

A general method to quantify ligand-driven oligomerization from fluorescence-based images

Michael R. Stoneman, Gabriel Biener, Richard J. Ward, John D. Pediani, Dammar Badu, Annie Eis, Ionel Popa, Graeme Milligan and Valerică Raicu

Supplementary Note 1: Molecular brightness calculation. The molecular brightness, ϵ , can be extracted from the variance of the intensity distribution, σ^2 , when using analog detectors using the following relation¹:

$$\sigma^2 = \sigma_o^2 + SG\epsilon N_{oligo} + G^2\gamma\epsilon^2 N_{oligo} \quad (S1)$$

where G is the analog gain in digital levels /photon, S is the slope of the variance vs. intensity plot for a constant intensity source, N_{oligo} is the average number of particles (or oligomers) within the observation volume, γ is a shape factor which depends on the shape of the laser PSF as well as the geometry of the sample², and σ_o^2 is the readout noise variance of the detection electronics, i.e., the variance in the absence of signal.

Dividing each term in (S1) by the average background corrected intensity of the distribution, $\langle I_s \rangle = G\epsilon_{app}N_{oligo}$, gives:

$$\frac{\sigma^2}{\langle I_s \rangle} = \frac{\sigma_o^2}{\langle I_s \rangle} + S + G\gamma\epsilon \quad (S2)$$

In order to accurately extract the molecular brightness from a measured intensity histogram using an analog detector, we correct the measured distribution of intensities for the noise characteristics of the detector. The detector noise variance depends on the slope S , average background corrected intensity $\langle I_s \rangle$, and readout variance σ_o^2 as follows:

$$\sigma_D^2(\langle I_s \rangle) = \sigma_o^2 + S\langle I_s \rangle \quad (S3)$$

The signal variance arising due to the detector, σ_D^2 , was measured for our detector using a constant light source at varying intensities (see Supplementary note 4). Therefore, for any average intensity measured, we have the corresponding σ_D^2 value. Substitution of σ_D^2 into equation (S2) gives:

$$\frac{\sigma^2 - \sigma_D^2}{\langle I_s \rangle} = G\gamma\epsilon \quad (S4)$$

Dividing both sides of equation (S4) by the shape factor γ leads to the relation for the effective molecular brightness:

$$\epsilon_{eff} \equiv G\epsilon = \frac{\sigma^2 - \sigma_D^2}{\gamma\langle I_s \rangle}. \quad (S5)$$

In order to determine the total number of protomers within the beam area, we used the relationships

$$N_{oligo} = \frac{N_{proto}}{n}, \quad (S6)$$

$$\epsilon_{oligo} = n\epsilon_{eff}^{proto}, \quad (S7)$$

to derive the following relation:

$$N_{proto} = \frac{\langle I_s \rangle}{\epsilon_{eff}^{proto}}, \quad (S8)$$

where n is the average number of protomers comprising the oligomers being measured. The molecular brightness of single protomers, ϵ_{eff}^{proto} , was determined from separate measurements on a calibration sample known to be monomeric.

Supplementary Note 2: Determination of receptor concentrations from fluorescence signal level.

The rate of fluorescence per unit volume that is detected by the measurement system is directly proportional to the spatially dependent average molecular excitation rate, $\langle W(\vec{r}) \rangle$, multiplied by the molecular concentration $C(\vec{r}, t)$. The total measured fluorescence per unit time can be found by integrating this signal over all space, as follows:

$$F(t) = \kappa \iiint \langle W(\vec{r}) \rangle C(\vec{r}, t) dV \quad (\text{S9})$$

The parameter κ incorporates the quantum yield of the fluorophores as well as the detection efficiency of the measuring system. The time averaged molecular excitation rate of the fluorophore is given by multiplying the excitation probability of a single laser pulse by the laser repetition rate, f_p :

$$\langle W(\vec{r}) \rangle = f_p \int_0^{t_p} W(\vec{r}, t) dt \quad (\text{S10})$$

where t_p is the duration of a single laser pulse. In equation (S10), $W(\vec{r}, t)$ is the well known instantaneous rate of absorption of photon pairs, given by:

$$W(\vec{r}, t) = \sigma_2 P_0^2(t) PSF^2(\vec{r}) / 2 \quad (\text{S11})$$

where σ_2 is the two-photon absorption cross section, $P_0^2(t)$ the squared laser flux at the center of the focused laser beam, and $PSF^2(\vec{r})$ the laser distribution function representing the point-spread function of the focused laser beam. The laser point spread function can be well approximated by a Gaussian-Lorentzian spatial profile:

$$PSF(\vec{r}) = \frac{\omega_0^2}{\omega^2(z)} e^{-2\left(\frac{x^2+y^2}{\omega^2(z)}\right)} \quad (\text{S12})$$

where ω_0^2 defines the e^{-2} laser beam radius at the focal plane,

$$\omega^2(z) = \omega_0^2 \left[1 + \left(\frac{z}{z_R} \right)^2 \right], \quad (\text{S13})$$

and $z_R = \frac{\pi \omega_0^2}{\lambda}$ is the Rayleigh range, or the effective axial dimension of the excitation volume.

The total collected fluorescence intensity for a given measurement is found by integrating equation (S9) over the exposure time of the acquisition:

$$I = \int_0^T F(t) dt \quad (\text{S14})$$

If the integration time of the acquisition, T , is chosen to be much shorter than the fluctuation time scale of the particular process being studied, then we can assume that during an acquisition, the molecules are static, and hence the concentration is constant in time within the exposure. Under this condition of a short integration time, equation (S9) and (S14) can be combined to give:

$$I = \varepsilon \iiint PSF^2(\vec{r}) C(\vec{r}) dV \quad (\text{S15})$$

The parameter ε represents the average number of detected photons per sampling time per molecule, i.e. the molecular brightness, and is given as follows:

$$\varepsilon = \frac{\kappa \sigma_2 \langle I_0^2 \rangle T}{2} \quad (\text{S16})$$

For the case of a cell expressing fluorescently labeled proteins in the plasma membrane, we assume that all the fluorescent molecules are bound to the membrane and are together positioned within the focal plane of the laser beam, i.e., at $z = 0$. Therefore, there are no particles located at positions given by $z \neq 0$. The concentration of fluorescent molecules in equation (S15) is replaced by

$$C(\vec{r}) \equiv C_m(x, y, z) \delta(z), \quad (\text{S17})$$

where $C_m(x, y, z)$ is the area concentration of the membrane, and $\delta(z)$ is the Dirac delta function which has units of the inverse of its argument (in this specific case, 1/distance). For such a 2D distribution of molecules, equation (S15) becomes:

$$I_{mem} = \varepsilon \iiint PSF^2(\vec{r}) C_m(x, y, z) \delta(z) dx dy dz \quad (S18)$$

Using the sifting properties of the delta function, integration along the z axis gives $\iint PSF^2(x, y, 0) C(x, y, 0) \delta(z) dx dy$, and assuming that the distribution of fluorophores within the membrane region is uniform, $C_m \equiv C_m(x, y, 0)$ can be taken out of the area integral. Thus:

$$I_{mem} = \varepsilon C_m \iint PSF^2(x, y, 0) dx dy \quad (S19)$$

In order to extract the value of C_m from intensity measurements, we need to determine the molecular brightness of the fluorescent protein. This physical parameter combines properties of the molecule itself, along with properties inherent to the measuring system, as is seen from equation (S16). Two methods have been used to determine the molecular brightness of monomeric species and then the concentration of molecules in the sample.

The first method relies on separate measurements of uniform solutions of known concentration of the fluorophore of interest. From equation (S15), the measured intensity of a solution of uniform concentration is given by:

$$I_{sol} = C_{sol} \varepsilon \iiint PSF^2(\vec{r}) dV \quad (S20)$$

The fluorescence intensities of numerous solutions of varying molecular concentration were plotted against their corresponding concentration; the slope of this plot, Θ , represents the amount of fluorescence intensity detected per molecular concentration of fluorophore. Therefore, according to (S22), Θ is dependent on the molecular brightness of the fluorophores and the volume integral of the laser PSF:

$$\Theta = \varepsilon \iiint PSF^2(\vec{r}) dV \quad (S21)$$

With the measurable quantity Θ , and a model for the PSF, see equation (S12), the molecular brightness can be written:

$$\varepsilon = \frac{\Theta}{\iiint PSF^2(\vec{r}) dV} \quad (S22)$$

Plugging this relation for the molecular brightness into (S21):

$$I_{mem} = \Theta C_m \frac{\iint PSF^2(x, y, 0) dx dy}{\iiint PSF^2(\vec{r}) dV} \quad (S23)$$

Solving for the area concentration of the membrane gives:

$$C_m = \frac{I_{mem}}{\Theta \delta_{PSF}} \quad (S24)$$

Here the parameter δ_{PSF} has been introduced as follows:

$$\delta_{PSF} = \frac{\iint PSF^2(x, y, 0) dx dy}{\iiint PSF^2(\vec{r}) dV} \quad (S25)$$

A second method to extract protomer concentration within the membrane from the measured value of fluorescence intensity relies on obtaining the measured molecular brightness value of a monomeric standard which is localized to the plasma membrane, $\varepsilon_{eff}^{proto}$ (see Fig. 1). With this calibration, the concentration can be calculated again by inserting $\varepsilon_{eff}^{proto}$ into equation (S19):

$$C_m = \frac{I_{mem}}{\varepsilon_{eff}^{proto} \iint PSF^2(x, y, 0) dx dy} \quad (S26)$$

Supplementary Note 3: Image segmentation algorithm. The fluorescence intensity fluctuation spectrometry software suite contains two different segmentation methods: (i) the moving square method and (ii) an algorithm that has been previously used in the literature and is denoted as simple linear iterative clustering (SLIC)³⁻⁵. In the moving square method, a polygonal ROI, which has been

drawn on the fluorescence image by the user, is bordered by a rectangle that encompasses all vertices of the ROI polygon. The bordering rectangle is broken into a number of smaller squares, i.e. segments, with sides of length l_s pixels, where l_s is an input parameter to the algorithm. For each segment, only the pixels that also reside within the original ROI polygon are include in the intensity distribution used to calculate the molecular brightness and concentration of said segment. A number of the segment squares will have zero pixels that overlap with the ROI; these segments are, of course, not used for further analysis. There are also segments, close to the edge of the ROI polygon, which partially overlap with the ROI and therefore have less than l_s^2 usable pixels. In order to avoid including segments with a number of pixels which is insufficient for accurately extracting molecular brightness and concentration, we introduced a threshold value for the minimum number of overlapping pixels a given segment must contain in order to be included in the segments intensity distribution. The moving squares has the benefit of executing very quickly. However, because of the need for including a threshold value, there are some pixels along the edges of the polygon, which end up not being used.

In order to solve the drawback inherent to the moving squares method, we have made available in the software suite a segmentation method that is based on the SLIC algorithm. Initially, this algorithm breaks an image, again containing the polygonal ROIs drawn by the user, into smaller segment squares of length l_s pixels. The center of mass, (x_i, y_i) , of each segment is first calculated. Then, the center positions of each segment are iteratively changed by reassigning pixels to different segments based on a distance criteria, D_p , that takes intensity, I , into account along with the spatial distance between a given pixel, p , located at (x, y) and the nearest segment center. The general formula for D_p is given as follows:

$$D_p = \sqrt{\frac{(x-x_i)^2+(y-y_i)^2}{l_s^2} + \frac{(I-I_i)^2}{w_I^2}}, \quad (S27)$$

where w_I is a weighting factor that determines the amount of influence of the intensity. In our analysis we used $w_I = 1$. Explicitly, a pixel p would be reassigned to a new segment, j , only if the distance between p and the center of said segment (x_j, y_j) is shorter than the distance to the center of the current segment of which p is assigned. For this algorithm, the intensity values I are binary, and depend on whether the pixel is located within a region enclosed by a polygonal ROI. In other words, the intensity used for the calculation of distances in S1 is 1 if p is inside the ROI and 0 outside and can therefore be replaced by a Kronecker delta, δ_{xyI} . This simplifies D_p to:

$$D_p = \sqrt{\frac{(x-x_i)^2+(y-y_i)^2}{l_s^2} + \delta_{xyI}}; \quad \delta_{xyI} = \begin{cases} 1 & \text{if } p \text{ is inside ROI} \\ 0 & \text{if } p \text{ is outside ROI} \end{cases} \quad (S28)$$

The outline of the algorithm is provided below.

1. Segment image into squares of area l_s^2
2. Calculate center of mass of each segment; remove segments with centers located outside ROI.
3. Move segment centers away from large intensity gradient positions by shifting the center point one pixel into the ROI
4. **While** difference between center positions in step k and step $k-1$ is greater than threshold, do.
5. **For** each segment, i , centered at (x_i, y_i)
6. Calculate distances of each pixel within neighboring segments to segment centers (x_i, y_i) using (S2).
7. If distance is smaller than the recorded one for the specific pixel, refresh distance value and mark the pixel as the member of the i^{th} segment.
8. **End for**
9. Recalculate segments centers using all the pixels contained within a segment.
10. **End While**

Supplementary Note 4: Determination of the detector characteristics. The background corrected intensity, I_s , introduced in equation (2), represents the portion of the readout intensity attributed only to the source of light in the sample, e.g. the fluorescent molecules attached to proteins of interest. This value was obtained for each measurement by subtracting the intensity readout from the detector when there is no sample present, I_{back} , from the total signal readout from the detector in the region of interest, I_{meas} , as follows:

$$I_s = I_{meas} - I_{back} \quad (S29)$$

Contributions to I_{back} include the electronic offset (i.e. bias level) added to the output signal of the detector as well as dark noise. I_{back} was determined by measuring the mean intensity in multiple ~10,000 pixel subregions of the acquisition where there was no cells/fluorophores present. If the sample measured was uniformly fluorescent, e.g. a solution of fluorescent molecules, the mean background intensity was measured from multiple acquisitions where the laser was turned off. The parameters S and σ_o^2 of equation (S3) in the Methods section were determined from measurements of a light source with constant intensity (i.e. no temporal fluctuations), such that any fluctuations in detected signal are attributed to the detector readout. The relationship between the detector variance and intensity is linear and therefore could be fit with a straight line of slope S , and intercept σ_o^2 . From these two parameters, σ_D^2 could be calculated for any value of $\langle I_s \rangle$ and subtracted from the variance of the intensity distribution. It should be noted that the intensities measured from the constant light source must still be corrected for background signal, as shown in equation (S29). Slope and intercept values of $S=37.49$ intensity units and $\sigma_o^2 = 0$ were found for image sets acquired using single photon excitation and $S=76.3$ and $\sigma_o^2 = 13,400$ for images acquired using the two-photon excitation setup.

Supplementary Note 5: Design and synthesis of solutions of monomeric, dimeric and tetrameric constructs used to induce dimer, tetramer, and octamer formation through side-by-side YFP association. All fluorescent protein constructs used in experiments on fluorescent protein solutions were engineered in a modified pQE-80L expression vector, with a (His)₆tag at the N-terminus and a (Cys)₂ at the C-terminus. For the dimeric constructs, an eight-amino acid flexible linker (GGSGGSR) was added in-between two fluorescent proteins, to prevent any misfolding of the fusion proteins. The expression vectors were transformed into C41(DE3) cells (Sigma), which were grown at 37°C in the presence of carbenicillin up to an OD⁶⁰⁰ of 0.7. The protein overexpression was induced with 1 mM IPTG overnight at 25°C. Cells were resuspended in sodium phosphate buffer pH 7.0, 300 mM NaCl (EW buffer), and disrupted using lysozyme and sonication. Following separation of the soluble fraction, the proteins were immobilized on a Ni-NTA affinity column (GE Healthcare) via the polyhistidine tag, washed with the same EW buffer with 7.5 % Imidazole and eluted with EW buffer containing 250 mM imidazole. Finally, the proteins were passed through a size-exclusion chromatography column (Superdex-200 10/300 GL column GE Healthcare) and the main fraction was collected in HEPES buffer 10 mM, 150 mM NaCl, pH 7.2.

Several different protein constructs were engineered: monomeric EGFP (denoted mEGFP here, and incorporating the non-dimerizing A206K mutation⁶), dimerization-prone EGFP (without the A206K mutation), dimerizing YFP, and dimerizing tandem YFP dimers ((YFP)₂, where the YFPs are linked by the flexible amino acid sequence GGSGGSR). The two Cysteins added to the C-terminus played a crucial role in our experiments on fluorescent protein solutions, in that they allowed slow end-to-end association via disulfide bridges of (YFP)₂ dimers in the presence of small amounts of oxygen normally dissolved in the aqueous solvent. This led to about 20% of the dimers forming disulfide

bridges, and hence irreversible tetramers (referred to as “(YFP)₂ss(YFP)₂” in Supplementary Fig. 1 and the main text). Using non-A206K YFP domains expressed as dimers allowed us to induce formation of reversible tetramers via side-by-side interactions at two binding loci (i.e., represented by Alanine 206) on the barrel of each of the two YFP molecules in the (YFP)₂ss(YFP)₂ tandem dimers, as well as formation of octamers via four side-by-side interactions at four binding loci on the barrels of YFP in the (YFP)₂ss(YFP)₂ tetramers. Addition of 1mM Dithiothreitol (DTT) led to complete dissociation of the disulfide bridges, so that only (YFP)₂ and side-by-side tetramers were present in the (YFP)₂ solution. Taken together, these different solutions of YFP and tandem YFP provided means to test the feasibility of using fluorescence fluctuation analysis to determine various oligomeric sizes.

Supplementary Note 6: Model fitting for the fluorescent protein solution

measurements. Figure 1(b) provides the plot of average ε_{eff} vs. concentration for solutions containing yellow fluorescent proteins (YFP) or YFP tandem dimers, (YFP)₂, in the presence of 1 mM Dithiothreitol (DTT) to dissociate any existing disulfide bridges, as well as a (YFP)₂ss(YFP)₂ solutions minus the addition of DTT. The experimental curves were fitted to theoretical model based on the Law of Mass Action, as described herein.

The formation of dimers through side-by-side binding of individual YFP monomers (see Note 5) in the presence of DTT is described by the following reaction:



The total concentration of the YFP molecules (i.e., protomers) in monomeric as well as dimeric forms is given by the expression:

$$C = [m] + 2[d] \quad (S30)$$

where $[m]$ is the concentration of the monomeric YFP, $[d]$ is the concentration of the side-by-side dimer $\begin{matrix} YFP \\ YFP \end{matrix}$. The Law of Mass Action states that in a steady state the following relation holds:

$$K_d = \frac{[m]^2}{[d]}, \quad (S31)$$

where K_d is the dimeric dissociation constant. By combining Eqs. (S31) and (S30) and solving for $[m]$, we obtain:

$$[m] = \frac{K_d}{4} \left(\sqrt{1 + \frac{8C}{K_d}} - 1 \right). \quad (S32)$$

By inserting $[d]$ from Eq. (S31) into the customary effective brightness expression⁷,

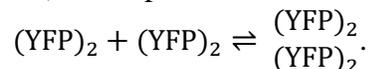
$$\varepsilon_{eff} = \frac{\varepsilon_m^2[m] + \varepsilon_d^2[d]}{\varepsilon_m[m] + \varepsilon_d[d]}, \quad (S33)$$

with $\varepsilon_m = \varepsilon_{eff}^{proto}$ and $\varepsilon_d = 2\varepsilon_{eff}^{proto}$, we obtain:

$$\varepsilon_{eff} = \frac{\varepsilon_m^2[m] + 4\varepsilon_m^2[d]}{\varepsilon_m[m] + 2\varepsilon_m[d]} = \varepsilon_{eff}^{proto} \left(2 - \frac{[m]}{C} \right), \quad (S34)$$

where $[m]$ is computed from equation (S32).

For the second solution of YFP, the formation of tetramers through side-by-side binding of YFP duplexes (i.e., (YFP)₂, see Note 5) in the presence of DTT is described by the following reaction:



Following a similar path as above, we have

$$C = 2[d] + 4[t], \quad (S35)$$

for the total concentration of YFP protomers present in solution in dimeric and tetrameric forms, and

$$K_t = \frac{[d]^2}{[t]}, \quad (\text{S36})$$

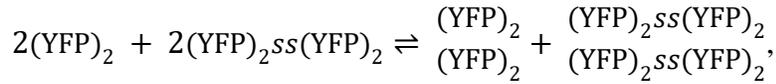
for the dissociation constant. Solving these equations for the dimer concentration, we obtain

$$[d] = \frac{K_t}{4} \left(\sqrt{1 + \frac{4C}{K_t}} - 1 \right). \quad (\text{S37})$$

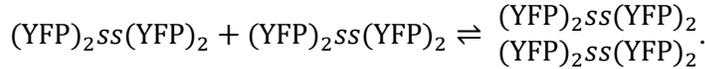
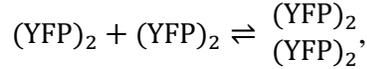
Using an expression similar to (S33) for the effective brightness of the solution, as well as $\varepsilon_d = 2\varepsilon_{eff}^{proto}$, $\varepsilon_t = 4\varepsilon_{eff}^{proto}$, we obtain

$$\varepsilon_{eff} = \frac{\varepsilon_d^2[d] + \varepsilon_t^2[t]}{\varepsilon_d[d] + \varepsilon_t[t]} = 4\varepsilon_{eff}^{proto} \left(1 - \frac{[d]}{c} \right). \quad (\text{S38})$$

As discussed in Note 5 above, the solution of YFP duplexes, $(\text{YFP})_2$, in the absence of DTT also contains a fraction, α , of tetrameric constructs (or tetraplexes) formed by end-to-end fusion via disulfide bridges of $(\text{YFP})_2$ duplexes. Therefore, the side-by-side interactions between YFP protomers leads to formation of tetramers and octamers. The reaction leading to the two oligomeric species,



may be decoupled, to a first approximation (i.e., by ignoring cross-interactions between duplexes and tetraplexes), into two separate reactions:



An interaction described by the first reaction has been dealt with above, and gives here:

$$2[d] + 4[t_p] = (1 - \alpha)C, \quad (\text{S39})$$

$$[t_p] = \frac{[d]^2}{K_t}, \quad (\text{S40})$$

$$[d] = \frac{K_t}{4} \left(\sqrt{1 + \frac{4(1-\alpha)C}{K_t}} - 1 \right) \quad (\text{S41})$$

where $[t_p]$ is the concentration of “parallel” tetraplexes formed through side-by-side interactions between $(\text{YFP})_2$ duplexes, and the total concentration of protomers in dimeric form is now replaced by $(1 - \alpha)C$, with $1 - \alpha$ being the fraction of duplexes unbound by disulfide bridges. The second reaction occurring within this fluorescent protein solution, and describing the association of $(\text{YFP})_2SS(\text{YFP})_2$ (i.e., “linear”) tetraplexes of concentration $[t_l]$ into side-by-side octamers of concentration $[o]$, is characterized by the following relations:

$$4[t_l] + 8[o] = \alpha C, \quad (\text{S42})$$

$$K_o = \frac{[t_l]^2}{[o]}, \quad (\text{S43})$$

$$[t_l] = \frac{K_o}{4} \left(\sqrt{1 + \frac{2\alpha C}{K_o}} - 1 \right). \quad (\text{S44})$$

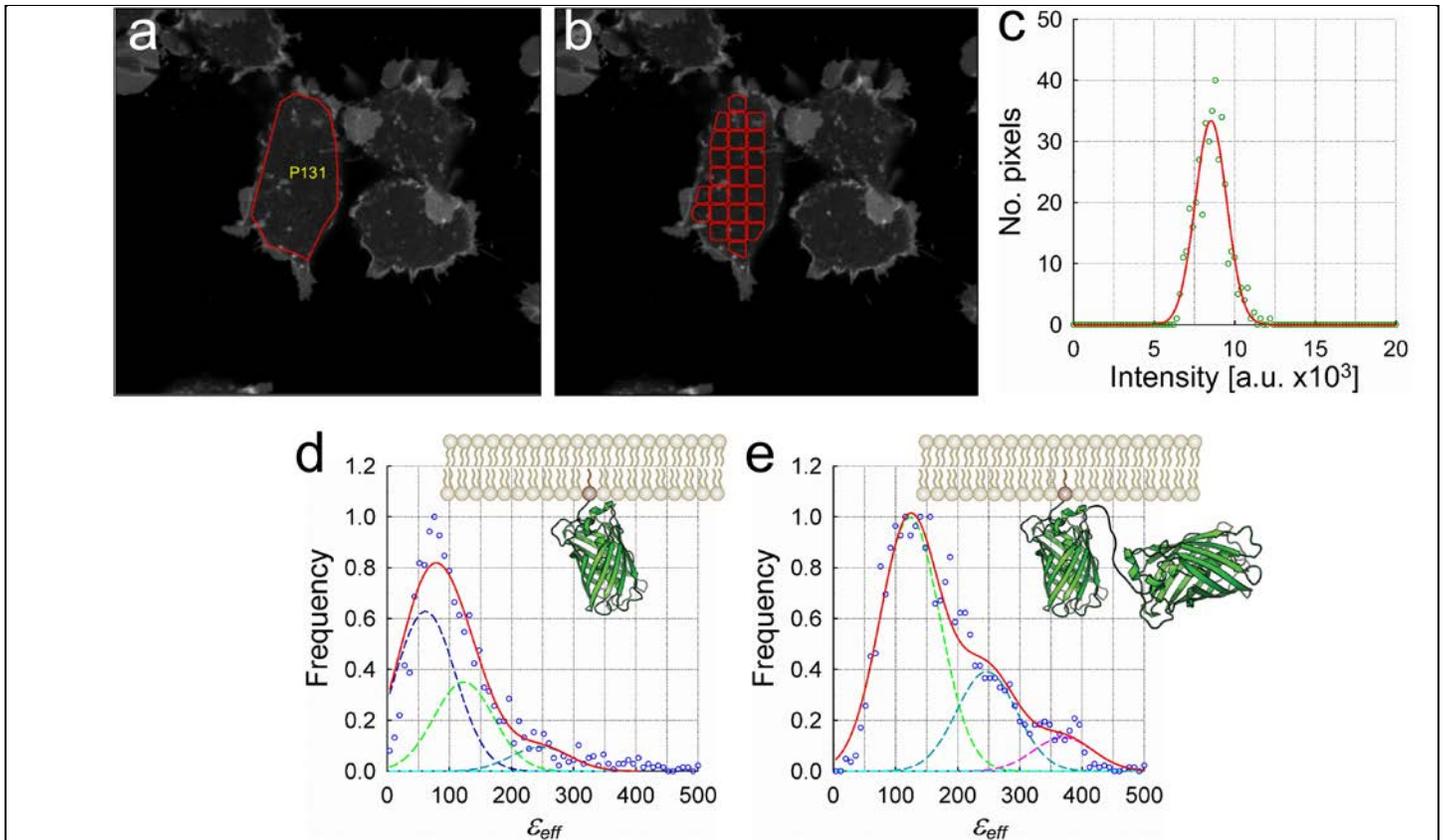
Using equations (S39) and (S42), as well as $\varepsilon_d = 2\varepsilon_{eff}^{proto}$, $\varepsilon_t = 4\varepsilon_{eff}^{proto}$, and $\varepsilon_o = 8\varepsilon_{eff}^{proto}$ the effective brightness of the solution is

$$\varepsilon_{eff} = \frac{\varepsilon_d^2[d] + \varepsilon_t^2\{[t_p] + [t_l]\} + \varepsilon_o^2[o]}{\varepsilon_d[d] + \varepsilon_t\{[t_p] + [t_l]\} + \varepsilon_o[o]} = 4\varepsilon_{eff}^{proto} \left(1 + \alpha - \frac{[d]}{c} - \frac{4[t_l]}{c} \right). \quad (\text{S45})$$

Equations (S34), (S38) and (S45) have been used to fit the experimental data shown in Supplementary Fig. 1, with α , ε_m , K_d , K_t , and K_o as adjustable parameters.

References

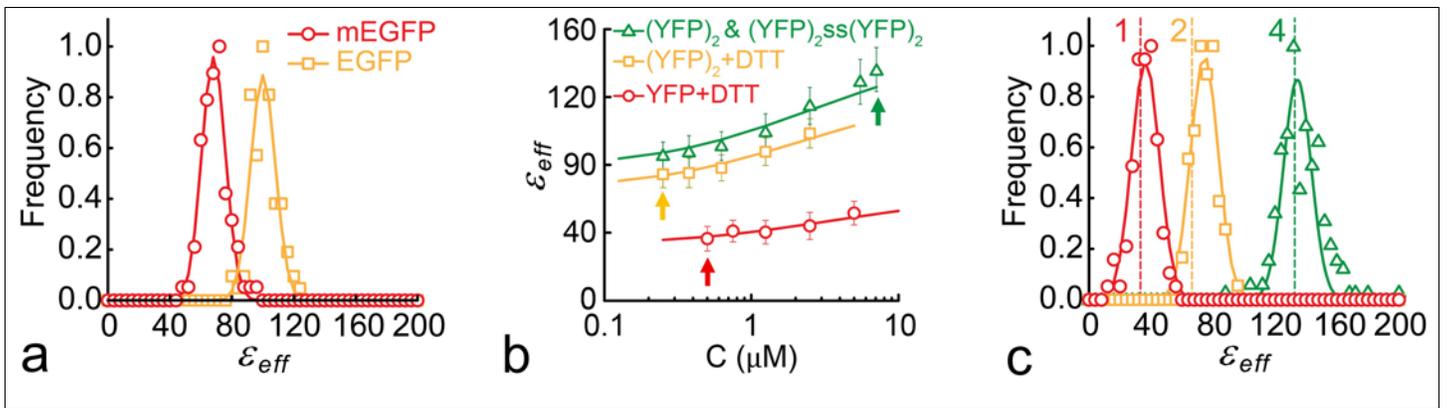
- 1 Unruh, J. R. & Gratton, E. Analysis of molecular concentration and brightness from fluorescence fluctuation data with an electron multiplied CCD camera. *Biophysical journal* **95**, 5385-5398 (2008).
- 2 Chen, Y., Muller, J. D., So, P. T. & Gratton, E. The photon counting histogram in fluorescence fluctuation spectroscopy. *Biophysical journal* **77**, 553-567 (1999).
- 3 Achanta, R. *et al.* SLIC Superpixels Compared to State-of-the-Art Superpixel Methods. *Ieee T Pattern Anal* **34**, 2274-2281 (2012).
- 4 Achanta, R. *et al.* SLIC Superpixels. *EPFL Technical Report*, 149300 (2010).
- 5 Zhang, Y. X., Li, X. M., Gao, X. F. & Zhang, C. M. A Simple Algorithm of Superpixel Segmentation With Boundary Constraint. *Ieee T Circ Syst Vid* **27**, 1502-1514 (2017).
- 6 Zacharias, D. A., Violin, J. D., Newton, A. C. & Tsien, R. Y. Partitioning of lipid-modified monomeric GFPs into membrane microdomains of live cells. *Science* **296**, 913-916 (2002).
- 7 Chen, Y., Wei, L.N. & Muller, J. D. Probing protein oligomerization in living cells with fluorescence fluctuation spectroscopy. *P Natl Acad Sci USA* **100**, 15492–15497 (2003).



Supplementary Figure 1

Illustration of the data reduction process in two-dimensional Fluorescence Fluctuation Spectrometry (2D-FFS) using two-photon excitation.

a, Typical fluorescence image obtained from two-photon excitation of Flp-In™ T-REx™ cells expressing fluorescently labeled plasma membrane targeted mEGFP construct (PM-1-mEGFP), and an overlaid polygon (P131) indicating a region of interest (ROI) which comprises a patch of the basolateral membrane of a cell. **b**, Software-generated image segmentation of the ROI in **(a)** using the Moving Square method (see Methods and Supplementary Note 3). **c**, A fluorescence intensity histogram (green circles) of an image segment selected at random, alongside the Gaussian curve (solid red line) used to fit the experimental histogram by adjusting the mean and width of the Gaussian. The intensity binning was set to 25 intensity counts (in arbitrary units). **d-e**, Normalized frequency distribution obtained from the **(d)** PM-1-mEGFP expressing cells (2,803 total segments) was simultaneously fit (solid red line) along with a distribution **(e)** constructed similarly from measurements of cells expressing dimeric, tandem linked mEGFP constructs (2,832 total segments) using a sum of Gaussian functions in order to find brightness of single protomers of mEGFP, $\epsilon_{eff}^{proto} = 61.4$, when measured using the two-photon optical micro-spectroscopy. **These results were confirmed using at least two additional sets of experiments.**



Supplementary Figure 2

Comparison of ϵ_{eff} distributions obtained, using two-photon excitation, from solutions of monomers, dimers, tetramers and octamers of fluorescent proteins.

a, Frequency distribution of ϵ_{eff} for solutions of monomeric EGFP monomers containing an oligomerization-inhibiting mutation (A206K) as well as EGFP without that mutation, which is prone to self-association. The average laser power was 15 mW and the integration time was 100 μ s per pixel. Each distribution was fit with a Gaussian function to obtain the mean (μ) and standard deviation (SD) of the distribution. Best-fit parameters for mEGFP (red line) were $\mu = 68.3$ and $SD = 7.7$, and for EGFP (yellow line) were $\mu = 100.1$ and $SD = 8.0$.

b, Average ϵ_{eff} vs. concentration for solutions containing **(i)** yellow fluorescent protein monomers (YFP) in the presence of 1 mM Dithiothreitol (DTT) (data labeled as “YFP + DTT”), **(ii)** YFP dimers or duplexes, $(YFP)_2$, treated with 1 mM Dithiothreitol (data labeled as “ $(YFP)_2 + DTT$ ”) to remove any disulfide bonds formed between the dimers (see Supplementary Note 5), and **(iii)** a mixture consisting of $(YFP)_2$ dimers plus $(YFP)_2$ dimers fused to each other end to end (i.e., through their termini) via disulfide bonds to form $(YFP)_2ss(YFP)_2$ tetramers (data labelled as “ $(YFP)_2$ & $(YFP)_2ss(YFP)_2$ ”). Data points and error bars represent $\mu \pm SD$ obtained by fitting brightness distributions with single Gaussian functions, similar to **(a)**, and as exemplified in **(c)**. As the concentration of molecules increased, self-association via side-by-side interactions between fluorescent proteins occurred, which led to dimer, tetramer, and octamer formation.

From the fitting of each curve with an appropriate theoretical model, the best-fit parameter values have been estimated as follows (see Supplementary Note 6). For “YFP + DTT,” i.e., $YFP + YFP \xrightleftharpoons{K_d} YFP_2$: $K_d = 5.8 \mu M$, $\epsilon_{eff}^{proto} = 33.3$; for “ $(YFP)_2 + DTT$,” i.e., $(YFP)_2 +$

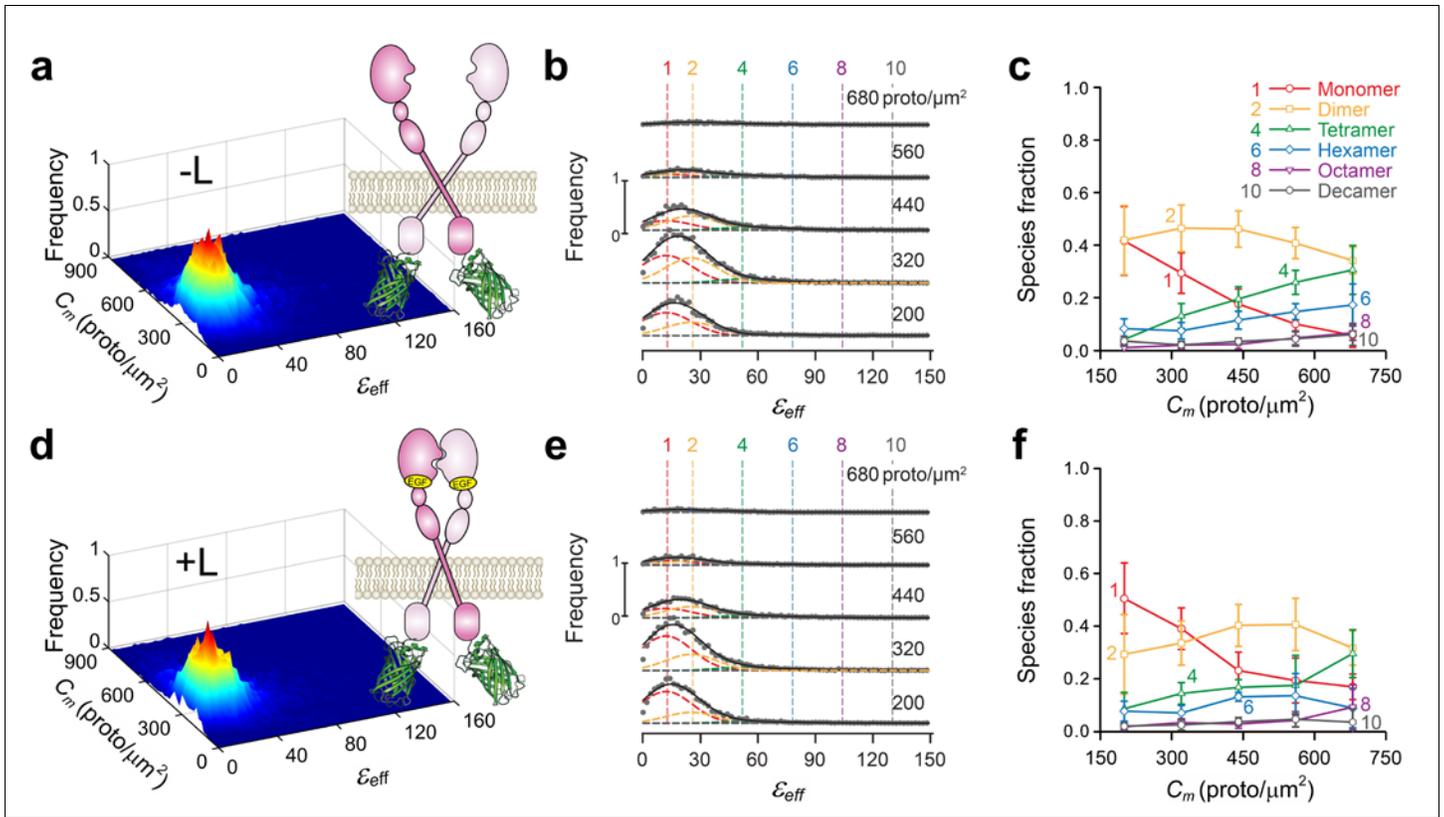
$(YFP)_2 \xrightleftharpoons{K_t} (YFP)_2$: $K_t = 3.6 \mu M$, $\epsilon_{eff}^{proto} = 33.2$; for “ $(YFP)_2$ & $(YFP)_2ss(YFP)_2$ ” the model accounts for two quasi-independent reactions,

$(YFP)_2 + (YFP)_2 \xrightleftharpoons{K_t} (YFP)_2$ and $(YFP)_2ss(YFP)_2 \xrightleftharpoons{K_o} (YFP)_2ss(YFP)_2$, and gives $K_t = 3.4 \mu M$, $K_o = 0.4 \mu M$, $\epsilon_{eff}^{proto} = 33.2$, and the fraction of

YFP protomers within $(YFP)_2ss(YFP)_2$ tetramers or octamers, $\alpha = 20\%$. The agreement between the ϵ_{eff}^{proto} values determined independently from the three experimental curves as well as between the two K_t values is indeed remarkable, confirming the validity of our measurement method. **c**, Brightness distributions corresponding to the single points indicated by vertical arrows in **(b)**. The vertical dashed lines indicate ϵ_{eff} values of 1x, 2x, and 4x the brightness of monomeric YFP (i.e., ϵ_{eff}^{proto}), as determined from fitting the

binding curves in **(b)**, and which correspond to the brightness values of monomers, dimers, and tetramers, respectively. Sample size, i.e., number of measurements in each data set (n).

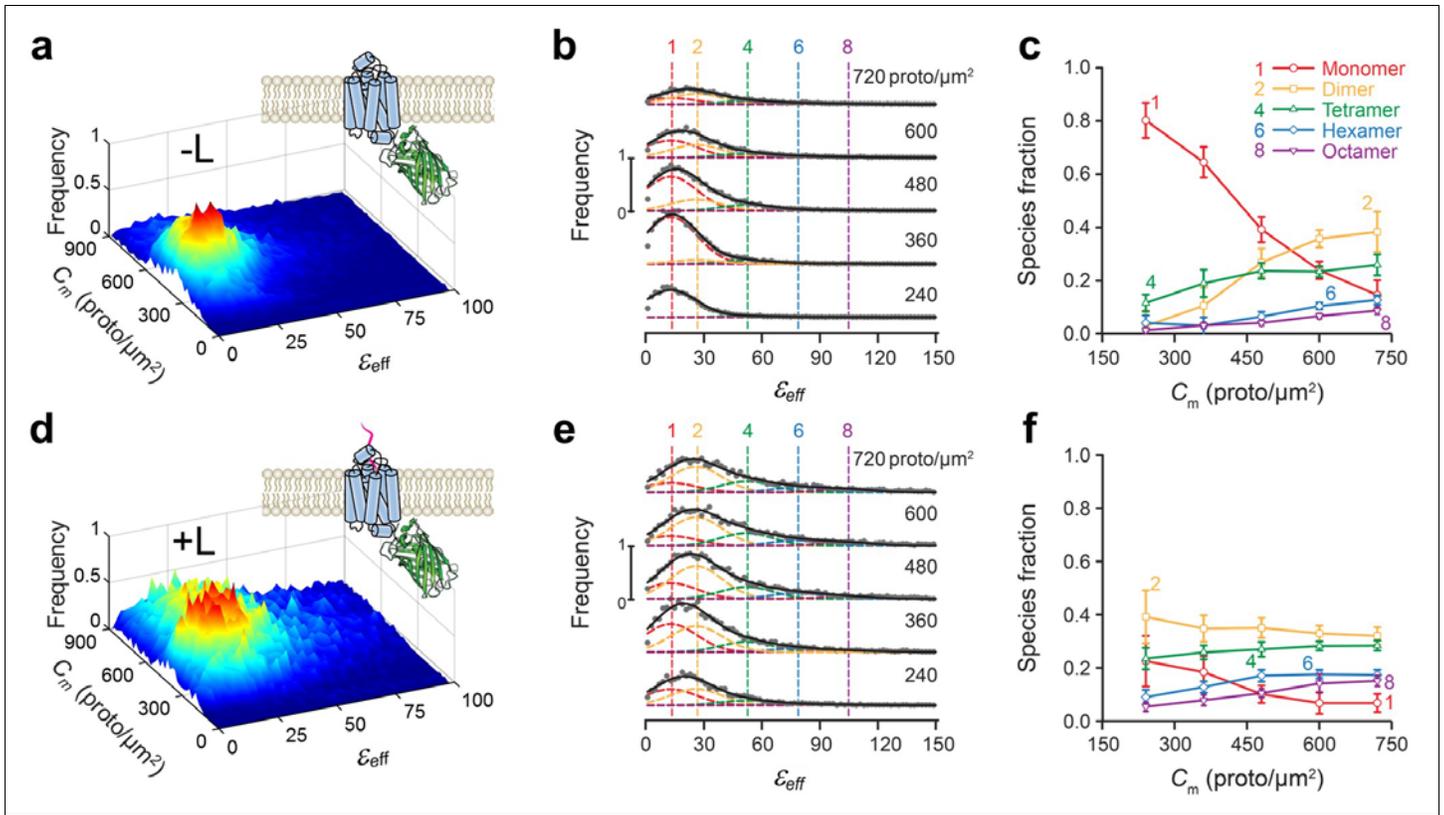
Similar results were obtained from several additional experiments using mEGFP, YFP, as well as other fluorescent protein variants (such as mCitrine).



Supplementary Figure 3

2D-FIF results obtained from single-photon excitation of Flp-In™ T-REx™ cells expressing the Tyr²⁵¹Ala,Arg²⁸⁵Ser mutant of the epidermal growth factor receptor (EGFR) in the absence of ligand (-L) or after ten-minute treatment with ligand (+L).

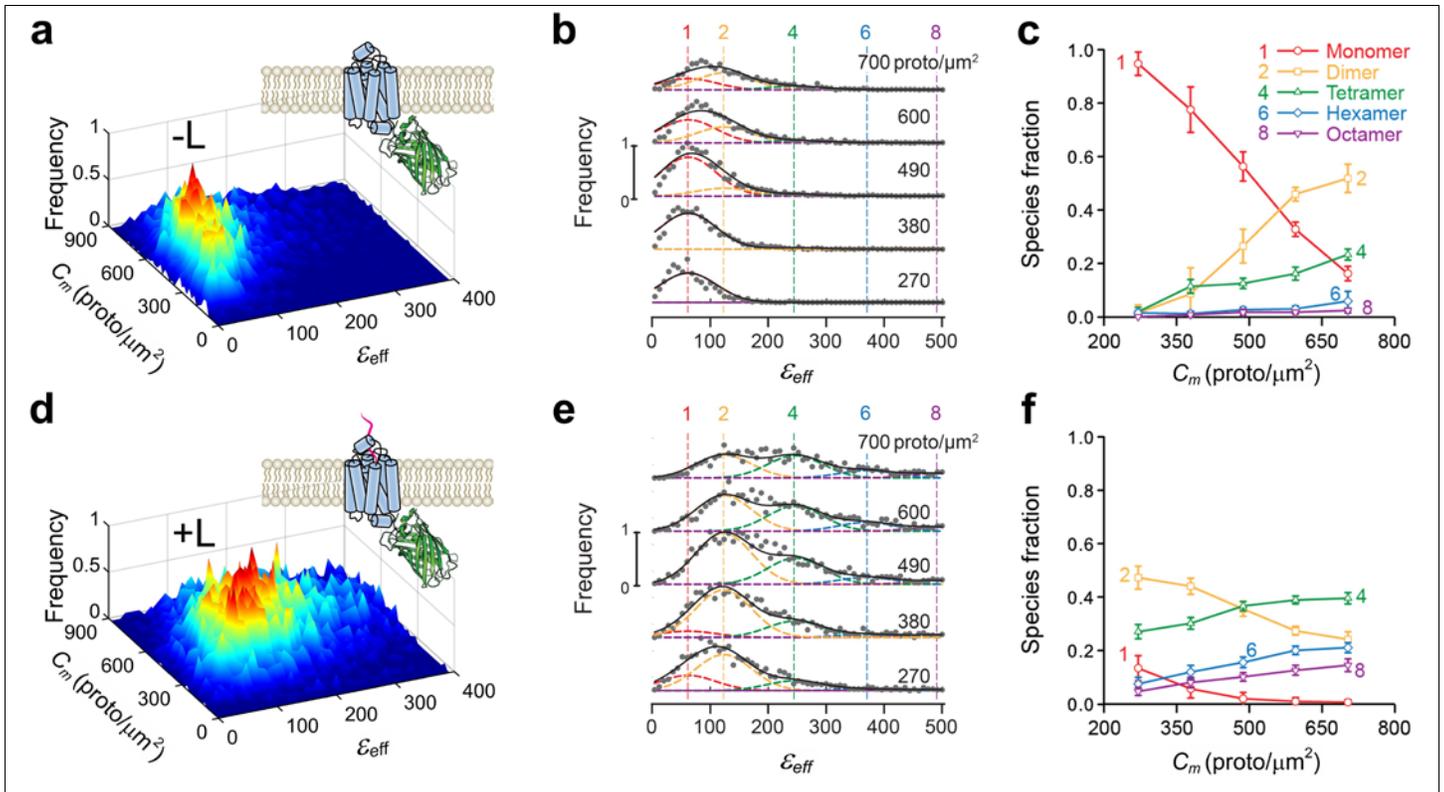
a,d (column 1), Surface plots of the frequency of occurrence of effective brightness (ϵ_{eff}) of each protomer concentration using **(a)** 22,283 and **(d)** 19,736 total segments to construct the distribution, extracted from 48 and 36 images, respectively (each consisting of several cells). The data were collected from at least two separate experiments. **b,e (column 2)**, Stacks of cross sections through the surface plots in panels **(a)** and **(d)** for different concentration ranges; average concentration for each range (in protomer/ μm^2) is indicated above each plot. Vertical dashed lines indicate the peak positions for the brightness spectra of monomers, dimers, etc., obtained from analysis of the brightness spectrograms of monomeric and tandem mEGFP (see main text), which were used as standards of brightness in the analysis. **c,f (column 3)**, Relative concentration of protomers in each oligomeric species vs. total protomer concentration, as derived from unmixing of the curves in column 2 into Gaussian components. Each data point and its error bar represent the mean \pm standard deviation, respectively, of 1,500 different relative fraction values resulting from bootstrapping and refitting the original set of images as described in the Methods section. The ϵ_{eff} distribution for each concentration range was fitted with a sum of six Gaussians; the peak of each Gaussian was set to a value of $n\epsilon_{eff}^{proto}$, where n is an integer denoting the number of protomers in a given oligomer size (e.g., 1, 2, 4, etc.), while the standard deviations (SD) were fixed. The ϵ_{eff}^{proto} and SD values were obtained from brightness measurements on cells expressing monomeric or dimeric forms of the fluorescent protein mEGFP (see Fig. 1 and Methods). Only the Gaussian amplitudes (A_n) were adjusted in the process of data fitting in **(b)**, which gave the fraction of protomers (shown in **c**) corresponding to each oligomeric species via $n_i A_i / \sum_n n A_n$.



Supplementary Figure 4

2D-FIF results obtained from single-photon excitation of CHO cells expressing wild-type secretin receptor in the absence of ligand (-L) or after ten-minute treatment with 100 nM agonist ligand (+L).

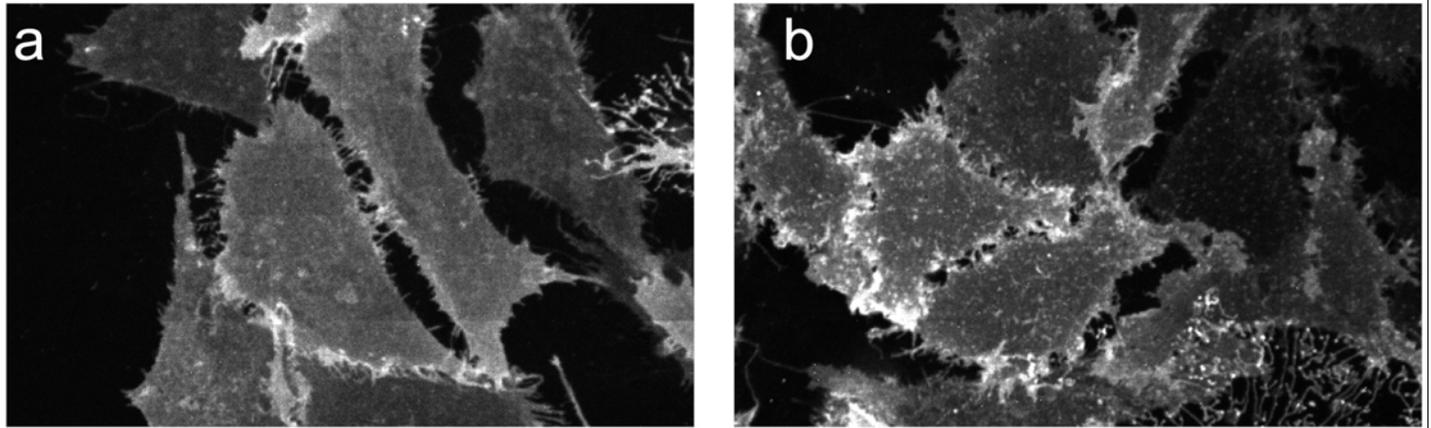
a,d (column 1), Frequency of occurrence of effective brightness (ϵ_{eff}) for each protomer concentration using **(a)** 64,619 and **(d)** 29,839 total segments to construct the distribution, extracted from 113 and 42 images, respectively (each of which contain several cells). The data were collected from at least two separate experiments. **b,e (column 2)**, Stacks of cross sections through the surface plots in panels **(a)** and **(d)**; average concentration for each range (in protomer/ μm^2) is indicated above each plot. Vertical dashed lines indicate peak positions for the brightness spectra of monomers, dimers, etc., obtained from (or predicted from) the simultaneous fitting of the PM-1- and PM-2-mEGFP spectrograms used as standards of brightness (see caption to Fig. 2). **c,f (column 3)**, Relative concentration of protomers within each oligomeric species vs. total concentration of protomers, as derived from unmixing of the curves in column 2 into different Gaussian components. Samples were as follows: wild-type secretin receptor treated with vehicle (-L) **(a-c)** or secretin (+L) for 10 minutes **(d-f)**. Each data point and its error bar represent the mean \pm standard deviation, respectively, of 1,500 different relative fraction values resulting from bootstrapping and refitting the original set of images as described in the Methods section. The ϵ_{eff} distribution for each concentration range was fitted with a sum of six Gaussians; the peak of each Gaussian was set to a value of $n\epsilon_{eff}^{proto}$, where n is an integer denoting the number of protomers in a given oligomer size (e.g., 1, 2, 4, etc.), while the standard deviations (SD) were fixed. The ϵ_{eff}^{proto} and SD values were obtained from brightness measurements on cells expressing monomeric or dimeric forms of the fluorescent protein mEGFP (see Fig. 1 and Methods). Only the Gaussian amplitudes (A_n) were adjusted in the process of data fitting in **(b)**, which gave the fraction of protomers (shown in **c**) corresponding to each oligomeric species via $n_i A_i / \sum_n n A_n$.



Supplementary Figure 5

2D-FIF results obtained from two-photon excitation of CHO cells expressing wild-type secretin receptor in the absence of ligand (-L) or treatment with ligand (+L) for non-fixed cells.

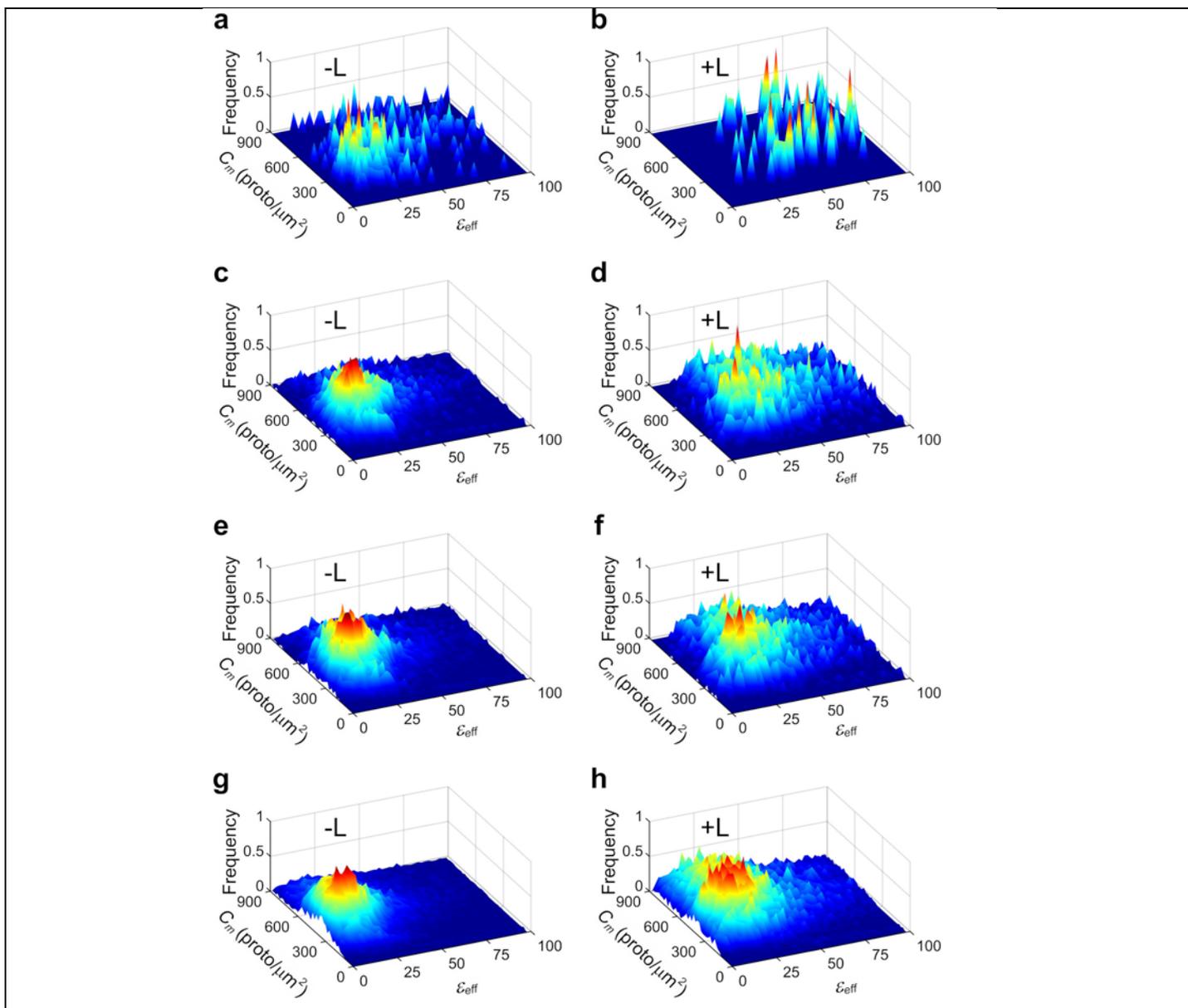
a,d (column 1), Surface plots of the frequency of occurrence of ϵ_{eff} for each concentration of protomers using **(a)** 11,251 and **(d)** 10,282 total segments to construct the distribution, extracted from 60 images for each row (each image consisting of several cells). The data were collected from at least two separate experiments. **b,e (column 2)**, Stacks of cross sections through the surface plots in panels **(a)** and **(d)**, i.e., frequency of occurrence vs. effective brightness for different concentration ranges; average concentration for each range (in protomer/ μm^2) is indicated above each plot. Vertical dashed lines indicate the peak positions for the brightness spectra of monomers, dimers, etc., obtained from analysis of the brightness spectrograms of monomeric and tandem EGFP (see main text), which were used as standards of brightness in the analysis. **c,f (column 3)**, Relative concentration of protomers within each oligomeric species vs. total concentration of protomers, as derived from unmixing of the curves in column 2 into different Gaussian components. Measurements were performed on cells for a range of 10-40 minutes after treatment with either the vehicle or secretin. Each data point and its error bar represent the mean \pm standard deviation, respectively, of 1,500 different relative fraction values resulting from bootstrapping and refitting the original set of images as described in the methods section. The ϵ_{eff} distribution for each concentration range was fitted with a sum of six Gaussians; the peak of each Gaussian was set to a value of $n\epsilon_{eff}^{proto}$, where n is an integer denoting the number of protomers in a given oligomer size (e.g., 1, 2, 4, etc.). The, ϵ_{eff}^{proto} and SD values were obtained from brightness measurements on cells expressing monomeric or dimeric forms of the fluorescent protein mEGFP (see Supplementary Fig. 2 and Methods). Only the Gaussian amplitudes (A_n) were adjusted in the process of data fitting in **(b)**, which gave the fraction of protomers (shown in **c**) corresponding to each oligomeric species via $n_i A_i / \sum_n n A_n$.



Supplementary Figure 6

Illustration of agonist-induced concentration of wild type secretin into small focal regions in the membrane for 30-minute treated cells.

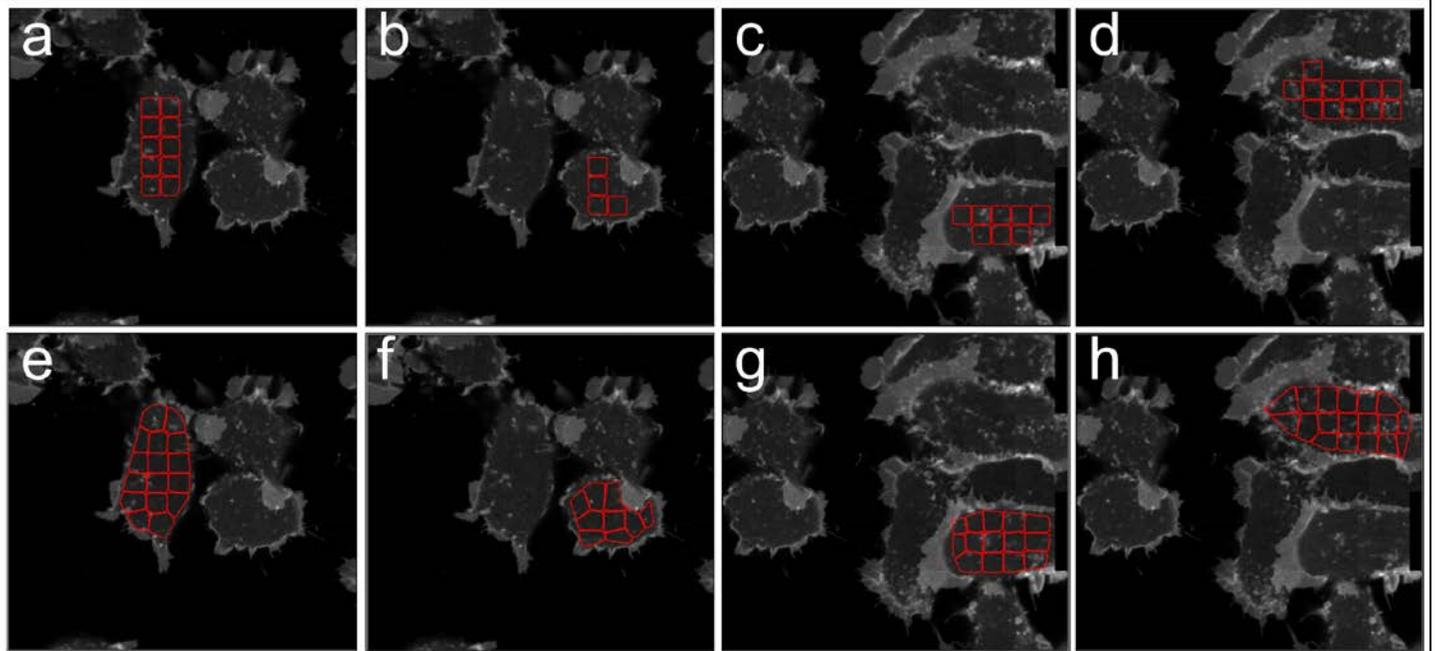
Typical fluorescence images of CHO cells expressing wild-type secretin receptor treated with **(a)** vehicle for 10 minutes or **(b)** secretin for 30 minutes. Note the “punctated” appearance of membranes in **(b)**, following long-term (i.e., 30 minutes) treatment with agonist, which was caused by accumulation of secretin receptors into clathrin-coated pits and endocytic vesicles. This gave rise to inhomogeneous concentration distributions even within the <500 pixel segments used to generate ϵ_{eff} distributions. Such inhomogeneous distribution caused the effective brightness of each ROI segment to take on a continuum of values, likely leading to the observed smearing of the brightness spectrograms (see Fig. 3h).



Supplementary Figure 7

Comparison of the effect of segment size on visualization and quantification of ϵ_{eff} for the data presented in Supplementary Fig. 4.

a-h, Surface plots of the frequency of occurrence of ϵ_{eff} for each concentration of protomers are shown for CHO cells expressing wild-type secretin receptor. The same dataset, secretin receptor treated with vehicle (-L), is presented in all four plots of **column 1 (a,c,e,g)**. Likewise, in **column 2 (b,d,f,h)**, data obtained from cells expressing secretin receptor and treated with ligand (+L) for 10 minutes is displayed in all four graphs. The difference between the plots in a column is the image segment size used to extract effective brightness and concentration values from the cell images. The typical segment size for each of the four rows of graphs is as follows: **a,b** - entire cells; **c,d** - 5000 pixel²; **e,f** - 2000 pixels²; and **h,i** - 500 pixel². The figure illustrates that the size of the segment where intensity distributions are being extracted from affects not only the smoothness of the distribution, but also very dramatically the width of the distribution.



Supplementary Figure 8

Comparison between two different methods of segmentation for Flp-In™ T-REx™ cells cells expressing fluorescently labeled plasma membrane targeted mEGFP construct

a-d, Software-generated image segmentation of different regions using the Moving Square method (see Supplementary Note 3). **e-h**: Software-generated image segmentation of the same ROIs selected for **a-d** using the SLIC method (see Supplementary Note 3). **Similar comparisons were made using hundreds of additional cells from different experiments.**