AUTOMATING SINGLE SUBUNIT COUNTING OF MEMBRANE PROTEINS IN MAMMALIAN CELLS*

Hugo McGuire[‡,§,1], Mark R. P. Aurousseau[‡,ll,1], Derek Bowie[‡,ll] and Rikard Blunck[‡,§,¶,2]

[‡]Groupe d'Étude des Protéines Membranaires (GÉPROM), Departments of [§]Physics and [¶]Physiology, Université de Montréal, Montréal, QC, H3C 3J7 and
[ll]Department of Pharmacology and Therapeutics, McGill University, Montréal, QC, H3G 0B1, Canada

[1] *Both authors contributed equally to this work*

Running Title: Automated Subunit Counting of Membrane Proteins

To whom correspondence should be addressed: Dr. Rikard Blunck, Département de physique, C.P. 6128 succ. Centre-ville, Université de Montréal, Montréal, QC, H3C 3J7, Email: rikard.blunck@umontreal.ca

Key words: single-molecule; step detection; ionotropic glutamate receptors; cys-loop receptors; superfolder GFP

Background: Although powerful, single subunit counting is time consuming, prone to user-bias and largely restricted to Xenopus expression.
Results: *PIF* is an automated analysis program that identifies subunit stoichiometry of any fluorescently-tagged membrane protein from TIRF-recordings.
Conclusion: *PIF* is accurate to more than 90% even in noisy data typical for mammalian expression system.
Significance: *PIF* approach is generalizable to any membrane protein and TIRF-microscope.

SUMMARY

Elucidating subunit stoichiometry of neurotransmitter receptors is preferably carried out in a mammalian expression system where the rules of native protein assembly are strictly obeyed. Although successful in *Xenopus* oocytes, single subunit counting – manually counting photobleaching steps of GFP-tagged subunits – has been hindered in mammalian cells by high background fluorescence, poor control of expression and low GFP maturation efficiency. Here, we present a fully-automated single-molecule fluorescence counting method that separates tagged proteins on the plasma membrane from background fluorescence and contaminant proteins in the cytosol or ER and determines the protein stoichiometry. Lower GFP maturation rates observed in cells cultured at 37°C were partly offset using a monomeric version of superfolder GFP. We were able to correctly identify the stoichiometry of GluK2 and α1 glycine receptors. Our approach permits the elucidation of stoichiometry for a wide variety of plasma-membrane proteins in mammalian cells with any commercially available TIRF microscope.

INTRODUCTION

The vast majority of membrane-bound proteins are expressed as oligomeric complexes in either homo- or heteromeric forms. Given this complexity, a number of biochemical and imaging techniques have been employed to elucidate subunit number and stoichiometry. As a first approximation, stoichiometry is often assessed by bulk biochemical methods, such as blue native polyacrylamide gel electrophoresis (1). However for a quantitative estimate, more recent approaches have turned to spectroscopic techniques such as fluorescence energy transfer (2) or protein x-ray crystallography (3,4). Though helpful, methods that infer subunit stoichiometry from ensemble measurements ultimately assume that copy number is invariant amongst protein complexes. To take this issue into account, stoichiometry can be determined by studying protein complexes at the single molecule level either with atomic-force

1

microscopy (5) or fluorescence spectroscopy (6-11). In particular, the single subunit counting technique has proven to be a powerful tool (7,11-18). For this technique, each subunit is first tagged with a fluorescent marker and then viewed by total internal reflection fluorescence (TIRF) microscopy which limits background noise in favor of signals coming from the plasma membrane (19). By minimizing surface expression levels, fluorescence emitted by individual protein complexes can be resolved. For each complex, the number of photobleaching steps can then be counted and, by association, the subunit stoichiometry determined (9,10). This strategy is particularly appealing when expressing the protein of interest in the *Xenopus laevis* oocyte since it permits the use of GFP-tagged subunits as well as providing measurements with high signal-to-noise ratio (7).

Despite its value, an important caveat of this approach is that the oocyte's biochemical machinery does not always assemble mammalian proteins faithfully. For example, the stoichiometry of human neuronal nicotinic acetylcholine receptors (nAChRs) was found to be different between Xenopus oocytes and mammalian cells (20,21). An added complication is that proteins nascent to the oocyte can co-assemble with exogenous proteins, as shown for nAChRs (22) and NMDA-type ionotropic glutamate receptors (23). This situation is unlikely to be limited to a few exceptions since the Xenopus oocyte expresses a plethora of ion-channels and transporters that are orthologs of mammalian proteins (24-26). Although the ideal solution would be to perform subunit counting in mammalian cells, this approach has its own challenges. First, background noise is greater in mammalian cells making it more difficult to discern the desired signal emitted from transfected cells. Second, surface protein density is more difficult to control in mammalian cells compared to Xenopus oocytes leading to heterogenous expression patterns and to occasional colocalization of proteins. This means that in spite of the time-consuming nature of the analysis, practically all proteins expressed on the cell surface have to be analyzed to prevent biasing the outcome to a certain subpopulation. Third and finally, post-translational formation of the GFP chromophore is less efficient at the higher

temperatures needed for mammalian cell cultures (i.e. 37°C) (27,28) placing limits on the accuracy of the counting method.

To respond to these challenges, we developed the necessary tools to make single subunit counting experiments in mammalian cells possible. Concerns with low chromophore formation were addressed by constructing a monomeric version of GFP (29) based on superfolder GFP (30) (msfGFP; see *Experimental Procedures*). To enable unbiased analysis of all proteins expressed on the cell membrane, we have developed a fully-automated counting program that is capable of resolving steps typical of noisy *in situ* photobleaching experiments. It does so by separating relevant signals from contaminants, extracting the bleaching traces also from high background levels and classifying these according to number of bleaching steps. The counting program has been named *PIF* to reflect the "**P**rogressive **I**dealization and **F**iltering" functions of step detection and to recall the French expression *avoir du pif*, meaning "to have intuition" or literally "to have a good nose for something". Using this approach, we have successfully determined the tetrameric and pentameric stoichiometry of GluK2 kainate and α1-glycine neurotransmitter receptors, respectively. Given this, we propose that *PIF* can be viewed as a universal tool for the automated analysis of subunit counting data of membrane-bound proteins expressed in mammalian cells.

EXPERIMENTAL PROCEDURES

CELL CULTURE AND TRANSFECTION
HEK293T cells were plated on poly-D-lysine-coated 35mm #1 glass bottom dishes (MatTek Corporation) at a density of 2.5×10^4 cells/mL in DMEM (Invitrogen) supplemented with fetal bovine serum (2%; Gibco). Cells were maintained in a humidified incubator at 37°C containing 5% CO_2. After 25-30 hours, cells were transfected by the calcium phosphate method. Precipitate was allowed to form for 4 hours before cells were washed twice with PBS and placed back into fresh DMEM. Cells were allowed to recover between 16-22 hours post-transfection before being fixed in EM-grade 4% formaldehyde (Ladd Research) in PBS for 24-36 hours at 4°C.

PLASMID CONSTRUCTION

Monomeric superfolder GFP (msfGFP) contains a lysine residue at amino acid position 206 (V206K), known to substantially reduce fluorophore dimerization (29). The monomeric form of the superfolder GFP (msfGFP) was inserted into the coding sequence of the rat ionotropic kainate receptor GluK2 (NM_019309; pRK5-msfGFP-GluK2) and the human glycine receptor GlyRα1 (NM_000171; pRK5-msfGFP-GlyRα1 utilizing a large-insert mutagenesis protocol (31). For both constructs, msfGFP was inserted N-terminally between the first and second or third and fourth amino acids of the mature GluK2 or GlyRα1 proteins, respectively. The fusions were generated by large-insert site directed mutagenesis. Briefly, primers were designed to incorporate a 20-23 nucleotide overhang into both the 5' and 3' ends of the msfGFP sequence. These overhangs are complementary to regions directly before and after the desired point of msfGFP insertion in expressions plasmids containing either the GluK2 or GlyRα1 sequence. Primers are used to generate a megaprimer by PCR, which, after purification, is in turn utilized as a primer for large-insert site-directed mutagenesis of the target sequence. All constructs were verified by sequencing.

IMAGING

Fluorescence was measured using an inverted microscope (Axiovert 200, Zeiss, Jena, Germany). Excitation was achieved in total internal reflection fluorescence (TIRF) configuration using a 60 mW 488 nm laser (PhoXx, Omicron-Laserage, Germany). TIRF allows a thin excitation height of about 200 nm above the coverslip, thus eliminating a huge proportion of the background emerging from different cell compartments. Excitation light (~10.5 mW) was defocused to a diameter of approx. 35 mm and was partially cut to a diameter of ~7 mm using an iris, thus allowing a very uniform excitation intensity over the region observed. The total excitation intensity at the exit of the objective was about 550 μW. Emission was collected with a 63x NA 1.49 objective (Olympus, Richmond Hill, ON) and filtered by a high efficiency 525/50 nm emission filter in combination with a Z488 dichroic mirror (Chroma Technologies, Rockingham, VT). Images were recorded with a backlit EMCCD Camera (iXon+

860BV, Andor Technology, South Windsor, CT) at a sampling rate of 20 Hz.

GENERATING SIMULATED DATA

Data featuring variably photobleaching, variations in step size and blinking were simulated using a hidden Markov model (Supplemental Fig. 1A). For simplicity, blinking events following single transition kinetics were considered. The mean dwell time τ in each state was calculated by the inverse of the sum of rates leaving the state. The duration t in which a state will be visited is calculated using a generated random number between 0 and 1, exclusively, to which an exponential decay function with characteristic time τ is associated. The result is then sampled at a frequency of 20 Hz, except for the simulations with fast blinking events ($\alpha \geq 50$), which are first sampled at 200 Hz before being re-sampled at 20 Hz. After this period in a given state, the next state is randomly selected based on the rate constants, acting as weights. This process is repeated until a duration of 100 sec is reached. Poisson noise is finally added to the idealized trace.

DEFINITION OF FLUORESCENT SPOTS

The fluorescence intensity of each spot had a contributing region of 3x3 pixels defined by the point spread function of the imaging system. The signal δF is obtained by fitting the initial total fluorescence intensity F of each spot to a 2D Gaussian profile and subtracting the estimated baseline from the total fluorescence. A spot was detected if the signal to total fluorescence ratio $\delta F/F$ overcame a pre-defined cut-off. In addition, an "overlap limit" ensured that individual spots do not influence each other's fluorescence intensity. The limit was defined as the distance where the Gaussian profile has reached 20% of its maximum and had to be free of neighboring spots.

TRACE FILTERING

Background fluorescence was removed by convoluting a *Laplacian of Gaussian* (LoG) kernel with the matrix of each selected spot (Fig. 1B). The filtering step effectively applied a band-pass filter similar to a deconvolution algorithm applied to confocal images to increase resolution and remove blurring (32). The LoG kernel has a standard deviation corresponding to the mean value obtained from the Gaussian fit over a set of

spots. This filter is especially useful in low SNR systems because it efficiently removes high background fluorescence.

Individual traces were further filtered with the *Chung-Kennedy (C-K) filter* (33). This non-linear filter has been designed to reveal transients dominated by noise and has been shown to be an efficient aid for step detection (33,34). Briefly, each of the points comprising a trace was filtered by a weighted average with the neighboring points directly preceding and following within a fixed time window. Weighting was evaluated for each point and was based on the variance of the trace before and after that point. The C-K filter improved the SNR by a factor of approx. 3.0 ± 0.5 and did not filter out any bleaching steps. Interestingly, for SNR > ~2, the filter reduces the overall accuracy by up to 7%. A C-K filter should therefore only be used if traces display a SNR below ~2. According to our estimations, C-K filtering should be applied for experimental traces improving their filtered SNR to values typically around 4-5 (Supplemental Fig. 1B).

STEP DETECTION
The program's step detection algorithm identifies segments of constant intensity. This is an iterative process where initially short segments are averaged, and averaged data are exposed to another iteration of the same algorithm. This gradual increase in segment duration ensures that steps are readily distinguished. The main task of the step detection algorithm is to separate bleaching steps from noise or blinking events. This requires information on noise and step amplitudes, both of which are initially unknown. As a direct determination of the noise level is prevented by the step-wise decrease of the baseline during bleaching events, the noise is estimated by evaluating variations of the fluorescence fluctuations (N_{FF}). Briefly, N_{FF} is obtained by calculating the standard deviation of the intensity differences between adjacent frames (for details see *Supplemental Text*). This value serves as the basis for the calculation of the SNR of the trace. Since also the step amplitudes (the signal) are initially unknown, the algorithm assumes a very small SNR (~0.25-1) for the first step detection iteration. Starting from the first frame, it continuously calculates a "floating" baseline as the

average of the intensity data. If the intensity falls below a limit $\sigma = (SNR \cdot N_{FF})/\varphi$ from the calculated average, a step is detected and the algorithm assumes a new baseline. The "sensitivity" φ – used to determine the threshold – is an empirical calibration factor dependent on the SNR of the current trace (Supplemental Fig. 1C). For subsequent iterations, the SNR is calculated using the step amplitudes resulting from the previous iteration.

In the first iteration of a trace, many more steps will be detected than actually present. However, the "idealized" trace obtained in each run will undergo subsequent iterations with a lower sensitivity φ – calculated each round from the newly obtained idealized trace – so that noise and arbitrary fluctuations are progressively eliminated until only true bleaching steps remain (Supplemental Fig. 2). This process stops if no further reduction of steps is observed in successive iterations of the algorithm. A more detailed description of the algorithm can be found in the supplemental information.

Three step properties are evaluated in the detection algorithm in order to reduce false detections caused by artifacts created occasionally when working in mammalian cells. These properties are the minimal time permitted between steps, the maximal step amplitude variation and the minimal step amplitude (Supplemental Fig. 3, A-D). Notably, the minimal time between steps must last at least two frames because each individual photobleaching event is likely to occur during any given exposure interval. The new fluorescence intensity is therefore only registered in the second frame whereas the first will assume an intermediate value. The variation of the step amplitudes was limited to exclude false steps, whose amplitudes were small in comparison to those of msfGFP bleaching steps within the same trace. The msfGFP bleaching steps displayed similar amplitudes amongst each other. This criterion proved most important and ruled out most of the detection artifacts in typical recordings of HEK293 cells. Limiting the maximal amplitude variation to 60% of the largest step in the trace removed the artifacts while still tolerating variations due to blinking events. 60% variation also accepts two fluorophores photobleaching

during one frame exposure. Steps that did not fulfill all three criteria were not counted.

QUALITY CONTROL

Despite PIF's accuracy in step detection, not all artifacts were ruled out. In particular, some traces were still analyzed that did not display any step-wise behavior (e.g. from "dirt" particles). This is a problem specific to automated algorithms. While a user would by eye ignore these traces at once, the algorithm will accept them as long as the detected "steps" do not violate any of the above step detection criteria. The algorithm has to be instructed specifically which traces to discard. For this purpose, a set of five evaluation criteria were built into the software as a final quality control. These rejection criteria were found to be good indicators for the presence of contaminants contributing to trace artifacts. They include threshold values for a Chi-squared goodness-of-fit evaluation ($\chi^2 > 1.5$), for a minimal SNR, for the step duration as well as for the step amplitude. The χ^2 of the traces also had to be smaller than the χ^2 of a counterfit (35) (for details see *Supplemental Text*). If an idealized trace fails to meet the requirements of any of these criteria, it is omitted from analysis.

DETERMINATION OF ACCURACY

The algorithm's step detection accuracy p_{acc} was evaluated for each SNR. Evaluation of the accuracy was achieved by computing the probability of correctly detecting exactly one step where exactly one step truly existed. Briefly, for each SNR, a step distribution histogram is obtained, from which the proportion of each detected step number is calculated. To correctly determine the accuracy, it is necessary to consider that simultaneous false step detections and missed events within the same trace would compensate for one another. Such errors could result in "falsely" finding the correct overall step number present in a given trace. Consequently, the calculation of the accuracy takes the probability of missed events and false step detections into account.

SIMULATIONS: CALCULATION OF THE ACCURACY P_{ACC}

We define the accuracy p_{acc} as the probability of correctly detecting one step when exactly one step exists. If PIF was applied to simulated photobleaching data containing exactly 11 steps, p_{acc} could be derived from the resulting observed step distribution. As a first approximation based on the frequency of identifying 11 steps (*P(11)*), one could reason that $P(11) \sim p_{acc}^{11}$. However, there are several ways to attain a final count of 11 steps. For instance, step detection errors could result in including a step that does not truly exist (false detection) and/or in missing an existing step (missed event). A false detection and a missed event occurring in the same trace would result in an apparently accurate final step count, leading to an overestimation of p_{acc}. To correct for such situations in the context of calculating p_{acc}, we defined the theoretical frequency distribution of detecting N steps for N_{real} real steps:

$$P(N) = \sum_{k=0}^{floor(N/2)} \frac{N_{real}!}{(N-2k)!(k+N_{real}-N)!k!} p_{acc}^{N-2k} p_0^{k+N_{real}-N} p_2^{k} \qquad \text{if } N < N_{real}$$

$$P(N) = \sum_{k=N-N_{real}}^{floor(N/2)} \frac{N_{real}!}{(N-2k)!(k+N_{real}-N)!k!} p_{acc}^{N-2k} p_0^{k+N_{real}-N} p_2^{k} \qquad \text{if } N \geq N_{real}$$

where p_0 and p_2 are the probabilities of missing a step and counting 2 steps for one real step, respectively. The factor at the beginning of each term represents all the arrangements of p_{acc}, p_0 and p_2, as each event could happen in any order. The resulting theoretical distribution $P(N)$ was then fitted to the observed step distribution as a function of p_{acc}, p_0 and p_2. This fit was performed by minimizing the difference between our calculations and the actual distribution (Levenberg-Marquardt algorithm).

KINETICS OF FLUORESCENT PROTEIN BLINKING

Two rate constants α and β determine the transitions out of and into the non-fluorescent state, respectively. For a constant photobleaching rate of $\kappa = 0.1$ s^{-1}, we varied blinking frequency and duration ($\alpha = 1$, 5, 10 and 50 s^{-1}; $\beta = 0.1$ and 1 s^{-1}; Supplemental Fig. 4). For both β values tested, step detection accuracy decreased slightly with increasing blinking duration (decreasing α). For blinking events with long duration ($\alpha \leq 5$ s^{-1}), the decrease was significant even at SNR > 2. Although for a moderate blinking rate ($\beta = 0.1$ s^{-1}) the detection accuracy always remained above 90% for SNR > 2, for a high blinking rate ($\beta = 1$ s^{-1}), 90% accuracy could only be achieved when

5

$\alpha \geq 10 \text{ s}^{-1}$, i.e. if the events were not too long-lived.

The above values for simulated blinking events were chosen to test the limitations of the algorithm and do not actually represent the conditions typically observed in experimental fluorescence recordings. To verify the blinking properties under typical conditions, we obtained experimental traces recorded from HEK293T cells expressing a monomeric form of "superfolder GFP" (msfGFP) (30) fused to GluK2 (see below). To obtain the experimental blinking rates α and β the traces were fitted to a hidden Markov model using QuB (36,37). We chose traces with particularly prominent blinking events to ensure precise computation of the blinking rate constants. Analysis resulted in two different phenotypic populations; a first population typified by long low frequency blinking events with mean values for $\langle\alpha\rangle$ and $\langle\beta\rangle$ of $1.27 \pm 1.05 \text{ s}^{-1}$ and $0.013 \pm 0.07 \text{ s}^{-1}$, respectively (N = 7). By contrast, a second population represented by the majority of traces exhibited a greater number of shorter blinking events with mean values for $\langle\alpha\rangle$ and $\langle\beta\rangle$ of $12 \pm 14 \text{ s}^{-1}$ and $0.12 \pm 0.11 \text{ s}^{-1}$, respectively (N = 12).

SOFTWARE
PIF was entirely programmed and is run using Matlab (Mathworks, Nattick, MA).

RESULTS

IDENTIFICATION OF RELEVANT SPOTS
When performing single molecule fluorescence in mammalian cells, a multitude of spots varying in intensity and distribution are found (Fig. 1A). Manually selecting a number of these spots introduces a user bias such that a certain sub-population (the "good-looking" ones) of proteins is favored at the expense of another. This is particularly pertinent when considering neuronal cells which invariably have a heterogenous receptor expression pattern. It is therefore imperative, first, to detect all spots in the region of interest and, then, to separate relevant spots, originating from the tagged protein located on the plasma membrane, from contaminant spots by objective criteria independent of the respective user. In order to process the large number of spots found on a single mammalian cell, each relevant fluorescent spot had to be detected and analyzed in an automated fashion (Fig. 1, A-D).

In our automated analysis, spot identification was limited to a user-defined region of interest (ROI) delineating the borders of a cell expressing the fluorescently-tagged protein (Fig. 1A). To resolve single spots, background intensity originating from within cells was suppressed by utilizing total internal reflection fluorescence (TIRF) microscopy which limits excitation to the areas close to the plasma membrane (38). However, TIRF configuration does not prevent excitation of any contaminant spots located close to the plasma membrane or within the penetration depth of the evanescent wave into the cytosol (approx. 0.3 µm). Both factors gave rise to contaminant spots that were difficult to distinguish from relevant spots corresponding to individual protein complexes (Fig. 2A).

Common contaminant sources include macromolecules present at the surface of the glass coverslip as well as endogenous proteins of the cell. For instance, flavins are abundantly expressed in the cytosol and are known to have spectral properties similar to GFP, albeit with much lower extinction coefficient and quantum yield (39,40). Thus, they fluoresce with significantly lower intensity than proteins tagged with a fluorescent protein variant. Similarly, proteins located at a distance of more than 200 nm from the plasma membrane in the cytosol fluoresce with very low intensities when using TIRF excitation. This allowed us to rule out many contaminant spots by defining a threshold for the fluorescence intensity of the spots (Supplemental Fig. 1D).

Other contaminants that exhibited fluorescence intensity comparable to our proteins of interest were not so easily distinguished. These were contaminants positioned so close to the membrane that their fluorescence intensity was not reduced by TIRF excitation. Some contaminants were also observed in the absence of tagged proteins and thus attached to the coverslip surface (Fig. 2A, *orange*). Also tagged proteins retained in the ER were still detected in mammalian cells (Fig. 2A, *violet*). In oocytes, the ER is positioned at a distance of more than 1 µm from the plasma

6

membrane (41) and is optically shielded by the strong pigmentation just below the membrane. In contrast, in mammalian cells, the ER comes so close to the plasma membrane that proteins therein are still excited in TIRF configuration. We tested this by transfecting HEK293 cells with a Kainate receptor that is retained in the ER (GluK5 fused to msfGFP). Even under TIRF excitation, proteins in the ER were still clearly detectable with intensities not significantly lower than proteins expressed at the plasma membrane. These spots had to be excluded because they might have been retained in the ER exactly because they did not exhibit the correct subunit composition.

We found that both types of contaminants, spots on the coverslip and tagged proteins in the ER, were despite their comparable fluorescence intensity distinguishable from relevant spots in their photobleaching decay times. On clean coverslips devoid of cells, a *fast* fluorescence decay was observed and could accordingly be attributed to background arising from impurities (Fig. 2A, *orange*). When expressing the trafficking deficient mutant GluK5-msfGFP, we found a very *slow* decay. The longer distance for proteins in the ER, and thus lower TIRF excitation, leads to lower excitation intensities and, by association, slow bleaching time constants (Fig. 2A, *violet*). The remaining *medium* component was the relevant component associated with the photobleaching of the msfGFP-fluorophores (Fig. 2A, *green*).

To remove fluorescence from the undesired fast and slow components, we applied a *Laplacian-of-Gaussian* (LoG) filter (Fig. 1B), which is known from deconvolution of confocal images to increase resolution and remove blurring (see Experimental Procedures, 32). Prior to filtering, the contaminant fast and slow components each accounted for 15-20% of the overall fluorescence leaving 60-70% attributed to the relevant msfGFP fluorescence. Filtering completely removed the fast background component and substantially attenuated the slow contaminant component, resulting in 90% fluorescence from the proteins of interest (Fig. 2A). At this time, we separated the complete ensemble of spots originating from tagged proteins on the plasma membrane from any contaminants. The associated fluorescence traces were finally filtered to increase their signal to noise ratio (Chung-Kennedy filter; Fig. 1B) before entering the step detection phase.

STEP DETECTION ALGORITHM PIF

In the second phase, the automated step detection algorithm is applied to every fluorescent trace from relevant spots identified above (Fig. 1C; see *Experimental Procedures* and S*upplemental Text* for a detailed description of the algorithm). Other step detection algorithms, designed to work at the high noise levels that are typically observed in mammalian cells, detect deviations larger than the noise level and compare them to the "signal" amplitude. In the case of photobleaching, this strategy is not practicable as both signal to noise ratio and the expected intensity levels (or even their number) cannot be determined *ad hoc* but are irreversibly intertwined with the number of steps that we wish to measure. In the PIF algorithm, this problem was solved by considering the variations of intensity rather than the noise (*Supplemental Text*) and by using a slow iterative process to detect the steps. We found that direct detection is more likely to miss a step or not to filter out artifacts. PIF, on the other hand, is initially overly sensitive and detects any "ripple", but, with each iteration and decreasing sensitivity, these are smoothed out until only the real steps are crystallized out. Steps in the idealized traces were finally analyzed according to a number of objective criteria to rule out any artifacts (see *Quality Control* in *Experimental Procedures*; Fig. 1D). PIF recognized each step in data typical for experimental conditions with an accuracy of more than 90% (Fig. 2B, see below).

TESTING STEP DETECTION BY PIF IN SIMULATIONS

We tested how well the PIF algorithm finds the correct photobleaching steps from noisy traces in simulated data. We simulated traces traversing 11 equidistant steps following a hidden Markov model with signal to noise ratios (SNR) between 0.33 and 5.75 (see *Experimental Procedures*). We determined the accuracy of step fitting to be higher than 90% for SNRs over ~2 without filtering (Fig. 2C, *black trace*). The accuracy was significantly improved by C-K filtering at low SNRs, and 90% accuracy was achieved for SNR > 1.2 (Fig. 2C, *red trace*).

Since fluorophores exhibit variable intensities (42), also the amplitudes of photobleaching steps are not necessarily identical but exhibit variations to some degree. Amplitudes vary typically by 33% under experimental conditions. We allowed the step amplitudes to vary in simulated data 33% and 75% representing a typical and extreme value, respectively. In both cases, step detection accuracy was not substantially altered and remained above 90% (Fig. 2B).

Most subunit counting experiments are performed using fusion constructs with fluorescent proteins such as GFP and its derivatives. These fusion constructs evade any problems of unspecific labeling or accessibility, as expected for labeling with organic fluorophores (15,43) or antibodies (11), respectively. However, GFP and its derivatives have been shown to exhibit recurrent short dark periods in their emission known as blinking (44,45). The onset of each dark period is comparable to a photobleaching step, but differs in that it recovers to the original intensity. We tested the effect of blinking on step detection accuracy of PIF in simulated traces containing four steps of equal amplitude (Fig. 2B). The kinetics of photobleaching were obtained from TIRF recordings with an iGluK2 fusion construct with the monomeric version of superfolder GFP (msfGFP, see below) expressed in HEK293 cells (see *Experimental Procedures*). Both sets of kinetic parameters found were tested in simulations, and neither had a significant effect on the accuracy of step detection of the PIF algorithm (Fig. 2B). Combining values of both the variations in step amplitude and the blinking rates typical for photobleaching data obtained from mammalian cells, PIF's step detection accuracy, still, was found to be over 90% (Fig. 2B).

We, finally, tested how well PIF determines the correct stoichiometry for different oligomers. We simulated photobleaching data with experimental parameters of different stoichiometries, i.e. 500 traces each for monomers, dimers up to hexamers, and let PIF analyze these data in an automated fashion. On average, each step was found with an accuracy of over 90%, and the maxima of the distributions were exactly at the correct value (Fig. 2D). These simulations confirmed the capability of

PIF to determine the correct stoichiometry of a multimeric protein under experimental conditions.

FLUORESCENT SUBUNIT COUNTING IN MAMMALIAN CELLS

Fluorescent subunit counting data were collected from fixed HEK293T cells transiently transfected with a fluorescently-tagged ligand-gated ionotropic glutamate receptor GluK2. HEK293T cells as recombinant expression system are typically used for expression of membrane proteins and offer particular advantages with respect to the neuronal ion channels analyzed in this study. Unlike *Xenopus laevis* oocytes (46), HEK293T cells do not endogenously express kainate receptors. Furthermore, despite their origin as kidney cells, HEK293T cells share a common mRNA transcription profile with neurons (47). This final point gives confidence that the ion channels such as GluK2, which are natively expressed almost exclusively in neurons, keep their inherent stoichiometry when expressed either in homo- or in heteromeric form.

N-terminal to GluK2 a msfGFP was fused, and cells were imaged using TIRF microscopy. Under these conditions, cells exhibiting low msfGFP-GluK2 expression levels showed single molecule fluorescence clearly visible above background fluorescence levels (Fig. 1A). If expression density or background fluorescence was substantial, cells were automatically removed from analysis because too many traces were rejected by *PIF*. Typically, $50 \pm 8\%$ of traces were excluded from counting analysis by the program, however, recordings resulting in more than 70% trace rejection were not considered for analysis, as such recordings indicated inappropriate cell selection. For all other cells, PIF automatically analyzed the recordings and accumulated the step number histogram (Fig. 3A).

Finally, the step number histogram that we obtained from the automated analysis remained to be interpreted. If all GFP molecules were fluorescent, the step number would match the number of subunits per protein assembly. Unfortunately, however, each fluorescent protein has only a certain probability (p_m) of being fluorescent defined by the ability of the chromophore to fold and mature correctly (7,48).

8

As all identical fluorophores have an equal probability of fluorescing (p_m), the resulting step number histogram has to follow a binomial distribution dependent on this maturation probability. Previous reports counting eGFP-fusion proteins expressed in *Xenopus laevis* oocytes matured at ~18°C have estimated p_m values between 75% and 80%. Such maturation rates were not achieved at 37°C required for mammalian cell culture. To optimize maturation rates as much as possible, we constructed a monomeric version of the superfolder GFP msfGFP (29,30). The monomeric version of GFP prevented protein-protein interactions between GFPs which might influence stoichiometry.

Consistent with the known homotetrameric assembly of GluK2-msfGFP, the step histogram generated automatically by PIF was best fit by a binomial distribution of 4th order (Fig. 3A, *blue trace*). Interestingly, a fraction of spots exhibited between 5 and 8 steps, which were not considered for the fit. They strongly suggested some cases where two receptors occupied a single spot. Whether this is a result of inadequate system resolution or is biologically relevant remains currently unknown. To account for the fraction of spots with two receptors, the binomial distribution was adjusted to consider both possibilities (*Supplemental Text*), which significantly improved the fit (Fig. 3A, *red trace*), and we were able to estimate that approximately 23% of the spots contained two receptors.

The maturation probability of GFP only depends on the expression system and the temperature during maturation and not on the protein it is fused to (30). This value has to be determined once for a given expression system with a protein of known subunit composition (here 53%), and will remain constant for any further protein. The distribution will then only depend on the order of the binomial given by the number of subunits. To confirm the universal validity of the maturation probability p_m, PIF was applied to another ion channel of known stoichiometry. The above experiments defined $p_m = 0.53$ for our expression system. We now expressed a msfGFP-tagged alpha1 glycine receptor (GlyRα1) and automatically analyzed the subunit counting data. As before, the best fit was obtained assuming two

receptors could occupy a single resolvable spot, evident in the detection of several traces containing 6 to 9 photobleaching steps. Similar to GluK2, ~84% of spots were estimated to contain individual GlyRα1 receptors. The step number histogram was fit with double-binomial distributions of 5th (pentamers) as well as 6th order (hexamers; Fig. 3B). However, only the pentameric stoichiometry reported a p_m of 53%, a value identical to that obtained in GluK2 experiments, whereas the hexameric stoichiometry required $p_m = 46\%$. A hexameric distribution with a $p_m = 53\%$ was clearly not able to follow the obtained distribution (Fig. 3B). By calibrating the system with a defined maturation rate p_m, we can thus determine unambiguously the stoichiometry of a membrane protein in spite of the low maturation rate of fluorescent proteins in mammalian cells.

DISCUSSION

In this study, we provide a complete framework to study the stoichiometry of any membrane protein in a mammalian expression system using single subunit counting. First, we have developed a fully-automated subunit counting algorithm, *PIF*, that greatly minimizes user-bias in discerning photobleaching steps from background noise. The automation of step detection has the added advantage of permitting the analysis of large datasets over a short time period minimizing the risk of analyzing only part of a heterogenous population. Second, *PIF* is able to reliably extract the relevant signal from records with relatively low signal-to-noise ratio enabling to routinely study the stoichiometry of a wide variety of membrane proteins expressed in mammalian cells. This feature resolves the emerging concern of performing similar experiments in *Xenopus laevis* oocytes where mammalian protein assembly is sometimes problematic. Third, and finally, we introduced in the analysis the use of a fixed maturation rate specific to each expression system and double-binomial distributions in response to low GFP maturation rate and co-localization of single photobleaching spots, respectively. Interestingly, co-localization was not reported in prior work where the manual counting of photobleaching steps was performed on *Xenopus* oocytes. Whether this finding is unique to

9

mammalian cells or has been neglected given the nature of the manual method will require further study.

COMPARISON WITH PREVIOUS STUDIES

Other step detection algorithms perform well for numerous applications (34). To evaluate how these other step detection algorithms performed for subunit counting compared to *PIF*, we applied the STEPFINDER algorithm (35) to simulated traces exhibiting variable SNRs in the absence of blinking. STEPFINDER is a chi-squared based algorithm originally developed to analyze the stepwise behavior of molecular motor displacements and was recently applied as a tool to analyze fluorescent subunit counting data (49). Despite STEPFINDER's proficiency as a tool for determining molecular motor displacements, the computed accuracy of *PIF* was significantly higher than STEPFINDER in particular at low SNR (< 4; Supplemental Fig. 1E). Like STEPFINDER, *PIF* could also be applied in the context of molecular motor displacements or similar experiments, especially since blinking would not be an issue. Similarly, PIF could also analyze data derived from oocytes or even purified proteins (7,14,15,18).

Some previous studies deduced the number of subunits from the fluorescence intensity distribution of the individual spots rather than counting the number of photobleaching steps as in this study (8,16,17,50). Evaluation of the fluorescence intensities depends on relatively constant quantal intensity. Our data indicate, however, that a mean variation of 35% is typically observed in mammalian cells using fluorescent proteins. These variations may result from microenvironment changes (e.g. quencher concentration) (14,51), dipole orientation (42) and uneven excitation intensity over the field of view.

THE CHALLENGES FOR FUTURE WORK ON MAMMALIAN CELLS

Although the ideal situation would be to study membrane-bound mammalian proteins in their native environment, this is difficult for a number of reasons. For example, the study of neurotransmitter receptors, as performed here, would be most relevant if performed in cultured neurons or acutely-isolated explant tissue where

the normal clustering and assembly of neurotransmitter receptors are regulated by a variety of intrinsic and extrinsic factors (52,53). The successful execution of subunit counting using *PIF* would prove difficult under these conditions due to high protein densities which would make single-protein resolution unlikely. Given that *PIF* remains highly accurate over a wide range of SNRs, automated subunit counting analysis utilizing neurons may be possible, for example at extrasynaptic sites, provided receptor density is sufficiently low. An added complication however would be the formation of protein complexes between endogenous and recombinant protein subunits which would interfere with estimates of stoichiometry. This problem can only be overcome either by silencing the endogenous expression or by directly analyzing the endogenous proteins with PIF fluorescently tagging them with specific antibodies or toxins (11). In view of these difficulties, HEK293T cells were selected as the expression system in this study. Despite the challenge of background fluorescence resulting in low SNR, this cell line is used routinely in many laboratories. Most importantly, it shares a surprisingly similar mRNA expression profile with neurons (47) making it the cell of choice for subunit counting of neurotransmitter receptors.

A more tenable issue that may be resolved in the near future is to maximize the post-translational formation of the chromophore of GFP. In this study, we constructed a monomeric form of superfolder GFP (msfGFP) fused to the protein of interest since GFP dimerization would likely influence protein assembly (29,30). Compared to other GFP variants, the superfolder GFP has the advantages of superior folding efficiency and chromophore maturation properties at 37°C (30). In addition, it is also brighter thus increasing SNR in single molecule studies. Despite the enhanced folding and maturation, reported p_m values in *Xenopus* oocytes (75-80%) (7), are significantly greater than the values we obtained expressing msfGFP (53%) in mammalian cells. This is a critical factor to resolve in future work since maximizing p_m optimizes the confidence in the binomial fit of the photobleaching step distribution. Why maturation rates are lower in mammalian cells is still not clear though it is

likely related to the higher temperature needed for mammalian cell culture which in turn affects the temperature-sensitive folding of GFP (27). Working at decreased temperature would only solve the problem to a limited extend as not only folding of GFP but possibly also folding, assembly and trafficking of the protein of interest would be affected. As we learn more about the structure of GFP and its variants, it may be possible to make mutations that allow the formation of the chromophore to be optimized for 37°C (27,28).

CONCLUSION

The last two decades has seen unprecedented advances in our understanding of a plethora of membrane-bound protein families in mammals; from their initial molecular identification by cloning to elucidation of their function with genetically-modified animals. Despite these advances, many basic properties of these proteins, such as subunit stoichiometry, remains poorly understood. Here, we have developed a reliable subunit counting method specifically for the study of proteins expressed in mammalian cells. The universality of this technique makes it possible to re-visit the widely anticipated role of subunit composition in fine tuning the response of individual cells and tissues to their external environment.

REFERENCES

1. Schagger, H., Cramer, W. A., and von, J. G. (1994) Analysis of molecular masses and oligomeric states of protein complexes by blue native electrophoresis and isolation of membrane protein complexes by two-dimensional native electrophoresis. *Anal. Biochem.* **217,** 220-230

2. Zheng, J., Trudeau, M. C., and Zagotta, W. N. (2002) Rod cyclic nucleotide-gated channels have a stoichiometry of three CNGA1 subunits and one CNGB1 subunit. *Neuron* **36,** 891-896

3. Doyle, D. A., Cabral, J. M., Pfuetzner, R. A., Kuo, A., Gulbis, J. M., Cohen, S. L., Chait, B. T., and MacKinnon, R. (1998) The Structure of the Potassium Channel: Molecular Basis of K+ Conduction and Selectivity. *Science* **280,** 69-77

4. Sobolevsky, A. I., Rosconi, M. P., and Gouaux, E. (2009) X-ray structure, symmetry and mechanism of an AMPA-subtype glutamate receptor. *Nature* **462,** 745-756

5. Barrera, N. P., Betts, J., You, H., Henderson, R. M., Martin, I. L., Dunn, S. M., and Edwardson, J. M. (2008) Atomic force microscopy reveals the stoichiometry and subunit arrangement of the alpha4beta3delta GABA(A) receptor. *Mol Pharmacol.* **73,** 960-967

6. Park, M., Kim, H. H., Kim, D., and Song, N. W. (2005) Counting the Number of Fluorophores Labeled in Biomolecules by Observing the Fluorescence-Intensity Transient of a Single Molecule. *Bulletin of the Chemical Society of Japan* **78,** 1612-1618

7. Ulbrich, M. H. and Isacoff, E. Y. (2007) Subunit counting in membrane-bound proteins. *Nat. Methods* **4,** 319-321

8. Godin, A. G., Costantino, S., Lorenzo, L. E., Swift, J. L., Sergeev, M., Ribeiro-da-Silva, A., De Koninck, Y., and Wiseman, P. W. (2011) Revealing protein oligomerization and densities in situ using spatial intensity distribution analysis. *Proceedings of the National Academy of Sciences* **108,** 7010-7015

9. Schmidt, T., Schütz, G. J., Gruber, H. J., and Schindler, H. (1996) Local Stoichiometries Determined by Counting Individual Molecules. *Anal. Chem.* **68,** 4397-4401

10. Leake, M. C., Chandler, J. H., Wadhams, G. H., Bai, F., Berry, R. M., and Armitage, J. P. (2006) Stoichiometry and turnover in single, functioning membrane protein complexes. *Nature* **443,** 355-358

11. Simonson, P. D., Deberg, H. A., Ge, P., Alexander, J. K., Jeyifous, O., Green, W. N., and Selvin, P. R. (2010) Counting bungarotoxin binding sites of nicotinic acetylcholine receptors in mammalian cells with high signal/noise ratios. *Biophys. J.* **99,** L81-L83

12. Das, S. K., Darshi, M., Cheley, S., Wallace, M. I., and Bayley, H. (2007) Membrane Protein Stoichiometry Determined from the Step-Wise Photobleaching of Dye-Labelled Subunits. *ChemBioChem* **8,** 994-999

13. Shu, D., Zhang, H., Jin, J., and Guo, P. (2007) Counting of six pRNAs of phi29 DNA-packaging motor with customized single-molecule dual-view system. *EMBO J* **26,** 527-537

14. Blunck, R., McGuire, H., Hyde, H. C., and Bezanilla, F. (2008) Fluorescence detection of the movement of single KcsA subunits reveals cooperativity. *Proc. Natl. Acad. Sci. U. S. A* **105,** 20263-20268

15. Groulx, N., McGuire, H., Laprade, R., Schwartz, J. L., and Blunck, R. (2011) Single molecule fluorescence study of the Bacillus thuringiensis toxin Cry1Aa reveals tetramerization. *J Biol Chem.* **286,** 42274-42282

16. Kask, P., Palo, K., Ullmann, D., and Gall, K. (1999) Fluorescence-intensity distribution analysis and its application in biomolecular detection technology. *Proc. Natl. Acad. Sci. U. S. A* **96,** 13756-13761

17. Chen, Y., Muller, J. D., So, P. T., and Gratton, E. (1999) The photon counting histogram in fluorescence fluctuation spectroscopy. *Biophys J* **77,** 553-567

18. Reiner, A., Arant, R. J., and Isacoff, E. Y. (2012) Assembly Stoichiometry of the GluK2/GluK5 Kainate Receptor Complex. *Cell Rep.* **1,** 234-240

19. Axelrod, D. (2001) Total internal reflection fluorescence microscopy in cell biology. *Traffic.* **2,** 764-774

20. Sivilotti, L. G., McNeil, D. K., Lewis, T. M., Nassar, M. A., Schoepfer, R., and Colquhoun, D. (1997) Recombinant nicotinic receptors, expressed in Xenopus oocytes, do not resemble native rat sympathetic ganglion receptors in single-channel behaviour. *J. Physiol* **500 (Pt 1),** 123-138

21. Krashia, P., Moroni, M., Broadbent, S., Hofmann, G., Kracun, S., Beato, M., Groot-Kormelink, P. J., and Sivilotti, L. G. (2010) Human alpha3beta4 neuronal nicotinic receptors show different stoichiometry if they are expressed in Xenopus oocytes or mammalian HEK293 cells. *PLoS. One.* **5,** e13611

22. Buller, A. L. and White, M. M. (1990) Functional acetylcholine receptors expressed in Xenopus oocytes after injection of Torpedo beta, gamma, and delta subunit RNAs are a consequence of endogenous oocyte gene expression. *Mol. Pharmacol.* **37,** 423-428

23. Schmidt, C. and Hollmann, M. (2008) Apparent homomeric NR1 currents observed in Xenopus oocytes are caused by an endogenous NR2 subunit. *J. Mol. Biol.* **376,** 658-670

24. Hagiwara, S. and Jaffe, L. A. (1979) Electrical properties of egg cell membranes. *Annu. Rev. Biophys. Bioeng.* **8,** 385-416

25. Dascal, N. (1987) The use of Xenopus oocytes for the study of ion channels. *CRC Crit Rev. Biochem.* **22,** 317-387

26. Schmidt, C., Klein, C., and Hollmann, M. (2009) Xenopus laevis oocytes endogenously express all subunits of the ionotropic glutamate receptor family. *J. Mol. Biol.* **390,** 182-195

27. Siemering, K. R., Golbik, R., Sever, R., and Haseloff, J. (1996) Mutations that suppress the thermosensitivity of green fluorescent protein. *Curr. Biol* **6,** 1653-1663

28. Craggs, T. D. (2009) Green fluorescent protein: structure, folding and chromophore maturation. *Chem. Soc. Rev.* **38,** 2865-2875

29. Zacharias, D. A., Violin, J. D., Newton, A. C., and Tsien, R. Y. (2002) Partitioning of lipid-modified monomeric GFPs into membrane microdomains of live cells. *Science* **296,** 913-916

30. Pedelacq, J. D., Cabantous, S., Tran, T., Terwilliger, T. C., and Waldo, G. S. (2006) Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol.* **24,** 79-88

31. Geiser, M., Cebe, R., Drewello, D., and Schmitz, R. (2001) Integration of PCR fragments at any specific site within cloning vectors without the use of restriction enzymes and DNA ligase. *Biotechniques* **31,** 88-90, 92

32. Wallace, W., Schaefer, L. H., and Swedlow, J. R. (2001) A workingperson's guide to deconvolution in light microscopy. *Biotechniques* **31,** 1076-8, 1080, 1082

33. Chung, S. H. and Kennedy, R. A. (1991) Forward-backward non-linear filtering technique for extracting small biological signals from noise. *Journal of Neuroscience Methods* **40,** 71-86

34. Carter, B. C., Vershinin, M., and Gross, S. P. (2008) A Comparison of Step-Detection Methods: How Well Can You Do? *Biophysical Journal* **94,** 306-319

35. Kerssemakers, J. W., Munteanu, E. L., Laan, L., Noetzel, T. L., Janson, M. E., and Dogterom, M. (2006) Assembly dynamics of microtubules at molecular resolution. *Nature* **442,** 709-712

36. Qin, F., Auerbach, A., and Sachs, F. (2000) Hidden Markov modeling for single channel kinetics with filtering and correlated noise. *Biophys J* **79,** 1928-1944

37. Qin, F., Auerbach, A., and Sachs, F. (2000) A direct optimization approach to hidden Markov modeling for single channel kinetics. *Biophys J* **79,** 1915-1927

38. Axelrod, D., Burghardt, T. P., and Thompson, N. L. (1984) Total internal reflection fluorescence. *Annu. Rev. Biophys Bioeng.* **13,** 247-268

39. Benson, R. C., Meyer, R. A., Zaruba, M. E., and McKhann, G. M. (1979) Cellular autofluorescence--is it due to flavins? *Journal of Histochemistry & Cytochemistry* **27,** 44-48

40. Xiao, J. (2009) Single-Molecule Imaging in Live Cells. In Hinterdorfer, P. and Oijen, A. v., editors. *Handbook of Single-Molecule Biophysics*, Springer, New York

41. Terasaki, M., Runft, L. L., and Hand, A. R. (2001) Changes in organization of the endoplasmic reticulum during Xenopus oocyte maturation and activation. *Mol Biol Cell* **12,** 1103-1116

42. Ding, H., Wong, P. T., Lee, E. L., Gafni, A., and Steel, D. G. (2009) Determination of the Oligomer Size of Amyloidogenic Protein Beta-Amyloid(1−40) by Single-Molecule Spectroscopy. *Biophysical Journal* **97,** 912-921

43. Blunck, R., Starace, D. M., Correa, A. M., and Bezanilla, F. (2004) Detecting rearrangements of shaker and NaChBac in real-time with fluorescence spectroscopy in patch-clamped mammalian cells. *Biophys J* **86,** 3966-3980

44. Dickson, R. M., Cubitt, A. B., Tsien, R. Y., and Moerner, W. E. (1997) On/off blinking and switching behaviour of single molecules of green fluorescent protein. *Nature* **388,** 355-358

45. Pierce, D. W., Hom-Booher, N., and Vale, R. D. (1997) Imaging individual green fluorescent proteins. *Nature* **388,** 338

46. Schmidt, C., Klein, C., and Hollmann, M. (2009) Xenopus laevis oocytes endogenously express all subunits of the ionotropic glutamate receptor family. *J. Mol. Biol.* **390,** 182-195

47. Shaw, G., Morse, S., Ararat, M., and Graham, F. L. (2002) Preferential transformation of human neuronal cells by human adenoviruses and the origin of HEK 293 cells. *The FASEB Journal* **16,** 869-871

48. Garcia-Parajo, M. F., Koopman, M., van Dijk, E. M., Subramaniam, V., and van Hulst, N. F. (2001) The nature of fluorescence emission in the red fluorescent protein DsRed, revealed by single-molecule detection. *Proc. Natl. Acad. Sci. U. S. A* **98,** 14392-14397

49. Schuster, M., Lipowsky, R., Assmann, M. A., Lenz, P., and Steinberg, G. (2011) Transient binding of dynein controls bidirectional long-range motility of early endosomes. *Proceedings of the National Academy of Sciences* **108,** 3618-3623

50. Swift, J. L., Godin, A. G., Dore, K., Freland, L., Bouchard, N., Nimmo, C., Sergeev, M., De, K. Y., Wiseman, P. W., and Beaulieu, J. M. (2011) Quantification of receptor tyrosine kinase transactivation through direct dimerization and surface density measurements in single cells. *Proc. Natl. Acad. Sci. U. S. A* **108,** 7016-7021

51. Blunck, R., Cordero-Morales, J. F., Cuello, L. G., Perozo, E., and Bezanilla, F. (2006) Detection of the Opening of the Bundle Crossing in KcsA with Fluorescence Lifetime Spectroscopy Reveals the Existence of Two Gates for Ion Conduction. *The Journal of General Physiology* **128,** 569-581

52. Spitzer, N. C. (2012) Activity-dependent neurotransmitter respecification. *Nat. Rev. Neurosci.* **13,** 94-106

53. Turrigiano, G. G. (2008) The self-tuning neuron: synaptic scaling of excitatory synapses. *Cell* **135,** 422-435

15

Acknowledgements - We wish to thank Elise Faure for stimulating discussions and comments on the manuscript.

FOOTNOTES

[‡]Groupe d'Étude des Protéines Membranaires (GÉPROM)

Departments of [§]Physics and [¶]Physiology, Université de Montréal, Montréal, QC, H3C 3J7, Canada
[ǁ]Department of Pharmacology and Therapeutics, McGill University, Montréal, QC, H3G 0B1, Canada
[1]Both authors contributed equally to this work
[2]To whom correspondence should be addressed: Département de physique, C.P. 6128 succ. Centre-ville, Université de Montréal, Montréal, QC, H3C 3J7, Email: rikard.blunck@umontreal.ca
[3]The abbreviations used are: TIRF, Total Internal Reflection Fluorescence; HEK, Human Embryonic Kidney; SNR, signal to noise ratio.

FIGURE LEGENDS

Figure 1.   (A-C) PIF automated analysis. (A) Spots of interest are automatically selected from a user-defined region which encloses the cell (ROI). The fluorescence intensity of 3 spots are shown as examples. (B) Individual spots are processed through a background removal LoG-type kernel with a standard deviation set to 0.76 pixels. The resulting trace is filtered with the so-called C-K algorithm using a fixed time window of 3 frames (150 ms). (C) PIF next applies its step detection algorithm to build an idealization of these traces, from which the steps are counted. Ultimately, a quality control based on 5 criteria determines whether the final result should be included or not in the counting analysis.

Figure 2.   (A) Scheme of a HEK cell under TIRF illumination. GFP labeled proteins located in plasma membrane and endoplasmic reticulum (ER) can be excited. Other excited contaminant molecules are located directly on the coverslip. The global fluorescence decay of each three regions is shown before and after background removal. (B) The accuracy of the step detection algorithm was evaluated from simulations of various photobleaching step conditions. These include the simulation of traces mimicking the experimental conditions, which were C-K filtered because of the initially low SNR observed in msfGFP-based experiments. Each set of simulation comprises about 500 individual photobleaching traces (C) The accuracy is calculated for each SNR from simulation of traces exhibiting equidistant photobleaching steps in absence of blinking events. The relationship between the accuracy and SNR is illustrated both before (*black squares*) and after Chung-Kennedy (C-K) filtering (*red triangles*). (D) Typical experimental conditions were used to simulate traces containing 1 to 6 steps with a final SNR of ~ 4 after C-K filtering. The maxima of the detected step number distribution are plotted against the actual number of steps present in the trace. The error was evaluated using full-width at half maximum.

Figure 3.   (A-B) PIF automated analysis over GluK2 and GlyRα1. (A) The final step distribution of msfGFP-GluK2 (bars) is fitted to a binomial distribution of 4th order (tetramer) either for purely single-channel spots (blue circles) and allowing the possibility of two channels on the same spot (red squares). The resulting probabilities to detect fluorescence from a fluorescent protein were p_m = 57% and 53% for purely single and combined single and double occupancy, respectively. For the latter, the fraction of single occupancy was 77%. (B) The final step distribution of msfGFP-GlyRα1 was fitted to a 5th order binomial distribution (pentamer) and reveal an identical p_m = 0.53 as well as a similar single occupancy (84%) to msfGFP-GluK2. The distribution was also fitted to a 6th order binomial distribution (hexamer) with p_m set to 0.53.

Figure 1

A. Identification of relevant individual spots

δF/F > 10 %
δF > 1750 arb.u.

4 μm
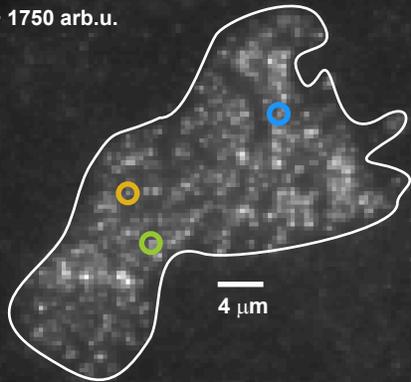
— **Region Of Interest (ROI)**
HEK293T cell

B. Trace filtering

LoG filter

HEK293T cell

+
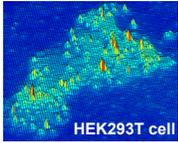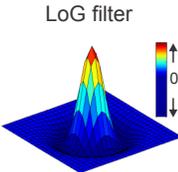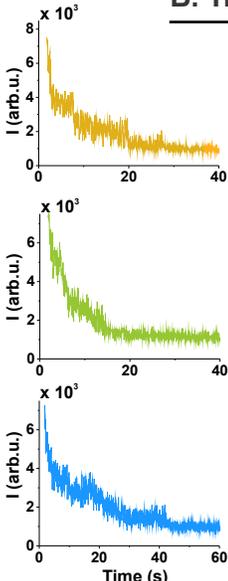
Chung-Kennedy
filter

C. Step detection
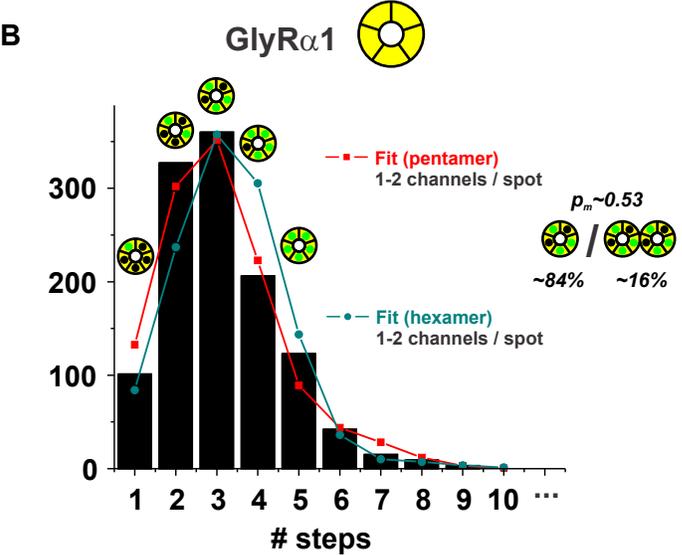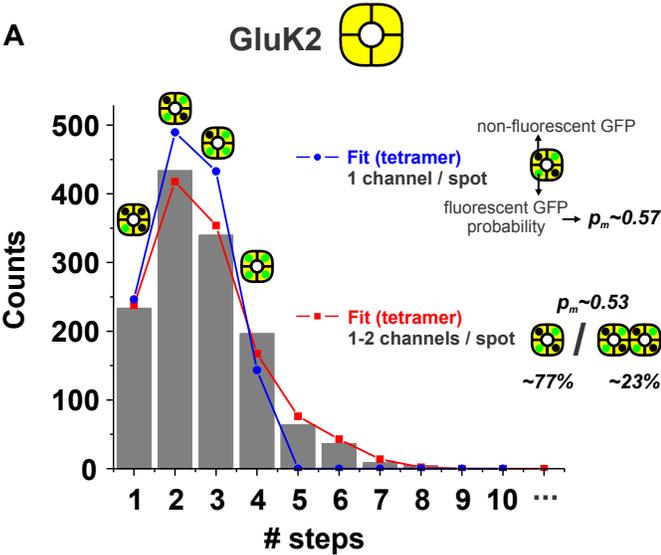
D. Quality control

3 steps

Accepted ●
Rejected ○

4 steps

Accepted ●
Rejected ○

3 steps

Accepted ○
Rejected ●

Figure 2

Figure 3

AUTOMATING SINGLE SUBUNIT COUNTING OF MEMBRANE PROTEINS IN MAMMALIAN CELLS*

Hugo McGuire[‡,§,1], Mark R. P. Aurousseau[‡,‖,1], Derek Bowie[‡,‖] and Rikard Blunck[‡,§,¶]

[‡]Groupe d'Étude des Protéines Membranaires (GÉPROM), Departments of [§]Physics and [¶]Physiology, Université de Montréal, Montréal, QC, H3C 3J7 and
[‖]Department of Pharmacology and Therapeutics, McGill University, Montréal, QC, H3G 0B1, Canada

[1] *Both authors contributed equally to this work*

*Running Title: *Automated Subunit Counting of Membrane Proteins*

To whom correspondence should be addressed: Dr. Rikard Blunck, Département de physique, C.P. 6128 succ. Centre-ville, Université de Montréal, Montréal, QC, H3C 3J7, Email: rikard.blunck@umontreal.ca

Key words: single-molecule; step detection; ionotropic glutamate receptors; cys-loop receptors; superfolder GFP

SUPPLEMENTAL INFORMATION

SUPPLEMENTAL TEXT

STEP DETECTION DETAILS OF PIF

Estimating noise

The basis of step detection by PIF rests on the ability to estimate the noise of a given trace in order to separate it from the bleaching steps. In brief, the noise is used as a "baseline" which a true step has to overcome in order to be recognized. In theory, the noise of a trace can be determined by calculating the difference between the parental raw trace and its idealization (noise free trace). The standard deviation of this resulting subtracted trace would reveal the global noise. However, it is impossible to obtain the idealized trace *a priori* because the idealized trace follows a baseline decreasing in discrete steps. Therefore, we estimated the noise using the standard deviation of the fluorescence fluctuations, as follows:

$$N_{FF} = \left(\frac{1}{T-1} \sum_{i=1}^{T-1} \left[\delta I - \overline{I}_{FF} \right]^2 \right)^{1/2} \quad \text{with} \qquad \overline{I}_{FF} = \frac{I_T - I_1}{T-1} \qquad \text{(eq. S1)}$$

$$N_{FF} = \sqrt{\left\langle \delta I^2 \right\rangle - \left\langle \delta I \right\rangle^2} \qquad \qquad \textit{(eq. S1b)}$$

where I_i is the fluorescence intensity of the frame i, $\overline{I}_{FF}$ represents the mean value of fluorescence fluctuations $\delta I = (I_{i+1} - I_i)$, N_{FF} is the noise from "fluorescence fluctuations" and T is the total number of frames within a recording. Since these fluctuations should represent the direct random nature of noise measured directly, $\overline{I}_{FF}$ is ~ 0. However, photobleaching events contribute to $\overline{I}_{FF}$ to a negative value. If k steps of height ΔI occur, $I_{FF} \approx - k\,\Delta I\,/\,(T-1)$. This would be exactly the expected value for each fluctuation, were the decay continuously distributed over each sample point.

Although this calculation tends to overestimate the noise of a trace, the precise knowledge of the noise is not essential since step detection was optimized with reference to the definition of N_{FF} here. More precisely, the optimal step detection sensitivity φ, which represents the basis of the step detection algorithm (*see below*), was determined using simulated trace data where N_{FF} is controlled. φ is expressed as a function of the signal-to-noise ratio *SNR*, using the step amplitudes as the signal and N_{FF} as noise

Progressive Step Detection

Although N_{FF} is initially known, we would also need the yet-unknown step amplitudes to evaluate the *SNR* of the trace. To circumvent this obstacle, the algorithm starts with a very small value for the *SNR* (user-defined parameter to ~0.25-1.0) and thus for the *signal* = *SNR* · N_{FF}. The step detection algorithm compares the current fluctuations to a threshold factor σ = *signal* / φ, where φ is defined as the step detection sensitivity φ. φ is an empirical calibration factor relating the step threshold σ to the current signal to noise ration *SNR*. The step detection sensitivity φ is well described by an exponential function that starts saturating for SNR > ~2.5 (Supplemental Fig. 1C). To keep the initial threshold σ small, *SNR* is set to a small and by association φ to a large value. Initially, the threshold is set to an even much smaller value $\sigma_1 = \sigma/1000$ (see *Annealing phase* below).

- 1 -

The step detection algorithm itself progresses in an iterative manner and can be broken down into the three major steps described below (Supplemental Fig. 2).

1. SCANNING PHASE: The first data point is set as reference. Starting from here, the algorithm compares the intensity of each subsequent point to that of the reference. Three outcomes are possible; (i) If a point of larger intensity than the reference is found, all data points between the reference and this larger one are set to their average intensity; these points create one idealized "level" (Supplemental Fig. 2A). Then the algorithm proceeds to the next data point, which is set as new reference and begins a new level. Subsequent points are compared to this new reference. (ii) If the value of any subsequent point is lower than the reference but their difference still smaller than the threshold ($I_{reference} - I_i < \sigma_1$), the algorithm sets all points between the reference to the current one to their average intensity (Supplemental Fig. 2B). The reference obtains a new value, but the *same* level is continued with the next data point. (iii) Finally, if the intensity falls below the reference by more than the threshold, the previous level is ended. This lower point itself becomes the new reference and the following points are compared to it (Supplemental Fig. 2C). The algorithm continues to the end of the trace and is then restarted with the idealized ("leveled") data points with identical σ. This is repeated until the number of levels is no longer decreased (Supplemental Fig. 2, D and E).

2. ANNEALING PHASE: Once the idealization for a given threshold value no longer changes, the threshold is progressively increased ($\sigma_1...\sigma_{1000} = \sigma/1000, \sigma/999, \sigma/998, ..., $ to σ) and each time the scanning phase repeated on the previous idealization (Supplemental Fig. 2F). Although one might think that the result should be similar to directly using the final value of σ defined through *SNR* and *φ*, it turned out that the accuracy was higher if the system was allowed to slowly anneal its final sensitivity.

3. REEVALUATION: Once the levels are no longer reduced even with the final value of σ, the new idealized values are used to re-compute N_{FF}. Now, also "signals", step heights between levels, are available, so that *SNR* and *φ* can be re-evaluated. With these values and the current idealized data points, scanning and annealing phase are repeated on the previous idealization – again until levels are no longer reduced (Supplemental Fig. 2G).

Initially, many small fluctuations are interpreted as steps (Supplemental Fig. 2D, *left*). By association all three step amplitudes, *SNR* and σ remain small, too. With progressing idealization, the number of levels decreases and step amplitudes increase until finally only the "real" photobleaching steps overcome the threshold σ. The algorithm, however, does not actually aim to detect steps, but inversely scans for periods of constant intensity.

The slow progression reduces the probability of detecting variations in the trace due to noise or blinking events as photobleaching events. Also, the slow progression reduces the probability of missing low amplitude photobleaching events. Despite the algorithm's iterative nature, final idealizations are generally obtained within fractions of a second using a modern PC. The computing time decreases with increasing idealization as only differences between levels are evaluated and not each data point.

TRACE REJECTION

The step detection algorithm will attempt to find step-wise behaviour in any provided trace, even though that trace may not actually contain any steps. Therefore, after step detection, traces must be evaluated by a set of criteria as an estimate of trace quality. Five criteria were implemented into PIF for the evaluation of the step-wise behaviour: Thresholds for the maximal Chi-squared, for *SNR*, for the maximal time allowed for all fluorophores to bleach and for the maximal step amplitude.

a. The Chi-squared ratio between the actual fit and a counter-fit for a given trace is a strong indicator of the step-wise behaviour. The counter-fit is computed as the average value between the centers of adjacent levels, as if the steps were located in the middle of two real consecutive steps (1). Small Chi-squared values are indicative of a better fit, and high values, such as those expected of a counter-fit, are indicative of poor fits. Consequently, a high ratio (> 1) between chi-squared values of the counter-fit and of the actual fit generally correspond to traces exhibiting clear step-wise behaviour (1). Accordingly, a ratio < 1 implies that the counter-fit is a better representation of the trace than the actual fit itself. Therefore, traces with a ratio < 1 are excluded from analysis.

b. Evaluating the quality of the fit for a given trace is achieved by calculating a Chi-squared value for the fit. The ability of PIF to fit a trace that does not exhibit clear step-wise behaviour (e.g. resulting from random fluorescence variations) is poor. In such cases, a goodness of fit (χ^2_{fit}) based on a Chi-squared calculation can be evaluated as follows:

$$\chi^2_{fit} = \sum_i \frac{\left(Exp_i - Ideal_i \right)^2}{var_i} \qquad \qquad (eq.\ S3)$$

where Exp_i and $Ideal_i$ represent the data and the idealization at frame *i*, respectively. After idealization, the variance *var* is calculated for each intensity level found. If the trace contains too many false detections or missed events, χ^2_{fit} is expected to be smaller or larger than 1, respectively. False detections are generally not observed in our measurements because only the step-wise behaviour is fitted, not the blinking events or any imperfection within the trace. As a result, the threshold for rejection by the goodness of fit χ^2_{fit} was empirically determined to be 1.5

c. Simulation experiments established that the accuracy of step detection of PIF is significantly reduced in traces where the SNR is <~2.5 after applying C-K filtering (corresponding to ~0.85 in the raw trace). Therefore, SNR was used as a rejection criterion where traces with SNRs < 2.5 after filtering were discarded. This should not influence the resulting step distributions because the SNR is independent of the step number. In experimental data almost every trace had SNRs above 2.5 after applying the C-K filter (Supplemental Fig. 1B).

d. The fourth criterion is based on the time required to bleach *n* fluorophores, which is set using the lifetime of the fluorophores. To determine this lifetime, an exponential decay curve is fitted to the summed intensity of all selected spots within a region of interest. This decay leads directly to the probability fulfilled by any fluorophore that it has photobleached after excitation time *t*:

$$P_{bleach}\left(t \right) = \left(1 - e^{-t/\tau} \right) \qquad \qquad (eq.\ S4)$$

From this relation, one can compute the time required to have a probability $P_{time}(n)$ of getting all *n* fluorophores photobleached:

$$P_{time}(n) = \left(1 - e^{t/\tau}\right)^{n} \qquad\qquad (eq.\ S5)$$

$$t = -\tau \cdot \ln\left(1 - P_{time}^{1/n}\right) \qquad\qquad (eq.\ S6)$$

The last relation can be used to determine the time when with a certain probability P_{time} all fluorophores have photobleached dependent on the number of steps detected. If a trace still shows steps after this allowed period of time, it is rejected. We found that setting P_{time} between 90% and 95% does not appear to reject clear traces, but mainly those where the exponential decay background was not successfully removed, leading to detection of *false* steps typically towards the end of the traces. It also helps to reject contaminant spots, as they are often represented as "unclear" single photobleaching step traces with long duration.

e. This single photobleaching event of the background fluorescence often exhibited an amplitude that was consistently significantly larger than steps attributed to the photobleaching of the fluorophores. These contaminants were removed by setting a rejection criterion based on the maximum allowable step amplitude. Traces showing step amplitudes greater than 2550 arb.u. (~3 times the amplitude of the average step size obtained from a step amplitude distribution) were rejected (Supplemental Fig. 3E).

CALCULATING THE PROBABILITY OF MISSED EVENTS

As missed photobleaching events could have an influence on the step frequency distribution (*see* Supplemental Fig. 3B), here we show the derivation of the probability of missing a photobleaching event ($P_{miss}(\Delta t)$) for a given time resolution Δt. As during Δt only a single step may be detected even though 2 or more occurred, P_{miss} is described by the probability that 2 or more photobleaching events occurred within the same interval Δt, while the other remaining fluorophores do not photobleach during this time interval. To evaluate this, the previously described probability $P_{bleach}(t)$ is used as a starting point. This probability represents the cumulative probability distribution of the probability density function $P_d(t)$ as follow:

$$P_d(t) = \frac{d}{dt} P_{bleach}(t) = \frac{e^{-t/\tau}}{\tau} \qquad\qquad (eq.\ S7)$$

$P_d(t)$ is useful to determine the probability that a fluorophore photobleaches during a given time interval by a simple integration. For instance, the integral of $P_d(t)$ between 0 and t returns the relation $P_{bleach}(t)$, as it represents the probability of being photobleached after time t. Also, the integral from 0 to infinity results to 1, as the fluorophore will photobleach eventually. From this probability density, we estimate the probability of missing 1 photobleaching event by the multiplication of the probability that 2 fluorophores photobleach during the same time interval i [t_i, $t_i+\Delta t$] and the probability that the other n-2 remaining fluorophores do not photobleach during this time interval, as shown in the equation S8. Finally, all time intervals i must be summed up to consider the whole trace. As fluorophores are not distinguishable, we have to consider their permutation as another multiplicative factor

- 4 -

$$P^1_{miss}(\tau, \Delta t) = \sum_i \frac{n!}{2!(n-2)!} \left(\int_{t_i}^{t_i+\Delta t} \frac{e^{-t/\tau}}{\tau} dt \right)^2 \left(1 - \int_{t_i}^{t_i+\Delta t} \frac{e^{-t/\tau}}{\tau} dt \right)^{n-2}$$

$$= \sum_i \frac{n!}{2!(n-2)!} \left(e^{-t_i/\tau} \left(1 - e^{-\Delta t/\tau} \right) \right)^2 \left(1 - e^{-t_i/\tau} \left(1 - e^{-\Delta t/\tau} \right) \right)^{n-2}$$

(*eq. S8*)

To calculate the probability of missing *k-1* events, the formula has to be adjusted in the following way:

$$P^{k-1}_{miss}(\tau, \Delta t) = \sum_i \frac{n!}{k!(n-k)!} \left(e^{-t_i/\tau} \left(1 - e^{-\Delta t/\tau} \right) \right)^k \left(1 - e^{-t_i/\tau} \left(1 - e^{-\Delta t/\tau} \right) \right)^{n-k}$$

(*eq. S9*)

So, the probability of missing 1 or more photobleaching events is the sum of all the possibilities:

$$P_{miss}(\tau, \Delta t) = \sum_{k=2}^{n} \left(\sum_i \frac{n!}{k!(n-k)!} \left(e^{-t_i/\tau} \left(1 - e^{-\Delta t/\tau} \right) \right)^k \left(1 - e^{-t_i/\tau} \left(1 - e^{-\Delta t/\tau} \right) \right)^{n-k} \right)$$

(*eq. S10*)

In typical experiments, the lifetime τ is approximately 7 s and Δt = 150 ms. For n = 5 existing fluorophores, we then estimate P_{miss} to about 10.4%. If n = 4 or n = 3, P_{miss} is calculated to be about 6.3% and 3.2%, respectively.

Simulations using τ = 7 s and Δt = 150 ms were also done to estimate P_{miss}. 500 photobleaching traces were simulated 5 times for n = 5, n = 4 and n = 3. Each fluorophore was associated to a 2-state Markov model where the transition from the fluorescent state and the photobleached state occurs according to the rate constant $1/\tau$. Missed events were observed when 2 or more fluorophores photobleached within the same time resolution Δt. The results correspond nicely to the previous calculation. For n = 5, n = 4 and n = 3, P_{miss} was obtained to be 10.1 ± 1.3%, 6.4 ± 0.8% and 2.8 ± 1.1%, respectively.

REFERENCES

1. Kerssemakers, J. W., Munteanu, E. L., Laan, L., Noetzel, T. L., Janson, M. E., and Dogterom, M. (2006) Assembly dynamics of microtubules at molecular resolution. *Nature* **442,** 709-712

FIGURE LEGENDS

Supplemental Figure 1. (A) Illustration of the Markov model used to simulate photobleaching data. Only single transition blinking is allowed, described by the rate constants α and β. These two parameters respectively define the transition of a fluorophore from a non-fluorescent to a fluorescent state, and vice versa. The last parameter, the rate constant κ, describes the transition of a fluorophore from a fluorescent state to its photobleaching and is therefore unidirectional. The model allows up to 11 fluorophores, but it is possible to start from a state where x (<11) fluorophores are left (not photobleached) in order to simulate traces with different step number. (B) Effect of C-K filtering on the signal-to-noise ratio. Chung-Kennedy filter has the ability to increase the signal-to-noise ratio of a trace while keeping a steep edge for relevant intensity variations. To evaluate the effect of C-K filtering on SNR, we used the fully-automated analysis of msfGFP-GluK2, as the SNR was computed with and without C-K filtering for the same idealization. The resulting distributions reveal a SNR increase by a factor of approximatively 3.0 ± 0.5. (C) The sensitivity φ used in the step detection algorithm was optimized from simulations by choosing the value corresponding to the larger accuracy for a given *SNR*. φ was subsequently plotted as a function of *SNR* and fitted with a single exponential decay curve. The resulting exponential decay equation then became the calibration formula in the step detection algorithm. (D) Comparison of spots on-cell and off-cell and effect of LoG-type kernel filtering. Using a *δF/F* cut-off of 10%, the normalized intensity distributions on-cell (*blue*) and off-cell (*red*) are compared by calculating the average intensity of the spots over the first 100 frames (5 sec). From these distributions, the *δF* threshold was set to 1750 (arb.u.) for the actual system used. (E) The comparison in accuracy of PIF step detection and STEPFINDER is illustrated as a function of *SNR*. This evaluation was obtained using simulated traces containing 11 equidistant traces without blinking. Over the tested range of *SNR*, PIF reveals a larger accuracy than STEPFINDER. In the particular case of STEPFINDER, the accuracy was not precisely obtained for *SNR* lower than ~4, as the step distribution was very broad and accordingly very difficult to fit as previously described in "*Simulations: Calculation of the Accuracy p_{acc}*". STEPFINDER detected step number was obtained by maximizing the ratio between the Chi-squared value of a counter-fit and the one corresponding to the fit (1).

Supplemental Figure 2. PIF step detection algorithm described as conditions for segment averaging (see text for details). These conditions are iteratively repeated over averaged segments until convergence of the idealization. The progression of the algorithm can be divided into three main sections (phase 1,2,3), which are illustrated for a simulated trace containing 11 steps. Each algorithm section gradually removes the false detections within the trace.

Supplemental Figure 3. (A-D) Effect of step detection parameters on PIF automated analysis over GluK2. Step detection parameters were optimized by adding each parameter one at a time to the previous settings, as to check their influence on the overall step distribution. (A) Step distribution without using the parameters, except for the minimum time between steps which has to be set to 2 frames by default (see *Step Detection*). (B) The minimum time between steps is set to 3 frames (150 ms). This tends to decrease the number of false detections possibly due to C-K filtering and slight changes in the baseline, likely produced by an incomplete background removal (see Trace filtering in main text). This restriction does not significantly alter the probability of missing a step event, which remains below ~6% (see *Calculating the Probability of Missed Events* in the *Supplemental Text*). It is important to keep this value relatively small, as it could alter the final observed step distribution frequency by introducing a shift

toward smaller values. Depending on each situation, larger probabilities of missing a step might complicate the interpretation of the results. (C) the maximal step amplitude variation is in addition set to 60%. As false steps were small in comparison to the maximal step size detected within a trace, setting the maximal amplitude variation to 60% of the largest step effectively removed these artifacts. This limit still tolerates variations in amplitudes due to blinking events and allows the possibility of two fluorophores photobleaching during one frame exposure. (D) According to the step size distribution (see E), the minimum step size amplitude of 375 (arb.u.) is added to the previous settings, while traces containing steps over 2550 (arb.u.) are rejected. We found that the most common step amplitude detected in all the traces is about 750-850 (arbitrary units specific to our system, see E). Since excitation intensity remains uniform over the entire recorded field of view (maximum intensity variations of ~ 15%), we reasoned that it was unlikely to observe a fluorophore emitting less than half the intensity of the most commonly observed fluorescence intensity. Therefore, steps with intensities smaller than 375 (arb.u.) are attributed to artifacts of a decreasing baseline. Since concomitant multiple photobleaching steps are possible, though very infrequent, an upper step amplitude limit was also set as 3 x 850 = 2550 (arb.u.). Traces containing steps outside this intensity range were rejected. (E) The step size distribution of detected steps was obtained from accepted traces of the distribution obtained in C.

Supplemental Figure 4. Effect of blinking on step detection accuracy. Sets of 500 simulated traces containing 4 photobleaching steps with different blinking properties were analyzed by PIF step detection algorithm for different SNRs. The accuracy curves using $\alpha = 1$ (black squares), 5 (red triangles), 10 (green circles), 50 (blue stars) are illustrated for $\beta = 0.1$ s^{-1} (*left*) and $\beta = 1$ s^{-1} (*right*).

Supplemental Figure 1
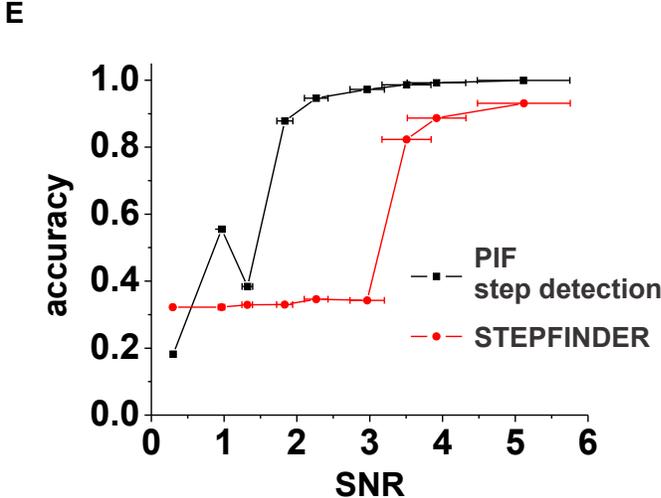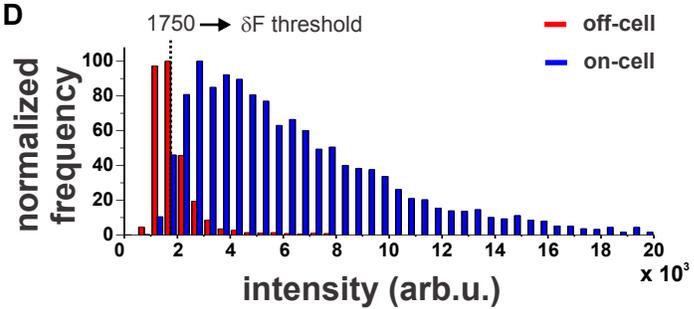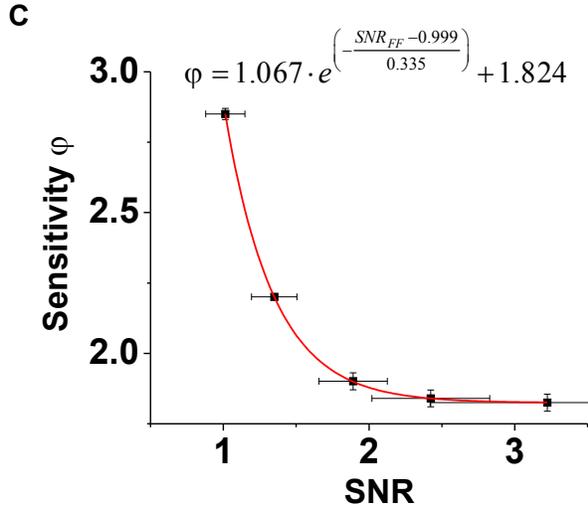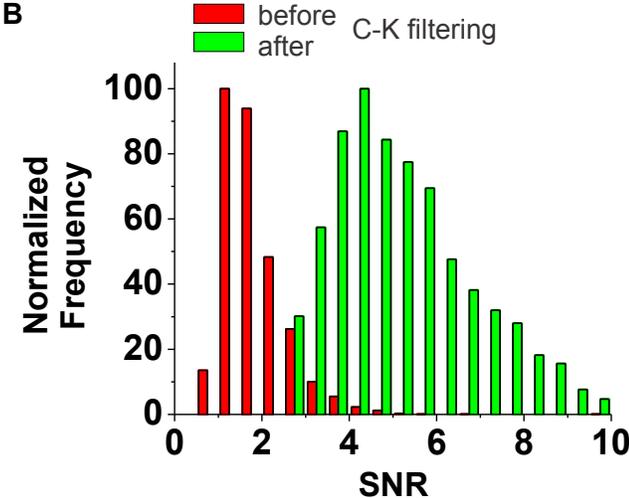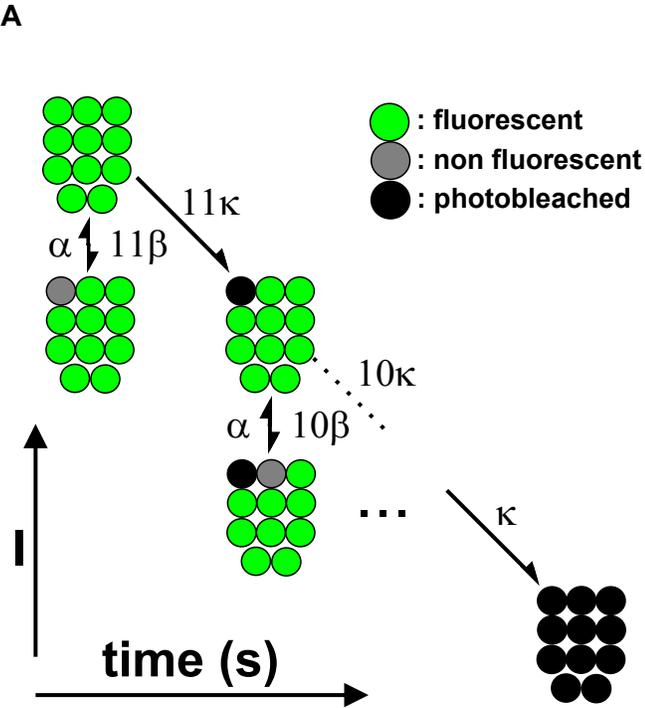
Supplemental Figure 2

A **if larger intensity is found then average segment**
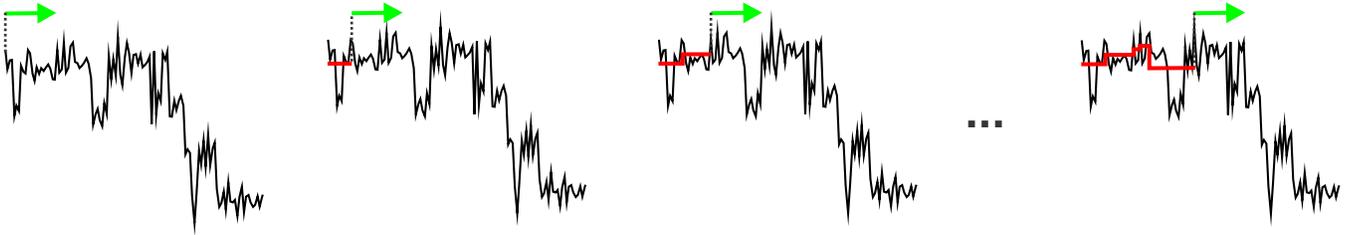
B **if intensity drop < σ_x then average segment**
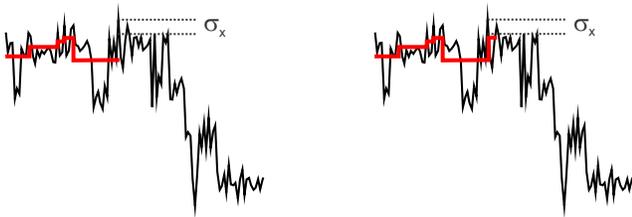
C **last position where intensity drop < σ_x level is ended; go to the next one**
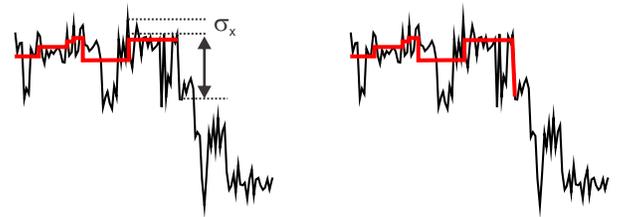
D **After the whole trace is segmented, subsequent iterations start over using the previous averaged segments until the idealization no longer changes**

E phase 1

↓ : Over-fits

F phase 2

↓ : Over-fits
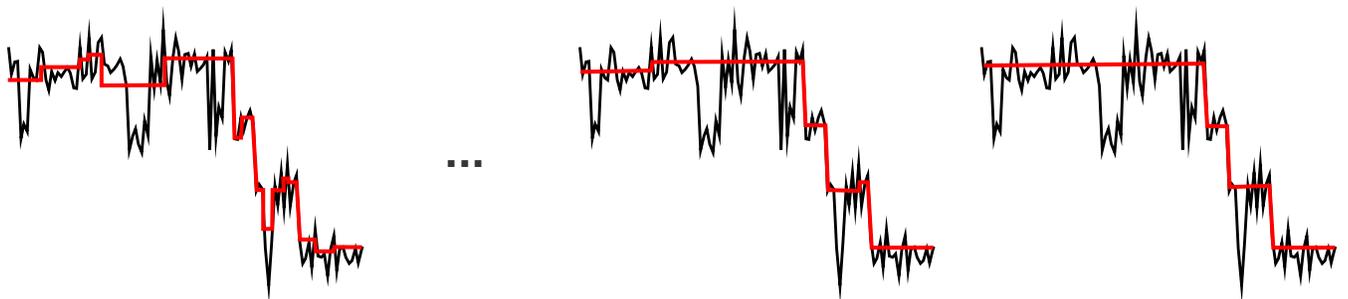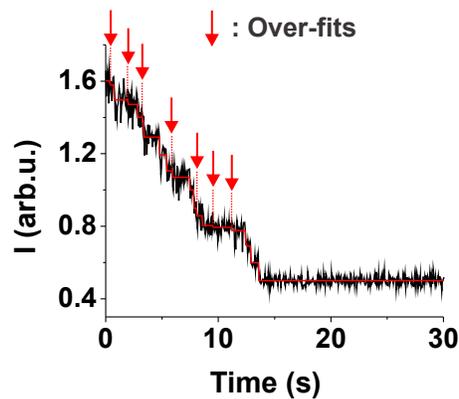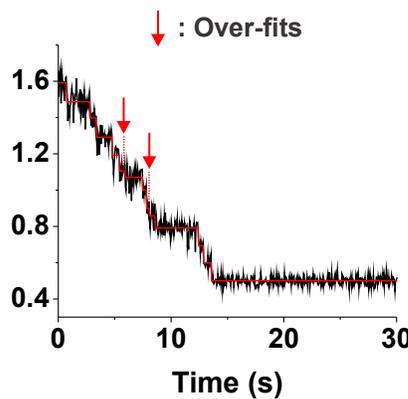
G phase 3

Supplemental Figure 3