

1 Article

# 2 Impact of molecule concentration, diffusion rates and surface 3 passivation on single-molecule fluorescence studies in solution

4 Olessya Yukhnovets <sup>1,\*</sup>, Henning Höfig <sup>1</sup>, Nuno Bustorff <sup>2</sup>, Alexandros Katranidis <sup>2</sup> and Jörg Fitter <sup>1,2,\*</sup>

5 <sup>1</sup> RWTH Aachen University, I. Physikalisches Institut (IA), AG Biophysik, Aachen, Germany

6 <sup>2</sup> Forschungszentrum Jülich, ER-C-3 Structural Biology / IBI-6 Cellular Structural Biology, Jülich, Germany

7 \* Correspondence: [yukhnovets@physik.rwth-aachen.de](mailto:yukhnovets@physik.rwth-aachen.de) (O.Y.) and [fitter@physik.rwth-aachen.de](mailto:fitter@physik.rwth-aachen.de) (J.F.)

8 **Abstract:** For single-molecule studies in solution, very small concentrations of dye-labelled mole-  
9 cles are employed in order to achieve single molecule sensitivity. In typical studies with confocal  
10 microscopes, often concentrations in the pico-molar regime are required. For various applications  
11 which make use of single-molecule Förster resonance energy transfer (smFRET) or two-color coinci-  
12 dence detection (TCCD), the molecule concentration must be set explicitly to targeted values and  
13 furthermore needs to be stable over several hours. As a consequence, specific demands must be  
14 imposed on the surface passivation of the cover slides during the measurements. The aim to have  
15 only one molecule in the detection volume at the time is not only affected by the absolute molecule  
16 concentration, but also by the rate of diffusion. Therefore, we discuss approaches to control and to  
17 measure absolute molecule concentrations. Furthermore, we introduced an approach to calculate  
18 the probability of chance coincidence events and demonstrate that measurements with challenging  
19 smFRET samples require a strict limit of maximal sample concentrations in order to produce useful  
20 results.

21 **Citation:** Yukhnovets, O.; Höfig, H.;

22 Bustorff, N.; Katranidis, A.; Fitter, J.

23 Impact of molecule concentration,

24 diffusion rates and surface pas-

25 sivation on single-molecule fluo-

26 rescence studies in solution. *Biomole-*

27 *cules* **2022**, *12*, x.

28 <https://doi.org/10.3390/xxxxx>

29 Academic Editor: Firstname Last-  
30 name

31 Received: date

32 Accepted: date

33 Published: date

34 **Publisher's Note:** MDPI stays neu-  
35 tral with regard to jurisdictional  
36 claims in published maps and institu-  
37 tional affiliations.



38 **Copyright:** © 2022 by the author  
39 Submitted for possible open access  
40 publication under the terms and  
41 conditions of the Creative Commons  
42 Attribution (CC BY) license  
43 (<https://creativecommons.org/licenses/by/4.0/>).

21 **Keywords:** single-molecule Förster resonance energy transfer; burst analysis; Two-color coinci-  
22 dence detection; confocal fluorescence microscopy; chance coincidence probability

## 24 1. Introduction

25 One major approach to investigate biomolecule conformations, dynamics or molec-  
26 ular interactions on single molecule level is given by fluorescence based techniques which  
27 make use of confocal microscopy. In this case a diffraction limited small detection volume  
28 is created by an optical setup which allows to measure fluorescence emission from indi-  
29 vidual diffusing molecules during a rather short time of the molecule's transit through  
30 this detection volume [1,2]. For medium sized proteins and confocal detection volumes in  
31 the order of a few femtoliter, the dwell time of the diffusing molecule is in the millisecond  
32 time regime. During this time, the emitted "burst of photons" from individual molecules  
33 represents the essential feature of a single molecule [2,3]. With the help of powerful burst  
34 analysis tools, the relevant data can be extracted from typical time trace measurements  
35 [4,5]. In this way, up to a few thousand bursts can be measured within hours, which de-  
36 livers reasonable counting statistics for a proper single molecule analysis [6,7].

37 An important prerequisite to achieve single molecule sensitivity by using this ap-  
38 proach requires that the absolute molecule concentration has to be kept to a value, that  
39 corresponds to the presence of not more than one molecule (or one type of molecule) at  
40 the time in the confocal detection volume. Due to the stochastic character of the Brownian  
41 diffusion the molecules perform in solution, it is typically assumed that this requirement  
42 is fulfilled if the average number of molecules in the detection volume (~ few femtoliters)  
43 is given by  $\langle N \rangle \sim 10^{-2}$ , see for example [8]. In most applications with diffraction limited

detection volumes in confocal microscopy, this requirement is related to molecule concentrations of a few ten pico-molar. For two-color coincidence detection (TCCD) studies [7, 9], which are for example employed to measure bimolecular binding affinities, one need not only to fulfill the mentioned requirement. In addition, also the absolute concentrations of the two differently labelled species need to be set to targeted values. Furthermore, the average number of molecules in the detection volume at the time depends not only on the molecule concentration but also on the diffusion rate of the molecules. Here we discuss specific single molecule Förster resonance energy transfer (smFRET) and TCCD applications for which even higher demands on sample properties are required in order to obtain reliable results. In this respect methodical approaches are introduced which help to monitor and control the molecule concentration. Finally, a reduced quality of the obtained results is demonstrated for data which is measured with samples of a sub-optimal concentration regime.

## 2. Materials and Methods

### 2.1 Double-stranded DNA

Hybridization procedure and detailed sample preparation procedures were published earlier [10]. Briefly, dsDNA samples were prepared by hybridizing a labelled ssDNA strand 5'-GGA CTA GTC TAG GCG AAC GTT TAA GGC GAT CTC TGT TTA CAA CTC CGA-3' with an unlabelled ssDNA strand 5'-TCG GAG TTG TAA ACA GAG ATC GCC TTA AAC GXT CGC CTA GAC TAG TCC-3' (IBA, Göttingen, Germany). The labelled strand was either having an Alexa488 dye at 5' end and an Atto647N dye at 3' end or just a single Atto647N dye at 3' end to produce a double or a single labelled dsDNA, respectively. After hybridization, samples were aliquoted and stored at  $-20^{\circ}$  C. If not stated otherwise, the dsDNA samples were measured in a DNA standard buffer: 20 mM TRIS pH 7.5, 100 mM NaCl, 10 mM MgCl<sub>2</sub>.

### 2.2 Phosphoglycerate kinase

Phosphoglycerate kinase (PGK) expression procedure was described in detail in [11]. Protein labelling and purification is described in [12]. Briefly, single cysteine mutant PGK C97S Q135C was produced by site-directed mutagenesis. Protein samples were frozen in liquid nitrogen and stored at  $-80^{\circ}$  C. The single cysteine mutants, was labelled either with maleimide functionalized Alexa647 or Alexa488 by incubating 10  $\mu$ M PGK solution with 5-fold excess of dye. After labelling, the protein was purified from unbound dye excess by size exclusion chromatography, using Sephadex G25 packed column. Labelled protein samples were stored at  $4^{\circ}$ C for a maximum of a few days.

### 2.3 Ribosomes

Isolation and labelling of ribosomes is described in [7]. Ribosomes were isolated by zonal centrifugation and incubated with 20-fold excess of Cy5-NHS-ester functionalized dye (GE Healthcare Life Sciences, Little Chalfont, UK). Labeled ribosomes were purified by pelleting the ribosomes through a 1.1 M sucrose cushion in Tico buffer: 20 mM HEPES pH 7.6, 10 mM magnesium acetate (Ac<sub>2</sub>Mg), 30 mM ammonium acetate (AcNH<sub>4</sub>), 4 mM  $\beta$ -mercaptoethanol ( $\beta$ -ME). The concentration of Cy5 and ribosomes was determined spectroscopically in a NanoDrop spectrophotometer (Thermo Fisher Scientific, Waltham, USA) using the absorption coefficients  $\epsilon_{\text{cy5}} = 2.5 \times 10^5 \text{ M}^{-1} \text{ cm}^{-1}$  (at  $\lambda = 647 \text{ nm}$ ) and  $\epsilon_{70\text{S}} = 4.2 \times 10^7 \text{ M}^{-1} \text{ cm}^{-1}$  (at  $\lambda = 254 \text{ nm}$ ), respectively. The label ratio was calculated to be  $\sim 6$  Cy5 dyes per 70S ribosome.

### 2.4 Ribosome nascent chain of Calmodulin

Ribosome nascent chains (RNCs) of Calmodulin (CaM) were synthesized to full polypeptide chain length (149 amino acid residues) using a cell free protein synthesis system

[13]. For keeping the nascent chain bound to the ribosome, an enhanced arrest peptide sequence from *E. coli* protein SecM, named SecMstr (FSTPVWIIWWPRIRGPP) was used [14]. The arrest peptide was introduced downstream of CaM connected via a linker composed of a 30 amino acid long sequence of glycines and serines spanning the length of the ribosomal tunnel. At the N-terminus, a twin strepII tag (WSHPQFEKGGGSGGGSGGSAWSHPQFEK) was also introduced for affinity purification. Two unique unnatural amino acids (UAAs), namely CpK and AzF were incorporated co-translationally at determined positions (34 and 110). The side chains of these two UAAs have functional groups that can react with functionalized dyes to produce double labelled RNCs, with an inter-dye distance suitable to observe FRET [15]. The reactions were performed using an *E. coli* high yield (Biotechrabbit) or a PURExpress (NEBiolabs) cell free system. The reaction was stopped by adding two volumes of RNC buffer (20 mM HEPES pH 7.5, 50 mM Ac<sub>2</sub>Mg, 30 mM AcNH<sub>4</sub>, 0.5 mM TCEP, 0.005% Tween20). RNCs were purified using magnetic Strep-TactinXT beads and labelled sequentially with 50 µM of red AF647-DBCO and blue AF488-tetrazine (Click Chemistry Tools, Scottsdale, USA) dyes. After each labelling step, the excess of dye was removed using Zeba columns (Thermo Fisher Scientific, Waltham, USA), previously equilibrated with apo-buffer (10 mM EGTA, 50 mM MOPS, 150 mM KCl, 0.005 % Tween20) [16], for subsequent smFRET measurements.

### 2.5 Confocal microscopy

Confocal measurements were performed using a MicroTime200 (PicoQuant, Berlin, Germany). The fluorophores were excited using LDH-D-C 485B and LDH-D-C 640B lasers with 485 nm and 640 nm emission (PicoQuant, Berlin, Germany) and a power of 21 µW and 18 µW, respectively. For smFRET and BTCCD measurements, lasers were operated in a pulsed-interleaved excitation (PIE) scheme, in which blue and red excitation is alternated in order to directly excite both channels [17]. The excitation light was focused and collected by a high numerical aperture water immersion objective (UPLSAPO 60x; Olympus, Hamburg, Germany) and directed through a 75 µm pinhole. The emission signal was separated by a dichroic mirror (T600lpxr, Chroma Technology, Olching, Germany) and filtered by band pass filters of 535 nm (FF01-535/55-25, Semrock, Rochester, NY, USA) and 685 nm (ET685/80m, Chroma Technology, Olching, Germany) for the blue and the red channel, respectively. Photons were detected by single-photon avalanche diodes (τ-SPAD, PicoQuant, Berlin, Germany; COUNT-T, Laser Components, Olching, Germany).

### 2.2 Sample preparation and data acquisition

Unless stated otherwise, all samples were measured on PEGylated cover slides. For PEGylation procedure high precision cover glasses of 170 µm thickness (Marienfeld, Germany) were cleaned with Piranha solution, plasma cleaned, treated with silane and left to react overnight with NHS-functionalized PEG (MW = 750 Da, Rapp Polymere, Tübingen, Germany) solution. Concentration of all samples was first determined with FCS and then samples were diluted to aimed single-molecule concentrations. Afterwards, the real single-molecule concentration was determined with bursts analysis. Single-molecule sample aliquots were measured within 10-20 min time intervals which were summed up afterwards for the analysis. Typical data acquisition time was about 60-180 min for a dataset. In order to maintain constant concentration and avoid evaporation during long measurement time, sample holders were sealed with parafilm.

### 2.7 Burst analysis

The inter photon distance (IPD) trace was calculated for acquired intensity traces [18]. Single bursts in both, the red and the blue, channels were discriminated from the background by applying a suitable threshold (usually ~ 100 µs). Bursts that contain only one photon were discarded because they would induce an artificially small dwell time and low molecular brightness. The start time of each burst corresponds to the macro time tag

of the first photon of that burst and, accordingly, the end time of a burst is defined as the macro time tag of the last photon. The burst duration is defined as the difference between its start and end time. Bursts with the duration time of more than 100-fold longer than the average bursts duration time and bursts with the number of photons of more than 100-fold more than the mean number of photons per bursts were considered as aggregates or contamination and were removed from the analysis. Typical data sets contain  $10^3$ - $10^4$  number of accepted bursts.

### 2.8 Brightness-gated two-color coincidence detection

Brightness-gated two-color coincidence detection (BTCCD) is a method realized by means of simultaneous single-molecule two-color confocal detection to quantify the fraction of bound (coincident) molecules. In contrast to conventional two-color coincidence detection (TCCD, see for example [9]), BTCCD overcomes the problem of coincidence fraction underestimation, caused by incomplete detection volume overlap for different excitation wavelength and lens aberrations. In order to estimate the coincidence fraction precisely, only molecules that diffused through both confocal volumes should be considered for analysis. In practice, it is assumed that the corresponding molecule trajectories resulted in bright bursts with high number of photons are more likely to touch both volumes, whereas dim bursts with only small number of photons are more likely to touch one of the volumes only slightly. For each accepted burst the burst intensity, i.e. number of photons detected between the start and end time and the mean number of photons per burst is calculated. To perform a coincidence analysis, the brightness threshold  $n_{br}$ , defined as the number of photons in a burst, normalized to the mean number of photons, is continuously increased. The coincidence is calculated for the red channel ( $f_{RB}$ ) and the blue channel ( $f_{BR}$ ) independently with

$$f_{RB}(n_{br}) = \frac{N_{RB}(n_{br})}{N_R(n_{br})}, \quad f_{BR}(n_{br}) = \frac{N_{BR}(n_{br})}{N_B(n_{br})} \quad (1)$$

where  $N_{RB}$  and  $N_{BR}$  are the number of coincident bursts in the red and blue channel, and  $N_R$  and  $N_B$  are the total number of selected red and blue bursts, respectively. For each value of the brightness threshold, only bursts that have more photons as defined by the brightness threshold are considered for analysis. Two bursts are considered as coincident if the start or end time tag of one burst is within the start and end time tags of the other burst. Coincidence fraction increases with the increase of  $n_{br}$  and eventually saturates once all bursts considered for the analysis correspond to molecule trajectories through both volumes. A complete description of the BTCCD method can be found in [7].

### 2.9 Single molecule FRET analysis

After selecting bursts by means of burst analysis, photon counts are calculated for each accepted burst. The use of PIE excitation scheme allows to calculate photon counts for acceptor emission after donor excitation ( $I_{AD}$ ), acceptor emission after acceptor excitation ( $I_{AA}$ ) and donor emission after donor excitation ( $I_{DD}$ ) and eliminate donor-only and acceptor-only populations from the analysis. For bursts with  $I_{AD} + I_{DD} > 20$  photon counts, the energy transfer efficiency

$$E = \frac{I_{AD}}{I_{AD} + \gamma \cdot I_{DD}} \quad (2)$$

was calculated, where  $\gamma$  is a correction factor accounting for differences in dye quantum yields and detection efficiencies in donor and acceptor channels. The stoichiometry parameter is given by

$$S = \frac{I_{AD} + I_{DD}}{I_{AD} + I_{DD} + I_{AA}} \quad (3)$$

Furthermore,  $E$  and  $S$  values were corrected for cross-talk and direct excitation as described in [19]. In order to evaluate the quality of the obtained smFRET data, FRET efficiency histograms and two-dimensional efficiency-stoichiometry histograms are plotted. Finally, FRET efficiency histograms were fitted with Gaussians in order to characterize individual subpopulations.

### 2.10 Concentration determination in single-molecule measurement

In order to calculate the molar concentration of a sample  $C = N_{av}/(N_A \cdot V_{eff})$  in a single-molecule experiment, the average number of detected molecules  $N_{av}$  (for simplicity from now on depicted only by  $N$ ) and the dimension of confocal detection volume  $V_{eff} = \pi^{3/2} \omega_0^2 z_0$  need to be known ( $N_A$ : Avogadro constant). The latter is directly accessible from FCS calibration measurements.  $N$  can be calculated from the total number of detected bursts,  $B_{meas}$ , and the duration time of detected bursts,  $\tau_d$ , by considering total fluorescence time  $t_F$  [4]. Total fluorescence time is defined as the product of total measurement time  $t_{meas}$  and the probability to detect a molecule ( $1 - \exp(-N)$ ) or by the product of  $B_{meas}$  and  $\tau_d$

$$t_F = [1 - \exp(-N)] \cdot t_{meas} = B_{meas} \cdot \tau_d. \quad (4)$$

This equation can now be used to determine the average number of detected molecules:

$$N = -\ln\left(1 - B_{meas} \frac{\tau_d}{t_{meas}}\right). \quad (5)$$

Typical values for calculated molar concentrations as obtained from confocal measurements with different pinhole diameters are shown in Table 1.

**Table 1.** Comparison of different molar molecule concentrations as determined from  $N$  for different confocal detection volumes.

Pinhole $\emptyset$	30 $\mu\text{m}$	50 $\mu\text{m}$	75 $\mu\text{m}$	150 $\mu\text{m}$
	<sup>1</sup> $V \approx 0.5$ fL	$V \approx 1$ fL	$V \approx 2$ fL	$V \approx 4$ fL
$N$	C [pM]			
0.001	1.66	0.83	0.55	0.33
0.005	8.31	4.15	2.77	1.66
0.01	16.61	8.31	5.54	3.32
0.02	33.22	16.61	11.07	6.64
0.03	49.83	24.92	16.51	9.97
0.05	83.06	41.53	27.69	16.61
0.1	166.11	83.06	55.37	33.22

<sup>1</sup> Effective detection volumes depend on the excitation and fluorescence emission wavelengths. The values given here represent a mean value, as obtained by averaging from 485 nm and 640 nm wavelengths conditions.

### 2.11 Chance coincidence events

In order to perform single-molecule experiment, the probability to detect more than one molecule must be negligibly low [20]. Knowing the probability of the multi-molecule event detection (i.e., having a number of current molecules  $N_c > 1$  in the detection volume) allows to maintain a proper single-molecule concentration regime. As shown in [21], it is calculated from Poisson distribution, with the probability to detect two molecules at a time is given by

$$pm_{n=2}(N) = \exp(-2N)(1 - \exp(-N)). \quad (6)$$

The probability to detect more than two molecules ( $N_c > 2$ ) is negligibly small for  $N \ll 1$ . Therefore, we define chance coincidence events as the presence of two molecules at a time in the detection volume.

For BTCCD, chance coincidence only matters for cross-color multi-molecule events, i.e. when a blue-labelled molecule enters the volume while the red-molecule is in the volume and vice versa. The probability to detect a blue-labelled molecule during the dwell time of a red-labelled molecule will depend on burst duration time in red and blue channels,  $\tau_d^R$  and  $\tau_d^B$ , and on the average number of molecules in blue channel  $N_B$

$$p_{RB} = 1 - \exp\left(-\left(\frac{\tau_d^R}{\tau_d^B}\right)N_B\right). \quad (7)$$

Blue-labelled molecule can either already be in the detection volume or enter during the dwell time of a red-labelled molecule, meaning that both probabilities need to be added. Also, because only cross-color events should be considered, only red-only molecules will cause a chance coincidence detection and a  $(1 - f_{RB})$ -factor should be introduced. The chance coincidence fraction is then calculated as

$$f_{RB}^{Chance} = (1 - f_{RB}) \left[ 1 - \exp\left(-N_B \left(\frac{\tau_d^R}{\tau_d^B} + 1\right)\right) \right]. \quad (8)$$

In the same manner, the chance coincidence fraction accounting for detecting a red-labelled molecule during the dwell time of a blue-labelled molecule is

$$f_{BR}^{Chance} = (1 - f_{BR}) \left[ 1 - \exp\left(-N_R \left(\frac{\tau_d^B}{\tau_d^R} + 1\right)\right) \right]. \quad (9)$$

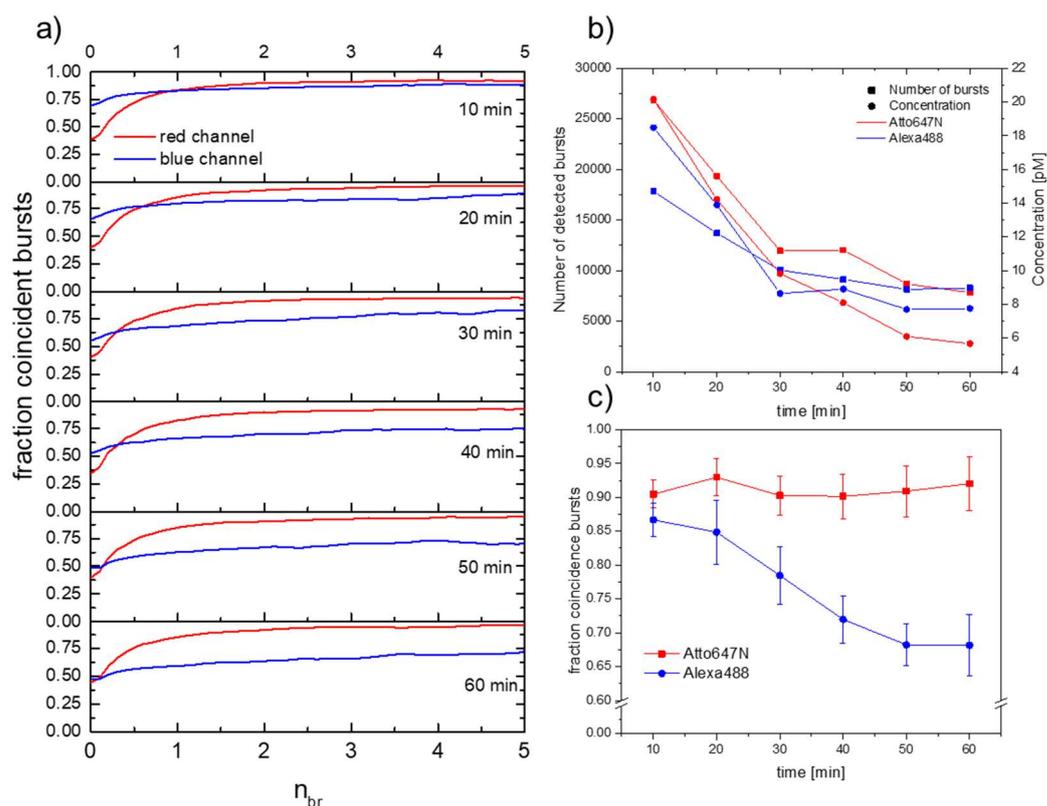
### 3. Results and Discussion

In practice samples for confocal fluorescence spectroscopy often consist of a rather heterogeneous ensemble of labelled molecules. This heterogeneity can originate from the fact, that (i) the biological molecules of interest are present in different conformational states or that (ii) individual biomolecules exhibit different degrees of labelling. For example, in the case of FRET studies which require donor and acceptor double-labelled species, donor only or acceptor only species might be present in the sample solution. In most single-molecule studies, problems related to such a sample labelling heterogeneity can either be avoided by highly productive sample preparation and purification procedures [22,23] or circumvented by elaborated molecule sorting algorithms [17, 19]. However, in some cutting-edge applications sample heterogeneity can still cause serious problems. In this respect, we will focus here mainly on two types of samples: (i) Proteins that are produced by cell-free synthesis systems allow the incorporation of unnatural amino acids for more selective double dye conjugation. Although these smFRET samples offer the opportunity for studying co-translational protein folding or multi-protein complex assembly, the obtained proteins suffer often from rather incomplete labelling and low protein yields [15, 24]. As a consequence, the typically applied protein purification procedures (for example to remove free dyes) also work less efficient. (ii) The strength of the BTCCD approach is to determine bi-molecular binding affinities in the low pico-molar regime. This requires also low molecule concentrations of the individual complementary single labelled binding partners. For both mentioned types of samples, a rather low molecule concentration in combination with a certain tendency for unspecific molecule attachment to the cover-slide surface may cause problems and artifacts which will be discussed in the following sub-chapters.

#### 3.1. Unbalanced reduction of molecule concentrations during extended measurement times

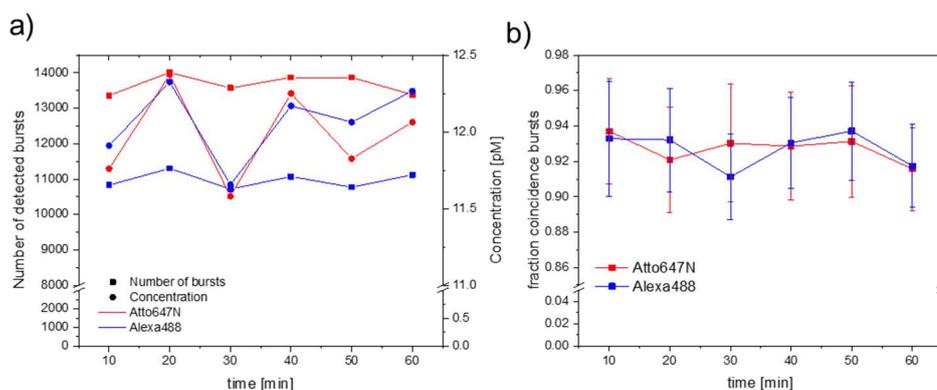
In typical BTCCD studies, calibration measurements have to be performed in order to validate that brightness-gating (the use of increasing  $n_{br}$  values) reduces the effects of the confocal detection volume mismatch (related to the two different excitation wavelengths) reasonably [7]. For this purpose, a double stranded DNA molecule labelled with two different dyes is employed. Since the label ratio for each color (i.e. each dye with its

respective absorption wavelength) is very high, typically larger the 95 %, we obtain very high coincidence fractions  $f_{RB}$  and  $f_{BR}$ , respectively (see subsection 2.8, eq. 1). In the case for a dsDNA (length 48 bp) labelled with Alexa488 and Atto647N the expected high coincidence fractions are visible for  $n_{br} > 2$  (see Fig. 1a, topmost panel). A closer look on the panels in Fig. 1a demonstrates that the coincidence fraction  $f_{RB}$  (blue) decrease substantially over a time span of 60 minutes, while the coincidence fraction  $f_{BR}$  (red) remains more or less unchanged during the same time span. This result is shown more clearly in Fig. 1c, where  $f_{BR}$  drops from 87% to 68 % and  $f_{RB}$  exhibits a constant value of about 90%. Such a behavior can be explained by the following facts: (i) There is a small fraction of single labelled species, 3-5 % only labelled with Alexa488 or with Atto647N. (ii) There is always a certain, often rather small, fraction of molecules that nonspecifically sticks to the surface, even in the case of PEGylated cover-slides. In our case, one fluorescent dye seems to show a tendency for the stickiness and obviously, it is the Atto647N which exhibits this tendency. As a consequence, the total molecule concentration of dsDNA molecules drops down over time. Since the majority of the diffusing molecules is still double labelled, the concentration for both colors is reduced from approximately 18-20 pM to 6-8 pM after 60 minutes (see Fig. 1b). However, a further consequence is that the single labelled Atto647N molecules are systematically extracted from the solution, while at the same time the single labelled Alexa488 molecules remain in the solution. Exactly this mismatch causes a decreasing  $f_{RB}$  (blue) and a constant  $f_{BR}$  (red) over time.



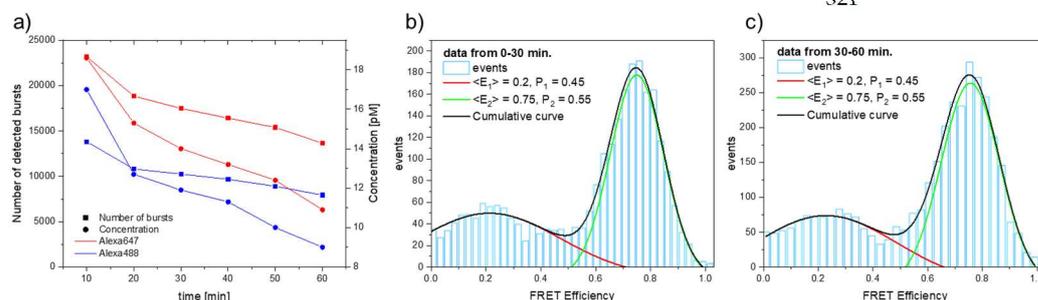
**Figure 1.** BTCCD results of double-labelled dsDNA samples as obtained from measurements at standard buffer conditions (only PEGylated surfaces). (a) Coincidence fractions,  $f_{RB}$  (blue) and  $f_{BR}$  (red) are shown as a function of the brightness-gating parameter  $n_{br}$  for different time intervals after the measurements were started (from top to bottom). The measurements have been performed with PEGylated cover-slides (see subsection 2.6) and in standard DNA-buffer (see subsection 2.1). (b) Based on the data shown in (a), the time course of the dye-concentration in solution is given in terms of “total number of bursts” as obtained in the respective time interval (see left y-axis) and in molar concentration (see right y-axis). (c) Based on the data shown in (a), the time course of the individual coincidence fractions  $f_{RB}$  (blue) and  $f_{BR}$  (red) are shown.

By adding a small amount of the detergent Tween20 to the buffer solution, the stickiness of Atto647N almost vanishes (see Fig. 2a,b). The corresponding results from measurements with Tween20 exhibit nearly constant molecule concentrations for all species, as well as stable and almost identical coincidence fractions  $f_{RB}$  and  $f_{BR}$  (for  $n_{br}$  values above a certain threshold), as expected for this type of sample, see [7].



**Figure 2.** (a,b) For data which was obtained from measurements in buffers containing 0.005% Tween20, the same type of graphs is shown as in Fig. 1b and c, respectively.

In another case we re-evaluate smFRET data from measurements, performed in a multidomain protein folding study [12]. In this study the conformational structure of phosphoglycerate kinase (PGK) was measured as a function of the chemical denaturant GndHCl. As observed already in various studies, proteins typically undergo a structural expansion upon increasing GndHCl concentrations which represents an unfolding transition [3,25]. The corresponding succession of smFRET histograms shows first a population for the native state (compact high FRET state). With increasing GndHCl concentrations a further population of the unfolded state shows up, while well above the GndHCl half-transition concentration (for PGK  $C_{1/2} \sim 0.65$  M) only the peak of the unfolded state (expanded low FRET state) remains. Since GndHCl is not only an efficient chemical denaturant, but also a good solvent (for molecules that expose hydrophobic regions), the GndHCl concentration determines not only the degree of unfolding, but also the tendency of partly unfolded proteins to stick to the cover-slides during extended measuring times.



**Figure 3.** Results as obtained from smFRET measurements with PGK at 0.65 M GndHCl denaturant concentration. (a) The number of detected bursts and the related molar protein concentration is given during the total measuring time (60 minutes) for time intervals of ten minutes. (b,c) The corresponding smFRET transfer efficiency histograms are shown for the first 30 minutes (b) and for the last 30 minutes (c). In both histograms the unfolded state population (at  $\langle E_1 \rangle \sim 0.2$ ) and the folded state population (at  $\langle E_2 \rangle \sim 0.75$ ) have within the limits of error the same statistical weights  $P_1 \sim 0.45$  and  $P_2 \sim 0.55$  (area under the respective fitted Gaussian peaks).

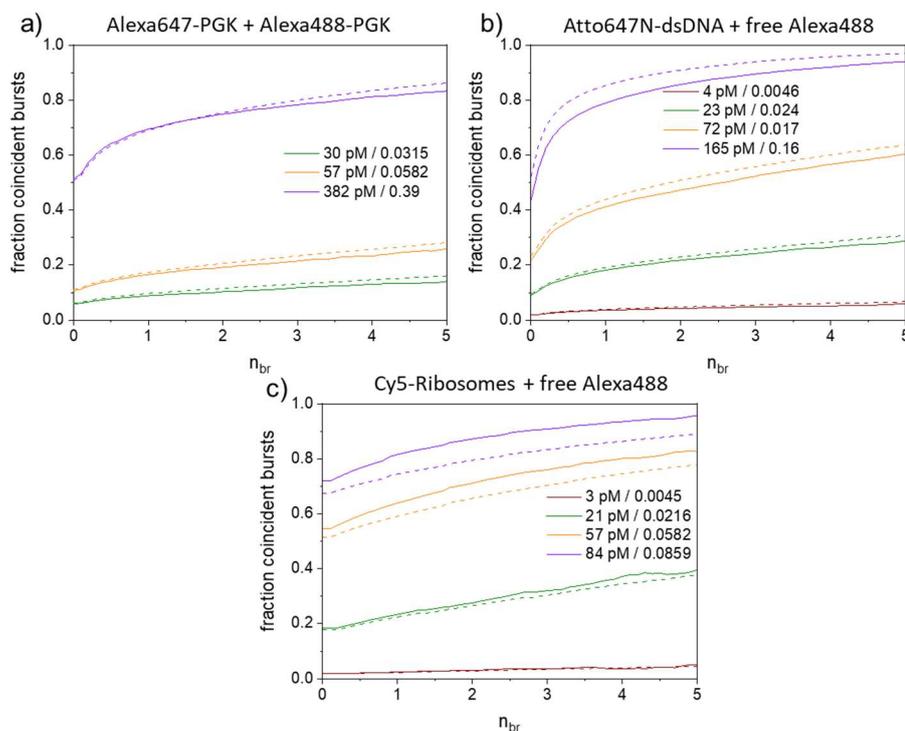
In this respect we observed for proteins in buffers with GndHCl concentrations close to  $C_{1/2}$  drastically reduced count rates during confocal smFRET measurements. A more detailed inspection of the data measured at  $C_{1/2}$ , revealed that the number of detected bursts and thereby the molecule concentration is reduced, mainly caused by unspecific protein binding to the cover-slide surface (see Fig. 3a). Interestingly, we do not observe a similar count rate reduction in other GndHCl concentration regimes (for example, 0-0.5 M for mainly folded states and > 0.9 M mainly unfolded states). Since at  $C_{1/2}$  we are dealing with samples which include two different subpopulations (folded and unfolded states) the question rises, whether one of the subpopulations is more prone for surface sticking and thereby would bias the statistical weights of the obtained subpopulations. In order to answer this question, we separated the data set which accumulated bursts over 60 minutes into two parts. The first part includes bursts of the first 30 minutes (Fig. 3b) and a second part that of the second 30 minutes (Fig. 3c). This rough estimate (a more finely subdivision of the data was not possible due to the limited number of bursts per time interval) exhibits more or less identical smFRET histograms. Importantly, the statistical weights ( $P_1$  and  $P_2$ ) for the individual subpopulations do not vary during the total measuring time. This indicates that both protein conformations have more or less the same tendency for surface sticking. Therefore, at least in the case presented here, unspecific surface sticking of the investigated biomolecules does not bias the obtained results.

### 3.2. Impact of chance coincidence on BTCCD and smFRET results

In studies with diffusing molecules targeting single molecule sensitivity, the molecule concentration must be low enough to ensure that we detect only single molecules. On the other hand, we would like to work with molecule concentrations as high as possible, in order to obtain a high number of bursts within a certain time interval. For achieving a reasonable trade-off between both requirements, it is worth to establish an approach to identify optimal target concentrations, depending on the specific boundary conditions. In confocal microscopy, in principle several freely diffusing molecules can be present in the detection volume at the same time, also known as chance coincidence. Here, we make use of mathematical terms for calculating fractions of chance coincidence, derived from some basic principles (see *subsection 2.11*).

In order to quantify chance coincidence in a straight forward manner, we performed BTCCD analyses with samples containing a pair of two distinct molecules which do not exhibit an appreciable binding affinity to each other. The two involved molecules were labelled with different dyes and the measured coincidence fraction can be interpreted as a pure chance coincidence. In a first example, two PGK species were labelled either with Alexa488 or with Alexa647, mixed in a certain stoichiometry and adjusted to different sample concentrations. In Fig. 4a the measured coincidence fractions  $f_{RB}$  (solid lines) and the corresponding calculated values (dashed lines, see eq. 7) are shown for three different molecule concentrations (here the concentration is given for the Alexa488 labelled species). It is obvious from this figure, that only for a concentration of 30 pM the chance coincidence fraction is below 0.1, while for higher concentrations  $f_{RB}$  values are larger than 0.2 (for 57 pM) or even larger than 0.8 (for 382 pM). The obtained results indicate, that only the 30 pM sample to some extent allows a reliable single molecule data interpretation (with approximately 90% of all bursts originating from single molecules). In the other cases (57 and 382 pM), too many of detected bursts contain photons from more than one molecule (20-80 %). Furthermore, the presented analysis demonstrates a rather good agreement between measured and calculated  $f_{RB}$  values. According to equation 7 (see *subsection 2.11*) the chance coincidence given by  $f_{RB}$  values depends not only on the molecule concentration (given by  $N$ ), but also on the ratio of the involved burst duration times  $\tau_d^R$  and  $\tau_d^B$  of both diffusing species. Therefore, the relative sizes of the diffusing molecules have a clear impact on the obtained chance coincidence. This leads to a more pronounced chance coincidence (i.e. larger  $f_{RB}$  values) if the red labelled molecules diffuse much slower than the

blue labelled ones. A confirmation of such a behavior is illustrated in Fig. 4b, showing data obtained from a molecule pair consisting of red labelled dsDNA and free Alexa488 (blue). In this case, even a molecule concentration of 23 pM gives more than 20% of all bursts which contain photons from more than one molecule (i.e. a single molecule interpretation of individual burst is no longer valid). Such a behavior is even more pronounced for a sample with a very large red labelled ribosome and free Alexa488 (see Fig. 4c).

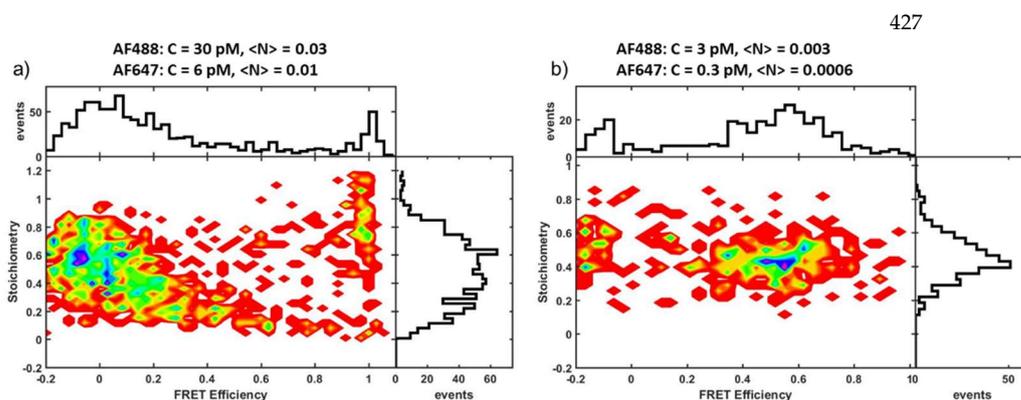


**Figure 4.** Measured (solid lines) and calculated (dashed lines)  $f_{RB}$  values obtained from samples with a pair of two distinct individually and freely diffusing molecules, one red (Alexa647/Atto647N/Cy5) labelled and the other blue (Alexa488) labelled. The respective molecule concentrations are given for the blue labelled species (including the parameter  $N$ , the corresponding average number of molecules in the detection volume at the same time). The concentration of the red labelled species (not given explicitly) is in the same regime as the blue ones. The presented examples exhibit a clear variation in the relative sizes, and thereby in the corresponding burst duration times  $\tau_d$ , of the respective molecules in the mixed samples: (a) similar size of red and blue labelled molecules, (b) larger red labelled molecule versus smaller blue label (i.e. only Alexa488) and finally (c) very large red labelled molecule versus small blue label.

The last two cases exemplarily show, that only samples with a molecule concentration of a few pM allow for a burst analysis with single molecule interpretation. It is noteworthy, that for these conditions the required  $N$  value is below 0.01 (for smFRET studies often the recommend  $N$  value is 0.01- 0.03). Furthermore, we observe a slightly lower agreement level between calculated and measured  $f_{RB}$  values for the samples shown in Fig. 4b and 4c, in particular for the higher chance coincidences probability.

Finally, we will analyze smFRET data from a special case study where the impact of chance coincidence is clearly visible. In a challenging study to investigate protein conformations of ribosome bound nascent chains (RNCs), smFRET data have been measured for double labelled calmodulin (CaM) which remained bound to the ribosome during the measurements. The respective samples were produced by using cell free protein synthesis (CFPS). For a specific dye attachment of the donor and the acceptor only to the CaM protein, we made use of unnatural amino acids, in order to establish a dye binding chemistry which is orthogonal to typically used cysteine based dye binding (see subsection 2.4). Due

to the rather small protein yields of CFPS we obtained only a small fraction of double labelled molecules (i.e. donor and acceptor). Furthermore, the sample purification methods did not work as efficient as compared to standard applications. The smFRET results from a burst analysis (including PIE, for details see *subsections 2.5 and 2.9*) are shown for a first sample in Fig. 5a. The obtained data was measured with a sample containing double labeled molecules, incomplete labeled molecules, and free dye. For the donor color a concentration of 30 pM was measured, while for the acceptor color a concentration of 6 pM. It is obvious that the corresponding transfer efficiency histogram (upper panel) does not show any reasonable FRET population which would correspond to one of the known CaM conformations (with dyes at positions 34 and 110, see for example [16, 26].



**Figure 5.** Here stoichiometry versus FRET efficiency plots are shown for RNC samples as measured with different molecule or dye concentrations. (a) At higher concentrations, the corresponding FRET efficiency histogram (upper panel) shows only peaks at  $\langle E \rangle \sim 1$  and  $\langle E \rangle \sim 0$  which can be considered as artifacts. (b) At lower concentrations, the corresponding efficiency histogram exhibits a meaningful FRET population at  $\langle E \rangle \sim 0.55$  which represents the apo-CaM state.

For a second sample the labelled molecules were much more diluted (donor color with 3 pM and acceptor color with 0.3 pM) and the corresponding smFRET histogram differs significantly from the first one (see Fig. 5b). Although the number of detected bursts is still low, we can now clearly identify a population centered at  $\langle E \rangle \sim 0.55$  which represents a calcium-free state of calmodulin (apoCaM)[16,26]. This is in full agreement with the buffer conditions used in these measurements (see *subsection 2.4*). This example demonstrates that at higher molecule concentration the impact of chance coincidence can cause artificial FRET populations in the efficiency histogram. In particular, the peak at  $\langle E \rangle \sim 0$  (which is often denominated “donor only” peak, but should not appear here due to the application of PIE sorting) is caused by the excess of free unbound donor dyes (which could not be removed efficiently during the sample purification). As demonstrated in Fig. 4c this can cause a significant amount of chance coincidence with artificially too high donor emission. Bursts exhibiting the correct transfer efficiency (i.e.  $\langle E \rangle \sim 0.55$ ) of the rather few “really double labelled” molecules are hidden at these high molecule concentrations and only become visible at much lower molecule concentrations, where chance coincidence vanishes.

#### 4. Conclusions

Single-molecule studies can give valuable and unique information about the ensemble of molecules in a sample. By the identification and the separation of subpopulations, detailed knowledge about properties of individual molecules is accessible. Furthermore, the statistical weights of different subpopulations can be determined quite precisely. However, unrealized sample conditions can cause artifacts which may bias the obtained results significantly.

462 Here, we have demonstrated in a BTCCD analysis that it is worth to determine the  
463 absolute molecule concentrations for all different species present in the sample and to  
464 monitor these values during extend measuring times. With this information at hand, un-  
465 balanced nonspecific molecule attachment to cover-slide surfaces can be identified and  
466 often prevented by slightly optimized environmental sample conditions. Furthermore,  
467 this information gives direct access to the probability of chance coincidence appearance.  
468 As shown in a smFRET study, the obtained results can suffer from chance coincidence and  
469 only a reduction of the molecule concentration can help to overcome this problem for a  
470 given sample. Following the presented approach is in particular helpful for cutting-edge  
471 applications where the sample quality is well below that of standard applications (see for  
472 example [27,28]). Compared to smFRET studies, in BTCCD analyses the situation is more  
473 favorable. We introduced an approach to calculate chance coincidence probabilities ab  
474 initio that exhibit a reasonable agreement with experimentally determined values. There-  
475 fore, we can make use of these calculations in typical BTCCD applications, for example  
476 the determination of binding affinities based of measured coincidence fractions [7,29].  
477 Here the advantage is that we can tolerate chance coincidence in these kind of measure-  
478 ments as long as we can quantify chance coincidence fractions accurately, because finally  
479 we can subtract them from the measured coincidence fractions which determine the bind-  
480 ing affinity in the given sample.

481 **Supplementary Materials:** Not applicable.

482 **Author Contributions:** Conceptualization, J.F., O.Y. and H.H.; methodology, A.K. and H.H.; soft-  
483 ware, H.H. and O.Y.; measurements, O.Y. and N.B.; writing—original draft preparation, J.F., N.B.  
484 and O.Y.; writing—review and editing, all authors.; project administration, J.F. and A.K.; funding  
485 acquisition, J.F. and A.K. All authors have read and agreed to the published version of the manu-  
486 script.

487 **Funding:** N.B. A.K. and J.F. acknowledge funding by the DFG Grants FI 841/10-1 and KA 4388/2-1.  
488 This work was supported the International Helmholtz Research School of Biophysics and Soft Mat-  
489 ter “BioSoft” (to O.Y.).

490 **Institutional Review Board Statement:** Not applicable.

491 **Informed Consent Statement:** Not applicable.

492 **Data Availability Statement:** Data which is contained within the article is available on request from  
493 the corresponding authors.

494 **Acknowledgments:** The authors thank Prof. K. Alexandrov and Dr. Z. Cui (QUT, Brisbane, Aus-  
495 tralia) for substantial support by producing RNC samples with CFPS systems.

496 **Conflicts of Interest:** The authors declare no conflict of interest.

## 497 References

- 498 1. Eggeling, C.; Berger, S.; Brand, L.; Fries, J. R.; Schaffer, J.; Volkmer, A.; Seidel, C. A. M., Data registration and selective single-  
499 molecule analysis using multi-parameter fluorescence detection. *Journal of Biotechnology* **2001**, *86* (3), 163-180.
- 500 2. Sisamakris, E.; Valeri, A.; Kalinin, S.; Rothwell, P. J.; Seidel, C. A., Accurate single-molecule FRET studies using multiparameter  
501 fluorescence detection. *Methods Enzymol.* **2010**, *475*, 455-514.
- 502 3. Deniz, A. A.; Laurence, T. A.; Beligere, G. S.; Dahan, M.; Martin, A. B.; Chemla, D. S.; Dawson, P. E.; Schultz, P. G.; Weiss, S.,  
503 Single-molecule protein folding: diffusion fluorescence resonance energy transfer studies of the denaturation of chymotrypsin  
504 inhibitor 2. *Proc.Natl.Acad.Sci.U.S.A* **2000**, *97* (10), 5179-5184.
- 505 4. Fries, J. R.; Brand, L.; Eggeling, C.; Kollner, M.; Seidel, C. A. M., Quantitative identification of different single molecules by  
506 selective time-resolved confocal fluorescence spectroscopy. *J Phys Chem A* **1998**, *102* (33), 6601-6613.
- 507 5. Zhang, K.; Yang, H., Photon-by-photon determination of emission bursts from diffusing single chromophores. *J Phys Chem B*  
508 **2005**, *109* (46), 21930-21937.
- 509 6. Deniz, A. A.; Laurence, T. A.; Dahan, M.; Chemla, D. S.; Schultz, P. G.; Weiss, S., Ratiometric single-molecule studies of freely  
510 diffusing biomolecules. *Annu Rev Phys Chem* **2001**, *52*, 233-253.
- 511 7. Höfig, H.; Yukhnovets, O.; Remes, C.; Kempf, N.; Katranidis, A.; Kempe, D.; Fitter, J., Brightness-gated two-color coincidence  
512 detection unravels two distinct mechanisms in bacterial protein translation initiation. *Commun Biol* **2019**, *2*, 459.

- 513 8. Dahan, M.; Deniz, A. A.; Ha, T. J.; Chemla, D. S.; Schultz, P. G.; Weiss, S., Ratiometric measurement and identification of single  
514 diffusing molecules. *Chemical Physics* **1999**, *247* (1), 85-106.
- 515 9. Li, H. T.; Ying, L. M.; Green, J. J.; Balasubramanian, S.; Klenerman, D., Ultrasensitive coincidence fluorescence detection of single  
516 DNA molecules. *Anal Chem* **2003**, *75* (7), 1664-1670.
- 517 10. Höfig, H.; Gabba, M.; Poblete, S.; Kempe, D.; Fitter, J., Inter-Dye Distance Distributions Studied by a Combination of Single-  
518 Molecule FRET-Filtered Lifetime Measurements and a Weighted Accessible Volume (wAV) Algorithm. *Molecules* **2014**, *19* (12),  
519 19269-19291.
- 520 11. Rosenkranz, T.; Schlesinger, R.; Gabba, M.; Fitter, J., Native and unfolded states of phosphoglycerate kinase studied by single-  
521 molecule FRET. *Chemphyschem : a European journal of chemical physics and physical chemistry* **2011**, *12* (3), 704-10.
- 522 12. Cerminara, M.; Schone, A.; Ritter, I.; Gabba, M.; Fitter, J., Mapping Multiple Distances in a Multidomain Protein for the Identifi-  
523 cation of Folding Intermediates. *Biophys J* **2020**, *118* (3), 688-697.
- 524 13. Lamprou, P.; Kempe, D.; Katranidis, A.; Buldt, G.; Fitter, J., Nanosecond dynamics of calmodulin and ribosome-bound nascent  
525 chains studied by time-resolved fluorescence anisotropy. *Chembiochem* **2014**, *15* (7), 977-85.
- 526 14. Kempf, N.; Remes, C.; Ledesch, R.; Zuchner, T.; Hofig, H.; Ritter, I.; Katranidis, A.; Fitter, J., A Novel Method to Evaluate Ribo-  
527 somal Performance in Cell-Free Protein Synthesis Systems. *Sci Rep-Uk* **2017**, *7* (7), 467535.
- 528 15. Sadoine, M.; Cerminara, M.; Kempf, N.; Gerrits, M.; Fitter, J.; Katranidis, A., Selective Double-Labeling of Cell-Free Synthesized  
529 Proteins for More Accurate smFRET Studies. *Anal Chem* **2017**, *89* (21), 11278-11285.
- 530 16. Sadoine, M.; Cerminara, M.; Gerrits, M.; Fitter, J.; Katranidis, A., Cotranslational Incorporation into Proteins of a Fluorophore  
531 Suitable for smFRET Studies. *Acs Synth Biol* **2018**, *7* (2), 405-411.
- 532 17. Muller, B. K.; Zaychikov, E.; Brauchle, C.; Lamb, D. C., Pulsed interleaved excitation. *Biophys.J* **2005**, *89* (5), 3508-3522.
- 533 18. Watkins, L. P.; Yang, H., Detection of intensity change points in time-resolved single-molecule measurements. *J Phys Chem B*  
534 **2005**, *109* (1), 617-628.
- 535 19. Hohlbein, J.; Craggs, T. D.; Cordes, T., Alternating-laser excitation: single-molecule FRET and beyond (vol 43, pg 1156, 2014).  
536 *Chem Soc Rev* **2014**, *43* (17), 6472-6472.
- 537 20. Gopich, I. V., Concentration effects in "Single-Molecule" spectroscopy. *J Phys Chem B* **2008**, *112* (19), 6214-6220.
- 538 21. Enderlein, J.; Robbins, D. L.; Ambrose, W. P.; Keller, R. A., Molecular shot noise, burst size distribution, and single-molecule  
539 detection in fluid flow: Effects of multiple occupancy. *J Phys Chem A* **1998**, *102* (30), 6089-6094.
- 540 22. Ratner, V.; Kahana, E.; Eichler, M.; Haas, E., A general strategy for site-specific double labeling of globular proteins for kinetic  
541 FRET studies. *Bioconjugate Chem* **2002**, *13* (5), 1163-1170.
- 542 23. Lerner, E.; Hilzenrat, G.; Amir, D.; Tauber, E.; Garini, Y.; Haas, E., Preparation of homogeneous samples of double-labelled  
543 protein suitable for single-molecule FRET measurements. *Anal Bioanal Chem* **2013**, *405* (18), 5983-5991.
- 544 24. Katranidis, A.; Fitter, J., Single-Molecule Techniques and Cell-Free Protein Synthesis: A Perfect Marriage. *Anal Chem* **2019**, *91*  
545 (4), 2570-2576.
- 546 25. Schuler, B.; Eaton, W. A., Protein folding studied by single-molecule FRET. *Curr.Opin.Struct.Biol* **2008**, *18* (1), 16-26.
- 547 26. Johnson, C. K., Calmodulin, conformational states, and calcium signaling. A single-molecule perspective. *Biochemistry-Uk* **2006**,  
548 *45* (48), 14233-14246.
- 549 27. Hamadani, K. M.; Howe, J.; Jensen, M. K.; Wu, P.; Cate, J. H. D.; Marqusee, S., An in vitro tag-and-modify protein sample  
550 generation method for single-molecule fluorescence resonance energy transfer. *J Biol Chem* **2017**, *292* (38), 15636-15648.
- 551 28. Cui, Z. L.; Mureev, S.; Polinkovsky, M. E.; Tnimov, Z.; Guo, Z.; Durek, T.; Jones, A.; Alexandrov, K., Combining Sense and  
552 Nonsense Codon Reassignment for Site Selective Protein Modification with Unnatural Amino Acids. *Acs Synth Biol* **2017**, *6* (3),  
553 535-544.
- 554 29. Orte, A.; Clarke, R.; Balasubramanian, S.; Klenerman, D., Determination of the fraction and stoichiometry of femtomolar levels  
555 of biomolecular complexes in an excess of monomer using single-molecule, two-color coincidence detection. *Anal Chem* **2006**,  
556 *78* (22), 7707-7715.