bioRxiv preprint doi: https://doi.org/10.1101/083287; this version posted October 26, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Applying Corrections in Single-Molecule FRET

Antonino Ingargiola*[1]

[1]Dept. Chem. & Biochem, Univ. California Los Angeles, Los Angeles, CA, USA.

October 26, 2016

Abstract

Single-molecule Förster resonance energy transfer (smFRET) experiments can detect the distance between a donor and an acceptor fluorophore on the 3-10nm scale. In ratiometric smFRET experiments, the FRET efficiency is estimated from the ratio of acceptor and total signal (donor + acceptor). An excitation scheme involving two alternating lasers (ALEX) is often employed to discriminate between singly- and doubly-labeled populations thanks to a second ratiometric parameter, the stoichiometry S. Accurate FRET and S estimations requires applying three well-known correction factors: donor emission leakage into the acceptor channel, acceptor direct excitation by the donor excitation laser and the "gamma factor" (i.e. correction for the imbalance between donor and acceptor signals due to different fluorophore's quantum yields and photon detection efficiencies).

Expressions to directly correct both raw FRET and S values have been reported in [1] in the context of freely-diffusing smFRET. Here we extend Lee *et al.* work providing several expressions for the direct excitation coefficient and highlighting a clear interpretation in terms of physical parameters and experimental quantities. Moreover, we derive a more complete set of analytic expressions for correcting FRET and S. We aim to provide a clear and concise reference for different definitions of correction coefficients and correction formulas valid for any smFRET experiment both in immobilized and freely-diffusing form.

Contents

1 Introduction **2**

2 Definitions **3**

2.1 Fluorescence intensities 3

2.2 FRET and Stoichiometry 4

2.3 Definition of direct excitation 5

2.4 Discussion of Definitions 1-5 7

2.5 Correction formulas 8

*ingargiola.antonino@gmail.com

bioRxiv preprint doi: https://doi.org/10.1101/083287; this version posted October 26, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

3 Conclusion **8**

1 Introduction

Förster resonance energy transfer (FRET) is a Coulombic interaction between the dipoles of two fluorophores, which results in the resonant and non-radiative transfer of excitation energy from a donor to an acceptor fluorophore (and the energy transfer probability decreases with the sixth power of the distance). Donor de-excitation via FRET competes with the donor's intrinsic radiative and non-radiative de-excitation paths. Therefore, in the presence of a nearby acceptor, the lifetime of the donor is reduced. The quantum yield, or efficiency, of the FRET process can be computed as [cite Clegg and Lakowicz books]:

$$E = 1 - \frac{\tau_{FRET}}{\tau_D} \tag{1}$$

where τ_{FRET} is the D lifetime in presence of FRET and τ_D is the intrinsic D lifetime without any acceptor nearby. Computing E following eq. 1 requires measuring the D excited-state lifetime, for example using a TCSPC setup. A simpler method of estimating E consists in measuring only the intensity of donor and acceptor fluorescence (F_D and F_A respectively) and computing the FRET efficiency ratiometrically as:

$$E = \frac{F_A}{F_A + F_D} \tag{2}$$

The previous eq. 1 and 2 require that the fluorescence lifetimes or intensities be relative to a single specie. In biological samples, where almost inevitably multiple FRET populations are present, single-molecule FRET (smFRET) experiments allows identifying different sub-populations and, for each of them, estimating the FRET efficiency [cite Shimon's science review].

The ratiometric approach of computing E is very common in smFRET owing to its modest hardware requirements (compared to TCSPC measurements) and has been extensively applied both to freely-diffusing and to surface-immobilized experiments. Unfortunately, unlike lifetime-based experiments, ratiometric FRET is affected by three systematic errors (or biases) intrinsic to the way F_A and F_D are measured. The first, a fraction of the donor emission spectrum almost inevitably falls in the acceptor detection band, causing spurious increase in acceptor-channel signal named "donor leakage". Additionally, the acceptor signal is contaminated by a fraction of fluorescence due to direct excitation of the acceptor fluorophore by the donor laser (ideally the acceptor should only be excited by the donor). Finally, the relative (detected) donor and acceptor fluorescence intensity is biased because of the different fluorescence quantum yields and photon detection efficiencies in the two detection channels (requiring the so-called "gamma factor" correction). These biases are well-known and expressions for their correction have been derived [1].

Contrary to ensemble measurements, single-molecule experiments can resolve different subpopulations and recover mean/peak FRET efficiencies of each single conformational or binding state (at least in cases where there are no conformations that

bioRxiv preprint doi: https://doi.org/10.1101/083287; this version posted October 26, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

interconvert much faster than diffusion times). However, obtaining accurate mean FRET efficiencies also requires applying corrections for the aforementioned biases.

This paper extends the ratiometric FRET corrections reported in Lee *et al.* [1]. We define the acceptor direct excitation as a function of different observable. For each definition we derive the direct excitation coefficient as a function of physical parameters and discuss its physical interpretation. We also derive a complete set of formulas for computing E or S as a function of the raw E and S as well as of the aforementioned correction factors. Note that the expression here presented are valid for any ratiometric smFRET or ALEX-smFRET experiment, being it immobilized or freely diffusing [2, 3].

2 Definitions

2.1 Fluorescence intensities

We start by defining the fluorescence intensity signal as a function of the physical parameters. For surface-immobilized measurements the signal can be donor or acceptor counts acquired in a camera frame for a given fluorophore. For freely-diffusing experiments the signal can be the counts detected in the donor and acceptor channel during a "burst" (i.e. a single molecule crossing the excitation volume). Following [1] we define:

$$n_d = I_{D_{ex}}\, \sigma^D_{D_{ex}}\, \phi_D\, \eta^{D_{em}}_{D_{det}}\, (1 - E) \tag{3}$$

$$n_a = I_{D_{ex}}\, \sigma^D_{D_{ex}}\, \phi_A\, \eta^{A_{em}}_{A_{det}}\, E \tag{4}$$

$$n_{aa} = I_{A_{ex}}\, \sigma^A_{A_{ex}}\, \phi_A\, \eta^{A_{em}}_{A_{det}} \tag{5}$$

Eq. 3 and 5 are the detected quantities (e.g. counts, or camera intensity) after background correction in the DexDem and AexAem photon streams respectively. The n_a quantity (eq. 4) needs to be estimated correcting the measured counts n^*_a in the DexAem stream (see eq. 10). The factors I, σ, ϕ and η are, respectively, the excitation intensity, the absorption cross-section, the fluorophore quantum yield and the photon detection efficiency. The label D_{ex} (resp. A_{ex}) indicates a coefficient computed at the donor-laser excitation wavelength. D_{det} (resp. A_{det}) indicated the donor detection band. Finally, in the σ coefficient the superscript D or A indicates the fluorophore. In addition to these quantities, we need to introduce the correction coefficient γ and β which are defined as follows:

$$\gamma = \frac{\phi_A\, \eta^{A_{em}}_{A_{det}}}{\phi_D\, \eta^{D_{em}}_{D_{det}}} \tag{6}$$

$$\beta = \frac{I_{A_{ex}} \sigma^A_{A_{ex}}}{I_{D_{ex}} \sigma^D_{D_{ex}}} \tag{7}$$

Briefly, γ makes the DexDem and DexAem signals commensurable (i.e. on the same scale) taking into account difference in dyes quantum yields and photon detection

bioRxiv preprint doi: https://doi.org/10.1101/083287; this version posted October 26, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

efficiencies. Similarly, the β factor is used to make the total Dex signal commensurable with the AexAem signal by taking into account the differences in excitation intensities ($I_{A_{ex}}$ vs $I_{D_{ex}}$) and in dyes absorption cross-sections ($\sigma^{A}_{A_{ex}}$ vs $\sigma^{D}_{D_{ex}}$). This expression of the β coefficient has been derived in [1] during the derivation of the fitting procedure for γ-factor.

It is also useful to introduce the total corrected signal during D-excitation which we can define equivalently with one of the following expressions:

$$n_d + n_a/\gamma = I_{D_{ex}}\, \sigma^{D}_{D_{ex}}\, \phi_D\, \eta^{D_{em}}_{D_{det}} \tag{8}$$

$$\gamma\, n_d + n_a = I_{D_{ex}}\, \sigma^{D}_{D_{ex}}\, \phi_A\, \eta^{A_{em}}_{A_{det}} \tag{9}$$

The choice between eq. 8 and 9 is only matter of convention. Finally, in a real experiment we cannot measure n_a directly, instead we acquire a value n^*_a that is contaminated by donor leakage (Lk) and acceptor direct excitation (Dir). We define n^*_a and the correction terms as follows:

$$n^*_a = n_a + Lk + Dir \tag{10}$$

$$Lk = I_{D_{ex}}\, \sigma^{D}_{D_{ex}}\, \phi_D\, \eta^{D_{em}}_{A_{det}}\, (1 - E) \tag{11}$$

$$Dir = I_{D_{ex}}\, \sigma^{A}_{D_{ex}}\, \phi_A\, \eta^{A_{em}}_{A_{det}} \tag{12}$$

Consistently with the n_d and n_{aa} definitions (eq. 3 and 5), the quantity n^*_a is assumed already background corrected.

2.2 FRET and Stoichiometry

We start defining the FRET efficiency E and the proximity ratios E_{PR} and E_R:

$$E_R = \frac{n^*_a}{n^*_a + n_d} \tag{13}$$

$$E_{PR} = \frac{n_a}{n_a + n_d} \tag{14}$$

$$E = \frac{n_a}{n_a + \gamma\, n_d} \tag{15}$$

where n_d, n_a are the donor and acceptor detected counts after all the corrections (see eq. 3 and 4), while n^*_a are the acceptor counts with only background correction of eq. 10 (no leakage and direct excitation corrections).

Similarly, for the stoichiometric ratio we can have different definitions depending on the degree on corrections that are applied:

bioRxiv preprint doi: https://doi.org/10.1101/083287; this version posted October 26, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

$$S_R = \frac{n_d + n_a^*}{n_d + n_a^* + n_{aa}} \tag{16}$$

$$S_{PR} = \frac{n_d + n_a}{n_d + n_a + n_{aa}} \tag{17}$$

$$S_\gamma = \frac{\gamma n_d + n_a}{\gamma n_d + n_a + n_{aa}} \tag{18}$$

$$S_{\gamma\beta} = \frac{\gamma n_d + n_a}{\gamma n_d + n_a + n_{aa}/\beta} \tag{19}$$

S_R (eq. 16) is the raw stoichiometry without any correction except for background (see definition of n_a^* in eq. 10). S_{PR} (eq. 17) is the stoichiometry corrected for leakage and direct excitation (see n_a definition in eq. 4). S_γ (eq. 18) is the stoichiometric ratio corrected for leakage, direct excitation and γ (so that FRET populations have stoichiometry centered around a constant value, typically close to 0.5). $S_{\gamma\beta}$ (eq. 19) includes a β correction ensuring that FRET populations have stoichiometry centered around 0.5. Since β (eq. 7) is equal to the ratio $n_{aa}/(n_a + \gamma n_d)$ (eq. 4 and 9) it follows that:

$$S_\gamma = \frac{1}{1+\beta} \tag{20}$$

From eq. 20 follows that when β is known it is possible to compute the S_γ value around which all FRET populations are distributed. As noted before, for $S_{\gamma\beta}$ this value is always 0.5.

2.3 Definition of direct excitation

The term Dir can be equivalently expressed as a fraction of any fluorescence intensity components (i.e. counts in the donor or acceptor channel during donor or acceptor excitation). Here we present five different definitions and their physical interpretation.

Definition 1

Defining Dir as a function of n_{aa} we have:

$$Dir = d_{AA} \cdot n_{aa} \tag{21}$$

The coefficient d_{AA} can be computed from an acceptor-only population in ALEX measurement, because in this case eq. 10 becomes $n_a^* = Dir$. In terms of physical parameters, recalling eq. 12 and 5, we can express d_{AA} as:

$$\begin{aligned} d_{AA} &= \frac{Dir}{n_{aa}} = \\ &= \frac{I_{D_{ex}}}{I_{A_{ex}}} \frac{\sigma^A_{D_{ex}}}{\sigma^A_{A_{ex}}} \end{aligned} \tag{22}$$

bioRxiv preprint doi: https://doi.org/10.1101/083287; this version posted October 26, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Since computing Dir through d_{AA} requires the spectroscopic quantity n_{aa} (see eq. 5), it cannot be used in case of single-excitation measurements. In this case the definitions in the next section can be used. Note that d_{AA} is indicated as d in [1].

Definition 2

Defining Dir as a function of the "corrected total signal" as defined in eq. 8 ($(n_a + \gamma\, n_d)$) results in:

$$Dir = d_T \cdot (n_a + \gamma\, n_d) \tag{23}$$

From eq. 23, it follows that:

$$d_T = \frac{Dir}{n_a + \gamma\, n_d} \tag{24}$$

To derive the expression of d_T as a function of physical parameters, consider the case of 100% FRET molecule. In this case, knowing that $n_d = 0$ and recalling the expression of n_a from eq. 4, we obtain:

$$d_T = \frac{\sigma^A_{D_{ex}}}{\sigma^D_{D_{ex}}} \tag{25}$$

Noting that, for $E < 1$ the "corrected total signal" $n_a + \gamma\, n_d$ (e.g. the corrected burst size in freely-diffusing measurements) will not change when γ is constant. Therefore the previous expression is valid for any E.

In ALEX measurements is easier to estimate d_{AA} from the data. Therefore, expressing d_T as a function of d_{AA} allows to easily estimate the former coefficient from the data. From the definitions of eq. 22, 25 and 7 we obtain:

$$d_T = \beta\, d_{AA} \tag{26}$$

This relation follows from the definition of β reported in the previous section and originally defined in [1].

Definition 3

Defining Dir as a function of the "corrected total signal" as defined in eq. 9 ($n_a/\gamma + n_d$) we have:

$$Dir = d_{T'} \cdot (n_a/\gamma + n_d) \tag{27}$$

The coefficient $d_{T'}$ can be obtained from the d_T expression noting that we simply divide the "corrected total signal" by γ:

$$d_{T'} = \frac{\sigma^A_{D_{ex}}}{\sigma^A_{D_{ex}}}\, \gamma = d_T\, \gamma \tag{28}$$

The coefficient $d_{T'}$ is indicated as d' in [1] (main text p. 2943 and SI). Note that the definition of d' given in eq. (27) of [1] has been derived for a $E = 0$ population (for

bioRxiv preprint doi: https://doi.org/10.1101/083287; this version posted October 26, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

which $n_d + n_a/\gamma = n_d$). However, by using the corrected total signal, it is possible to use the same coefficient to express the Dir contribution for any FRET population (and independently from E).

Definition 4

Defining Dir as a function of n_d we have:

$$Dir = d_D \cdot n_d \tag{29}$$

The coefficient d_D is a function of E as well as the physical parameters. Taking the ratio of the physical definitions of Dir and n_d we obtain:

$$d_D = \frac{\sigma^A_{D_{ex}}}{\sigma^A_{D_{ex}}} \frac{\gamma}{1-E} = d_T \frac{\gamma}{1-E} \tag{30}$$

Definition 5

Defining Dir as a function of n_a:

$$Dir = d_A \cdot n_a \tag{31}$$

The coefficient d_A is a function of E as well as the physical parameters. Taking the ratio of the physical definitions of Dir and n_a we obtain:

$$d_A = \frac{\sigma^A_{D_{ex}}}{\sigma^D_{D_{ex}}} \frac{1}{E} \tag{32}$$

2.4 Discussion of Definitions 1-5

Definitions 4 and 5 are inconvenient because the coefficient depends on E. Definition 3 does not depend on E but depends on γ, while Definition 2 depends only on the ratio of two absorption cross sections and is therefore the most general form. Definition 1 can only be used in an ALEX measurement but it is easy to fit from the S value of the A-only population.

So, for non-ALEX measurements, Definition 2 (d_T) gives the simplest and most general coefficient. It can be computed from datasheet values or from d_{AA} estimated from an ALEX measurement using the same dyes pair and D-excitation wavelength ($d_T = \beta\, d_{AA}$).

As physical interpretation, definitions 2 and 3 are similar. In Definition 2, when $E = 1$, the "corrected total signal" is n_a. When $E < 1$, the "corrected total signal" does not change (at the same excitation intensity, and fixed γ) being the sum of acceptor and γ-corrected donor counts. Similar considerations hold for Definition 3 (starting from $E = 0$). Note that using eq. 25 to estimate Dir requires the knowledge of the corrected total signal of eq. 8 (including A-direct excitation correction). For practical purposes, using a signal only corrected for γ and leakage to compute Dir via eq. 25, is a very good approximation. Alternatively, using eq. 33 (see next section) it is possible to compute corrected E values without any approximation.

bioRxiv preprint doi: https://doi.org/10.1101/083287; this version posted October 26, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

2.5 Correction formulas

We can expressing E as a function of E_R and the three correction factors as follows:

$$\begin{aligned} E &= f(E_R) = \\ &= \frac{E_R\,(L_k + d_T\,\gamma + 1) - L_k - d_T\,\gamma}{E_R\,(L_k - \gamma + 1) - L_k + \gamma} \end{aligned} \tag{33}$$

This expression is the same of eq. S9 in [1] when we replace $d_T\,\gamma$ with d'.

Similarly we can express S as a function of S_R, but in this case the expression will also depend on E_R in addition of the correction parameters:

$$\begin{aligned} S_\gamma &= f(E_R, S_R) = \\ &= \frac{S_R\,(E_R L_k - E_R\gamma + E_R - L_k + \gamma)}{E_R L_k S_R - E_R S_R \gamma + E_R S_R - L_k S_R - S_R d_T + S_R \gamma - S_R + d_T + 1} \end{aligned} \tag{34}$$

A similar formula has been reported in [1] (SI) expressing S as a function of S_{PR} and E_{PR}. Here the expression is simply expanded as a function of E_R and S_R, resulting in an explicit dependence on lk and d_T. The derivation of these formulas only involves using algebraic manipulations of the E and S expression. To avoid trivial errors, these expression have been derived with computer-assisted algebra (CAS). We also provide text-based version of the formula (python syntax) that is tested and easy to copy and paste in most other text-based language. For derivation details see Appendix: Derivation of the formulas.

3 Conclusion

We have introduced five definitions of acceptor direct excitation as a function of different experimental observable, and discussed that out of the five, two have the most useful in practice. In particular, eq. 25 can be used to correct for A-direct excitation even in single-laser measurements provided the coefficient d_T can be estimated independently. Furthermore, eq. 33 and 34 allows to apply corrections to E and S values, only knowing the raw E and S and the correction factors. With eq. 33 and 34 it is possible to correct the fitted E or S values as a last independent step of the analysis, without the need to modify (i.e. correct) the distributions prior fitting. This is important because, from a statistical point of view, the fit of the raw E and S peaks can provide more reliable estimates due to simpler modeling (e.g. using a Binomial distribution) which requires less assumptions. For example, methods such as shot-noise [4] and probability distribution analysis [5,6] and Gopich-Szabo likelihood analysis [7,8] can be directly applied to raw FRET distributions. Conversely, applying these methods to the corrected FRET distributions requires unnecessary complex statistical models which include the effect of each correction factor. In practice, the benefit of a more complex model is dwarfed by the inaccuracies arising from the additional approximations (even implicit) and from reliance on estimated correction parameters in the

bioRxiv preprint doi: https://doi.org/10.1101/083287; this version posted October 26, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

model itself. Using eq. 33 and 34, instead, allow to decouple the correction of E and S values from the population-level statistical modeling, resulting in more robust models and more accurate estimates.

Acknowledgments

We thank Dr. Eitan Lerner for a critical reading of the manuscript. Support provided by NIH R01 GM095904.

References

[1] Nam Ki Lee, Achillefs N. Kapanidis, You Wang, Xavier Michalet, Jayanta Mukhopadhyay, Richard H. Ebright, and Shimon Weiss. Accurate FRET measurements within single diffusing biomolecules using alternating-laser excitation. *Biophysical Journal*, 88(4):2939–53, April 2005. doi:10.1529/biophysj.104.054114.

[2] Shimon Weiss. Fluorescence spectroscopy of single biomolecules. *Science (New York, N.Y.)*, 283(5408):1676–1683, March 1999. doi:10.1126/science.283.5408.1676.

[3] Joseph R. Lakowicz, editor. *Principles of Fluorescence Spectroscopy*. Springer US, Boston, MA, 2006. URL: http://link.springer.com/10.1007/978-0-387-46312-4.

[4] Eyal Nir, Xavier Michalet, Kambiz M Hamadani, Ted A. Laurence, Daniel Neuhauser, Yevgeniy Kovchegov, and Shimon Weiss. Shot-noise limited single-molecule FRET histograms: comparison between theory and experiments. *Journal of Physical Chemistry B*, 110(44):22103–24, November 2006. doi:10.1021/jp063483n.

[5] Matthew Antonik, Suren Felekyan, Alexander Gaiduk, and Claus A. M. Seidel. Separating Structural Heterogeneities from Stochastic Variations in Fluorescence Resonance Energy Transfer Distributions via Photon Distribution Analysis. *The Journal of Physical Chemistry B*, 110(13):6970–6978, April 2006. doi:10.1021/jp057257+.

[6] Stanislav Kalinin, Suren Felekyan, Matthew Antonik, and Claus A. M. Seidel. Probability Distribution Analysis of Single-Molecule Fluorescence Anisotropy and Resonance Energy Transfer. *The Journal of Physical Chemistry B*, 111(34):10253–10262, August 2007. doi:10.1021/jp072293p.

[7] Irina V Gopich and Attila Szabo. Decoding the pattern of photon colors in single-molecule FRET. *The Journal of Physical Chemistry B*, 113(31):10965–10973, 2009. doi:10.1021/jp903671p.

bioRxiv preprint doi: https://doi.org/10.1101/083287; this version posted October 26, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

[8] Irina V Gopich and Attila Szabo. Theory of Single-Molecule FRET Efficiency Histograms. In Tamiki Komatsuzaki, Masaru Kawakami, Satoshi Takahashi, Haw Yang, and Robert J. Silbey, editors, *Single-Molecule Biophysics: Experiment and Theory, Volume 146*, volume 146 of *Advances in Chemical Physics*, pages 245–297. John Wiley & Sons, Inc., Hoboken, NJ, USA, November 2011. URL: `http://doi.wiley.com/10.1002/9781118131374`.

Appendix: Derivation of the formulas

See Derivation of FRET and S correction formulas.