Supporting Information

Combining Graphical and Analytical Methods with Molecular Simulations to Analyze Time-resolved FRET-measurements of Labeled Macromolecules Accurately

Thomas-Otavio Peulen, Oleg Opanasyuk, Claus A.M. Seidel*
Lehrstuhl für Molekulare Physikalische Chemie, Heinrich-Heine-Universität, Universitätsstraße 1, 40225 Düsseldorf, Germany
*: corresponding author: cseidel@hhu.de

DOI: 10.1021/acs.jpcb.7b03441

Table S1. Fluorescence parameters of the fluorescent dye Alexa647 coupled to various proteins.

Exponential components of fluorescence intensity decay and residual anisotropies of Alexa647 attached to different labeling sites to T4 Lysozyme (T4L), *P. aeruginosa* lipase foldase (LiF), a cysteine free Cyclin-dependent kinase inhibitor 1B variant (p27), Post-synaptic density protein 95 (PSD-95).

		Fluorescence lifetime distribution[b]						Species weighted lifetime[c]	Residual anisotropy [d]
Protein	**Variant**	$x^{(1)}$	$\tau^{(1)}$ / ns	$x^{(2)}$	$\tau^{(2)}$ / ns	$x^{(3)}$	$\tau^{(3)}$ / ns	$\langle\tau\rangle_x$ / ns	r_∞
T4L	Q69C	0.40	1.72	0.60	0.96			1.26	0.23
	P86C	0.41	1.60	0.59	0.89			1.18	0.18
	R119C	0.43	1.66	0.57	0.93			1.25	0.21
	N132C	0.52	1.67	0.48	0.90			1.30	0.29
	I150C	0.47	1.73	0.53	0.94			1.29	0.23
LiF	A215C	0.24	2.10	0.67	1.36	0.08	0.60	1.47	0.25
	Q263C	0.52	2.08	0.41	1.46	0.06	0.42	1.71	0.33
	R296C	0.44	2.00	0.50	1.38	0.07	0.34	1.58	0.28
	I297C	0.41	2.24	0.49	1.42	0.11	0.39	1.64	0.28
p27	C29C/E54C	0.60	1.80	0.40	0.86			1.13	0.10
	E54C/R93C	0.62	1.79	0.38	0.83			1.23	0.09
	E75C/S110C	0.60	1.76	0.40	0.94			1.43	0.14
PSD-95[a]	S142C/A230C	0.81	1.77	0.19	0.89			1.61	0.30
	S142C/Y236C	0.56	1.77	0.44	1.17			1.51	0.28
	E135C/Y236C	0.30	1.90	0.70	1.30			1.48	0.27
	Q107C/Y236C	0.68	1.77	0.32	1.02			1.53	0.27
	E135C/A230C	0.71	1.78	0.29	0.83			1.50	0.28
	S142C/M159C	0.78	1.76	0.22	0.97			1.59	0.32
	Q107C/H182C	0.91	1.80	0.09	1.00			1.73	0.32
	D91C/A230C	0.48	2.09	0.52	1.55			1.81	0.33
	S142C/K202C	0.83	1.84	0.17	1.00			1.70	0.31
	E135C/M159C	0.65	1.78	0.35	0.93			1.48	0.30

[a] Different variants were labeled by Alexa647 C2 maleimide (order number: A20347). The naming scheme highlights introduced mutations and potential labeling position by the original amino acids, sequence numbers and the introduced mutations. In PSD-95, hGBP1, HIV-RT, LiF and T4L native amino acids were replace by cysteines. These cysteines were labeled by using maleimide chemistry. In PSD-95 and p27 two cysteines were present. Thus, Alexa647 is distributed among two potential labeling sites. [b]The fluorescence lifetimes were determined by a fitting a multi-exponential relaxation model $f(t) = k_{F,D}\sum_{i=1}^{N} x^{(i)} \exp(-t/\tau^{(i)})$ to experimental fluorescence decays. [c]The species averaged lifetimes $\langle\tau\rangle_x = \sum_{i=0}^{n} x_i \tau^{(i)}$ were calculated using the fitted species fractions $x^{(i)}$ and lifetimes $\tau^{(i)}$. [d]The residual anisotropies r_∞ were determined by the offset of the time-resolved anisotropy decays $r(t)$.

Table S2. Fluorescence parameters of the fluorescent dye Alexa488 coupled to various proteins.

Exponential components of fluorescence intensity decay and residual anisotropies of Alexa488 attached to different labeling sites to T4 Lysozyme (T4L), Post-synaptic density protein 95 (PSD-95), a cysteine free variant of human guanylate binding protein 1 (hGBP1), HIV reverse transcriptase (HIV-RT), and *P. aeruginosa* lipase foldase (LiF).

		Fluorescence lifetime distribution[b]						Species weighted lifetime[c]	Residual anisotropy [d]	Simulated species weighted lifetime [e]
Protein	**Variant[a]**	$x^{(1)}$	$\tau^{(1)}$/ns	$x^{(2)}$	$\tau^{(2)}$/ns	$x^{(3)}$	$\tau^{(3)}$/ns	$\langle\tau\rangle_x$ / ns	r_∞	$\langle\tau\rangle_x$ / ns
T4L	R8pAcF	0.89	4.10	0.11	1.60			3.83	0.25	3.98
	R19pAcF	0.88	3.85	0.12	1.68			3.58	0.25	3.67
	S44pAcF	0.96	4.32	0.04	1.44			4.20	0.19	4.12
	N55pAcF	0.92	4.14	0.08	1.48			3.93	0.19	4.07
	K60pAcF	0.88	4.20	0.12	1.76			3.92	0.20	3.99
	K60pAcF	0.89	4.20	0.11	1.47			3.90	0.21	4.08
	Q69pAcF	0.88	4.08	0.12	1.12			3.73	0.17	4.17
PSD-95	D91C	0.66	3.66	0.34	0.41			2.56	0.25	2.45
	Q107C	0.66	3.76	0.34	0.27			2.57	0.16	2.94
	E135C	0.77	3.89	0.23	0.27			3.06	0.18	3.35
	S142C	0.73	3.96	0.27	0.22			2.95	0.16	3.92
	A230C	0.74	3.73	0.26	0.46			2.88	0.21	3.23
	Y236C	0.71	3.74	0.29	0.35			2.76	0.18	2.11
hGBP1	N18C	0.82	4.15	0.18	1.35			3.65	0.11	3.38
	Q254C	0.69	3.60	0.31	0.53			2.65	0.12	3.12
	Q344C	0.94	3.78	0.06	1.00			3.61	0.11	3.04
	T481C	0.93	3.78	0.07	0.37			3.54	0.09	2.97
	A496C	0.84	3.68	0.16	0.32			3.14	0.25	3.57
	Q525C	0.80	3.51	0.20	0.66			2.94	0.09	3.56
	V540C	0.85	3.87	0.15	1.50			3.51	0.19	3.78
	Q577C	0.91	4.15	0.09	1.49			3.91	0.14	4.05
HIV-RT	p66Q6C	0.78	3.67	0.22	0.88			3.05	0.24	2.09
	p66T27C	0.76	3.60	0.24	0.73			2.91	0.20	3.53
	p66E194C	1.00	3.87					3.87	0.13	3.44
	p66K287C	0.70	3.50	0.30	0.78			2.67	0.27	3.24
	p51Q6C	0.80	3.65	0.20	0.88			3.09	0.24	2.31
	p51K173C	0.79	3.70	0.21	0.98			3.12	0.26	2.37
	p51E194C	0.82	3.86	0.18	1.07			3.37	0.13	3.62
	p51K281C	0.82	3.77	0.18	0.93			3.25	0.25	3.00
LiF	A215C	0.85	4.17	0.09	1.98	0.06	0.25	3.74	0.14	
	Q263C	0.83	4.28	0.11	2.27	0.06	0.35	3.82	0.18	
	R296C	0.81	4.22	0.12	2.20	0.07	0.38	3.72	0.14	
	I297C	0.76	4.09	0.14	1.90	0.11	0.28	3.39	0.14	

[a] Different variants were site-specifically labeled either by Alexa488 C5 maleimide (order number: A10254) or Alexa488 hydroxylamine (order number: A30632). The naming scheme highlights the labeling position by the original amino acid, its sequence number and the introduced mutations. In PSD-95, hGBP1, HIV-RT, LiF native amino acids were replace by cysteines. These cysteines were labeled with Alexa488 C5 maleimide. In T4L the unnatural amino acid p-acetyl-L-phenylalanine (pAcF) was introduced. The keto group of pAcF was labeled by Alexa488 hydroxylamine. As HIV-RT is complex consisting of two sub-units p51 and p66, the respective subunit name additionally given. **[b]** The fluorescence lifetimes were determined by a fitting a multi-exponential relaxation model $f(t) = k_{F,D}\sum_{i=1}^{N} x^{(i)} \exp(-t/\tau^{(i)})$ to experimental fluorescence decays. **[c]**The species averaged lifetimes $\langle\tau\rangle_x = \sum_{i=0}^{n} x_i \tau^{(i)}$ were calculated using the fitted species fractions $x^{(i)}$ and lifetimes $\tau^{(i)}$. **[d]** The residual anisotropies r_∞ were determined by the offset of the time-resolved anisotropy decays *r(t)*. **[e]** The simulated species averaged lifetimes were determined by simulating the fluorescence decay using parameters as given in Fig. 8 and protein structures as given in Table S3.

Table S3. Crystal structures used in the BD simulations presented in Figure 8

Protein	PDB
T4 Lysozyme (T4L) [(a)]	148L, 172L
Human guanylate binding protein 1 (hGBP1)	1F5N
HIV reverse transcriptase (HIV-RT)	1RTD
PSD-95	3ZRT

[(a)] In case of T4L it was assumed that 50% is in the "closed" conformation 148L and 50% in the open conformation 172L.

Note S1. Decay analysis by normally distributed distances

As demonstrated in Fig. 13, the distance distribution between a donor D and acceptor dye A can be approximated by a normal distribution. If in a mixture of N states the D and A distance distribution $x(R_{DA}^{(j)})$ of an individual state (j) is normal width a width w_{DA}, and expected distance $\overline{R_{DA}^{(j)}}$ and the fraction of each state is $x_{DA}^{(j)}$ the total distribution of the distances $x(R_{DA})$ is given by the species weighted average:

$$x(R_{DA}) = \frac{1}{w_{\mathrm{DA}}\sqrt{\pi/2}} \sum_{j=1}^{N} \left[\mathrm{x}_{\mathrm{DA}}^{(\mathrm{j})} \cdot \exp\left(-2 \left[\frac{R_{\mathrm{DA}} - \overline{R_{DA}^{(j)}}}{w_{DA}} \right]^2 \right) \right] \tag{S1}$$

By combining the above equation with eq. (7), (9) and (17) a model function for the time-resolved fluorescence decay of the donor in presence of an acceptor $f_{D|D}^{(DA)}(t)$ is obtained.

Note S2. Time-dependent yield of FRET

The steady-state transfer efficiency is defined by:

$$E = 1 - \frac{F_{D|D}^{(DA)}}{F_{D|D}^{(D0)}} \tag{S2}$$

Here, $F_{D|D}^{(D0)}$ and $F_{D|D}^{(DA)}$ are the time-resolved fluorescence intensities integrated over the time (steady-state):

$$F_{D|D}^{(DA)} = \int_{t=0}^{t=\infty} f_{D|D}^{(DA)}(t)dt \quad \text{and} \quad F_{D|D}^{(D0)} = \int_{t=0}^{t=\infty} f_{D|D}^{(D0)}(t)dt \tag{S3}$$

To obtain a time-resolved quantifier which provides the steady-state transfer-efficiency E in the limit $t \to \infty$, $F_{D|D}^{(D0)}$ and $F_{D|D}^{(DA)}$ have to be replaced by the cumulative intensities:

$$F_{D|D}^{(DA)}(T) = \int_{t=0}^{T} f_{D|D}^{(DA)}(t)dt \quad \text{and} \quad F_{D|D}^{(D0)}(T) = \int_{t=0}^{T} f_{D|D}^{(D0)}(t)dt \tag{S4}$$

Using the cumulative intensities, a time-dependent quantity is obtained with the meaning of a transfer-efficiency:

$$E(T) = 1 - \frac{\int_0^T f_{D|D}^{(DA)}(t)dt}{\int_0^T f_{D|D}^{(D0)}(t)dt} \quad \text{(S5)}$$

This quantity describes the time-dependent yield of the FRET-process up to the time *T*.
Van der Meer defines the "time-dependent transfer-efficiency" (*TRE*) as:

$$TRE(t) = 1 - \frac{f_{D|D}^{(DA)}(t)}{f_{D|D}^{(D0)}(t)} \quad \text{(S6)}$$

This does not quantify the *yield* of the FRET-process. Thus, the *TRE* is not a FRET-efficiency. In a mixture of fluorescent species its asymptote provides the species fraction of FRET-active molecules and not the FRET-efficiency.

Note S3. Accessible volume simulations to assess the effect of labeling symmetry

Since, the dyes were tethered to the protein by long linkers, the spatial distribution of the flurophores had to be considered. The dye distributions were modeled by the accessible-volume (AV) approach according to [75, 108, 109]. The AV-approach uses a geometric search algorithm to determine all dye positions within the linker-length from the attachment point which do not cause steric clashes with the macromolecular surface. The dyes were approximated by ellipsoids. The center of each ellipsoid was connected to its attachment point by a flexible linkage of a length L_{link}. Here, the C_β-atoms were used as attachment points. The linker-length is given by the longest distance from the attachment point (C_β-atom of the cysteine) to the center of the dye. It includes the reactive group, a spacer and the internal linker of the dye. Both, Alexa Fluor 488 C5 maleimide (Alexa488) and Alexa Fluor 647 C2 maleimide (Alexa647) were modeled with a linker width of 4.5 Å. As linker-lengths L_{link} 20.5 Å and 22 Å were used for Alexa488 and Alexa647, respectively. The radii of the ellipsoid (R_{Dye1}, R_{Dye2} and R_{Dye3}) were determined by the spatial dimensions of the dyes. Alexa488 was modeled using radii of 5.0 Å, 4.5 Å and 1.5 Å. Alexa647 was modeled using radii of 11.0 Å, 4.7 Å and 1.5 Å. To study the effect of the linker-length on the symmetry, the fluorophore pair BodipyFL C1 iodacetamine (Bodipy) and Alexa647 was simulated. To simulate Bodipy a linker-length of 10.8 Å and width of 4.5 Å were used while the dye shape was approximated by radii of 4.5 Å, 3.2 Å and 0.9 Å.

To determine the effect of the labeling symmetry as shown in Fig. 13, a set of 5592 protein structures with at least 360 amino acids in the chain, a minimum resolution of 1.8 Å and no unresolved amino acids was selected from the protein databank using the program "PDBselect"[103]. For each structure at least 180 random amino acid pairs were chosen.

Next, for each pair of amino acids the accessible volumes of the pair DA, where the donor is located at the first amino acid, and the AD-pair, where the donor is located at the second amino acid, were simulated. Using the AV-simulations, for both pairs the mean and the width of the distance distribution were calculated. In case if one of the two amino acids was buried within the structure and inaccessible for the dye, the amino acid pair was discarded. To discriminate inaccessible labeling sites FRET-pairs were discarded if a volume of an AV was smaller than 3.0% of the average AV-volume over all structures.

In absence of surface interactions, a main peak and a shoulder are visible (Fig. 9A, top). In presence of surface interaction, the width of $x(R_{app})$ increases, its mean distance shifts by ~3 Å and the shoulder is less pronounced (Fig. 9A, bottom). The features of $x(R_{app})$ depend on the diffusion coefficients. In case the dyes interact with the surface, they diffuse in average slower and thus, the differences between $x(R_{DA})$ and $x(R_{app})$ are less pronounced. In both cases the mean distance of $x(R_{app})$ is shifted by 2 Å indicating that the mean of $x(R_{DA})$ can be approximated by the mean of $x(R_{app})$. However, the width of the $x(R_{app})$ is decreased by ~3 Å for the chosen diffusion coefficient. Such narrowing was previously experimentally observed [9, 75].

Note S4. The estimation of statistical errors

To estimate statistical errors due to the photon noise we use the Fisher information matrix (FIM) and the Cramér–Rao inequality. The Cramér–Rao inequality states that the variance-covariance matrix Σ is bigger than the inverse of the FIM I ($\Sigma \geq I^{-1}$). For two model parameters α and β the elements of the FIM are given by:

$$I_{\alpha,\beta} = E\left[\partial_\alpha \ln \ell(\alpha,\beta)\, \partial_\beta \ln \ell(\alpha,\beta)\right] \tag{S7}$$

Here $\ell(\alpha,\beta)$ is a likelihood function which quantifies the agreement between the model and the experiment. In TCSPC experiments with n-detection bins and N detected photons with N_i photons per bin the likelihood is given by a multinomial probability mass function:

$$\ell(\alpha,\beta) = N! \prod_{i=1}^{n} \frac{f_i^{N_i}(\alpha,\beta)}{N_i!} \tag{S8}$$

Here f_i is the probability of detecting a photon in a bin i assumed by the model-function. Thus, Nf_i is the expectation value in the bin i. The probability f_i is obtained from the continuous model function $f(t,\alpha,\beta)$ describing the experiment by piecewise integration and normalization:

$$f_i(\alpha,\beta) = \frac{\int_{(i-1)\Delta t}^{i\Delta t} f(t,\alpha,\beta)dt}{\int_0^{n\Delta t} f(t,\alpha,\beta)dt} \tag{S9}$$

Here, Δt is the bin-width of the fluorescence intensity decay histogram. As the likelihood function is multinomial the FIM takes the simple form:

$$I_{\alpha,\beta} = N\sum_{i=1}^{n} \frac{\partial_\alpha f_i(\alpha,\beta)\,\partial_\beta f_i(\alpha,\beta)}{f_i(\alpha,\beta)} \tag{S10}$$

Under these conditions the FIM does not contain the experimental information. Hence, the variances and co-variances can be predicted *a priori* given a model function.