

Nanoscale

Accepted Manuscript

This article can be cited before page numbers have been issued, to do this please use: A. M. Szalai, L. F. López, M. Á. Morales-Vásquez, F. D. Stefani and P. F. Aramendia, *Nanoscale*, 2020, DOI: 10.1039/C9NR10805J.



This is an Accepted Manuscript, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this Accepted Manuscript with the edited and formatted Advance Article as soon as it is available.

You can find more information about Accepted Manuscripts in the [Information for Authors](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the [Ethical guidelines](#) still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this Accepted Manuscript or any consequences arising from the use of any information it contains.

Analysis of sparse molecular distributions in fibrous arrangements based on the distance to the first neighbor in single molecule localization microscopy.

View Article Online
DOI: 10.1039/C9NR10805J

Alan M. Szalai^{1,2}, Lucía F. Lopez¹, Miguel Ángel Morales-Vásquez^{1,3}, Fernando D. Stefani^{*1,4},
Pedro F. Aramendía^{*1,2}

¹Centro de Investigaciones en Bionanociencias “Elizabeth Jares-Erijman” (CIBION), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Godoy Cruz 2390, C1425FQD Ciudad Autónoma de Buenos Aires, Argentina

²Departamento de Química Inorgánica, Analítica y Química Física, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Güiraldes 2620, C1428EHA Ciudad Autónoma de Buenos Aires, Argentina

³Materias Primas Medicinales (MAPRIMED, S.A); Av. Directorio 6155; C1440ATA, Ciudad Autónoma de Buenos Aires, Argentina.

⁴Departamento de Física, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Güiraldes 2620, C1428EHA Ciudad Autónoma de Buenos Aires, Argentina

* Corresponding authors. pedro.aramendia@cibion.conicet.gov.ar;
fernando.stefani@cibion.conicet.gov.ar

Abstract

Single Molecule Localization Microscopy (SMLM) currently attains lateral resolution around 10 nm approaching the molecular size. Together with increasingly specific fluorescent labeling it opens the possibility to quantitatively analyze molecular organization. When the labeling density is high enough, SMLM provides clear images of the molecular organization. However, either due to limited labeling efficiency or due to intrinsically low molecular abundance, SMLM delivers a small set of sparse and highly precise localizations. In this work, we introduce a correlation analysis of molecular locations based on the functional dependence of the complementary cumulative distribution function (CCDF) of the distance to the first neighbor (r_1). We demonstrate that the $\log(-\log(\text{CCDF}(r_1)))$ vs. $\log(r_1)$ is characterized by a scaling exponent n that takes extreme values of 2 for a random 2D distribution and 1 for a strictly linear arrangement, and find that n is a robust and sensitive metric to distinguish characteristics of the underlying structure responsible for the

molecular distribution, even at very low labeling density. The method enables the detection of fibrillary organizations and the estimation of the diameter of host fibers under conditions where a visual inspection provides no clue.

Introduction

Single molecules inside biological cells may be located in different topological environments such as the cytosol, the lumen or the membrane of organelles and fibrils, among others. In this context, biomolecular organization takes place over length scales that range from a few nanometers to micrometers, therefore covering several orders of magnitude in the spatial domain and in the number of molecules participating in these superstructures. Biological function is quite dependent on the molecular distribution and association states. Proteins, either individually or clustered, can be located in specific organelles, distributed over membrane domains, associated to cytoskeleton, or forming higher order supramolecular structures. Each particular spatial organization is linked to specific biological functions.

The advent of super resolution microscopy, also known as fluorescence nanoscopy, has dramatically changed biological imaging as it allows the observation of biomolecules organized at the nanoscale.¹⁻⁷ Probes and labeling techniques of increasing performance and specificity⁸⁻¹³, together with the development of numerous nanoscopy methods and analysis algorithms,¹⁴⁻¹⁶ extended and deepened the insight that optical nanoscopy provides for biological research.

A first generation of optical nanoscopy methods, composed of coordinate-targeted methods such as Stimulated Emission Depletion (STED)¹⁷, and coordinate-stochastic methods such as Stochastic Optical Reconstruction Microscopy (STORM)² and Photactivated Localization Microscopy (PALM)¹⁰, are able to deliver lateral resolutions in the 10-20 nm range, limited by the photon budget available from organic fluorophores. DNA-Points Accumulation for Imaging in Nanoscale Topography (DNA-PAINT)¹⁸ is becoming increasingly popular as it increases the fluorescence photon budget enabling sub-10 nm lateral resolution. Lately, a second generation of methods combines the advantages of coordinate-stochastic and coordinate-targeted methods to achieve higher resolutions, well below 10 nm. The recent molecular localization with Minimal emission Fluxes (MINFLUX) technique¹⁹ has lowered this limit to 1 nm, practically the molecular size of the fluorescent marker. More recently, a method called Supercritical Illumination Microscopy by Photometric z-Localization Encoding (SIMPLE)²⁰, and also MINFLUX²¹, have demonstrated a more challenging sub-10 nm axial resolution. In summary, fluorescence nanoscopy has reached a level of resolution that is comparable to the size of structural proteins and even to the size of the fluorescent marker, giving access to the supramolecular organization of biomolecules at molecular

scale in their native environment.

However, the direct visualization of protein organization is not always possible due to intrinsic biological conditions or limitations of the fluorescent labeling process. Labeling each and every moiety that is aimed to be visualized in a supramolecular protein structure is often challenging, if not impossible. The maximum density of labeling through (bio-) chemical interactions is limited by the interaction affinity as well as by steric or electrostatic repulsions. The effectiveness of endogenous labeling is also imperfect and actually hard to quantify. In addition, the organization and abundance of biomolecules varies depending on their expression level and their structural or physiological role. In the case of highly abundant structural proteins sufficiently high labeling densities can be achieved and fine structural details can be observed. This is the case, for example of microtubules, actin filaments or the Membrane-associated Periodic Skeleton (MPS) of mature neurons^{22–25}. By contrast, when a particular enzyme or receptor is the target, fluorescence nanoscopy images are usually composed of a sparse distribution of molecules where no evident organization is observed^{26–28}. Still, the location and function of these sparsely distributed biomolecules may be determined by a subjacent structure. For example, the location of CD44 receptor for cell-cell interaction and adhesion, whether tethered to linear or branched actin filaments, regulates its mobility^{22,23}. Motor proteins transporting cargos along microtubules experience a great hindrance to motion at intersections^{29,30}. The cell adhesion protein P-selectin redistributes into circular nanostructures in platelets when incubated with cancer cells due to accumulation in alpha-granules, while no pattern is observed when normal cells are used.³¹

Then, the question arises as to whether it is possible to obtain information about the underlying structure responsible for a particular protein distribution, even under conditions where sparse molecular localizations are acquired, at such a low density that no molecular organization can be recognized by direct visual inspection.

In view of its importance, efforts have been recently devoted to the quantitative characterization of molecular distribution within cellular images⁷. In particular, many algorithms were developed to extract information from the correlation of molecular positions determined by Single Molecule Localization Microscopy (SMLM)^{15,32–43}. These analyses were recently reviewed¹⁵. As examples, they include the probability density function of nearest neighbors⁴⁴. Those algorithms based on Ripley's K analysis⁴⁵ or on the pair correlation function³⁵ measure the distribution function in concentric disks or rings from a central point, respectively, providing information on clustering. Area sectioning based on Voronoi polygons^{46,47} has been used to quantify regions of clusters and of sparsely distributed molecules. Cluster-type analysis gives valuable quantitative information on the extent and number of partners forming the aggregates^{35,37,46,48–50}. In all these cases, the basis for the

analysis is an array of molecular (x, y) positions derived from different SMLM techniques, and the random distribution of molecules in two dimensions (2D) is used as a reference. Nevertheless, biomolecules are often organized in fibrillary structures. Fibrils and reticules are structures profusely present in biological cells that extend predominantly in one dimension (1D) and are characterized by linear-like molecular distributions. Recently, the group of Owen reported an analysis strategy to interrogate fibrillary organization in SMLM data using an angular version of the Ripley's K function^{38,43}. This analysis extracts several quantitative descriptors such as the number of fibers, their length, as well as the locations and angles of branching points, although it only works well in situations where straight fibers are the dominant structure and the labeling density is high.

In many cases, however, biological structures that show a linear organization in the short range tend to curve and intertwine in a longer range, or form part of higher order structures, giving rise to higher dimensionality sub-regions where preferential directions are blurred. For that reason, they cannot be treated either as pure 1D- or 2D-like distributions.

Here, we develop an analysis method to obtain information about molecular organization from SMLM data at very low densities. Based on simulated and experimental SMLM data, we demonstrate that the distance dependence of the complementary cumulative distribution function (CCDF) of the distance to the first neighbor is an adequate means to obtain quantitative indexes that correlate with the dimensionality and complexity of the underlying structure responsible for the molecular distribution. Even at low localization densities, where visual inspection provides no clue about the molecular organization, this method reveals if the molecular distribution is two-dimensional or fibrillary, providing also information about the degree of intercrossing and the width of the host fibers.

Experimental

Primary hippocampal neuronal cultures: Cell cultures had the same origin and were treated in the same way as previously described⁵¹. Briefly, mouse (CD1) hippocampal neurons were harvested from embryonic day 17 pups, following the general guidelines of the National Institute of Health (NIH, USA) and approval of the National Department of Animal Care and Health (SENASA, Argentina). Neurons were then cultured in Neurobasal medium (Gibco) supplemented with 5 mM GlutaMAX-I (Gibco) and 2% B27 supplement (Gibco) at 37 °C and 5% CO₂, and seeded on #1.5 thickness glass coverslips of 18 mm diameter. Glass coverslips were previously coated with 0.05 mg/mL poly-L-lysine (overnight at 37°C) (Sigma Aldrich) and 1 µg/µL Laminin (3 h at 37°C) (Sigma Aldrich). Neurons were incubated for either 3 or 40 days, for microtubules and spectrin labeling, respectively.

Immunofluorescence: Neurons were simultaneously fixed and permeabilized for 20 min at RT in PHEM buffer (60 mM PIPES, 25 mM HEPES, 5 mM EGTA, 1 mM MgCl_2) containing 0.25% glutaraldehyde, 3.7% paraformaldehyde, 3.7% sucrose, and 0.1% Triton X-100. Samples were quenched with 0.1 M glycine in PBS for 15 min and subsequently blocked for 1 h in 5% BSA solution in PBS containing 0.01% Triton X-100. For spectrin labeling, mouse anti- β -Spectrin II monoclonal primary antibody (Clone 42/B-Spectrin II, BD Biosciences) was diluted 1:400 in blocking solution and incubated with the samples overnight at 4 °C. After 3 washes with PBS, Alexa Fluor 647 conjugated goat anti-mouse secondary antibody (Invitrogen, # A-21236) was incubated for 1 h at room temperature in 1:750 dilution in blocking solution.

For microtubules staining, neurons were treated with mouse monoclonal anti- α -Tubulin (clone TUB-A4A Sigma Aldrich) for 1 h at room temperature using a 1:400 dilution in 5% BSA in PBS, followed by 3 washes with PBS. Alexa Fluor 647-conjugated goat anti-mouse secondary antibody (Invitrogen, # A-21236) was then incubated for 1 h at room temperature in 1:300 dilution in blocking solution.

Poly-(methylmethacrylate) (PMMA) film with Rhodamine 6G single molecules embedded. Glass coverslips of 18 mm diameter were sonicated at 30°C for 10 min submerged in acetone, 10 min in aqueous solution of Hellmanex III (0.05%), and 10 min in Milli-Q water. After drying at 100°C for 2 h, coverslips were treated in a Plasma Cleaner (Zepto, Diener Electronic), with air-filtered plasma (15 minutes at 20% power). A solution containing Rhodamine 6G 0.01 nM (Sigma Aldrich) and 0.5 % w/w PMMA in toluene (Merck Millipore, spectroscopic grade) was spin coated onto the cleaned coverslip. The films were dried in vacuum at 50°C for 12 h before imaging.

STORM setup and imaging. The experimental setup was described in a previous work⁵¹. Briefly, the setup was custom-built around an Olympus IX-73 inverted microscope, and a 642 nm 1.5 W laser (MPB Communications 2RU-VFL-P-1500-642) was used for fluorescence. A 405 nm 50 mW laser (RGB Photonics Lambda Mini) was used to re-activate fluorophores from their dark states. Both lasers were combined with dichroic mirrors (Semrock LM01-427 and LM01-552), magnified and focused on the back focal plane of an oil immersion objective Olympus PlanApo 60x NA 1.42. A dichroic mirror (Semrock Di03-R 405/488/532/635-t1), a multi-edge notch filter (Semrock NF03-405/488/532/635E) and a band pass filter (Chroma ET 700/75m) were used in order to selectively detect the fluorescence light. The emission light was expanded with a 2x telescope so that the pixel size of the EMCCD camera (Andor iXon3 897 DU-897D-CS0-#BV) was 133 nm in the sample

plane. The camera and lasers were controlled with a custom software developed in the laboratory and described previously.⁵² STORM imaging buffer (IB), pH 8, contained 50 mM TRIS (Sigma Aldrich), 10 mM NaCl, 10% w/v D-glucose, 10 mM mercaptoethylamine, 1 mg/mL glucose oxidase (Sigma Aldrich), and 40 μ g/mL catalase (Sigma Aldrich). Typically, 20,000-30,000 frames at 30 ms of exposure time were acquired with a laser power density of ~ 20 kW/cm² at the sample.

Simulation and data analysis: Simulation of random location of molecules in different environments, as well as processing routines were performed in self-written programs in Matlab. STORM movies were analyzed with ThunderSTORM⁵³ plugin for Fiji-ImageJ⁵⁴, taking into account the link of multiple localizations of the same molecule in order to get the list of single molecule localizations necessary to reconstruct the final super-resolved images.

Results and Discussion

Working principle of the analysis

The central concept of our method is that the number of molecules encountered as a function of the distance r from any starting point scales differently with the distance, depending on the way molecules are distributed. For a random 1D distribution it scales linearly with r (Figure 1a) and for a random 2D distribution it scales as r^2 (Figure 1b). In intermediate cases, such as a reticule with different branching or crossing degrees (Figure 1c), the operative dependence of the number of encountered molecules scales as r^n with $1 < n < 2$.

The analysis we present is based on the distribution of distances from every detected molecule to its nearest detected neighbor, r_l . For a random distribution of molecules in 2D, the distribution function of the distance to the first neighbor, $P_{2D}(r_l)$, is given by:

$$P_{2D}(r_1) = 2.\pi.D_2.r_1.\exp(-\pi.D_2.r_1^2) \quad (1)$$

D_2 being the area density of molecules. The corresponding function in the case of a random arrangement of molecules along a line, $P_{1D}(r_l)$, is given by:

$$P_{1D}(r_1) = D_1.\exp(-D_1.r_1) \quad (2)$$

where D_l is the linear density of molecules. The n-dimensional Complementary Cumulative Distribution Function (CCDF) is defined as:

$$CCDF_{nD}(r_1) = 1 - \int_0^{r_1} P_{nD}(r_1') \cdot dr_1' \quad (3)$$

The function $CCDF(r_1)$ represents the probability of finding a neighbor at a distance greater than r_1 from the reference point. Consequently, $CCDF(0) = 1$. The expressions of $CCDF(r_1)$ for the random distributions in 1D and 2D are:

$$CCDF_{1D}(r_1) = \exp(-D_1 \cdot r_1) \quad (4)$$

$$CCDF_{2D}(r_1) = \exp(-\pi \cdot D_2 \cdot r_1^2) \quad (5)$$

The algebraic treatment of the localization data uses a plot of $\log(-\log(CCDF(r_1)))$ vs. $\log(r_1)$, as depicted in Figure 1d. In this plot, the pure 1D and 2D random organizations are represented by straight lines with a slope of 1 and 2, respectively. Intermediate cases display distance-scaling exponents, n , between 1 and 2. In general, we expect:

$$CCDF(r_1) = \exp(-B \cdot r_1^n) \quad 1 \leq n \leq 2 \quad (6)$$

where B is an adimensionalization coefficient, playing a role equivalent to a density.

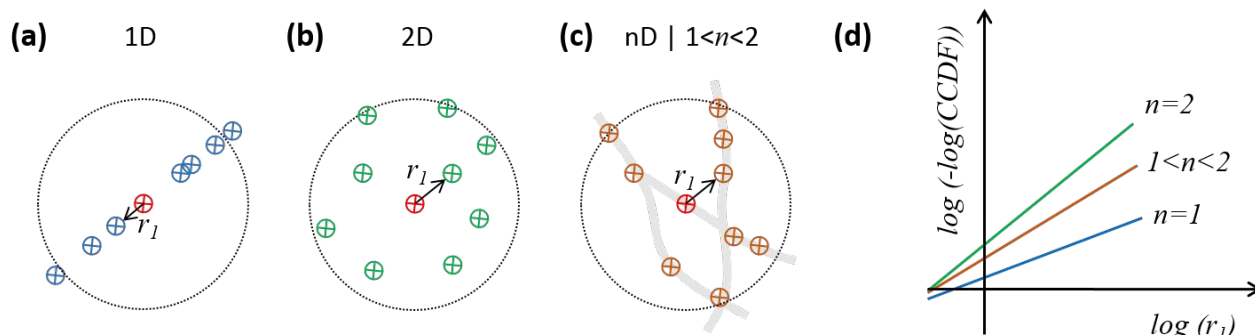


Figure 1. Schematic representation of three different distributions of molecules. (a) Distributed randomly over a line. (b) Distributed randomly on the plane. (c) Distributed randomly in a fibrillar network. (d) Representation of the expected slopes of the $\log(-\log(CCDF(r_1)))$ vs $\log(r_1)$ for the three cases.

In the next sections, we test the power of $CCDF(r_1)$ to discern different molecular organizations sampled with low density localizations.

We will assume that the target molecules are labeled and detected with uniform probability. We will concentrate the analysis in *a priori* random distributions. On the one side they represent the less

biased case, while on the other, they can be analytically studied and serve as reference to compare with actual distributions.

View Article Online
DOI: 10.1039/C9NR10805J

2D random distributions

Figure 2a shows two example simulations of molecular localizations distributed randomly in 2D with two different localization densities. It is important to note that the density of localizations has different meanings in 1D or 2D distributions. We choose to use the average distance to the first neighbor $\langle r_1 \rangle$ as a measure of the density of localizations, because this parameter can be obtained for any distribution independently of its dimensionality and provides an unbiased comparison. SMLM data is characterized by the average uncertainty in the localization of single molecules σ . The value of σ varies depending on the particular SMLM method and the experimental conditions. In order to withdraw general conclusions, we perform our analysis in units of σ . The two simulations shown in Figure 2a correspond to $\langle r_1 \rangle = 7.6 \sigma$ and $\langle r_1 \rangle = 20.2 \sigma$.

Figure 2b shows the plots of $\log(-\log(CCDF(r_1)))$ as a function of $\log(r_1)$, obtained from the distributions of Figure 2a. The slope of these plots yields a value for n , the distance scaling exponent of the distribution, according to Equation 6. Linear fits retrieve values of $n = 1.97$ and 1.96 for the high and low density situations, respectively. Figure 2c shows the dependence of n (mean value and standard deviation) on $\langle r_1 \rangle$ obtained from simulations. Over a range of $\langle r_1 \rangle$ spanning from 5 to 8000 σ , the average value of n is consistently found near the expected value of 2. We note that in order to obtain the theoretically expected value of 2, it is necessary to exclude localizations near the borders. If this is not taken into account, slightly smaller values are obtained (being the difference less than 1%). Both the standard deviation of $\langle n \rangle$ as well as the effect of the border increase as the number of localizations decreases. We performed 1000 simulations, each one using the same number of localizations sufficiently away from the borders, and obtained for 3250 localizations $\langle n \rangle = 2.00$ and $\sigma_n = 0.06$, and for 650 localizations $\langle n \rangle = 2.00$ and $\sigma_n = 0.13$. As it will be analyzed in more detail below, the standard deviation of n is determined by the number of localizations and not by the number of simulations.

An experimental realization of the 2D random distribution of molecules was achieved by spin-coating a PMMA film containing well dispersed Rhodamine 6G (R6G) molecules. An example of 187 R6G molecules localized in a $70 \times 70 \mu\text{m}^2$ area of the film is shown in Figure 2d. These localizations were obtained with a $\sigma = 16 \text{ nm}$, and $\langle r_1 \rangle = 2247 \text{ nm} = 140 \sigma$. Figure 2e shows the corresponding distance scaling plot and linear fit, that yields $n = 2.04$. Figure 2f displays the distribution of n obtained from the analysis of 100 images of the R6G-doped PMMA film, which is

well described by a Gaussian function with an average of 2.02 ± 0.04 (95% confidence interval), and $\sigma_n = 0.14$. The average number of R6G molecules per image was 141. Observations are consistent with the simulation predictions.

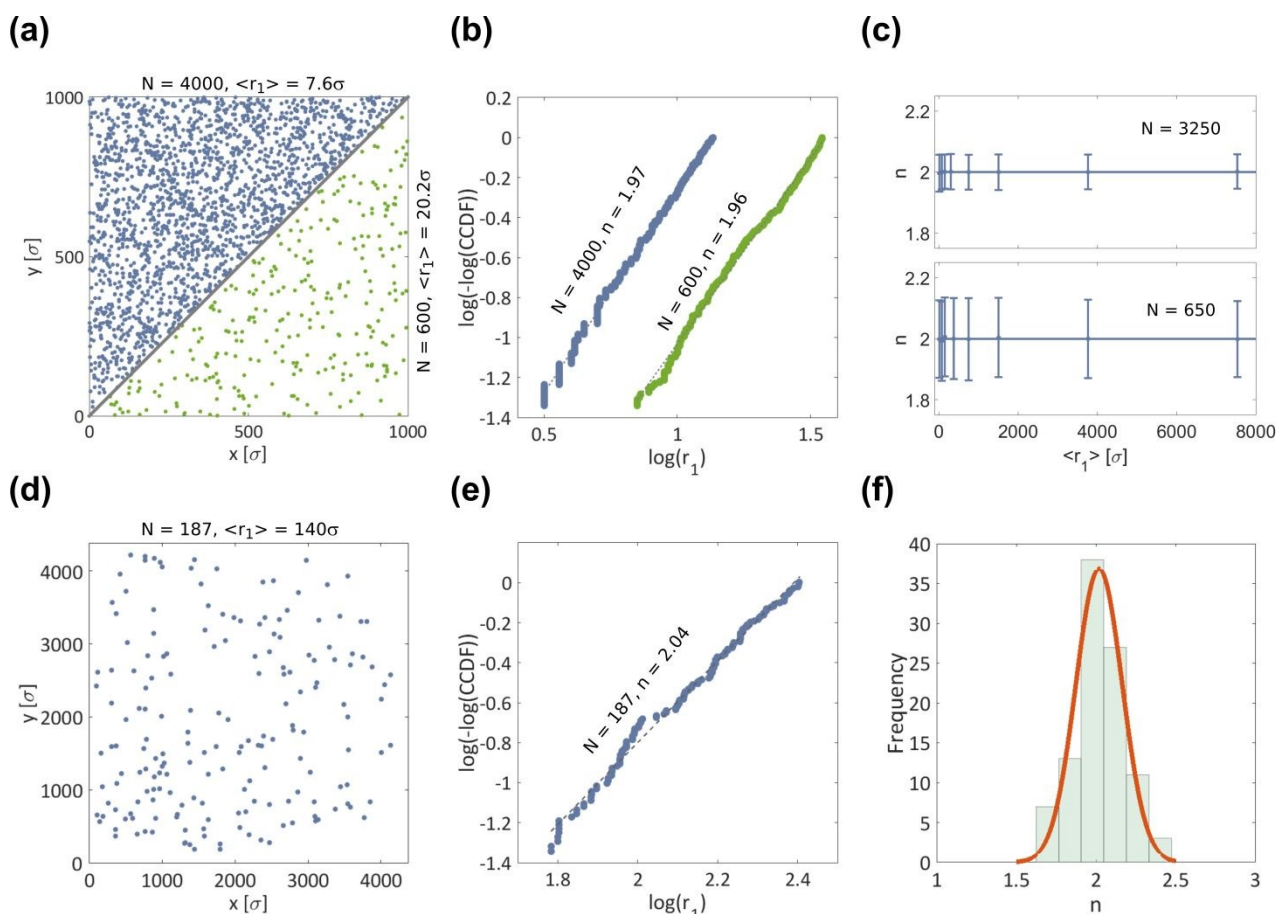


Figure 2: (a). Example of two random distributions in 2D with different density. The total number of points, N , and the average distance to the first neighbor, $\langle r_1 \rangle$, characterize each distribution. Distance expressed in units of the average uncertainty in SMLM, $\sigma = 16$ nm (see Figure S1) (b) The corresponding plots according to Equation 6 to obtain the distance scaling exponent, n . (c) Plot of n as a function of $\langle r_1 \rangle$. Error bars represent the standard deviation. Each point is the result of 1000 simulations. (d) SMLM image of a PMMA spin coated film of Rhodamine 6G (one image shown from a sequence of 100). (e) The corresponding plot of the distribution according to Equation 6. (f) Distribution histogram of the value of the distance scaling exponent, n , in the 100 images of the sequence.

Fibrillar organizations

Next, we performed a similar characterization of the *CCDF* analysis for the other extreme case: molecules distributed randomly along a single straight line. We note that, while 2D random

distributions can be encountered in real life experiments, the perfectly linear distribution is an idealization because it would correspond to molecules organized along a fiber of null width and localized with infinite precision. This situation can, of course, be simulated and a distance scaling exponent of 1.00 is obtained. Real cases of molecular organization along fibers are more complex and require a more profound analysis. Molecules placed along a single straight fiber will display distance scaling exponents greater than 1 when the probability of localizing molecules in the direction perpendicular to the fiber is increased. This, in turn, can be favored by the increase of three factors: fiber width, localization uncertainty, and localization density. The effect of the first two factors is illustrated in Figure 3a, where the situation of molecules distributed along a straight fiber of width d , that are localized with uncertainty σ , is schematically shown. Clearly, both parameters contribute to the fact that the detected positions depart from a perfect line. An effective width, D , can be defined as $D = d + 2\sigma$. Figure 3b shows the influence of increasing d , while keeping $\langle r_1 \rangle$ constant. Three simulated fibrillar distributions are displayed (left) together with their corresponding plots to obtain n (right). As expected, for a given $\langle r_1 \rangle$, an increase of d results in a larger value of n . Analogously, Figure 3c shows the influence of increasing the density of localizations in a fiber of constant width. It also results in an increase of n . Finally, Figure 3d summarizes the behavior of n as a function of d and $\langle r_1 \rangle$ in fibrillary structures. The influence of the localization uncertainty is implicitly included as d and $\langle r_1 \rangle$ are expressed in terms of σ . High densities (low $\langle r_1 \rangle$ values) as well as thick fibers favor values of n approaching 2, as expected. Nevertheless, in cases of high density, the fibrillary shape of the underlying structure will be evident to the eye and there would be no need to perform this analysis.

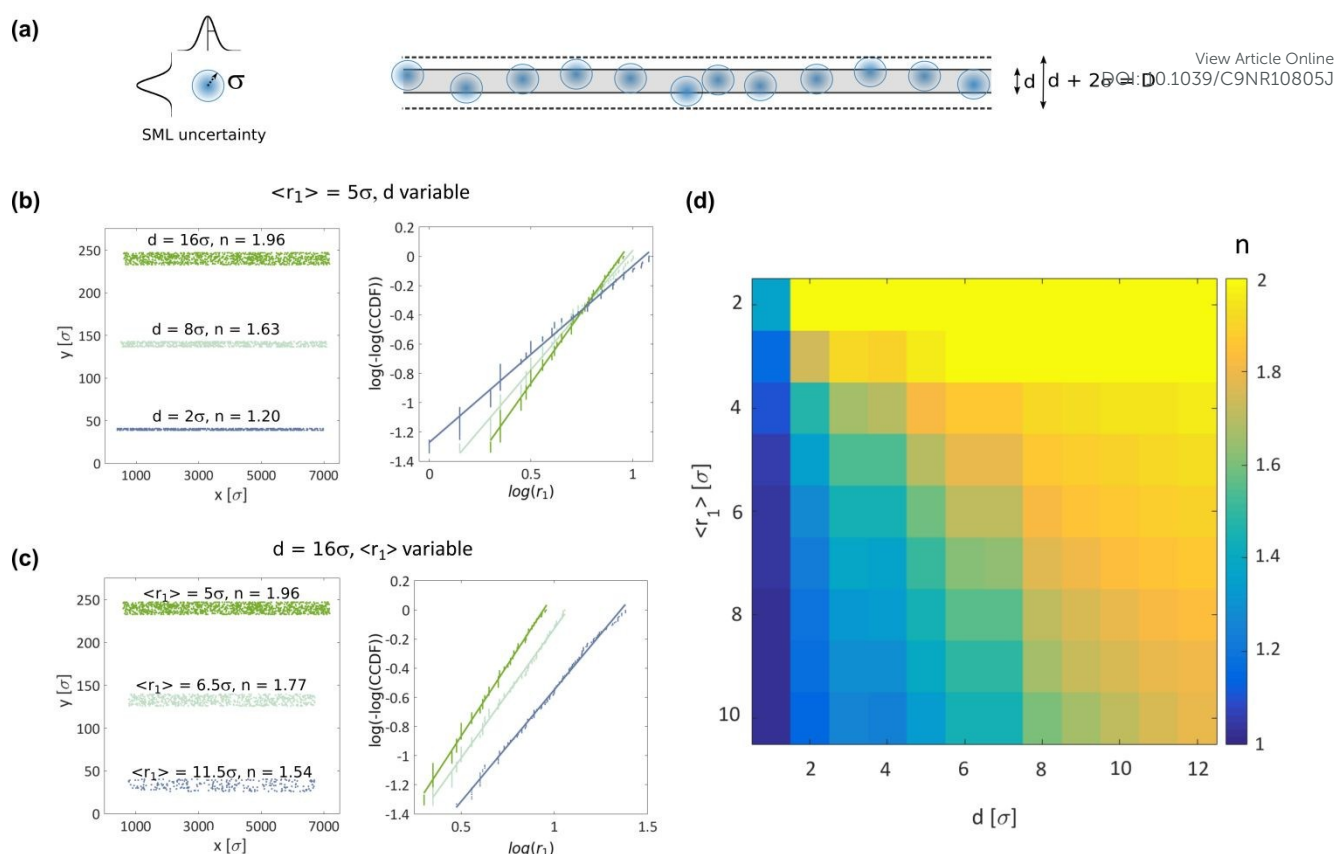


Figure 3: (a) Schematic representation of molecules distributed randomly in a fiber of width d , localized uncertainty σ . (b) Left: Simulations of random distributions with fixed mean distance to first neighbor, $\langle r_1 \rangle = 5\sigma$, and increasing fiber width, d , (from bottom to top). Right: Corresponding plots to obtain the distance scaling exponent of each distribution. (c) Left: Simulations of random distributions with fixed $d = 16\sigma$, and decreasing $\langle r_1 \rangle$ (from bottom to top). Right: Corresponding plots to obtain the distance scaling exponent of each distribution. (d) Histogram of the distance scaling exponent (expressed in color code) derived from simulations of random distributions in linear environments, as a function of $\langle r_1 \rangle$ and d .

In addition to the localization uncertainty and the width of the host fiber, the value of the distance scaling exponent n of a fibrillary distribution can be increased by other factors. The presence of more than one fiber in the region analyzed (even if they are parallel), as well as the crossing or bending of the fibers contribute to increase the probability of nearby locations with respect to the case of a single straight fiber. Some of these cases are illustrated in the SI (Figure S2).

In order to test the performance of our analysis in a real fibrillary structure, we applied it to single molecule localizations on microtubules of hippocampal neurons. Figure 4a displays a STORM image reconstructed with around 300,000 localizations of single AlexaFluor647 molecules immuno-conjugated to α -tubulin in a hippocampal neuronal cell. Figure 4b shows, in a divided

image, two randomly selected subsets of 2,000 and 200 localizations within the same region. The respective plots of $\log(-\log(CCDF(r_1)))$ vs $\log(r_1)$ used to obtain the distance scaling exponent are displayed in Figure 4c and the behavior of n as a function of $\langle r_1 \rangle$ is shown in Figure 4d. In the latter, each point is the average of 1000 different random choices of subsets of a given number of localizations. Remarkably, the same value of n is obtained even when considering only 200 localizations. As expected for a reticular pattern with intercrossing fibers, an intermediate value between 1 and 2 is obtained for n . Again, the analysis based on the $CCDF(r_1)$ provides a robust measure of the molecular organization: in this case a fibrillary network labelled at very low localization densities, where it is difficult or impossible to infer any information from a visual inspection of the images.

Another possible scenario that delivers an intermediate value of n is the case of two populations of molecules: fiber-associated molecules and molecules randomly distributed on the areas not occupied by the fibers. In this case, the value of n increases with the proportion of non-associated molecules. A simulation of this possible scenario is shown in Figure S4, where a subset of 388 localizations from Figure 4a are taken as a pure fibrillar organization case to which non associated molecules are added and assumed to be randomly distributed.

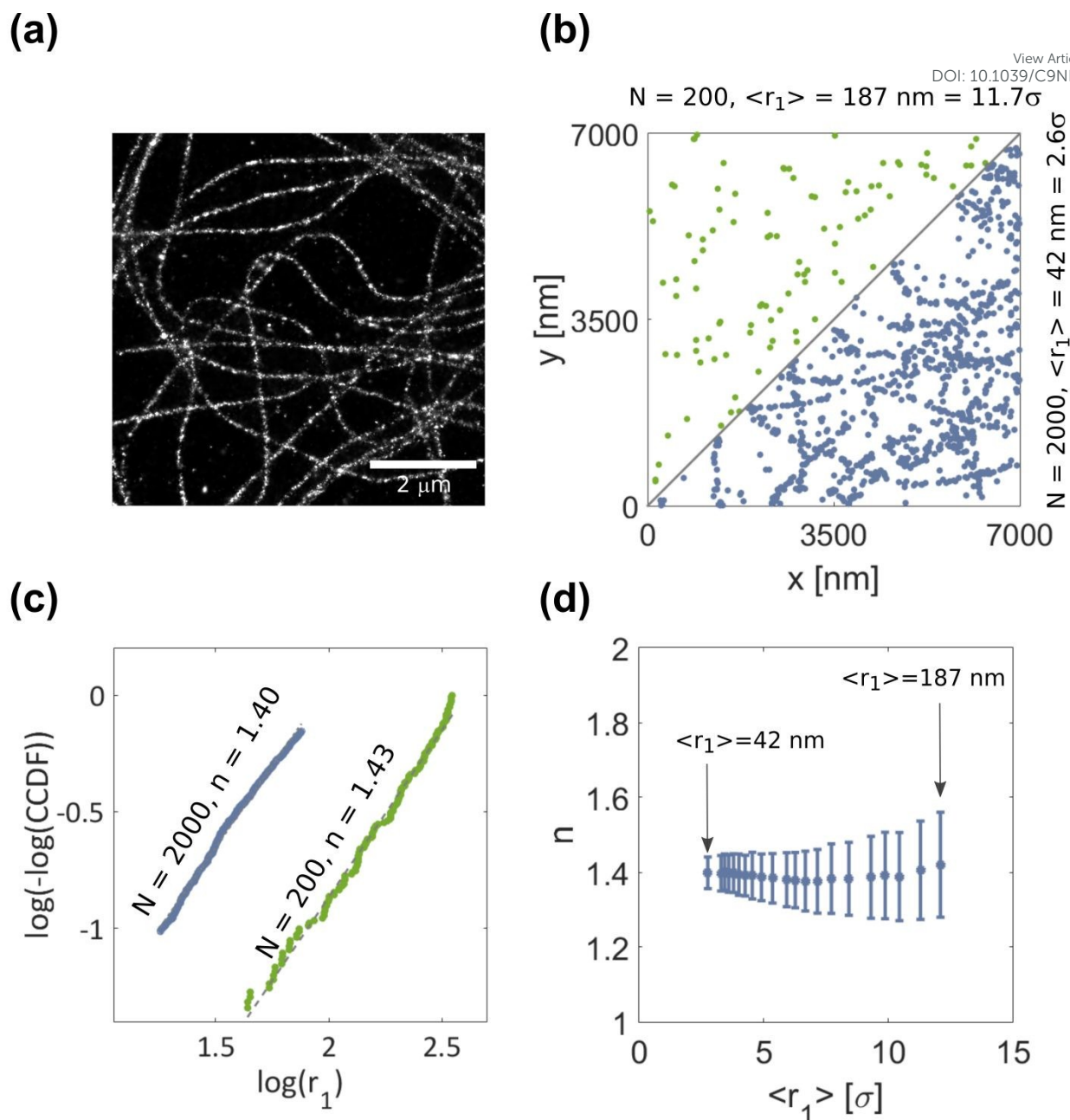


Figure 4: (a) Super resolution image of microtubules in a hippocampal neuron with AlexaFluor647 labeled α -tubulin. (b) Random selection of subsets of localizations of the whole image. The image is divided in two, each half being a representative part of the whole image: an upper triangle showing the corresponding half of 200 localizations of the whole image, and a lower triangle showing the corresponding other half of 2000 localizations of the whole image. (c) The respective plots for the distance scaling exponent of the distributions of panel (b). (d) Dependence of n on $\langle r_1 \rangle$. Each point is the average of 1000 randomly selected subsets of localizations and error bars represent the standard deviation. The arrows indicate the conditions of $\langle r_1 \rangle$ of the two examples shown in panel (b).

Spatial information

View Article Online

DOI: 10.1039/C9NR00053J

Up to now, we characterized the molecular distribution over a whole region of interest by an average value of n . Nevertheless, biomolecules in real samples may present different types of organization, degrees of complexity, and local concentrations in different regions of a cell. Figure 5a illustrates this fact with a SMLM (STORM) image of microtubules in a hippocampal neuron. Regions 1 to 4 were specifically chosen as examples of different local complexity scenarios of the microtubule network within the same cell area. For each of these regions the exponent n is plotted in Figure 5b as a function of the mean first neighbor distance. To build this plot, smaller subsets were taken out of the total localizations in each region, while each point is the average of 100 different random choices. From Figure 5b, we can observe two facts. First, the value of n increases with the complexity of the selected region (in this case given by the number of crossings and changes of direction of microtubules). Second, for very low and very high values of $\langle r_1 \rangle$ the probability of finding neighbors in directions transverse to each fiber increases, and therefore the obtained value of n increases slightly. For the high density situation, the reason for this is the same as the one described for a single straight fiber: increasing the number of molecules raises the probability of finding neighbors in all directions, and at some point the distances between neighbors in the transverse and longitudinal direction are even. On the other hand, when the molecular density is too low, the molecules located in the same fiber are so far from one another that the first neighbor can be found in a different host fiber, located at a random direction, due to fiber intercrossing. With this information at hand, we aim to obtain a map of local values of n . On doing this we are faced to the following compromise. For a given labelling density, increasing the spatial resolution implies a higher uncertainty in the determination of n because it must be determined using a smaller set of localizations (N). The value of σ_n grows for smaller N . Figure S3 shows that, while the relative uncertainty σ_n/n does not change significantly with increasing number of simulations of the same experiment, it shows a quasi Poissonian dependence with N , i.e. $\log(\sigma_n/n) \approx -0.50 \log(N)$, with a small offset. This means that $\sigma_n/n \approx N^{-1/2}$. For the following, we have taken a lower limit of 100 molecules in the considered area, which enables the determination of n with an uncertainty of ca 15%, both for the 2D and for the fibrillary distributions (Figure S3).

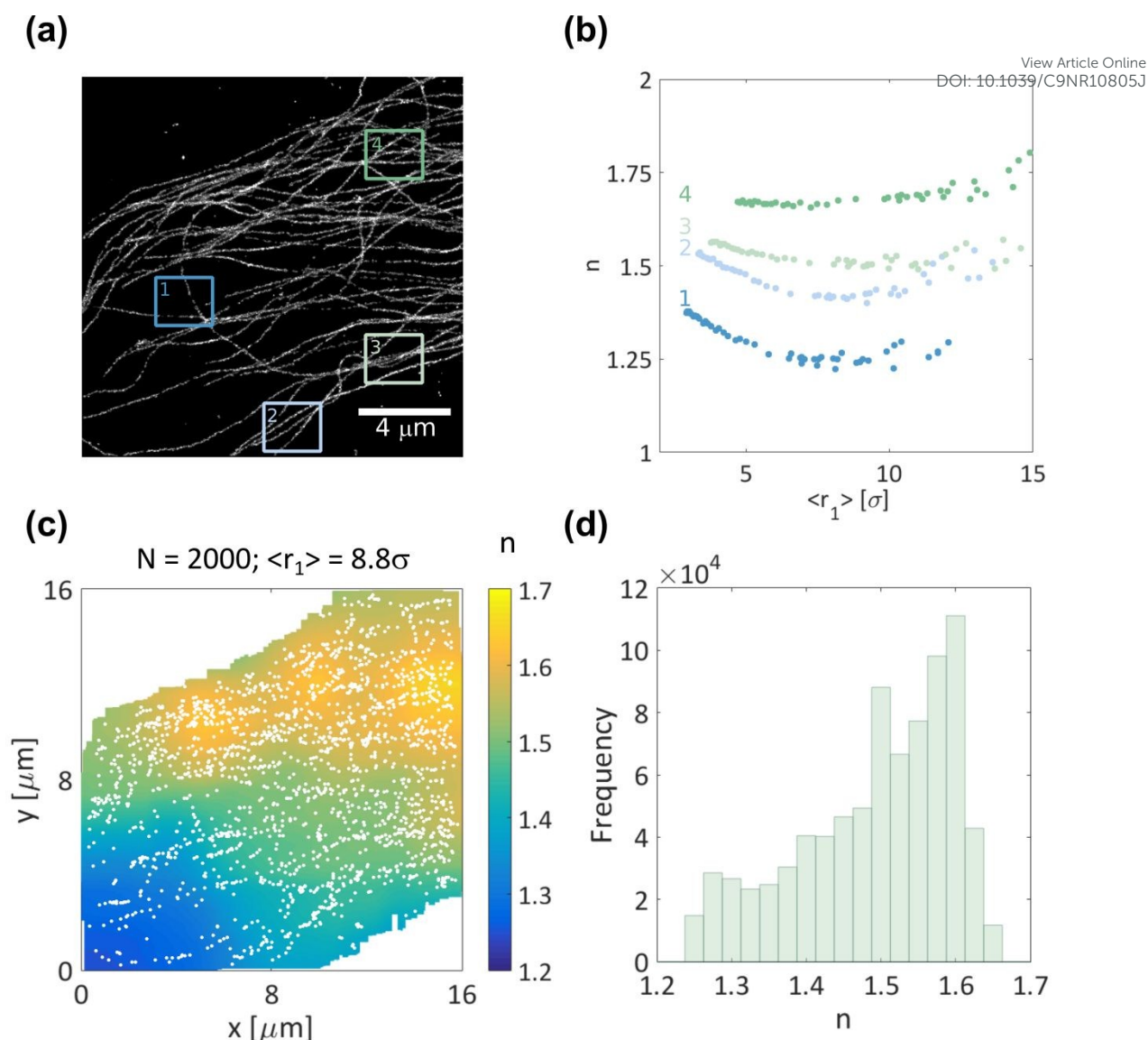


Figure 5: (a) SMLM image of AlexaFluor647 labeled α -tubulin in neuronal hippocampal cells. Regions 1 – 4 were analyzed separately. (b) Plot of the distance scaling exponent, n , as a function of the average distance to the first neighbor, $\langle r_1 \rangle$, in the distribution of each region 1 – 4 as marked in panel (a). Each point is the average of 100 simulations performed with different random subsets within each region. (c) Color map representing the local average value of n (see text) of 2000 localizations taken randomly from the image in (a). White dots show the molecular localizations in one of the 100 simulations averaged to build the color plot. White regions were not computed due to insufficient number of localizations. (d) Histogram of the distribution of n values of panel (c).

In Figure 5c we consider a restricted set of 2000 localizations, $\langle r_1 \rangle = 8.8\sigma$, out of the 1.4 million localizations that form the full STORM image shown in Figure 5a. Using these 2000 localizations, a color map of the local values of n was calculated as follows. We divided the whole area in square pixels of σ^2 size. A value of n was assigned to each one of these pixels according to the following

procedure. For each localized molecule, we identified its nearest 100 neighbors. As explained above, this number of locations determines a 15% uncertainty in the value of n . These 100 localizations defined a rectangular area by the extreme values of the (x,y) coordinates of the set. We only considered sets when $\langle r_1 \rangle < 15 \sigma$. We did not make an exhaustive analysis of the influence of this upper limit but clearly setting it at a very high value will extend the influence of low density regions whereas a low value will restrict the analysis to the more dense areas. We set this value at twice the $\langle r_1 \rangle$ of the whole set, still in a range where n displays a constant behavior with $\langle r_1 \rangle$ (Figure 5b). We computed the value of n and assigned it to the whole area determined by the 100 molecules. Finally, the value of n in each pixel of the whole n -map is the average of all determinations for each particular location. The map shown in Figure 5c is an average of 100 different choices of 2000 molecules.

The spatial resolution of this map can be defined by the square root of the area containing the first 100 neighbors of each molecule. With this criterion, resolution depends locally according to the molecular density. If the latter is too low, the rectangles will be larger than for cases of higher densities and, in consequence, the spatial resolution will be lower. For Figure 5c, where 2000 molecules occupy $\sim 60 - 70 \%$ of the total area (i. e. $\sim 150 - 180 \mu\text{m}^2$), the resolution lies in the range of $2.5 - 3 \mu\text{m}$.

The distribution of n values is displayed in the histogram of Figure 5d. Extreme values of 1.25 and 1.65 are obtained. Figure S5 compares the average distribution of n of Figure 5c with four individual determinations. Values of n around 1.5 are predominant and reflect environments like the ones of regions 2 and 3 of Figure 5a, with multiple crossing fibers. Complex environments, like the ones of region 4 of Figure 5a represent less than 5% of the area. Values of n in the vicinity of 1.2 represent areas with single filaments, like the region around the bottom left corner of Figure 5a. This value coincides with the minimum value of n obtained in simulations of linear distributions. Values of $n < 1.2$ are not obtained because even the thin immunolabeled microtubules have a thickness ($40 - 45 \text{ nm}$) greater than the localization uncertainty of SMLM ($10 - 20 \text{ nm}$), and therefore do not represent a strictly linear host. If we were analyzing a protein located in an unknown fibrillar environment, we could reasonably assume that the lowest n value corresponds to the simplest structure, represented by a single linear fiber. In this image, the average value of σ is 16 nm . A value of $n = 1.2$ when $5\sigma < \langle r_1 \rangle < 12\sigma$, is compatible with d of $2 - 3 \sigma$ (Figure 3d), in agreement with the thickness of the immunolabeled microtubule.

Multi-scale organization

Biomolecules may be organized in fibrillary structures over multiple length scales. As an example of biomolecular multiscale arrangement, we considered the distribution of spectrin in the MPS of a mature hippocampal neuronal axon. In the MPS, spectrin forms rings that determine the axonal diameter and are separated by 190 nm from one another along the axon²⁴. Both the length and width of axons are several times larger than the lateral resolution in SMLM (10-20 nm) and therefore they represent an environment for single molecule localizations closer to a two-dimensional situation than to a fibrillar one. However, the inner organization of spectrin is expected to result in a dimensionality lower than 2.

View Article Online
DOI: 10.1039/C9NR00053J

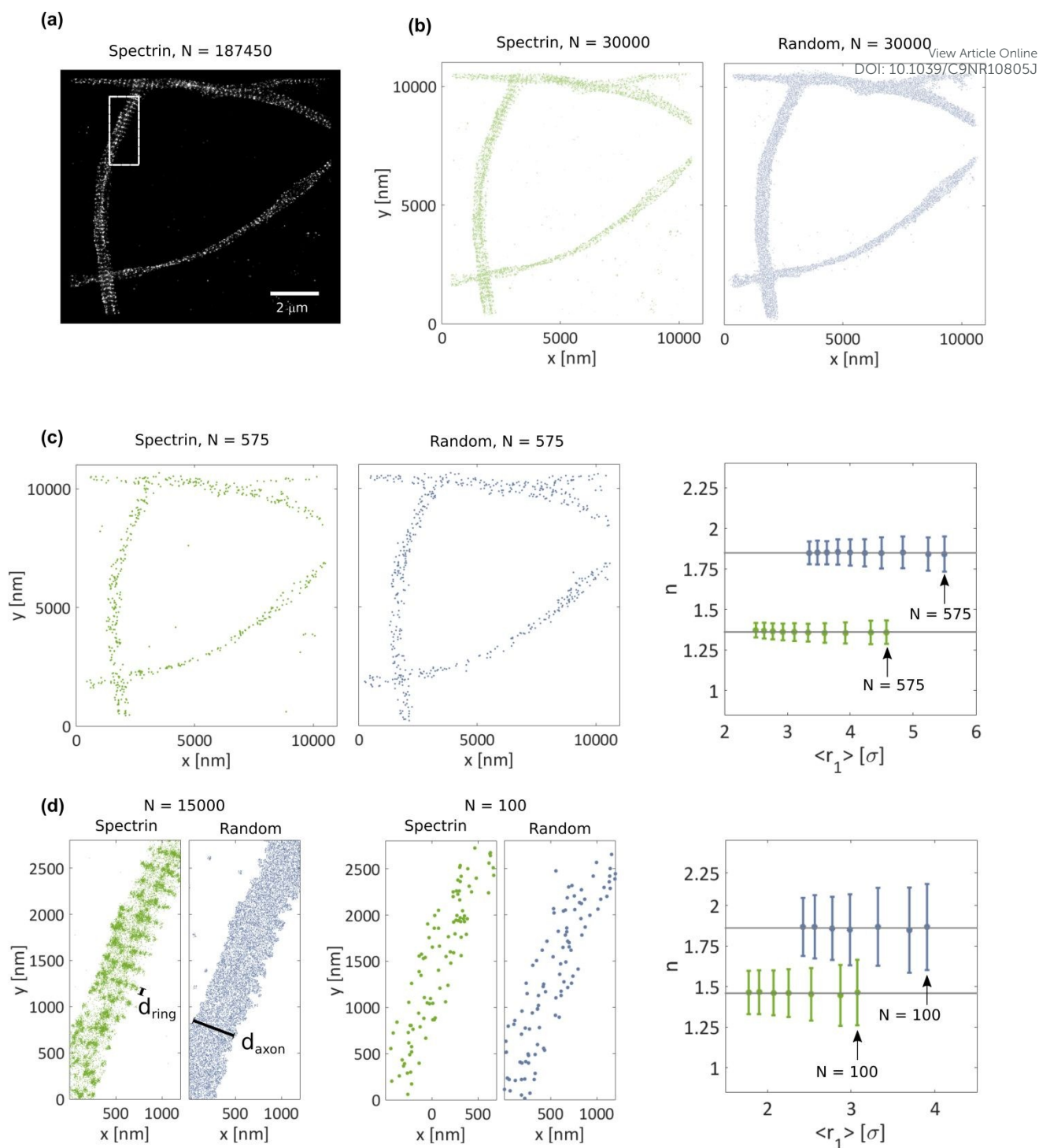


Figure 6: (a) Super resolution image of spectrin in hippocampal neurons labeled with Alexa Fluor 647. (b). In green, to the left, random subset of 30,000 molecules out of the localizations that form the STORM image (a). In blue, to the right, 30,000 localizations distributed randomly over the same region. (c) Random subset of 575 molecules for each distribution of panel (b) and the plot showing the dependence of n on the molecular density for the two distributions considering subsets of 2000 to 575 molecules from the corresponding distributions on panel (b). Each point is the average of 1000 different random choices of subsets. Average n values are 1.85 for the random distribution and

1.36 for the real sample. (d) Detail of the region defined by the white rectangle in panel (a); left: 15,000 localizations in the actual (green) and a random (blue) distribution. Center: Subsets of 100 molecules for each distribution. The marked lengths d_{ring} and d_{axon} are the respective characteristic lengths that determine the distribution in each case. Right: Plot of n as a function of $\langle r_1 \rangle$, showing the value corresponding to the distributions of the central panel (d). Average n values are 1.86 for the random distribution and 1.45 for the actual one.

Figure 6a shows a SMLM image of AlexaFluor647 labeled spectrin in hippocampal neurons obtained from 187450 localizations. A subset of 30,000 localizations is shown in Figure 6b (green) together with the same number of localizations randomly distributed in the same region (blue). The internal organization of spectrin in the MPS is evident in the real distribution, while it is not in the random distribution over the same area. Next we considered the case of a downsampled image, using just 575 molecules of each distribution, as shown in Figure 6c. Under this condition, the MPS structure is no longer evident to the eye, and the real sample becomes indistinguishable from a random distribution of molecules. Figure 6c also shows the plot of n as a function of $\langle r_1 \rangle$. For each density condition, 1,000 different random subsets of localizations were used in order to calculate the mean value and the standard deviation of n . Once more, values of n are significantly different even though the order in the green distribution cannot be detected by the eye. Figure 6d enlarges the portion defined by the white rectangle in Figure 6a. The rings perpendicular to the axon direction are evident, as well as the difference to the random distribution in the same space and with the same amount of localizations, that is depicted in blue. The ring width d_{ring} and the axon width d_{axon} are the characteristic lengths that determine the distribution. Also shown is a subset of 100 molecules of each distribution. The distance scaling exponent of the axonal distribution of spectrin is significantly smaller than the one of the 2D random simulation even at very low densities ($N = 100$, $\langle r_1 \rangle = 3 - 4 \sigma$) where the visual inspection is meaningless.

Mitochondria are another example where molecules can be found organized in multiscale fibrillary structures. As described by Jakobs and Wurm⁵⁵, three main different labeling scenarios can be expected: a mitochondrion with a regular 100 nm separation between cristae; a helical structure circumventing the matrix; and randomly distributed molecules in the outer membrane. Figure S6 illustrates simulations of these three types of molecular distributions at different density and the dependence of nn as a function of the average distance to the first neighbor. In the first two cases, the (x,y) projections of molecular positions form fibrillary organizations hosted in an elongated structure, and hence can be treated as a multi-scale organization. The fibrillar arrangements give rise to n values in the range of 1-1.4. For the case of molecules located randomly in the outer

membrane, a higher scaling exponent is expected. Whether the exponent n reaches a value near 2 or not will depend on the aspect ratio of each analyzed mitochondrion and on the molecular density.

View Article Online
DOI: 10.1039/C9NR10805J

All the examples presented demonstrate that the distance scaling exponent n is a robust and sensitive measure to discern different degrees of confinement of molecules that are randomly distributed over a given surface or fibrillary structure. Nevertheless, clustering of proteins is very often encountered^{36,40,46,49,50}. Figure S7 compares clustered, random, and mixed random and clustered distributions of the same average density. The analysis based on the distance-scaling exponent renders an easily differentiable diagnosis of these three types of distribution, and indicates that the method is also useful to detect if clusters are present, as well as to estimate the proportion of clustered molecules and the aspect ratio of clusters. Further studies of varying clustering scenarios must be carried out in the future in order to completely understand the potential and applicability of the distance-scaling exponent to characterize clustered protein organization.

Finally, we can think of potential applications where this method might be useful in the analysis of sparse distributions, besides the already mentioned as a way to distinguish if a given protein is randomly distributed either in the cytoplasm or in the membrane, or if it is associated to fibers or elongated structures. It might be helpful to monitor time evolution of low density molecular distributions: for example, if a protein originally located in a fibrous arrangement is released to the whole cell, n is expected to increase along with this process. The same is true if there is an increase in the branching of the fibrous arrangement as a function of time. As mentioned in the Introduction of the manuscript, receptors such as CD44 can be associated to actin and be affected in its function by the degree of branching of the actin filaments. If spatial distribution of n can be performed in the image, it renders information on the distribution of fibrillar complexity, and the corresponding histogram of n values can point to the evolution of this complexity.

Conclusions

In this work, we propose the use of the distance scaling exponent n of the *CCDF* of the distance to the nearest neighbor as an indicator of the molecular organization. Though it is not expected that only one parameter will fully characterize any molecular distribution, n is a robust and sensitive indicator that provides information about the subjacent molecular organization even at low density of localizations, where visual inspection is meaningless. Remarkably, the value of n in a given region of interest remains constant over a broad range of localization density and reveals if the molecules are organized in a random 2D distribution or form part of a fibrillary structure. Analysis of subsets of localizations, both simulated and extracted from a real cases, demonstrate that single determinations of n show variabilities that are mainly dependent on the number of localizations

considered, in a quasi Poissonian fashion. Values of n with a statistical error of 15% can be obtained with just 100 localizations. Local analysis allows the identification of areas of high and low complexity (crossings and bendings) of a fibrillary network. The value of n increases with the possibility of different radial locations of molecules around a given point. Thus, many factors contribute to the increase of the value of n , namely, fiber complexity, surface distribution, mixed specific and non-specific location. On the other side, particular attention should be given to values of n smaller than 1.5, as displayed in the images of Figures 4 to 6, either globally or locally. They are indicative of stringent limitations to the location of nearby molecules, as in fibrillar distribution. Furthermore, the width of the host fibers can be estimated from the minimum value of n and the localization uncertainty, by using the plot of Figure 3d, as we exemplified for the microtubules of Figure 5 with only 2000 localizations in the whole image.

We demonstrate that n is also useful to detect the presence of fibrillary structures confined to other supra-structures, as in the case of the organization of spectrin in neuronal axons. Here we note that further developments of the method, such as considering higher order neighbors or analyzing the dependence of n on $\langle r_n \rangle$, should provide information about the internal structure with respect to the supra-structure.

Acknowledgements

AMS and LFL are fellows, MAM-V is a former fellow, and FDS and PFA are research staff from Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET, Argentina). The authors acknowledge Dr. Federico Barabás and Luciano Masullo (CIBION) for acquiring the STORM images and Dr. Nicolás Unsain, Dr. Mariano Bisbal, Dr. Alfredo Cáceres (INIMEC, CONICET), Dr. Sebastián Giusti, and Dr. Damián Refojo (IBioBA-MPSP-CONICET) for preparing and providing the biological samples. Research was performed under support from grants PIP-0397, PIP-0626 (CONICET), and PICT 2013-1391, PICT2013-0792, PICT2014-3634, PICT2014-0739 from Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT, Argentina). FDS thanks the support from the Max-Planck Society and Alexander von Humboldt Foundation.

References

- 1 S. W. Hell, M. Dyba and S. Jakobs, *Curr. Opin. Neurobiol.*, 2004, **14**, 599–609.
- 2 M. J. Rust, M. Bates and X. Zhuang, *Nat. Methods*, 2006, **3**, 793–795.
- 3 L. Schermelleh, A. Ferrand, T. Huser, C. Eggeling, M. Sauer, O. Biehlmaier and G. P. C. Drummen, *Nat. Cell Biol.*, 2019, **21**, 72–84.
- 4 Y. M. Sigal, R. Zhou and X. Zhuang, *Science*, 2018, **361**, 880–887.

- 5 G. Vicidomini, P. Bianchini and A. Diaspro, *Nat. Methods*, 2018, **15**, 173–182.
- 6 S. J. Sahl, S. W. Hell and S. Jakobs, *Nat. Rev. Mol. Cell Biol.*, 2017, **18**, 685–701.
- 7 M. S. Dietz and M. Heilemann, *Nanoscale*, 2019, **11**, 17981–17991.
- 8 M. J. Hinner and K. Johnsson, *Curr. Opin. Biotechnol.*, 2010, **21**, 766–776.
- 9 Y. Hori and K. Kikuchi, *Curr. Opin. Chem. Biol.*, 2013, **17**, 644–650.
- 10 E. Betzig, G. H. Patterson, R. Sougrat, O. W. Lindwasser, S. Olenych, J. S. Bonifacino, M. W. Davidson, J. Lippincott-Schwartz and H. F. Hess, *Science*, 2006, **313**, 1642–1645.
- 11 N. George, H. Pick, H. Vogel, N. Johnsson and K. Johnsson, *J. Am. Chem. Soc.*, 2004, **126**, 8896–8897.
- 12 A. Gautier, A. Juillerat, C. Heinis, I. R. Corrêa, M. Kindermann, F. Beaufils and K. Johnsson, *Chem. Biol.*, 2008, **15**, 128–136.
- 13 A. Keppler, S. Gendreizig, T. Gronemeyer, H. Pick, H. Vogel and K. Johnsson, *Nat. Biotechnol.*, 2003, **21**, 86–89.
- 14 S.-H. Lee, J. Y. Shin, A. Lee and C. Bustamante, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 17436–17441.
- 15 P. R. Nicovich, D. M. Owen and K. Gaus, *Nat. Protoc.*, 2017, **12**, 453–460.
- 16 D. M. Owen, C. Rentero, J. Rossy, A. Magenau, D. Williamson, M. Rodriguez and K. Gaus, *J. Biophotonics*, 2010, **3**, 446–454.
- 17 S. W. Hell and J. Wichmann, *Opt. Lett.*, 1994, **19**, 780–782.
- 18 A. Sharonov and R. M. Hochstrasser, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 18911–18916.
- 19 F. Balzarotti, Y. Eilers, K. C. Gwosch, A. H. Gynnå, V. Westphal, F. D. Stefani, J. Elf and S. W. Hell, *Science*, 2017, **355**, 606–612.
- 20 A. M. Szalai, B. Siarry, J. Lukin, D. J. Williamson, N. Unsain, R. Becerra, D. Refojo, A. Caceres, M. Pilo-Pais, G. Acuña, D. M. Owen, S. Simoncelli and F. D. Stefani, *bioRxiv*, 2019, 693994.
- 21 K. C. Gwosch, J. K. Pape, F. Balzarotti, P. Hoess, J. Ellenberg, J. Ries and S. W. Hell, *Nat. Methods*, 2020, **17**, 217–224.
- 22 P. Sil, N. Mateos, S. Nath, S. Buschow, C. Manzo, K. G. N. Suzuki, T. Fujiwara, A. Kusumi, M. F. Garcia-Parajo and S. Mayor, *Mol. Biol. Cell*, 2019, mbc.E18-11-0715.
- 23 S. A. Freeman, A. Vega, M. Riedl, R. F. Collins, P. P. Ostrowski, E. C. Woods, C. R. Bertozzi, M. I. Tammi, D. S. Lidke, P. Johnson, S. Mayor, K. Jaqaman and S. Grinstein, *Cell*, 2018, **172**, 305–317.
- 24 K. Xu, G. Zhong and X. Zhuang, *Science*, 2013, **339**, 452–457.

- 25 N. Unsain, F. D. Stefani and A. Cáceres, *Front. Synaptic Neurosci.*, 2018, **10**, 1–8.
- 26 D. Baddeley and J. Bewersdorf, *Annu. Rev. Biochem.*, 2018, **87**, 965–989.
- 27 J. M. H. Goudsmits, A. M. Van Oijen and A. Robinson, *Biophys. J.*, 2016, **110**, 1708–1715.
- 28 A. P. Curd, J. Leng, R. E. Hughes, A. J. Cleasby, B. Rogers, C. Trinh, M. A. Baird, Y. Takagi, C. Tiede, C. Sieben, S. Manley, T. Schlichthaerle, R. Jungmann, J. Ries, H. Schroff, M. Peckham. *BiorXiv*, 2020, DOI 10.1101/2020.02.13.947135.
- 29 Š. Bálint, I. V. Vilanova, Á. S. Álvarez and M. Lakadamyali, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 3375–3380.
- 30 J. P. Bergman, M. J. Bovyn, F. F. Doval, A. Sharma, M. V. Gudheti, S. P. Gross, J. F. Allard and M. D. Vershinin, *Proc. Natl. Acad. Sci. U. S. A.*, 2018, **115**, 537–542.
- 31 J. Bergstrand, L. Xu, X. Miao, N. Li, O. Öktem, B. Franzén, G. Auer, M. Lomnytska and J. Widengren, *Nanoscale*, 2019, **11**, 10023–10033.
- 32 W. Vandenberg, M. Leutenegger, T. Lasser, J. Hofkens and P. Dedecker, *Cell Tissue Res.*, 2015, **360**, 151–178.
- 33 P. Annibale, S. Vanni, M. Scarselli, U. Rothlisberger and A. Radenovic, *PLoS One*, 2011, **6**, e22678.
- 34 P. Sengupta, T. Jovanovic-Talisman, D. Skoko, M. Renz, S. L. Veatch and J. Lippincott-Schwartz, *Nat. Methods*, 2011, **8**, 969–975.
- 35 P. Sengupta, T. Jovanovic-Talisman and J. Lippincott-Schwartz, *Nat. Protoc.*, 2013, **8**, 345–354.
- 36 F. Baumgart, A. M. Arnold, K. Leskova, K. Staszek, M. Fölser, J. Weghuber, H. Stockinger and G. J. Schütz, *Nat. Methods*, 2016, **13**, 661–664.
- 37 M. A. Kiskowski, J. F. Hancock and A. K. Kenworthy, *Biophys. J.*, 2009, **97**, 1095–1103.
- 38 R. Peters, M. Benthem Muñoz, J. Griffié, D. J. Williamson, G. W. Ashdown, C. D. Lorenz and D. M. Owen, *Bioinformatics*, 2017, **33**, 1703–1711.
- 39 Z. Zhang, Y. Nishimura and P. Kanchanawong, *Mol. Biol. Cell*, 2017, **28**, 333–345.
- 40 J. Scurll, L. Abraham, D. W. Zheng, R. Tafteh, K. Chou, M. R. Gold and D. Coombs, *bioRxiv*, 2019, **0450**, 515627.
- 41 S. Malkusch and M. Heilemann, *Sci. Rep.*, 2016, **6**, 34486.
- 42 S. Culley, D. Albrecht, C. Jacobs, P. M. Pereira, C. Leterrier, J. Mercer and R. Henriques, *Nat. Methods*, 2018, **15**, 263–266.
- 43 R. Peters, J. Griffié, G. L. Burn, D. J. Williamson and D. M. Owen, *Sci. Rep.*, 2018, **8**, 1–8.
- 44 F. C. Clark, Philip J. Evans, *Ecology*, 1954, **35**, 445–453.
- 45 B. D. Ripley, *J. R. Stat. Soc. Ser. B*, 1977, **39**, 172–212.

- 46 L. Andronov, I. Orlov, Y. Lutz, J. L. Vonesch and B. P. Klaholz, *Sci. Rep.*, 2016, **6**, 1–9.
- 47 R. G. Sampayo, A. M. Toscani, M. G. Rubashkin, K. Thi, L. A. Masullo, I. L. Viol, J. N. Lakins, A. Cáceres, W. C. Hines, F. Coluccio Leskow, F. D. Stefani, D. R. Chialvo, M. J. Bissell, V. M. Weaver and M. Simian, *J. Cell Biol.*, 2018, jcb.201703037.
- 48 A. Getis and F. Janet, *Ecology*, 1987, **68**, 473–477.
- 49 M. Dai, R. Jungmann and P. Yin, *Nat. Nanotechnol.*, 2016, **11**, 798–807.
- 50 J. Rossy, E. Cohen, K. Gaus and D. M. Owen, *Histochem. Cell Biol.*, 2014, **141**, 605–612.
- 51 F. M. Barabas, L. A. Masullo, M. D. Bordenave, S. A. Giusti, N. Unsain, D. Refojo, A. Cáceres and F. D. Stefani, *Sci. Rep.*, 2017, **7**, 1–10.
- 52 F. M. Barabas, L. A. Masullo and F. D. Stefani, *Rev. Sci. Instrum.*, 2016, **87**, 126103.
- 53 M. Ovesný, P. Křížek, J. Borkovec, Z. Svindrych and G. M. Hagen, *Bioinformatics*, 2014, **30**, 2389–2390.
- 54 C. T. Rueden, J. Schindelin, M. C. Hiner, B. E. DeZonia, A. E. Walter, E. T. Arena and K. W. Eliceiri, *BMC Bioinformatics*, 2017, **18**, 1–26.
- 55 S. Jakobs and C. A. Wurm, *Curr. Opin. Chem. Biol.*, 2014, **20**, 9–15.

View Article Online
DOI: 10.1039/C9NR10805J