Supplementary Information

Mapping the energy and diffusion landscapes of membrane receptors at the cell surface using high-density single-molecule imaging and Bayesian inference

J.-B. Masson[1,2], P. Dionne[‡3,4*], C. Salvatico[‡5*], M. Renner[5], C.G. Specht[5], A. Triller[5] and M. Dahan[3,6]

1 Institut Pasteur, Physics of Biological Systems, 28 rue du Dr Roux, 75724, Paris Cedex 15, France.

2 CNRS, URA 2171, F-75015 Paris, France.

3 Laboratoire Kastler Brossel, Centre National de la Recherche Scientifique UMR8552, Ecole normale supérieure, Université Pierre et Marie Curie-Paris 6, 75005 Paris, France;

4 Centre de Recherche Universite Laval Robert-Giffard, Quebec, Canada.

5 Biologie Cellulaire de la Synapse, Institut Nationale de la Santé et de la Recherche Médicale U789, Ecole normale supérieure, 75005 Paris, France.

6 Present address: Laboratoire Physico-Chimie, Institut Curie, CNRS UMR168, Paris France

* These authors contributed equally

1. Elements on the inference scheme

General Procedure of the Bayesian inference. The inference scheme features two generic steps: first, the derivation of the *posteriori* probability distribution of the unknown parameters given the experimental observations and second the sampling of the posterior distribution to estimate the parameters. It follows from Bayes rule that the posterior probability distribution $P(\{U_l\}|\{T_k\})$ of the set of the unknown parameters $\{U_l\}$ given the set of observed trajectories $\{T_k\}$ reads

$$P\left(\{U_l\}|\{T_k\}\right)=\frac{P\left(\{T_k\}|\{U_l\}\right)P\left(\{U_l\}\right)}{P\left(T_k\right)} \quad \text{(S1)}$$

where $P(\{T_k\}|\{U_l\})$ is the likelihood of the trajectories given the set of parameters $\{U_l\}$, $P(\{U_l\})$ is the prior probability of the set of parameters and $P(\{T_k\})=\int P(\{T_k\}|\{U_l\})P(\{U_l\})d\{U_l\}$ is the evidence of the model. Without any prior knowledge we used Jeffrey's prior as a prior distribution (discussed below). The estimator of the set of parameters $\{U_l\}$ (in this paper) is the Maximum A Posteriori (MAP), *i.e.* the values of $\{U_l\}$ that maximize the posterior distribution.

Prior Information. There are various choices of priors for the inference depending on its expression and on what characteristics of the diffusion and potential fields are to be extracted.

Jeffrey's rule: Results displayed in the paper were inferred using the following prior. Experimentally, there are no prior information on the diffusion and potential fields. Yet, in Bayesian inference, prior knowledge can be extracted from the symmetries of the likelihood distribution [1]. Jeffrey's rule states that:

$$P\left(\{U\}\right)\propto\sqrt{|J|} \quad \text{(S2)}$$

where J is the matrix defined by

$$J = \left(\partial_U \partial_{U'}^T \int d\mathbf{r} \sqrt{P\left(\mathbf{r}|\{U\}\right) P\left(\mathbf{r}|\{U'\}\right)} \right)_{U=U'} \tag{S3}$$

and ∂_U is the gradient with respect to the components U_l of the N-dimensional hypothesis

$$\partial_U = \begin{pmatrix} \partial/\partial U_1 \\ \cdot \\ \cdot \\ \partial/\partial U_N \end{pmatrix} \tag{S4}$$

with N the total number of parameters. The operator $\partial_U \partial_{U'}^T$ is the dyadic product of the gradients, *i.e.* the matrix of the second derivatives. Note that $\sqrt{|J|}$ is proportional to the Fisher Information of the likelihood [2,3].

Inside a mesh domain the likelihood reads

$$P\left(\Delta\mathbf{r}, \Delta\mathbf{t} \middle| D_{i,j}, \nabla_x V_{i,j}, \nabla_y V_{i,j}\right) =$$

$$\frac{1}{4\pi\left(D_{i,j} + \frac{\sigma^2}{\Delta t}\right)\Delta t} \exp\left(-\frac{\left(\Delta x - D_{i,j}\nabla_x V_{i,j}/k_B T\right)^2}{4\left(D_{i,j} + \frac{\sigma^2}{\Delta t}\right)\Delta t} - \frac{\left(\Delta y - D_{i,j}\nabla_y V_{i,j}/k_B T\right)^2}{4\left(D_{i,j} + \frac{\sigma^2}{\Delta t}\right)\Delta t} \right) \tag{S5}$$

where all the sources of positioning noise, such as Poisson noise, background fluorescence, algorithm to fit the intensity profile… are modelled into one Gaussian distribution with null average and σ standard deviation. This leads to the prior

$$P\left(D_{i,j}, \nabla_x V_{i,j}, \nabla_y V_{i,j}\right) \propto \frac{D_{i,j}^2}{\left(D_{i,j}\Delta t + \sigma^2\right)^2} \tag{S6}$$

An example of prior is plotted in Fig. S1:

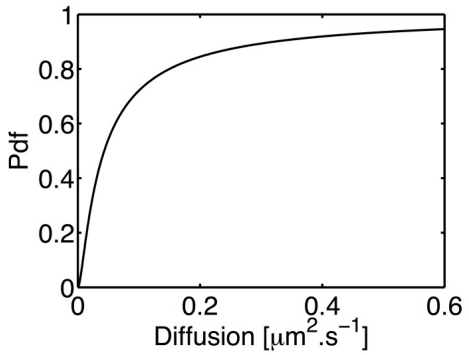


Figure S1. Prior distribution of the parameters with Δt= 0.05s and σ= 30nm. The Prior distribution cannot be normalized.

Interestingly, this prior is still improper (it cannot be normalized) and if there were no positioning noise the prior would be flat which would lead to maximum likelihood optimization for the estimation of the diffusion and potential fields.

The smoothing prior: In the version of the inference scheme used in the analysis of the data, high gradient damping is introduced in the potential optimization scheme. It was also possible to introduce it directly, through a prior penalizing high gradients values:

$$P\left(\left\{\nabla V_{i,j}\right\}_{(i,j)=(1..n_x,1..n_y)},\left\{D_{i,j}\right\}_{(i,j)=(1..n_x,1..n_y)}\middle|\left\{T_l\right\}\right)=\prod_{(i,j)=(1..n_x,1..n_y)}P\left(\nabla V_{i,j},D_{i,j}\middle|\left\{T_l\right\}\right)\times P(\nabla V)$$

$$\propto\prod_{(i,j)=(1..n_x,1..n_y)}\prod_{\mu:r_\mu\in S_{i,j}}\frac{e^{\left(-\left(\mathbf{r}_{\mu+1}-\mathbf{r}_\mu-D_{i,j}\nabla V_{i,j}\Delta t_\mu/k_BT\right)^2\Big/4\left(D_{i,j}+\frac{\sigma^2}{\Delta t_\mu}\right)\Delta t_\mu\right)}}{4\pi\left(D_{i,j}+\frac{\sigma^2}{\Delta t_\mu}\right)\Delta t_\mu}\times e^{\left(-\beta\left\|\nabla V_{i,j}\right\|^2\right)} \tag{S7}$$

It is worth noting that in that case, the potential evaluation could be performed by minimizing:

$$\xi\left(\left\{V_{i,j}\right\}\middle|(i,j)\in\left\{\left|N(i,j)\right|\neq 0\right\}\right)=\sum_{(i,j)}\left(\nabla V_{i,j}-\nabla V_{i,j}^{\inf}\right)^2 \tag{S8}$$

For our measurements, the two damping methods led to similar results.

Experimental Prior: Experimental data accumulated on various cells can be used as priors for future inference. Statistics of diffusion values in the mesh cells can be used as a global prior on diffusion values. Similarly, the smoothness of the diffusion and potential fields can be extracted from experimental data by fitting the distribution of $\left\|\nabla V_{i,j}\right\|^2$ and $\left\|\nabla D_{i,j}\right\|^2$ with the distribution $p_\beta(x)=e^{-\beta x}$ to find the optimal β_{opt} and then use $p_{\beta_{opt}}$.

Likelihood of the diffusivity and gradient potential over the cell surface. The local surface of the cell is meshed with regular subdomains whose dimensions are proportional to the average jump lengths. The typical length of the subdomains is designed so that the biomolecule motion between two consecutive frames happens either inside the same mesh subdomain or between two nearing subdomains. Note that nothing prevents irregular meshing of the surface of the cell, as long as there is no undersampling of internal mesh rectangles. Within each mesh subdomain the potential gradient is approximated to be constant. So inside the mesh square *(i,j)* the solution of Fokker-Planck equation reads

$$P\left(\mathbf{r},t\middle|\mathbf{r}_0,t_0\right)=\frac{e^{\left(-\left(\mathbf{r}-\mathbf{r}_0-D_{i,j}(\nabla V)_{i,j}(t-t_0)\right)/k_BT\right)^2\Big/4\left(D_{i,j}+\frac{\sigma^2}{(t-t_0)}\right)(t-t_0)\right)}}{4\pi\left(D_{i,j}+\frac{\sigma^2}{(t-t_0)}\right)(t-t_0)} \tag{S9}$$

As all the mesh domains are independent, the global posterior distribution *P* is the product of the posteriori inside each of them [4]:

$$P\left(\left\{\nabla V_{i,j}\right\}_{(i,j)=(1..n_x,1..n_y)},\left\{D_{i,j}\right\}_{(i,j)=(1..n_x,1..n_y)}\middle|\left\{T_l\right\}\right)=\prod_{(i,j)}P\left(\nabla V_{i,j},D_{i,j}\middle|\left\{T_l\right\}\right)$$

$$=\prod_{(i,j)}\left(\prod_{l}\prod_{\mu:\mathbf{r'}_\mu\in S_{i,j}}\frac{\exp\left(-\left(\mathbf{r'}_{\mu+1}-\mathbf{r'}_\mu-D_{i,j}\nabla V_{i,j}\Delta t/k_BT\right)^2\Big/4\left(D_{i,j}+\frac{\sigma^2}{\Delta t}\right)\Delta t\right)}{4\pi\left(D_{i,j}+\frac{\sigma^2}{\Delta t}\right)\Delta t}\right) \tag{S10}$$

where μ designates the index for which the points $\mathbf{r}^{l}_{\mu}$ of trajectory l are in $S_{i,j}$, σ is the experimental localization accuracy (~30 nm) and Δt the acquisition time. The estimators $\left(D^{\mathrm{MAP}}_{i,j},\ \nabla V^{\mathrm{MAP}}_{i,j}\right)$ of the local diffusivity and force field are the Maximum a Posteriori (MAP) of the posterior distribution of the parameters.

Potential optimization from the gradient field. The potential at the surface of the cell is extracted from the MAP values of its gradient field. The estimation is performed by minimizing $\xi(\{V_{i,j}\})$, which is defined as

$$\xi\left(\left\{V_{i,j}\right\}\middle|(i,j)\in\left\{\left|N(i,j)\right|\neq 0\right\}\right)=\sum_{(i,j)}\left(\nabla V_{i,j}-\nabla V^{\mathrm{MAP}}_{i,j}\right)^2+\beta(\delta)\sum_{(i,j)}\left(\nabla V_{i,j}\right)^2 \qquad \text{(S9)}$$

with |N(i,j)| the number of neighboring occupied mesh domains, β(δ) a constant (optimized on numerically generated trajectories) depending on δ that is the ratio between the size of the domains and the average length of the protein motion during Δt. In order to diminish the bias that the mesh could induce in the evaluation of $\nabla V_{i,j}$, the gradient along the two directions x and y was fitted, using two neighboring sites, by parabolic functions $\Im_x(x,y)$ and $\Im_y(x,y)$ so that $\nabla_x V_{i,j}=\partial\Im_x(x,y)/\partial x$ and $\nabla_y V_{i,j}=\partial\Im_y(x,y)/\partial y$. $\xi(\{V_{i,j}\})$ is the sum of two terms: the first one aims to minimize the difference between the gradient of the potential field and the inferred gradient field while the second one penalizes the strong gradients. This penalization allows better convergence towards the true value of the potentials; it also prevents anomalous generation of very high potential areas that would modify the potential on a large scale.

We emphasize that typical experimental trajectories lead to several hundred to thousands of variables. The quality of the optimization on such a large number of variables was investigated numerically. Typical procedures, similar to the ones used in [5-9], have been employed to test the convergence and the quality of the inference. Here we discuss the choice of β(δ). Numerical simulations with the same diffusivities, the same average number of points per mesh rectangle, the same average trajectory duration and globally the same potential energy landscape were used to generate trajectories that were subsequently inferred for various values of β. The optimal β was the one that minimized the average square difference (ψ) between the input and the inferred potential field.

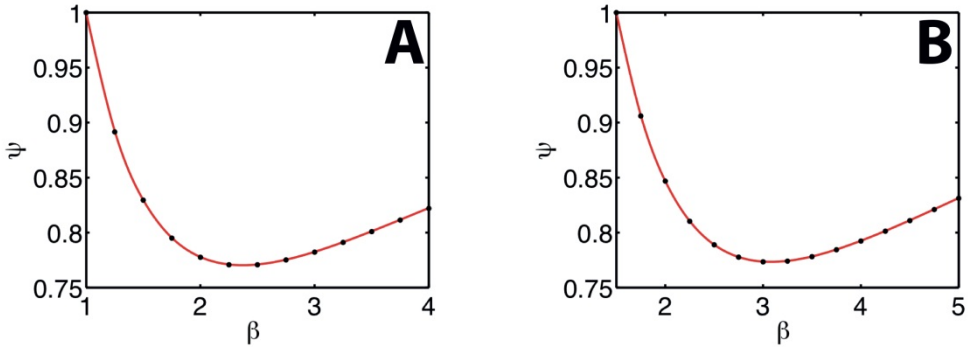


*Figure S2. **A)** Evolution of the average difference ψ (normalized here) between the theoretical potentials and inferred ones with β for a mesh subdomain with δ=2. **B**) Evolution of the average difference ψ (normalized here) between the theoretical potentials and inferred ones with β for a mesh subdomain with δ=3.*

β was found to mainly depend on the choice of δ, which in our experiments was always set between 1 and 3 to minimize the number of holes in the mesh structure. Note that the choice of β may also be driven by the search of some specific structures in the potential field, low values of β would favor large local variations of the potential whereas high values favor low variations over large scale and damp the large local variations of the potential.

Inference applied to a synaptic cluster. When dealing with local motion (*i.e.* inside a cluster), the inference scheme was derived from the ones introduced in [5-8]. Trajectories entering the synaptic area, identified as the translocations entering the fluorescent signal of the tagged gephyrin, were grouped together to be analysed. The potential was developed on a basis of function so that the inference is performed on the coefficient of the development. A simple polynomial basis provided good results (see refs [5-9]):

$$V_{2D}(x,y) = \sum_{l+m\le N} \alpha_k (x-x_c)^l (y-y_c)^m \tag{S11}$$

with (x_x, y_c) the ~~barycenter~~ center of mass of the trajectory points, $k = \frac{(l+m)(l+m+1)}{2} + l$ and N the order the polynomial. The posteriori reads:

$$\begin{aligned} P\left(\{\alpha_k\}_{k=0..\frac{N^2+3N}{2}}, \{D_{i,j}\}_{(i,j)=(1..n_x,1..n_y)} \middle| \{T_l\}\right) &= \left(\prod_{(i,j)=(1..n_x,1..n_y)} P\left(\{\alpha_k\}_{k=0..\frac{N^2+3N}{2}}, D_{i,j} \middle| \{T_l\}\right)\right) \times P(\nabla V, D) \\ &\propto \prod_{(i,j)=(1..n_x,1..n_y)} \prod_{\mu \mathbf{r}_\mu \in S_{i,j}} \frac{\exp\left(-\left(\mathbf{r}_{\mu+1} - \mathbf{r}_\mu - D_{i,j}(\nabla V)_{i,j}\Delta t / k_B T\right)^2 \Big/ 4\left(D_{i,j} + \frac{\sigma^2}{\Delta t_\mu}\right)\Delta t\right)}{4\pi\left(D_{i,j} + \frac{\sigma^2}{\Delta t}\right)\Delta t} \times \frac{D_{i,j}^2}{\left(D_{i,j}\Delta t + \sigma^2\right)^2} \end{aligned} \tag{S12}$$

With $S_{i,j}$ the mesh rectangle *(i,j)*, μ the index where the biomolecule is in *(i,j)*, Δt the acquisition time. Each mesh rectangle was treated independently and they were coupled together by the potential polynomial development. In the numerical implementation of the scheme, positions were expressed in μm, the diffusivity in $\mu m^2 s^{-1}$ and the potential in $k_B T$. Theoretically, the choice of the optimal order of the polynomial is made by comparing, for varying value of N, the evidence of the model. As the order increases, the space of integration gets larger. Due to the shape of the posteriori (product of Gaussians) integrals may be approximated by the Laplace method, *i.e.* the posteriori is approximated by a Gaussian probability distribution centered on the MAP and so the integral is evaluated directly. Yet, in a practical manner, the algorithm runs fast and so the criterion to choose the order of the polynomial may simply be the absence of change in the potential shape and the lowest error on the coefficient of the development.

Note that simple polynomial development is not the only possible basis. Depending on the type of motion and on the global geometry of the trajectory, 2D Fourier series and 2D orthogonal Hermitte functions are useful bases. The latter is interesting when dealing with local confinement surrounded by free motion. Yet, both are much less versatile in their use than the simple polynomial basis.

Error estimation. An estimate of the error on the various parameters can be obtained using one of the two following general strategies: first, the direct sampling of posterior distribution using Monte Carlo algorithm and second the evaluation of the eigenvalues of the log-posteriori Hessian at the MAP values. The latter method takes advantage that most of the relevant quantities are extracted from the neighborhood of the maximum of the distribution. However, the

Hessian, which is much faster to compute than the Monte Carlo Algorithm, will tend to slightly over-estimate the noise of the diffusivities. Furthermore, if one is interested in the error evaluation for the difference of potential between two points, Monte Carlo sampling of the posteriori is the more efficient way.

Initialization of the parameters before optimization. In the two inference schemes, the potential and the diffusivity fields have to be initialized to accelerate the convergence of the optimization processes. When the inference is performed on large surfaces, the potential is initialized by the thermal equilibrium values, *i.e.* $V_{i,j} = -k_B T \log\left(N_{i,j}/N_{Max}\right)$ with $N_{i,j}$ the number of points in the mesh subdomain *(i,j)* and N_{Max} the maximal number of points inside one of the mesh subdomains. The diffusivity fields can be initialized in two ways leading to approximately the same computation time. The most common consists in approximating the diffusivity along the arbitrarily defined x-axis and y-axis as $D_x = \left\langle l_x \right\rangle^2 / \Delta t$ and $D_y = \left\langle l_y \right\rangle^2 / \Delta t$ with $\left\langle l_x \right\rangle$, $\left\langle l_y \right\rangle$ that are the average step sizes along x and y during Δt. Then, the diffusivity field is initialized to $D_{i,j} = 1/3\left(D_x + D_y\right)$. The other one consisted on performing the same calculus but locally on groups of mesh subdomains. When the inference is performed locally, the diffusivity field is initialized in the same manner as for the inference on large surfaces. The coefficient of the polynomial development of the potential are initialized on an harmonic approximation of the potential with the spring constant $k = 2V_0 \big/ \left|\mathbf{r}_{\max} - \mathbf{r}_{mean}\right|^2$ with $\mathbf{r}_{mean}$ the average position, $\mathbf{r}_{\max}$ the position corresponding to the largest distance from $\mathbf{r}_{mean}$ and V_0 a user defined potential value, usually set to $5k_BT$.

Possible bias during the potential landscape optimization. Here, we want to emphasize an important point. In all the possible applications of the different schemes shown here, it is essential to quantify their behaviors with extensive simulations. Furthermore, most of the bias that we have encountered (see for example [7,8]) can be analytically corrected. Fortunately, the overdamped Langevin equation framework is fast to simulate, allowing direct testing of the various types of inference schemes. Theoretically, the quality, the rate of convergence and the optimality have to be evaluated by computing the Fisher Information [2,3]. Unfortunately, in most of these systems, the summation over all possible parameters state is impossible due to its large size. Hence, we have to rely on extensive simulations to study the behavior of the inference.

In order to illustrate the bias that can appear on the potential maps, we show on the next figure an example of complex field inference. As an example of complex field, we chose an oscillating potential because it induces a very incomplete sampling of space due to the proximity of low potential region to high potential region:

$$V(\mathbf{r}) = k_B T\left(2 + \cos\left(\frac{2\pi}{\lambda}x\right) + \cos\left(\frac{2\pi}{\lambda}y\right)\right) \tag{S13}$$

with λ=800 nm. The inferred potential (Fig S3. A,B) matches the structure and the amplitude of the potential used to generate the trajectories. The advantage of using symmetric potential fields is that it highlights (by propagating them) the defects of the optimization procedure. Interestingly, the potential is well inferred near the boundaries of the domains,

while in some parts near the center, there are small differences between the inferred and the simulated potential but these errors do not propagate thanks to the damping term $\beta\left[\vec{\nabla}V(\mathbf{r})\right]^2$ in the optimization function. Overall, the inference is able to recover fast variation of the potentials with local bias near high potential structures.

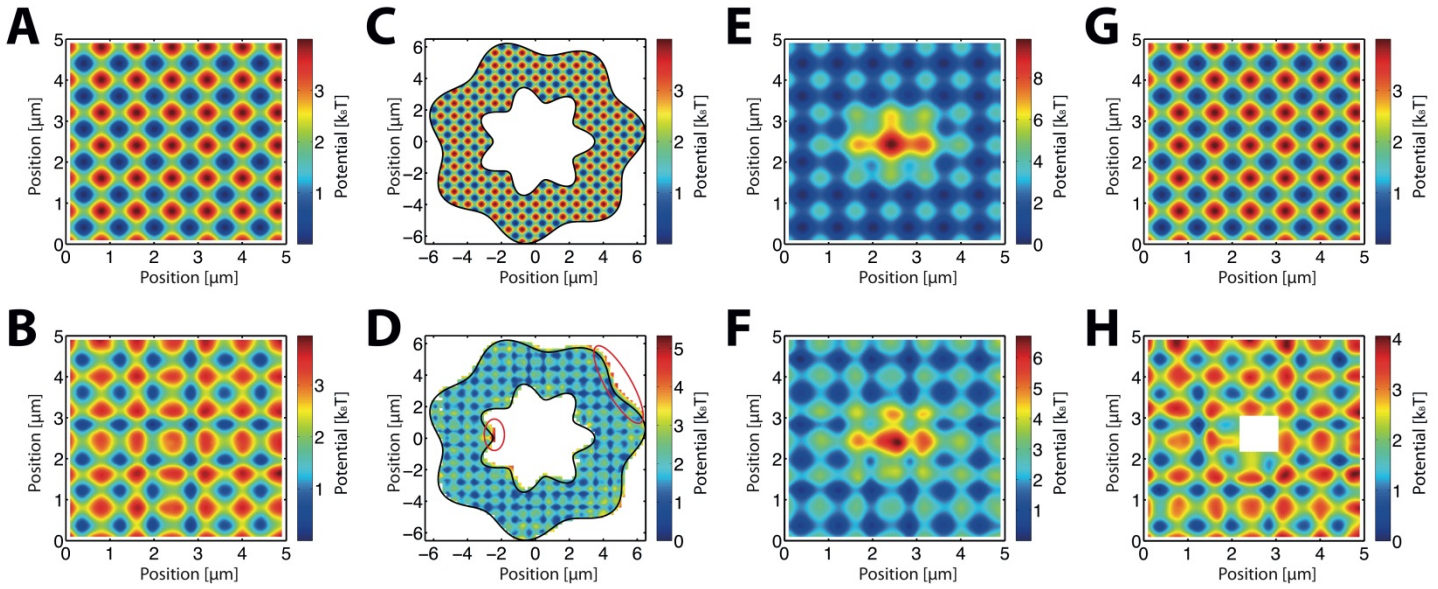


*Figure S3. **A)** Potential map used to generate the trajectories. **B)** The inferred potential field with trajectories simulated in **A**). **C)** Potential map used to generate the trajectories. **D)** The inferred potential field with trajectories simulated in **C**). The red ellipses point the two local areas where there is a significant bias. Simulations to generate the trajectories were made with reflecting boundary conditions on the borders, shown here by the black curves. **E)** Potential map used to generate the trajectories. **F)** The inferred potential field with trajectories simulated in **E**). **G)** Potential map used to generate the trajectories. **H)** Inferred potential field with trajectories simulated in **G**). The white square is the un-sampled part of space due to the removal of the trajectory points. Simulations are performed with a diffusion field with constant value $D=0.1\mu m^2 s^{-1}$ and $\Delta t=50ms$. Note that the colorbar may be different between the simulated and inferred landscape.*

More generally, the borders tend to generate limited or very localized bias. In Fig.S3C,D we show on a complex geometry (with an identical potential as the one used in Fig.S3A,B) that the borders do not shift significantly the inferred potential from their true values in the internal part of the maps, and that mainly on the top right of the geometry, in a region mostly unvisited the potential is locally biased. In most parts of the border regions there are no biases in the inferred potentials. Note that the differences in the color of the maps are induced by the high potential region on the top of Fig.S3D. The inference is not going to detect every possible structure. Obviously, diffusion or interaction structures on scales largely inferior to the scale of the diffusion motion ($\sim\sqrt{4D_{i,j}\Delta t}$) cannot be seen. Note that it does not mean that other estimators could not be used to extract characteristics of these structures, yet in order to access these scales prior knowledge would be necessary. One of the key interests of the mapping scheme is that no prior information is needed to extract information from the random motion. Some structures of potential are also going to shield others from being accessible to the inference. In Fig.S3E,F we show an example on a more heterogeneous potential. The potential is well inferred except in the region of high potential that is partially inaccessible due to its high value. Yet, the average

difference between local maxima and minima of the potential, in the periodic structure, is the theoretical value (4 k_BT). This proves that some potential structures can prevent the inference from accessing local information but that this will not have large scale effects. Finally, some regions may be unvisited for technical reasons such as regions that are out of focus or simply the absence of experimental points. In Fig.3G,H we show the effect of un-sampled areas by removing all the points in the trajectories inside a square. We made the empty square larger than the ones met experimentally to induce a noticeable effect. Again the inference is slightly biased near the anomaly (lack of points) but remains of good quality on the rest of the surface. There are no biases in the estimation of the difference between the maxima and the minima of the potentials (4 k_BT).

Finally, we tested the inference scheme in the absence of potentials to ensure that optimization on large surfaces did not generate local potentials (this test had already been performed for the confined trajectories in local potentials [7,8]).

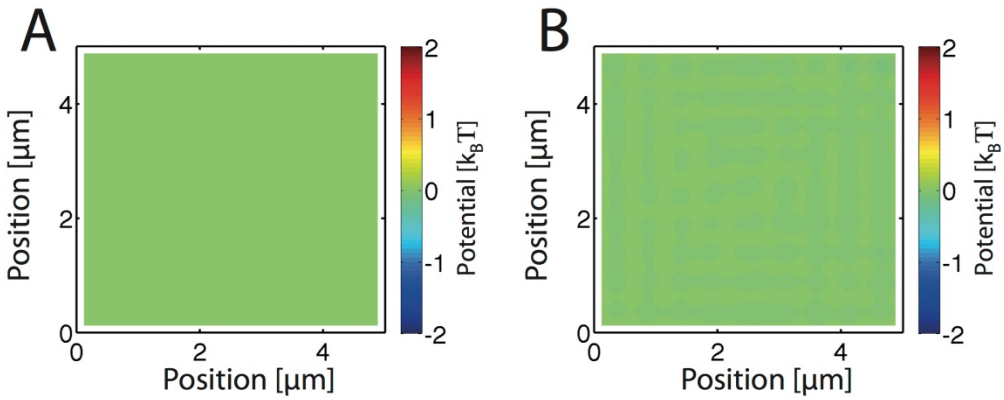


Figure S4. A) Constant potential used to generate trajectories. B) Inferred potential. Note the low values of the potential value; the noise in the potential is 0.15 k_BT; it is the consequence of the mesh with small variation of potential near neighboring mesh subdomains.

Some of the possible structures that may appear when there is a constant potential are small oscillations of potential values between neighboring mesh subdomains. They are mostly the consequence of the finite mesh size. Yet, if experimentally we would expect small potential values or large areas without potentials variations, the value of β should be raised. Another possible way to deal with flat regions and other possibly complex structures would be to build non-regular types of mesh. Meshes generated to have identical number of trajectory points lead to very good results by diminishing the bias in the inferred potential. Yet, the relations between the numbers of points in each mesh subdomain, the characteristics of the potential and diffusion landscape and the value of β are less direct than for regular meshes. In Fig.S5, we show an example of these structures in the experimental energy landscape for the β^--TM construct

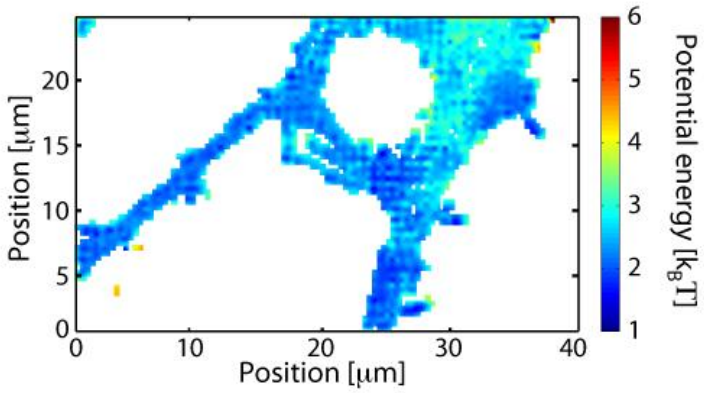


Figure S5. Energy Maps for membrane construct β-TM.

2. **Simulations and measurements in the landscapes**

Why performing Simulations in the Maps. Probing a biological environment using single molecule technics (SPT, PALM, UPaint etc.) involve a tradeoff between 3 main phenomena:

- Spatial Sampling
- Duration of Recording
- Typical time of biological variations

Efficient spatial sampling, especially for large areas (>1μm^2), can either be achieved by long trajectories or by numerous short trajectories. It is also worth noting that the nature of the motion (fractional Brownian Motion, Continuous Time Random Walk etc.) may prevent efficient spatial sampling. Precise measures of temporally averaged estimators (MSD, first time passage) require long trajectories to be discriminative. Yet, the possible heterogeneities of the surrounding media are going to be inserted in the estimators preventing their direct use as parameters estimators or as model discriminant. Finally, All measures have to be made with minimal variations of the biological system. Again, it is worth noting that in most biological systems all the characteristics times of biological variations are not known, hence measures based on the fluctuations of interests estimators are useful to detect significant variations of the biological system.

In order to analyse, at multiple scales, the neuronal membrane, we combined the UPaint recording, the inference scheme and the simulations in the maps. Large numbers of short trajectories allow efficient space sampling even for complex media. Short individual recordings prevent the use of estimators on individual trajectories. The total time of recording (few minutes) does not lead to significant changes in the measures (done by the inference). The inference allows the measure of the diffusion and potential field on the membrane. Hence, we have shifted the balance towards space mapping and short time recording. Simulations in the maps allow studying any estimator without limitations in time or in number of trajectories. Furthermore, they allow the use of “Ensemble Averaged” estimators that are more selective than temporally estimated estimators.

Simulations in the landscapes. The maps of the diffusion and energy landscapes, D(r) and V(r), can be used to simulate the behavior of the molecules at different time and space scales. In each mesh sub-domain *(i,j)* a diffusivity $D_{i,j}$ is associated with a potential energy value $V_{i,j}$. The dynamics of the molecules are described by the Fokker-Planck equation:

$$\frac{\partial P\left(\mathbf{r},t|\mathbf{r}_0,t_0\right)}{\partial t} = -\nabla \bullet \left(-\frac{\nabla V(\mathbf{r}) P\left(\mathbf{r},t|\mathbf{r}_0,t_0\right)}{\gamma(\mathbf{r})} - \nabla\left(D(\mathbf{r}) P\left(\mathbf{r},t|\mathbf{r}_0,t_0\right)\right)\right) \tag{S14}$$

where $P\left(\mathbf{r},t|\mathbf{r}_0,t_0\right)$ is the conditional transition probability from ($\mathbf{r}_0$,t_0) to ($\mathbf{r}$,t). Fokker-Planck equations can always be approximated by Master equations:

$$\frac{dP_{(i,j)}(t)}{dt} = \sum_{(i',j')\in\mathcal{N}(i,j)} W_{(i,j),(i',j')} P_{(i',j')}(t) - \sum_{(i',j')\in\mathcal{N}(i,j)} W_{(i',j'),(i,j)} P_{(i,j)}(t) \tag{S15}$$

with in our case

$$W_{(i,j),(i',j')} = \frac{D_{(i',j')}}{\Delta x^2} \exp\left(-\frac{\Delta x F^x_{(i,j),(i',j')}}{2\gamma_{(i',j')} D_{(i',j')}} \right) \tag{S16}$$

if the transition happens in x direction and

$$W_{(i,j),(i',j')} = \frac{D_{(i',j')}}{\Delta y^2} \exp\left(-\frac{\Delta y F^y_{(i,j),(i',j')}}{2\gamma_{(i',j')} D_{(i',j')}} \right) \tag{S17}$$

if the transition happens in the y direction and with $W_{(i,j),(i',j')}$ the transition rate from the *(i',j')* site to the *(i,j)*, Δx (Δy) the mesh size in the x (y) direction, and $F^{x(y)}_{(i,j),(i',j')}$ the potential gradient acting on the random walker in the x (y) direction when moving from *(i',j')* to *(i,j)*. The motion of the molecule following equation (S19) was simulated using the Gillespie scheme [10]. When the molecule was at the site *(i,j)*, the transitions rates, rewritten a_υ to match Gillespie formalism, υ taking values from 1 to 4, were evaluated on all neighboring sites. We define $a_0=\Sigma_\upsilon a_\upsilon$. The time, τ, to move from the site *(i,j)* to a neighboring site is extracted from an exponential probability density function of rate a_0, so that

$$\tau = \frac{1}{a_0} \log\left(\frac{1}{r_1}\right) \tag{S18}$$

with r_1 a random number in [0,1]. The destination site, k, is chosen to satisfy

$$\sum_{\upsilon=0}^{k-1} a_\upsilon \le r_2 a_0 < \sum_{\upsilon=0}^{k} a_\upsilon \tag{S19}$$

with r_2 a random number in [0,1]. Limits of the neuronal cells and unvisited sites are defined as inaccessible sites. Note that the trajectory generation process leads to trajectories with non-constant time steps. In order to evaluate the different estimators (see below), trajectories were regularized to obtain the molecule position at regular time lags by imposing that as long as each τ was not reached the molecule did not move.

Computations of the estimators of the protein movement. Here, we explain how the various estimators of the protein movement were extracted and also show other possible estimators that were not included in the main text. Unless mentioned otherwise, the estimators were obtained by averaging trajectories coming from different neuronal maps or from different clusters. Using the simulated data, we computed:

- The propagator $\Pi(d,t)$, *i.e.* the probability of moving a distance d in a time t.
- The scaling factor $\chi(t)$ (Fig.4c), computed from $\Pi(d,t)$
- The ensemble-averaged mean squared displacement (MSD) defined as

$$\left\langle \mathbf{r}^2(t) \right\rangle = \frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{r}_i(t)-\mathbf{r}_i(0)\right)^2 \tag{S20}$$

with N the total number of trajectories.

- The time statistics to escape gephyrin clusters (Fig.S6A), computed by generating trajectories beginning in the clusters and measuring the time needed to exit the clusters.

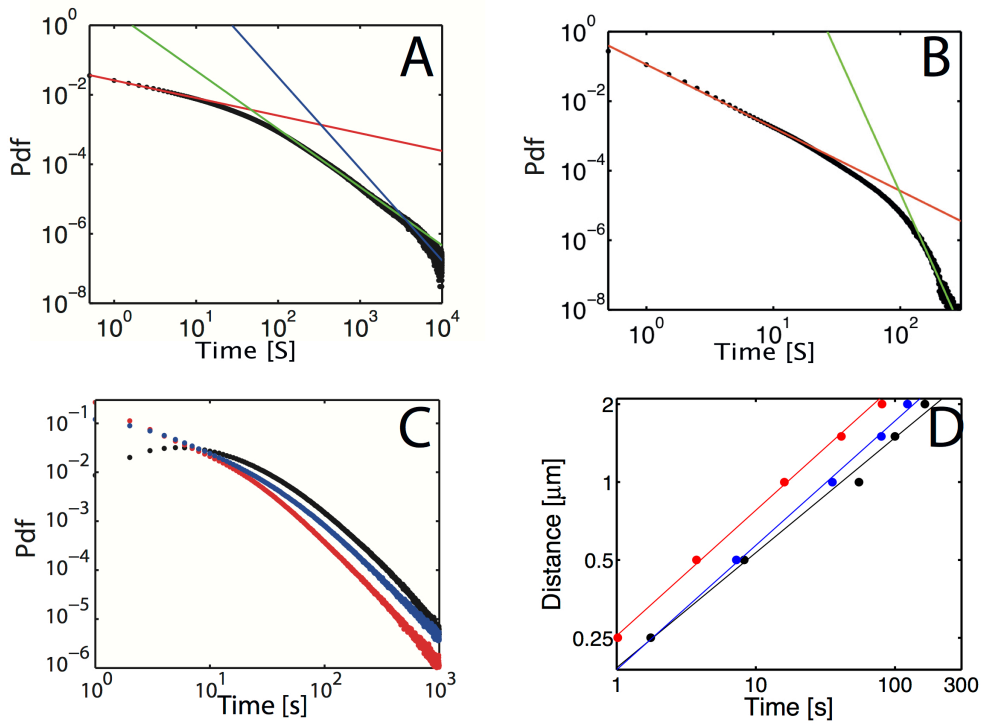


*Figure S6. **A)** Example of "time to escape gephyrin clusters" statistics based on simulations on subsets of β^{WT}-TM molecules, shown in log-log scale. Black points are results of simulation and colored lines are linear fits (in log-log) with slope α=-1.8 in red and in α=-8. in green. **B)** Example of "time to reach a gephyrin cluster" statistics based on simulation on subsets of β^{WT}-TM expressing neurons in log-log scale. Black points are results of simulation and colored lines are linear fits (in log-log) with slope α=-0.5 in red, α=-1.7 in green and α= -2.6 in blue. **C)** Statistics of first time passage for a distance of 1µm, average over the complete set of neurons. Colors are associated to receptors, β^{WT}-TM in black, β^{S403D}-TM in Blue and β^{-}-TM in Red. **D)** Evolution of the distance that receptors first reach with the average first time to reach it. Colors are associated in the same fashion as for **C)**. the distance scales as t^{η} with η=0.44 for β^{WT}-TM, with η=0.48 for β^{S403D}-TM and with η=0.49 for β^{-}-TM.*

- The time statistics to enter a gephyrin cluster (Fig.S6B), computed by measuring the time needed to get to a cluster, starting from any point outside of clusters.
- The first passage time statistics to reach a specific distance (Fig.S6C), computed by generating trajectories starting from any point at the surface of the neuron and measuring the time needed to reach the defined distance for the first time.

- The evolution of the distance to reach with the average first time to reach it (Fig.S6D), computed from the first passage time statistics.

Simulations of the fluctuations of receptor numbers. Fluctuations of the number of receptors in the gephyrin clusters were (numerically) measured by placing receptors at the surface of a neuron with experimentally measured densities, *i.e.* 500 receptors per μm^2 inside gephyrin clusters and 5 receptors per μm^2 outside, and then letting them evolve using the Gillespie scheme. We thus deduced the time course of the number of receptors at individual synapses (see results in Fig.4). Using the same method, we also computed a map of relative variations of receptor numbers (Fig.S7).

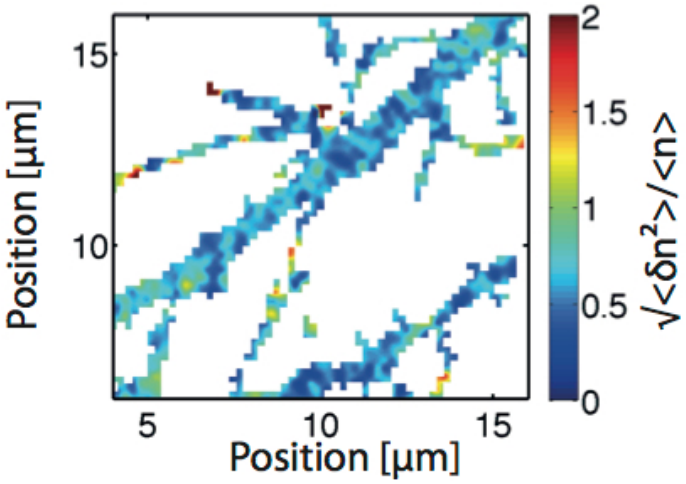


Figure S7. Map of the relative variation (standard deviation over mean) of receptor number during a 100 s temporal window.

3D Biases. The inference scheme is not limited to 2D motion and may be applied in 3D. In our case the transmembrane proteins movements were analyzed in 2D although the membrane can exhibit curvature. However, we show that the possible local curvature of the membrane has a limited effect on the inferred values.

We first simulated the random walk on a cylinder with high curvature $\varkappa$=4 μm^{-1} (R=250nm) and inferred the diffusivity field using the inference scheme developed in the paper. The motion was not limited along the axis of the cylinder. We show, on figure S8A, the statistics of diffusivity values inside all mesh squares

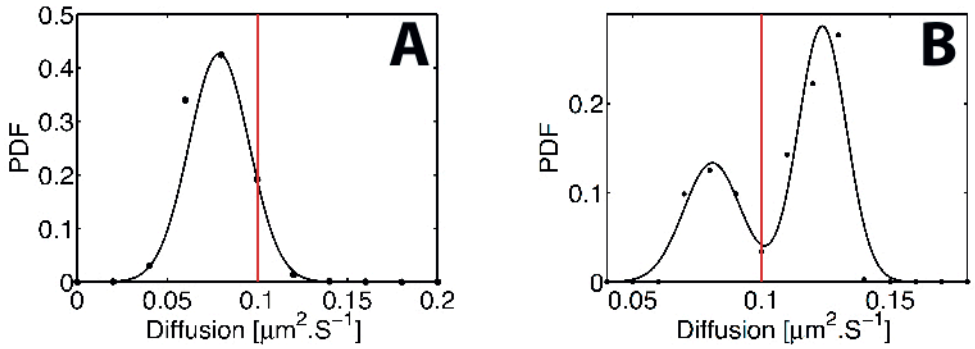


Figure S8 . Inferred diffusivity statistics of a random walker on a cylinder A) and on a sphere B) of radius 250nm with Δt=50ms. The diffusion used to generate the trajectories is D=0.1$\mu m^2 s^{-1}$ and is indicated by the red lines. Trajectories are recorded on (x,y) plane. The statistics is made with the MAP values of the diffusivity in each mesh square of the mesh.

We next simulated the random walk on sphere with high curvature $\varkappa$=4 μm^{-1} (R=250nm). Results are shown on Fig.S8B. The statistics show two local maxima centered on the true value of diffusivity, and the average value of the distribution is <D>=0.11±0.02 $\mu m^2 s^{-1}$. We observe that for high curvature the inference scheme tends to slightly shift the diffusivities. Interestingly, when there is curvature the diffusivity statistics have characteristics that differ from the ones on flat surfaces. Furthermore, the statistics of the points' positions that are concentrated on the limit of the cylinder or the sphere indicate the presence of curvature. So, here we deliberately applied the inference without taking the noticeable curvature effect (very inhomogeneous repartition of points) into account. Note that in both cases if the curvature is taken into account, by modifying the expression of the likelihood, both statistics become unbiased and centered on the true value of diffusivities.

Here, we show how to generate the trajectories on the cylinder and the sphere. The simulation of the random walk on the cylinder was straightforward because of the parameters values:

$$\begin{aligned} \theta_{t+dt} &= \theta_t + \sqrt{\frac{2Ddt}{R^2}} dU_t \\ y_{t+dt} &= y_t + \sqrt{2Ddt}\, dV_t \end{aligned}, \tag{S21}$$

with R the radius of the cylinder, θ the radial angle, y the direction of the cylinder, D the diffusivity, Δt the time between each move, and $\left(dU_t, dV_t\right)$ Gaussian random numbers of null average and of unit standard deviation.

In that case the propagator reads

$$P\left(\theta, z, t \middle| \theta_0, z_0, t_0\right) = \left[1 + \sum_{n\geq 1} \cos\left(2\pi n\left(\theta - \theta_0\right)\right) e^{-\frac{4\pi^2 D(t-t_0)}{R^2}}\right] e^{-\frac{(z-z_0)^2}{4D(t-t_0)}} \Big/ \sqrt{4\pi D\left(t-t_0\right)}\,, \tag{S22}$$

The simulation of the random walk on the sphere could not be simulated with a similar scheme

$$\begin{aligned} \varphi_{t+dt} &= \varphi_t + \sqrt{\frac{2Ddt}{R^2}} dU_t \\ \theta_{t+dt} &= \theta_t + \frac{Ddt}{R^2 \tan\left(\theta_t\right)} + \sqrt{2Ddt}\, dV_t \end{aligned}, \tag{S23}$$

Since a suitable dt would have to be such as $\sqrt{dt} \approx R^2 / \sqrt{D}$, diffusion was simulated on a shell of width *h* such that $\varepsilon = h/R < 0.01$ with reflecting boundary conditions on the inner and outer shell. dt was adjusted so that the relation $\left\langle \cos\left(\theta(t)\right)\right\rangle = e^{-\frac{2Dt}{R^2}}$ was respected. Note that another way to efficiently simulate the random walk on the sphere, or more generally on a n sphere in an n+1 Euclidian space, can be found in [11].

In that case the propagator reads.

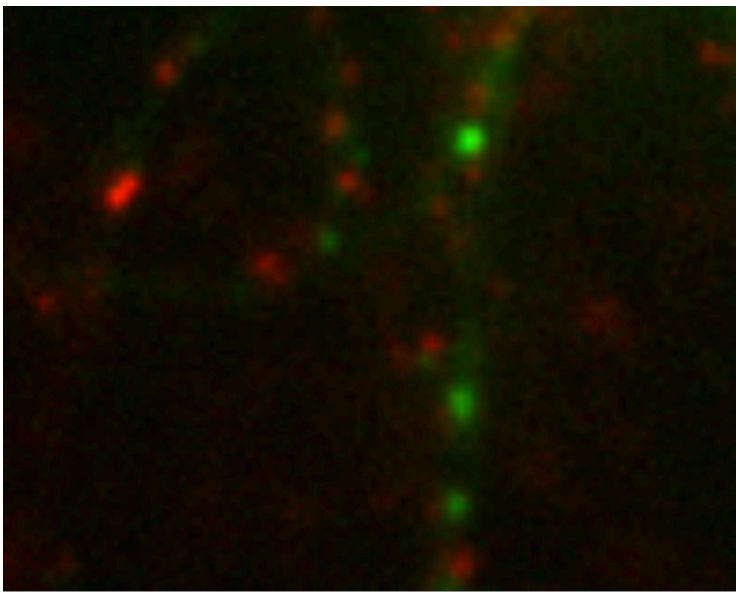
$$P\left(\theta,\varphi,t\middle|\theta_0,\varphi_0,t_0\right)=\sum_{k=0}^{\infty}\sum_{m=-k}^{k}Y_k^m\left(\theta,\varphi\right)Y_k^m\left(\theta_0,\varphi_0\right)e^{-\frac{k(k+1)D\left(t-t_0\right)}{R^2}}, \tag{S24}$$

with

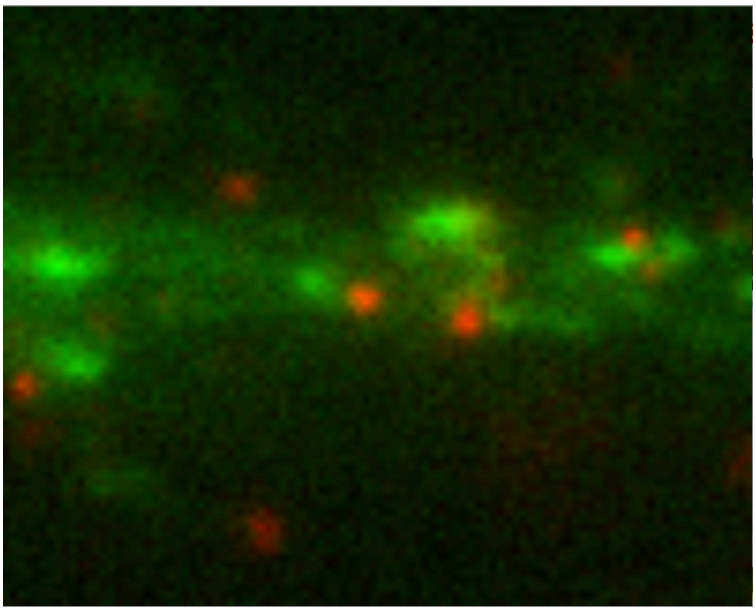
$$Y_k^m\left(\theta,\varphi\right)=e^{im\varphi}P_k^m\left(\cos\left(\theta\right)\right), \tag{S25}$$

with P_k^m the Legendre functions.

3. **Supplementary Videos**



Sup Movie 1: example of a uPaint movie for the beta- TM construct (red), labeled with anti-GFP antibodies coupled to Atto647N dues, and gephyrin molecules (green). Acquisition time: 50 ms/image.



Sup Movie 2 : example of a uPaint movie for the betaWT- TM construct (red), labeled with anti-GFP antibodies coupled to Atto647N dues, and gephyrin molecules (green). Acquisition time: 50 ms/image.

References

[1] H.L. Harney, *Bayesian Inference: Parameter Estimation and Decisions*, Springer 2003
[2] D. J.C. Mackay, *Information Theory, Inference and Learning Algortihms*, Cambridge University Press 2003
[3] T. M. Cover & J.A. Thomas, *Elements of Information Theory,* Wiley Series in Telecommunication and Signal Processing 2006
[4] H. Risken, *The Fokker-Planck Equation, Methods of Solution and Applications,* Springer (1996)
[5]. Masson J-B, Casanova D, Türkcan S, Voisinne G, Popoff MR, Vergassola M, Alexandrou A (2009) Inferring maps of forces inside cell membrane microdomains. Phys Rev Lett 102**:** 048103 .
[6] Voisinne G, Alexandrou A, Masson J-B (2010) Quantifying Biomolecule Diffusivity Using an Optimal Bayesian Method. Biophys J 98: 596-605 .
[7] Türkcan S, Masson J-B, Casanova D, Mialon G, Gacoin T, Boilot J-P, Popoff MR, Alexandrou A (2012) Observing the confinement potential of bacterial pore-forming toxin receptors inside rafts with non-blinking Eu^{3+}-doped oxide nanoparticles. Biophys J, 102: 2299-2308.
[8] Türkcan S, Alexandrou A, Masson J-B (2012) A Bayesian inference scheme to extract diffusivity and potential fields from confined single-molecule trajectories. Biophys J 102: 2288-2298.
[9] Türkcan S, Richly M.U.,Alexandrou A.,Masson J.-B, *Probing membrane protein interactions with their lipid raft environment using single molecule tracking and Bayesian inference analysis,* PloS ONE 8(1): e53073
[10] Gillespie D.T., *Exact Stochastic Simulation of coupled chemical reactions,* The Journal of Physical Chemistry, 81, No 25, p2340, (1977)
[11] J. Nissfolk, T. Ekholm & C. Elvingson, *Brownian dynamics simulations on a hypersphere in 4-space*, Journal of Chemical Physics, **119**, 6423 (2003)