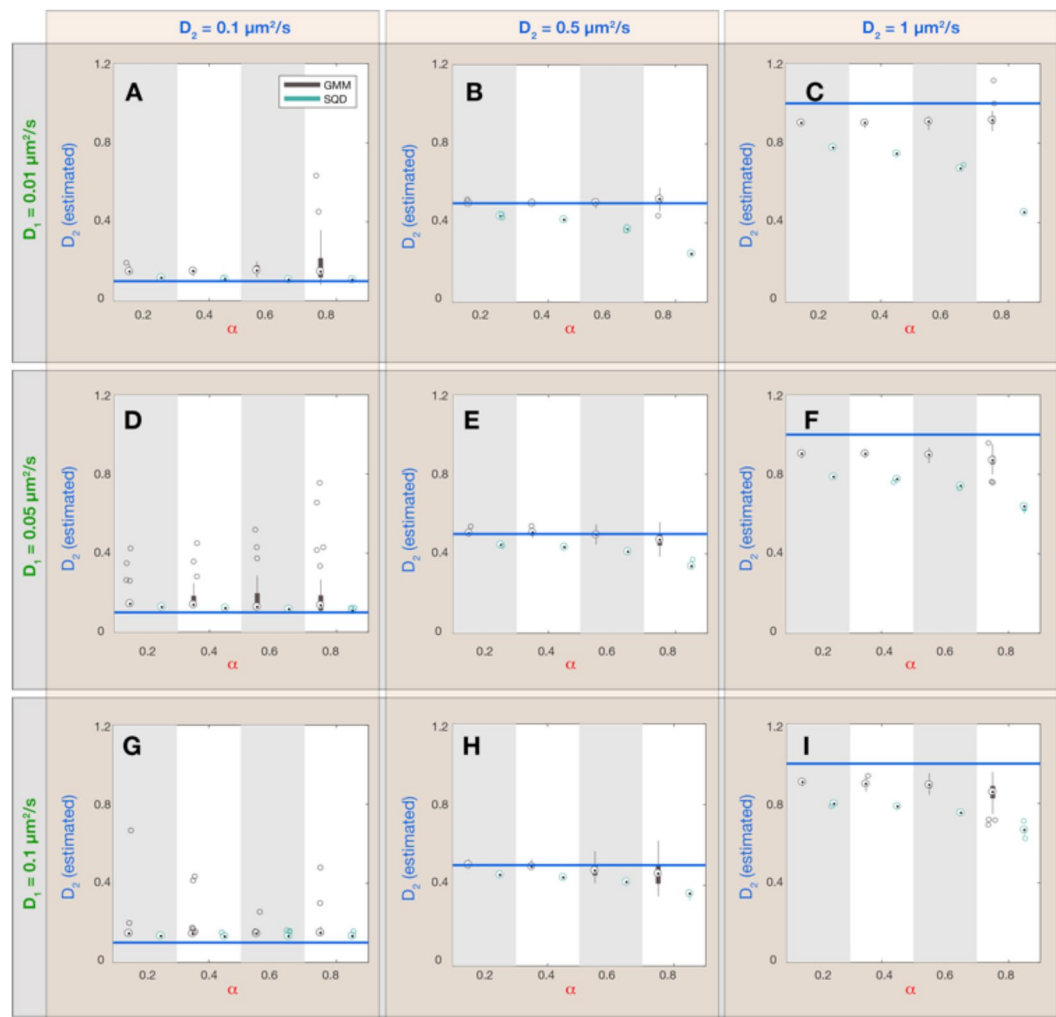


Figure 3. Dependence of estimated diffusion constant D_2 (fast subpopulation) on simulation parameters and inference method. Each panel (A–I) corresponds to a combination of simulation parameters D_1 and D_2 , as indicated by the green and blue labels outside the plot panels, respectively. Within each panel 4 values of the fraction sizes α were chosen and synthetic simulations were performed on all parameter combinations. For each parameter combination (D_1 , D_2 and α) 30 simulations with 2000 tracks of 20 frames duration (see Table 1 for all simulation parameters) were executed and the resulting data was analyzed with the GMM and SQD method, yielding two different estimates for the value of D_2 (GMM: black boxplots; SQD: cyan boxplots). On each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points not considered outliers and the outliers are plotted as individual circles. The blue horizontal lines indicate the true value for D_2 in the simulation.

that for the static subpopulation of AbrB ($0.084 \mu\text{m}^2/\text{s}$), suggesting that this value might be the result of a mixture between two diffusive states, for instance by AbrB frequently interacting with DNA in a non-specific manner, in search of its binding sites. Indeed, a more fine-grained analysis using the 3-state SQD approach support the presence of 3 populations: $D = 0.019/0.1/0.81 \mu\text{m}^2/\text{s}$; $F = 35/40/25\%$, suggesting that 35% of AbrB molecules are bound to their target DNA sites, while 40% move along DNA in a constrained manner, and 25% freely diffuse as tetramers.

Figure 5D shows SMT-derived dwell time determination of AbrB-YFP, revealing two distinct fractions – one with 38 ms, and the other with 110 ms. The latter fraction likely represents the DNA-bound AbrB fraction, and because its dwell time is shorter than the average life time of AbrB tracks, it reveals that binding of AbrB is very dynamic *in vivo*. It should be noted that actual dwell time *in vivo* will be somewhat longer, because of YFP bleaching due to the experimental setup. To test if bleaching strongly influences dwell time determination, we used a YFP-MreB fusion and determined bleaching kinetics of static MreB molecules (Fig. 6), which are bound within filaments that run underneath the cell membrane mostly perpendicular to the long axis of cells³⁴, and play an important role in the maintenance of cell morphology. Average half-lifetime of YFP molecules was 1.4 s, much longer than the estimated dwell time of AbrB-YFP, showing that YFP bleaching does not strongly affect dwell time determination of AbrB. Interestingly, an average dwell time of 110 ms was found for 30% of AbrB-YFP molecules (Fig. 5E), in good agreement with the 35% of static AbrB molecules as deduced from SQD analyses. Thus, AbrB

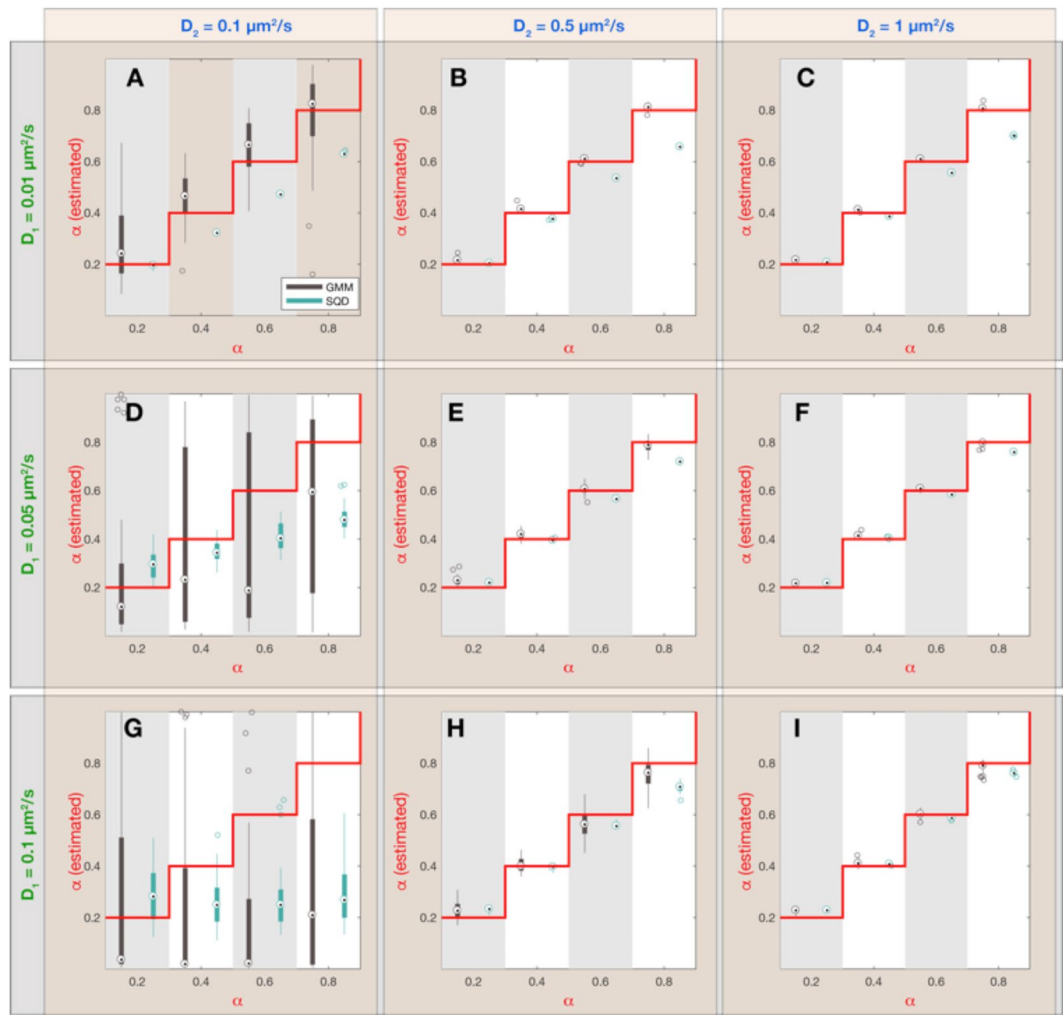


Figure 4. Dependence of estimated fraction α (slow subpopulation) on simulation parameters and inference method. Each panel (A–I) corresponds to a combination of simulation parameters D_1 and D_2 , as indicated by the green and blue labels outside the plot panels, respectively. Within each panel 4 values of the fraction sizes α were chosen and synthetic simulations were performed on all parameter combinations. For each parameter combination (D_1 , D_2 and α) 30 simulations with 2000 tracks of 20 frames duration (see Table 1 for all simulation parameters) were executed and the resulting data was analyzed with the GMM and SQD method, yielding two different estimates for the value of α (GMM: black boxplots; SQD: cyan boxplots). On each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points not considered outliers and the outliers are plotted as individual circles. The red lines indicate the true values for α in the simulation.

binding sites are occupied by individual AbrB molecules for only short times (LacI repressor has average DNA binding times of few minutes³⁵), revealing high molecule turnover on the chromosome.

A further interesting finding is shown by overlaying all tracks obtained by SMT acquisition, which reveals the position of static tracks. Figure 5B shows that AbrB-YFP forms static foci in cells, either one or two, on the nucleoids. Although we can presently not determine if these sites correspond to any specific site on the chromosome (e.g. origin regions), and if they confer an important function, it is interesting to note that this is a parallel to global transcriptional silencer H-NS in *E. coli*, an important nucleoid-associated protein with many hundreds of binding sites. In contrast to other nucleoid-associated proteins such as HU, Fis and IHF, which are scattered throughout the genome, H-NS forms one or two subcellular clusters on the nucleoids that are important for nucleoid shaping³⁶. It is therefore possible that AbrB confers a similar structural function in *B. subtilis*.

Conclusions

SMTTracker software provides a user-friendly interface for comprehensive visualization and analysis of SMT data that can be readily printed in publication-quality figures in an effortless and automated way. It provides an intuitive pipeline to compare the diffusion of molecules under varying conditions in terms of fraction size, spatial distribution and binding times. Our software was benchmarked on a computer-simulated set of tracking data, exposing the advantages of individual SMT analyses under different experimental conditions, and thereby

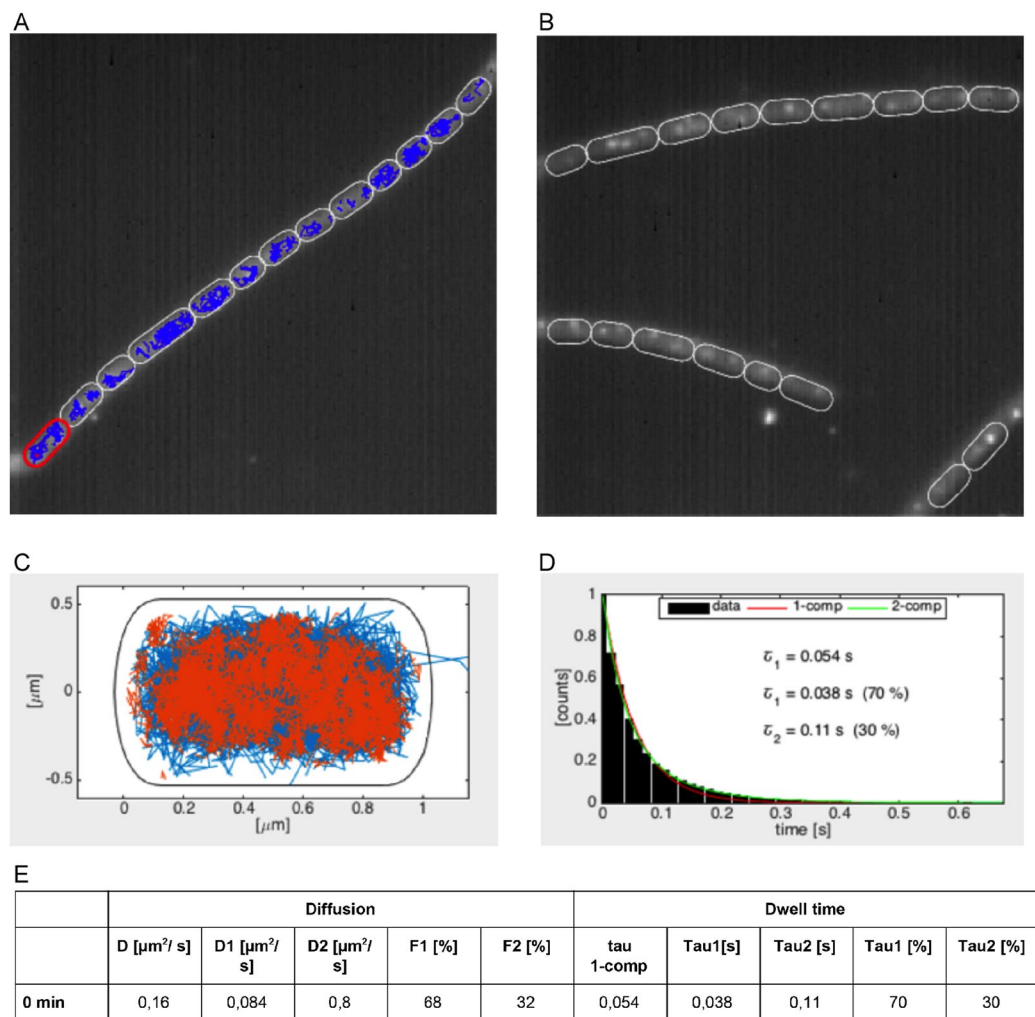


Figure 5. SMT analyses of AbrB-YFP from *Bacillus subtilis*. **(A)** Projection of all frames of a typical stream acquisition overlaid with tracks (in blue) detected by U-track (left panel), **(B)** and a projection of frames without tracks. **(C)** Heat map showing mobile AbrB-YFP tracks in blue, and static tracks in red. **(D)** dwell time determination, **(E)** Diffusion constants and average dwell times of AbrB-YFP.

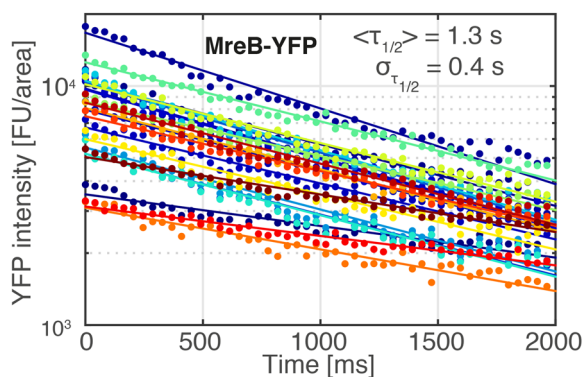


Figure 6. Bleaching of YFP-MreB in single *B. subtilis* cells. Image acquisition parameters were identical to those used in Fig. 5. Exponential fits (solid curves) to the bleaching curves of 23 single cells (dots) revealed an average bleaching half-life time of $\langle \tau_{1/2} \rangle = 1.3 \text{ s}$ with a standard deviation of $\sigma_{\tau_{1/2}} = 0.4 \text{ s}$.

highlighting the usefulness of combining multiple approaches of SMT data analysis in the SMTracker suite. Using this pipeline, we show that the genome-wide repressor AbrB in *B. subtilis* shows three fractions of different mobility, with one third of the molecules being bound to DNA throughout the genome, about one third sliding along

DNA in search of binding sites, and another third freely diffusing. As AbrB has very short dwell times on DNA, our data suggest that there is a high exchange between molecules being bound and diffusing along DNA, and between the latter fraction and freely diffusing molecules. Although SMTracker was used for SMT in rod-shaped bacterial cells, SMTracker can be applied to SMT experiments in cells with different morphologies even up to eukaryotic cells (e.g. *Schizosaccharomyces pombe*). We expect that SMTracker will enable non-specialized scientists to answer advanced questions in (bacterial) cell biology without the need to deeply engage in programming and the underlying mathematical analyses. Finally, SMTracker allows data export to other tools such as SMMtrack⁹ and the vbSPT software³⁷ (see suppl. material), which analyzes SMT data on the basis of a Hidden Markov model to extract diffusion coefficients and transition rates between diffusive states.

Materials and Methods

Bacterial strains and growth conditions. AbrB was fused to YFP by amplification of 500 bp from the 3' end of the *abrB* gene (excluding the stop codon) and by Gibson assembly into pSG1164³⁸. *B. subtilis* strain PY79 was transformed with the resulting plasmid, such that a fusion of the *abrB* gene with *yfp* at its C-terminus was generated at the original gene locus, which was under control of the *abrB* promoter. As *abrB* is monocistronic, no downstream gene was affected. For microscopy, cells were grown in S7₅₀ minimal medium at room temperature until mid-exponential growth, and were imaged under laid by agarose containing growth medium.

Single molecule tracking of AbrB. Imaging was performed with a Nikon Ti-E microscope configured with a high numerical aperture objective (CFI Apochromat TIRF 100XC Oil, NA 1.49), an EM-CCD camera (ImagEM X2, Hamamatsu) and an filter set for imaging YFP molecules (YFP HC Filterset; BrightLine 500/24, Beamsplitter 520 and BrightLine 542/27). Specimen were continuously illuminated with the central part of an expanded laser beam (TOPTICA Beam Smart, 515 nm, max. power 100 mW) with an intensity of 160–300 W/cm² and streams were recorded at a frame rate of ~67 Hz using VisiView (Visitron Systems).

Software implementation. SMTracker was developed for Windows and Mac OS X running MATLAB R2014b and later. The graphical user interface (GUI) of SMTracker relies on the GUI layout Toolbox (Ben Tordoff and David Sampson, Consulting Services group at MathWorks), which is freely available at <https://www.mathworks.com/matlabcentral/fileexchange/47982-gui-layout-toolbox>. In addition, SMTracker requires the following MATLAB toolboxes: Curve Fitting, Statistics and Machine Learning, Optimization, Image Processing and Parallel Computing. SMTracker is available under GPLv3 license at <https://sourceforge.net/projects/singlemoleculetracker>, where updated versions can be downloaded for direct use as well as for further extension and modification by experienced developers. The authors request acknowledgment of the use of SMTracker in published works.

Data Availability Statement

All data are available within the manuscript.

References

- Shashkova, S. & Leake, M. C. Single-molecule fluorescence microscopy review: shedding new light on old problems. *Biosci. Rep.* **37**, <https://doi.org/10.1042/BSR20170031> (2017).
- Stracy, M., Uphoff, S., Garza de Leon, F. & Kapanidis, A. N. *In vivo* single-molecule imaging of bacterial DNA replication, transcription, and repair. *FEBS Lett.* **588**, 3585–3594, <https://doi.org/10.1016/j.febslet.2014.05.026> (2014).
- Kusumi, A., Tsunoyama, T. A., Hirose, K. M., Kasai, R. S. & Fujiwara, T. K. Tracking single molecules at work in living cells. *Nat. Chem. Biol.* **10**, 524–532, <https://doi.org/10.1038/nchembio.1558> (2014).
- Zawadzki, P. *et al.* The Localization and Action of Topoisomerase IV in *Escherichia coli* Chromosome Segregation Is Coordinated by the SMC Complex, MukBEF. *Cell Rep.* **13**, 2587–2596, <https://doi.org/10.1016/j.celrep.2015.11.034> (2015).
- Kleine Borgmann, L. A., Ries, J., Ewers, H., Ulbrich, M. H. & Graumann, P. L. The bacterial SMC complex displays two distinct modes of interaction with the chromosome. *Cell Rep.* **3**, 1483–1492, <https://doi.org/10.1016/j.celrep.2013.04.005> (2013).
- Stracy, M. *et al.* Live-cell superresolution microscopy reveals the organization of RNA polymerase in the bacterial nucleoid. *Proc. Natl. Acad. Sci. USA* **112**, E4390–4399, <https://doi.org/10.1073/pnas.1507592112> (2015).
- English, B. P. *et al.* Single-molecule investigations of the stringent response machinery in living bacterial cells. *Proc. Natl. Acad. Sci. USA* **108**, E365–373, <https://doi.org/10.1073/pnas.1102255108> (2011).
- Li, W. *et al.* Effects of amino acid starvation on RelA diffusive behavior in live *Escherichia coli*. *Mol. Microbiol.* **99**, 571–585, <https://doi.org/10.1111/mmi.13252> (2016).
- Schenk, K. *et al.* Rapid turnover of DnaA at replication origin regions contributes to initiation control of DNA replication. *PLoS genetics* **13**, e1006561, <https://doi.org/10.1371/journal.pgen.1006561> (2017).
- Stracy, M. *et al.* Single-molecule imaging of UvrA and UvrB recruitment to DNA lesions in living *Escherichia coli*. *Nat. Commun.* **7**, 12568, <https://doi.org/10.1038/ncomms12568> (2016).
- Uphoff, S., Reyes-Lamothe, R., Garza de Leon, F., Sherratt, D. J. & Kapanidis, A. N. Single-molecule DNA repair in live bacteria. *Proc. Natl. Acad. Sci. USA* **110**, 8063–8068, <https://doi.org/10.1073/pnas.1301804110> (2013).
- Liao, Y., Schroeder, J. W., Gao, B., Simmons, L. A. & Biteen, J. S. Single-molecule motions and interactions in live cells reveal target search dynamics in mismatch repair. *Proc. Natl. Acad. Sci. USA* **112**, E6898–6906, <https://doi.org/10.1073/pnas.1507386112> (2015).
- Chenouard, N. *et al.* Objective comparison of particle tracking methods. *Nat. Methods* **11**, 281–289, <https://doi.org/10.1038/nmeth.2808> (2014).
- El Beheiry, M., Dahan, M. & Masson, J. B. InferenceMAP: mapping of single-molecule dynamics with Bayesian inference. *Nat. Methods* **12**, 594–595, <https://doi.org/10.1038/nmeth.3441> (2015).
- Vallotton, P. *et al.* Diatrack particle tracking software: Review of applications and performance evaluation. *Traffic* **18**, 840–852, <https://doi.org/10.1111/tra.12530> (2017).
- Olson, A. L. *et al.* Structure and DNA-binding traits of the transition state regulator AbrB. *Structure* **22**, 1650–1656, <https://doi.org/10.1016/j.str.2014.08.018> (2014).
- Hamon, M. A., Stanley, N. R., Britton, R. A., Grossman, A. D. & Lazazzera, B. A. Identification of AbrB-regulated genes involved in biofilm formation by *Bacillus subtilis*. *Mol. Microbiol.* **52**, 847–860, <https://doi.org/10.1111/j.1365-2958.2004.04023.x> (2004).

18. Strauch, M. A. *et al.* The transition state transcription regulator *abrB* of *Bacillus subtilis* is a DNA binding protein. *EMBO J.* **8**, 1615–1621 (1989).
19. Oliveira, P. & Lindblad, P. An *AbrB*-Like protein regulates the expression of the bidirectional hydrogenase in *Synechocystis* sp. strain PCC 6803. *J. Bacteriol.* **190**, 1011–1019, <https://doi.org/10.1128/JB.01605-07> (2008).
20. Lieman-Hurwitz, J. *et al.* A cyanobacterial *AbrB*-like protein affects the apparent photosynthetic affinity for CO₂ by modulating low-CO₂-induced gene expression. *Environ. Microbiol.* **11**, 927–936, <https://doi.org/10.1111/j.1462-2920.2008.01818.x> (2009).
21. Chumsakul, O. *et al.* Genome-wide binding profiles of the *Bacillus subtilis* transition state regulator *AbrB* and its homolog *Abh* reveals their interactive role in transcriptional regulation. *Nucleic Acids Res.* **39**, 414–428, <https://doi.org/10.1093/nar/gkq780> (2011).
22. Jaqaman, K. *et al.* Robust single-particle tracking in live-cell time-lapse sequences. *Nat. Methods* **5**, 695–702, <https://doi.org/10.1038/nmeth.1237> (2008).
23. Tinevez, J. Y. *et al.* TrackMate: An open and extensible platform for single-particle tracking. *Methods* **115**, 80–90, <https://doi.org/10.1016/j.ymeth.2016.09.016> (2017).
24. Sliusarenko, O., Heinritz, J., Emonet, T. & Jacobs-Wagner, C. High-throughput, subpixel precision analysis of bacterial morphogenesis and intracellular spatio-temporal dynamics. *Mol. Microbiol.* **80**, 612–627, <https://doi.org/10.1111/j.1365-2958.2011.07579.x> (2011).
25. Paintdakhi, A. *et al.* Oufiti: an integrated software package for high-accuracy, high-throughput quantitative microscopy analysis. *Mol. Microbiol.* **99**, 767–777, <https://doi.org/10.1111/mmi.13264> (2016).
26. Kass, R. E. & Raftery, A. E. Bayes Factors. *J. Am. Stat. Assoc.* **90**, 773–795 (1995).
27. Manzo, C. & Garcia-Parajo, M. F. A review of progress in single particle tracking: from methods to biophysical insights. *Rep. Prog. Phys.* **78**, 124601, <https://doi.org/10.1088/0034-4885/78/12/124601> (2015).
28. Michalet, X. Mean square displacement analysis of single-particle trajectories with localization error: Brownian motion in an isotropic medium. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **82**, 041914, <https://doi.org/10.1103/PhysRevE.82.041914> (2010).
29. Schutz, G. J., Schindler, H. & Schmidt, T. Single-molecule microscopy on model membranes reveals anomalous diffusion. *Biophys. J.* **73**, 1073–1080, [https://doi.org/10.1016/S0006-3495\(97\)78139-6](https://doi.org/10.1016/S0006-3495(97)78139-6) (1997).
30. Kuzmenko, A. *et al.* Single molecule tracking fluorescence microscopy in mitochondria reveals highly dynamic but confined movement of Tom40. *Sci. Rep.* **1**, 195, <https://doi.org/10.1038/srep00195> (2011).
31. Yang, F. *et al.* Single-molecule dynamics of the molecular chaperone trigger factor in living cells. *Mol. Microbiol.* **102**, 992–1003, <https://doi.org/10.1111/mmi.13529> (2016).
32. Haas, B. L., Matson, J. S., DiRita, V. J. & Biteen, J. S. Single-molecule tracking in live *Vibrio cholerae* reveals that *ToxR* recruits the membrane-bound virulence regulator *TcpP* to the *toxT* promoter. *Mol. Microbiol.* **96**, 4–13, <https://doi.org/10.1111/mmi.12834> (2015).
33. Schibany, S. *et al.* Single molecule tracking reveals that the bacterial SMC complex moves slowly relative to the diffusion of the chromosome. *Nucleic Acids Res.* **46**, 7805–7819 (2018).
34. Reimold, C., Defeu Soufo, H. J., Dempwolff, F. & Graumann, P. L. Motion of variable-length *MreB* filaments at the bacterial cell membrane influences cell morphology. *Mol. Biol. Cell* **24**, 2340–2349, <https://doi.org/10.1091/mbc.E12-10-0728> (2013).
35. Elf, J., Li, G. W. & Xie, X. S. Probing transcription factor dynamics at the single-molecule level in a living cell. *Science* **316**, 1191–1194, <https://doi.org/10.1126/science.1141967> (2007).
36. Wang, W., Li, G. W., Chen, C., Xie, X. S. & Zhuang, X. Chromosome organization by a nucleoid-associated protein in live bacteria. *Science* **333**, 1445–1449, <https://doi.org/10.1126/science.1204697> (2011).
37. Persson, F., Linden, M., Unoson, C. & Elf, J. Extracting intracellular diffusive states and transition rates from single-molecule tracking data. *Nat. Methods* **10**, 265–269, <https://doi.org/10.1038/nmeth.2367> (2013).
38. Lucena, D., Mauri, M., Schmidt, F., Eckhardt, B. & Graumann, P. L. Microdomain formation is a general property of bacterial membrane proteins and induces heterogeneity of diffusion patterns. *BMC Biol.* **16**, 97, <https://doi.org/10.1186/s12915-018-0561-0> (2018).

Acknowledgements

This work was supported by the Deutsche Forschungsgemeinschaft (SFB 987) and the SYNMIKRO Center for Synthetic Microbiology, funded by the LOEWE program of the state of Hessen.

Author Contributions

T.C.R. conceived of the program, performed data analyses and wrote the manuscript, L.M.O.B. performed simulation experiments and helped develop SMTracker, G.F. conceived of the G.M.M. method, helped develop SMTracker and wrote the manuscript, P.L.G. conceived of the study and wrote the manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-33842-9>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018