A DNA origami platform for quantifying protein copy number in super-resolution

Francesca Cella Zanacchi1,6, Carlo Manzo1,2,7, Angel S Alvarez1,7, Nathan D Der3,4, Maria F Garcia-Parajo1,5 & Melike Lakadamyali1,6

Single-molecule-based super-resolution microscopy offers researchers a unique opportunity to quantify protein copy number with nanoscale resolution. However, while fluorescent proteins have been characterized for quantitative imaging using calibration standards, similar calibration tools for immunofluorescence with small organic fluorophores are lacking. Here we show that DNA origami, in combination with GFP antibodies, is a versatile platform for calibrating fluorophore and antibody labeling efficiency to quantify protein copy number in cellular contexts using super-resolution microscopy.

Single-molecule localization microscopy has become an important tool for nanoscale imaging, and immense effort has been dedicated to quantifying protein copy number in super-resolution images. However, exact copy number quantification is impaired by the stochasticity of labeling and the complex photophysics of the fluorescent probes. Therefore, recent work has focused on developing analytical approaches and calibration standards to overcome this challenge in order to calibrate and count photoactivatable fluorescent proteins (FPs) in small organic fluorophores. Because of their high photon budget (compared with that of FPs), small organic fluorophores are popular probes for many super-resolution studies. Targeting such probes to the protein of interest typically requires immunofluorescent labeling with primary and secondary antibodies. In this case, both the antibody labeling efficiency and the number of fluorophores conjugated to the antibodies are highly stochastic. Additionally, fluorophores undergo repeated reactivation events. Together, these issues pose major challenges for protein copy number quantification. Partial solutions to these challenges have been reported. For example, fluorophore photophysics can be modeled or characterized using fluorophore-labeled antibodies or images of sparse spots on the sample. In the case of DNA-PAINT, which relies on ‘on-off’ binding of fluorophore-labeled small oligos, the binding kinetics can be modeled and accounted for in the quantification. Nonetheless, in all cases the unknown labeling stoichiometry, which results from the stochasticity of fluorophore–antibody and antibody–target binding, affects the precision of the final quantification. Ad hoc calibration standards have allowed researchers to quantify complex structures; however, there is no general solution for calibrating labeling stoichiometry.

To develop versatile calibration standards that can be used for quantifying protein copy number in intracellular contexts, we used a previously developed 3D DNA origami chassis. The handles projecting out from the chassis provide site- and sequence-specific attachment points for single fluorophores as well as for proteins of interest and allow testing of several labeling strategies (Fig. 1a). We first attached complimentary antihandle sequences labeled with Alexa Fluor 647 to the three handles located at positions 1, 7 and 13 of helix 0 to establish a baseline for the efficiency of handle–antihandle attachment. This attachment efficiency should be independent of the fluorophore used and only depend on the sequence of the oligos. A single TAMRA fluorophore attached at position 14 of the outer helices (h3, h4, h7, h8, h11) was used to identify the DNA origami structures on the glass slide (Fig. 1b). Single-step photobleaching of Alexa Fluor 647 spots that colocalized with TAMRA revealed single, double and triple steps (Fig. 1c); and the distribution of the number of counted steps fit to a binomial giving a handle–antihandle attachment efficiency of 48% (Supplementary Fig. 1a). Similarly, STORM images of Alexa Fluor 647 spots that colocalized with TAMRA revealed single, double or triple clusters (Fig. 1d). We segmented these clusters using a previously developed algorithm and found that the intercluster distances matched the expected distance between the individual handles used for the labeling (Supplementary Fig. 1b).

The number of localizations detected from individual clusters showed a broad distribution (Fig. 1e); and the median value for one, two and three fluorophores increased roughly linearly (Fig. 1f and Supplementary Table 1).

We next purified a modified dimeric *Saccharomyces cerevisiae* dynein motor (see Online Methods and Supplementary Fig. 2) whose individual protomers contained both the SNAP-tag and GFP. Because dynein is a homodimer, each motor has two GFP copies. Dynein was covalently linked to antihandle sequences using the SNAP-tag at the same three positions (see Online Methods and Fig. 1a). We then immunostained the GFP with primary anti-GFP antibodies and photoswitchable dye pair (Alexa 488). The number of localizations from single, double and triple sequences of the oligos. A single TAMRA fluorophore attached at position 14 of the outer helices (h3, h4, h7, h8, h11) was used to identify the DNA origami structures on the glass slide (Fig. 1b). Single-step photobleaching of Alexa Fluor 647 spots that colocalized with TAMRA revealed single, double and triple steps (Fig. 1c); and the distribution of the number of counted steps fit to a binomial giving a handle–antihandle attachment efficiency of 48% (Supplementary Fig. 1a). Similarly, STORM images of Alexa Fluor 647 spots that colocalized with TAMRA revealed single, double or triple clusters (Fig. 1d). We segmented these clusters using a previously developed algorithm and found that the intercluster distances matched the expected distance between the individual handles used for the labeling (Supplementary Fig. 1b).

The number of localizations detected from individual clusters showed a broad distribution (Fig. 1e); and the median value for one, two and three fluorophores increased roughly linearly (Fig. 1f and Supplementary Table 1).

We next purified a modified dimeric *Saccharomyces cerevisiae* dynein motor (see Online Methods and Supplementary Fig. 2) whose individual protomers contained both the SNAP-tag and GFP. Because dynein is a homodimer, each motor has two GFP copies. Dynein was covalently linked to antihandle sequences using the SNAP-tag at the same three positions (see Online Methods and Fig. 1a). We then immunostained the GFP with primary anti-GFP antibodies and photoswitchable dye pair (Alexa 488). The number of localizations from single, double and triple
fluorescence microscopy and clustering analysis (Fig. 1g). Only those clusters that colocated with the TAMRA and had the expected intercluster distances were selected (see Online Methods). The counted number of single, double and triple clusters showed a binomial distribution with a labeling efficiency of 38% (Supplementary Fig. 1c), which was only slightly lower than the labeling efficiency for attaching single fluorophores (48%, Supplementary Fig. 1a)—this suggested that the main limitation in labeling is the efficiency of dynein attachment, rather than the antibody labeling efficiency. DNA origami immobilized on a biotin–streptavidin functionalized glass substrate and on cells displayed similar localization distributions (Supplementary Fig. 3). A calibration curve corresponding to the median number of localizations for clusters 1, 2 and 3 (GFPs 2, 4 and 6) was roughly linear (Fig. 1h and Supplementary Table 1), which suggests that the binding of primary antibody to its target GFP did not reach saturation levels. This calibration curve can be used to extract average protein copy numbers in a given image by comparing the median number of localizations obtained in the cellular context to the curve (Supplementary Note 1).

To determine whether this method can also be used to quantify the percentage of each oligomeric state, we fit the localization distributions to a functional form. The distributions could be simultaneously fit using only two free parameters (μ and σ), assuming that they correspond to the convolutions of respectively two, four and six functions f_i, where f_i is a log-normal distribution describing the probability distribution of number of localizations from monomeric GFP (see Online Methods and Fig. 1i). Therefore, for a general distribution of localizations containing an unknown mixture of oligomeric states, it should be possible to extract both the oligomeric state and the percentage of oligomers corresponding to that particular state by fitting the data to a linear combination of calibration distributions f_n obtained by recursively convoluting f_1 n times. We validated this idea in multiple ways. First, we generated a synthetic distribution of localizations by combining a known fraction of single, double and triple clusters from the DNA origami images, which we then fit to a linear combination of log-normal functions f_n (Fig. 2a)—where only even values for n were considered, given the dimeric nature of each motor. The optimal number of functions used for the fit (N_max) was chosen as the number minimizing the fit.
Fig. 2 | Validation of stoichiometry determination. (a,b) Estimation of the stoichiometry for a synthetic sample with known percentage of single, double and triple motors generated from the DNA origami images, fit to a linear combination of log-normal distributions up to three dimers. (c) The objective function for different stoichiometries with a minimum corresponding to three motors (\(N_{\text{max}} = 3\)). (d,e) Estimation of the stoichiometry for a synthetic sample with equal percentage (25%) of 1, 4, 8 and 16 motors (2, 8, 16 and 32 GFPs) generated from the DNA origami images, fit to a linear combination of log-normal distributions up to 20 dimers. (f) The objective function for a number of stoichiometries with a minimum corresponding to 20 motors (\(N_{\text{max}} = 20\)). (g) Clustering analysis of STORM images for DNA chassis functionalized with five motors. Scale bar, 200 nm. (h) Distribution showing the total number of localizations per five dynein motors (red) and the corresponding fit to a linear combination of log-normal functions up to five dimers (black line). (i) The objective function for a number of stoichiometries with a minimum corresponding to a stoichiometry of five motors (\(N_{\text{max}} = 5\)) and (j) distribution of number of motors at each experimental image (one to five rings) (black line). The objective function for a number of stoichiometries with a minimum corresponding to a stoichiometry of five motors (\(N_{\text{max}} = 5\)) and (j) distribution of number of motors at each experimental image (one to five rings) (black line).

Objective function (see Online Methods). The extracted fraction of single, double and triple motors was in excellent agreement with the expected fractions, given sufficient statistics (Fig. 2b and Supplementary Fig. 4a), for \(N_{\text{max}} = 3\) (Fig. 2c). Second, to extend the stoichiometry beyond 3, we generated synthetic distributions of localizations comprising an equal fraction of 1, 4, 8 and 16 motors (Fig. 2d). The peaks in the stoichiometry distribution were in agreement with the chosen oligomeric states (Fig. 2d,e), providing a good correlation with the theoretical data for \(N_{\text{max}} = 20\) (Fig. 2f, Supplementary Fig. 4b–d, and Supplementary Notes 2 and 3). Finally, we attached dynein to five handles on the chasis. In this case, because of the short distance between the handle positions (28 nm), we could no longer distinguish clusters corresponding to individual dynein motors (Fig. 2g). We thus combined all the localizations corresponding to each DNA origami structure (identified by the presence of TAMRA signal) and plotted the distribution of localizations (Fig. 2h). The obtained stoichiometry distribution for \(N_{\text{max}} = 5\) contained 37% single, 44% two, 14% three, 4% four, 1% five dynein motors (Fig. 2i), fitting well to a binomial distribution for a labeling efficiency of 33% in close agreement with the 38% labeling efficiency obtained for triple handles (Fig. 2j).

We finally tested this method on the nuclear pore complex (NPC) subunit Nup133 fused to GFP, which is expressed in U2OS cells in the presence of siRNA to knock down the endogenous Nup133. We chose Nup133 as its stoichiometry has been previously characterized20 (Fig. 3a). Super-resolution images showed ring-like structures, as expected (Fig. 3b,c), albeit with lower than the eight-fold symmetry per NPC. This is likely due to the incomplete siRNA knockdown and potential limitations with an antibody’s access to its target for high stoichiometries. We first manually sorted the Nup133 images, taking into account the number of Nup133 clusters that were visible by eye. We could reliably sort up to five Nup133 clusters, since the images of individual clusters started significantly merging together within the resolution limit of STORM for higher order structures. This manual sorting is prone to some errors, as multiple clusters in close proximity may be counted as a single cluster. We then fit the distribution for the number of localizations corresponding to one to five Nup133 clusters with \(f_{\text{i}}\) (where \(f_{\text{i}}\) corresponds to monomeric GFP) (Fig. 3d–h) and obtained the expected range of stoichiometries considering up to four GFPs per cluster (Fig. 3d–h and Supplementary Fig. 5a–e). Cluster analysis and fitting of NPC ring data without manual sorting (Fig. 3c) showed a broad distribution of stoichiometries, with a maximum stoichiometry of ~30 and a mean stoichiometry of ~12 (Fig. 3i, black bars; Supplementary Fig. 5f, Supplementary Fig. 6 and Supplementary Note 2). The maximum stoichiometry is consistent with the expected stoichiometry of 32. Given that the majority of the NPCs contained less than eight Nup133 clusters (Fig. 3i), we expected to obtain an average stoichiometry lower than 32. The most predominantly observed NPC rings contained around three Nup133 clusters (Fig. 3i), which is consistent with the mean stoichiometry of 12. In addition, weighing the stoichiometries obtained from the sorted data (one to five rings) (Fig. 3d–h) with their occurrence in the super-resolution images (Fig. 3i) gave a distribution that matched remarkably well to the experimentally obtained distribution (Fig. 3i, red line). Finally, similar results were obtained for another subunit of the NPC, Nup107, which belongs to the same subcomplex as that of Nup133 (Supplementary Fig. 7). In conclusion, we show that DNA origami can be used as a versatile calibration standard to quantify protein copy number in
immunolabeled samples imaged with super-resolution. The use of GFP antibodies provides a versatile strategy for quantifying a large number of proteins of interest using the calibration curve reported here. In order to do so, it is important to point out that the same imaging and image analysis conditions should be used as detailed in the Online Methods. We used standard imaging buffers, laser powers and acquisition settings that are typical for STORM experiments. Finally, this method is not limited to GFP antibodies; and it is applicable to antibodies against any endogenous protein as well as nanobody, Halo or SNAP-tag labeling and photoactivatable fluorescent proteins.

**METHODS**

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

**ACKNOWLEDGMENTS**

We thank P. Gomez (ICFO, Barcelona) for helpful discussions. The siRNA-resistant Nup133-GFP and Nup107-GFP plasmids were a kind gift from J. Ellenberg (EMBL, Heidelberg). M.L. acknowledges funding from the Fundació Cellex Barcelona, European Union Seventh Framework Programme under the European Research Council grants 337191-MOTORS and Spanish Ministry of Economy, Competitiveness and the Fondo Europeo de Desarrollo Regional (FEDER) grant FIS2015-63550-R (MINECO/FEDER). F.C.Z. acknowledges funding from the ‘Severo Ochoa’ Programme for Centres of Excellence in R&D (SEV-2015-0522). C.M. acknowledges funding from the Spanish Ministry of Economy and Competitiveness and the European Social Fund (ESF) through the Ramón y Cajal program 2015 (RYC-2015-17896).

**AUTHOR CONTRIBUTIONS**

F.C.Z. performed experiments and analyzed data. C.M. wrote software and analyzed data. A.S.A. performed the dynein purification and gave support with sample preparation. N.D.D. provided DNA origami materials. M.L. conceived the idea, and M.L. and M.F.G.-P. supervised the research. M.L. and F.C.Z. wrote the manuscript. All authors provided feedback on the manuscript.

**COMPETING FINANCIAL INTERESTS**

The authors declare competing financial interests: details are available in the online version of the paper.

Reprints and permissions information is available online at [http://www.nature.com/reprints/index.html](http://www.nature.com/reprints/index.html). Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Oddone, A., Vilanova, I.V., Tam, J. & Lakadamyali, M. Microsc. Res. Tech. 77, 502–509 (2014).
2. Durisic, N., Laparra-Cuervo, L., Sandoval-Alvarez, A., Borbely, J.S. & Lakadamyali, M. Nat. Methods 11, 156–162 (2014).
3. Annibale, P., Vanni, S., Scarselli, M., Rothlisberger, U. & Radenovic, A. Nat. Methods 8, 527–528 (2011).
4. Finan, K., Rauf, A. & Heilemann, M. Angew. Chem. Int. Edn Engl. 54, 12049–12052 (2015).
5. Fricke, F., Beaudouin, J., Ells, R. & Heilemann, M. Sci. Rep. 5, 14072 (2015).
6. Hummer, G., Fricke, F. & Heilemann, M. Mol. Biol. Cell 27, 3637–3644 (2016).
7. Jungmann, R. et al. Nat. Methods 13, 439–442 (2016).
8. Lee, S.H., Shin, J.Y., Lee, A. & Bustamante, C. Proc. Natl. Acad. Sci. USA 109, 17436–17441 (2012).
9. Puchner, E.M., Walter, J.M., Kasper, R., Huang, B. & Lim, W.A. Proc. Natl. Acad. Sci. USA 110, 16015–16020 (2013).
10. Rollins, G.C., Shin, J.Y., Bustamante, C. & Pressé, S. Proc. Natl. Acad. Sci. USA 112, E110–E118 (2015).
11. Ricci, M.A., Manzo, C., García-Parajo, M.F., Lakadamyali, M. & Cosma, M.P. Cell 160, 1145–1158 (2015).
12. Nieuwenhuizen, R.P. et al. Nat. Methods 10, 557–562 (2013).
13. Ehmann, N. et al. Nat. Commun. 5, 4650 (2014).
14. Bakker, G.J. et al. Proc. Natl. Acad. Sci. USA 109, 4869–4874 (2012).
15. Torreno-Pina, J.A. et al. Proc. Natl. Acad. Sci. USA 111, 11037–11042 (2014).
16. Derf, N.D. et al. Science 338, 662–665 (2012).
17. Beck-Peterson, S.L. et al. Cell 126, 335–348 (2006).
18. Schmidt, T., Schutz, G.J., Gruber, H.J. & Schindler, H. Anal. Chem. 68, 4397–4401 (1996).
19. Moertelmaier, M., Brameshuber, M., Linimeier, M., Schutz, G.J. & Stockinger, H. Appl. Phys. Lett. 87, 263903 (2005).
20. von Appen, A. et al. Nature 526, 140–143 (2015).
STORM optical setup. Imaging was performed on an inverted Nikon Eclipse Ti microscope (Nikon Instruments). The excitation module is equipped with four excitation laser lines: 405 nm (100 mW, OBIS Coherent, California), 488 nm (200 mW, Coherent Sapphire, California), 561 nm (500 mW MPB Communications, Canada) and 647 nm (500 mW MPB Communications, Canada). The laser-beam power was regulated through AOMs (AA Opto Electroncs MT80 A1, 5 Vis), and different wavelengths were mixed and coupled into the microscope objective through dichroic mirrors. The focus was locked through the Perfect Focus System (Nikon), and imaging was performed on an EmCCD camera (Andor iXon X3 DU-897, Andor Technologies). Fluorescence-emitted signal was spectrally filtered by a Quad Band filter (ZT405/488/561/647rpc-UF2, Chroma Technology) and selected by an emission filter (ZET405/488/561/647m-TRF, Chroma). For single-molecule detection, the emitted light was acquired at 25 Hz by an oil-immersion objective (Nikon, CFI Apo TIRF 100x, NA 1.49, Oil) providing a corresponding pixel size of 157 nm.

DNA origami structure assembly. 12-helix-bundle DNA origami chassis structures were prepared using p8064 scaffold and oligonucleotide staple sequences as previously described. Briefly, 100 nM scaffold (Tilibit Nanosystems) was mixed with 600 nM core staples (Life Technologies), 3.6 µM handle staples (IDT), and 9 µM TAMRA-labeled fluorophore antihandles (IDT). Folding was performed in DNA origami folding buffer (5 mM Tris pH 8.0, 1 mM EDTA and 16 mM MgCl2) with heating to 80 °C and cooling in single-degree increments to 65 °C for 75 min, followed by cooling in single-degree increments to 30 °C for 17.5 h. Folded chassis were purified by glycerol gradient sedimentation by centrifugation through a 10–45% glycerol gradient in TBE buffer supplemented with 11 mM MgCl2 for 130 min at 242,704 g in a SW50.1 rotor (Beckman) at 4 °C. Folded scaffolds were resuspended in DNA origami folding buffer (5 mM Tris pH 8.0, 1 mM DTT and 16 mM MgCl2) for 1 h at 16 °C with slow rotation. Beads were removed with centrifugal filters and protein concentrated with Amikon 100K and frozen in LiqN2. Concentration of the purified dynein was determined by spectrophoresis; 4–20% Criterion TGX Precast Protein Gels stained with SYPRO Ruby gel stain following the rapid-stain protocol. We used a Precision Plus Protein unstained Standard (Biorad) that offers absolute molecular-weight accuracy confirmed by mass spectrometry with a known amount of protein in each band to allow approximation of protein concentration in your sample. The gel has been imaged using the Molecular Imager Gel Doc XR+ System and analyzed using the Image Lab Software.

DNA origami structures: sample preparation. A LaBtek chamber (no. 1.0, eight well) was rinsed with KOH (1 M) and PBS three times. Coverglass was incubated with 100 µL of streptavidin (0.5 mg/ml in PBS) for 20 min and washed three times with PBS. The coverglass was subsequently incubated with 100 µL of BSA-Biotin (0.5 mg/mL in PBS) for 20 min, extensively washed in PBS, and incubated with fiducial markers (Carboxyl Fluorescent Particles, yellow, 1% w/v Spherotec SPH-CFP-0252-2, diameter 111 nm, diluted 1:25,000 in PBS). Blocking of coverglass was performed in blocking buffer containing 10% (w/v) BSA (Sigma) in DAB solution (30 mM Hepes, 50 mM KAcetate, 2 mM MgAcetate, 1 mM EGTA 7.5, 10% glycerol, 1 mM DTT, 0.5 mM Mg-ATP, 1 mM Pefabloc) and then incubated (20 min at RT) with BG oligonucleotides (20 µM). After functionalization, beads were washed three times in TEV buffer and incubated in TEV protease (1:100 in TEV buffer) for 1 h at 16 °C with slow rotation. Beads were removed with centrifugal filters and protein concentrated with Amikon 100K and frozen in LiqN2. Concentration of the purified dynein was determined by spectrophoresis; 4–20% Criterion TGX Precast Protein Gels stained with SYPRO Ruby gel stain following the rapid-stain protocol. We used a Precision Plus Protein unstained Standard (Biorad) that offers absolute molecular-weight accuracy confirmed by mass spectrometry with a known amount of protein in each band to allow approximation of protein concentration in your sample. The gel has been imaged using the Molecular Imager Gel Doc XR+ System and analyzed using the Image Lab Software.

DNA origami structures: sample preparation. A LaBtek chamber (no. 1.0, eight well) was rinsed with KOH (1 M) and PBS three times. Coverglass was incubated with 100 µL of streptavidin (0.5 mg/ml in PBS) for 20 min and washed three times with PBS. The coverglass was subsequently incubated with 100 µL of BSA-Biotin (0.5 mg/mL in PBS) for 20 min, extensively washed in PBS, and incubated with fiducial markers (Carboxyl Fluorescent Particles, yellow, 1% w/v Spherotec SPH-CFP-0252-2, diameter 111 nm, diluted 1:25,000 in PBS). Blocking of coverglass was performed in blocking buffer containing 10% (w/v) BSA (Sigma) in DAB solution (30 mM Hepes, 50 mM KAcetate, 2 mM MgAcetate, 1 mM EGTA 7.5, 10% glycerol, 1 mM DTT, 0.5 mM Mg-ATP, 1 mM Pefabloc) and then incubated (20 min at RT) with BG oligonucleotides (20 µM). After functionalization, beads were washed three times in TEV buffer and incubated in TEV protease (1:100 in TEV buffer) for 1 h at 16 °C with slow rotation. Beads were removed with centrifugal filters and protein concentrated with Amikon 100K and frozen in LiqN2. Concentration of the purified dynein was determined by spectrophoresis; 4–20% Criterion TGX Precast Protein Gels stained with SYPRO Ruby gel stain following the rapid-stain protocol. We used a Precision Plus Protein unstained Standard (Biorad) that offers absolute molecular-weight accuracy confirmed by mass spectrometry with a known amount of protein in each band to allow approximation of protein concentration in your sample. The gel has been imaged using the Molecular Imager Gel Doc XR+ System and analyzed using the Image Lab Software.

DNA origami structures: sample preparation. A LaBtek chamber (no. 1.0, eight well) was rinsed with KOH (1 M) and PBS three times. Coverglass was incubated with 100 µL of streptavidin (0.5 mg/ml in PBS) for 20 min and washed three times with PBS. The coverglass was subsequently incubated with 100 µL of BSA-Biotin (0.5 mg/mL in PBS) for 20 min, extensively washed in PBS, and incubated with fiducial markers (Carboxyl Fluorescent Particles, yellow, 1% w/v Spherotec SPH-CFP-0252-2, diameter 111 nm, diluted 1:25,000 in PBS). Blocking of coverglass was performed in blocking buffer containing 10% (w/v) BSA (Sigma) in DAB solution (30 mM Hepes, 50 mM KAcetate, 2 mM MgAcetate, 1 mM EGTA 7.5, 10% glycerol, 1 mM DTT, 0.5 mM Mg-ATP, 1 mM Pefabloc) and then incubated (20 min at RT) with BG oligonucleotides (20 µM). After functionalization, beads were washed three times in TEV buffer and incubated in TEV protease (1:100 in TEV buffer) for 1 h at 16 °C with slow rotation. Beads were removed with centrifugal filters and protein concentrated with Amikon 100K and frozen in LiqN2. Concentration of the purified dynein was determined by spectrophoresis; 4–20% Criterion TGX Precast Protein Gels stained with SYPRO Ruby gel stain following the rapid-stain protocol. We used a Precision Plus Protein unstained Standard (Biorad) that offers absolute molecular-weight accuracy confirmed by mass spectrometry with a known amount of protein in each band to allow approximation of protein concentration in your sample. The gel has been imaged using the Molecular Imager Gel Doc XR+ System and analyzed using the Image Lab Software.
were then washed twice and blocked in blocking buffer for 15 min at 4 °C. Immunostaining was performed by incubation with primary antibody (chicken polyclonal anti-GFP, Abcam 13970) diluted 1:2,000 in blocking buffer for 1 h at 4 °C. Samples were rinsed three times in blocking buffer and incubated for 1 h at 4 °C with donkey-anti chicken secondary antibodies (Jackson ImmunoResearch, 703-005-155) (1:50 in blocking buffer) labeled with photoactivatable dye pairs for STORM Alexa Fluor 405-Alexa Fluor 647. For experiments on BSC-1 cells (from ATCCC, #CLL-26), cells were plated (30,000 seeding density) on eight-well Lab-tek 1 coverglass chamber (Nunc) and grown under standard conditions and fixed with methanol-ethanol (1:1) at 20 °C for 2 min and incubated for 5 min with DNA origami (one motor attached) and rinsed three times in DAB solution.

Nup107 and Nup133: sample preparation. Human osteosarcoma U2OS cells (from ATCC, HTB-96) were plated (30,000 seeding density) on eight-well LabTek chambered coverglass (Nunc) and grown under standard conditions (DMEM, high glucose, pyruvate (Invitrogen 41966052) supplemented with 10% FBS). U2OS were chosen, since they perform well for transfection and siRNA KD of Nup. For GFP-tagged Nup107 and GFP-tagged Nup133 experiments, cells were transfected with the construct(s) (plasmid from J. Ellenberg, EMBL, Heidelberg, pEGFP-Nup107-s 32272res, Euroscarf plasmid ref. P0729 and pMGEFP-Nup133-s31401res, Euroscarf plasmid ref. P30728) using Fugene (FUGENE HD Transfection Reagent, Roche 4070975001). Incorporation into the pore of the GFP-tagged Nup was facilitated by depletion of the endogenous protein ENREF_5, performed by RNA interference, transfecting after 24 h the cells with a matching siRNA (Nup107 SiRNA s32272 and Nup133 SiRNA s31401, Thermo Fisher, Silencer Select siRNA s32272 and Silencer Select siRNA s31401, Nup107 and Nup133, 3 pmol of siRNA per well was used). After 70 h cells were rinsed with PFA 3%, extracted with 0.2% Triton X-100 in PBS for 2 min and fixed with PFA (3%) for 7°. Immunostaining of Nup107–GFP fusion protein was performed using immunofluorescence as described above. Cells lines were regularly tested for mycoplasma contamination by PCR-based standard methods (ATCC, Universal Mycoplasma Detection Kit, 30-1012K).

STORM imaging conditions. The imaging conditions were kept constant for all the experiments. Imaging was performed using TIRF illumination with an excitation intensity of ~1 kW/cm² for the 647 nm readout laser line and ~25 W/cm² using the 405 nm laser line. 85,000 frames at 25 Hz frame rate were acquired. For dual-color imaging of DNA origami structures, fluorescence signal from TAMRA was acquired with 561 nm laser (intensity of ~200 W/cm²). STORM imaging buffer was used that contained GLOX solution as oxygen scavenging system (40 mg/mL), Sigma, 0.5 mg/ml glucose oxidase, 10% glucose in PBS) and MEA 10 mM (Cysteamine MEA, Sigma-Aldrich, 30070-50G, in 360 mM Tris–HCl).

Data analysis. Analysis and reconstruction of super-resolution images was performed using custom software (Insight3, kindly provided by B. Huang, University of California) by Gaussian fitting of the single-molecules images to calculate the molecular localization coordinates. Molecules are identified by a threshold, and the radial positions x and y are extracted by fitting with a simple Gaussian function. The final image is obtained by plotting each identified molecule as a Gaussian spot with a width corresponding to the localization precision (9 nm) and finally corrected for drift. Molecules appearing within a distance of 9 nm are merged and considered the same molecule. Spatial clusters of localizations were identified based on a distance-based clustering algorithm by means of custom-written code written in Matlab11. The localizations list was first binned to 20 nm pixel size images that were filtered with a square kernel (7 × 7 pixels²) and thresholded to obtain a binary mask. Specifically, a density map was built by 2D convolution of the localization images with a square kernel (7 × 7 pixels²), and a constant threshold was used to digitize the maps into binary images. The low-density areas, where the density is lower than the threshold value and a value of 0 was assigned, are discarded from further analysis. Only the components of the binary image, where adjacent (six-connected neighbors) nonzero pixels were found, are analyzed. A peak-finding routine provides the clusters number and the relative centroid coordinates from the maxima of the density map in the connected regions. Molecular localizations lying over connected regions of the mask were assigned to each cluster using a distance-based algorithm depending on their proximity to the cluster centroids. For each cluster, its centroid position is iteratively recalculated and saved for further analysis until convergence of the sum of the squared distances between localizations and the associated cluster is reached. The cluster centroid positions, the number of localizations obtained per cluster and the cluster size are saved.

For DNA origami calibration, dual-color cluster analysis first allowed the identification of TAMRA signal (used as a reference to identify the DNA origami structures) and dynein clusters attached to the same DNA origami. In order to consider only the signal belonging to motors attached to DNA origami structures, only the clusters with a relative distance shorter than 200 nm between the clusters in the two channels were considered for further analysis. Clusters identifying single, double and triple motors were then sorted depending on the number motors attached. Additional filter was applied to select structures with the expected handle-to-handle distance (85 ± 7 nm and 157 ± 17 nm). To ensure the statistical significance, we chose a sample size able to ensure a power value close to 1 (the total number of DNA origami considered was N = 3,077; N = 1,153; N = 250 for single, double and triple motors, respectively.

The distributions of the number of localizations per cluster obtained for DNA origami structures showing 1, 2 and 3 dyneins (corresponding to 2, 4 and 6 GFPs, respectively) were used as a calibration standard. To this aim, we considered that the distribution of the number of localizations for a structure composed by n GFP can be recursively obtained as

\[ f_n = f_{n-1} \otimes f_1 \]

where \( \otimes \) represents the convolution, and \( f_1 \) is a log-normal distribution:

\[ f_1(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}} \]

(1)

The distributions of localizations obtained for 1, 2, 3 dyneins \((n = 2, 4, 6)\) were simultaneously fitted to the functions \( f_2, f_4, f_6 \) obtaining the parameters \( \mu = 3.3 \) and \( \sigma = 0.83 \). The same parameters were used for all the other fittings. The log-normal distribution
was chosen because, among several tested distributions, it provided the best data model.

For a general distribution of number of localizations, the copy number of a given protein can thus be estimated by fitting the distributions to a linear combination of the ‘calibration’ distributions $f_n$:

$$ g(x) = \sum_{n=1}^{N_{\text{max}}} \alpha_n f_n(x) \quad (2) $$

where $\alpha_n$ represents the weight of the distribution of $n$-mers, and

$$ \sum_{n=1}^{N_{\text{max}}} \alpha_n = 1. $$

To estimate motors attached to the DNA origami chassis, the fit was performed considering only dimers (linear combination of distributions $f_n$ with even values $n = 2, 4, 6, 8, \ldots, 2k$) given the dimeric nature of the motors containing two copies of GFP per motor; while for NPC estimation the fit was performed considering $n$ monomers (linear combination of distributions $f_n$ with values $n = 1, 2, 3, 4, \ldots, k$), where $k$ is the maximum stoichiometry.

Fittings are performed by a two-step numerical minimization of

$$ F = -w_t \sum_x p(x) \ln g(x) + w_E \sum_n \alpha_n \ln \alpha_n \quad (3) $$

which represents the sum of the negative log likelihood and the entropy. In the first term, $p(x)$ corresponds to the number of occurrences for number of localization $x$. In the first optimization step, we set $w_t = 1$ and

$$ w_E = \frac{\max(x) - \min(x)}{\langle x \rangle}, $$

with $\langle x \rangle$ representing the average value of the data; and we let the optimization run at varying $N_{\text{max}}$ until the minimum of the objective function $F_{\text{min}}$ is found. By means of this procedure, we calculate the maximum number of log-likelihood functions necessary to satisfactorily fit the data. Once this number is determined, we further refine the fit by performing a second step of optimization where the weight of the log likelihood is set to the inverse of its target value $w_t = 1/F_{\text{min}}$. When fitting distributions involving the linear combination of only dimeric terms ($n = 2, 4, 6, \ldots, 2k$), in the second step of optimization we further allow the parameters $\mu$ and $\sigma$ to slightly vary constrained to a maximum tolerance of 5% in order to supply to the reduced number of degrees of freedom. Calculation of the errors on the estimated weights $\alpha_n$ was based on the reciprocal of the diagonal elements of the Fisher information matrix and thus represent a lower bound to the standard error of the estimators. The code can be found at https://github.com/cmanzo/DECO.

For Nup133 and Nup107 quantification, clustering analysis is carried out to segment single nuclear pores and the distribution of the number of localizations per NPC ring was filtered considering a minimum average cluster radius of 40 nm. The total number of nuclear pores analyzed was $N = 1,460$ for Nup133 and $N = 855$ for Nup107.

Statistics and data analysis. The DNA origami data used for calibration are obtained by five independent experiments, and the total number of DNA origami structures imaged was $N_1 = 3,077$; $N_2 = 1,153$; $N_3 = 250$ for single, double and triple motors, respectively (Fig. 1g–i). In the case of validation experiments using DNA origami functionalized with five dynein motors, the number of DNA origami chassis imaged was $N = 934$ in $N = 4$ independent experiments (Fig. 2g–j).

Sorted data were used to quantify Nup133 (Fig. 3d–h) ($N = 1$ experiment, total number of NPC rings analyzed $N = 798$). Images corresponding to 1, 2, 3, 4, 5 clusters were sorted (NPC rings number analyzed $N_1 = 153$ for 1 cluster, $N_2 = 122$ for 2 clusters, $N_3 = 219$ for 3 clusters, $N_4 = 187$ for 4 clusters, $N_5 = 117$ for 5 clusters). For NPC quantification in the whole cell, the total number of NPC rings analyzed was $N = 1,460$ for Nup133 (Fig. 3i) and $N = 855$ for Nup107 (Supplementary Fig. 7). Cluster occurrence for NUP133 was estimated from the super-resolution images ($N = 1$ experiments; total number of NPC rings counted $N = 1,764$; Fig. 3i, inset).

The box plots (Fig. 1f,i and Supplementary Fig. 1b) show 25/75th percentile, the line is the median value, and the whiskers represent the s.d.

We performed a chi-square test to verify the matching of the data to a binomial distribution in all cases (Fig. 2j and Supplementary Fig. 1a,c).

Performances of the method and the correlation between estimated and actual values at varying statistics and stoichiometry have been characterized calculating the Pearson correlation coefficient R (Supplementary Fig. 4a,b,d,e).

The error bars on stoichiometry estimation correspond to the lower bound to the standard errors based on the Fisher Information Matrix (Fig. 2b,e,j, Fig. 3d–i, and Supplementary Figs. 6 and 7).

Data availability statement. Raw data for DNA origami calibration and validation have been provided. All other data are available upon request. Source data files for Figures 1 and 2 are available online.

21. Celli Zanacchi, F. et al. Protoc. Exch. http://dx.doi.org/10.1038/protoc.2017.089 (2017).
22. Lin, C., Perrault, S.D., Kwak, M., Graf, F. & Shih, W.M. Nucleic Acids Res. 41, e40 (2013).
23. Qiu, W. et al. Nat. Struct. Mol. Biol. 19, 193–200 (2012).
24. Szymborska, A. et al. Science 341, 655–658 (2013).