Supporting Material:

A Bayesian inference scheme to extract diffusivity and potential fields from confined single-molecule trajectories

Silvan Türkcan, Antigoni Alexandrou, Jean-Baptiste Masson

The supporting material includes 6 figures in 4 sections. Section S1: **Additional experimental CpεT receptor trajectories and inferred potentials.** Section S2: **Parameters for the correction of the inferred values** expands the investigation of the performance of the inference technique in section 4.2 of the manuscript. Section S3: **Robustness to Noise** examines the influence of positioning noise on the inferred variables.Here we examine the inferred values in terms of the noise level of the system through numerical trajectories. Section S4: **The Flat-Well Potential** expands section 4.4 of the manuscript where we show that the inferred forces in a flat-well potential are negligible. Here we explore a greater parameter space by varying the size of the confining box.

S1 Additional experimental CpεT receptor trajectories and inferred potentials

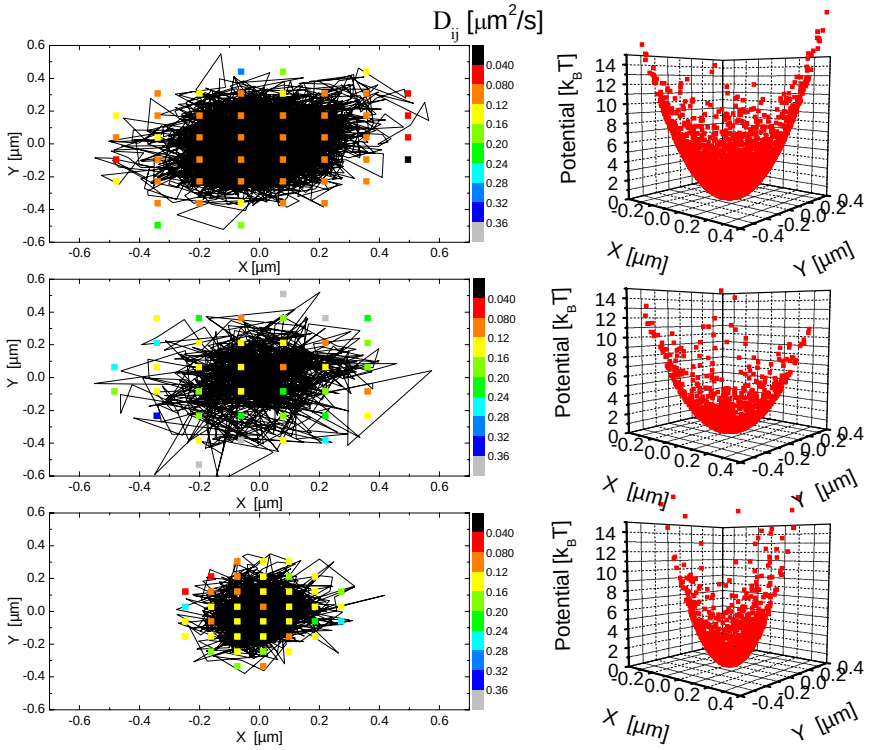

FIGURE S1: Inferred confining potentials (right column) and diffusivity maps (left column) from experimental trajectories (left column). Top row: a CSαT receptor trajectory with D_{Inf} of 0.11 ± 0.04μm^2/s and inferred k_r of 0.28 ± 0.02pN/μm. Middle row: a CPεT receptor trajectory with D_{Inf} of 0.2 ± 0.1μm^2/s and inferred k_r of 0.23 ± 0.04pN/μm.Bottom row: a CPεT receptor trajectory with D_{Inf} of 0.14 ± 0.04μm^2/s and inferred k_r of 0.48 ± 0.08pN/μm.

S2 Parameters for the correction of the inferred values

The performance of the inference technique in extracting the diffusion coefficient and spring constant is tested with numerical trajectories, whose input parameters are varied. A systematic deviation from the input value can be corrected for and we determined the two parameters A and B (Eqs. 11-12) that are required to correct the inferred values.

The deviation from the input values is due to the non-linearities in theoptimization process. Here we test the extent of the parameter space for which A and B can be used. The numerical trajectories have a step size of 51.3 ms, 1000 points (this parameter is varied in Fig. S2), and D_{Input}=0.075 $\mu m^2/s$ (varied in Fig. S3). We vary the confinement factor u by linearly varying the input spring constant k_r. 2500 trajectories were calculated for each set of parameters.

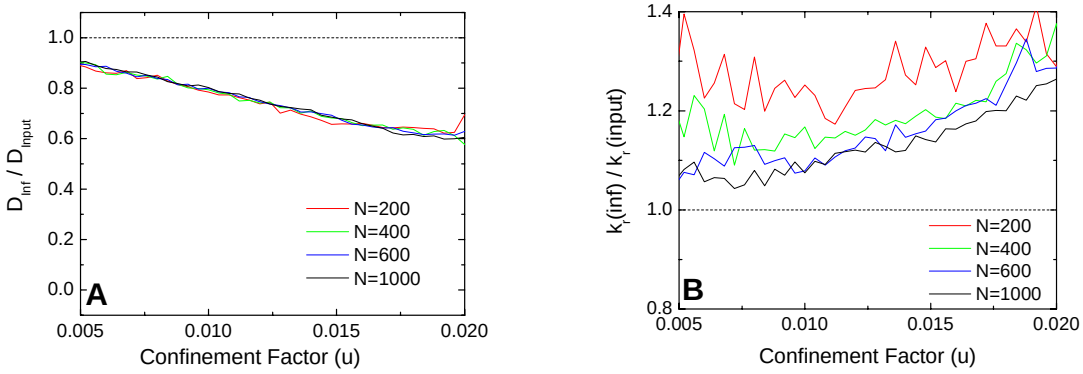

FIGURE S2: Normalized values of inferred parameters for diffusion coefficients (A) and radial spring constants (B) for trajectories of varying length N. The underestimation of the diffusion coefficient remains constant, even for short trajectories of 200 points. The correction factor A for the diffusion coefficient in Eq. 11 remains constant. The correction factor B in Eq. 12 for the spring constant does change for short trajectories, but remains the same, once trajectories are longer than 400 points.

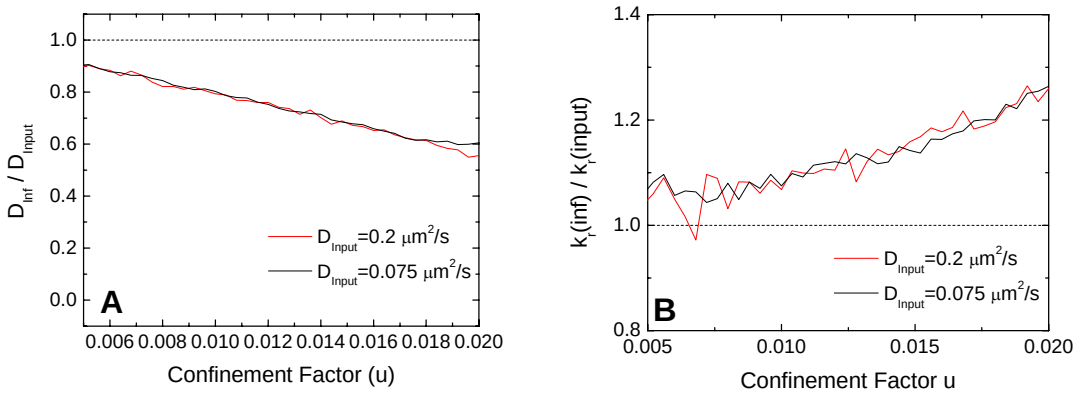

FIGURE S3: Normalized values of inferred parameters for diffusion coefficients (A) and radial spring constants (B) for trajectories of two input diffusion coefficients. The underestimation of the diffusion coefficient remains constant. The correction factor A for the diffusion coefficient in Eq. 11 and the correction factor B for the spring constant in Eq. 12 remain constant.

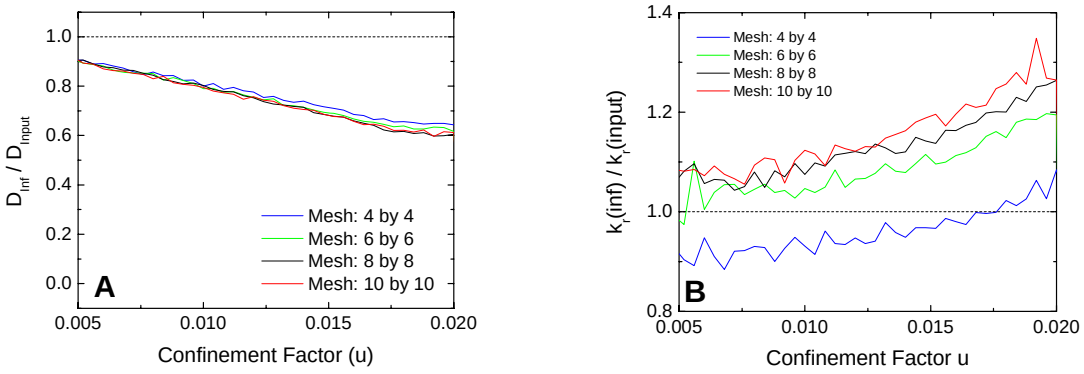

FIGURE S4: Normalized values of inferred parameters for diffusion coefficients (A) and radial spring constants (B) for trajectories analyzed with a different mesh size. The underestimation of the diffusion coefficient remains constant, because we infer a global diffusion coefficient. The correction factor A for the diffusion coefficient in Eq. 11 remains constant. The correction factor B for the spring constant in Eq. 12 also remains constant down to a coarse mesh of 6 by 6. The 6 by 6 mesh shows less bias because more data points fall in each mesh square. This is compatible with the results shown in Fig. S2, where fewer data points induce a greater deviation. For the 4 by 4 case, the factor B in Eq. 12 has to be reevaluated.

S3 Robustness to noise

If there were no noise on the biomolecule positioning, the probability of going from ($\mathbf{r}_1$,t_1) to ($\mathbf{r}_2$,t_2) would read:

$$P\left(\mathbf{r}_1,t_1\middle|\mathbf{r}_2,t_2\right)=\frac{e^{-\frac{\left(\mathbf{r}_2-\mathbf{r}_1-\mathbf{F}_{ij}\Delta t/\gamma_{ij}\right)^2}{4D_{ij}\Delta t}}}{4\pi D_{ij}\Delta t}$$

with (ij) the subdomain where the motion takes place, D_{ij} the diffusivity, F_{ij} the force, γ_{ij} the friction coefficient and $\Delta t = t_2 - t_1$. All sources of noise, i.e. Poissonian photon shot noise due to signal and background, pixelization effects, detector noise, and error of the localization algorithm can be modeled by a Gaussian noise with standard deviation σ. Therefore, the transition probability should read:

$$P\left(\mathbf{r}_1,t_1\middle|\mathbf{r}_2,t_2\right)=\iint d\mathbf{u}_1 d\mathbf{u}_2\,\frac{e^{-\frac{(\mathbf{u}_1-\mathbf{r}_1)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}\frac{e^{-\frac{\left(\mathbf{r}_2-\mathbf{r}_1-\mathbf{F}_{ij}\Delta t/\gamma_{ij}\right)^2}{4D_{ij}\Delta t}}}{4\pi D_{ij}\Delta t}\frac{e^{-\frac{(\mathbf{u}_2-\mathbf{r}_2)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}=\frac{e^{-\frac{\left(\mathbf{r}_2-\mathbf{r}_1-\mathbf{F}_{ij}\Delta t/\gamma_{ij}\right)^2}{4\left(D_{ij}+\frac{\sigma^2}{\Delta t}\right)\Delta t}}}{4\pi\left(D_{ij}+\frac{\sigma^2}{\Delta t}\right)\Delta t}$$

where the transition probability has been convoluted by two Gaussians modeling positioning uncertainties. From the above equation, it is obvious that inference in the presence of noise will lead to an apparent diffusion coefficient $D_{ij}+\sigma^2/\Delta t$. For numerical trajectories, the input positioning noise is known and its effect on the inferred diffusion coefficient can be directly subtracted.

We explore the performance of the method for a range of positioning noise values. To this end, we define the noise level, which is the ratio between the apparent displacement due to the noise, $\sigma = 2B$, and the displacement due to the Brownian motion, L_D. We generate 50 trajectories with 1000 points, a diffusion coefficient of 0.075 $\mu m^2/s$, a radial spring constant of 0.5 pN/μm and an acquisition time of 51.3 ms. At each frame we add a static positioning noise from a Gaussian distribution.

We show that the introduced inference method is robust even for very high noise levels.

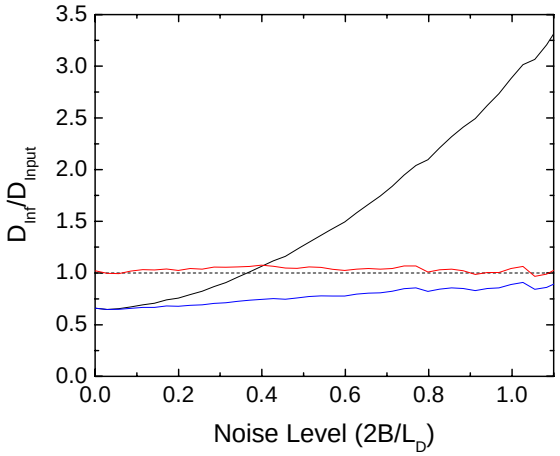

FIGURE S5: Normalized values of inferred diffusion coefficients as a function of noise level. The black (blue) curve shows the diffusion coefficient before (after) correction for the positioning noise. Then using the parameters A and B (Eq. 11-12), the diffusion coefficient can be corrected to give the final inferred value (red). These corrected values show that the method is very robust even when the noise masks the Brownian motion at the noise level of 1.

S4 The flat-well potential

We use this potential in section 4.3 in the main textto test if the algorithm that we use for the experimental data will show that there are no forces (or no significant forces) within this hard wall potential. Because the walker cannot leave the well, we have no information about the wall of the potential. That is why we do not expect to infer the potential with walls using our algorithm for the experimental trajectories. We rather expect to see the completely flat part of the bottom of the well. If this part is not flat, i.e. the algorithm shows significant forces within the domain, we cannot be confident in the inferred forces from the experimental trajectories. In contrast, if the inference algorithm, which is looking for a non-flat potential shows that the potential well is flat, i.e. extracts the correct input potential, we can be confident in the significance of the extracted forces from the experimental trajectories.

We generate numerical trajectories in a flat-well potential, where we vary the length of the confining box L from 100 nm to 1000 nm. The position is captured every 51.3 ms and the input diffusion coefficient is 0.1 $\mu m^2/s$. The trajectory length is 100, 400 or 1000 points with 1000 sub-steps that are not averaged. Before the position is recorded, we add a positioning noise B=60 nm from a Gaussian distribution. We then evaluate the potential assuming a 2nd order potential.In order to be confident that the k_r values extracted from experimental trajectories represent real potentials, the inferred k_r values must lie well above the k_r values that the inference extracts from these flat-well trajectories. Figure S6 represents the k_r values extracted from these numerical trajectories as a function of the length L of the confining box together with the experimental data for CpεT and CsαT receptors. For trajectory lengths of 400 or 1000 points, the k_r valuesinferred from experimental trajectories lie well above the spring constants inferred from numerical trajectories with a flat-well potential. Note that all the experimental trajectories consist of at least 500 points. Furthermore, the use of a higher order potential leads to the extraction of potentials much closer to zero. We here restricted ourselves to a second order potential to directly compare with the potential used on experimental trajectories.

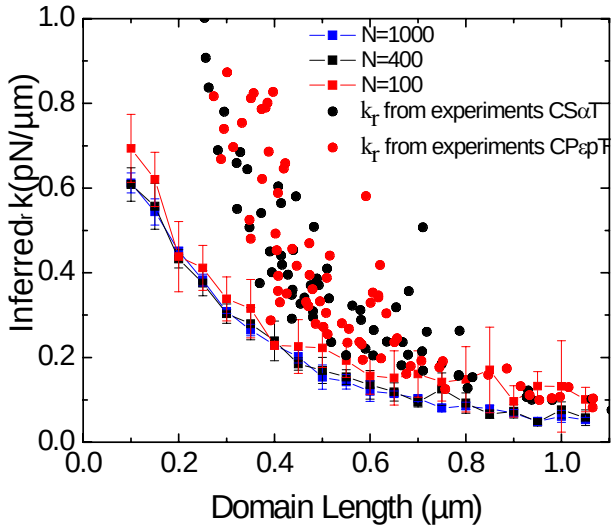

FIGURE S6: Comparison of spring constants extracted from numerical trajectories in a flat-well potential and spring constants extracted from experimental trajectories of CpεTand CsαTreceptors.Trajectories in a flat-well potential should yield an inferred k_r of 0. However, due to the optimization process and the reflections of the walls, the inferred potential, using the algorithm that is expecting a spring potential, is not perfectly flat. The determined k_r is shown for trajectories of varying length N. The calculated data points (squares) show the mean value and the standard deviation obtained from 100 numerical trajectories, the solid lines are a guide to the eye. The k_r values inferred from experimental trajectories are shown as red and black circles for CpεTand CsαTreceptors, respectively.Given that the experimental trajectories are circular or elliptical, the experimental data points are attributed to square-box domain lengths with equal confinement domain area.