Bayesian Approach to MSD-Based Analysis of Particle Motion in Live Cells

Citation: Monnier, Nilah, Syuan-Ming Guo, Masashi Mori, Jun He, Peter Lenart, and Mark Bathe. "Bayesian Approach to MSD-Based Analysis of Particle Motion in Live Cells." Biophysical Journal 103, no. 3 (August 2012): 616–626. © 2012 Biophysical Society

As Published: http://dx.doi.org/10.1016/j.bpj.2012.06.029

Publisher: Elsevier

Persistent URL: http://hdl.handle.net/1721.1/88695

Version: Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

Bayesian Approach to MSD-Based Analysis of Particle Motion in Live Cells

Nilah Monnier,[†‡] Syuan-Ming Guo,[†] Masashi Mori,[§] Jun He,[†] Péter Lénárt,[§] and Mark Bathe[†*]
[†]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts; [‡]Graduate Program in Biophysics, Harvard University, Cambridge, Massachusetts; and [§]Cell Biology and Biophysics Unit, European Molecular Biology Laboratory, Heidelberg, Germany

ABSTRACT Quantitative tracking of particle motion using live-cell imaging is a powerful approach to understanding the mechanism of transport of biological molecules, organelles, and cells. However, inferring complex stochastic motion models from single-particle trajectories in an objective manner is nontrivial due to noise from sampling limitations and biological heterogeneity. Here, we present a systematic Bayesian approach to multiple-hypothesis testing of a general set of competing motion models based on particle mean-square displacements that automatically classifies particle motion, properly accounting for sampling limitations and correlated noise while appropriately penalizing model complexity according to Occam's Razor to avoid over-fitting. We test the procedure rigorously using simulated trajectories for which the underlying physical process is known, demonstrating that it chooses the simplest physical model that explains the observed data. Further, we show that computed model probabilities provide a reliability test for the downstream biological interpretation of associated parameter values. We subsequently illustrate the broad utility of the approach by applying it to disparate biological systems including experimental particle trajectories from chromosomes, kinetochores, and membrane receptors undergoing a variety of complex motions. This automated and objective Bayesian framework easily scales to large numbers of particle trajectories, making it ideal for classifying the complex motion of large numbers of single molecules and cells from high-throughput screens, as well as single-cell-, tissue-, and organism-level studies.

INTRODUCTION

Advances in high spatial and temporal resolution imaging of fluorescently tagged biological molecules, organelles, and cells are increasingly enabling the collection of detailed time-series data on the positions of these particles over time within living cells, tissues, and embryos using conventional and superresolution microscopy (1–5). The resulting single-particle trajectories (SPTs) contain important information on the transport dynamics and local environments of individual biological molecules, the collective behaviors of molecules and cells, and the spatial and temporal regulation of these behaviors (6–15). In most biological applications, the underlying mode of particle motion is unknown a priori and must be inferred using mathematical models from data sets that are limited by experimental parameters including sampling rate, acquisition time, and number of trajectories (16). In addition, the stochastic nature of SPTs requires careful treatment of trajectory variability and noise properties to facilitate objective model evaluation (6).

Despite the importance of analyzing SPT motion in biological systems, systematic and automated means of evaluating multiple competing motion models are still lacking, with interpretation of SPTs typically relying on time-consuming data analysis with significant manual intervention that focuses on evaluating a particular motion model such as confined or anomalous diffusion. Although such approaches allow for the testing of a single hypothesis or a pair of competing hypotheses for defined biological applications (17–19), standardized approaches that allow side-by-side comparison of larger sets of generalized complex motion models without any constraints on model form, trajectory length, or number are needed for higher-throughput systematic biological studies, as well as to standardize the analysis of SPTs across laboratories (6). Here, we present an approach based on Bayesian inference for multiple-hypothesis testing (20–23), which has proven successful in handling noise and experimental limitations in other biological applications (24–29). This approach focuses on evaluating models of stationary, time-invariant physical processes governing the motion of single particles undergoing free, confined, or anomalous diffusion, with or without directed transport superimposed.

Stationary physical processes are characterized by ensemble average distribution functions such as the mean-square displacement (MSD), which is commonly used to evaluate particle motion because of the availability of closed-form analytical solutions to the dependence of MSD on time lag, τ, for a number of motion models (6,30,31). It is important to note that MSD curves from individual particles undergoing the same stochastic motion are typically highly variable due to limited sampling and strong correlations over τ, which can result in fitting erroneous, overly complex models (30–33). To avoid this over-fitting problem, here we account for these correlations by measuring the MSD and its noise covariance matrix using multiple independent MSDs, and we infer model probabilities using an empirical Bayesian approach similar to that

Submitted April 26, 2012, and accepted for publication June 19, 2012.
*Correspondence: mark.bathe@mit.edu
Editor: Paul Wiseman.

0006-3495/12/08/0616/11 $2.00
http://dx.doi.org/10.1016/j.bpj.2012.06.029

recently applied to fluorescence correlation spectroscopy data (28,29).

The Bayesian approach computes relative probabilities of an arbitrary set of competing motion models without any requirement on model form or nesting, in contrast to frequentist tests such as the F-test (21). This approach naturally handles experimental sampling limitations and heterogeneity between particles in a given biological data set, automatically identifying the simplest model consistent with the observed data according to the Principle of Parsimony or Occam's Razor, a well established property of Bayesian inference (20–23,28,29). Although Bayesian inference requires a choice of prior probabilities associated with each model and its parameters, this requirement objectifies the scientific process by formalizing and reporting these biases concisely in the mathematical form of a prior distribution (21,22). Given a set of priors, Bayesian inference can then be applied automatically, without user intervention.

THEORY

Mean-square displacement analysis

A single-particle trajectory consists of a sequence of N particle positions $\{\mathbf{r}_i\}_{i=1}^{N} = \{x_i, y_i, z_i\}_{i=1}^{N}$ observed at specific times $\{t_i\}_{i=1}^{N}$ separated by time step dt. The mean-square displacement is then computed for time lags τ according to

$$\mathrm{MSD}(\tau) \equiv \left\langle \Delta\mathbf{r}(\tau)^2 \right\rangle = \frac{1}{N-\tau}\sum_{i=1}^{N-\tau} |\mathbf{r}_{i+\tau} - \mathbf{r}_i|^2. \quad (1)$$

For stationary processes, the MSD is given in three dimensions by the following closed-form analytical solutions for free diffusion (D), anomalous diffusion (DA), confined diffusion (DR), and flow or directed motion (V),

$$\mathrm{MSD_D}(\tau) = 6D\tau \quad (2)$$

$$\mathrm{MSD_{DA}}(\tau) = 6D\tau^{\alpha} \quad (3)$$

$$\mathrm{MSD_{DR}}(\tau) = R_C^2\left(1 - e^{-6D\tau/R_C^2}\right) \quad (4)$$

$$\mathrm{MSD_V}(\tau) = \upsilon^2\tau^2, \quad (5)$$

where υ is the magnitude of the particle velocity, D is its diffusion coefficient, α is the anomalous exponent, and R_C is the radius within which the particle is confined (6).

Free diffusion is characteristic of unrestricted stochastic particle motion (33), confined diffusion is characteristic of trapped particles, for example, due to physical corralling by cytoskeletal polymers (6,34,35), and directed motion may result from molecular motor-driven transport (36,37) or cytoskeletal flows (11). Anomalous diffusion arises from a variety of underlying physical processes, including the presence of obstacles or transient binding events, making it difficult to interpret mechanistically (38–40). These diffusive modes often occur together with directed motion, yielding more complex motion models described by linear combinations of the above equations, such as $MSD_{DV}(\tau) = 6D\tau + \upsilon^2\tau^2$ for free diffusion plus flow (DV) (6). Experimental particle-position measurements typically contain a localization error characterized by a positional uncertainty with standard deviation σ_e, which adds a constant term of $6\sigma_e^2$ to the MSD (32) that can be easily incorporated into the proposed Bayesian framework (Fig. S2 in the Supporting Material). In some physical situations, such as confinement within a radius smaller than either the mean localization error or the mean diffusive step size given the sampling rate, the particle may appear stationary because the MSD curve is dominated by this constant term.

Model selection

Classical data regression fits an observed series of data, $\mathbf{y} = \{y_1, y_2, \ldots y_n\}$ (in this case, the MSD values), with a model function $\mathbf{f}(\mathbf{x}; \boldsymbol{\beta})$ (for example, Eqs. 2–5) according to $y_i = f(x_i; \boldsymbol{\beta}) + \varepsilon_i$, where $\mathbf{x} = \{x_1, x_2, \ldots, x_n\}$ are the sample points (in this case, the time lags τ), $\boldsymbol{\beta} = \{\beta_1, \beta_2, \ldots, \beta_p\}$ are the model parameters, and ε_i are the errors associated with the y_i measurements. Classical statistical approaches minimize the sum of the squared residuals, $\chi^2 = \sum_{i=1}^{n}[y_i - f(x_i, \boldsymbol{\beta})]^2/\sigma_i^2$, where the error terms, ε_i, are assumed to be uncorrelated and normally distributed, with mean zero and standard deviations σ_i. The chi-squared value, χ^2, can then be used to test the goodness of fit of models conditioned on a null hypothesis. However, MSD curves contain highly correlated errors that often result in their appearing overly complex (30,32,33) (Fig. 1 *A* and Note S1 in the Supporting Material). For example, MSD curves from purely diffusive trajectories may appear by eye to include directed motion or confinement (Fig. 1 *A*).

Here, we account for correlated errors directly by including an error covariance matrix in Bayesian inference, following previous work on temporal autocorrelation functions from fluorescence correlation spectroscopy data that suffer from similar correlated errors (28,29). For K possible models ($M_1, \ldots, M_K$), the probability of each model given the observed data, $\mathbf{y}$, is given by Bayes' theorem,

$$P(M_k|\mathbf{y}) = \frac{P(\mathbf{y}|M_k)\,P(M_k)}{P(\mathbf{y})} \propto P(\mathbf{y}|M_k), \quad (6)$$

where $P(\mathbf{y}) = \sum_{k=1}^{K} P(\mathbf{y}|M_k)\,P(M_k)$, and the proportionality holds if the prior model probabilities $P(M_k)$ are assumed equal for all k, which is suitable when no information is

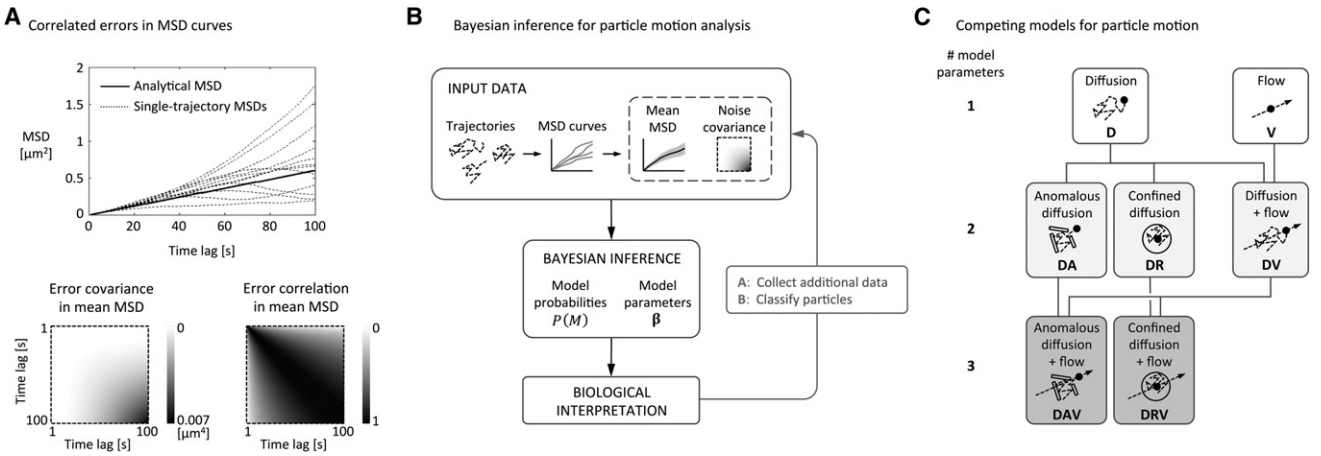

FIGURE 1 Proposed MSD-based Bayesian approach to analyzing particle trajectories. (*A*) *Top*, Example MSD curves (*dashed lines*) from individual simulated particle trajectories undergoing pure diffusion with $D = 0.001\ \mu m^2/s$, $dt = 1$ s, and $T = 200$ s. The analytical form of the MSD curve, $MSD_D(\tau) = 6D\tau$, is also shown (*solid line*). *Bottom*, Analytical form (30,32) of the noise covariance and correlation matrices for pure diffusion MSD curves with the above parameters. (*B*) Sequence of analysis steps to apply Bayesian inference. Starting from a set of particle trajectories, an MSD curve is calculated from each trajectory and then the set of MSD curves is used to calculate a mean MSD curve and its noise covariance matrix, which serve as inputs to the Bayesian procedure described in the main text. The output model probabilities and parameters can be interpreted in the context of the biological system, and, if necessary to improve resolution of complex models, additional trajectories can be collected or existing trajectories can be classified into less heterogeneous sub-groups. (*C*) Models of particle motion used in the Bayesian approach. The simplest (single-parameter) models are shown in the top row, followed by intermediate-complexity (two-parameter) models in the middle row and the most complex (three-parameter) models in the bottom row. Model abbreviations are chosen to specify the parameters in each model; for example, the diffusion-plus-flow model (DV) has both a diffusion coefficient and a velocity magnitude as parameters. Lines connecting the models indicate nesting relationships; for example, both DR and DV are nested in DRV, but DAV and DRV are not nested one in the other.

available to prefer one model over another. $P(\mathbf{y}|M_k)$ is calculated by marginalizing over the model parameters ($\boldsymbol{\beta}_k$),

$$P(\mathbf{y}|M_k) = \int P(\mathbf{y}|\boldsymbol{\beta}_k, M_k)\, P(\boldsymbol{\beta}_k|M_k) d\boldsymbol{\beta}_k, \qquad (7)$$

which inherently penalizes overly complex (overparameterized) models (20). The probability $P(\mathbf{y}|\boldsymbol{\beta}_k, M_k)$ of observing the data $\mathbf{y}$ for any given realization of the parameters $\boldsymbol{\beta}_k$ of model M_k with model function $\mathbf{f}_k(\mathbf{x};\ \boldsymbol{\beta}_k)$ is given by the general multivariate Gaussian function (20) including $\mathbf{C}$, the covariance matrix of the errors ε_i,

$$P(\mathbf{y}|\boldsymbol{\beta}_k, M_k) = \frac{1}{(2\pi)^{n/2}|\mathbf{C}|^{1/2}} \exp\left\{-\frac{1}{2}[\mathbf{y} - \mathbf{f}_k(\mathbf{x}, \boldsymbol{\beta}_k)]^T \times \mathbf{C}^{-1}[\mathbf{y} - \mathbf{f}_k(\mathbf{x}, \boldsymbol{\beta}_k)]\right\}. \qquad (8)$$

Estimation of $\mathbf{C}$ is essential for proper calculation of the model probabilities (29). We use multiple observations of the data $\mathbf{y}$ (independent MSD curves from multiple particle trajectories or subtrajectories) to estimate $\mathbf{C}$, as described in Note S1 in the Supporting Material. (In the case of anomalous diffusion, displacements along an SPT may be correlated, and thus, splitting a single trajectory into multiple independent subtrajectories may require estimation of decorrelation time using block transformation (29,41) or a related approach.) Regularization of the estimated covariance matrix is often required, because the number of available independent MSDs is typically less than the dimension of the matrix (42,43). A shrinkage approach to regularization (42) performs well when low numbers of MSD curves are available (Note S1.3 in the Supporting Material and Fig. S1).

Although numerical integration is required to evaluate the integral in Eq. 7 in general, the Laplace approximation may be used to perform this integration analytically by assuming that $P(\mathbf{y}|\boldsymbol{\beta}_k,\ M_k)$ is well approximated by a multivariate Gaussian distribution around the Bayesian point estimate $\widehat{\boldsymbol{\beta}}_{k,Bayes} = \arg\max_{\boldsymbol{\beta}_k}[P(\mathbf{y}|\boldsymbol{\beta}_k, M_k)\, P(\boldsymbol{\beta}_k|M_k)]$ (20,28). The Laplace approximation is asymptotically exact in the limit of high amounts of data, which is not true of derived metrics such as the Aikake Information Criterion (21). The Bayesian Information Criterion is an alternative, commonly used special case of the Laplace approximation (44) but does not sufficiently penalize model complexity in the case of small sample sizes (28). Here, we use a uniform prior parameter distribution, $P(\boldsymbol{\beta}_k|M_k)$, for which $\widehat{\boldsymbol{\beta}}_{k,Bayes}$ is equal to the maximum-likelihood point estimate $\widehat{\boldsymbol{\beta}}_{k,MLE} = \arg\max_{\boldsymbol{\beta}_k}[P(\mathbf{y}|\boldsymbol{\beta}_k, M_k)]$ (28). For additional details of the implementation, please see the Methods section in the Supporting Material.

The application of Bayesian inference to particle trajectories is summarized in Fig. 1 *B*. The set of competing models evaluated by this method (Fig. 1 *C*) can vary in complexity and include nesting relationships that cannot be treated using standard frequentist tests.

RESULTS AND DISCUSSION

Performance of the Bayesian approach on simulated trajectories

To evaluate the performance of the Bayesian procedure in a controlled setting, we applied it to simulated trajectories of particles undergoing Brownian motion with flow (Fig. 2) or within a confined spherical corral (Fig. 3 *A*). Although we use default simulation parameters comparable to the experimental conditions observed below for starfish chromosomes, namely $D = 0.005\ \mu m^2/s$, $dt = 2.5$ s, $T = 300$ s, $\tau_{max} = T/4$, and $n = 30$ trajectories per data set, we emphasize that the illustrated properties of the proposed multiple-hypothesis-testing procedure are general.

An important source of noise in MSD values is statistical sampling noise due to experimental limitations on the number, length, and sampling rate of available SPTs (6,33). Trajectories were simulated with the above default parameters (Fig. 2 *A*, *middle*), with lower noise (higher T and n; Fig. 2 *A*, *left*) or higher noise (lower T and n; Fig. 2 *A*, *right*), while systematically varying the value of a superimposed velocity, v. The relative contributions of diffusive

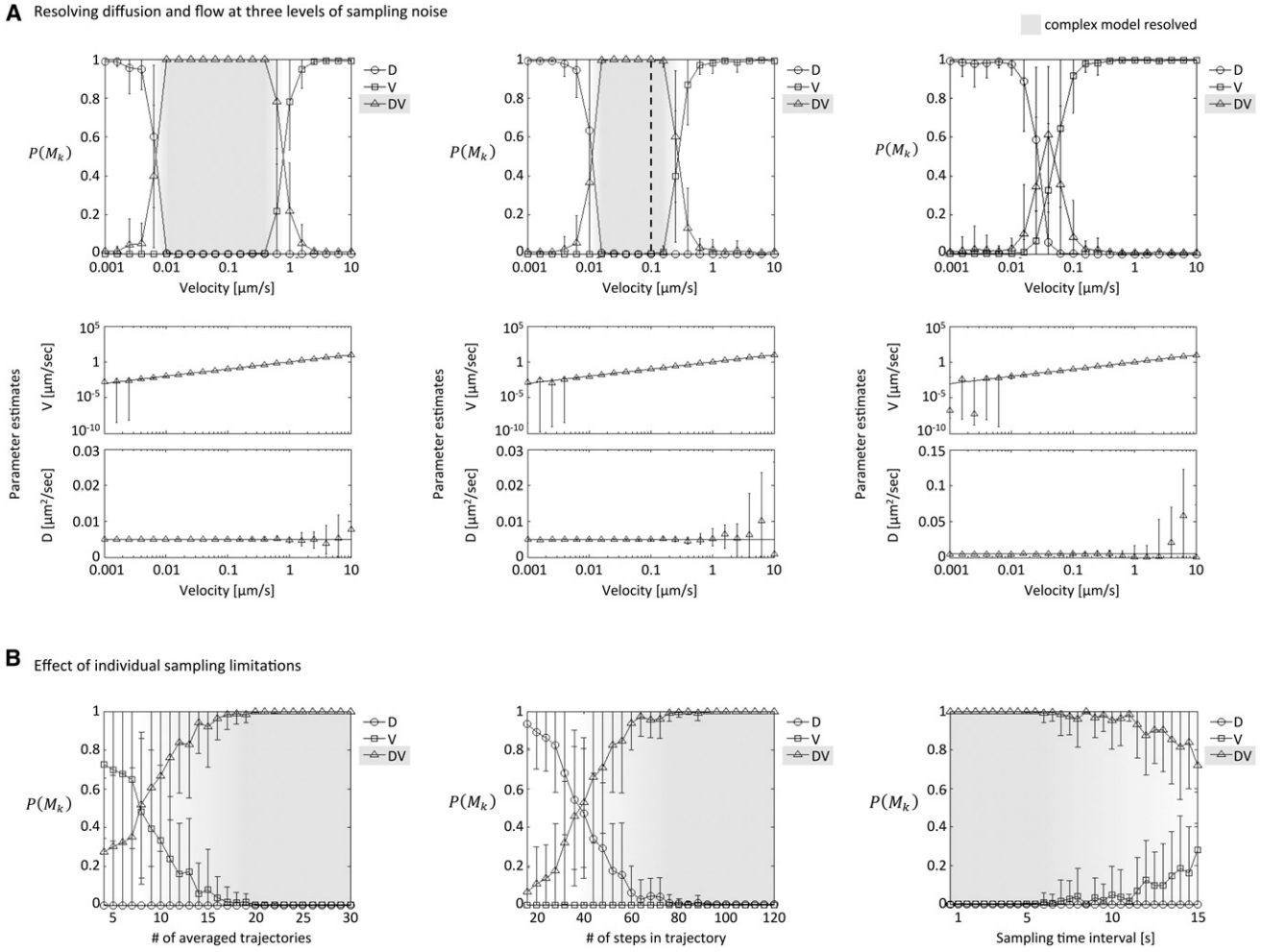

FIGURE 2 Diffusion plus flow simulations and effect of sampling limitations. (*A*) Model probabilities for simulated trajectories undergoing diffusion plus flow with $D = 0.005\ \mu m^2/s$, $dt = 2.5$ s, and varying v as shown along the x axis. T varies from 600 s (*left*) to 300 s (*middle*) to 150 s (*right*), and the number of trajectories per data set varies from $n = 60$ (*left*) to 30 (*middle*) to 5 (*right*). MSD curves with 30 points (up to $\tau_{max} = 75$ s) are calculated for all n trajectories and used to calculate a mean MSD curve and error covariance matrix as input to the Bayesian approach. The resulting model probabilities are shown as means and standard deviations over 50 repetitions of the simulations and inference procedure. Light gray shading indicates the range of velocities over which the true model (DV) can be resolved given the simulated experimental parameters. Estimated values of v and D obtained from fitting the DV model are plotted as medians and quartiles below the model probabilities, in comparison with the true values of v and D used in the simulation (*lines*). The thick dashed line in the top middle panel indicates the starting parameters used for the trajectory number and sampling rate limitation tests in *B*. (*B*) Model probabilities for trajectories simulated as in *A* (*middle*), but at a fixed velocity and systematically varying one of the sampling parameters. *Left*, Velocity is fixed at $v = 0.1\ \mu$m/s and the number of trajectories per data set, n, is varied from 30 down to 4. *Middle*, Velocity is fixed at $v = 0.02\ \mu$m/s and T is varied from 300 s down to 40 s (from 120 down to 16 steps/trajectory). The number of points in the MSD curves is held constant at 1/4 of the number of steps in the trajectory. *Right*, Velocity is fixed at $v = 0.1\ \mu$m/s, and dt is varied from 0.5 s up to 15 s with T fixed at 300 s. The number of points in the MSD curves is again held constant at 1/4 of the number of steps in the trajectory.

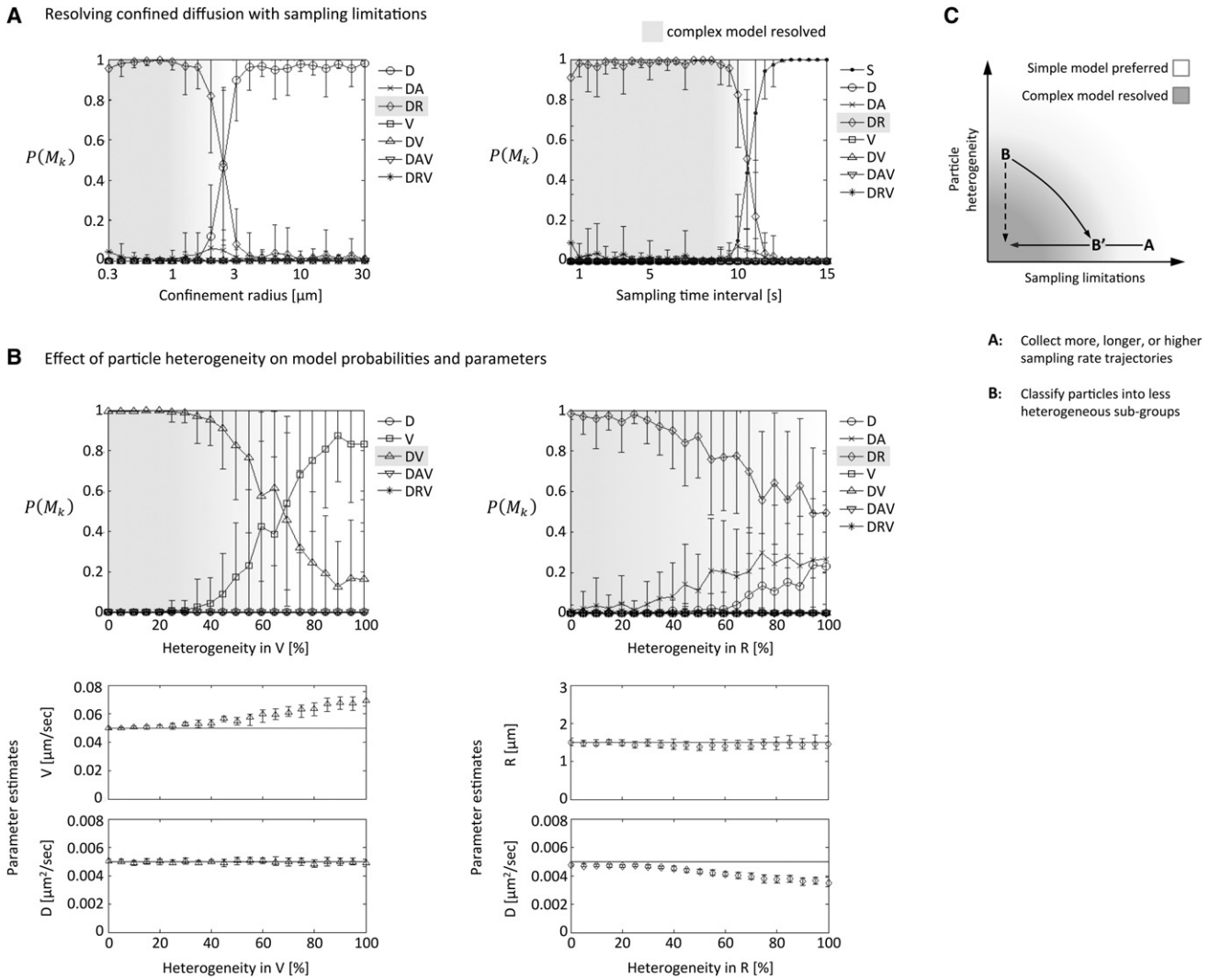

FIGURE 3 Confined diffusion simulations and effect of heterogeneity. (*A*) *Left*, Model probabilities for simulated trajectories undergoing confined diffusion inside a reflecting spherical boundary with $D = 0.005\ \mu m^2/s$, $dt = 2.5$ s, $T = 300$ s, and varying confinement radius R_C as shown along the x axis. Analysis is performed as in Fig. 2 but using the full set of motion models as in Fig. 1 *C*. *Right*, Trajectories are simulated as in *A* but at a fixed confinement radius of $R_C = 0.4\ \mu m$. dt is varied from 0.5 s to 15 s as in Fig. 2 *B*. S represents a stationary-particle model including only a constant term. (*B*) *Left*, Trajectories are simulated as in Fig. 2 *A* (*middle*) but with the velocity of each particle drawn from a normal distribution centered on $v = 0.05\ \mu m/s$, with standard deviation (as a percentage of the mean) as shown on the x axis. Analysis is performed as in Fig. 2. *Right*, Trajectories are simulated as in *A* but with the confinement radius of each particle drawn from a normal distribution centered on $R_C = 1.5\ \mu m$, with standard deviation (as a percentage of the mean) as shown on the x axis. Analysis is performed as in Fig. 2. (*C*) Complex models are most likely to be resolved when there is low heterogeneity between particles and low noise due to data collection limitations, such as the number, length, and sampling rate of the trajectories. The tradeoff between reducing particle heterogeneity and increasing sampling noise by splitting trajectories into smaller groups of fewer trajectories is illustrated by the transition from B to B′.

and directed motion to the diffusion plus flow (DV) MSD equation given above are of similar magnitude when $\tau \sim 6D/v^2 \equiv \tau_{DV}$. The Bayesian approach strongly prefers the DV model for v values corresponding to a timescale τ_{DV} that is comparable to the time lags covered by the MSD curve (Fig. 2 *A*, *middle*). The simpler D and V models are preferred at low and high v, respectively, where the contribution of the v or D parameter to the more complex DV model is not significant given the level of noise in the mean MSD curve. The locations of these crossovers to simpler preferred models at low and high v depend on the level of sampling noise (Fig. 2 *A*, *left* and *right*). Examination of the fit parameter values for the true DV model in Fig. 2 *A* shows that when the DV model probability is high, both parameter values are well estimated, whereas their values become poorly estimated when the model probability is low. Thus, the Bayesian multiple-hypothesis-testing framework not only selects the appropriate model that is justified given the empirical level of noise, it also provides a prescreening filter for downstream physical or biological interpretation of model parameter values, which are only reliable when the model to which they belong is strongly preferred.

We next independently varied three contributing factors to the sampling noise—trajectory number, trajectory length, and sampling rate—at fixed values of v (Fig. 2 *B*). Starting with the default simulation parameters above and a fixed value of $v = 0.1$ μm/s near the righthand crossover point (Fig. 2 *A*, *middle*), decreasing the number of trajectories used to calculate each mean MSD curve and associated error covariance matrix from 30 to 4 results in loss of the ability to resolve the DV model over the simpler V model due to the increasing level of noise (Fig. 2 *B*, *left*). Decreasing T from 300 s to 40 s reduces the ability to resolve the v component of the motion (Fig. 2 *B*, *middle*), whereas increasing dt from 2.5 s to 15 s at a fixed T reduces the ability to resolve the D component of the motion (Fig. 2 *B*, *right*), due to the difference in the relative contributions of diffusion and flow to the MSD curve at high and low τ.

To test whether this Bayesian procedure applies generally to other motion models in addition to diffusion and flow, we repeated the above tests on simulations of confined diffusion (Fig. 3 *A* and Fig. S3). Here, we also included the full set of competing models shown in Fig. 1 *C* to test the robustness of the model selection procedure in the presence of both higher- and lower-complexity competing models. Confinement makes a significant contribution to the confined diffusion (DR) MSD equation (Eq. 4) when the ratio $6D\tau/R_C^2$ is on the order of 1, or when $\tau \sim R_C^2/6D \equiv \tau_{\mathrm{DR}}$. The Bayesian approach strongly prefers the DR model when this ratio τ_{DR} is below the maximum τ in the MSD curve, whereas the simpler D model is preferred for larger confinement radii (Fig. 3 *A*, *left*). As above, the exact crossover point depends on the level of noise in the mean MSD curve. For a fixed value of R_C, decreasing n or T results in loss of the ability to resolve the DR model (Fig. S3). Finally, increasing the trajectory time sampling interval, dt, past the ratio τ_{DR} for a fixed value of R_C results in loss of the ability to resolve the diffusive component of the motion, making the particle appear stationary (Fig. 3 *A*, *right*).

Effect of particle heterogeneity on model selection

The above results demonstrate that this Bayesian procedure can be used to detect both confinement and directed motion in a systematic manner that accounts appropriately for the sampling noise level, avoiding over-fitting of complex models. Heterogeneity in motion type, either within a single trajectory or between distinct particle trajectories, may also reduce the ability to resolve the underlying physical process (Note S2 in the Supporting Material). To explore the effect of heterogeneity, we simulated trajectories as above but allowed a single parameter (v or R_C) to vary randomly between particles according to a normal distribution. As a result, even perfectly measured MSD curves without any statistical sampling error would still vary between the different particles, introducing an apparent noise into the mean MSD curve estimate. As heterogeneity between particles is increased by increasing the standard deviation of the distribution of v or R_C values, the ability to resolve the true motion model diminishes in favor of simpler models due to this increase in apparent noise (Fig. 3 *B*). For normal diffusion plus flow, heterogeneity does not change the dependence of the mean MSD function on τ, but the estimated value of v obtained from the DV model is systematically higher than the true mean due to the quadratic dependence of MSD on v (Fig. 3 *B*, left, and Note S2.1 in the Supporting Material). For confined diffusion, heterogeneity in R_C changes the dependence of the mean MSD function on τ so that none of the models describes the resulting mean MSD curve satisfactorily (Fig. 3 *B*, *right*, and Note S2.3 in the Supporting Material). The apparent diffusion coefficient decreases with increasing heterogeneity in R_C because the diffusion timescale is affected disproportionately by larger confinement radii.

Analysis of actin-dependent chromosome transport in starfish oocytes

To test the performance of the proposed model-selection procedure on experimental biological data sets, we first applied it to the motion of chromosomes during meiosis I in starfish oocytes. Chromosomes are transported toward the spindle at the animal pole (AP) of the oocyte (Fig. 4 *A*) by homogeneous contraction of a large actin network that forms in the nuclear region after nuclear envelope breakdown (NEBD) (11,45). Chromosomes from four oocytes were imaged and tracked at 2.6-s time resolution, a more than fivefold improvement in resolution over previous studies (11), during the 6-min actin-dependent transport phase (Fig. 4 *A*). We analyzed the mean MSD curve over all 30 chromosome trajectories (Fig. 4 *B*) using the Bayesian inference approach to test the full set of motion models shown in Fig. 1 *C*. The DV model is strongly preferred over the other models, consistent with the previously proposed hypothesis that chromosomes diffuse within the actin network as they are transported in a directed manner toward the spindle (11). This result indicates that the chromosome trajectories provide significant evidence for both the diffusive and directed components of their motion but do not provide significant evidence for additional complexity such as confinement or anomalous diffusion, which could potentially result from steric interactions with the actin network structure (11) or the viscoelastic nature of the actin network (46). These more complex motions are not necessarily ruled out by the above result, however, because the additional complexity of confined or anomalous diffusive models might be masked by sampling noise or heterogeneity, as shown in the simulations above.

Since the chromosomes were previously shown to have significant heterogeneity in their velocities, which are

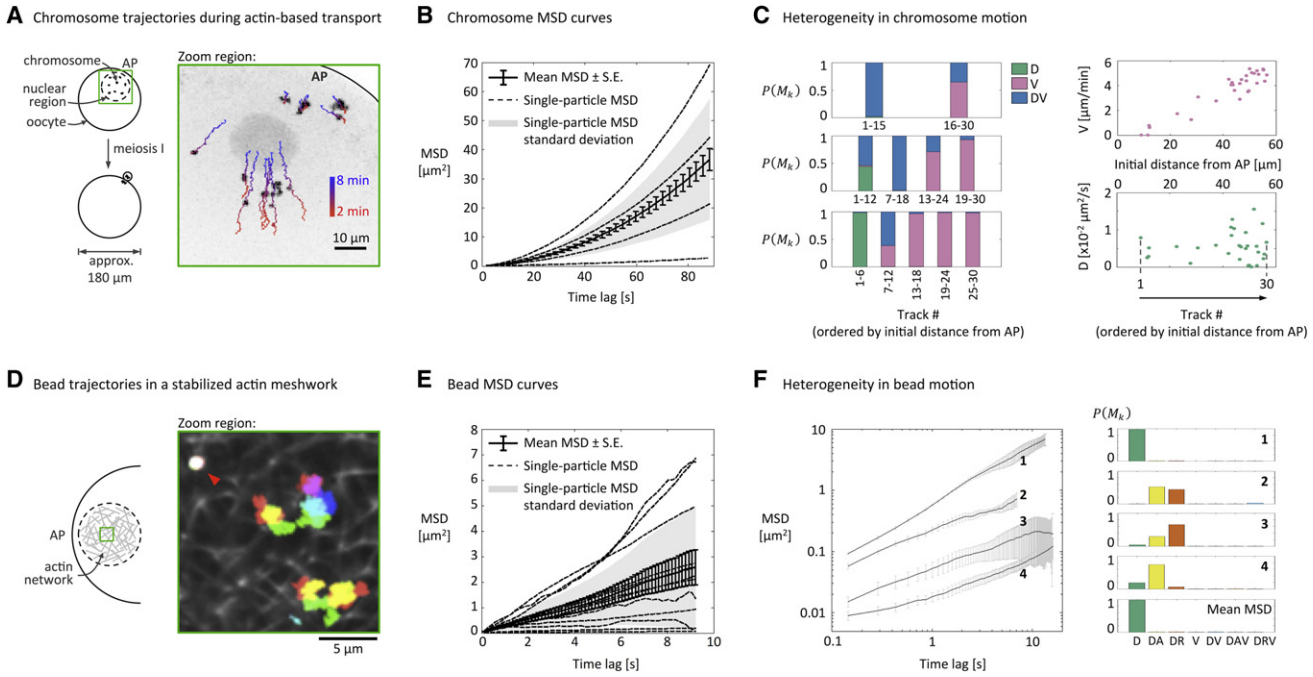

FIGURE 4 Analysis of chromosome and bead trajectories in dynamic and stabilized starfish oocyte actin networks. (*A*) *Left*, Cartoon of chromosome positions in the starfish oocyte at the start (*upper*) and end (*lower*) of meiosis I. AP indicates the animal pole of the oocyte toward which the chromosomes are congressing. *Right*, Maximum-intensity Z-projection through a starfish oocyte nuclear region showing chromosomes labeled with H2B-GFP at 4 min after NEBD. Chromosome trajectories over the full actin transport phase are superimposed, colored from 2 min after NEBD (*red*) to 8 min after NEBD (*blue*). (*B*) Mean MSD curve with standard errors (*solid line*) averaged over 30 chromosome trajectories from a total of four oocytes imaged at 2.6 s time resolution for the 6-min period from 2–8 min after NEBD. Four sample MSD curves from individual chromosome trajectories are shown (*dashed lines*), as is the standard deviation over all 30 of the individual-chromosome MSD curves (*gray region*). The preferred model by Bayesian inference is diffusion plus flow (DV) for the mean MSD curve. (*C*) *Left*, Model probabilities obtained by fitting mean MSD curves over subgroups of 15, 12, and 6 chromosomes (*top to bottom*), plotted from left to right in order of increasing initial distance from the AP. Only the D, V, and DV model probabilities are shown (all other model probabilities were negligible). *Right*, Velocity and diffusion coefficient estimates obtained from the DV model fit to individual-chromosome MSD curves, showing the correlation of velocity with initial distance from the AP. (*D*) *Left*, Cartoon of an actin network in the post-NEBD nuclear region of a starfish oocyte. *Right*, Time projection (*red to blue*) of the motions of 0.2-μm-diameter beads in a utrophin-GFP-stabilized actin network. Some beads appear transiently immobilized (*red arrowhead*). (*E*) Mean and individual MSD curves as in *A* from 12 bead trajectories in a utrophin-stabilized actin network. The preferred model by Bayesian inference is pure diffusion (D) for the mean MSD curve. (*F*) *Left*, Four sample MSD curves from individual beads in the actin network, shown on a log-log scale. *Right*, Model probabilities for the seven tested models fit to each of the four individual-bead MSD curves shown on the left, as well as to the mean MSD curve in *E*.

correlated with initial distance from the AP (11), we split the chromosome trajectories into equally sized groups to reduce this heterogeneity and reanalyzed their motions (Fig. 4 *C*). An initial split into two groups revealed that the DV model is preferred for chromosomes closer to the AP, whereas the simpler V model is preferred farther from the AP (Fig. 4 *C*, *upper left*), confirming that there is heterogeneity along this biological coordinate. Splitting trajectories into less heterogeneous subgroups has a tradeoff (Fig. 3 *C*) in that it reduces the number of trajectories per group, which was shown above to reduce the ability to resolve complex models. The effect of this tradeoff is apparent in the overall trend toward simpler models upon further subclassification of the chromosome trajectories (Fig. 4 *C*, *left*). Although the increase in sampling noise that results from the reduction in number of SPTs per subgroup outweighs the reduction in heterogeneity in this case, additional oocytes could in principle be added to the total pool of data to again resolve the more complex DV model. Finally, the increasing probability of the simpler V model for chromosomes far from the AP and the simpler D model for chromosomes close to the AP is comparable to moving to the right and left, respectively, along the horizontal axes in Fig. 2 *A* because of the difference in velocities between these chromosomes (Fig. 4 *C*, *right*).

Bead dynamics probe confinement by the actin network

We next sought an alternate means of probing the starfish actin network that is not complicated by the network's directed motion. We examined the diffusion of 0.2-μm beads within the network by injecting them into the oocyte nucleus just before NEBD while simultaneously overexpressing mEGFP-UtrCH to stabilize actin bundles to prevent network contraction (Fig. 4 *D*). Bead trajectories

have previously been used to characterize the density of obstacles, sizes of pores, and viscoelastic properties of cytoskeletal networks (46,47). We found that beads in the stabilized actin network exhibit a range of behaviors (Fig. 4 *D*) and that the mean MSD curve over multiple bead trajectories (Fig. 4 *E*) is best explained by the simple diffusion model, presumably due to this high heterogeneity. However, when individual bead trajectories are analyzed by splitting them into consecutive subtrajectories (assumed to be independent) to estimate the mean MSD and noise covariance matrix for each bead, then a variety of diffusive models are resolved, including the higher-complexity anomalous- and confined-diffusion models (Fig. 4 *F*). This data set therefore provides an example in which heterogeneity between particles is high enough that moving from a mean MSD curve over all particles to individual-particle MSD curves improves the ability to resolve complex models despite the associated increase in sampling noise. A more detailed analysis of these heterogeneous bead dynamics will be the subject of future work.

Dynamics of the membrane receptor CD36 in macrophages

As another example of detecting confinement in a very different biological system, we analyzed previously published trajectories of the membrane receptor CD36 (Fig. 5 *A*), which exhibits a range of behaviors including linear motion, confined diffusion, and unconfined diffusion (1,8). Testing the full set of motion models from Fig. 1 *C* with the Bayesian procedure reveals that the mean MSD curve over all CD36 trajectories (Fig. 5 *B*) is best fit by the anomalous-diffusion model, but that individual CD36 trajectories are best explained by either pure diffusion or by the stationary-particle model described above (Fig. 5 *C*). The high probability of the stationary model suggests that these receptors are confined within a radius smaller than the mean diffusive step size of the trajectory, as in Fig. 3 *A*, or are attached to a stationary structure such as a cytoskeletal matrix (48), consistent with the confined diffusion classification in previous analysis of the trajectories (8). Pure diffusion is the preferred model for nearly all trajectories

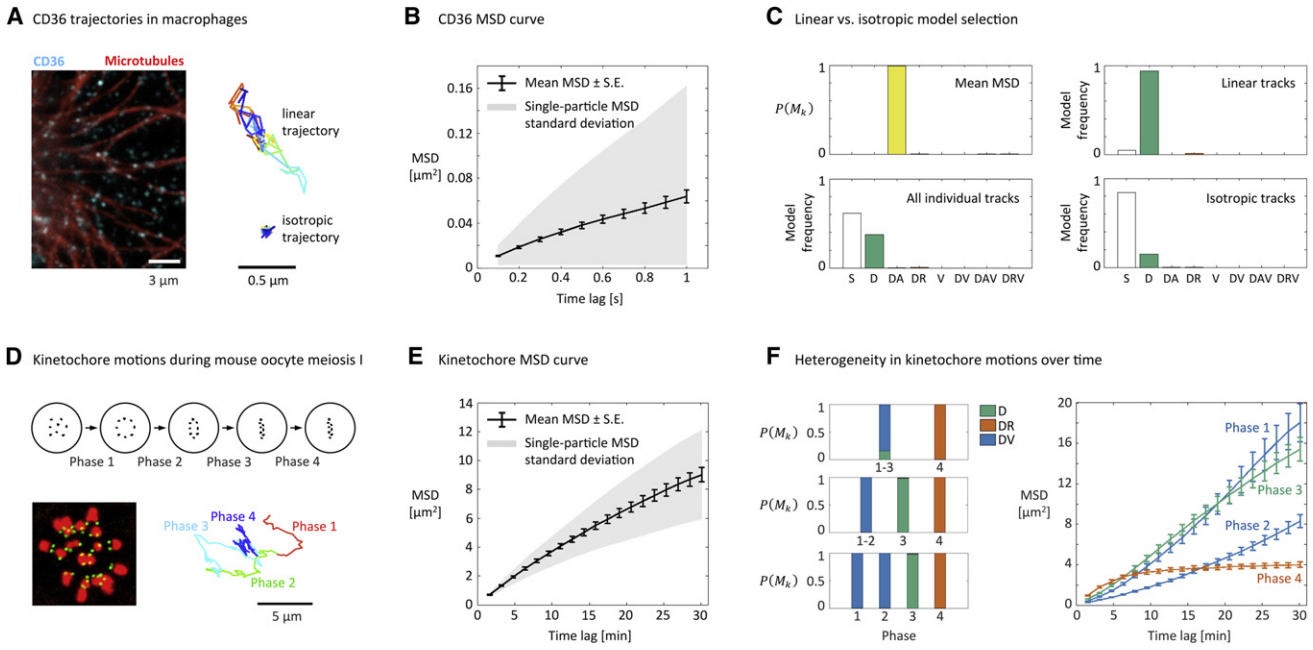

FIGURE 5 Analysis of CD36 trajectories and mouse oocyte kinetochore trajectories. (*A*) *Left*, Image of the membrane receptor CD36 (*blue*) and microtubules (*red*) in a macrophage. Image reprinted from Jaqaman et al. (8) with permission from Elsevier. *Right*, Example trajectories classified as linear (*upper*) and isotropic (*lower*) by the asymmetry metric used in Jaqaman et al. (8), colored over time from red to blue. (*B*) Mean MSD curve with standard errors and standard deviation over all of the CD36 trajectories that are at least 40 time steps in length (296 trajectories total). The preferred model by Bayesian inference is anomalous diffusion (DA). (*C*) Model probabilities for the mean MSD curve (*upper left*). Frequency with which each of eight tested models is selected as the most probable model for all CD36 trajectories (*lower left*), for the 84 linear CD36 trajectories (*upper right*), and for the 212 isotropic CD36 trajectories (*lower right*). As in Fig. 3 *A*, S represents a stationary-particle model including only a constant term. (*D*) *Upper*, Cartoon of kinetochore motions during the different time phases defined in Kitajima et al. (12) leading up to the first meiotic division in mouse oocytes. *Lower left*, Mouse kinetochores (*green*) and chromosomes (*red*) in a maximum-intensity Z-projection through the spindle at the beginning of phase 2. Image reprinted from Kitajima et al. (12) with permission from Elsevier. *Lower right*, Sample kinetochore trajectory showing the four phases of motion. (*E*) Mean MSD curve as in *B* over all 40 kinetochore trajectories from a single oocyte during the full 8.7-hr period of meiosis. The preferred model by Bayesian inference is confined diffusion (DR). (*F*) *Left*, Model probabilities obtained by fitting mean MSD curves over all 40 kinetochore trajectories split into the time phases shown in *D*. Only the D, DR, and DV model probabilities are shown (all other model probabilities were negligible). *Right*, Mean MSD curves over all 40 kinetochore trajectories for the individual time phases, colored by the preferred model.

previously classified as linear (Fig. 5 *C*), confirming that these motions are linear due to 1D diffusion (for example, along 1D tracks or within linear-shaped confinement zones), whereas the stationary model is preferred for most of the previously classified isotropic trajectories (Fig. 5 *C*). Only a small fraction of receptors exhibit isotropic unconfined diffusion.

Kinetochore trajectories during mouse meiosis I exhibit heterogeneity in time

The above examples illustrate that a single automated Bayesian approach can be used to detect either directed motion or confinement or anomalous diffusion in a variety of biological systems. We next sought to detect both types of motion within a single biological data set. Kinetochores in mouse oocytes (Fig. 5 *D*) were recently found to exhibit distinct complex motions during discrete time phases during meiosis (12). Analyzing the mean MSD curve over the entire period of meiosis (Fig. 5 *E*) with the Bayesian procedure reveals that the highest-probability model for the mean behavior of the kinetochores is confined diffusion. However, sequentially dividing the kinetochore trajectories into time periods corresponding to the previously described phases (12) reveals that confinement is localized to phase 4, whereas diffusion plus flow is preferred for phases 1 and 2, and pure diffusion is preferred for phase 3 (Fig. 5 *F*). In future studies or screens, this Bayesian procedure may be used in an automated manner to discover the above phase boundaries a priori by systematic evaluation of boundary locations and number of phases.

CONCLUSION

Like all experimental measurements, SPTs require the use of mathematical models for their physical interpretation. To enable analysis of many dozens or even hundreds or thousands of trajectories, often under both wild-type and perturbed biological conditions, a fully automated approach to systematic evaluation of competing motion models that does not require manual intervention or data curation is highly desirable. Bayesian inference, applied here to MSD-based analysis of SPTs, is a general theoretical framework that is useful for this purpose. The Bayesian approach handles multiple competing models for single-particle motion simultaneously, preferring simpler models when statistical noise and heterogeneity preclude the resolution of more complex models that are not justified by the data. Although fully objective in the computation of model probabilities, Bayesian inference still involves a subjective choice about what probability is considered convincing evidence for a given model or hypothesis to be accepted, a topic that is discussed at length by Raftery and colleagues (21,49). Although the emphasis here is on performing systematic multiple-hypothesis testing for particle motion, we also illustrate that computed model probabilities act as a reliability test for the downstream physical or biological interpretation of model parameter values.

Statistical noise due to sampling limitations and heterogeneity between particles limits the ability to resolve complex motion models. Sampling noise may be reduced by collecting more data, namely longer or more trajectories, to improve the statistical accuracy of estimates of the mean MSD and its correlated error (Fig. 3 *C*, *Case A*). However, heterogeneity within a trajectory or across multiple trajectories may only be reduced by appropriately segmenting trajectories into smaller subsets along a relevant biological axis (Fig. 3 *C*, *Case B*). Segmentation typically comes at the cost of increasing sampling noise, because the number of particle trajectories within each subgroup is reduced unless additional particle trajectories from the same system are acquired. Resolving meaningful, reproducible heterogeneity in biological systems is of central interest to understanding biological behavior, and therefore, automated classification schemes for this purpose will be the subject of future work. Accounting for heterogeneity directly in the inference process by the use of stochastic models instead of ensemble-averaged quantities such as the MSD is also of interest (34,35,50,51). Nevertheless, the approach presented here already enables the systematic and automated analysis of information-rich particle-trajectory data sets and can be applied to high-throughput screens involving cells, embryos, and whole animals by incorporation into automated screening platforms, such as Cell Profiler (52) and CellCognition (53), or in-house analysis programs via download from http://msd-bayes.org.

SUPPORTING MATERIAL

Supplementary methods, notes, three figures, and references (54–56) are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(12)00718-7.

We are grateful to Khuloud Jaqaman and Gaudenz Danuser (Harvard Medical School) and Tomoya Kitajima and Jan Ellenberg (European Molecular Biology Laboratory) for sharing particle-trajectory data sets and providing advice and critical reading of the manuscript. We also thank Korbinian Strimmer (University of Leipzig) for advice on covariance matrix regularization techniques.

This work was funded by Massachusetts Institute of Technology Faculty Start-up Funds and the Samuel A. Goldblith Career Development Professorship awarded to M.B.

REFERENCES

1. Jaqaman, K., D. Loerke, …, G. Danuser. 2008. Robust single-particle tracking in live-cell time-lapse sequences. *Nat. Methods.* 5:695–702.
2. Sergé, A., N. Bertaux, …, D. Marguet. 2008. Dynamic multiple-target tracing to probe spatiotemporal cartography of cell membranes. *Nat. Methods.* 5:687–694.
3. Betzig, E., G. H. Patterson, …, H. F. Hess. 2006. Imaging intracellular fluorescent proteins at nanometer resolution. *Science.* 313:1642–1645.

4. Rust, M. J., M. Bates, and X. W. Zhuang. 2006. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat. Methods.* 3:793–795.
5. Yildiz, A., J. N. Forkey, …, P. R. Selvin. 2003. Myosin V walks hand-over-hand: single fluorophore imaging with 1.5-nm localization. *Science.* 300:2061–2065.
6. Saxton, M. J., and K. Jacobson. 1997. Single-particle tracking: applications to membrane dynamics. *Annu. Rev. Biophys. Biomol. Struct.* 26:373–399.
7. Brandenburg, B., and X. Zhuang. 2007. Virus trafficking: learning from single-virus tracking. *Nat. Rev. Microbiol.* 5:197–208.
8. Jaqaman, K., H. Kuwata, …, S. Grinstein. 2011. Cytoskeletal control of CD36 diffusion promotes its receptor and signaling function. *Cell.* 146:593–606.
9. Chuang, C. H., A. E. Carpenter, …, A. S. Belmont. 2006. Long-range directional movement of an interphase chromosome site. *Curr. Biol.* 16:825–831.
10. Cremer, T., and C. Cremer. 2001. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat. Rev. Genet.* 2:292–301.
11. Mori, M., N. Monnier, …, P. Lénárt. 2011. Intracellular transport by an anchored homogeneously contracting F-actin meshwork. *Curr. Biol.* 21:606–611.
12. Kitajima, T. S., M. Ohsugi, and J. Ellenberg. 2011. Complete kinetochore tracking reveals error-prone homologous chromosome biorientation in mammalian oocytes. *Cell.* 146:568–581.
13. Gardner, M. K., C. G. Pearson, …, D. J. Odde. 2005. Tension-dependent regulation of microtubule dynamics at kinetochores can explain metaphase congression in yeast. *Mol. Biol. Cell.* 16:3764–3775.
14. Ehrlich, M., W. Boll, …, T. Kirchhausen. 2004. Endocytosis by random initiation and stabilization of clathrin-coated pits. *Cell.* 118:591–605.
15. Turner, L., W. S. Ryu, and H. C. Berg. 2000. Real-time imaging of fluorescent flagellar filaments. *J. Bacteriol.* 182:2793–2801.
16. Jaqaman, K., and G. Danuser. 2006. Linking data to models: data regression. *Nat. Rev. Mol. Cell Biol.* 7:813–819.
17. Simson, R., E. D. Sheets, and K. Jacobson. 1995. Detection of temporary lateral confinement of membrane proteins using single-particle tracking analysis. *Biophys. J.* 69:989–993.
18. Rajani, V., G. Carrero, …, C. W. Cairo. 2011. Analysis of molecular diffusion by first-passage time variance identifies the size of confinement zones. *Biophys. J.* 100:1463–1472.
19. Tejedor, V., O. Bénichou, …, R. Metzler. 2010. Quantitative analysis of single particle trajectories: mean maximal excursion method. *Biophys. J.* 98:1364–1372.
20. Sivia, D. S., and J. Skilling. 2006. Data Analysis: A Bayesian Tutorial, 2nd ed. Oxford University Press, Oxford, UK.
21. Raftery, A. E. 1995. Bayesian model selection in social research. *Sociol. Methodol.* 25:111–163.
22. Posada, D., and T. R. Buckley. 2004. Model selection and model averaging in phylogenetics: advantages of Akaike Information Criterion and Bayesian approaches over likelihood ratio tests. *Syst. Biol.* 53:793–808.
23. Carlin, B. P., and T. A. Louis. 2009. Bayesian Methods for Data Analysis, 3rd ed. CRC Press, Boca Raton, FL.
24. Beaumont, M. A., and B. Rannala. 2004. The Bayesian revolution in genetics. *Nat. Rev. Genet.* 5:251–261.
25. Sachs, K., O. Perez, …, G. P. Nolan. 2005. Causal protein-signaling networks derived from multiparameter single-cell data. *Science.* 308:523–529.
26. Friedman, N. 2004. Inferring cellular networks using probabilistic graphical models. *Science.* 303:799–805.
27. Bronson, J. E., J. Fei, …, C. H. Wiggins. 2009. Learning rates and states from biophysical time series: a Bayesian approach to model selection and single-molecule FRET data. *Biophys. J.* 97: 3196–3205.
28. He, J., S. M. Guo, and M. Bathe. 2012. Bayesian approach to the analysis of fluorescence correlation spectroscopy data I: theory. *Anal. Chem.* 84:3871–3879.
29. Guo, S. M., J. He, …, M. Bathe. 2012. Bayesian approach to the analysis of fluorescence correlation spectroscopy data II: application to simulated and in vitro data. *Anal. Chem.* 84:3880–3888.
30. Qian, H., M. P. Sheetz, and E. L. Elson. 1991. Single particle tracking. Analysis of diffusion and flow in two-dimensional systems. *Biophys. J.* 60:910–921.
31. Kusumi, A., Y. Sako, and M. Yamamoto. 1993. Confined lateral diffusion of membrane receptors as studied by single particle tracking (nanovid microscopy). Effects of calcium-induced differentiation in cultured epithelial cells. *Biophys. J.* 65:2021–2040.
32. Michalet, X. 2010. Mean square displacement analysis of single-particle trajectories with localization error: Brownian motion in an isotropic medium. *Phys. Rev. E.* 82:041914.
33. Saxton, M. J. 1997. Single-particle tracking: the distribution of diffusion coefficients. *Biophys. J.* 72:1744–1753.
34. Das, R., C. W. Cairo, and D. Coombs. 2009. A hidden Markov model for single particle tracks quantifies dynamic interactions between LFA-1 and the actin cytoskeleton. *PLoS Comput. Biol.* 5:e1000556.
35. Cairo, C. W., R. Das, …, D. E. Golan. 2010. Dynamic regulation of CD45 lateral mobility by the spectrin-ankyrin cytoskeleton of T cells. *J. Biol. Chem.* 285:11392–11401.
36. Bormuth, V., V. Varga, …, E. Schäffer. 2009. Protein friction limits diffusive and directed movements of kinesin motors on microtubules. *Science.* 325:870–873.
37. Elting, M. W., Z. Bryant, …, J. A. Spudich. 2011. Detailed tuning of structure and intramolecular communication are dispensable for processive motion of myosin VI. *Biophys. J.* 100:430–439.
38. Petrov, E., and P. Schwille. 2008. State of the art and novel trends in fluorescence correlation spectroscopy. *In* Standardization and Quality Assurance in Fluorescence Measurements II: Bioanalytical and Biomedical Applications. U. Resch-Genger, editor. Springer, Berlin. 145–197.
39. Weber, S. C., A. J. Spakowitz, and J. A. Theriot. 2010. Bacterial chromosomal loci move subdiffusively through a viscoelastic cytoplasm. *Phys. Rev. Lett.* 104:238102.
40. Wang, X., T. Wohland, and V. Korzh. 2010. Developing in vivo biophysics by fishing for single molecules. *Dev. Biol.* 347:1–8.
41. Flyvbjerg, H., and H. G. Petersen. 1989. Error estimates on averages of correlated data. *J. Chem. Phys.* 91:461–466.
42. Schäfer, J., and K. Strimmer. 2005. A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Stat. Appl. Genet. Mol. Biol.* 4:e32.
43. Ledoit, O., and M. Wolf. 2004. A well-conditioned estimator for large-dimensional covariance matrices. *J. Multivar. Anal.* 88:365–411.
44. Kass, R. E., and L. Wasserman. 1995. A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion. *J. Am. Stat. Assoc.* 90:928–934.
45. Lénárt, P., C. P. Bacher, …, J. Ellenberg. 2005. A contractile nuclear actin network drives chromosome congression in oocytes. *Nature.* 436:812–818.
46. Wong, I. Y., M. L. Gardel, …, D. A. Weitz. 2004. Anomalous diffusion probes microstructure dynamics of entangled F-actin networks. *Phys. Rev. Lett.* 92:178101.
47. Caspi, A., R. Granek, and M. Elbaum. 2000. Enhanced diffusion in active intracellular transport. *Phys. Rev. Lett.* 85:5655–5658.
48. Shin, J. H., M. L. Gardel, …, D. A. Weitz. 2004. Relating microstructure to rheology of a bundled and cross-linked F-actin network in vitro. *Proc. Natl. Acad. Sci. USA.* 101:9636–9641.
49. Hoeting, J. A., D. Madigan, …, C. T. Volinsky. 1999. Bayesian model averaging: a tutorial. *Stat. Sci.* 14:382–401.

50. Voisinne, G., A. Alexandrou, and J. B. Masson. 2010. Quantifying biomolecule diffusivity using an optimal Bayesian method. *Biophys. J.* 98:596–605.

51. Dey, A. 2011. Hidden Markov Model Analysis of Subcellular Particle Trajectories. Massachusetts Institute of Technology, Cambridge, MA.

52. Carpenter, A. E., T. R. Jones, …, D. M. Sabatini. 2006. CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol.* 7:R100.

53. Held, M., M. H. A. Schmitz, …, D. W. Gerlich. 2010. CellCognition: time-resolved phenotype annotation in high-throughput live cell imaging. *Nat. Methods.* 7:747–754.

54. Burkel, B. M., G. von Dassow, and W. M. Bement. 2007. Versatile fluorescent probes for actin filaments based on the actin-binding domain of utrophin. *Cell Motil. Cytoskeleton.* 64:822–832.

55. Lénárt, P., G. Rabut, …, J. Ellenberg. 2003. Nuclear envelope breakdown in starfish oocytes proceeds by partial NPC disassembly followed by a rapidly spreading fenestration of nuclear membranes. *J. Cell Biol.* 160:1055–1068.

56. Daniels, B. R., B. C. Masi, and D. Wirtz. 2006. Probing single-cell micromechanics in vivo: the microrheology of C. elegans developing embryos. *Biophys. J.* 90:4712–4719.