Extracting intracellular diffusive states and transition rates from single-molecule tracking data

Fredrik Persson[1,3], Martin Lindén[2,3], Cecilia Unoson[1] & Johan Elf[1]

We provide an analytical tool based on a variational Bayesian treatment of hidden Markov models to combine the information from thousands of short single-molecule trajectories of intracellularly diffusing proteins. The method identifies the number of diffusive states and the state transition rates. Using this method we have created an objective interaction map for Hfq, a protein that mediates interactions between small regulatory RNAs and their mRNA targets.

Most of our knowledge about the kinetics of intracellular processes is based on observed responses to perturbations. This strategy is severely limited by the difficulty of inducing gentle and specific perturbations *in vivo*. Single-particle tracking (SPT), in contrast, offers the possibility of monitoring intracellular kinetics of some processes directly. SPT using photoactivatable fluorescent proteins (spt-PALM) is capable of following the movements of individual protein molecules on cell membranes and in the cytoplasm of living cells[1–3]. It is also possible to classify individual trajectories into coarsely predefined states of diffusion corresponding, for example, to different states of binding, through the size dependence of the diffusion constant[3]. This suggests that it should be possible to determine intracellular state transition rates under steady-state conditions simply by determining the average time an individual molecule spends in one diffusion state before changing to another. In this way, spt-PALM makes it possible to overcome the averaging that normally masks all information about transition rates in a steady-state ensemble.

Two major challenges have to be overcome to study kinetics by tracking individual cytoplasmic proteins. First, there has been no objective way to determine the number of diffusive states on the basis of experimental data alone. Instead, it is commonly assumed that there exist two states of diffusion, bound and free, and then the data are simply fitted under these assumptions[3,4]. Second, intracellular single-molecule diffusion trajectories from fluorescent fusion proteins are typically very short in contrast to trajectories from membrane proteins that can be extracellularly labeled with photostable dyes. The challenge is to correctly combine the information from thousands of short trajectories (fewer than ~20 steps) instead of following a few long trajectories with multiple state transitions.

RESULTS

vbSPT for single-molecule tracking data

We provide an analytical tool, variational Bayes SPT (vbSPT), based on a variational Bayesian treatment of a hidden Markov model (HMM)[5,6] that describes diffusing particles making memoryless jumps between different states of diffusion.

Previous HMMs for diffusing particles use a fixed number of states (most often two or three) and individual long trajectories[7–9]. In contrast, our vbSPT method is capable of learning model parameters such as transition rates, as well as the number of diffusive states, from the experimental data. Furthermore, the method is able to extract useful information even from data sets with only a few points per trajectory.

The vbSPT method uses a maximum-evidence criterion to determine the underlying parameters and the number of diffusive states from the observed data. In finding the correct number of states, there is a trade-off between goodness of fit and model complexity. On the one hand, models with more states and parameters can generally achieve a better fit to the data. On the other hand, these models also have a broader descriptive repertoire spread out over a larger parameter space; therefore, 'simple' data are more unlikely to originate from a complex model. The maximum-evidence criterion strikes a balance between these tendencies, analogous to Occam's razor. The vbSPT method (**Supplementary Software**; http://www.sourceforge.net/projects/vbspt/) computes the evidence using a mean-field approximation (Online Methods).

Validation and performance

We validated vbSPT using simulated reaction diffusion trajectories in a cellular geometry with the same trajectory length distribution as that of our experimental data and using realistic localization error (20 nm), diffusion and transition parameters (**Supplementary Note 1**). The method accurately recovers the parameters used for simulating the data (**Fig. 1** and **Supplementary Tables 1** and **2**), including mean dwell times several times longer than the typical trajectory length. Confronted

[1]Department of Cell and Molecular Biology, Science for Life Laboratory, Uppsala University, Uppsala, Sweden. [2]Center for Biomembrane Research, Department of Biochemistry and Biophysics, Stockholm University, Stockholm, Sweden. [3]These authors contributed equally to this work. Correspondence should be addressed to J.E. (johan.elf@icm.uu.se).

RECEIVED 11 JULY 2012; ACCEPTED 20 DECEMBER 2012; PUBLISHED ONLINE 10 FEBRUARY 2013; DOI:10.1038/NMETH.2367

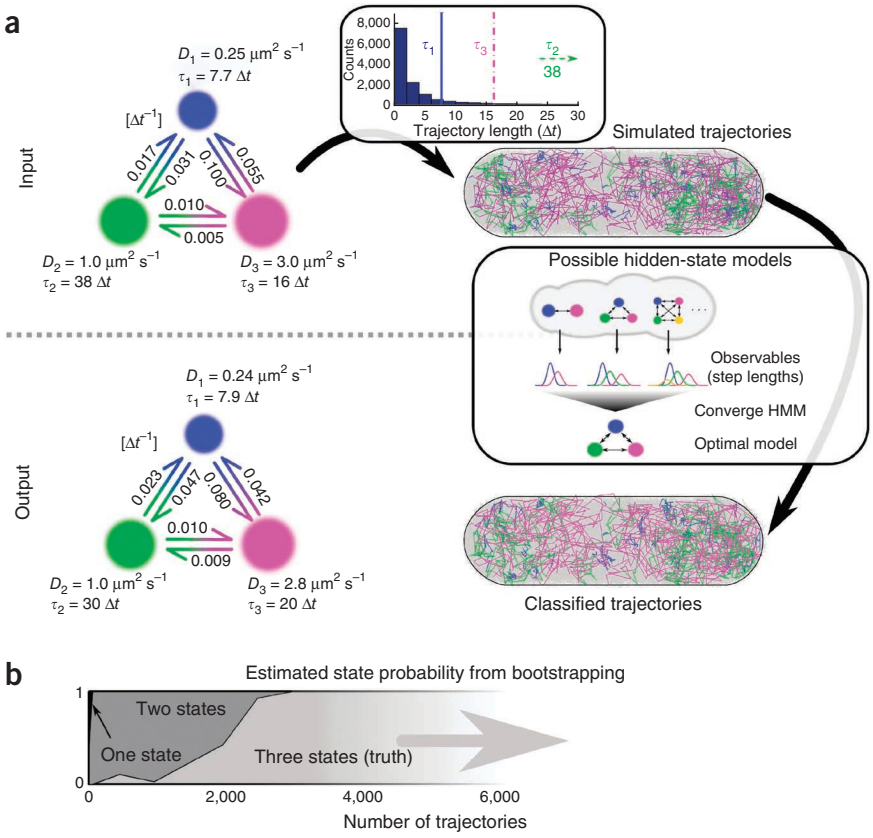

Figure 1 | A test of the vbSPT approach using simulated data. Shown are 12,130 trajectories simulated with an experimental length distribution and in a realistic cell geometry (length, 2.8 μm). (**a**) Top, model, parameters and resulting trajectories for the synthetic data set. In the model, the area of each state represents its relative occupancy. A subset of 1,000 trajectories is color coded on the basis of diffusive state, defined by diffusion constant (D) and mean dwell time (τ). Bottom, model and parameters found by vbSPT analysis. Standard errors from a bootstrap analysis are typically less than 20% (**Supplementary Table 2**). The same trajectories as above are color coded according to the inferred most likely state. (**b**) Distribution of the most likely number of states with an increasing number of trajectories, from bootstrap analysis (truth = number of states in input data).

with synthetic data in which some transitions are absent (**Supplementary Note 1** and **Supplementary Table 3**), the algorithm assigns rates for these transitions corresponding to approximately one transition in 100 frames, which sets a scale for the slowest transitions we consider significant within our experimental data.

To determine how many trajectories are needed to identify the underlying model, we varied the number of simulated trajectories included in the analysis. We observed that the algorithm correctly identified three underlying states when more than 3,000 trajectories were used (**Fig. 1**).

The diffusion constants, state occupancies and transition probabilities for the states in the three-state model converge for an increasing number of trajectories sampled from a length distribution corresponding to that of our experimental single-molecule tracking data (**Fig. 2**). We performed this analysis for three different localization errors (s.d. of 20 nm, 30 nm and 40 nm, respectively; **Supplementary Note 1**). We found that the diffusion rates and the state occupancies converge at a few thousand trajectories, whereas transition probabilities require more than ~10,000 trajectories.

The transition probabilities (**Fig. 2c**) display a clear bias toward 0.05 per time step for small amounts of data. This value represents an a priori assumption about the dynamics (prior value), an input parameter for the analysis. The decay in the bias reflects how an increasing amount of data drowns out such assumptions. In addition, keeping the amount of data constant while increasing the average trajectory length gives a modest further improvement (**Supplementary Note 1**).

The probability that the vbSPT algorithm can resolve two similar states depends not only on the amount of data and the localization error but also on how close the diffusive states are and how often the states interconvert in relation to the frame rate (**Supplementary Note 1**). For example, with average state lifetimes of ~10 time steps, and diffusion sampled in one spatial dimension, vbSPT can resolve differences of about 50% in diffusion constants given the number of trajectories and sampling conditions in **Figure 1**. The resolution is improved with analysis of the intracellular diffusion not only along the long axis of the cell but in two or three dimensions as well. The vbSPT algorithm supports multiple spatial dimensions, but in our case, boundary artifacts in the narrow *Escherichia coli* cells compromised the improvement.

Changes in the diffusion constant of a protein can occur for reasons other than binding or dissociation reactions. We show two examples of how vbSPT behaves in response to synthetic data in which the diffusion constant depends on the spatial coordinates (**Fig. 3**). When applied to a model with a nucleoidal region with slower diffusion (**Fig. 3a** and **Supplementary Table 4**), the method returns a two-state model with the slower state enriched in the nucleoidal region and at the cell poles. The latter enrichment is an artifact caused by confinement[3], which alone is not enough to induce a spurious slower state in the case of simple uniform diffusion (**Fig. 3b**). In the case of a continuously varying diffusion constant (**Fig. 3c** and **Supplementary Table 5**), the analysis returns a three-state model, representing the range of the gradient and indicating that the transition between the states is largely sequential. When we classified the individual steps into the most likely states, we found that their localization highly resembles the simulated diffusion gradient. Trajectories from the fastest state occupy the largest space, which is expected because molecules that move fast reach further before they have been sampled sufficiently to be classified into a specific state.

These data (**Fig. 3**) also illustrate two potential overinterpretations of the results of vbSPT analysis. First, the algorithm may group similar diffusive states into one state unless there is sufficient evidence to tell them apart. Second, the transition rates

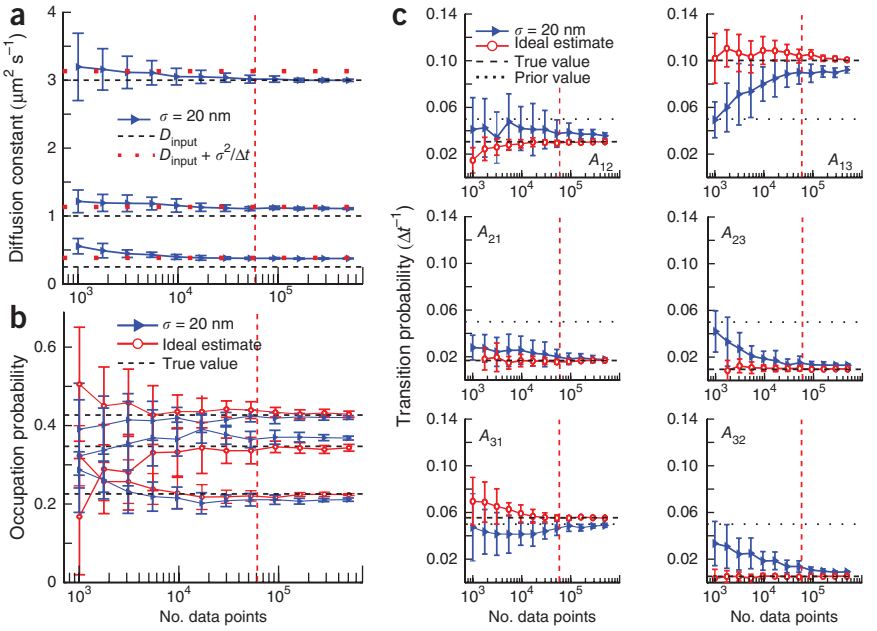

Figure 2 | Convergence properties for the model presented in **Figure 1** with localization error (σ) of 20 nm. All plotted values are averages of posterior mean values from the vbSPT analysis (Online Methods), from 50 independent simulated data sets, as a function of the number of data points. Error bars, 1 s.d. (n = 50). (**a**–**c**) Diffusion constants (**a**), occupation probabilities (**b**) and transition probabilities (**c**). In **a**, diffusion constants converge rapidly toward the apparent value (dotted red lines), for which the contribution of the localization error has been included, except for the fastest state, whose diffusion constant converges to the true value, D_{input} (dashed black lines). This reflects a downward bias in diffusion constants due to the finite cell size[3], an effect that increases with increasing diffusion constant, and, in this case, cancels the contribution from the localization error. In **b**, occupation probabilities converge toward the true values (dashed black lines) but not as fast as the ideal estimates (red circles) based on perfect knowledge of the hidden states at every data point. In **c**, the six subpanels represent the probabilities (A_{ij}) of transitioning from state i to state j during one time step (frame). Dotted lines indicate the mean transition probabilities of the prior. The dashed vertical red lines indicate the amount of experimental data.

do not necessarily represent chemical reactions but rather the probability per time step to change between the identified states: in this case, simply the rate of moving between cellular regions with substantially different diffusion properties. Furthermore, dynamics in two-dimensional reaction diffusion systems, such as membranes, should be interpreted with caution. This is because the underlying assumption in vbSPT, that state transitions are memoryless, will often be false in two-dimensional systems[10].

The performance and limitations of vbSPT depend on system-specific factors such as geometry, underlying kinetics and measurement noise. Hence, although the above examples (see also **Supplementary Note 1**) indicate the behavior in our particular system, applications to other systems should be guided by tests on synthetic data related to the specific system. We include a basic simulation tool for this purpose with the software (**Supplementary Software** and on SourceForge, http://www.sourceforge.net/projects/vbspt/).

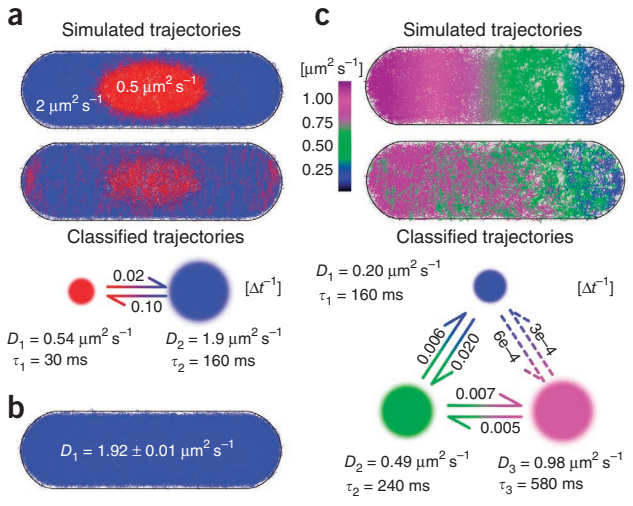

Figure 3 | Test of vbSPT on synthetic data with a spatially varying diffusion constant in a cell geometry. All synthetic data have an experimental length distribution containing 23,756 trajectories and an average length of five frames. (**a**) Analysis of synthetic data including a nucleoidal region with a diffusion constant one-fourth that of the surrounding area. The model and classified trajectories show the two states found by vbSPT analysis. All visualizations represent a 250-nm-thick central cross-section with slower trajectories overlaid for visibility. (For standard errors from a bootstrap analysis, see **Supplementary Table 4**.) (**b**) Trajectories classified by vbSPT analysis for a synthetic data set lacking the nucleoidal region. (**c**) Analysis of synthetic data with a continuously increasing diffusion rate along the cell long axis. The classified trajectories represent a subset of 5,000 trajectories overlaid on the rest for visibility. (For standard errors from a bootstrap analysis, see **Supplementary Table 5**.)

Application to Hfq kinetics

To test the method on a biological system, we studied the RNA helper protein Hfq[11,12] (UniProtKB: P0A6X3), which mediates post-transcriptional gene regulation by facilitating interactions between mRNA and noncoding small RNA (sRNA). Hfq was chromosomally expressed as a fusion protein with a monomeric photoconvertible protein, Dendra2 (**Supplementary Note 2**). Cells expressing the Hfq-Dendra2 fusion protein exhibited wild-type growth rates, whereas an *hfq* deletion strain was growth impaired (**Supplementary Note 3**). We tracked the Hfq-Dendra2 protein at a frame rate of approximately 300 frames per second (**Supplementary Note 4**), which is necessary for tracking freely diffusing proteins[3]. This enabled us to collect 1,000–3,000 single-molecule diffusion trajectories per *E. coli* cell and 10,000–30,000 trajectories per experiment.

Hfq is a hexameric protein that binds to both sRNAs and mRNAs to mediate base-pairing. Considering that mRNA can be found either free or coupled to other macromolecules such as DNA during transcription or ribosomes during translation, we expected to find Hfq in binding states of very different diffusivity.

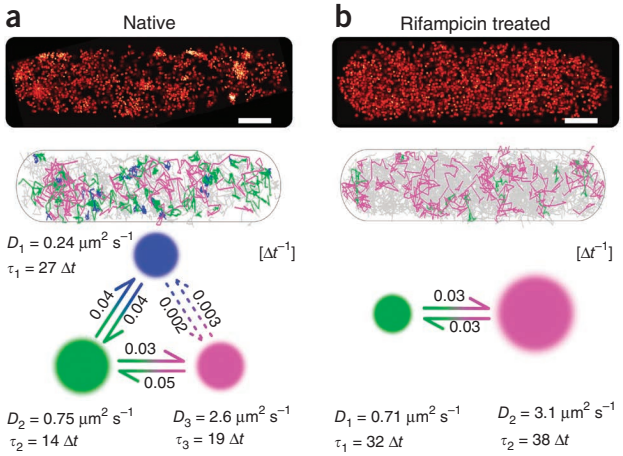

Figure 4 | vbSPT analysis of experimental tracking data on Hfq. (**a**,**b**) Untreated cells (**a**) and cells treated with the transcription inhibitor rifampicin (**b**). Top, typical super-resolution localization plot from a single cell. Scale bars, 0.5 μm. Center, ~1,500 trajectories taken from an arbitrary cell. Trajectories longer than seven steps are color coded by the most likely state found by vbSPT analysis. Bottom, models including transition rates, diffusion constants (*D*) and mean dwell times (τ) found by the analysis. The area represents the relative occupation of each state. (For standard errors from a bootstrap analysis, see **Supplementary Tables 6** and **7**.) The dashed transitions represent rare and therefore uncertain transitions (**Supplementary Note 1**).

To extract the number of binding states and transition rates from the Hfq tracking data, we applied the vbSPT analysis with an uninformative prior distribution to a data set of 12,130 trajectories from a single experiment on approximately ten cells. The best fit was achieved by a three-state model (**Fig. 4a**), with diffusion constants of approximately 3 $\mu m^2 s^{-1}$, 0.7 $\mu m^2 s^{-1}$ and 0.2 $\mu m^2 s^{-1}$. We repeated the experiment, obtaining an unaltered outcome, and determined the accuracy of the parameter estimates by bootstrap analysis[13] (**Supplementary Table 6**). Next we performed experiments on cells treated with the antibiotic rifampicin (rif), which inhibits transcription and leads to global RNA decay, as we confirmed by RNA decay studies (**Supplementary Note 5**). For the rif-treated cells we obtained a two-state model, in which the two states have diffusion constants corresponding to the two faster states for untreated cells (3 $\mu m^2 s^{-1}$ and 0.7 $\mu m^2 s^{-1}$) (**Fig. 4b** and **Supplementary Table 7**).

To interpret the model, we mapped the different states of Hfq diffusion to states of Hfq binding. As Hfq is found mostly in its hexameric form[14,15], we propose that the fastest state (3 $\mu m^2 s^{-1}$) corresponds to the 223-kDa Hfq-Dendra2 hexamer. After rif treatment, the slowest state vanishes, and the fraction occupying the intermediate state decreases substantially (**Fig. 4**), indicating that these states may represent interactions of Hfq with different RNA species, either free or under processing. The slowest state may represent Hfq bound to RNA being transcribed, which would be expected to disappear upon transcription inhibition.

The decrease in the transition rate from the fast to the slow state after rif treatment (**Fig. 4**) could be explained by the decreasing number of RNAs available for binding. The conversion from the slow to the fast state could reflect the rates of Hfq dissociation from RNA and/or of RNA degradation, both of which should not be altered much by rif treatment.

When analyzing the combined data from both Hfq experiments on untreated cells (23,756 trajectories), we obtained sufficient evidence to split the intermediate state into two states (**Supplementary Note 1**). This is consistent with our interpretation that the intermediate state represents binding to mRNA molecules of varying lengths and consequently varying rates of diffusion. After we added another 18,903 trajectories from an additional experiment, the four state model remained the most likely, but we observed by bootstrap analysis that the likelihood of higher-state models increased in this case (**Supplementary Note 1**). We did not observe state splitting upon pooling data from rif-treated cells, most probably because the RNA-associated pool of Hfq is diminished under this condition because of global RNA decay (**Supplementary Note 5**).

DISCUSSION

The vbSPT method opens new possibilities for objective characterization of intracellular processes on the basis of single-molecule tracking data. With vbSPT it is possible to obtain kinetic information from a system at steady state, which is important because most intracellular systems are impossible to perturb gently and with high specificity. We note that the intracellular environment in many cases will include interactions that cannot be found using purified components in the test tube. These include transient or rare interactions as well as complex interactions in uncharted pathways that can be probed only by altering the genetic background or growth conditions.

The vbSPT algorithm currently defines states by diffusion constants and state lifetimes. However, the variational framework is flexible and could be extended to many types of more complex state descriptions by, for example, taking state-dependent localization errors into account.

The current analysis method can resolve states with average lifetimes of one time step, but accurate inference of kinetics is limited to lifetimes of several time steps. Limitations on slow kinetics are set by the data and factors including the underlying model, trajectory lengths, localization error and amount of data. The range of accessible kinetics might be extended by modifying the algorithm to combine data acquired at different frame rates.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Supplementary information is available in the online version of the paper.

ACKNOWLEDGMENTS

We thank I. Barkefors for her careful and critical reading of the manuscript. M.L. is grateful to C.H. Wiggins and J.-W. van de Meent for insightful discussions. This work was supported by the European Research Council (J.E.), the Knut and Alice Wallenberg Foundation (J.E.), Vetenskapsrådet (J.E.), the Göran Gustafsson Foundation (J.E.), the Wenner-Gren Foundations (M.L.) and the Center for Biomembrane Research (M.L.).

AUTHOR CONTRIBUTIONS

J.E. and M.L. conceived the method, M.L. designed the vbSPT algorithm, M.L. and F.P. implemented and tested the algorithm, F.P. designed and implemented the image analysis and particle-tracking algorithms, and C.U. cloned and characterized the bacterial strains. F.P. and J.E. built the optical setup. F.P., C.U. and J.E. designed the experiments, and C.U. and F.P. performed the experiments. F.P., M.L., C.U. and J.E. wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/doifinder/10.1038/nmeth.2367.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Manley, S. *et al.* High-density mapping of single-molecule trajectories with photoactivated localization microscopy. *Nat. Methods* **5**, 155–157 (2008).
2. Niu, L. & Yu, J. Investigating intracellular dynamics of FtsZ cytoskeleton with photoactivation single-molecule tracking. *Biophys. J.* **95**, 2009–2016 (2008).
3. English, B.P. *et al.* Single-molecule investigations of the stringent response machinery in living bacterial cells. *Proc. Natl. Acad. Sci. USA* **108**, E365–E373 (2011).
4. Bakshi, S., Siryaporn, A., Goulian, M. & Weisshaar, J.C. Superresolution imaging of ribosomes and RNA polymerase in live *Escherichia coli* cells. *Mol. Microbiol.* **85**, 21–38 (2012).
5. Bronson, J.E., Fei, J., Hofman, J.M., Gonzalez, R.L. Jr. & Wiggins, C.H. Learning rates and states from biophysical time series: a Bayesian approach to model selection and single-molecule FRET data. *Biophys. J.* **97**, 3196–3205 (2009).
6. Bishop, C.M. *Pattern Recognition and Machine Learning* (Springer, 2006).
7. Das, R., Cairo, C.W. & Coombs, D. A hidden Markov model for single particle tracks quantifies dynamic interactions between LFA-1 and the actin cytoskeleton. *PLoS Comput. Biol.* **5**, e1000556 (2009).
8. Chung, I. *et al.* Spatial control of EGF receptor activation by reversible dimerization on living cells. *Nature* **464**, 783–787 (2010).
9. Beausang, J.F. *et al.* DNA looping kinetics analyzed using diffusive hidden Markov model. *Biophys. J.* **92**, L64–L66 (2007).
10. Mahmutovic, A., Fange, D., Berg, O.G. & Elf, J. Lost in presumption: stochastic reactions in spatial models. *Nat. Methods* **9**, 1163–1166 (2012).
11. Vogel, J. & Luisi, B.F. Hfq and its constellation of RNA. *Nat. Rev. Microbiol.* **9**, 578–589 (2011).
12. Waters, L.S. & Storz, G. Regulatory RNAs in bacteria. *Cell* **136**, 615–628 (2009).
13. Efron, B. Bootstrap methods: another look at the jackknife. *Ann. Stat.* **7**, 1–26 (1979).
14. Link, T.M., Valentin-Hansen, P. & Brennan, R.G. Structure of *Escherichia coli* Hfq bound to polyriboadenylate RNA. *Proc. Natl. Acad. Sci. USA* **106**, 19292–19297 (2009).
15. Fender, A., Elf, J., Hampel, K., Zimmermann, B. & Wagner, E.G. RNAs actively cycle on the Sm-like protein Hfq. *Genes Dev.* **24**, 2621–2626 (2010).

ONLINE METHODS

The statistical approach of vbSPT is to model the state kinetics by a hidden Markov model (HMM) for diffusing particles with memoryless jumps in diffusion constants, which we analyze by an approximate approach to maximum-evidence model selection, known as variational Bayes or ensemble learning[6,16]. Variational algorithms for HMMs have been derived earlier[5,16–20] and applied successfully in a biophysical setting to, for example, single-molecule FRET data[5,17,21]. The main advantage of the variational maximum-evidence approach over the more common maximum-likelihood approach to HMMs is the inherent complexity control, i.e., the ability not only to learn parameter values from data but also to make model selection and learn the number of hidden states[17]. An implementation of vbSPT for MATLAB (MathWorks), including a graphical user interface, code to generate synthetic data and a user manual, is included as **Supplementary Software** and can be downloaded as open source from SourceForge (http://www.sourceforge.net/projects/vbspt/).

Model selection by maximum evidence. In this section, we briefly introduce our model selection criterion, maximum evidence. For simplicity, we use x, s, θ and N to denote tracking data, hidden-states sequences corresponding to the data, unknown parameters in the model and the number of hidden states, respectively. For readers new to Bayesian statistics, we also recommend the brief introduction by Eddy[22] (or textbooks[6,16]).

A probabilistic model specifies the probability of obtaining the data x and hidden states s, given a model N and some parameter values θ: $p(x,s|\theta,N)$. To use this for Bayesian model selection, we treat all parameters as random variables; use the laws of probability to derive the inverse probability $p(N|x)$, which we interpret as a statement about our degree of confidence in model N given the data x; and prefer the most likely model. This will require us to specify prior distributions, of the form $p(N)$ and $p(\theta|N)$, that express our beliefs about the models and their parameters before seeing the data x. Indeed, the willingness to assign a probability distribution to unknown parameters and treat them on an equal footing with other random variables is a characteristic of Bayesian statistics.

Returning to the derivation of $p(N|x)$, we first combine our model and prior belief about the parameters into a probability distribution over the data and all unknown quantities, $p(x,s,\theta|N) = p(x,s|\theta,N)p(\theta|N)$. We then get rid of θ and s by marginalizing over them (or, in physics jargon, integrating them out), leading to

$$p(x|N) = \sum_s \int d\theta\, p\left(x,s\middle|\theta,N\right) p(\theta|N) \tag{1}$$

The sought inverse probability is then obtained from Bayes' rule, $p(N|x) = p(x|N)p(N)/p(x)$, or

$$p(N|x) = \frac{1}{p(x)} \int d\theta \sum_s p\left(x,s\middle|\theta,N\right) p(\theta|N) p(N) \tag{2}$$

where the denominator is a normalization constant

$$p(x) = \sum_{N'} \int d\theta' p\left(x\middle|\theta',N'\right) p\left(\theta'|N'\right) p(N')$$

If we further assume that $p(N)$ is constant, i.e., that all models (in some interval $0 < N < N_{\max}$) are equally probable a priori, and note that the denominator $p(x)$ is independent of parameters and models, then the best model is the one that maximizes the numerator in equation (2), also known as the 'evidence'

$$N_{\mathrm{ME}} = \mathrm{argmax}_N \int d\theta \sum_s p\left(x,s\middle|\theta,N\right) p(\theta|N) \tag{3}$$

where $\mathrm{argmax}_y f(y)$ denotes the value of y that maximizes the function f.

In practice, the integrals and sums in the evidence are intractable for almost all interesting models, and further progress requires good approximations and computers. The difficulties resemble those of computing partition functions in statistical physics, and the two common fallbacks in that field—Monte Carlo simulations[16,23,24] and mean-field theory[6,16,17]—work here as well. The variational or ensemble learning approach we will follow corresponds to mean-field theory, where we seek a separable approximation $(q(\theta)q(s))$ to the posterior distribution over parameters and hidden states

$$p\left(s,\theta\middle|x,N\right) = \frac{p\left(s,x\middle|\theta,N\right)p(\theta|N)}{p(x|N)} \approx q(s)q(\theta) \tag{4}$$

i.e., where the hidden states s and parameter values θ are statistically independent. We refer the interested reader to the vbSPT software documentation for details and derivations.

Estimated parameter values. The approximative posterior distribution for the parameters and hidden states produced by the variational approach, $q(s)q(\theta)$, can be used for approximate inference but is inconvenient to visualize. We therefore use point estimates to communicate and visualize our results. Among several possible choices (mean values, most likely values, medians, etc.), we present mean values unless stated otherwise. Explicit expressions are given in the software documentation.

The posterior distributions could also supply estimates of s.d., but the variational approximation tends to underestimate the width of the posterior distributions (ref. 16, Ch. 33). Instead, we use the bootstrap method[13] for this purpose, with bootstrap sets from a random selection (with replacement) of individual trajectories.

Diffusion model. We assume that we can track the particles in d dimensions, and we call $\mathbf{x}_t$ the position at time t. The time between consecutive measurements is Δt, but we are going to use t as an integer index as well. We model the movement by simple diffusion

$$\mathbf{x}_{t+1} = \mathbf{x}_t + \sqrt{2D_{s_t}\Delta t}\,\mathbf{w}_t \tag{5}$$

where $\mathbf{w}_t$ are independent d-dimensional Gaussian variables with uncorrelated components of zero mean and unit variance.

Binding and unbinding events are modeled as jumps in the diffusion constant D_{s_t}, as indicated by the degree of freedom $s_t \in \{1, 2, \ldots, N\}$, where N is the number of diffusive states. This degree of freedom constitutes our hidden state, and we model it

doi:10.1038/nmeth.2367

as a discrete Markov process with a transition matrix A and initial state probabilities $\boldsymbol{\pi}$, i.e.

$$p(s_1 = j) = \pi_j, p\left(s_t = j \middle| s_{t-1} = i\right) = A_{ij}, \text{ for } t > 1 \tag{6}$$

where normalization demands that

$$\sum_{k=1}^{N} \pi_k = \sum_{k=1}^{N} A_{jk} = 1$$

The number of hidden states N is a parameter to be determined from the data. The vector of diffusion constants corresponding to the different hidden states is denoted $\mathbf{D} = (D_1, D_2, \ldots, D_N)$. We will also use the shorthand notation $\mathbf{x}_{1:T} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T\}$ and $s_{1:T-1} = \{s_1, s_2, \ldots, s_{T-1}\}$ for the positions and hidden states of a whole trajectory (because s_T does not influence the position data, we exclude it from the model).

For our analysis, we need expressions for the joint distribution of positions, hidden states and parameters, $p(\mathbf{x}_{1:T}, s_{1:T-1}, \mathbf{D}, A, \boldsymbol{\pi} | N)$, which can be factorized as

$$\begin{aligned} p(\mathbf{x}_{1:T}, s_{1:T-1}, \mathbf{D}, A, \boldsymbol{\pi} | N) &= p(\mathbf{x}_{1:T} | s_{1:T-1}, \mathbf{D}) p(s_{1:T-1} | A, \boldsymbol{\pi}) \\ &\times p(\mathbf{D} | N) p(A | N) p(\boldsymbol{\pi} | N) \end{aligned} \tag{7}$$

The first two factors on the above right-hand side are specified by the model. In particular, the distribution of hidden state sequences is

$$p(s_{1:T-1} | A, \boldsymbol{\pi}) = \pi_{s_1} \prod_{t=1}^{T-2} A_{s_t s_{t+1}} \tag{8}$$

and the distribution of positions, conditional on the hidden states, is given by

$$\begin{aligned} p(\mathbf{x}_{1:T} | s_{1:T-1}, \mathbf{D}) &= p(\mathbf{x}_1) \prod_{t=1}^{T-1} p(\mathbf{x}_{t+1} - \mathbf{x}_t | s_t, \mathbf{D}) \\ &= p(\mathbf{x}_1) \prod_{t=1}^{T-1} \frac{1}{\left(4\pi D_{s_t} \Delta t\right)^{d/2}} e^{-\frac{1}{4\pi D_{s_t} \Delta t} \mathbf{x}_t^2} \end{aligned} \tag{9}$$

The initial position distribution $p(\mathbf{x}_1)$ does not depend on any parameter of interest to us here, and we will drop it from the analysis.

The last three factors in equation (7) are the prior distributions for the parameters, expressing our beliefs about the parameter values for different model sizes (values of N), before seeing the data. For computational convenience, independent priors, for example, $p(\mathbf{D}|N)p(A|N)p(\boldsymbol{\pi}|N)$, are used instead of the more general $p(\mathbf{D}, A, \boldsymbol{\pi}|N)$. For the same reason, priors' functional forms are chosen as conjugate priors (ref. 6, Ch. 2.4.2), which in this case means inverse gamma distributions for the diffusion constant priors, and Dirichlet distributions (a multidimensional version of the beta distribution) for the initial state and transition probability priors. Within these constraints, we strive to choose uninformative, or weak, priors to let the data speak for themselves as much as possible. More details are given in **Supplementary Note 1**, where we spell out and motivate our prior choices in detail and also investigate the robustness of our inference results against variations in the transition probability prior.

Treatment of many trajectories. Typical *in vivo* SPT experiments produce many rather short trajectories. Most trajectories contain fewer than 20 consecutive positions, and probably very few (≤2) transitions. If considered in isolation, such trajectories are not very informative, and transitions are difficult to identify accurately. To extract meaningful information, we need to pool many trajectories. The simplest way to do that is to assume that they are statistically independent and governed by the same model and parameter set, i.e., that all molecules (including those from different cells in the same batch) are equivalent and that the interaction dynamics do not change over time. In that case, the probability distribution for a set of M trajectories is a product of single-trajectory densities governed by the same model

$$p\left(\left\{\mathbf{x}_{1:T}^i, s_{1:T-1}^i\right\}_{i=1}^{M} \middle| \mathbf{D}, A, \boldsymbol{\pi}, N\right) = \prod_{i=1}^{M} p\left(\mathbf{x}_{1:T}^i, s_{1:T-1}^i \middle| \mathbf{D}, A, \boldsymbol{\pi}, N\right) \tag{10}$$

where different trajectories are indicated by the index i.

16. MacKay, D.J.C. *Information Theory, Inference, and Learning Algorithms* (Cambridge University Press, 2003).
17. Bronson, J.E., Hofman, J.M., Fei, J., Gonzalez, R.L. Jr. & Wiggins, C.H. Graphical models for inferring single molecule dynamics. *BMC Bioinformatics* **11** (suppl. 8), S2 (2010).
18. MacKay, D.J.C. Ensemble learning for hidden Markov models. ⟨http://www.inference.phy.cam.ac.uk/mackay/abstracts/ensemblePaper.html⟩ (1997).
19. Ghahramani, Z. An introduction to hidden Markov models and Bayesian networks. in *Hidden Markov Models: Applications in Computer Vision* (eds. Bunke, H. & Caelli, T.) 9–42 (World Scientific, River Edge, New Jersey, USA, 2001).
20. Beal, M.J. *Variational Algorithms for Approximate Bayesian Inference*. PhD thesis, Univ. College London (2003).
21. Okamoto, K. & Sako, Y. Variational Bayes analysis of a photon-based hidden Markov model for single-molecule FRET trajectories. *Biophys. J.* **103**, 1315–1324 (2012).
22. Eddy, S.R. What is Bayesian statistics? *Nat. Biotechnol.* **22**, 1177–1178 (2004).
23. Green, P.J. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82**, 711–732 (1995).
24. Robert, C.P., Rydén, T. & Titterington, D.M. Bayesian inference in hidden Markov models through the reversible jump Markov chain Monte Carlo method. *J. R. Stat. Soc., B* **62**, 57–75 (2000).