

METHODOLOGY ARTICLE

Open Access

A Bayesian method for inferring quantitative information from FRET data

Catherine A Lichten^{1,2} and Peter S Swain^{2*}

Abstract

Background: Understanding biological networks requires identifying their elementary protein interactions and establishing the timing and strength of those interactions. Fluorescence microscopy and Förster resonance energy transfer (FRET) have the potential to reveal such information because they allow molecular interactions to be monitored in living cells, but it is unclear how best to analyze FRET data. Existing techniques differ in assumptions, manipulations of data and the quantities they derive. To address this variation, we have developed a versatile Bayesian analysis based on clear assumptions and systematic statistics.

Results: Our algorithm infers values of the FRET efficiency and dissociation constant, K_d , between a pair of fluorescently tagged proteins. It gives a posterior probability distribution for these parameters, conveying more extensive information than single-value estimates can. The width and shape of the distribution reflects the reliability of the estimate and we used simulated data to determine how measurement noise, data quantity and fluorophore concentrations affect the inference. We are able to show why varying concentrations of donors and acceptors is necessary for estimating K_d . We further demonstrate that the inference improves if additional knowledge is available, for example of the FRET efficiency, which could be obtained from separate fluorescence lifetime measurements.

Conclusions: We present a general, systematic approach for extracting quantitative information on molecular interactions from FRET data. Our method yields both an estimate of the dissociation constant and the uncertainty associated with that estimate. The information produced by our algorithm can help design optimal experiments and is fundamental for developing mathematical models of biochemical networks.

Background

Proteins work together continuously in the cells of all living things, generating cascades of reactions that are vital for life. To fully understand each individual protein's task requires discovering the timing, location, and strength of its interactions. To acquire this detailed information, fluorescence microscopy methods are ideal because they can provide dynamic, single-cell data at high spatial resolution [1,2]. One fluorescence tool that enables researchers to observe protein interactions in living cells is Förster resonance energy transfer (FRET). FRET data has the potential to yield biochemical constants, which are critical for modeling biological

systems, but measuring protein interactions from FRET data requires careful quantitative analysis.

FRET is a physical process where a molecule in an excited energetic state (the donor) transfers energy to a nearby ground-state molecule (the acceptor). The chance that an excited donor will transfer its energy to an acceptor (known as the FRET efficiency, E_{fr}) depends on the distance between the donor and acceptor (which must be between 1 and 10 nm for FRET to occur), their relative orientation, and the extent to which the donor's fluorescence emission peak overlaps the excitation peak of the acceptor [3]. Optimal conditions for energy transfer occur when the distance between donor and acceptor is minimal, the molecules' electric dipoles are aligned, and the spectral overlap is significant.

When FRET is used to study protein interactions in living cells, the proteins under investigation are fused to fluorescent tags (often variants of the green fluorescent

* Correspondence: peter.swain@ed.ac.uk

²Centre for Systems Biology at Edinburgh, University of Edinburgh, Edinburgh EH9 3JD, UK

Full list of author information is available at the end of the article

protein) that act as the donors and acceptors. When the proteins interact, they bring the fluorescent tags together so that FRET may occur. FRET increases the number of photons emitted by acceptors and reduces both the number of photons emitted by donors and the donor's fluorescence lifetime. To observe these effects, the most common techniques for collecting FRET data include fluorescence lifetime imaging (FLIM) and using a fluorescence microscope or spectrofluorometer to record fluorescence intensity after exposing samples to light that mainly excites either donors or acceptors [4-6].

For FRET data to reveal information about the underlying protein interactions, complicating factors must be dealt with. One confounding issue is spectral contamination, which arises from the requirement that the donor and acceptor must have overlapping spectra for FRET to occur. Due to the overlap, excitation light intended to excite one fluorophore may also excite the other (spectral crosstalk), and, conversely, one fluorophore may emit photons in the emission range of the other (spectral bleed-through). Several FRET analysis methods address this issue by calculating a FRET index, which is the FRET signal corrected for contamination from spectral overlap and normalised by donor or acceptor concentrations [7-12]. Although they are straightforward to calculate, FRET indices often bear an indirect and nonlinear relationship to the underlying concentrations and strength of protein interactions [13,14]. Alternatively, the ratio of donors to acceptors and the apparent FRET efficiency, which is the product of the intrinsic FRET efficiency and the fraction of donors (or acceptors) in complex, can be calculated [15-19,14]. Such quantities can be measured, provided calibrations are carried out using constructs consisting of a donor linked to an acceptor [17,19], but again do not relate directly to the dissociation constant (K_d) of an interaction.

Another significant challenge is that both the K_d and the FRET efficiency, E_{fr} , affect the observed signals, and so neither can be determined independently of the other based on data from a single sample. E_{fr} has been estimated from separate acceptor photobleaching [14] and FLIM experiments [15], but these approaches have drawbacks. Acceptor photo-bleaching is slow, irreversible, often fails to bleach all acceptors, and yields the intrinsic FRET efficiency only when all donors are in complex with acceptors [4-6]. FLIM requires specialized equipment, is often slower than standard fluorescence imaging, and analysis of FLIM data is complicated by the multiexponential fluorescence decays of fluorescent proteins [20]. There is however an alternative to directly measuring E_{fr} : Many values of K_d and E_{fr} will be consistent with data from a single cell or sample, but, by taking a set of data from samples that contain varying

concentrations of donors and acceptors and analyzing it altogether, it is possible to find the unique values of K_d and E_{fr} that are consistent with the ensemble [16,21].

Given these different approaches, it is not always obvious which one should be applied in different situations and there is no consensus on the statistical analysis, with each method processing the data differently and most giving no procedure to test the reliability of any estimates. A general method for inferring the K_d along with the uncertainty of that inference is necessary for the standardization of quantitative FRET measurements, the design of informative experiments, and for providing *in vivo* parameters for developing models of protein networks [22-24].

Here we propose a systematic analysis method that explicitly includes models of the photophysics and underlying chemical interactions and of measurement noise. Building on a spectral model for FRET [25], we develop a Bayesian algorithm to infer both the interaction strength (in terms of the *in vivo* K_d) and the FRET efficiency. Applying our algorithm to simulated data, we gain insight into how both experimental design factors such as measurement noise, number of measurements, fluorophore concentrations and ratios, and prior information impact our estimate and its uncertainty.

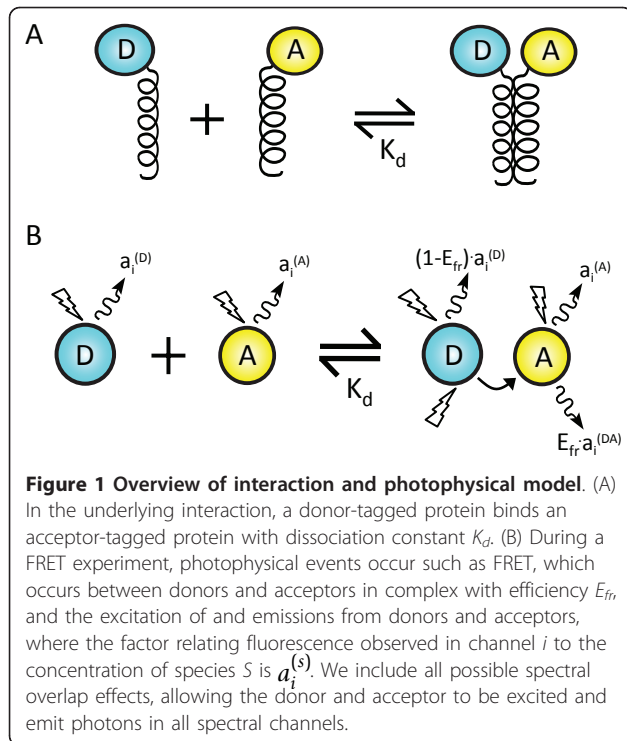
Results and Discussion

Algorithm

Overview of the mathematical model

FRET enables the study of molecular interactions in diverse settings. To design a widely applicable analysis technique, we consider a general system containing proteins (or other molecules) that form complexes and are labelled with fluorophores that act as FRET donors and acceptors (Figure 1A). We assume that all the molecules of interest are fluorescently tagged, so instead of referring to 'donor-tagged proteins' or 'acceptor-tagged proteins', we simply refer to 'donors' and 'acceptors'.

Our model relies on a few other assumptions. First, we assume that donors and acceptors may be free or form bimolecular, donor-acceptor complexes (i.e. $[D] + [A] \rightleftharpoons [DA]$) with dissociation constant $K_d = \frac{[D] \cdot [A]}{[DA]}$. Whether free or in complex, the fluorophores undergo fluorescence excitation and emission (Figure 1B). In complexes, the donor and acceptor are together and FRET may also occur. We assume that FRET due to random collisions between donors and acceptors occurs seldom enough to be neglected. Finally, we assume that donor-acceptor complexes all have the same FRET efficiency, which is denoted E_{fr} and refers to the fraction of instances where exciting a donor in complex leads to excitation of the acceptor.



For our general mathematical description of the FRET system, we use a previously described spectral mixing model [25] with slightly modified matrices. This model relates concentrations of fluorescent molecules to the fluorescence observed at various excitation wavelengths and ranges of emission wavelengths (spectral channels). It can be written as a matrix equation in the following form [6]:

$$M \times C = I_{obs}, \quad (1)$$

where C is a vector of the concentrations of fluorophores ($C = ([D], [A], [DA])^T$), I_{obs} is a vector of the fluorescence intensities observed for each spectral channel, and M is a matrix of the photophysical parameters that relate the concentrations of fluorophore to the observed intensities (Figure 1B).

The observed fluorescence intensities could represent data obtained from fluorescence microscopy, spectrophotometry or a flow cytometer. Microscopy images would require standard image processing steps for quantification, such as subtracting background fluorescence and defining regions of interest such as areas within the cytosol or nucleus for localized fluorescence [26].

Photophysics in the mathematical model

The matrix M contains all the relevant information about the photophysical processes we model. In addition to the expected direct excitation and emission, we include resonance energy transfer, cross-talk (unintended excitation) and bleed-through (acceptors emitting in the donors' typical emission range).

The spectral mixing framework can represent any number of spectral channels by increasing the dimensions of the matrix, but we focus on the particular case of three, which describes the most common 'three-cube' FRET experiments [12,7-9,14,18,17,21,16]. The first spectral channel, known as the donor channel, consists of an excitation wavelength that primarily excites donors and an emission filter that collects photons from the donor's emission range. The second channel, the acceptor channel, consists of an excitation wavelength that primarily excites acceptors and an emission filter that collects photons from the acceptor's emission range. The third channel, the FRET channel, combines the donor channel's excitation wavelength with the acceptor channel's emission filter, so that donors are excited preferentially and the emissions are filtered to primarily collect photons from acceptors. For three-cube FRET, the detailed form of Eq (1) is

$$\begin{bmatrix} a_1^{(D)} & a_1^{(A)} & f_1(E_{fr}) \\ a_2^{(D)} & a_2^{(A)} & f_2(E_{fr}) \\ a_2^{(D)} & a_3^{(A)} & f_3(E_{fr}) \end{bmatrix} \times \begin{bmatrix} [D] \\ [A] \\ [DA] \end{bmatrix} = \begin{bmatrix} I_D^{pred} \\ I_A^{pred} \\ I_F^{pred} \end{bmatrix}, \quad (2)$$

where $f_i(E_{fr}) = a_i^{(D)} \cdot (1 - E_{fr}) + a_i^{(A)} + a_i^{(DA)} \cdot E_{fr}$. The parameters $a_i^{(S)}$ each represent a combination of excitation and emission information [25]:

$$a_i^{(S)} = I_i \varepsilon_i^{(S)} Q^{(S)} S_i^{(S)}. \quad (3)$$

Here the subscript i refers to the spectral channel (for the three-cube case, it is 1 for the donor channel, 2 for the acceptor channel, and 3 for the FRET channel) and the superscript (S) refers to the fluorescent species (D , the donor; A , the acceptor; and DA , the complex). The variable I_i is the illumination intensity for the excitation wavelength of channel i ; $\varepsilon_i^{(S)}$ is the molar extinction coefficient of species (S) with excitation wavelength from channel i ; $Q^{(S)}$ is the quantum yield of species (S); and $S_i^{(S)}$ is the product of the emission of species (S) at the wavelengths of channel i and the sensitivity of the detector to the emitted photons. We describe a calibration procedure for obtaining these values in the Methods.

To illustrate Eq (2), consider the expression it yields for I_D^{pred} , the intensity in the donor channel. The intensity will be the sum of contributions from free donors, free acceptors, and donor- acceptor complexes. The contribution from free donors is $a_1^{(D)} \cdot [D]$ and the contribution from free acceptors is $a_1^{(A)} \cdot [A]$, which results from cross-talk and bleed-through. The complex, DA , can potentially produce contributions from donors not undergoing FRET ($a_1^{(D)} \cdot (1 - E_{fr}) \cdot [DA]$), from

acceptors ($a_1^{(A)} \cdot [DA]$) and from FRET ($a_1^{(DA)} \cdot E_{fr} \cdot [DA]$). Adding these contributions together, we obtain the following expression for the predicted intensity in the donor channel, I_D^{pred} ,

$$I_D^{\text{pred}} = a_1^{(D)} \cdot [D] + a_1^{(A)} \cdot [A] + \left(a_1^{(D)} \cdot [1 - E_{fr}] + a_1^{(A)} + a_1^{(DA)} \cdot E_{fr} \right) [DA]. \quad (4)$$

Continuing the matrix multiplication yields analogous expressions for I_A^{pred} and I_F^{pred} .

Molecular interactions in the mathematical model

To include K_d in our model, we model the chemical equilibrium, $[D] + [A] \rightleftharpoons [DA]$. Defining $[D_0] = [D] + [DA]$ and $[A_0] = [A] + [DA]$, and given that $K_d = \frac{[D][A]}{[DA]}$, we can compute the values of $[DA]$, $[D]$, and $[A]$ in terms of $[D_0]$, $[A_0]$, and K_d .

$$[DA] = \frac{1}{2} \left(([D_0] + [A_0] + K_d) - \sqrt{([D_0] + [A_0] + K_d)^2 - 4 \cdot [D_0] \cdot [A_0]} \right), \quad (5)$$

with $[D] = [D_0] - [DA]$ and $[A] = [A_0] - [DA]$.

Bayesian inference

Our aim is to infer the values of K_d and E_{fr} from the data. To this end, Bayesian inference is an ideal tool. Given a model, it quantifies the knowledge gained from a set of experimental observations about the parameters of interest, including the uncertainty of that knowledge [27]. That is, it allows us to update the prior probability distribution, $P(K_d, E_{fr})$, which represents what is known about the parameters before data has been acquired, to the posterior probability distribution, $P(K_d, E_{fr} | \text{data})$, which reflects both the prior information and what is learned from the data. According to Bayes' theorem,

$$P(K_d, E_{fr} | \text{data}) \propto P(\text{data} | K_d, E_{fr}) P(K_d, E_{fr}). \quad (6)$$

The likelihood function, $P(\text{data} | K_d, E_{fr})$, can be explicitly expressed through our photophysical model and a model of measurement noise. It quantifies the likelihood of different values for K_d and E_{fr} given the current data. We assume that the data follows a Gaussian distribution centred around the predicted intensity given by Eq (2). For three measurements, one from each of the three spectral channels, the likelihood is:

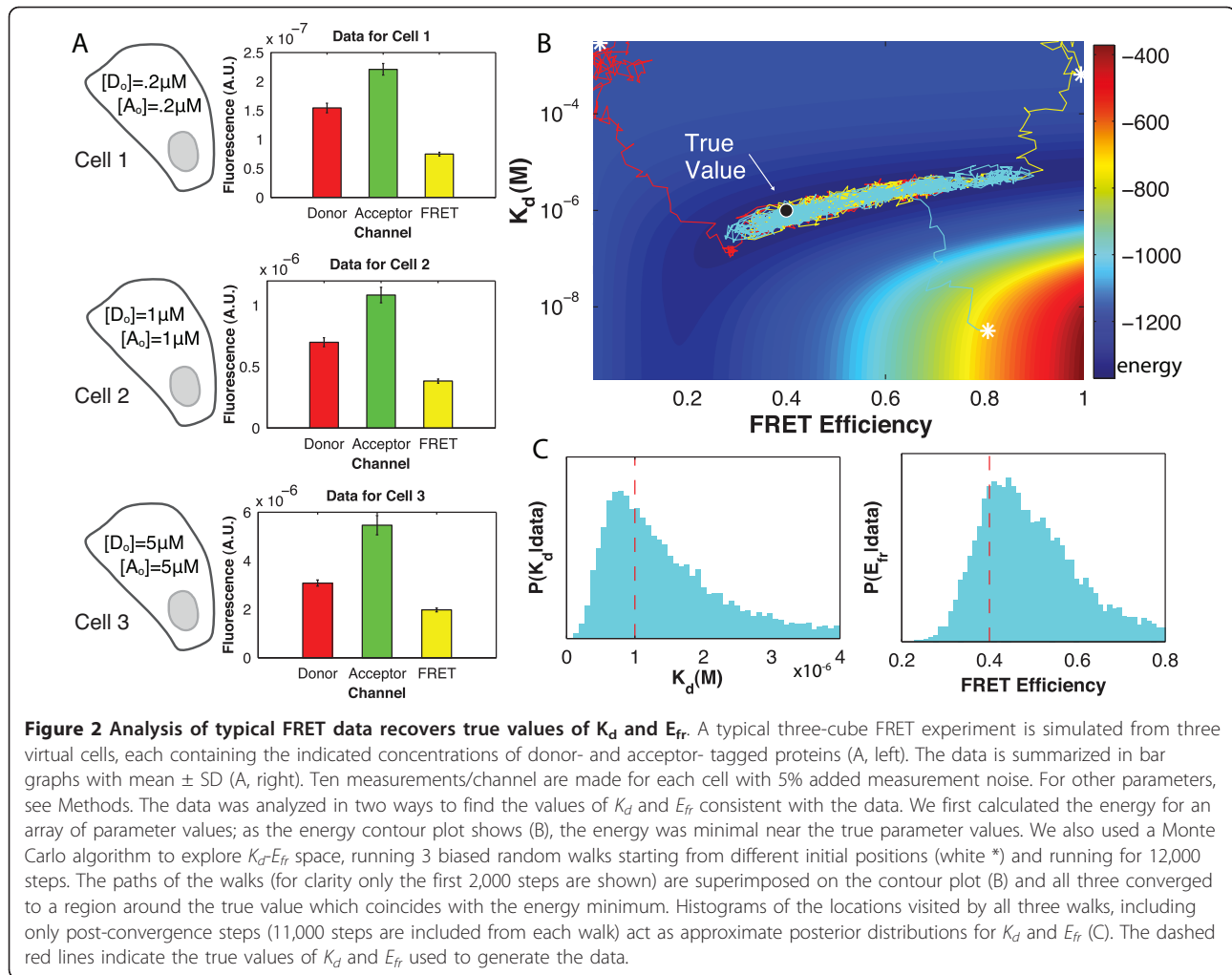
$$P(\text{data} | K_d, E_{fr}) \propto \frac{1}{\sigma_D \sigma_A \sigma_F} e^{-\frac{(I_D^{\text{pred}} - I_D)^2}{2\sigma_D^2} - \frac{(I_A^{\text{pred}} - I_A)^2}{2\sigma_A^2} - \frac{(I_F^{\text{pred}} - I_F)^2}{2\sigma_F^2}}. \quad (7)$$

The observed intensities for each channel are I_D , I_A , and I_F . I_D^{pred} , I_A^{pred} and I_F^{pred} denote the predicted intensities and σ_D , σ_A and σ_F measure the magnitude of the measurement noise in each channel and are approximated by the variances of the data in that channel (Methods). For fitting, we use a Gaussian approximation of Eq (7) based around the values of D_0 and A_0 that maximize the likelihood (see section on marginalization of D_0 and A_0 in the Methods). Combining the likelihood and prior probability distribution allows us to calculate the posterior probability distribution and determine the most probable values for our parameters of interest.

Illustration of method

To illustrate our method, we simulated data from three cells (Figure 2A, left), although in practice, data could come instead from cellular compartments or other sub-cellular regions of interest. For simplicity, each cell contains equal concentrations of donor ($[D_0]$) and acceptor ($[A_0]$). In the first cell, $[D_0] = [A_0] = 0.2 \mu\text{M}$; in the second, $[D_0] = [A_0] = 1 \mu\text{M}$; and in the third, $[D_0] = [A_0] = 5 \mu\text{M}$. Other parameter values are given in the Methods. We then simulated three-cube FRET measurements, generating data for the donor, acceptor, and FRET channels (Figure 2A, right). We made a series of ten measurements for each channel and each cell, to which we added Gaussian noise so that the standard deviation of the measurements was around 5% of the mean signal strength. We define r as the ratio of the standard deviation of the measurements to the mean signal strength; in this case, $r \approx 0.05$.

To obtain the posterior probability distribution that corresponds to this data for two parameters of interest, K_d and E_{fr} , we used two methods, the results of which should agree. First, because we are restricting our analysis to two parameters, we can visualize the distribution by computing the energy, the negative logarithm of the posterior probability, over a grid of values of K_d and E_{fr} . The resulting energy surface, shown in Figure 2B, was minimal around the true values of the parameters. If we were to extend our analysis to more than two parameters, including uncertainty in the $a_i^{(S)}$ constants for instance, we would use a Markov chain Monte Carlo (MCMC) algorithm to sample from the parameter space. To show that the MCMC algorithm gives results consistent with the energy surface, we used the algorithm to sample the posterior probability distribution for K_d and E_{fr} . We ran the algorithm three times, starting each time at different initial values. All three walks converged on the same small region of equally probable values: they are superimposed on the energy surface in Figure 2B. Both methods therefore recover the true values used to generate the data as optimal parameter



values. The approximate posterior probability distributions for K_d and E_{fr} are shown in Figure 2C. The distributions are estimated as the histograms of the three walks, excluding the burn-in period before convergence [28]. They peak near the parameter values used to simulate the data, indicating that the highly probable values for K_d and E_{fr} predicted by the methods are accurate.

Inferring other quantities from FRET data

To further illustrate the versatility of our method, we show that we can also use our model and Bayesian framework to estimate other parameters. Two instrument-independent parameters that have been a focus of interest are the apparent FRET efficiency, $E_d = E_{fr} \cdot \frac{[DA]}{[D_0]}$,

and the ratio, $r_{da} = \frac{[D_0]}{[A_0]}$ [15-19,14]. The analogous

apparent FRET efficiency for the acceptor, $E_a = E_{fr} \cdot \frac{[DA]}{[A_0]}$, is the product of E_d and r_{da} . We can estimate these quantities using our method, provided

that a calibration has been carried out with cells expressing only donors, only acceptors, and only a donor-acceptor construct. The FRET efficiency of the construct need not be known. Supposing we have calibrated the parameters $a_i^{(S)}$ of Eq 2 relative to $a_2^{(A)}$, which is taken to be unity, then we can express Eq 2 just in terms of $[A_0]$ with $[D_0] = r_{da}[A_0]$ and $E_{fr}[DA] = E_d r_{da}[A_0]$. We can analytically maximize the posterior probability of r_{da} and E_d as a function of $[A_0]$ by setting the derivative of the probability with respect to $[A_0]$ equal to zero and solving for this optimal $[A_0]$ in terms of r_{da} and E_d . Our MCMC algorithm can be used to sample values of r_{da} and E_d that are consistent with the data.

Testing

Algorithm reflects data quality

An algorithm for estimating parameter values should report not just an estimate of the most probable parameter values but also the reliability of that estimate. To evaluate this aspect of our algorithm, we measured

both the error and the uncertainty of its output. We defined the error as the discrepancy between the mean of the posterior probability distribution and the true value. We calculated error as $\log_2 \left(\frac{\text{Mean value}}{\text{True value}} \right)$, meaning that a perfect estimate would have zero error. Because under experimental conditions we would not know the true value, we also calculated the uncertainty of the estimate as an indicator of the reliability even when true values are unknown. The uncertainty corresponds to the width of the posterior probability distributions for K_d and E_{fr} . To prevent the magnitude of the parameter of interest from skewing the uncertainty, we use the relative uncertainty, calculated as the coefficient of variation (the standard deviation divided by the mean).

We first tested our algorithm in the presence of varying levels of measurement noise. As the examples in Figure 3A suggest, when measurement noise increased, the optimal regions in K_d and E_{fr} -space grew wider and longer and the peak of the posterior probability distribution for K_d often deviated from the true value (see Figure 3A inset). In Figures 3C and 3E, the error (upper plot) and uncertainty (lower plot) for estimating K_d are shown and confirm that both the error and uncertainty of the estimate increased with increasing noise. Similar results were obtained for the estimate of E_{fr} (data not shown).

Next, we verified that gathering more data would improve the quality of the estimates of K_d and E_{fr} . Such data could be obtained by, for example, collecting several images of the same sample or dividing a region of

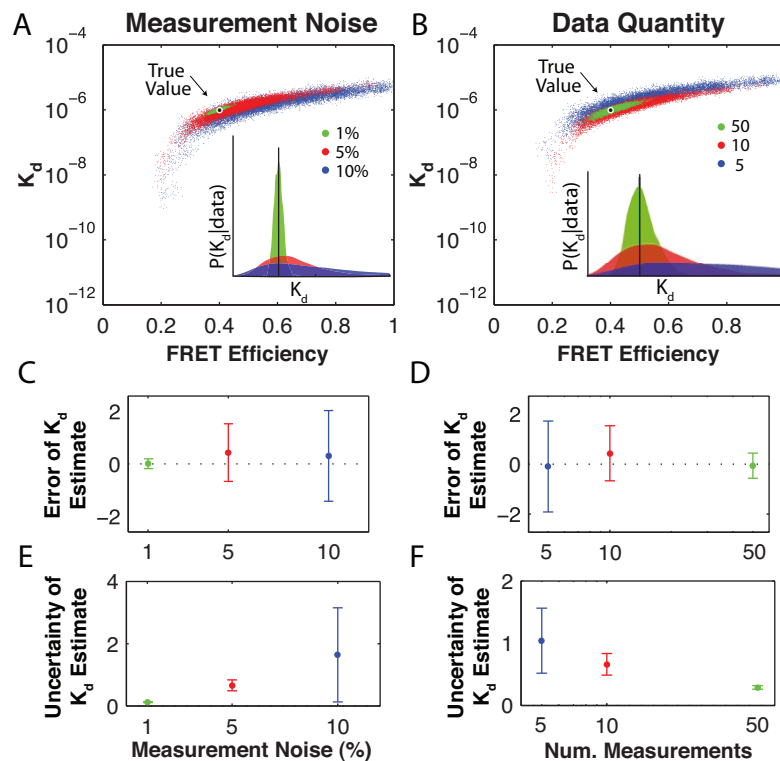


Figure 3 K_d estimates reflect data quality and quantity. We simulated and analyzed data with varying levels of measurement noise and numbers of measurements/cell/channel. The locations visited by MCMC walks (dots) for the three noise levels (A) and the three numbers of measurements/cell/channel (B) show that the highly probable region grows as measurement noise increases and as the number of measurements decreases. The 'True Value' (black and white spot) indicates values used to generate the data. Histograms of the locations visited by the walks (insets, histograms smoothed for clarity) approximate the corresponding posterior probability distributions for K_d , with true values indicated by black lines. The plots of error ($\log_2 \frac{\text{mean}}{\text{true value}}$) vs. measurement noise (C) and error vs. amount of data (D) illustrate that, in general, accuracy decreases with increasing noise and decreasing number of measurements although the mean of the error remains centered at zero. Even when the true value is unknown, the relative uncertainty of the parameter estimate (coefficient of variation of variation of locations visited by a walk) is measurable; it also grows with increasing noise (E) and decreasing number of measurements/cell/channel (F). In (C-F), error bars are mean \pm SD. 50 data sets were analyzed for each noise level or number of measurements, and each dataset was analyzed once with a random walk running for 20,000 steps, starting once the walk converged. Except when otherwise indicated, data had 10 measurements/cell/channel and 5% added Gaussian noise. For other parameters, see Methods.

interest into subregions. The examples in Figure 3B show that with more measurements, the optimal region became smaller and the peak of the posterior probability distribution for K_d moved closer to the true value (see Figure 3B inset). Figures 3D and 3F show, as expected, that increasing the number of measurements decreases the error and uncertainty for the K_d estimate. The effect of data quantity on the error and uncertainty of the E_{fr} estimate was similar (data not shown).

Overall, our algorithm reliably reflects the quality of the data analyzed in the uncertainty it gives of its estimate and accurately infers the values of K_d and E_{fr} for 10 measurements per cell per channel with 10% measurement noise. We also note that in Figures 3C and 3D, the mean error in K_d remained close to zero, indicating that neither measurement noise nor number of measurements systematically biased the location of the peak.

The total amount of donor, $[D_0]$, and the total amount of acceptor, $[A_0]$, affect inference of K_d and E_{fr}

Inferring unique values of K_d and E_{fr} requires variations in $[D_0]$ and $[A_0]$ A challenge in analyzing FRET

data is that both E_{fr} and K_d affect the extent of FRET that may occur, making it difficult to tease out the contribution of each parameter. In general, for data from a single pair of donor and acceptor concentrations, a weak E_{fr} and strong K_d can give rise to the same observations as a strong E_{fr} and weak K_d . However, because our algorithm can analyze data from multiple cells at once, it can overcome this challenge.

To demonstrate the problem, we analyzed a set of simulated data from a single sample with $[A_0] = [D_0] = 1 \mu M$. As Figure 4A shows, the resulting region of optimal values had an elongated shape, indicating that many equally probable solutions exist for $\{K_d, E_{fr}\}$ for that dataset. To show that collecting more data from similar cells does not help, we repeated the experiment, simulating data from 3 cells with the same concentrations as the first. As Figure 4B shows, including this data made the region slightly narrower but did not change its elongated shape.

The difficulty in determining K_d without knowing E_{fr} or vice versa can be overcome by analyzing data from

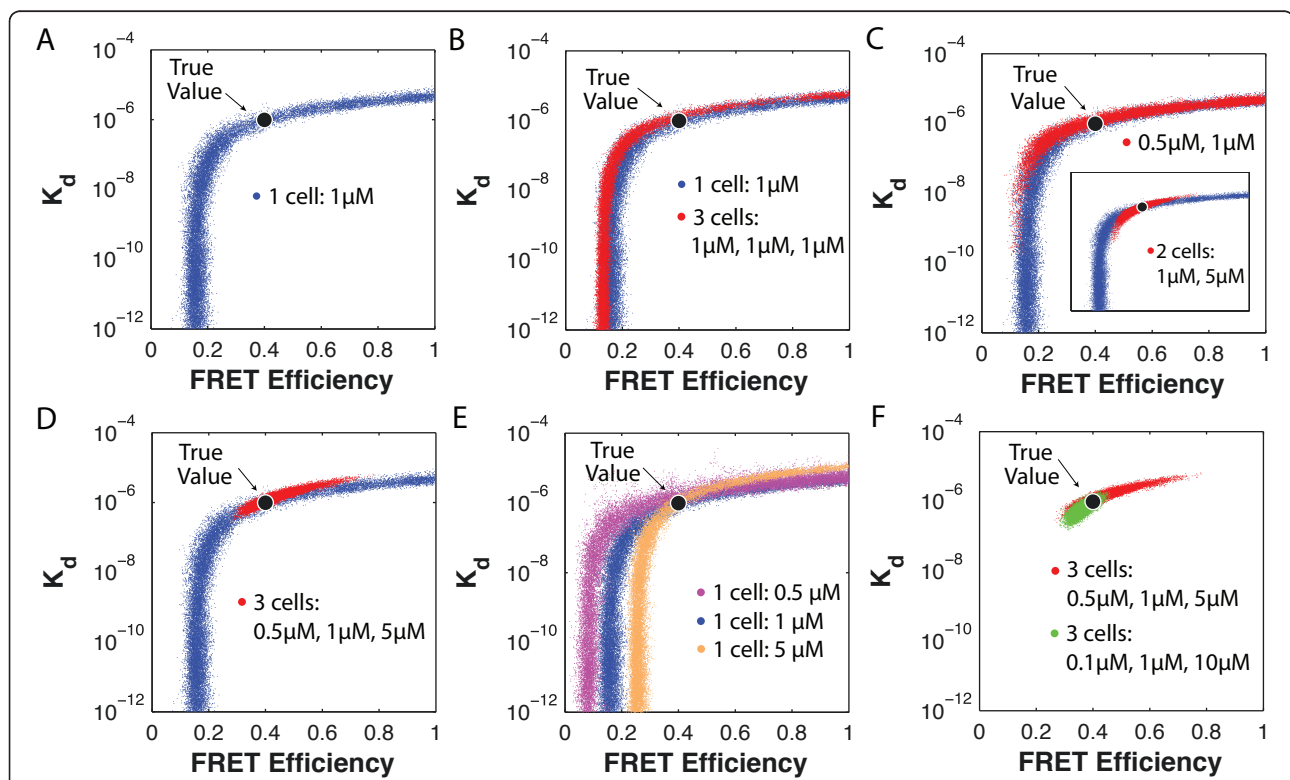


Figure 4 Concentration variations determine shape of region of high probability. Using the Monte Carlo algorithm to analyze data from a single cell containing $[D_0] = [A_0] = 1 \mu M$ produced an elongated region of highly probable values (A), indicated by the collection of dots showing the locations in (K_d, E_{fr}) visited by the biased random walk. These single-cell results are shown again in (B-E) (blue dots), with the results from analyzing other data sets superimposed. Analyzing three identical cells instead of one produced a narrower but still elongated region (B). However, when analyzing data from two cells (C and C inset) and three cells (D) with different concentrations of fluorophores, the resulting regions were contracted. Analyzing cells individually showed that the elongated highly probable regions for each all intersected near the true value (E). More extreme variation in concentrations led to an even smaller optimal region (F). In (A-F), there were 10 measurements/cell/channel, 3% added Gaussian noise, and 31,000 steps are shown for each walk. For other parameter values, see Methods.

samples containing varying fluorophore concentrations [16,21]. To quantitatively justify how varying concentrations facilitates parameter fitting, we simulated data from three cells containing different concentrations of donors and acceptors. As Figure 4C shows, analyzing data from two different cells decreased the size of the region of equally probable values. Combining data from a cell containing $[D_0] = [A_0] = 1 \mu\text{M}$ with data from a cell containing $0.5 \mu\text{M}$ eliminated high values of K_d (Figure 4C); combining data from a $1 \mu\text{M}$ cell and a $5 \mu\text{M}$ cell eliminated low values of K_d (Figure 4C, inset). As Figure 4D indicates, analyzing data from all three

cells ($0.5 \mu\text{M}$, $1 \mu\text{M}$, and $5 \mu\text{M}$) together eliminated higher and lower K_d values, leaving a small region of equally probable values that included the true values. To show why varying fluorophore concentrations is effective, we analyzed data from each of the three cells in Figure 4D individually. The optimal regions for each cell are plotted in Figure 4E and clearly intersect at the true value. Varying the concentrations more broadly can further shrink the solution region (Figure 4F).

Inference is limited if $[D_0] = [A_0] \gg K_d$ or $[D_0] = [A_0] \ll K_d$ Plots of complex formation vs. K_d (Figure 5A, insets) show that only K_d values within a limited range

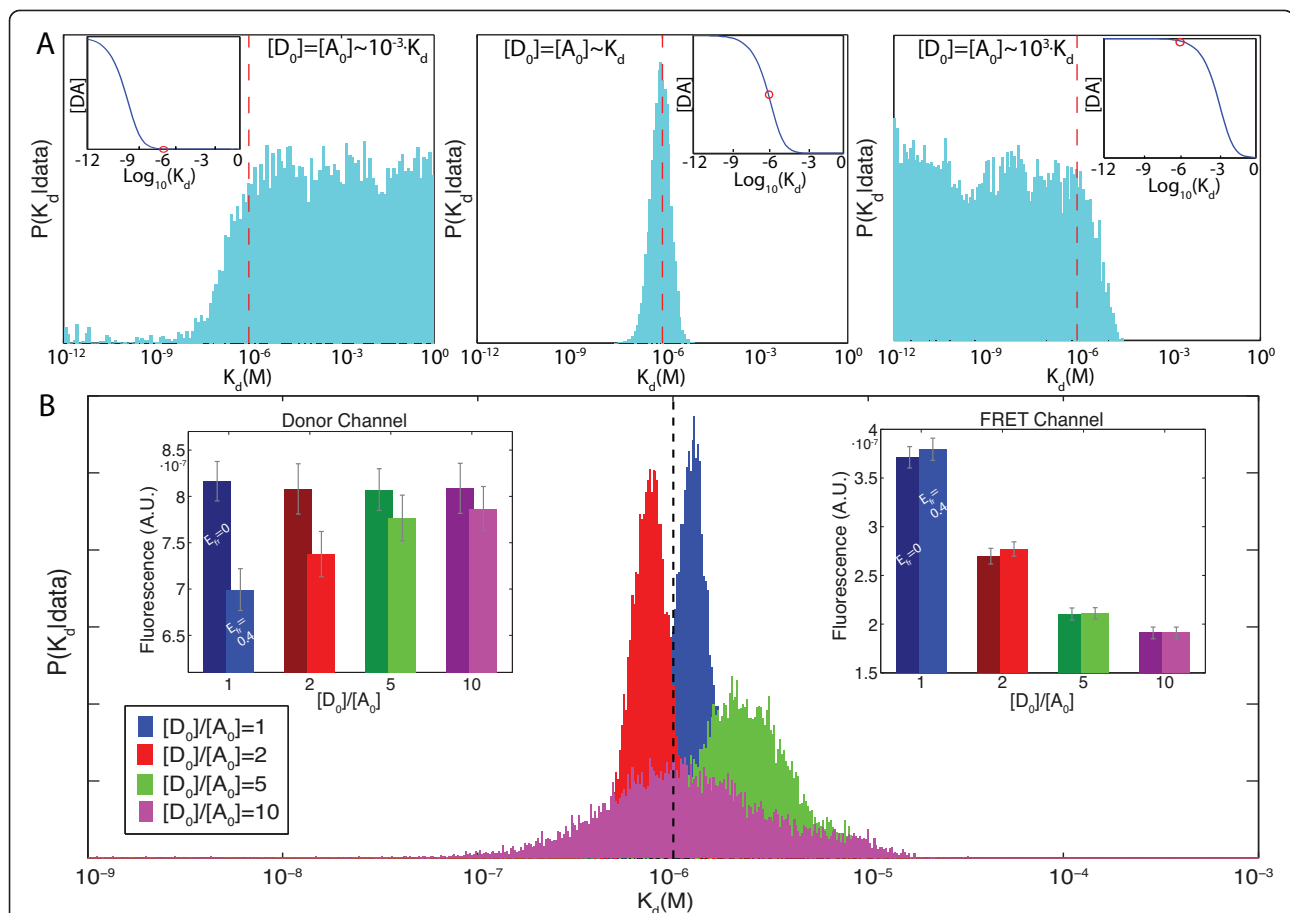


Figure 5 Gaining insight into optimal experimental design. The approximate posterior probability distributions for K_d (A) have different shapes if the data analyzed was simulated from three cells containing equal concentrations of donors and acceptors which are much lower than (left), higher than (right) or about equal to K_d (centre). For the data analyzed in the left panel, for instance, the cells contained the concentrations $[D_0] = [A_0] = 0.2 \cdot 10^{-3} \mu\text{M}$, $[D_0] = [A_0] = 1 \cdot 10^{-3} \mu\text{M}$, and $[D_0] = [A_0] = 5 \cdot 10^{-3} \mu\text{M}$. Insets show amount of complex formed as a function of K_d for the indicated concentrations, demonstrating that where complex formation is insensitive to K_d corresponds to plateaus in the posterior probability distributions for K_d . In each plot (or inset), a vertical dashed line (or red circle) indicates 10^{-6}M , the true value of K_d . (A) had 36,000 steps/walk and 5% added noise. Exploring another aspect of fluorophore concentrations, increasing the ratio $[D_0] : [A_0]$ increases the uncertainty in fitting K_d (B). As the ratio was increased (by keeping $[D_0]$ constant for the three cells at $0.2 \cdot 10^{-6}\text{M}$, $1.0 \cdot 10^{-6}\text{M}$ and $5.0 \cdot 10^{-6}\text{M}$ while decreasing $[A_0]$ according to the ratio), posterior probability distributions for K_d broadened (true value indicated by dashed vertical line). Insets show data used for fitting (bars marked ' $E_{fr} = 0.4$ ') from the donor channel (left) and FRET channel (right) contrasted with data from the same cells simulated with $E_{fr} = 0$, demonstrating that the relative contribution of FRET decreases as $[D_0] : [A_0]$ increases. (B) had 50 measurements/cell/channel, 36,000 steps/walk and 3% added noise. Bars show mean \pm SD. For other parameters, see Methods.

significantly affect complex formation: outside this range, further changes in K_d have little effect. For instance, when $[D_0] = [A_0] \approx K_d$, small changes in K_d in either direction affect $[DA]$ (Figure 5A centre, inset). However, when $[D_0] = [A_0] \ll K_d$ (Figure 5A, left inset) or $[D_0] = [A_0] \gg K_d$ (Figure 5A, right inset), only changes to K_d in one direction affect the amount of $[DA]$. When a range of K_d values are consistent with the data, K_d cannot be fit to a single value.

By finding the posterior probability distribution of K_d , our algorithm predicts when a range of K_d values will be consistent with the data. First, we simulated data with equal concentrations of acceptors and donors that were very low or very high relative to K_d . As shown in Figure 5A, the approximate posterior probability distributions of K_d for data simulated with relatively low or high concentrations form plateaus instead of single peaks. The plateaus indicate that a range of K_d values is equally probable and consistent with the data. The posterior distribution obtained for $[D_0] = [A_0] \ll K_d$ reveals that K_d could be any number greater than approximately $10^{-6}M$; the distribution for $[D_0] = [A_0] \gg K_d$ reveals that $K_d < 10^{-6}M$.

This analysis illustrates how posterior probability distributions for the parameters of interest can be more informative than single values. For example, information from posterior distributions can be useful for improving experimental design. If one were to obtain plateaued distributions like the two shown in Figure 5A, one could recognize that the fluorophore concentrations were too low or high relative to K_d and devise a more informative experiment.

The ratio $[D_0]:[A_0]$ affects inference We also used our algorithm to explore the effects of varying the ratio, $[D_0] : [A_0]$, on the algorithm's ability to infer K_d (Figure 5B) and how these effects arise. While other authors have reported that FRET data becomes increasingly unreliable as the ratio $[D_0] : [A_0]$ deviates from unity [13], they measured how noise propagates in FRET formulas and did not directly address how the ratio impacts the measurement of interaction strength. As Figure 5B shows, the width of the approximate posterior distributions for K_d broaden as the ratio increases, meaning that the uncertainty in the estimate increases. The peaks of the distributions however remain close to the true values. The same effect occurs when the ratio decreases (data not shown).

The insets in Figure 5 illustrate why the ratio affects the uncertainty, showing data from the donor and FRET channels in the presence and absence of FRET. Here, FRET contributes only a fraction of the observed signals. This small contribution is particularly evident in the FRET channel, where the change in the signal due to FRET is small compared to the measurement error:

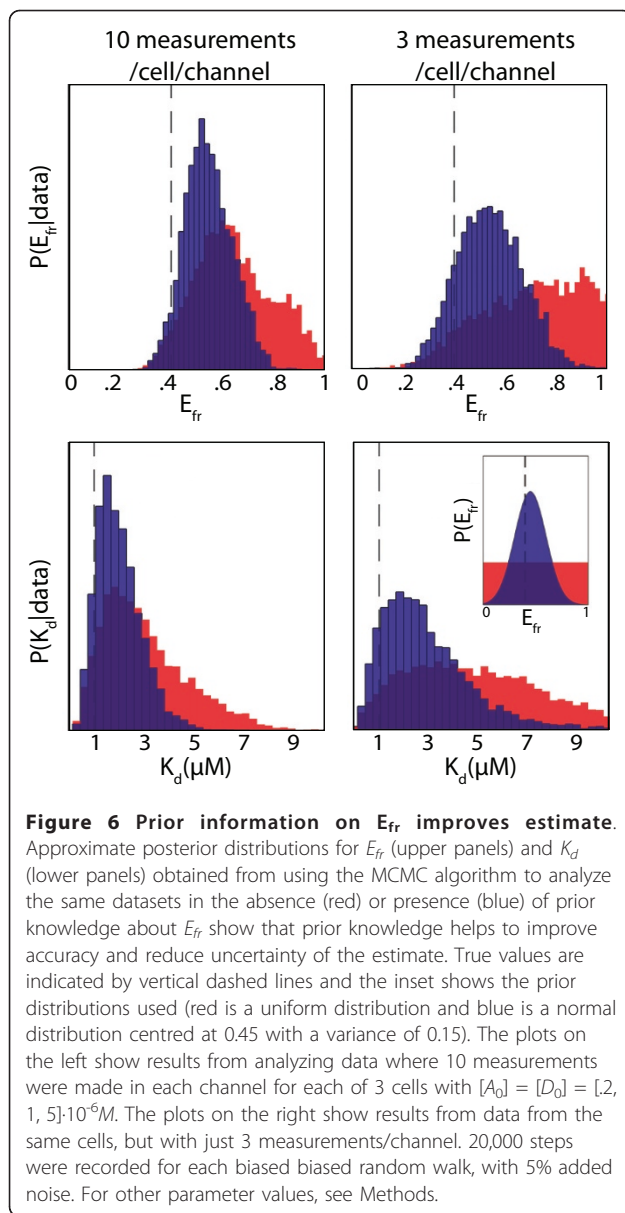
measurement error is 3% and the increase in the FRET channel from FRET is only 2.9% for $[D_0] : [A_0] = 1$ and drops to 0.7% for $[D_0] : [A_0] = 10$ (FRET is included by changing the FRET efficiency from 0 to 0.4). As the ratio increased, the relative shortage of acceptors meant that fewer complexes could form, making the contributions from FRET smaller and causing the uncertainty of the estimate to grow. In the absence of measurement noise, however, our algorithm can estimate K_d even from very extreme ratios. These data also demonstrate that with our method, we can make inferences about K_d even when FRET causes perceptible changes in only one channel (e.g. the donor channel here).

Knowing how spectral overlap and the ratio of donors to acceptors affect the inference of protein interaction strength helps make it possible to design informative experiments. For example, we find that in each channel, the difference generated by FRET depends on spectral overlap. In the donor channel, where FRET decreases the signal, the relative change is $(a_1^{(DA)} - a_1^{(D)}) / (a_1^{(A)}A_0 + a_1^{(D)}D_0)[DA]E_{fr}$, meaning that the observed change depends on the difference between acceptor bleed-through, $a_1^{(DA)}$, and donor fluorescence, $a_1^{(D)}$. Analogous changes occur in the other channels. When $[D_0] = [A_0]$, the change in the donor channel is proportional to $(a_1^{(DA)} - a_1^{(D)}) / (a_1^{(A)} + a_1^{(D)})$, the change in the acceptor channel to $(a_2^{(DA)} - a_2^{(D)}) / (a_2^{(A)} + a_2^{(D)})$ and the change in the FRET channel to $(a_3^{(DA)} - a_3^{(D)}) / (a_3^{(A)} + a_3^{(D)})$. In each channel, increasing the difference between $a_i^{(DA)}$ and $a_i^{(D)}$ would make experiments more informative.

Prior information improves estimate

A further advantage of our method is its flexibility. The Bayesian framework makes it possible to incorporate additional details we know about the system as prior information, allowing us to more accurately represent the system being analyzed and potentially improve our estimate of the parameters of interest. To exploit this feature, we tested whether including additional information about the FRET efficiency would improve our inference of E_{fr} and K_d . Information about E_{fr} could be obtained, for example, from lifetime imaging experiments [29].

Figure 6 shows the approximate posterior probability distributions for E_{fr} and K_d that result from analyzing the same data in the presence and absence of prior information for E_{fr} . We included slightly inaccurate prior information with large uncertainty, supposing that E_{fr} was measured to be 0.45 (instead of the true value, $E_{fr} = 0.4$) with an uncertainty of 0.15. Even with this limited information, the resulting histograms for both



K_d and E_{fr} have taller, narrower peaks which are closer to the true values compared to the histograms produced by analyzing the same data in the absence of the prior information. As expected, this improvement was more substantial when we analyzed a set of data consisting of just 3 measurements per cell per channel (Figure 6, right column): prior information becomes more important the less relevant information there is in the data. As another form of prior information, we could also include error in the calibration measurements for the constants that relate molecular concentration to fluorescence in Eq (2). Incorporating this information would make it possible to monitor the impact of that error on the uncertainty of the final estimate of K_d .

Summary of testing

We have shown, using typical, simulated FRET data, that our algorithm accurately recovers the parameters of interest. It responded consistently and intuitively to changes in the amount of measurement noise present in the data and the quantity of data. We also used the posterior probability distributions we obtain for the parameters to gain insight into how the magnitude and variation of the donor and acceptor concentrations affect the ability of the algorithm to infer K_d . The algorithm gave informative results even when given data simulated with very high or low concentrations of donors and acceptors relative to the K_d although, as expected, with this data it could not converge on a unique pair of values for K_d and E_{fr} but gave upper or lower bounds. Gaining an understanding of how experimental parameters such as measurement noise, number of measurements, spectral overlap and fluorophore concentrations affect inference allows experimenters to design optimal experiments that yield information with less uncertainty [24].

Conclusions

Fluorescence microscopy and FRET open a window onto the cell, allowing us to observe protein interactions as the cell functions as a complete system. However, for protein interaction information from FRET data to be integrated into models and improve our understanding of biological systems, it must be reliably quantified, including the uncertainty in the estimates produced [27].

For this purpose, we have presented an algorithm for inferring the most probable values of the absolute or relative *in vivo* dissociation constant and the FRET efficiency from three-cube FRET data. Our algorithm produces estimates in terms of the posterior probability distribution for the parameters of interest. Posterior probability distributions yield more extensive information than a single value can provide, conveying the reliability of the estimates through the shape and width of the distribution and giving upper or lower bounds on parameters if the data is not more informative. Our method requires only basic three-cube FRET data but has the flexibility to use data from more spectral channels and incorporate other types of data, such as a separate measure of the FRET efficiency. We have focused on using our algorithm to fit E_{fr} and K_d , but it can also infer the apparent FRET efficiency and ratio of donors to acceptors, other quantities sometimes used as measures of FRET [15-19,14].

In the examples described here, we make a few assumptions, but these are not necessarily part of our methodology. First, we use molar extinction coefficients that must be measured separately or taken from the literature and assume that the literature measurements are

valid in the cellular environment. Second, we assume Gaussian measurement noise, but our model could be straightforwardly adapted to include log-normal or other types of noise. Finally, we have not taken into consideration photo-bleaching, incomplete labelling or dark states, but our model could readily be extended to include these factors.

We have used simulated data to illustrate principles that apply to real data, showing how, in practice, it is best to infer the dissociation constant from FRET data. Three-cube FRET experiments are a well-established experimental technique [3-6]. For our method, the calibration procedure we describe in the Methods for quantifying the $a_i^{(S)}$ constants of Eq (2) is also necessary. The procedure is similar to existing calibration methods. Measuring relative values of $a_i^{(D)}$ and $a_i^{(A)}$ is a standard part of FRET data collection, necessary for calibrating the extent of spectral bleed-through and cross-talk. We require calibration with a donor-acceptor construct to relate the brightness of the donor to that of the acceptor, which is necessary for obtaining the relative K_d . This calibration is not always a part of standard FRET data collection, but it has been used in several studies [8,18,17,4,14,19,21]. Alternatively, measuring the absolute brightness per molecule, which is necessary for measuring the absolute K_d , is more technically challenging but has been achieved [30,31].

We have demonstrated that to infer the values of K_d and E_{fr} , it is important that the FRET data analyzed come from samples containing varying concentrations of donor- and acceptor-tagged proteins. The precise amount of variation required is difficult to predict because it depends on the measurement noise present in the data, the number of measurements made, and the extent of spectral bleed-through and cross-talk. Naturally occurring variation in protein expression from cell to cell generated sufficient variability for estimating the FRET efficiency and relative dissociation constant [21], at least in transiently transfected mammalian cells.

While a number of *ad hoc* methods exist for quantifying protein interactions using FRET, our method contributes something new in that it makes plain the spectral information being used and the bio-chemical assumptions made about the system. Our model is general: it is not specific to three-cube data and could be used to analyze data with any number of spectral channels. It can be straight-forwardly adapted to other experimental situations, for instance to measure the dimerization affinity of a homodimeric protein. Our approach is a Bayesian analysis that reveals the uncertainty in estimated parameters and produces informative results for data from a wide-range of experimental set-ups. It focuses on determining what we can learn from experimental

observations about the parameters of interest given clear assumptions and a systematic, statistical analysis based only on those assumptions.

Methods

Calibration of fluorescence constants, $a_i^{(S)}$

The $a_i^{(S)}$ constants, which relate concentrations of single fluorophores to observed fluorescence intensity, can be obtained by calibration. Calibration of spectral cross-talk and bleed-through is a standard part of FRET quantification [4-6]. If the fluorophore concentrations can be determined [30,31], then the values for the three $a_i^{(D)}$ and three $a_i^{(A)}$ constants can be obtained from samples containing only donors or only acceptors. This calibration makes it possible to measure the absolute K_d . Three-cube data from donor-only and acceptor-only samples correspond to the following equations (from Eq (2)):

$$\begin{aligned} I_D^{(D),\text{pred}} &= a_1^{(D)} \cdot [D_0]; & I_A^{(D),\text{pred}} &= a_2^{(D)} \cdot [D_0]; \\ I_F^{(D),\text{pred}} &= a_3^{(D)} \cdot [D_0]; \end{aligned} \quad (8)$$

$$\begin{aligned} I_D^{(A),\text{pred}} &= a_1^{(A)} \cdot [A_0]; & I_A^{(A),\text{pred}} &= a_2^{(A)} \cdot [A_0]; \\ I_F^{(A),\text{pred}} &= a_3^{(A)} \cdot [A_0]. \end{aligned} \quad (9)$$

To infer the values of the constants $a_i^{(D)}$ and $a_i^{(A)}$ and their uncertainty (due to measurement error), we can sample from posterior distributions for $a_i^{(D)}$ and $a_i^{(A)}$ given the calibration data. We will illustrate this procedure for a more complicated example below.

Determining the $a_i^{(DA)}$, which relate a complex undergoing FRET to its fluorescence in channel i , is more complicated because the $a_i^{(DA)}$ depend on properties of both the donor and the acceptor. However, from Eq (3), $a_i^{(DA)} = I_i \varepsilon_i^{(D)} Q^{(A)} S_i^{(A)}$, but $a_i^{(A)} = I_i \varepsilon_i^{(A)} Q^{(A)} S_i^{(A)}$. Therefore, $a_i^{(DA)} = \frac{\varepsilon_i^{(D)}}{\varepsilon_i^{(A)}} \cdot a_i^{(A)}$, and so $a_i^{(DA)}$ can be obtained through knowledge of the ratio of molar extinction coefficients at the excitation wavelength for channel i . The values of the molar extinction coefficients, $\varepsilon^{(D)}$ and $\varepsilon^{(A)}$, may not always be available at these wavelengths, but they can be estimated from literature values of the molar extinction coefficients (usually measured at the fluorophore's excitation peak) and the excitation spectra of the donor and acceptor [25]. Interpreting the excitation spectra as the probability of the fluorophore becoming excited and assuming that the extinction coefficient is proportional to the probability of excitation, we can use the excitation spectra to rescale the literature value of the extinction

coefficient to the excitation wavelength of channel i . This estimate is valid provided that the molar extinction coefficients obtained from the literature are not significantly different from the molar extinction coefficients in the cellular environment of the experiment (or provided that the change in environment affects both donor and acceptor similarly).

When the relationship between brightness and concentration cannot be determined absolutely, relative values for $a_i^{(D)}$ and $a_i^{(A)}$ can be obtained from samples containing only donors and only acceptors, along with samples containing linked donor-acceptor constructs. The constructs consist of a donor and an acceptor fluorophore separated by a short linker of 5-10 amino acids and have been used for FRET calibration in several studies [8,18,17,4,14,19,21]. The construct's FRET efficiency, E_{fr}^c , need not be known and can be determined using the procedure we describe below.

The three-cube measurements on samples containing only donors or only acceptors would correspond to Eqs (8) and (9). Three-cube data obtained from samples containing donor-acceptor constructs would correspond to the following equations:

$$\begin{aligned}
 I_D^{(DA),\text{pred}} &= \\
 &\left(a_1^{(D)} [1 - E_{fr}^c] + a_1^{(A)} + \frac{\varepsilon_1^{(D)}}{\varepsilon_1^{(A)}} a_1^{(A)} E_{fr}^c \right) [DA], \\
 I_A^{(DA),\text{pred}} &= \\
 &\left(a_2^{(D)} [1 - E_{fr}^c] + a_2^{(A)} + \frac{\varepsilon_2^{(D)}}{\varepsilon_2^{(A)}} a_2^{(A)} E_{fr}^c \right) [DA], \\
 I_A^{(DA),\text{pred}} &= \\
 &\left(a_3^{(D)} [1 - E_{fr}^c] + a_3^{(A)} + \frac{\varepsilon_3^{(D)}}{\varepsilon_3^{(A)}} a_3^{(A)} E_{fr}^c \right) [DA],
 \end{aligned} \tag{10}$$

where we have replaced $a_i^{(DA)}$ by $\frac{\varepsilon_i^{(D)}}{\varepsilon_i^{(A)}} a_i^{(A)}$. To infer the values of the $a_i^{(S)}$ constants given calibration data consisting of three-cube measurements made on samples with only donors, only acceptors, and donor-acceptor complexes, we first define the general likelihood function of these experiments for a single sample:

$$P(\text{data} | a_i^{(S)}, E_{fr}^c) = \prod_{k=1}^3 \frac{1}{\sigma} \exp \left[\frac{-1}{2\sigma^2} \sum_{i=1}^n (I_k^{(i)} - I_k^{\text{pred}})^2 \right], \tag{11}$$

where $I_1^{(i)}$, $I_2^{(i)}$, and $I_3^{(i)}$ are the i^{th} measurements in the donor, acceptor, and FRET channels, respectively. For a

given sample, we assume that an equal number of measurements, n , will be made in each channel and σ quantifies the error of the measurements. The predicted intensities in channel k , I_k^{pred} , are given by Eqs (8), (9) and (10) and are a function of the constants $a_i^{(S)}$, the concentrations of the species (donors, acceptors, or donor-acceptor complexes), and also E_{fr}^c in the case of the donor-acceptor construct. The full likelihood for all the calibration data together would be the product of three instances of Eq (11), one specified for each three-cube experiment (donor-only, acceptor-only, or donor-acceptor complex).

We are interested in the values of the $a_i^{(S)}$ but indifferent to the concentrations of fluorophores and complexes in the samples. For this reason and because the concentrations of fluorophores and complexes will vary for different samples, it is useful to marginalize the calibration likelihood over $[A_0]$, $[D_0]$, and $[DA]$. We can also eliminate σ by marginalization [28], assuming the measurement error is the same for all three spectral channels. If this assumption does not hold, one can define σ_k for the measurement error in each channel and either approximate each σ_k as equal to the standard deviation of the measurements in the k^{th} channel or include the σ_k as parameters to be inferred in the procedure we describe below. Assuming a prior probability that only specifies positive values for the $a_i^{(S)}$ and E_{fr}^c , marginalizing Eq (11) over the concentrations and σ yields the posterior probability of E_{fr}^c and relative values of $a_i^{(S)}$:

$$\begin{aligned}
 P_{\text{calib}} \approx & \\
 & \left[\left\{ \left(a_1 \sum_{i=1}^n I_2^{(i)} - a_2 \sum_{i=1}^n I_1^{(i)} \right)^2 + \right. \right. \\
 & \left. \left(a_2 \sum_{i=1}^n I_3^{(i)} - a_3 \sum_{i=1}^n I_2^{(i)} \right)^2 + \right. \\
 & \left. \left(a_1 \sum_{i=1}^n I_3^{(i)} - a_3 \sum_{i=1}^n I_1^{(i)} \right)^2 + \right. \\
 & \left. (a_1^2 + a_2^2 + a_3^2) \sum_{k=1}^3 \sum_{i=1}^n \sum_{j=i+1}^n (I_k^{(i)} - I_k^{(j)})^2 \right\} \\
 & \left. / ((a_1^2 + a_2^2 + a_3^2) \sigma^2 n) \right]^{\frac{n-2}{2}} (a_1^2 + a_2^2 + a_3^2)^{-\frac{1}{2}}, \tag{12}
 \end{aligned}$$

which is valid for the three cube experiments for each sample (donor, acceptor, or construct). For the donor-only sample, $a_i = a_i^{(D)}$; for the acceptor-only sample, $a_i = a_i^{(A)}$; and for the sample with the donor-acceptor

construct, the a_i are replaced with $a_i^{(D)}(1 - E_{fr}^c) + a_i^{(A)} + \frac{\varepsilon_i^{(D)}}{\varepsilon_i^{(A)}} a_i^{(A)} E_{fr}^c$. The posterior probability, including all the calibration data together, is obtained by multiplying the forms of Eq (12) for each of the three 3-cube experiments.

For a given data set, we can infer E_{fr}^c and five of the six $a_i^{(S)}$ constants relative to the remaining one, for instance $a_2^{(A)}$, which is set to unity. The numerator of Eq (12) remains unaltered by such a rescaling of the $a_i^{(S)}$, and the alteration in the denominator is cancelled by the Jacobian required for the change in variables. One can use a Markov chain Monte Carlo method to sample the $a_i^{(S)}$ and E_{fr}^c from P_{calib} . Alternatively, one can use a numerical solver to find the most probable values by solving for the roots of the system of equations consisting of the derivatives of Eq (12) with respect to each of the six variables. Using this relative calibration procedure, the final K_d values obtained would relate to the true K_d by the scaling factor, for instance $a_2^{(A)}$.

Data simulation

We designed our simulated data to mimic the key features of experimental data, which could come from various sources, such as fluorescence reader measurements of solutions of purified proteins or images of cells from fluorescence microscopy that have been processed and quantified. To simulate data, we wrote a function in Matlab (The Mathworks, Natick, MA) that takes as input E_{fr} , K_d , \mathbf{A}_0 , \mathbf{D}_0 , n , $a_i^{(S)}$, and r where \mathbf{A}_0 , \mathbf{D}_0 are vectors of length m (representing m cells or regions within cells), n is the number of measurements made on each cell or vesicle, and r is the strength of measurement noise. The function outputs a set of simulated three-cube FRET data from m samples with n measurements per sample.

For each pair of concentrations, $[A_0]$ and $[D_0]$, we calculate $[D]$, $[A]$, and $[DA]$ using Eq (5). Next, we calculate the simulated experimental intensities I_D , I_A , and I_F using Eq (1). Finally, we simulate measurement noise by adding Gaussian random numbers to the data. The variable r is a scalar between 0 and 1 that refers to the strength of the measurement noise relative to the mean of the observed signal. Noise with $r = 10\%$ would have Gaussian measurement noise with standard deviation that is 10% of the mean signal observed in each channel.

In our examples, we use $E_{fr} = 0.4$ and $K_d = 1 \mu\text{M}$. Unless otherwise indicated, we simulate ten measurements per cell per channel and use the following set of $a_i^{(S)}$ values, which were chosen to represent a donor-

acceptor pair with considerable spectral cross-talk and bleed-through:

$$\begin{aligned} a_1^{(D)} &= 0.800, & a_1^{(A)} &= 0.020, & a_1^{(DA)} &= 0.008, \\ a_2^{(D)} &= 0.100, & a_2^{(A)} &= 1.000, & a_2^{(DA)} &= 0.100, \\ a_3^{(D)} &= 0.170, & a_3^{(A)} &= 0.200, & a_3^{(DA)} &= 0.240. \end{aligned}$$

The constants corresponding to the donor, acceptor, and complex fluorescing in their respective channels ($a_1^{(D)}$, $a_2^{(A)}$, and $a_3^{(DA)}$) are assigned the largest values. As donors and acceptors are often not equally bright, we set $a_1^{(D)} < a_2^{(A)}$. In the FRET channel, the constants corresponding to bleed-through from donors ($a_3^{(D)}$) and cross-talk from acceptors ($a_3^{(A)}$) have been set to approximately 20% of the constants for donor and acceptor in their respective channels ($a_1^{(D)}$ and $a_2^{(A)}$) because spectral overlap contributes significantly in the FRET channel. In the acceptor channel, smaller values are assigned to $a_2^{(D)}$ and $a_2^{(DA)}$ to describe the donor undergoing cross-talk and, subsequently, either bleed-through ($a_2^{(D)}$) or FRET ($a_2^{(DA)}$) in the acceptor channel. It is unlikely that acceptors would emit photons detectable in the donor channel, so the constants corresponding to that process, $a_1^{(A)}$ and $a_1^{(DA)}$, are very small, but they are non-zero to show that any measurable spectral contamination can be included.

Incorporating prior information

The Bayesian framework makes it possible to incorporate prior information about any of the parameters, including uncertainty in $a_i^{(S)}$. As an example, we include prior knowledge about E_{fr} (see the Results). Such data could be collected in separate experiments under slightly different experimental conditions and in the presence of measurement noise. We define the prior probability of E_{fr} by a Gaussian distribution with mean at the measured value \hat{E}_{fr} and variance σ_E , which reflects the confidence in the measurement:

$$P(E_{fr}) = \frac{1}{\sigma_E} \exp \left[\frac{-(E_{fr} - \hat{E}_{fr})^2}{2\sigma_E^2} \right]. \quad (13)$$

Markov chain Monte Carlo (MCMC) estimation

To sample K_d and E_{fr} from the posterior probability distribution, we use a biased random walk. Although a simpler approach would be sufficient to explore two dimensions, we use this method because it will allow us to efficiently extend our search to sample other,

additional parameters we may wish to infer, such as the values of $a_i^{(S)}$.

We use the Metropolis-Hastings algorithm [32,33]. It begins at a random location in parameter space and takes random steps, moving in up to 3 dimensions at a time with proposed steps drawn from a distribution that is symmetric about the current location. Proposals that increase the posterior probability are always accepted; those that decrease it are accepted with probability $\frac{P(x^j)}{P(x^{j-1})}$, where $P(x^j)$ is the posterior probability of the proposed step and $P(x^{j-1})$ is the posterior probability of the current location. The walk is run for a sufficiently long time (about 10,000 steps) to generate independent samples from the posterior distribution and the step size chosen to maintain an acceptance rate of 40 - 60% [28].

Once the walk has converged with the energy fluctuating around a minimum value, we record the steps taken and use the histogram of these sampled values as our estimate of the posterior probability distribution. From this estimated distribution, we obtain the mean and standard deviation of each parameter being inferred. The algorithm is summarized below.

Marginalization of D_0 and A_0

We have defined the likelihood in Eq 7 and assume that the measurement noise in each channel is independent. Because we have a Gaussian model, the values of the measurement noise parameters, σ_D , σ_A and σ_F , that maximise the posterior probability are approximately equivalent to the observed variances of the data from the donor, acceptor, and FRET channels, respectively [28]. However, it would also be possible to fit σ_D^2 , σ_A^2 and σ_F^2 directly, along with K_d and E_{fr} .

Although K_d , E_{fr} , D_0 and A_0 are important parameters, we cannot measure them directly. For our purposes, we are interested in the values of K_d and perhaps E_{fr} , but not the values of D_0 and A_0 . Rather than fit D_0 and A_0 , we marginalize or integrate them out:

$$\begin{aligned} P(\text{data}|K_d, E_{fr}) &= \int dD_0 dA_0 P(\text{data}, D_0, A_0 | K_d, E_{fr}) \\ &= \int dD_0 dA_0 P(\text{data}|K_d, E_{fr}, D_0, A_0) \\ &\quad \times P(D_0, A_0). \end{aligned} \tag{14}$$

To indicate that we have no knowledge about the values of D_0 and A_0 , we set the prior, $P(D_0, A_0)$, to a constant for positive D_0 and A_0 (and 0 otherwise).

As a result,

$$\begin{aligned} P(\text{data}|K_d, E_{fr}) &\propto \int dD_0 dA_0 P(\text{data}|K_d, E_{fr}, D_0, A_0). \end{aligned} \tag{15}$$

Because this expression is difficult to integrate analytically, we consider the 'energy', $E = -\log(P(\text{data}|K_d, E_{fr}, D_0, A_0))$, and approximate E with a two-dimensional Taylor expansion about $\{D_0^*, A_0^*\}$, where $\{D_0^*, A_0^*\}$ minimizes E (and thus maximizes the likelihood). Ignoring terms above second order,

$$\begin{aligned} E &\simeq E(D_0^*, A_0^*) + \\ &\quad \frac{1}{2} \begin{pmatrix} D_0 - D_0^* \\ A_0 - A_0^* \end{pmatrix}^T \nabla \nabla E(D_0^*, A_0^*) \begin{pmatrix} D_0 - D_0^* \\ A_0 - A_0^* \end{pmatrix}, \end{aligned} \tag{16}$$

where $\nabla \nabla E$ is the Hessian, or matrix of second-derivatives of E with respect to D_0 and A_0 .

Therefore,

$$\begin{aligned} P(\text{data}|K_d, E_{fr}, D_0, A_0) &= \exp(-E) \\ &\simeq \exp \left[-E(D_0^*, A_0^*) - \frac{1}{2} \begin{pmatrix} D_0 - D_0^* \\ A_0 - A_0^* \end{pmatrix}^T \right. \\ &\quad \left. \times \nabla \nabla E(D_0^*, A_0^*) \begin{pmatrix} D_0 - D_0^* \\ A_0 - A_0^* \end{pmatrix} \right]. \end{aligned} \tag{17}$$

This approximation of the likelihood results in a Gaussian integrand, which we can then integrate analytically [34]. Note that we do not use the limits of integration $(0, \infty)$ which would be appropriate for non-negative concentration values. We instead use the limits $(-\infty, \infty)$ to make the integral simple. It is a valid approximation provided that D_0^* and A_0^* are sufficiently large and it consistently yields appropriate results in practice.

In summary,

$$\begin{aligned} P(\text{data}|K_d, E_{fr}) &= \int_{-\infty}^{\infty} dA_0 dD_0 P(\text{data}|K_d, E_{fr}, D_0, A_0) \\ &\approx \frac{\exp(-E(D_0^*, A_0^*))}{\sqrt{|H|}}, \end{aligned} \tag{18}$$

where $|H|$ is the determinant of the Hessian, $\nabla \nabla E(D_0^*, A_0^*)$. Although K_d and E_{fr} do not appear explicitly in the final form of Eq (18), both $E(D_0^*, A_0^*)$ and H depend on K_d and E_{fr} .

Algorithm summary

Our algorithm for sampling from the posterior probability distribution of (K_d, E_{fr}) with a prior probability distribution for E_{fr} consists of the following steps:

1. Perform calibration to obtain values (absolute or relative) for the $a_i^{(S)}$ constants.

2. Define prior probability distributions for parameters to be inferred based on initial information that is known, if any.

3. Run Markov chain Monte Carlo algorithm:

Choose initial step (K_d^1, E_{fr}^1) and compute the posterior probability, $P_1 = P(K_d^1, E_{fr}^1)$, as outlined in (b) below.

For $j = 2$ to n ,

(a) Choose proposal step (K_d^j, E_{fr}^j) .

(b) Compute the posterior probability, $P_j = P(K_d^j, E_{fr}^j)$, for that step:

i. Find $\{D_0^*, A_0^*\}$, the values of $\{D_0, A_0\}$ that minimise the energy of the posterior probability using nonlinear optimisation (the Nelder-Mead simplex algorithm implemented in Matlab's `fminsearch` function (The Mathworks, Natick, MA)).

ii. Compute $\nabla^2 E(D_0^*, A_0^*)$ using analytical expressions for the second derivatives of E computed in Mathematica (Wolfram Research, Champaign, IL) and exported to Matlab.

iii. Using Eq (18), compute the likelihood, $P(\text{data} | K_d^j, E_{fr}^j)$.

iv. Using Eq (13), compute the prior, $P(E_{fr}^j)$.

v. Compute the posterior probability distribution, P_j , which is the likelihood times the prior probability.

(c) Check whether to accept the move to (K_d^j, E_{fr}^j) :

• If $P_j > P_{j-1}$, accept.

• Otherwise, accept with probability $\frac{P_j}{P_{j-1}}$.

Steps (a)-(c) are repeated until $j = n$. After an initial burn-in period where the energy reaches a minimum and the walk achieves a stationary distribution, record samples of K_d and E_{fr} .

4. Repeat step 3, varying initial (K_d, E_{fr}) . To ensure convergence, the stationary distributions of all walks should overlap [28].

Availability

We have made our data simulation and analysis software available at <http://swainlab.bio.ed.ac.uk/software/FRET>.

Acknowledgements

We thank Marko Laine (Finnish Meteorological Institute) for providing his MCMC code and helpful feedback. We also wish to acknowledge Stephane Laporte, May Simaan and Jay Nadeau (McGill University) for useful discussions and Andrea Weisse, Bruno Martins, Christos Josephides, Clive Bowsher (University of Edinburgh), Nathan Scales (McGill University), and

Vahid Shahrezaei (Imperial College London) for helpful comments on the manuscript. CL and PSS are supported by the Scottish Universities Life Sciences Alliance.

Author details

¹Department of Physiology, McGill University, 3655 Promenade Sir William Osler, Montreal, Quebec H3G 1Y6, Canada. ²Centre for Systems Biology at Edinburgh, University of Edinburgh, Edinburgh EH9 3JD, UK.

Authors' contributions

PSS conceived of the study. CL and PSS developed the methodology. CL implemented and tested the algorithm. Both authors wrote the paper and read and approved the final manuscript.

Received: 14 February 2011 Accepted: 9 May 2011

Published: 9 May 2011

References

1. Verveer P, Bastiaens P: Quantitative microscopy and systems biology: seeing the whole picture. *Histochemistry and cell biology* 2008, **130**(5):833-843.
2. Megason S, Fraser S: Imaging in systems biology. *Cell* 2007, **130**(5):784-795.
3. Lakowicz J: *Principles of Fluorescence Spectroscopy* New York, NY: Springer; 2006.
4. Vogel SS, Thaler C, Koushik SV: Fanciful FRET. *Sci STKE* 2006, **2006**:re2.
5. Piston DW, Kremers GJ: Fluorescent protein FRET: the good, the bad and the ugly. *Trends Biochem Sci* 2007, **32**:407-14.
6. Hoppe A: Quantitative FRET microscopy of live cells. In *Imaging Cellular and Molecular Biological Functions*. Edited by: Shorte SL, Frischknecht F. Heidelberg: Springer; 2007:157-180.
7. Gordon G, Berry G, Liang X, Levine B: Quantitative fluorescence resonance energy transfer measurements using fluorescence microscopy. *Biophysical Journal* 1998, **74**:2702-2713.
8. Xia Z, Zhou Q, Lin J, Liu Y: Stable SNARE complex prior to evoked synaptic vesicle fusion revealed by fluorescence resonance energy transfer. *J Biol Chem* 2001, **276**:1766-71.
9. Nagy P, Vámosi G, Bodnár A, Lockett SJ, Szöllösi J: Intensity-based energy transfer measurements in digital imaging microscopy. *Eur Biophys J* 1998, **27**:377-89.
10. Elangovan M, Wallrabe H, Chen Y, Day RN, Barroso M, Periasamy A: Characterization of one- and two-photon excitation fluorescence resonance energy transfer microscopy. *Methods* 2003, **29**:58-73.
11. Trón L, Szöllösi J, Damjanovich S, Helliwell SH, Arndt-Jovin DJ, Jovin TM: Flow cytometric measurement of fluorescence resonance energy transfer on cell surfaces. Quantitative evaluation of the transfer efficiency on a cell-by-cell basis. *Biophysical Journal* 1984, **45**:939-46.
12. Youvan D, Silva C, Bylina E, Coleman W, Dilworth M, Yang M: Calibration of fluorescence resonance energy transfer in microscopy using genetically engineered GFP derivatives on nickel chelating beads. *Biotechnology et alia* 1997, **3**:1-18.
13. Berney C, Danuser G: FRET or no FRET: a quantitative comparison. *Biophysical Journal* 2003, **84**:3992-4010.
14. Zal T, Gascoigne NRJ: Photobleaching-corrected FRET efficiency imaging of live cells. *Biophysical Journal* 2004, **86**:3923-39.
15. Hoppe A, Christensen K, Swanson JA: Fluorescence resonance energy transfer-based stoichiometry in living cells. *Biophysical Journal* 2002, **83**(6):3652-64.
16. Erickson MG, Alseikhan BA, Peterson BZ, Yue DT: Preassociation of calmodulin with voltage-gated Ca(2+) channels revealed by FRET in single living cells. *Neuron* 2001, **31**(6):973-85.
17. Chen H, Puhl H, Koushik S, Vogel S, Ikeda S: Measurement of FRET efficiency and ratio of donor to acceptor concentration in living cells. *Biophysical Journal* 2006, **91**:L39-41.
18. Elder A, Domin A, Schierle GK, Lindon C, Pines J, Esposito A, Kaminski C: A quantitative protocol for dynamic measurements of protein interactions by FRET-sensitized fluorescence emission. *Royal Society Interface* 2009, **6**:1-24.
19. Włodarczyk J, Woehler A, Kobe F, Ponimaskin E, Zeug A, Neher E: Analysis of FRET signals in the presence of free donors and acceptors. *Biophysical Journal* 2008, **94**:986-1000.

20. Gadella T: *FRET and FLIM Techniques* Elsevier Science; 2009.
21. Chen H, III HP, Ikeda S: **Estimating protein-protein interaction affinity in living cells using quantitative Förster resonance energy transfer measurements.** *Journal of Biomedical Optics* 2007, **12**:054011.
22. Bruggeman F, Westerhoff H: **The nature of systems biology.** *TRENDS in Microbiology* 2007, **15**:45-50.
23. Piehler J: **New methodologies for measuring protein interactions in vivo and in vitro.** *Curr Opin Struct Biol* 2005, **15**:4-14.
24. Kreutz C, Timmer J: **Systems biology: experimental design.** *FEBS Journal* 2009, **276**(4):923-942.
25. Neher RA, Neher E: **Applying spectral fingerprinting to the analysis of FRET images.** *Microsc Res Tech* 2004, **64**:185-95.
26. Ljosa V, Carpenter AE: **Introduction to the Quantitative Analysis of Two-Dimensional Fluorescence Microscopy Images for Cell-Based Screening.** *PLoS Comput Biol* 2009, **5**(12):e1000603.
27. Wilkinson D: **Stochastic modelling for quantitative description of heterogeneous biological systems.** *Nat Rev Genet* 2009, **10**(2):122-133.
28. MacKay D: *Information Theory, Inference & Learning Algorithms* Cambridge: Cambridge University Press; 2002.
29. Wallrabe H, Periasamy A: **Imaging protein molecules using FRET and FLIM microscopy.** *Current Opinion in Biotechnology* 2005, **16**:19-27.
30. Rosenfeld N, Perkins TJ, Alon U, Elowitz MB, Swain PS: **A fluctuation method to quantify in vivo fluorescence data.** *Biophysical Journal* 2006, **91**(2):759-66.
31. Digman MA, Dalal RB, Horwitz AF, Gratton E: **Mapping the number of molecules and brightness in the laser scanning microscope.** *Biophysical Journal* 2008, **94**(6):2320-2332.
32. Metropolis N, Rosenbluth A, Rosenbluth M, Teller A, Teller E: **Equation of state calculations by fast computing machines.** *J Chem Phys* 1953, **21**:1087-1092.
33. Hastings W: **Monte Carlo sampling methods using Markov chains and their applications.** *Biometrika* 1970, **57**:97-109.
34. Sivia D, Skilling J: *Data Analysis: a Bayesian Tutorial* Oxford: Oxford University Press; 2006.

doi:10.1186/2046-1682-4-10

Cite this article as: Lichten and Swain: A Bayesian method for inferring quantitative information from FRET data. *BMC Biophysics* 2011 **4**:10.

**Submit your next manuscript to BioMed Central
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

