Quantifying nucleosome dynamics and protein binding with PIE-FCCS and spFRET

Martens, C.L.G.

Chapter 2

Methods and advances in single-molecule fluorescence microscopy and spectroscopy

2.1 Introduction

Experimental investigations in the life sciences have traditionally been performed on a population level[105][106]. Analysis of the ensemble average of molecular properties results in loss of information concerning molecular heterogeneity, and may ultimately lead to misinterpretations of the underlying physiological relevance of subpopulations of molecules. Focusing on molecules as the minimal 'functional' units in a biological system, single-molecule biophysics research has an important impact on a range of fields of biological investigation[107].
The true workhorse of the single-molecule methods is fluorescence microscopy. It is a widely-used, and low-invasive method, allowing the biomolecule to remain in an *in vivo* environment. Maintaining *in vivo* conditions saves most, if not all functionality of the biological system. Moreover, fluorescence microscopy methods allow access to useful measurable parameters on time and length scales relevant for the biomolecular processes[106][108].
Wide-field microscopy is how single-molecule microscopy is most easily achieved nowadays. However, confocal imaging has the advantage of less background signal. The principle of confocal imaging has been developed in the 1950's and overcomes several limitations of traditional wide-field fluorescence microscopy. In a conventional microscope, the entire specimen is illuminated evenly. As all parts of the sample in the optical path are excited at the same time, the resulting fluorescence includes a large unfocused background part. In contrast, a confocal microscope uses point illumination and a pinhole in the optically conjugate plane in front of the detector to eliminate out-of-focus signal (fig. 2.2a). The resolution, particularly in the sample depth direction, is much better than that of wide-field microscopes. Scanning the focused beam produces images with resolution close to the Abbe limit in scanning confocal microscopy (SCM)[109][110][111].
To quantify conformational changes of single molecules or interaction between molecules of length scales below the diffraction limit one can revert to super-resolution techniques in which the center of mass of individual molecules is determined[112][113]. However, the limited number of photons, together with background signal, typically limit the accuracy to several 10's of nanometers. For even better accuracy, one can revert to the non-radiative energy transfer between two fluorophores known as FRET

(Forster or Fluorescence Resonance Energy Transfer). FRET can resolve distances of 2-10 nanometers, making it very suited for resolving conformations of and interactions between biomolecules[114][115][116][117].
Though scanning is time consuming, the temporal resolution of the detection of photons can be sub-nanoseconds. Fluorescence correlation spectroscopy (FCS) highlights the temporal rather than the spatial resolution and computes the correlation times of the fluorescence, revealing the mobility of single molecules as well as time constants that characterize the photophysics of the fluorescent molecules[118][119][120][121][122].
When FRET is resolved at the single-molecule level, referred to as single-pair FRET (spFRET), an additional advantage over ensemble measurements comes up: it is possible to track the dynamics of conformational changes, or molecular interactions, even when they proceed stochastically, which most biomolecular processes do. Here we will focus on using spFRET to reveal conformational changes in nucleosomes. We will also address fluorophore choice and setup calibration, data handling and post-fit corrections necessary to obtain reliable data.

2.2 Methods and materials

2.2.1 FRET

Förster's Resonance Energy Transfer (FRET), named after Theodor Förster for first describing the phenomenon, is the energy transfer between two molecules by resonance. The mechanism of FRET can be described as the quantum analog of classically coupled mechanical oscillators. When the electronic states of two molecules are in phase, the energy of the first molecule transfers in a non-radiative way to the second molecule, promoting it from ground state to excited state (figure 2.1). When both molecules are capable of fluorescence, FRET is dubbed Fluorescence Resonance Energy Transfer. Note that the energy transfer is not through fluorescence.
The strength of this dipole-dipole interaction depends on the inter-molecular distance:

$$E = \frac{1}{1 + (\frac{R}{R_0})^6} \tag{2.1}$$

where R is the distance between donor and acceptor molecules and R_0 their Förster radius. The FRET efficiency E is derived from the intensities of the

donor and acceptor fluorescence, as we will show in later sections.

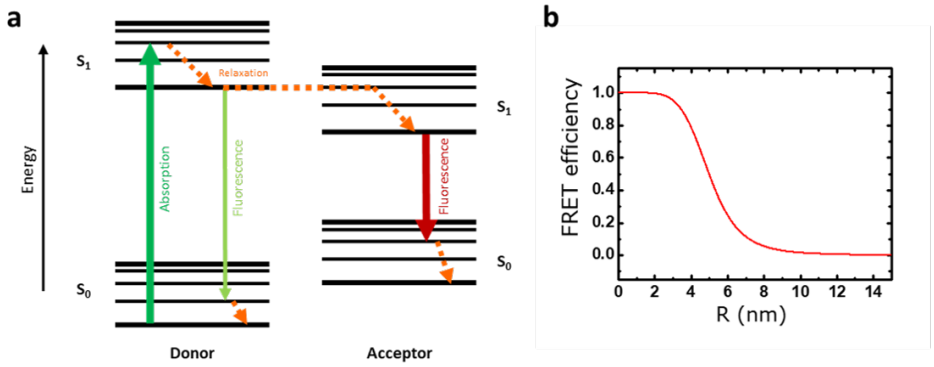

FIGURE 2.1: **Mechanism of FRET at the nanometer scale. a)** A Jablonski diagram of FRET, depicting the energy transfer, which only occurs when electronic singlet states S_0 and S_1 are in close proximity. **b)** FRET efficiency as a function of distance R. FRET efficiency E is 50% at characteristic distance R_0.

2.2.2 Choice of fluorophores

The right choice of fluorophores for single-molecule experiments is paramount; in order to follow a molecule, whether its diffusion or conformational dynamics, one needs a good signal of the fluorophore tagged to the molecule. Fluorophores need to be stable emitters (no bleaching), and their emission should be stable over time, i.e. minimal transitions into dark triplet states ($S_1 \rightarrow T_1$) should occur. Moreover, the fluorophore should have a high emission efficiency (quantum yield). For FRET, to follow dynamics, there should be sufficient spectral overlap between the emission of the donor fluorophore and the absorption of the acceptor fluorophore.
Two fluorophores qualifying this description are Cy3B and ATTO647N (fig. 2.4). Cy3B is a cyanine dye (synthetic polymethines) with a higher quantum yield (0.67) than its siblings Cy3 (0.15), Cy5 (0.27) and Cy7 (0.28)[123]. The high quantum yield is because Cy3B is not capable of cis-trans isomerization around the polymethine group, which can lead to loss of fluorescence after excitation. As a result, Cy3B is not subject to photo-isomerization and is both extremely bright and stable[124].
ATTO647N is derived from carbopyronin and also has a rigid structure, making cis-trans isomerization impossible. It is one of the brightest and

most photostable dyes available. Contrary to Cy5, ATTO647N remains stable in salt buffers and has a high ozone resistance. Cy3B and ATTO647N are an optimal FRET-pair; they have a Förster radius of 6.2 nm, one of the longest radii available at the moment[125][126].
In this thesis nucleosomes were labelled with Cy3B and Atto647N for following breathing dynamics. When quantifying protein binding affinity, DNA and nucleosomes were only labelled with ATTO647N, and proteins were labelled with fluorophores in excitation and emission spectra similar to Cy3B. Although these fluorophores, such as EYFP and ATTO532, are spectrally similar to Cy3B, they have distinct structural and functional features. EYFP is the Enhanced Yellow Fluorescent Protein and is a derivative of GFP, a naturally occurring fluorescent protein. As such, it can be fused into the vector DNA, transcribed, translated and genetically tagged to a protein, resulting in a population of proteins that is completely labelled. The downside of using Fluorescent Proteins for single-molecule studies, especially the YFP's, is their limited optical stability, a phenomenon known as flickering[127]. Such prolonged residence in a dark triplet state can last micro- to milliseconds. In organic fluorophores this relaxation process is much faster (sub-microsecond) and is called blinking. As diffusion times and lifetimes of different conformations are typically in the order of milliseconds, flickering can obscure events. EYFP does have a high quantum yield: 0.60[128].
ATTO532 is a carboxy-derived fluorophore like ATTO647N, and has excellent photostability. It is a strong absorber, has a very high quantum yield (0.90), remains stable in a wide thermal range and has excellent water solubility[129]. These characteristics make ATTO532 a good choice for protein labelling. ATTO532 and ATTO647N form a FRET pair with a Förster radius of 5.1 nm[130].
An other important feature to take into account when doing fluorescence microscopy is the size of the fluorophore. Many fluorescence-based methods to follow interactions rely either explicitly (FCS, stopped-flow cytometry) or implicitly (SCM, TIRFM) on (differences in) diffusion of proteins, DNA or nucleosomes. Organic dyes compare favourable in this respect. ATTO dyes are $\sim$2 nm long and have molecular weights of 0.8 - 1 kDa. Cy dyes are even smaller; Cy3B has a length of 1.4 nm[131] and weighs 0.6 kDa. Fluorescent proteins are large compared to ATTO and Cy dyes. The chromophore of the protein is only 1 nanometer long, but the rest of the protein is more than 4 nm long and has a diameter of 3 nm (fig. 2.2b). Its molecular

weight is 27 kDa[132]. This results in a larger hydrodynamic radius, when a biomolecule is tagged with the fluorophore, leading to larger diffusion constants. Moreover, the added structure may sterically hinder interactions

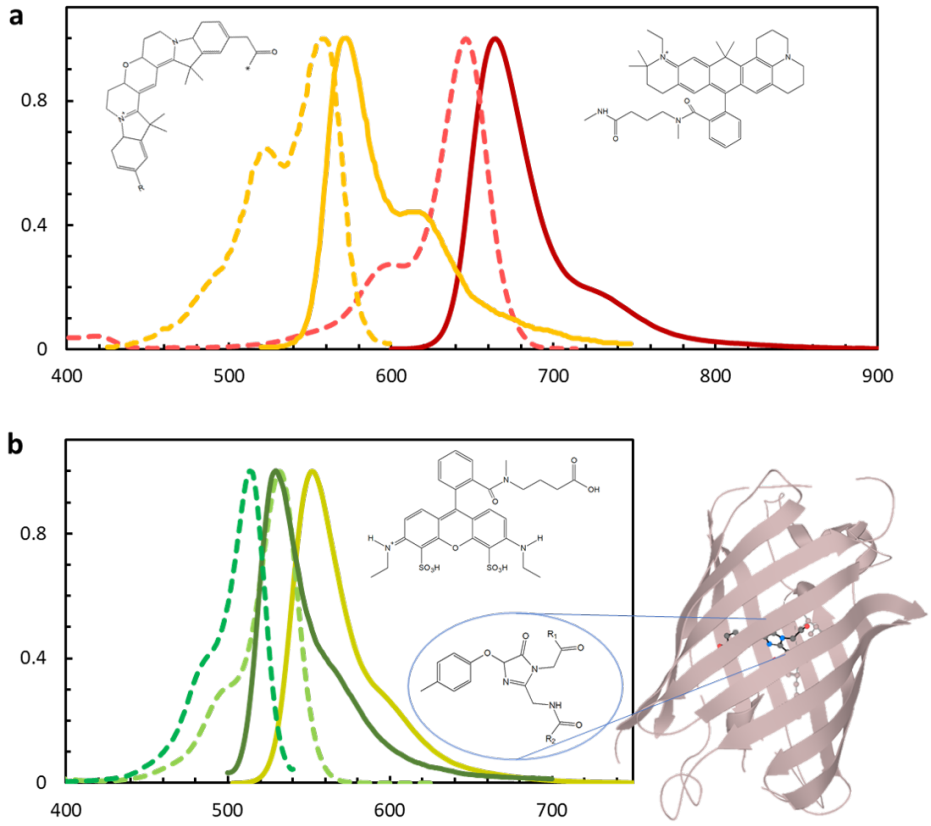

FIGURE 2.2: **Excitation and emission spectra of fluorophores and FRET pairs used in this thesis. a)** Excitation and emission spectra, as well as molecular structures of FRET pair Cy3B (left molecule, orange curves) and ATTO647N (right molecule, red curves). **b)** Molecular structures and excitation and emission spectra of ATTO532 (top molecule, darker green curves) and EYFP (bottom and right molecule, moss green curves). It should be noted that, even though EYFP emission and ATTO647N excitation spectra partial overlap, due to the large size of EYFP, FRET was not observed.

between or conformational changes within molecules. Finally, the chemical

properties of the fluorophore may induce aggregation of the biomolecules. Though we will not use the FRET efficiency to deduce absolute distances in this thesis, one should be aware that locking the orientations of the dyes may affect the FRET efficiency, and may lead to its own dynamics. However, we are not aware that this is significant in the results we present here.

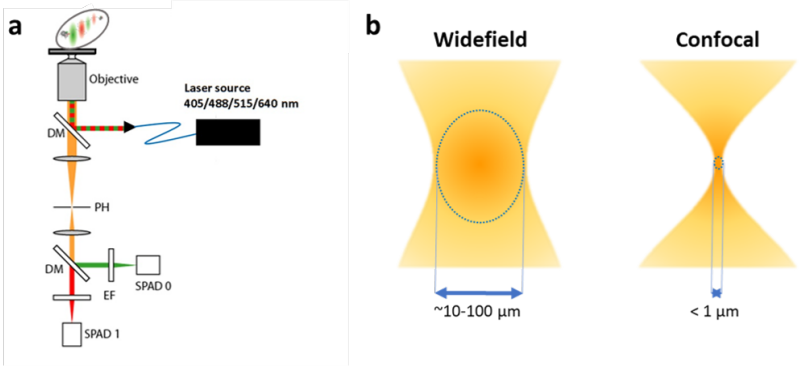

FIGURE 2.3: **Confocal microscopy reduces illuminated volume by orders of magnitude compared to widefield microscopy. a)** graphic depiction of the confocal microscope setup; the excitation beam from the laser is collimated and reflected by dichroic mirror 1 (DM1) into the objective. Fluorescence passes through DM1 and is focused onto pinhole PH. Next, the signal is collimated and wavelengths shorter than 640 nm are reflected by DM2 (640dcxr), filtered and focused onto SPAD 0 (Green photodiode). Wavelengths longer than 640 nm are directed to SPAD 1 (Red photodiode).**b)** Visualization of the difference in light collimation between widefield and confocal microscopy.

2.2.3 Confocal microscope setup

Measurements were performed on a home-built confocal microscope as depicted in figure 2.7 equipped with a 60x water-immersion objective (NA 1.2, Olympus). An ICHROME MLE-SFG laser was used as an 514 nm and 632 nm excitation source. From an optical fiber a collimated excitation beam was directed through the objective and focused 25 μm above the glass-sample interface. The excitation power was in the order of 5 μW. The collected fluorescence was focused with a 15mm tube lens (ThorLabs) and spatially filtered with a 50 μm pinhole. After collimating the beam with a 15mm lens (ThorLabs) it was split by a dichroic mirror (640dcxr, Chroma). The two emission beams were further filtered (hq570/100nm

and hq700/75m, respectively) and imagined on the active area of Single Photon Avalanche Photodiodes (SPADs, SPCM AQR-14, Perkin Elmer) using a 10mm lens (ThorLabs). The photodiodes were read out with a TimeHarp 200 photon counting board (PicoQuant). The acquired data was stored in time-tagged to time-resolved data files (*.t3r), which were further processed with a home-built Python program to read out the photon data according to arrival time and SPAD channel.

2.3 Single-molecule burst microscopy

Though it is not possible to discriminate acceptor fluorescence from direct excitation and FRET, one can intermittently use different different excitation wavelengths to check for the presence of both fluorophores. For Alternating Laser EXcitation (ALEX)-spFRET measurements, biological samples were diluted to picomolar concentrations to ensure that only one molecule is in focus during the bin time of 1 millisecond. Light pulses of 514 and 632 nm, each 25 microseconds long, were used to excite either the

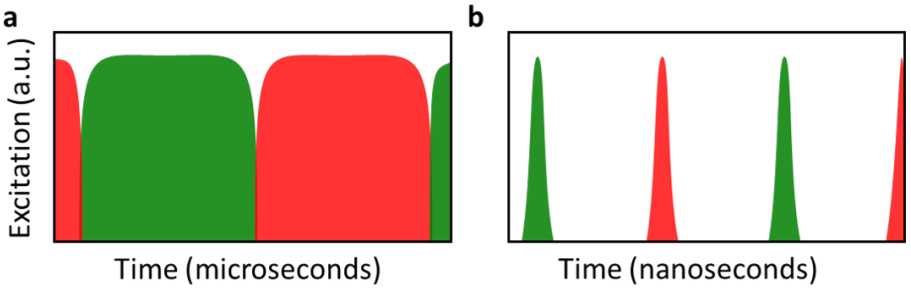

FIGURE 2.4: **Major differences between ALEX and PIE are timescale as well as partial pulse overlap due to equipment. a)** Alternating Laser EXcitation entails consecutive laser pulses at microsecond scale. **b)** Pulsed Interleaved Excitation means sub-microsecond pulses, separated by dark periods. These dark periods also assure pulses do not partially overlap.

acceptor fluorophore (632 nm) or the donor (514 nm). Pulsed Interleaved Excitation (PIE) is a similar technique of alternating several laser pulses of different wavelengths but with dark periods at a sub-microsecond timescale. This allows for almost simultaneous recording of the temporal behaviour. For our FCS analysis, FRET pair Cy3B-ATTO647N was excited with 100

ns pulses of wavelengths 514 and 632 nm, separated by intermittent dark periods of 300ns (figure **2.4b**). These dark periods prevent residual temporal leakage of the laser.

2.3.1 ALEX-spFRET burst analysis

As the diffusion time of the proteins, DNA and nucleosomes used in our experiments were ˜100 times slower than the pulse cycle of 50 microseconds, emitted photons from different excitation wavelengths were detected as quasi-continuous bursts in the two detectors. A single burst was defined as a series of at least 50 photons of any color, with an inter photon-photon time less or equal to 100 microseconds. The number of photons in each channel was used to calculate the stoichiometry of the fluorescent labels (S) and the FRET efficiency (E). E and S were plotted in an E,S-histogram as shown in Figure4.1b.
The approximate stoichiometry and FRET efficiency were defined as

$$E^{raw} = \frac{I^R_{514}}{I^R_{514} + I^G_{514}} \tag{2.2}$$

and

$$S^{raw} = \frac{I^R_{514} + I^G_{514}}{I^R_{514} + I^G_{514} + I^R_{632}} \tag{2.3}$$

were I^G_{514} is the signal on the green APD during 514 nm excitation, I^R_{514} acceptor emission after donor excitation and I^R_{632} the intensity of acceptor emission after acceptor excitation. Stoichiometry S discriminates the populations of molecules labeled with both fluorophores ($0.2 < S < 0.7$), from molecules labeled with either the donor ($S > 0.7$) or the acceptor fluorophore ($S < 0.2$).
The FRET efficiency and stoichiometry were corrected for several effects. The intensity of the background was subtracted from each burst, taking into account duration τ_b of the burst:

$$I^D = I^G_{514} - \tau_b \cdot i^G_{514},$$

$$I^A = I^R_{632} - \tau_b \cdot i^R_{632}$$

Depending on the donor fluorophore, a significant portion of the donor emission can leak into the acceptor channel. This phenomenon is called spectral leakage (α), and was calculated as

$$\alpha = \frac{E_{Donly}}{1 - E_{Donly}}$$

where E_{Donly} was determined by taking the mean FRET efficiency of the bursts where $S > 0.7$ (nucleosomes labeled with only donor fluorophore). Direct excitation of the acceptor fluorophore during donor excitation was calculated as

$$\delta = \frac{S_{Aonly}}{1 - S_{Aonly}}$$

with S_{Aonly} the mean stoichiometries of the population of bursts for $S < 0.2$. Finally, parameters α and δ were used to correct I^A_{514} and determine I^F:

$$I^F = I^A_{514} - \alpha \cdot I^D - \delta \cdot I^A$$

I^F represents the signal where the emission of the acceptor during donor excitation was only due to FRET. After these corrections a more accurate FRET efficiency and stoichiometry can be defined:

$$E^{pr} = \frac{I^F}{I^F + I^D} \tag{2.4}$$

and

$$S^{pr} = \frac{I^F + I^D}{I^F + I^D + I^A} \tag{2.5}$$

The last correction factor, γ, is to account for differences in quantum yield, excitation intensity and detection efficiencies of the donor and acceptor and can be calculated from tabulated or independently measured parameters. It can also be determined experimentally from the relation between E^{pr} and S^*_{pr} post-hoc. When fitting $\frac{1}{S_{pr}} = m \cdot E_{pr} + c$, γ is defined as $\frac{c-1}{c+m-1}$. The fully corrected FRET efficiency and stoichiometry are then computed as

$$E = \frac{I^F}{I^F + \gamma \cdot I^D} \tag{2.6}$$

and

$$S = \frac{I^F + \gamma \cdot I^D}{I^F + \gamma \cdot I^D + I^A} \tag{2.7}$$

We used the FRETBursts toolkit developed by Ingargiola et al.[133] to calculate E and S and to generate E,S-histograms.

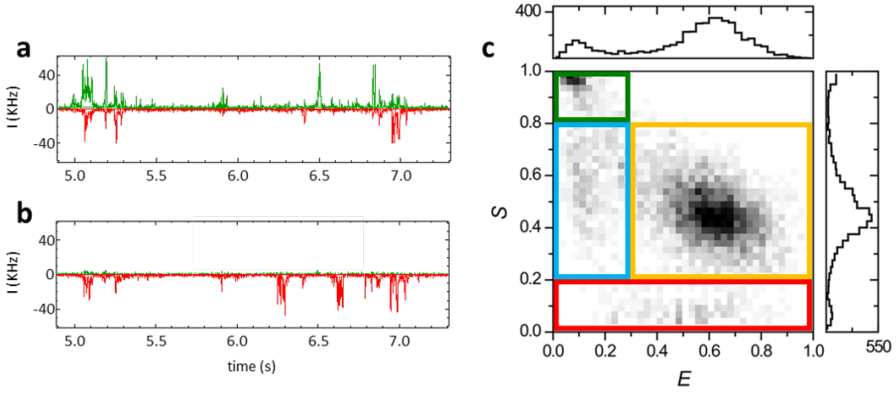

FIGURE 2.5: **E,S histograms are derived from single bursts in emission channels. a)** and **b)** the emission channels after green excitation (a) and red excitation (b). The signals were binned at 1 ms. **c)** The E,S histogram resulting from summing the photons from individual bursts shows the signal consisted of different populations of single- and double-labeled species; in the green rectangle is the population of molecules only labeled with the donor fluorophore, in red only labeled with the acceptor. Blue is the double-labeled population exhibiting no/low FRET, orange the double-labeled population showing high FRET.

2.3.2 Scanning Confocal Microscopy

Next to detecting freely diffusing molecules in a fixed focus, we used Scanning Confocal microscopy (SCM) to follow immobilized molecules over longer times than the diffusion time. The confocal setup as described in the previous section was used to perform scanning measurements combined with ALEX (figure **2.6a**) to generate images as figure **2.6b**. Laser intensities were 5 μW for both 514 and 632 nm excitations. The pixel frequency was 1 kHz (1ms per color) and the line scan rate was 100 Hz, so each pixel was excited 5 times with both colors. Piezo step resolution (P-517 3CD, Physik Instrumente) was 1 nm and travel range 100 μm for x,y direction. In figure **2.6b** the pixel size was 250x250 nm. Nucleosomes were immobilized on a PEG/biotinPEG-coated surface using a modified version of the protocol

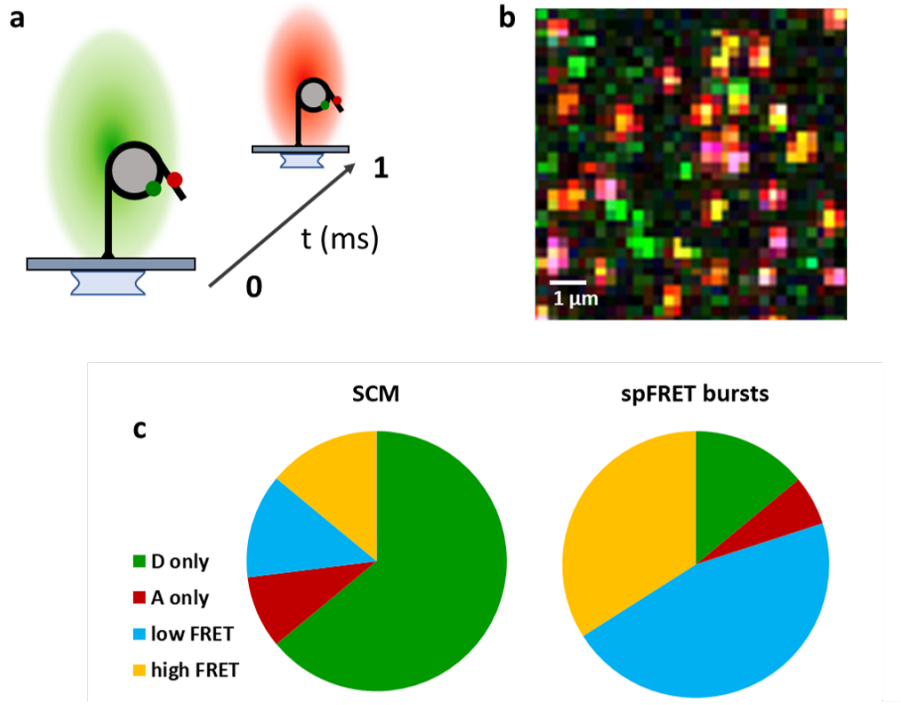

FIGURE 2.6: **Immobilization and ALEX-scanning confocal microscopy visualizes nucleosome dynamics at millisecond timescale but bleaches acceptor fluorophore. a)** Visualization of ALEX-scanning confocal microscopy. **b)** Nucleosomes (N=35) on a 72μm^2 surface show some molecules spectrally overlapping, making it difficult to automatize detection and characterization. **c)** Comparing subpopulations of SCM (N=200) with spFRET (N=8000) shows a significant part of the acceptor fluorophores quenched. The main cause for this is likely the proximity of ATTO647N to the surface.

from Luo et al.[100]. The complete protocol is discussed in Chapter 3 of this thesis.

Examination of label stoichiometry and FRET efficiency of immobilized nucleosomes showed a significant effect of immobilization on the ratio of populations (figure **2.6c**). Although an oxygen scavenger system as well as a photobleaching reductor (Trolox) were added, scanning immobilized nucleosomes showed 64% of the acceptor fluorophores was quenched or bleached, compared to 14% in spFRET burst experiments. The donor fluorophore appeared more resilient, as only 6% (burst) to 9% (SCM)

of nucleosomes was classified as acceptor-only population. Nucleosomes containing both fluorescent labels made up 80% of the total population when freely diffusing compared to only 27% when immobilized. The ratios between low and high FRET were similar; 46% and 34% or 3:2 for spFRET bursts against 13% and 14% or 1:1 for scanning confocal. It is important to note that nucleosomes were selected manually for SCM, as our search algorithm was unable to distinguish between molecules too close to eachother. Because homogeneous molecule dispersion could not be attained by altering nucleosome concentration or other steps of the immobilization protocol we did not further use SCM to investigate nucleosome dynamics and nucleosome-protein interactions. These results do show an important benefit of burst-analysis or FCS compared to SCM of immobilized molecules: the limited duration of excitation relieves bleaching effects and allows for higher excitation and emission rates, even in absence of oxygen scavenger and triplet state quenchers.

2.4 Fluorescence Correlation Spectroscopy

Having labeled molecules diffuse through the focus causes the intensity of the fluorescent signal to fluctuate in time. In Fluorescence Correlation Spectroscopy (FCS) these fluctuations are used to determine the concentration, diffusion constant and when possible dynamical properties of molecules. The fluctuations in intensity are analyzed by correlating photon arrival times over increasing time-lag τ:

$$G(\tau) = \frac{\langle \delta I_1(t) \cdot \delta I_2(t+\tau) \rangle}{\langle I_1(t) \rangle \cdot \langle I_2(t) \rangle} \tag{2.8}$$

To assess the diffusion of a molecule, photon arrival times of one channel are correlated to generate an autocorrelation curve ($I_1 = I_2$). The correlation function that fits the diffusional part of a autocorrelation curve is formulated in terms of the concentration and diffusion time of the population of molecules labeled with the same fluorophore, taking into account the confocal volume:

$$G_{diff}(\tau) = N^{-1} \cdot (1 + \tau/\tau_D)^{-1} \cdot (1 + a^{-2} \cdot \tau/\tau_D)^{-1/2} \tag{2.9}$$

where τ_D is the diffusion time and N the average number of molecules in the confocal volume. Parameter a is the ratio between the axial and radial size of the confocal volume. The value of a for the setup used for the measurements presented here was determined through calibration experiments to be 8. The diffusion time τ_D of a molecule is determined by its size and the viscosity of the solvent η. The parameter used to express the size of a molecule is its hydrodynamic radius r_H and can be obtained using the Stokes-Einstein equation:

$$r_H = \frac{k_B T}{6\pi\eta D} \tag{2.10}$$

where diffusion constant $D = \frac{w^2}{4\tau_D}$ with k_B the Boltzmann constant, T temperature and w the radius of the confocal spot in the radial (x,y) direction. Equation **2.10** shows the hydrodynamic radius scales proportional with diffusion time, implying that larger molecules move slower through the focus. This property was used to analyze correlation curves constructed from signals of molecules of different sizes. If we assume the molecule to have a spherical shape, the radius scales with the molecular mass as $r_H \propto M^{\frac{1}{3}}$. In practice this means for the diffusion time to increase two-fold, the mass of a molecule needs to increase a factor of 8.
To quantify the colocalization of two differently labeled molecules the signal of one molecule (I_1) is correlated with the signal of another molecule (I_2) to generate a crosscorrelation curve. Physical interpretation of the crosscorrelation functions requires additional calculations and will be discussed further on.
The Python module pycorrelate developed by Ingargiola et al.[133] was used to calculate all correlation curves. The correlation algorithm used in this module was developed by Laurence et al.[134]. The algorithm is based on rewriting the correlation as a counting operation on photon pairs and can be used with arbitrary bin widths and spacing (see figure **2.7**).

2.4.1 Photophysics

Photophysics of the fluorophore, i.e. transiting to a triplet or dark state, as well as afterpulsing effects from the APDs need to be included in the fit of a correlation curve

$$G_{total}(\tau) = G_{diff}(\tau) \cdot G_{tr}(\tau) \cdot G_{ap}(\tau) \tag{2.11}$$

where the latter two terms are defined as

$$G_{tr}(\tau) = 1 + \left(\frac{F_{tr}}{1 - F_{tr}} \cdot e^{\frac{-\tau}{\tau_{tr}}} \right)$$

and

$$G_{ap}(\tau) = 1 + \left(\frac{F_{ap}}{1 - F_{ap}} \cdot e^{\frac{-\tau}{\tau_{ap}}} \right)$$

with F_{tr}, F_{ap} the fractions of molecules associated with either triplet state (**tr**) or afterpulsing (**ap**), and τ_{tr}, τ_{ap} their characteristic timescales. As fluorophore photophysics and afterpulsing take place on different timescales[135][136] sensible boundaries were set for fitting these parameters (see Table 2.1 for all parameter boundaries). Figure 2.7a-d shows how a correlation curve is built from the fluorescent signal, culminating in a curve fitted with $G_{total}(\tau)$. Note that PIE leads to additional modulations in the correlation curve at short delays. We therefore refrained from fitting AP in those cases.

2.4.2 Microscope calibration and corrections

Also for FCS we applied post-fit corrections to compensate for the contributions of spectral leakage, background intensity, the difference of confocal volume for different excitation wavelengths and missing part of signal due to pinhole mis-alignment[137][138][139][140][141]. Spectral leakage from 514 nm excitation to the Red APD was corrected as

$$I_{514R} = I_{514R+leak} - (c_{leak} \cdot I_{514G}) \tag{2.12}$$

and eliminates false-positive detection of FRET. c_{leak} is the parameter denoting the percentage of leakage and depends on the fluorophore. For Cy3B $c_{leak} = 0.11$, for Atto532 $c_{leak} = 0.03$ and for EYFP $c_{leak} = 0.15$.
Background photons increase the apparent number of molecules in solution through correlation of random/noise photons (which appear often) with photons from fluorophores (appearing rarely when in low concentration).

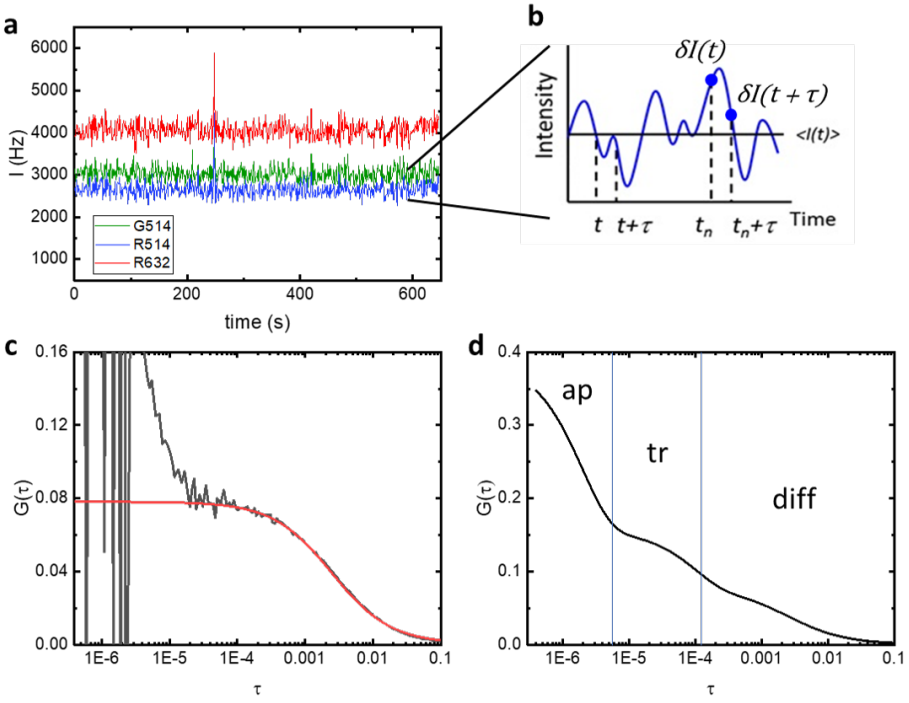

FIGURE 2.7: **Correlation curves are built-up by correlating fluorescent signal at increasing time-lag. a)** and **b)** Fluorescent signal from FCS-PIE measurements is correlated at time t at increasing time-lag τ, characterizing differences in intensity I over increasing timescales. **c)** Taking only the diffusion into account when fitting results in an overestimation of the number of particles, as fluorophores switching in and out of triplet state also contribute to the signal at the time scale of the fit's plateau. **d)** Visualization of the timescales at which diffusion (diff), triplet state photophysics (tr) and afterpulsing (ap) occur. It is shown that overlap of triplet state and diffusion decreases when diffusion time increases.

$$
\begin{aligned}
N_{514G,bgcor} &= N_{514G} \cdot \left(\frac{I_{514G} - bg_{514G}}{I_{514G}} \right)^2 \\
N_{514R,bgcor} &= N_{514R} \cdot \left(\frac{I_{514R} - bg_{514R}}{I_{514R}} \right)^2 \\
N_{632R,bgcor} &= N_{632R} \cdot \left(\frac{I_{632R} - bg_{632R}}{I_{632R}} \right)^2
\end{aligned}
\tag{2.13}
$$

where bg_i is the number of background photons / dark counts on APD G or R during excitation with 514 or 632 nm light of solvent without fluorescent sample. The difference in confocal volume and pinhole alignment were corrected by introducing correction factors based on the number of molecules found in the 632R channel:

$$
\begin{aligned}
N_{514G} &= N_{514G,bgcor} \cdot c_{ax514G} \\
N_{514R} &= N_{514R,bgcor} \cdot c_{ax514R} \\
N_{632R} &= N_{632R,bgcor}
\end{aligned}
\tag{2.14}
$$

Channel 632R was chosen because 632 nm excitation created a larger focal spot (i.e. larger **ax**ial radius) hence more emission than with 514 nm and alignment of the pinhole was more optimal for this channel. Table **2.1** and figure **2.8** show how correction factors were determined and used; figure **2.8a** shows the apparent difference in concentration of TetraSpec beads (Invitrogen, no. T7279, $d \sim 0.1\mu m$) in different channels (after correction for background). Table **2.1** shows the results of using equations **2.13** and **2.14**.

The same diffusion time, $t_D = 12.2 \pm 1.1$ ms, was found for all autocorrelations as well as for all cross-correlations. A DNA construct of 310 base pairs with fluorescent labels Cy3B and Atto647N placed 10 base pairs apart was constructed to provide a stable FRET signal. This construct, dubbed DNAFRET, was used to evaluate the effect of background contributions to the apparent number of molecules N as a function of the actual concentration C (figure **2.8c**). Using equations **2.13** and **2.14** on a dilution series of DNAFRET from 30 nM to 60 pM shows in figure **2.8d** that the contribution of background was properly corrected. Residual differences were attributed to incomplete double-labelling and photobleaching.

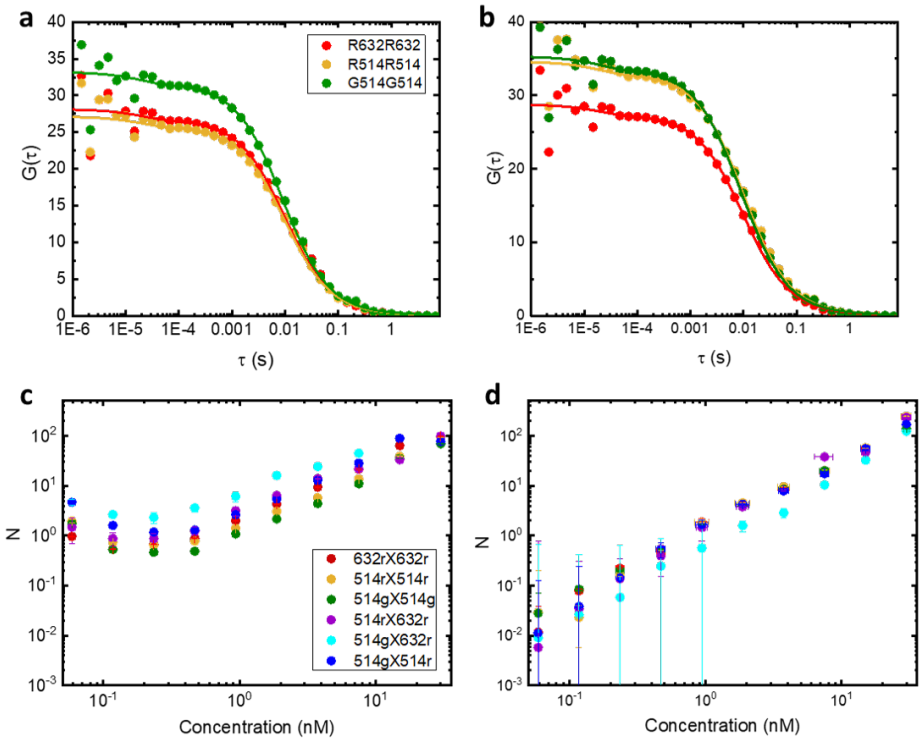

FIGURE 2.8: **TetraSpec beads used to determine degree of spot overlap, titration of DNAFRET effect of background to false increase in number of molecules when concentration decreased. a)** Autocorrelation curves for TetraSpec beads before and **b)** after correction for background. The resulting curves show the imperfect overlap between green (G514 and R514 curve) and red excitation spots. Also a small difference between donor emission from donor excitation (G514) and acceptor emission from donor excitation (R514) can be seen. For TetraSpec beads, R514 signal is not FRET but spectral leakage and shows the degree of overlap of the emission channels, which is nearly perfect (98%). **c)** Titrating DNAFRET from 30 nM to 0.05 nM. Below 1 nM the effect of background photons becomes visible. **d)** After background correction all channels, except for 514G632R, align. The underestimation of this outlier channel was also observed for TetraSpec beads and will be compensated in crosscorrelation fit corrections (next paragraph).

channel	bg (Hz)	I (Hz)	N_{bgcor}	c_{ax}
514G	42	896	0.0298 ± 0.0017	1.24
514R	23	212	0.0304 ± 0.0020	1.17
632R	24	886	0.0366 ± 0.0012	-

TABLE 2.1: **Calibration with TetraSpec beads shows partial focal spot overlap.** Beads were measured in PIE-mode for 1200 seconds.

2.4.3 Fluorescence Crosscorrelation Spectroscopy

Contrary to the value of G(0) of an autocorrelation curve, the amplitude of a crosscorrelation curve scales proportionally with the concentration of molecules in complex. Additionally, the amplitude of the crosscorrelation curve represents the complex molecules as a percentage of the population of molecules present at a higher concentration[142], implying that G_{cc} is reduced when one or more of the concentrations of single labeled molecules are increased. This is illustrated in figure **2.9a** and **2.9b** where the crosscorrelation curve (blue) is only a fraction of the autocorrelation curve of the Atto647N label on DNA (red).
To calculate the real number of complexed molecules from the crosscorrelation curve, N_{CC} need first be corrected for background photons from both channels involved in the cross-correlation[139]:

$$N_{CC,corr} = \frac{N_{CC} \cdot (I_{514G} - bg_{514G}) \cdot (I_{632R} - bg_{632R})}{I_{514G} \cdot I_{632R}} \tag{2.15}$$

We have used channels 514G and 632R as an example as in this thesis I_{514G} corresponds to the signal of proteins (when labeled), I_{532R} with DNA or nucleosomes. In principle any two different channels (f.i. 514R and 632G) can be cross-correlated. We also need to take into account that confocal spots from different excitation wavelengths do not completely overlap. Not compensating for this incomplete overlap would mean underestimating the number of molecules in complex[141]. The actual number of molecules in a complex is then calculated as

$$N_{514Gx632R} = c_{over}^{-1} \cdot \frac{N_{514G} \cdot N_{632R}}{N_{CC,corr}} \tag{2.16}$$

where c_{over} is used to correct for the incomplete overlap. Calibration experiments with TetraSpec beads performed after each alignment of the setup showed $c_{over} = 0.9 \pm 0.03$, implying without correction ~10% of the number of molecules in complex would be missed. With the actual number of molecules in complex determined, the dissociation constant K_d is calculated as

$$K_d = \frac{[molecule1] \cdot [molecule2]}{[complex1+2]} \tag{2.17}$$

The dissociation constant is a measure of the binding affinity and is equal to the concentration of molecule 1 at which half of its available binding sites are occupied by molecule 2. Figures **2.9e** and **2.9f** show how high and low affinity binding look like in FCS.

2.4.4 Conformational dynamics

Often, the highest time constant in the auto/cross-correlation curve corresponds to the average diffusion time of the molecules contributing to the signal. From the diffusion time the hydrodynamic r_H and molecular weight M can be estimated[143] (eq.s **2.10** and **2.11**). By extension it is possible to distinguish between molecules of different sizes combined in an correlation curve. A significant difference in the size or weight of a molecule (2*r_H or 8*M for 2*τ_D) is required to be able to detect such a difference[144]. This means that small changes such ligand binding or conformational dynamics would not be detectable with FCS.
However, FCS can readily be used for the detection of kinetics occurring on scales much smaller than r_H. In fact, one of the first uses of FCS was to determine the interaction rates of ethidium bromide (EtBr) with DNA[145]. By dividing the correlation curve corresponding to the complex or bound DNA by the curve corresponding to both free and bound DNA the Magde group cancelled the diffusional component, generating a correlation curve exclusively built from molecules of EtBr and DNA associating or dissociation while traversing through the confocal spot. More recently Widom and Langowski showed the same method can be applied to follow nucleosome dynamics[146]:

$$G_{dyn} = \frac{G_{514R}}{G_{632R}} = C(\tau) \cdot \left(1 + K \cdot e^{-\tau(k_o+k_c)}\right) \cdot G_{tr} \cdot G_{ap} \tag{2.18}$$

With $K = \frac{k_o}{k_c}$, the equilibrium between open and closed nucleosomes described by the ratio between their opening (k_o) and closing (k_c) rates. Assuming equal diffusion times for open and closed nucleosomes, as assumed by Magde et al. makes the correction factor $C(\tau)$ equal to 1. However, previous experiments by Koopmans and Buning showed that closed nucleosomes diffuse significantly faster than open nucleosomes[53]. Therefore when calculating G_{dyn} we did not assume $\tau_D^{514R} = \tau_D^{632R}$ and included a correction, resulting in

$$C(\tau) = \sqrt{\frac{\tau_D^{632R} \cdot (\tau_D^{514R} + \tau)}{\tau_D^{514R} \cdot (\tau_D^{632R} + \tau)}} \tag{2.19}$$

Figure 2.9 shows the application of eq. **2.18** to data of molecules with (**2.9-b,d**, nucleosomes) and without (**2.9-a,c**, DNAFRET) internal dynamics. At first glance the correlation curves of DNAFRET (fig. **2.9a**) and nucleosomes (fig. **2.9b**) look very similar, as the two construct have very similar diffusion times. The only observable differences are in the apparent concentrations (height of the curves), which could be due to various other reasons than internal dynamics (or the lack thereof); to the untrained eye these two collections of curves could have originated from similar molecules. In order to determine internal dynamics accurately, we cut up a long measurement into smaller sets (subsets), correlated the smaller sets and calculated the average curves. This decreased effects of random noise from the signal, especially in the time range of diffusion and dynamics, as these characteristic times are unrelated to noise. In order to optimize the fits of the curves, we took the inverse of the standard deviation of a curve as the fits weighing factor. These tactics are independent of the time range, making the fitting algorithm flexible over a large time range, i.e. several magnitudes of molecule size.

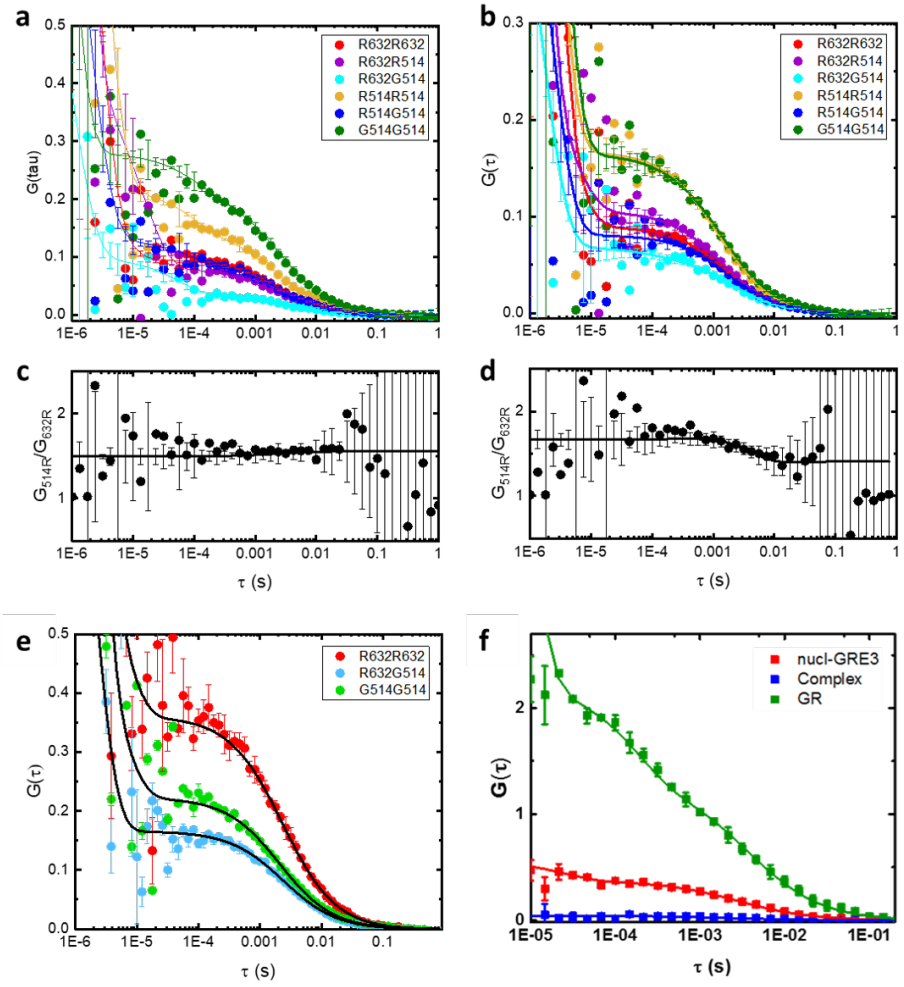

FIGURE 2.9: **Dynamics and binding in FCS are often subtle effects on correlation curves and hence need to be extracted through additional analysis. a)** and **b)** are correlation curves generated from signal DNAFRET and nucleosomes, respectively. Looking similar on first glance, DNAFRET displaying more particles in FRET channel, quantifying dynamics shows only nucleosomes (**d)**) switch between conformational states (K_{eq} for DNAFRET by fitting curve **c)** is $1.6 \cdot 10^9 \pm 6 \cdot 10^{17}$). **e)** Example of high affinity binding, observed in the crosscorrelation curve (blue) almost overlapping with the lowest autocorrelation curve. **f)** When binding is however (almost) absent, the crosscorrelation curve approaches $G(\tau) = 0$.

2.5 FCS data analysis

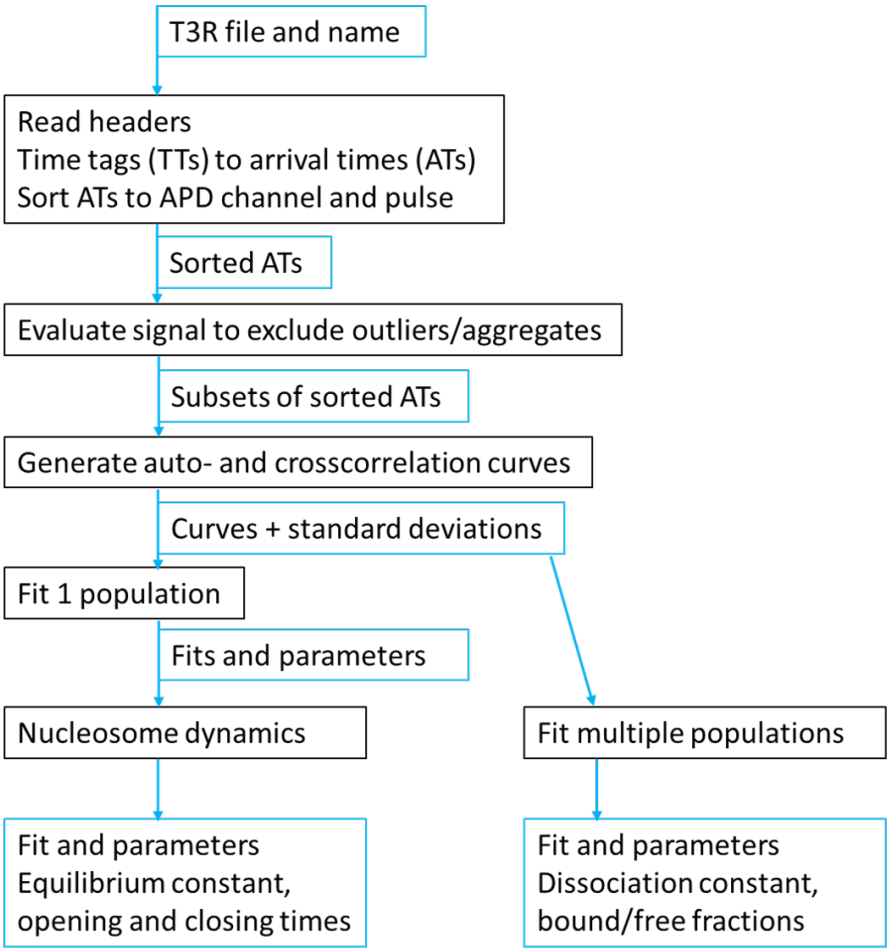

FIGURE 2.10: **Workflow of FCS data analysis.** Black boxes represent the parts of programs, in blue the data files and variables used as input and/or generated as output.

2.5.1 Data acquisition and processing

The signal generated by photons falling on the APD's during measurement was collected with a TimeHarp 200 (Time Measurement Histogram Accumulating Real-time Processor, software v. 6.1, Picoquant).The Time Resolved to Time Tagged (tttr or t3r) mode acquired signals at a 100 ns resolution, which was the also the pulse length of our PIE experiments. Measurements were saved in a 32-bit t3r file which consisted of several header lines and t3r records. After reading the header the t3r records were read as followed: first 16 bits containing the timetag, next 12 bits signifying the channel, next 2 bits the APD route, next 1 bit specifying if the timetag is valid. The last bit is reserved for system purposes. 16 Bits for the timetag and a resolution of 100 ns for each 'tick' means the photon arrival time that can be saved into a t3r record could only go to 2^{16} * 100 ns = 6553500 nanoseconds, or 6.5 milliseconds. This limitation is overcome by recording an overflow event. Each time the measurement exceeds n*65535 ticks, an event is recorded in the overflow channel and the 1 bit recording if a timetag was valid becomes 0 (= invalid). We counted the overflow events n and added n*65535 to the timetag. All time tags were multiplied with 100 ns to compute the actual photon arrival time in nanoseconds.

2.5.2 Assigning channels, PIE/ALEX phases and excluding aggregates

As no marker pulse was generated corresponding to the excitation pulses, we needed to relate the photon arrival times to their corresponding phase in the excitation cycle in another way. To accomplish this the excitation scheme and pulse resolution was included in the t3r filename. The arrival times were assigned a position in the excitation cycle.
The photon arrival times were assigned to their corresponding APD as this information is recorded by bits 29 and 30. The arrival times were then accumulated relative to the excitation phase. To determine the unknown phase shift, we evaluated which pulse position contained the highest number of arrival times and used the signals from the Green APD to align this phase with the 514 nm excitation. After determining the position of the 514G pulse, the sorted arrival times were aligned with a correction factor. The relevant photon arrival times were sorted in channels G514 (514 nm excitation, emission on Green APD), R632 (632 nm excitation, emission on Red APD)

and R514 (514 nm excitation, emission on Red APD).

parameter	channel	before	after
τ_D (ms)	R632	2.41 ± 0.7	1.64 ± 0.01
	R514	1.82 ± 1.5	1.18 ± 0.01
	G514	1.81 ± 0.13	1.40 ± 0.01
N	R632	12.24 ± 0.3	10.91 ± 0.1
	R514	6 ± 2	5.7 ± 0.1
	G514	6.4 ± 0.2	5.9 ± 0.5
dynamics	K_{eq}	1.3 ± 1837	2 ± 5
	t_{open} (ms)	361 ± 1262	5 ± 3
	t_{closed} (ms)	499 ± 1335	11 ± 5

TABLE 2.2: **A single aggregate can disturb correlation curves of molecules such, that their diffusion times appear 30% slower.** The fitted number of molecules is less affected by the presence of a single aggregate in figure **2.11a**. Kinetic rates were significantly affected, becoming 80 times slower. For all parameters, the fit errors reduced significantly upon exclusion of the aggregate.

Before correlating the photon signals we evaluated the signal stability over time of each channel. The mean intensity per second and standard deviation were calculated and used to exclude parts of the measurement exceeding a threshold of 2.5 of the standard deviation. The threshold could be changes in the main code of the program but was usually kept at 2.5. Time points where the mean signal exceeded n*std were saved in a separate list and used to divide the signal into shorter subsets of the photon arrival times. These subsets needed to span a measurement time large enough to generate a correlation curve full and accurate. For most measurements the length was set at 100 seconds, experience has shown to not set the subset length below 80 seconds.
The result of the exclusion algorithm is shown in Figure **2.11**. Some samples contained PEG from a previous purification step, which caused aggregation of the fluorescently labeled nucleosomes. Even the occurrence of a single aggregate during measurement can distort the resulting correlation curves and dynamics curve (figures **2.11-a,c,e,g**). Applying the exclusion algorithm resulted in deletion of 3.3% of the signals but yielded usable subsets and clean correlation curves that could be fitted accurately (figures

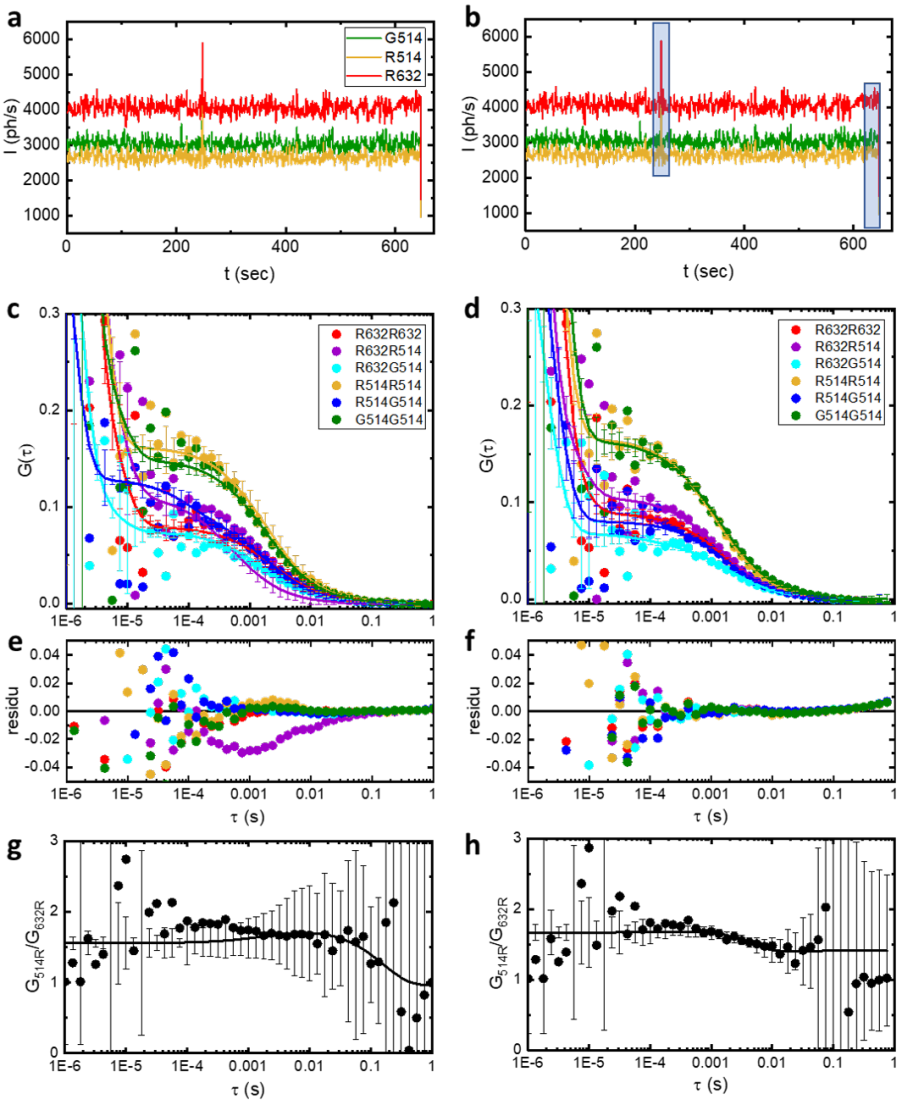

FIGURE 2.11: **A single aggregate can disturb fluorescence signals making fitting of diffusion and kinetics more inaccurate. a)** A single aggregate (at 250 seconds) changed the shapes of the correlation curves (**c**)) and hence decreased accuracy of the fits (**e**)). Also fitting kinetic rates became a more inaccurate exercise (**g**). Excluding 3.3 % (or 21 seconds) of the total 647 seconds shows fits becoming more accurate (**d**) and **h**)), and residues decreased (**f**)).

2.11-b,d,f,h). Comparing fit parameters shows that diffusion and kinetics are most affected by the exclusion algorithm (table **2.2**). The diffusion times and kinetic rates after exclusion resemble those found measurements that appear not disturbed by aggregates.

2.5.3 Correlation method and single-population fit

The hallmark of FCS is the quantification of the fluctuations from the mean intensity of a fluorescent signal, with the fluctuations representing molecules diffusing through a focus. From the correlation function described in equation **2.8** and visualized in figure **2.7** it appears there is always a signal ($I(t) \neq 0$) when binning over a large enough time. However, when reducing the bin size it becomes clear that the fluorescent signal is quite empty. Consider a signal of 1000 photons per second: reducing the bin size to one microsecond results in the signal becoming 0.001 photons per microsecond (or 1 photon being detected every 1000 microseconds). The temporal resolution used in our PIE experiments was 100 nanoseconds, meaning 1 in 10000 chance of detecting a photon (excluding background / dark photons) every 100 nanoseconds. As FCS measurements usually take minutes it is wise to choose an appropriate set of lag times τ that minimizes computational effort. A commonly used method to speed up calculations is the use of the multiple-τ algorithm[147]. This method entails calculating the first 16 time lags at full resolution, decreasing this resolution by half and grouping photons in bins accordingly. The process is repeated for 8 more time lags, and so forth, until a chosen point where minimal correlation is expected. The multiple-τ algorithm offers a dynamic range of time lags, but increasing resolution to sample more τ values also increases computation time. Hence we chose the algorithm developed by Laurence et al.[134] where the values of τ are not rigid but depend only on the time scale of the measurement; a base and the number of points (time lags) per base are defined in a time range in seconds. The scale invariance resulting from choosing an appropriate base speeds up computation time significantly.
As one cycle of excitation pulses in our PIE measurements was 800 nanoseconds long, the correlation time range was set from 10^{-6} to 10^{1} seconds. The base was set to 10 (meaning log^{10}) and number of points per base at 8. Oversampling by increasing the number of points per base lead to larger fluctuations in the correlation curve at small τ stemming from the dark periods in the pulse cycle. Undersampling by decreasing number of points

per base resulted in coarser correlation curves at large τ making fitting less accurate as also described by Laurence[134].
All correlation curves were fitted with equation **2.18**; parameter bounds and initial guesses were set as shown in table **2.3**. Only the value of a was fixed (at 8), all other parameters were free to be fitted. The standard deviation from dividing the arrival times into subsets was used as weight factor to minimize the residue of the data and fit. The `Minimizer` function in our Python program used the Levenberg-Marquardt algorithm to compute the best fit.

parameter	initial guess	$bound_{min}$	$bound_{max}$
N	1	$1E-4$	$1E3$
τ_D (s)	$1E-3$	$5E-4$	1
F_{tr}	0.1	0	0.9
τ_{tr} (s)	7E-6	5E-6	1E-4
F_{ap}	0.1	0	0.9
τ_{ap} (s)	1E-6	1E-7	5E-6

TABLE 2.3: **By choosing physical sensible initial guesses, bounds can be set wide to capture a wide range of possibilities.** Also setting bounds of characteristic times for diffusion, triplet states and afterpulsing to overlap did not result in fits capping on a bound.

parameter	initial guess	$bound_{min}$	$bound_{max}$
A	1	0.97	1.05
k_{on} (s^{-1})	120	0	2000
k_{off} (s^{-1})	80	0	2000
F_{tr}	0.1	0	0.9
τ_{tr} (s)	3E-6	1E-7	5E-5

TABLE 2.4: **Parameter bounds for fitting nucleosome dynamics.** The wide ranges made it possible to accurately fit the different dynamics when batch processing files of measurement of nucleosomes at different salt concentrations.

When differences in diffusion coefficient were expected, the relative correlation curve of R514/R632 was fitted with equation 2.19, with bounds and initial guesses as in table 2.4. We expected the contribution

of afterpulsing (ap) to vanish by dividing the correlation curves and leave only the effect of ATTO647N going into the triplet (tr) state. The fit was optimized using the LM algorithm and weighted with the relative errors of correlation curves R514 and R632.

2.5.4 Multi-population fit strategies

To determine stoichiometry of populations in a sample with two interacting molecules, or when protein labelling efficiency or concentration is known. The correlation curve is composed of the product of two fractions, for instance closed and open, or bound and free as represented in the diffusional part of G(tau) for different molecules:

$$G_{diff}(\tau) = N_{total}^{-1} \cdot \left(F_1 \cdot (1 + \tau/\tau_{D1}) \cdot (1 + a^{-2} \cdot \tau/\tau_{D1})\right)^{-1/2} \cdot \left(F_2 \cdot (1 + \tau/\tau_{D2}) \cdot (1 + a^{-2} \cdot \tau/\tau_{D2})\right)^{-1/2} \tag{2.20}$$

with $F_1 + F_2 = 1$. For nucleosomes the closed fraction is obtained from the ratio between the autocorrelation curve from the FRET channel and the direct excitation of the acceptor ($F_{closed} = \frac{N_{closed}}{N_{total}}$). From the fit of the FRET channel, we also obtained the characteristic diffusion time ($\tau_{D,closed}$), leaving only the diffusion time of open nucleosomes $\tau_{D,open}$ to be fitted by the multi-population fit.
The multi-population fit algorithm can also accurately fit more than one unknown parameter. When estimating labelling efficiency of a protein with only the diffusion time of free dye known, the algorithm is able to find from the autocorrelation curve the characteristic diffusion time of the labeled protein as expected from calculations. The ratio of the fractions was confirmed by gel electrophoresis. The multi-population algorithm can be expanded to estimate a third population, though this requires several different baseline experiments to reduce variable parameters.

2.6 Conclusions

In this chapter we have described the procedures for proper preparation, calibration, characterization and correction of samples and setup, as well as the analysis process for single-molecule experiments. Quantifying the intricacies of fluorophore signals in the microscope setup and optimizing alignment of the setup yielded correct physical parameters of fitting the correlation curves from various channels.
The use of ALEX or PIE allowed to distill in addition to the concentration and diffusion times, kinetic rates and multiple populations. The algorithm presented here to exclude aggregate effects from long measurements is a novelty in the FCS field. By using this algorithm, one does not have to measure 'around the aggregates' as was the recommended method[139]. The excluding method can also be used inversely to investigate condensates. Overall, using multiple excitation and detection colors in combination with multiple fluorophores and advanced correction and analysis protocols makes it possible to accurately quantify concentrations, diffusion times and interaction rates within a single experiment with relatively little sample. These features go beyond the capabilities of traditional single color FCS.