Title

Comparison between iMSD and 2D-pCF analysis for molecular motion studies on in vivo cells: The case of the epidermal growth factor receptor

Authors

Malacrida, Leonel
Rao, Estella
Gratton, Enrico

Publication Date

2018-05-01

DOI

10.1016/j.ymeth.2018.01.010

Peer reviewed

HHS Public Access

Author manuscript

Methods. Author manuscript; available in PMC 2019 May 01.

Published in final edited form as:

Methods. 2018 May 01; 140-141: 74–84. doi:10.1016/j.ymeth.2018.01.010.

Comparison between iMSD and 2D-pCF analysis for molecular motion studies on *in vivo* cells: the case of the epidermal growth factor receptor

Leonel Malacrida[1,2,#], **Estella Rao**[1,3,#], and **Enrico Gratton**[1]

[1]Laboratory for Fluorescence Dynamics, Department of Biomedical Engineering, University of California., Irvine, USA

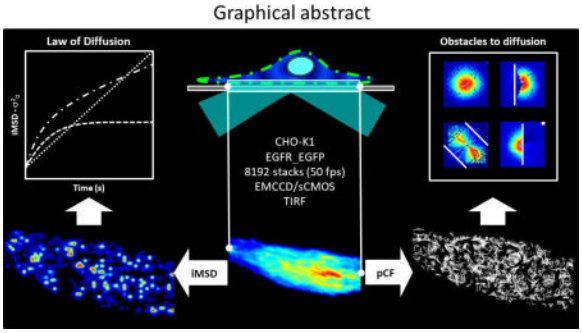
[2]Área de Investigación Respiratoria, Departamento de Fisiopatología, Hospital de Clínicas, Facultad de Medicina, Universidad de la República, Uruguay

[3]Dipartimento di Fisica e Chimica, Università di Palermo, Italy

Abstract

Image correlation analysis has evolved to become a valuable method of analysis of the diffusional motion of molecules in every points of a live cell. Here we compare the iMSD and the 2D-pCF approaches that provide complementary information. The iMSD method provides the law of diffusion and it requires spatial averaging over a small region of the cell. The 2D-pCF does not require spatial averaging and it gives information about obstacles for diffusion at pixel resolution. We show the analysis of the same set of data by the two methods to emphasize that both methods could be needed to have a comprehensive understanding of the molecular diffusional flow in a live cell.

Graphical abstract



*corresponding author egratton22@gmail.com.

#LM and ER contributed equally to this work

Keywords

Fluorescence Fluctuation spectroscopy; diffusion anisotropy; barrier to diffusion; connectivity maps

1. Introduction

In this article we describe two methods to analyze the same set of data. The iMSD (image Mean Square Displacement) method is used to obtain the diffusion law in small regions of interest (ROI) of the cell [1]. The term diffusion law indicates here that the iMSD plot can be fitted with a parabolic equation (flow), a straight line (pure diffusion), a function that bends at a given value of the iMSD (confined diffusion) or a combination of diffusion at a small spatial scale and slow apparent diffusion at a large spatial scale. These terms are discussed in [1]. The iMSD analysis provides information similar to the MSD obtained using single particle tracking, but is based on spatio temporal image correlation functions (STICS) and therefore do not require to isolate and track single particle. Since the correlation function is calculated in a given ROI, the diffusion law obtained refers to the ROI. The spatial resolution depends on the size of the ROI. In figure 1 we discuss the required size of the ROI as a function of the local diffusion coefficient. In figure 2 we show the concept of diffusion law in the contest of the iMSD method. The second method discussed in this article is named 2D-pCF (two dimensional pair correlation function) has the resolution of a pixel and provides the average path followed by molecules in the proximity of obstacle [2]. In the application of the 2D-pCF in article the 2D-pCF is not used to obtain values of the diffusion coefficient or the law of diffusion but it is used to visualize the path followed by molecules in the proximity of obstacles. The two approaches are complementary since they provide different information.

1.1 Theory/calculation

1.1.1 The purpose of the iMSD analysis—The purpose of the iMSD (image Mean Square Displacement) analysis is to determine the diffusion law in heterogeneous media. Commonly, the MSD is obtained in the single particle tracking from the analysis of a particle trajectory. Instead, for the iMSD, the MSD is obtained by a correlation function calculation without resolving the specific trajectory of a particle. The correlation function provides an estimation of the variance of a Gaussian describing the broadening of the spatial probability of finding a molecule in a given volume if it was at the center of the volume at time zero [1]. The Gaussian shape of the correlation function is a direct result of the Fick's second law for diffusion. In general, as molecules diffuse, they can be found at a distance from the origin that depends on the square of time, if the motion is pure Brownian diffusion. This is called the mean square displacement (MSD) which is a graph of the square of the displacement of a molecule as a function of time [3]. The slope of this graph is proportional to the diffusion coefficient. In the single particle tracking method, the MSD parameter is obtained in real space by tracking the position of particles one a time [3]. However, in heterogeneous media such as the cell interior, the distance from the center could be limited by the nature of the medium, barriers to diffusion, cavities and molecular interactions with fixed or mobile structures. Information about the mean-square displacement can be obtained

using spatial correlation functions in the correlation space rather than in the real space without the need to track individual particles [4]. Using image correlation methods we can calculate the spatiotemporal correlation function that shows a broadening as a function of time that can be related to a model of diffusion. Since this broadening is obtained using image correlation methods we call the parameter obtained by image correlation the iMSD. The broadening can in principle be different in different directions giving rise to diffusion anisotropy [5]. The intent of the iMSD analysis is to capture the heterogeneous broadening of the correlation function in different parts of the sample and at different times. If the motion is very fast, we must explore a large region around the origin to capture the displacement of the molecule. If the motion is very slow, the molecule (or particle) could not have moved or moved a very small quantity with respect to our sampling of the space around the molecule. For example, assume that a molecule has a diffusion coefficient of 100 $\mu m^2/s$, in 10^{-2}s, which is about the fastest frame time we can acquire with a EMCCD camera, the molecule will have moved within a Gaussian distribution with a variance of about $4\mu m^2$. Assuming that the camera pixel (of the image projected on the real camera pixel) is about 100nm, we will need a region of exploration of about 80 pixel-square to capture the broadening of the Gaussian to a factor of 2 in variance. At the other extreme, if the molecule has a diffusion coefficient of $0.001\mu m^2/s$, then the change of the variance of the Gaussian distribution after 1s will be only a fraction of a pixel ($10^{-5}\mu m$), so that we will need to collect data for a much longer delay times to see a sizeable increase of the variance. This concept is illustrated in figure 1. In figure 1A we show the broadening of the Gaussian (Fick's second law) which depends on time and on the diffusion coefficient. In figure 1B we show the calculations of the size of the ROI needed to detect a change of a value of 2 in the Gaussian variance as a function of the diffusion coefficient. The consideration of the size of the ROI is crucial for the proper calculation of the diffusion coefficient using the correlation function method. As shown in figure 1B, when the diffusion coefficient is large, we need to increase the region of analysis. If the diffusion coefficient is very small, we need to collect data for a longer time.

1.1.2 Algorithm and equations used for the iMSD analysis and data processing pipeline—The algorithm for the calculation of the iMSD is based on the STICS correlation function [4]

$$G(\xi,\chi,\tau)=\frac{\langle I(x,y,t)\cdot I(x+\xi,y+\chi,\tau+t)\rangle}{\langle I(x,y,t)\rangle^2}-1 \quad \text{(Eq 1)}$$

where the bracket indicates average over the x, y coordinates and time t. ξ and χ are the shift of the x and y coordinates and τ is the time delay. *I(x,y,t)* represents one image of an image stack of the same field of view acquired as a function of time. While this equation is valid for any type of images, in this article we deal with fluorescence intensity images. Assuming the shape of the illumination volume is Gaussian and the process of diffusion can be expressed by the Fick's second law, the resulting correlation function after the integration on the x, y and t coordinates is given by Eq 2 [1]

$$G(\xi,\chi,\tau) = \boldsymbol{g(\tau)} \cdot \exp\left(-\frac{\xi^2+\chi^2}{\boldsymbol{\sigma_r^2(\tau)}}\right) + \boldsymbol{g_\infty(\tau)} \quad \text{(Eq 2)}$$

The term $\boldsymbol{g(\tau)}$ represents the change of the amplitude of the correlation function and the term $\boldsymbol{\sigma^2_r(\tau)}$ represents the broadening as a function of time and position of the Gaussian correlation function.

1.1.2.1 Diffusion laws: The variance of the Gaussian is assumed to be composed of 3 terms as shown in equation EQ 3.

$$\sigma^2(\tau) = {\sigma_0}^2 + 4D_{macro}\tau + \frac{L^2}{3}(1 - e^{-k_{micro}\tau}) \quad \text{(Eq 3)}$$

The first term $\boldsymbol{\sigma^2_0}$ represents the Gaussian variance at time zero, the second term represents the macroscopic diffusion and the third term represents confined diffusion within a region of average length L. k_{micro} is the rate to achieve confinement and it has dimensions of inverse time.

The behavior of Eq 3 is illustrated in figure 4 where 3 common diffusion laws are considered [1]. In the case of pure diffusion, the value of L is zero and the variance changes linearly with the delay time τ. In the case of pure confinement, the rate to reach confinement is described by the k_{micro} term and the average size of the confinement is given by L. The term k_{micro} has the dimension of inverse time. The initial slope to reach confinement is given by the following expression.

$$D_{micro} = \frac{L^2}{12} * k_{micro} \quad \text{(Eq 4)}$$

which has the dimension of a diffusion coefficient and it is generally also called D_{micro}

1.1.3 The iMSD sprites—From the graphs of figure 1B, we can see that for a diffusion coefficient of 1 $\mu m^2/s$, the change of the Gaussian variance by a factor of 2 will occur in about 1s or 100 frames (green line in Figure 1B). Instead for a diffusion coefficient of $0.001 um^2/s$ it will require about 10,000 frames for the Gaussian to broaden by about a factor of 2. For this reason it is convenient to present the STICS correlation function using a log time axis. A sprite is a small image obtained calculating a series of correlation functions for a given ROI (typically 32×32 pixels or 16×16 pixels) using a log time axis with 32 time points equally spaced in a logarithmic time scale. The size of a sprite in terms of file size in bytes is, for example 32×32×32*(sizeoffloat32)=131072 bytes. Assume that a typical image was acquired in the format 256×256 pixels and that we calculate each iMSD sprite by moving the ROI of 32×32 pixels in such a way as to superimpose each ROI by at one half of the ROI. In each direction we will have (2*256/32)−1 sprites for a total of 15×15=225

sprites. If the frame size is 512×512 we will have 961 sprites, which is typical of our measurements (Figure 3).

1.1.4 Calculation time—Next we discuss the issue of the calculation time for each sprite. The computation of each sprite must be in the order of 10ms to give a reasonable waiting time for the 200 to 1000 calculations to complete. To calculate one sprite we use an ultrafast 3D FFT routine (C. Gohlke. "Pair Correlation Function Analysis of Fluorescence Fluctuations in Big Image Time Series using Python". Big Data Image Processing & Analysis (BigDIPA) Course, 19 September 2017, University of California, Irvine). A typical calculation involves a 3D FFT of size 32×32×32768 and it takes about 20ms (also taking into account the log averaging operation) for a total computational time of about 4–5 s for a 256×256 image stack and 16–20 s for a 512×512 image stack. This is a reasonable amount of time to wait for this calculation to be completed. The sprites are computed only once and then they are stored together with the original image stack. For a file of 256×256×32768 the total sprite file is approximately 28.8 MB starting with about 8GB of the original image stack, which is a very substantial decrease in file size. In the SimFCS program (available at www.lfd.uci.edu), the sprite files are stored in the same directory and with the name of the original image stack but with information added to the file name to indicated the size of the ROI and if the time was calculated in a log or linear scale.

A further reduction of the data size take place by fitting each of the planes of the correlation function using a 2D Gaussian with 7 parameters: the amplitude, zero offset, 2 central positions, 2 variances along 2 lab axis and the orientation of the 2D Gaussian (Figure 3, right panel). This reduces each plane of the sprite (32×32) to 7 parameters per each of the 32 time delays (Figure 3, left panel). The resulting file is about 200Kb and it is stored again in the same directory and with the same name but the end part of the name is changed. There are options in SimFCS to perform the fit of the correlation function with lesser parameters, but all parameters are stored in a vector of 8 values anyway.

1.1.5 Law of diffusion—Once the fit parameters are calculated, it is possible to use these parameters to find the “law of diffusion”. For example the inverse of the G(0) and the 2D Gaussian variance can be used to fit 3 models for diffusion at each ROI, linear or pure diffusion, confined diffusion or a mixture of confined diffusion and long range diffusion corresponding to the models of Figure 2. The models can be ranked according to the correlation parameter or the chi-square of the fit as it will be shown in the Results section. The parameters of the fit are then stored in a fit-file together with other information about the data set and the images for each of the parameters as it varies at every ROI. All images and maps can be saved in an Excel file for permanent storage and easy display.

1.2 The pair-correlation function analysis

As previous described by [6] for one dimension, we used the 2D pair-correlation function defined as:

$$pCF = G(\tau, r_0, r_1) = \frac{\langle F(t, r_0) \cdot F(t+\tau, r_1) \rangle}{\langle F(t, r_0) \cdot F(t, r_1) \rangle} - 1 . \quad \text{(Eq 5)}$$

where, **τ** is the time delay between acquisitions of the fluorescence intensity (*F*) at two points in the image (r_0 and r_1). The temporal average is indicated by the brackets. The 2D-pCF (Eq. 5 and Figure 7B) is able to detect barriers or obstacles to diffusion by the delay time on the correlation maximum in the presence of diffusion barriers or the absence of correlation between two points. The 2D-pCF approach is intended to measure the average path for a molecule diffusing in the cell with statistical spatio-temporal significance. To achieve this goal, we start with a stack of at least 8192 images of the same focal plane in the cell. From this stack we calculate the pCF at a given distance (δr) and at different angles around a given point for all the pixels in the image. A schematic illustration for an expected 2D-pCF with heterogeneous diffusion is represented in the Figure 4B. To analyze the 2D-pCF for each pixel in the image we use common image processing methods to obtain parameters of the angular distribution of the pCF [2]. We calculate first and second central moments of the spatial distribution. From these two parameters it is possible to obtain the angle of the 2D-pCF distribution (θ) and long (λ_1) and short (λ_2) axis using EQs 6–11. Equation 10 and 11 define the eccentricity and anisotropy, respectively, for the 2D-pCF and this parameter will be used to produce the connectivity maps.

$$M_{ij} = \frac{\sum_x \sum_y x^i y^j I(x,y)}{\sum_x \sum_y I(x,y)} \quad \text{(Eq 6)}$$

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q I(x,y) \quad \text{(Eq 7)}$$

$$\theta = \frac{1}{2} \arctan\left(\frac{2\mu_{11}}{\mu_{20} - \mu_{02}}\right) \quad \text{(Eq 8)}$$

$$\lambda_i = \frac{\mu_{20} + \mu_{02}}{2} \pm \frac{\sqrt{4\mu_{11}^2 + (\mu_{20} - \mu_{02})^2}}{2} \quad \text{(Eq 9)}$$

$$Eccentricity = \sqrt{1 - \frac{\lambda_2}{\lambda_1}} \quad \text{(Eq 10)}$$

$$Anisotropy = \frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2} \quad \text{(Eq 11)}$$

I(*x,y*) in EQ. 6 is the value of the 2D-pCF correlation function at pixels shifts *x* and *y*. *Mij* is the definition of the moments of the distribution of the 2D-pCF. The shift of the center of mass of the distribution (the μ parameters at the first order, p=q=1) can be used to obtain the net velocity (or the center of mass shift) of the particle. From EQ. 7, the displacement of the center of mass is expressed in the unit of pixels of the image. We use the ratio (or normalized difference) of the long and short axis of the ellipses (λ_1 and λ_2, respectively) to construct a map indicating where the pCF distribution is expanding unevenly, indicating the average local direction of the diffusional motion.

2. Material and Methods

2.1 Material

All media and reagent for cell culture were acquired at Invitrogen by Thermo Fisher Scientific Inc. (Waltham, MA-USA). All chemicals and solvents used were high-grade quality and acquired from Sigma-Aldrich (St. Louis, MO-USA). Plasmids were acquired from Addgene (Cambridge, MA-USA)[7].

2.2 Cell culture

Chinese Hamster Ovary (CHO-K1) cells were cultured at 37°C in media containing Dulbecco's Modified Eagle medium (DMEM)/ F-12 Nutrient mixture supplemented with 10% Fetal Bovine Serum (FBS) and 1% Penstrep, a 5% $C0_2$ atmosphere was used. Cells were plated on a 35 mm dish previously coated with fibronectin (Invitrogen, Thermo Fisher Scientific Inc., Waltham, MA-USA) and transfected using Lipofectemine® 2000 and according to manufactures instructions (Invitrogen, Thermo Fisher Scientific Inc., Waltham, MA-USA). The experiments were carried out at 37°C by a thermostat stage (Tokai Hit Co., Ltd., Fujinomiya-shi, Shizuoka-Japan) and 5% $C0_2$ atmosphere was included.

2.3 TIRF instrumentation

The measurements were done on a commercial Olympus TIRF microscope (Olympus, USA), for the EGFR-EGFP excitation a 488 nm line of the argon ion laser was used. In the emission path a filter set for EGFP a HQ500/20 and 515 nm long pass filter split by a Q515lp dichroic mirror (Chroma Technology Corporation., Bellows Falls, VT-USA) was used. The signal was collected using an Olympus TIRF UIS2 PlanApo N 60× (NA = 1.45) oil-immersion objective and the fluorescence fluctuation was recorded by an EMCCD camera 512B Cascade (Photometrics, Tucson, AZ, USA). The frame rate was 20 ms and the image size 256×256 pixels with a pixel size of 139 nm. For both methods (iMSD and 2D-pCF) we needed to acquire at least 8192 frames which correspond to about 80 seconds of data acquisition. Cells were CHO-K1 cells transiently transfected with EGFR-EGFP plasmid (#32751) from Addgene (Cambridge,MA)[7] All date sets were detrended for bleaching using the exponential detrend function of the SimFCS Software. Bleaching affects the

amplitude of the correlation function but much less the variance. However, in the specific data set used for the examples of this paper to total bleaching was about 10% from the initial to the final frame.

3. Results

3.1 iMSD laws of diffusion

For this illustration of the iMSD method we analyze a time stack of 8192 frame taken with the EMCCD camera at a rate of 50 frames/sec of a cell expressing the EGFR receptor in the TIRF microscope. The actual record is much longer but we analyze here only a small fraction since the cell was slowly moving to the right. However the movement during the first 80s was less than one pixel. The same data was also used for the 2D-pCF illustration (see next session in the Results section). For the sample, the EGFR receptor moves slowly so that we use a small ROI of exploration (16×16 pixels =2.22 μm-square) a total of 961 sprite. All pixel of intensity less than 1000 were disregarded for this analysis giving 264 sprites.

3.1.1. Linear model for diffusion—The data were first analyzed using the linear model for diffusion (Figure 4);

The EGFR receptor has a relatively small diffusion coefficient. Most of the values in the histogram are in the 0.01–0.1 $\mu m^2/s$ range which is typical for transmembrane receptors. In the histogram of the σ_0^2 parameter most of the occurrences are larger than the value of the intercept expected for the sole effect of the PSF. The PSF is about 0.3μm so that the square should be $0.09\mu m^2$. This finding indicates extensive clustering of the receptor.

3.1.2. Confined model for diffusion—The correlation coefficient histogram indicates that most of the fits are “reasonable” (Figure 5E). We use a threshold value of 0.7 of the correlation coefficient to accept all fits above that “reasonable” value. The histogram of the σ_0^2 parameter indicates again extensive clustering of the receptor. The histogram of the confinement parameter L, which is prevalent below 1μm, is in the range expected for a transmembrane receptor. The D_{micro} parameter is also very small in the same range of the D_{macro} found using the linear model. This result could indicate the in many sprites the fit is not reflecting the real model. For this reason we need a criterion to compare the models which is used next.

3.1.3. The “all models” for diffusion—Then we analyzed the data for the “all models”. In the “all models” the three models for diffusion are tested in sequence (linear model, confined model and partially confined, Eq 3). The correlation coefficient for the fit is calculated at each sprite. Then the models are ranked according to the correlation coefficient. The model with the highest correlation coefficient is then recorded in a table (Figure 6) if the correlation coefficient is larger than 0.7, otherwise the sprite is disregarded. The 3 models tested are in reference to Eq 3. Model 1 only considers the first and second terms in Eq 3 and it is called linear model. Model 2 considers term 1 and term 3 of Eq 3 and is called the confined model. Model 3, also called “all” corresponds to a partially confined model. For this data set, the linear model has a greater prevalence, with the confined model found in fewer sprites and the partially confined model found only in some sprites as indicated in the

map. For each model, the map of the parameters of the fit is shown. The maps are complementary because if the best ranking is found in one sprite it cannot be found for another model in the same sprite. The models have different type and number of parameters. Figure 6J shows the histogram of D_{micro} only for the sprites with best fit for the confined model. Comparing this panel with Figure 5H one can observe that the histogram of D_{micro} is now different indicating that forcing a specific model for the fit can give erroneous results. Instead ranking all the models according to the correlation parameter for the fit provides the best model fit for a given sprite.

The ranking method shows that most of the sprites are best fitted by the linear model and that forcing the fit to provide the parameters for the confined model as was done for in the previous subsection could lead to erroneous interpretations. For example, in the few sprites where the confined model fit the best, the D_{micro} parameter is much higher than the D micro found in subsection 3.1.2, as it should be expected (Figure 6J).

3.2 Diffusion and connectivity of EGFR by the 2D-pCF

Here we analyze by the 2D-pCF approach the same data set used for the illustration of the iMSD algorithm in section 3.1. Figure 7 shows the map for the anisotropy defined in EQ. 11 obtained for pCF at a distance of 4 and 8 pixels by the 2D-pCF analysis presented in section 1.2. EGFR in CHO-K1 cells shows clusters of high diffusion anisotropy (Figure 7C and D). The direction of the prevalent diffusion can be found by the map of the angles of the diffusion anisotropy (see Figure 7E and F). By calculating the pCF at increasing distances (4 and 8) we can identify the occurrence of different features in Figures 7C and F. The anisotropy histogram shows also a change in the profile of values at the two distances (Figure 7G and H). These results mean that the maps for the anisotropy are modified by the pCF distance calculation, thus associated the anisotropy with obstacles detected at different distances.

Hence, if we focus on the anisotropy direction (or angle) for pCF(8) and we split the anisotropy for range 0.15–0.3, 0.3–0.5, 0.5–0.7 and 0.7–1.0 and by fitting the histogram of the anisotropy directions using multi-Gaussian we can identify the main component of the diffusion in the cells (Figure 8). Anisotropies below 0.15 were disregarded because this is the anisotropy range expected for random diffusion. The 2D-histogram for the anisotropy vs anisotropy direction shows the occurrence of at least 2 maximum angles for anisotropies ranged 0.3–0.8 (Figure 8A). Then the fitted results were plotted in a polar plot and the main angle for the major axis of the cell is represented also on top of the plot (Figure 8B), therefore, it is possible to see the distribution of angles for the anisotropy with respect to the main cell axis. From the amplitude of the angle histogram, we can distinguish one main angle directions aligned with the cell axis and one more in the perpendicular direction.

The pCF analysis also provide an additional variable particular sensitive to the occurrence of barriers, which is the center of mass shift (CMS, see insert in Figure 9A for its definition). The high resolution images for the CMS for pCF at 4 and 8 pixel distances are shown in Figure 9. These regions of high CMS are associated with the displacement of the first central moment of the 2D-pCF distribution which occurs close to obstacles, shown in the inset of

Figure 9A. It is also possible so see that the features observed at pCF(4) are different for pCF(8).

3.2.1 The connectivity map and its meaning—The connectivity map is built by thresholding the anisotropy histogram excluding the values below 0.15, which are associated with isotropic diffusion. Then a segment at each pixel of length proportional to the value of the anisotropy and at the angle of the anisotropy direction. In Figure 10, we show the connectivity maps for pCF(8) using the same ranges for the anisotropy values defined in Figure 8. First column (A) in Figure 10 shows the connectivity maps for the four thresholds (0.15–0.3, 0.3–0.5, 0.5–0.7 and 0.7–1.0) for the anisotropy (left columns). For each anisotropy range it is possible to identify different sizes and directions in the clusters of anisotropy values. These results are even clearer when the angle for each vector is color-coded (see the second column in Figure 10B). By segmentation and masking it was possible to quantify different shape descriptors for the clusters within anisotropy ranges as describe above (column 3 in Figure 10C and 11). The shape descriptors selected were: Area (size of clusters) and Circularity (determined as 4π.area/perimeter^2, perfect circle = 1) using the ImageJ routine “Analyze particles”.

The analysis of the anisotropy map using increasing anisotropy thresholds selects clusters with increasing circularity (see Figure 11). This is not obvious and indicates that the connectivity map is not simply the results of a random distribution. The histogram for cluster sizes does not show significant changes (evaluated by the area) by the different anisotropy thresholds (Figure 11A). However, the shape descriptor for the clusters shows the differences in clusters shape by anisotropy thresholds (Figure 11B), thus illustrating the ability of the 2D-pCF approach to detect the diffusion path for a molecule at high resolution in space and time in the cell.

4. Discussion

Biological processes studied by image correlation analysis

We show that the law of diffusion can be obtained in ROIs of about 2 μm side. This is due to the iMSD method that is based on a spatial average. We schematically show in figure 1 that the ROI must have a minimum size in order to observe the broadening of the correlation function or in the case of very slow diffusion the correlation must be calculated at very long time delays. In order to better define the law of diffusion we need to compare the fits of the correlation function using different models. This is done in our software for the law of diffusion. The ranking of the model can be shown at each sprite and the chi-square of the fits can be compared. For the 2D-pCF we show that we can map barriers for diffusion which are not visible and that cannot be obtained with the iMSD method.

Necessary hardware

The iMSD and the pCF method work with images taken with a camera. However, several of the new confocal microscopes can acquire images at a speed comparable with EMCCD cameras in relatively small frames. If the image is acquired very fast in the confocal mode, it could appear as images acquired with a camera. However the interpretation of the intercept

parameter σ_0^2 must be re-evaluated. For the example shown in this paper, since the EGFR receptor diffuse relatively slowly, the TIRF microscope with a EMCCD camera provides sufficient resolution and speed to observe the diffusion of the receptor. Large frame sCMOS cameras could be used although the noise of these cameras must be accounted for using de-noising procedures. Otherwise, we must consider the 2D-pCF approach that has a larger dynamic range and better spatial resolution.

Computational power and size of the data set

The ideal data set should have good spatial resolution and the time stack should have at least 10,000 frames. The average file size for a single measurement is of several GB. Clearly, the computer must be able to maintain in core memory the data needed for the computation. We have been using laptops with a minimum of 8GB of core memory and at least 4 processors to use independent threads for the computation. Since the iMSD and pCF algorithms are based on fast Fourier transforms in 1D or 3D, the speed of computation is very high but we have to use double or quadrupole precision to compute FFT's with one of the axis on the order of 32,000 elements or larger sizes. The sprite method discussed in this article allows very large compression of the original stack. After the first computation of the correlation function the size of the data set is reduced by more than a factor of 100 and subsequent calculations of the fit of the correlation function using few parameters can reduce to data set size another factor of 100. This enormous compression without losing information makes the calculation fast and affordable to be further analyzed for different physical models.

Methods for data visualization

Visualization of data obtained in every pixel or small region of a cell is usually done under the form of maps, with the purpose of maintaining the spatial location of the original fluorescence image so that features of the original image can be associated with the property measured. We have used this visualization strategy for all data obtained by the iMSD and by the 2D-pCF techniques as shown in this article. However, for example diffusion coefficients are tensors which require more than one value to be displayed. Other quantities can be visualized using only one value and some of the maps do not refer to a specific physical quantity, like the map of the chi-square or the map of the best model. Visualization using maps requires the color palette to be specified and the correspondence between color and value to be given. In this work, for each map we provide a histogram of all values found in the map and we use the same color scale for all images. The color scale is reported for one histogram. The vector or tensor maps have associated two or more images. One way to visualize a vector is to use an arrow that indicate the direction and the modulus of the quantity is indicated by the length or color of the arrow. If this vector quantity is obtained at each pixel, it is not possible to have an arrow at every pixel. In this case we use two color-coded maps, one to display the modulus and the other the direction.

Using this visualization it is easy to recognize regions that have similar moduli and regions that show the same direction. The visualization methods provided by our software are intended to be the first step for a more detailed analysis. For example, the map of the σ^2_0 above a given threshold can be associated to clusters. The map (like all the maps in the

software) can be exported to be analyzed in other programs, for example in ImageJ, to obtain the number, size, shape, area and other parameters of the clusters.

Other approaches for image correlation analysis

There are been other approaches different from the iMSD and the 2D pCF techniques discussed in this article. Particularly important are the methods described by the Wohland lab [8–10] which also provide the law of diffusion at each point in an image and with relatively high resolution. In this approach, the law of diffusion is obtained by integrating the intensity at increasing areas around a pixel and calculating the time correlation function as a function of the area. This method is quite robust but similarly to the iMSD method it is limited in the spatial resolution by the need to integrating over increasing areas. A different method was proposed earlier based on the k-space analysis by the Wiseman lab [11].

5. Conclusions

We have presented some elements of the theory behind the iMSD and 2D-pCF approaches to image correlation spectroscopy as well as the algorithms and the software for data visualization. The two methods offer complementary information about the diffusion of molecules in live cells. The iMSD provides the law of diffusion in a certain ROI. In the iMSD method, the spatial resolution is limited by the size of the ROI since we cannot determine which part of the ROI has contributed to the overall results. We show that the size of the ROI is in part dictated by the rate of diffusion (Fig 1). Specifically, faster diffusion needs to be observed in a large ROI for a given frame rate. For slow diffusion, we need to collect data for a long time to observe a broadening of the correlation function. The 2D-pCF approach is not based on spatial averaging so that very high spatial resolution could be obtained. This is particularly important for the visualization of barriers to diffusion such as internal membranes or organelles that limit the space for local diffusion.

The algorithms used for iMSD and 2D-pCF have been published and also a program is available that implements the calculation and visualization shown in this article. As a final note, fluctuation analysis should not be used for fixed samples where fluctuations are absent.

We envision that the methods based on correlation functions will find a broad application in cases in which single particles cannot be isolated. The 2D-pCF has the highest resolution since it depends on the fluctuations in a pair of pixels. We are developing the 2D-pCF for the Zeiss Airy detector where the pixel waist is effectively in the range of 120nm, at the detector center. The methods iMSD and 2D-pCF techniques described in this article can also be used on line scans which can be very fast in the STED microscope.

Acknowledgments

The authors would like to thanks to Milka Stakic for the cells preparation and transfection. This work was supported by grants Grants NIH P41-GM103540 and NIH P50-GM076516. LM is supported for the Universidad de la República-Uruguay as a full time professor. The development of the algorithms and the routines for the SimFCS software were done during the training period of Miss Estella Rao at the Laboratory for Fluorescence Dynamics during 2016

Abbreviations

2D-pCF	two dimensional pair-Correlation Function
iMSD	image Mean Square Displacement
EGFR	epidermal growth factor receptor

References

1. Di Rienzo C, et al. Fast spatiotemporal correlation spectroscopy to determine protein lateral diffusion laws in live cell membranes. Proc Natl Acad Sci U S A. 2013; 110(30):12307–12. [PubMed: 23836651]
2. Leonel Malacrida PNH, Ranjit Suman, Cardarelli Francesco, Gratton Enrico. Visualization of barriers and obstacles to molecular diffusion in live cells by spatial pair-cross-correlation in two dimensions. Biomedical Optics Express. 2018; 9(1):303–321. [PubMed: 29359105]
3. Michalet X. Mean square displacement analysis of single-particle trajectories with localization error: Brownian motion in an isotropic medium. Phys Rev E Stat Nonlin Soft Matter Phys. 2010; 82(4 Pt 1):041914. [PubMed: 21230320]
4. Hebert B, Costantino S, Wiseman PW. Spatiotemporal image correlation spectroscopy (STICS) theory, verification, and application to protein velocity mapping in living CHO cells. Biophys J. 2005; 88(5):3601–14. [PubMed: 15722439]
5. Di Rienzo C, et al. Diffusion Tensor Analysis by Two-Dimensional Pair Correlation of Fluorescence Fluctuations in Cells. Biophys J. 2016; 111(4):841–851. [PubMed: 27558727]
6. Digman MA, Gratton E. Imaging barriers to diffusion by pair correlation functions. Biophys J. 2009; 97(2):665–73. [PubMed: 19619481]
7. Carter BZ, et al. gamma-glutamyl leukotrienase, a gamma-glutamyl transpeptidase gene family member, is expressed primarily in spleen. J Biol Chem. 1998; 273(43):28277–85. [PubMed: 9774450]
8. Sankaran J, et al. Diffusion, Transport, and Cell Membrane Organization Investigated by Imaging Fluorescence Cross-Correlation Spectroscopy. Biophysical Journal. 2009; 97(9):2630–2639. [PubMed: 19883607]
9. Wohland T, et al. Single Plane Illumination Fluorescence Correlation Spectroscopy (SPIM-FCS) probes inhomogeneous three-dimensional environments. Optics Express. 2010; 18(10):10627–10641. [PubMed: 20588915]
10. Singh AP, et al. SPIM-FCCS: A Novel Technique to Quantitate Protein-Protein Interaction in Live Cells. Biophysical Journal. 2013; 104(2):61a–61a.
11. Kolin DL, Ronis D, Wiseman PW. k-Space image correlation spectroscopy: A method for accurate transport measurements independent of fluorophore photophysics. Biophysical Journal. 2006; 91(8):3061–3075. [PubMed: 16861272]

Highlights

- The iMSD fluctuation correlation analysis method provides the mean square displacement of molecules in a crowded environment without the need to track individual molecules.
- iMSD requires averaging over small regions of the cell depending on the value of the diffusion coefficient, limiting the spatial resolution.
- The 2D-pCF analysis shows obstacles and barriers to diffusion and emphasizes the anisotropy of the diffusion flow in live cells.
- 2D-pCF analysis does not require spatial averaging and the position of barriers to diffusion can be obtained with pixel resolution

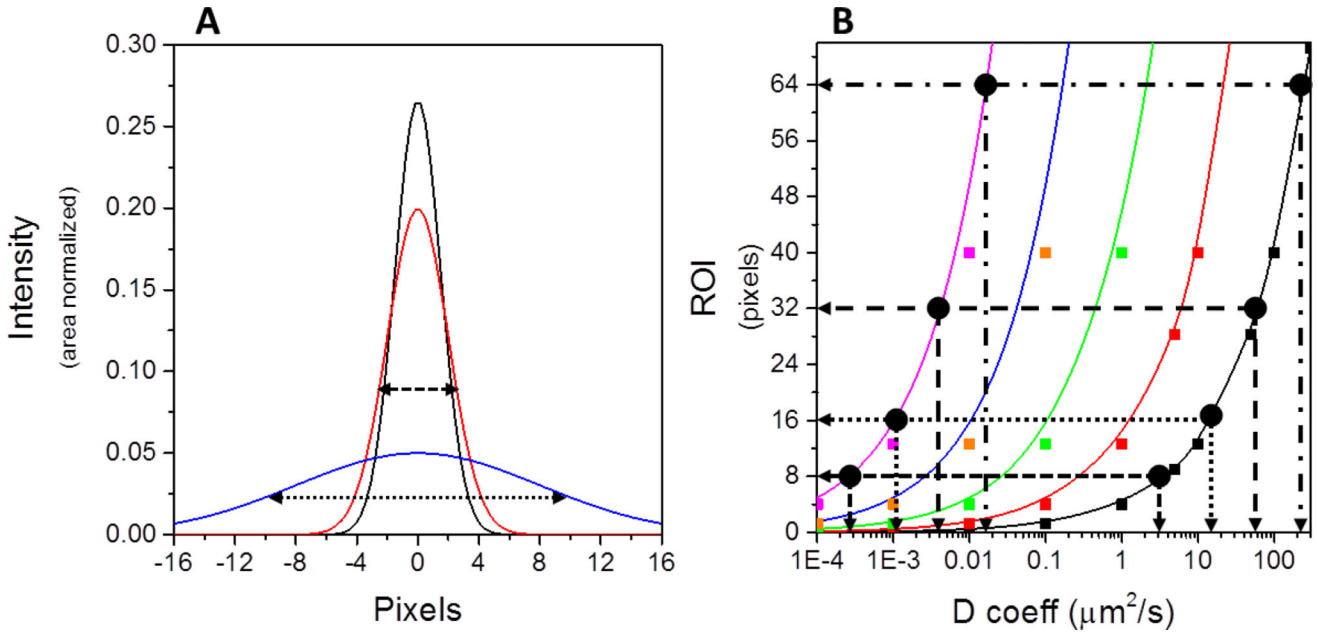


Figure 1.

A) Simulation of a Gaussian distribution for with a width of 3 pixels (black curve). The blue curve represents the spreading as a function of time of the original Gaussian function (black curve) by diffusion of a molecule with large diffusion coefficient. The dotted double arrow line represents FWHM for a fast diffusion (16 pixels wide). The red curve represents the spreading of the original Gaussian function (black curve) by diffusion of a molecule with slow diffusion coefficient. The spreading for the slow diffusion is almost unperceivable but the amplitude significantly decreases in the case of slow diffusion. B) Plot for the ROI (in pixels) in a camera-based image (pixel size 100 nm) needed to observe a change of a factor of 2 in the FWHM as a function of diffusion coefficient. From top to bottom the dashed/dotted lines indicate the measurable range in the diffusion coefficient (from 400 $\mu m^2/s$ to 3×10^{-4} $\mu m^2/s$) for a ROI (from 64 to 8 pixels) used in the iMSD analysis. From right to left (black to purple curves) the delay is increasing from 0.01 s black single frame, 0.1s red, 1s green 10s blue, 100s purple (delay after 10,000 frames).

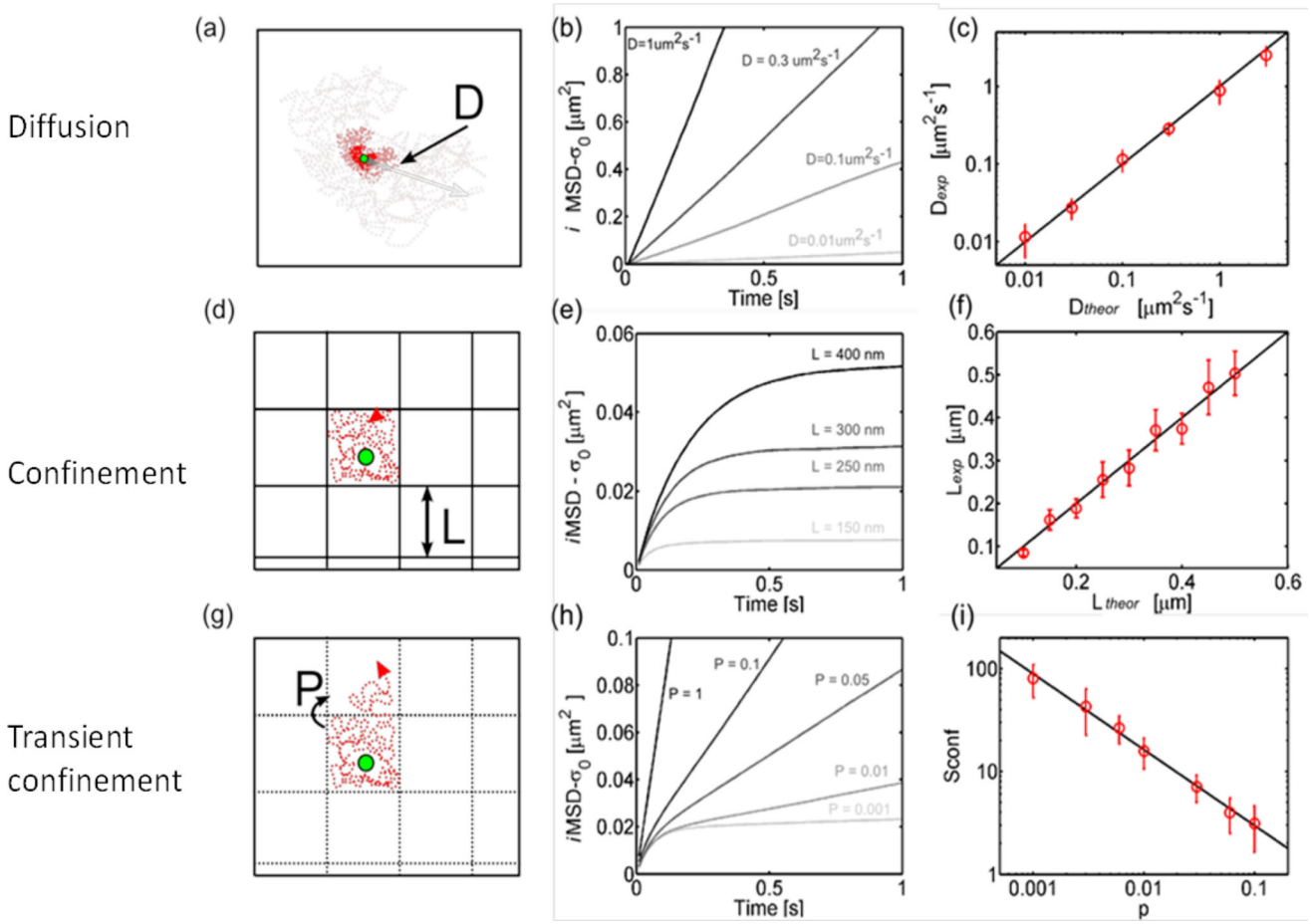


Figure 2.

iMSD analysis on simulated 2D diffusion. (a) Simulated condition: 2D isotropic diffusion, with diffusivity D. (b) iMSD is linear, with a higher slope for increasing D values. (c) Accordance between the theoretical D value and that recovered from the analysis. (d) Simulated condition: 2D isotropic diffusion in a meshwork of impenetrable barriers (probability P = 0 to overcome the barrier). (e) iMSD plot starts linear and then reaches a plateau that identifies the confinement area and the corresponding linear size L. (f) Accordance between the theoretical L value and that recovered from the analysis. (g) Simulated condition: 2D isotropic diffusion in a meshwork of penetrable barriers. Particles have probability P > 0 to overcome the barrier, thus generating a hop diffusion component. (h) iMSD plot starts linear (with a slope dependent on D_{micro}) and then deviates toward a lower slope which depends on P. (i) Calculated S_{conf} as a function of the imposed P. Part of this figure was previously published in [1].

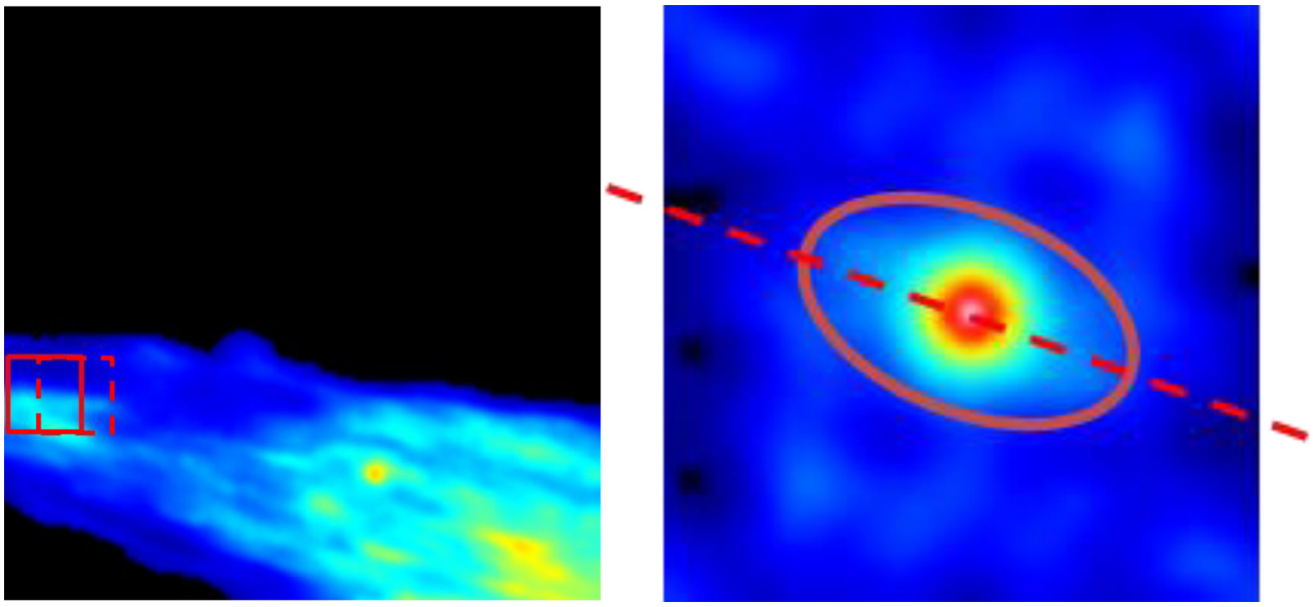


Figure 3.

Left: Part of an image of size 256×256 is shown (pixel size is 0.139 μm). The red solid square indicates an ROI of 16×16 pixels (2.22 μm square) and the red dashed square indicate to movement by 8 pixels in the x direction where the next sprite will be calculated. The sprites are calculated at each ROI and there is superposition between adjacent sprites. Right: example of one spatial correlation function at one sprite fitted with a 2D Gaussian function tilted. The red ellipse indicates the contour at half the amplitude. The size of the sprite is the same as that of the ROI.

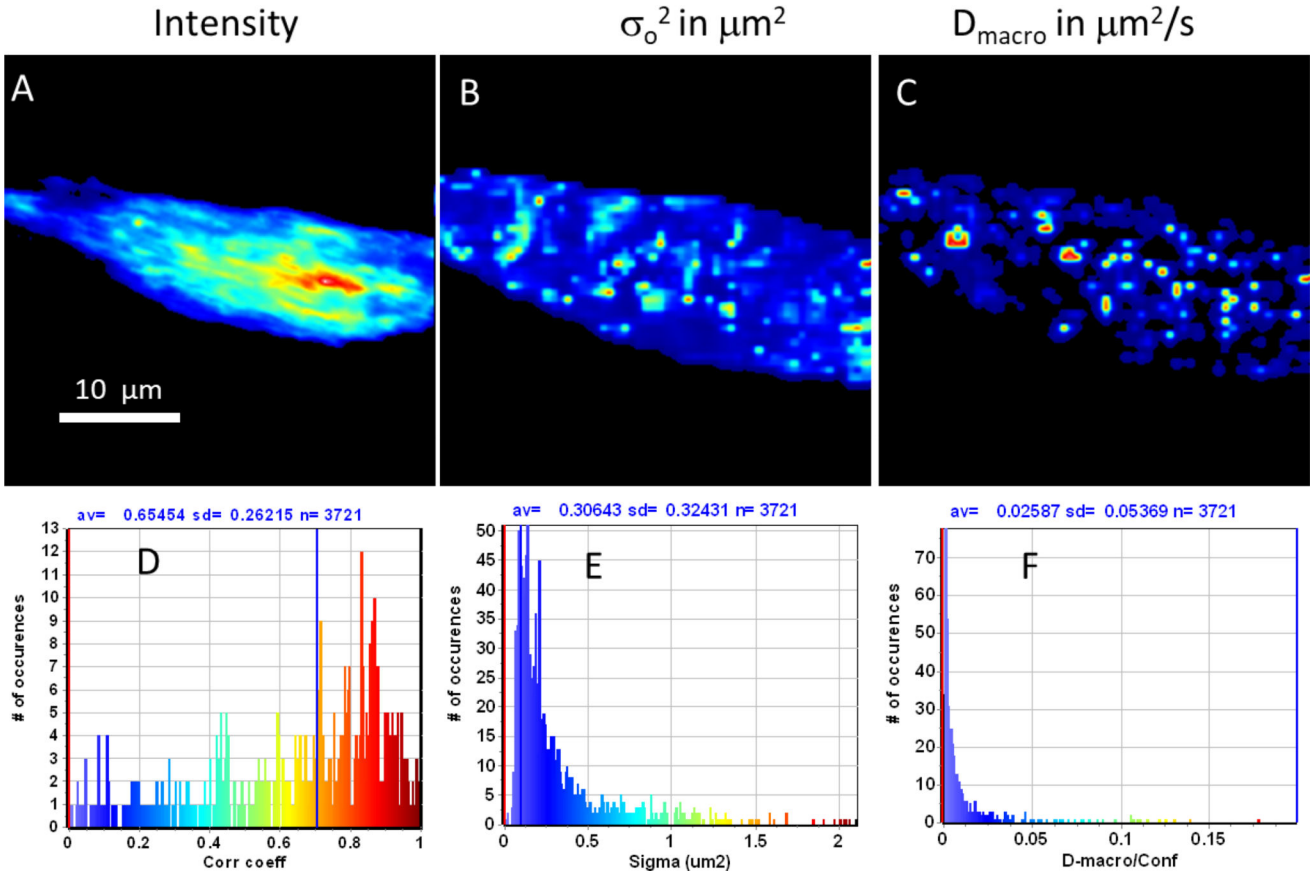


Figure 4.

iMSD analysis for linear model of diffusion. A) Intensity image. B) Map of the intercept parameter σ_0^2 as defined in EQ. 3. C) Map of the D_{macro} parameter defined in EQ. 3. D) Histogram of the correlation coefficient of the fit. The blue line is at 0.7. Values below 0.7 were not used to evaluate the model. E) Histogram of the σ_0^2 parameter. The black vertical line indicates the value expected if the intercept is due to the PSF. F) Histogram of the D_{macro} parameter. The color code in the histograms is the same color code used for the maps.

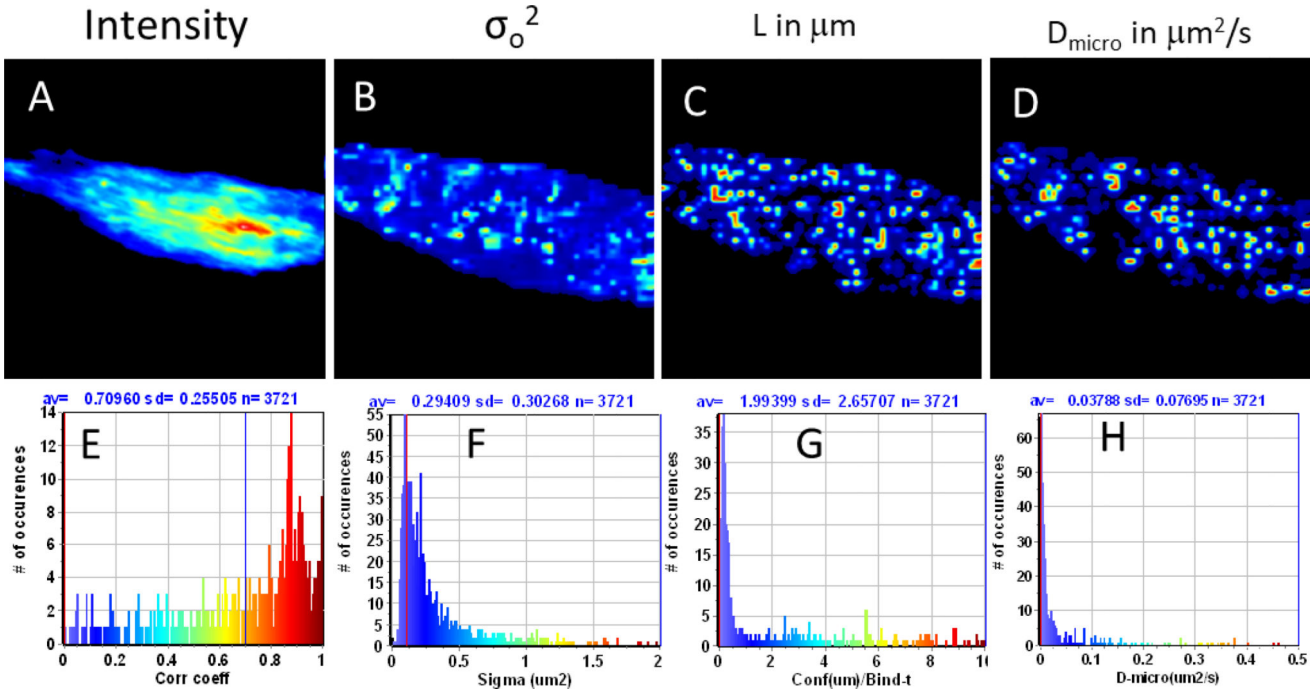


Figure 5.

iMSD analysis for confined model of diffusion. A) Intensity image. B) Map of the intercept parameter σ_0^2 as defined in EQ. 3. C) Map of the L parameter of confinement size as defined in EQ. 3. D) Map of the initial slope also called D_{micro} parameter as defined in EQ. 4. E) Histogram of the correlation coefficient of the confined model fit. Only sprites with values above the blue line in the histogram are selected. F) Histogram of the intercept parameter. The blue vertical line indicates the value expected if the intercept is due to the PSF. G) Histogram of the confinement length. H) Histogram of the initial slope also called D_{micro} parameter. The color code in the histograms is the same color code used for the maps.

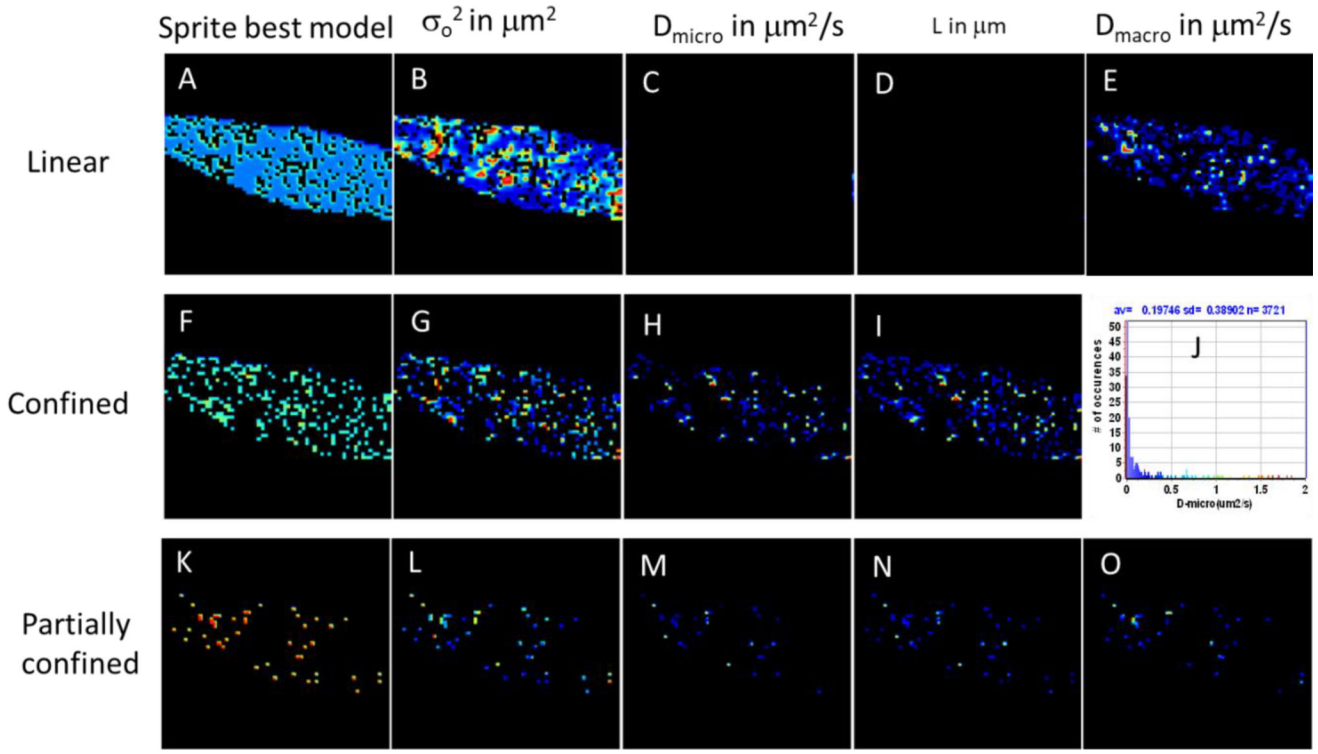


Figure 6.

Comparison and ranking of model for diffusion at each sprite. The first column labeled "Sprite best model" show where in the cell one of the 3 models ranks the best. The 3 models tested are in reference to Eq 3. Model one only considers the first and second terms in Eq3 and it is called linear model. Model 2 considers term 1 and term 3 of Eq 3 and is called the confined model. Model 3, also called "all" corresponds to a partially confined model. The linear model has a greater prevalence with the confined model found in fewer sprites and the partially confined model found only in some sprites as indicated in the map. For each model, the map of the parameters of the fit is shown. The maps are complementary because if the best ranking is found in one sprite it cannot be found for another model in the same sprite. The models have different type and number of parameters. Panel J show the histogram of D_{micro} only for the sprites with best fit for the confined model. Comparing this panel with panel H) of Figure 5 one can observe that the histogram of D_{micro} is now different indicating that forcing a specific model for the fit can give erroneous results. Instead ranking all the models according to the correlation parameter for the fit provides the best model fit for a given sprite.

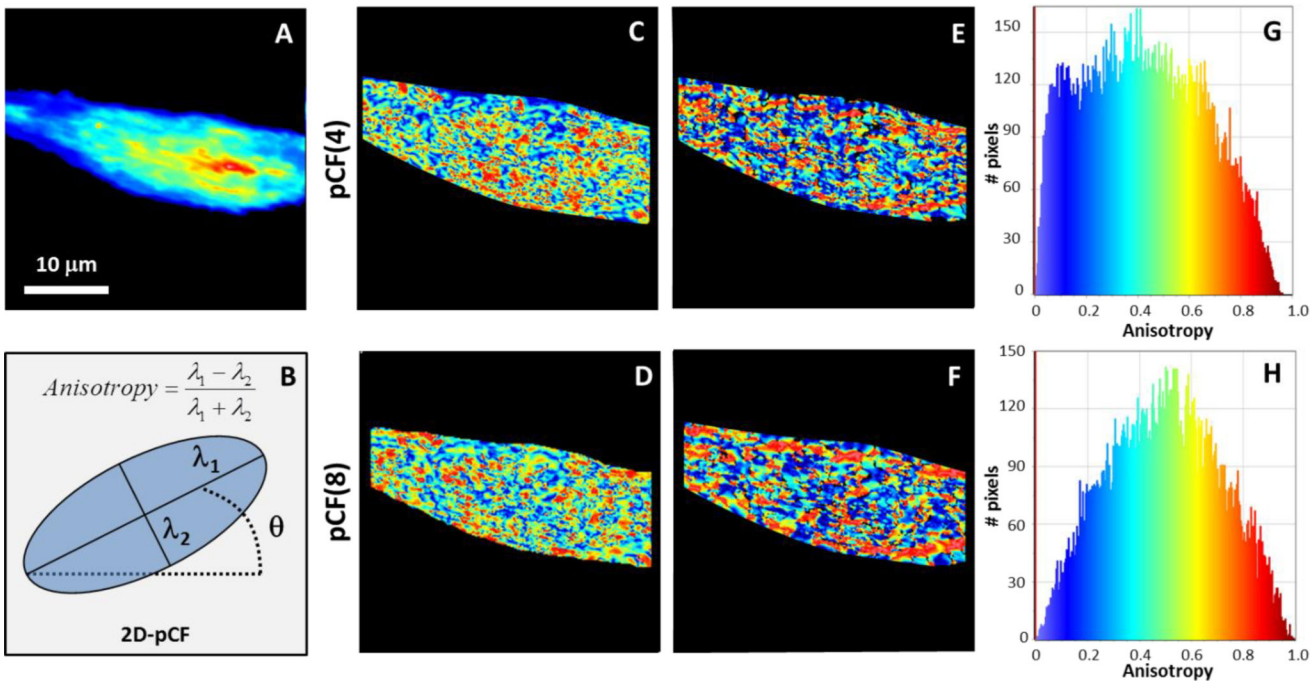


Figure 7. Analysis of the anisotropy diffusion of the EGFR on CHO-K1 cell based on 2D-pCF of TIRF data

A) Average intensity for the 8192 frames from EGFR-EGFP fluorescence used in the 2D-pCF analysis. B) 2D-pCF illustration (blue ellipse) and definition for the anisotropy calculation. C and D) Anisotropy of EGFR diffusion image for the pCF distance of 4 and 8, respectively. E and F) Angle of the anisotropy diffusion (anisotropy direction, θ) for the EGFR at pCF distance of 4 and 8, respectively. G and H) Histograms of the anisotropy distribution in the images C and D, respectively. The color scale in the histogram plots is the same color scheme used for the anisotropy map in the range 0–1 and for the anisotropy direction in the range 0–180 degrees (blue to red).

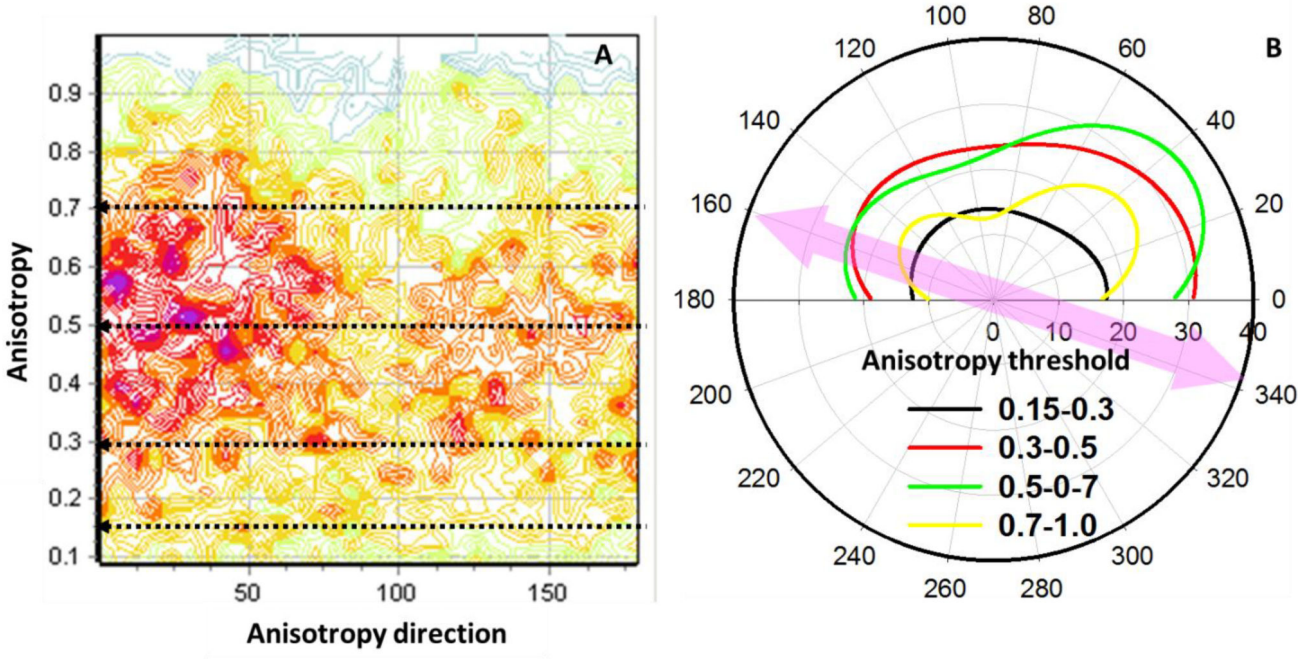


Figure 8. Analysis of the anisotropy direction for the pCF distance of 8

A) Two dimensional histogram of the anisotropy vs anisotropy directions (angle) for the pCF(8), Figure 7D and F. B) Polar histogram of the diffusion angle produced by the anisotropy direction histogram using thresholds for the anisotropy (the thresholds are indicated at the 2D-histograme in panel A). This plot is obtained by fitting multi-Gaussian components on the Cartesian histogram for the angles (not shown). The pink arrow represents the cell long axis orientation.

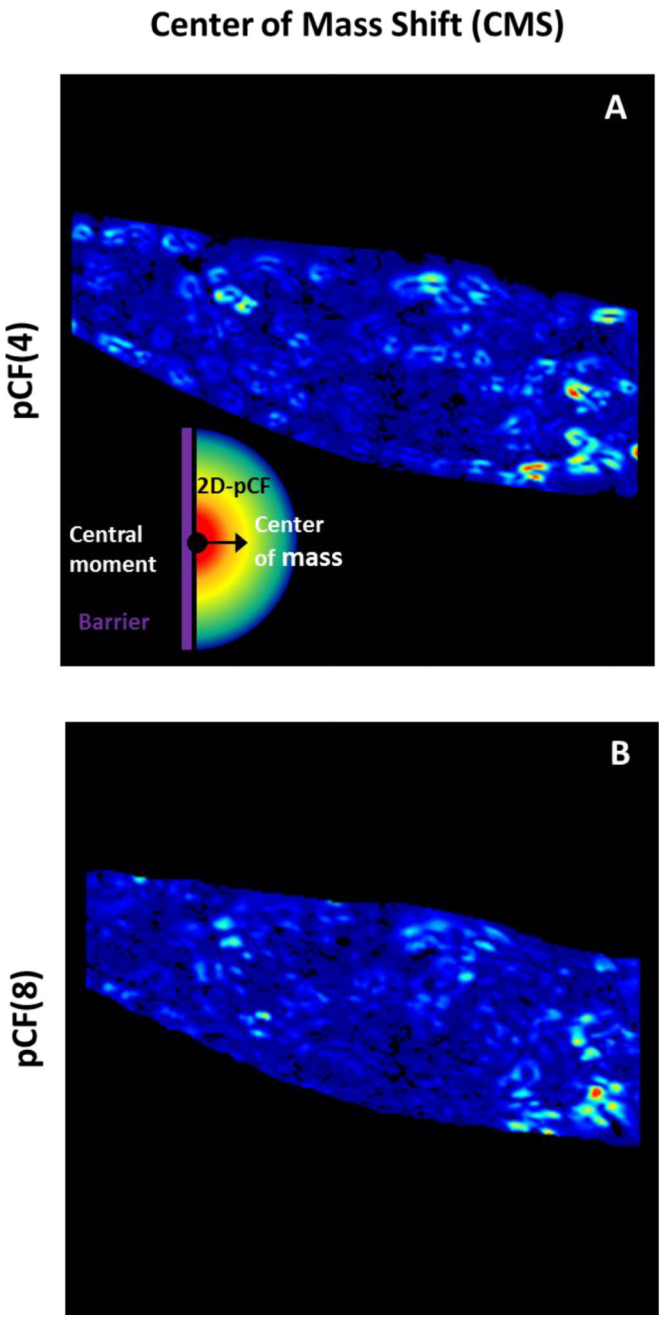


Figure 9. Center of mass shift (CMS) map for the EGFR diffusion in CHO-K1 cells

A and C) Images for the center of mass shift obtained for EGFR-EGFP on the plasma membrane of NIH3T3 cells transfected with the EGFR receptor for pCF distance of 4 and 8, respectively. Insert in figure A illustrate the concept of the CMS calculation.

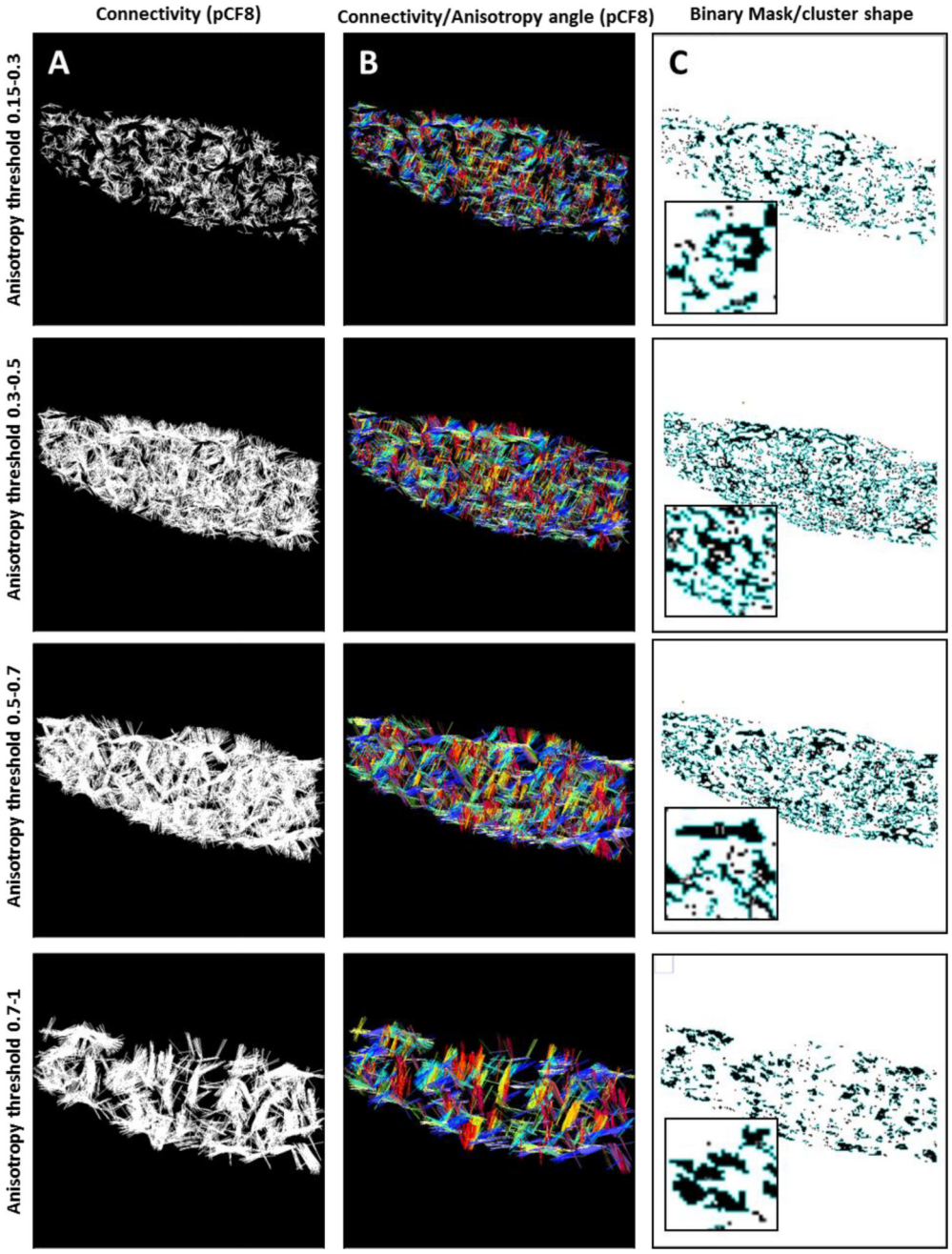


Figure 10. Connectivity maps and cluster shape analysis of EGFR-EGFP in CHO-K1 cells by pCF(8)

A) The left column shows the connectivity maps for 0.15–0.3, 0.3–0.5, 0.5–0.7 and 0.7–1.0 threshold in the anisotropy values, from top to bottom respectively. B) The connectivity maps are color-coded by the anisotropy angle used in the Figure 7. C) The connectivity clusters obtained by the anisotropy thresholds (0.15–0.3, 0.3–0.5, 0.5–0.7 and 0.7–1.0, from top to bottom respectively) were segmented and masked to be analyzed using the “Analyze Particles” routine in ImageJ for shape descriptors. The insert is a 3× zoom for a ROI on the corresponding map.

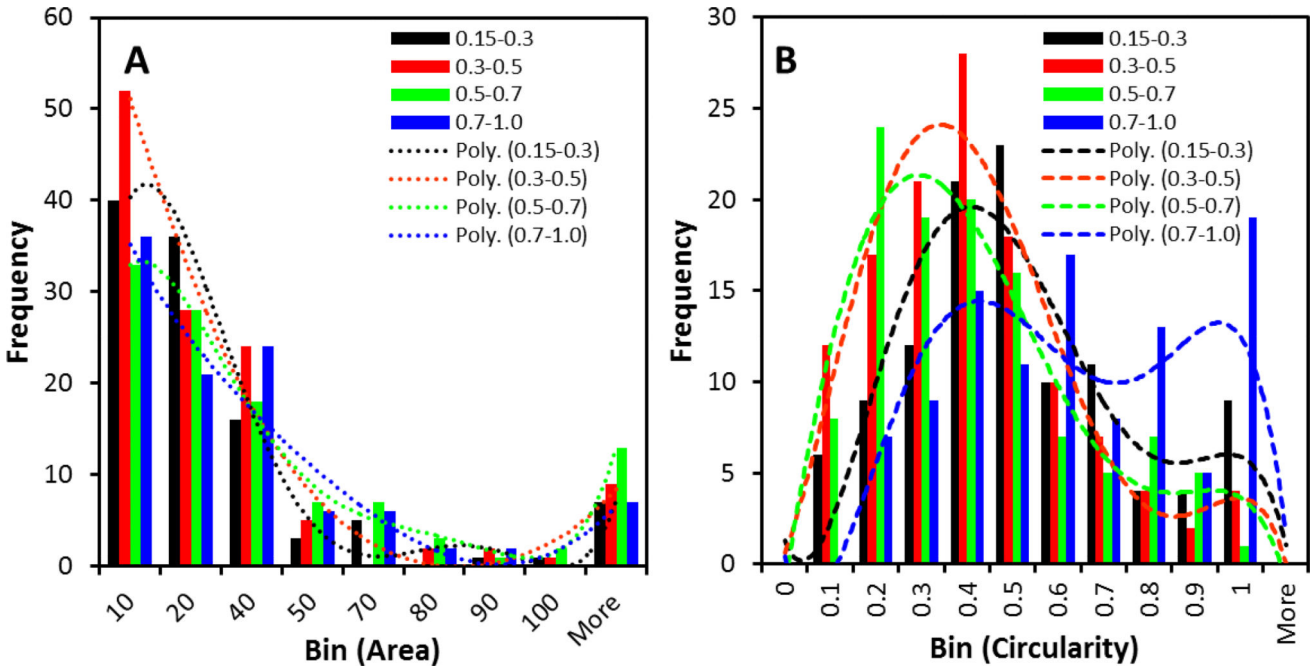


Figure 11.

A and B) Histograms produced by the shape descriptor of the connectivity clusters: Area and circularity, respectively. Dotted lines are polynomic function fitted to compare the histograms distributions.