

RESEARCH ARTICLE

A computational approach to inferring cellular protein-binding affinities from quantitative fluorescence resonance energy transfer imaging

Khamir Mehta¹, Adam D. Hoppe², Raghunandan Kainkaryam¹, Peter J. Woolf^{1,3}, and Jennifer J. Linderman^{1,3}

¹Department of Chemical Engineering, University of Michigan, Ann Arbor, MI, USA

²Department of Chemistry and Biochemistry, South Dakota State University, Brookings, SD, USA

³Program in Bioinformatics, University of Michigan, Ann Arbor, MI, USA

Fluorescence resonance energy transfer (FRET) microscopy can measure the spatial distribution of protein interactions inside live cells. Such experiments give rise to complex data sets with many images of single cells, motivating data reduction and abstraction. In particular, determination of the value of the equilibrium dissociation constant (K_d) will provide a quantitative measure of protein–protein interactions, which is essential to reconstructing cellular signaling networks. Here, we investigate the feasibility of using quantitative FRET imaging of live cells to estimate the *local* value of K_d for two interacting labeled molecules. An algorithm is developed to infer the values of K_d using the intensity of individual voxels of 3-D FRET microscopy images. The performance of our algorithm is investigated using synthetic test data, both in the absence and in the presence of endogenous (unlabeled) proteins. The influence of optical blurring caused by the microscope (confocal or wide field) and detection noise on the accuracy of K_d inference is studied. We show that deconvolution of images followed by analysis of intensity data at local level can improve the estimate of K_d . Finally, the performance of this algorithm using cellular data on the interaction between yellow fluorescent protein-Rac and cyan fluorescent protein-PBD in mammalian cells is shown.

Received: June 6, 2008

Revised: June 18, 2009

Accepted: September 1, 2009

Keywords:

Dissociation constant / Fluorescence resonance energy transfer / Image deconvolution / Optical blurring / Protein–protein interaction / Technology

1 Introduction

Protein–protein interaction networks form a fundamental regulatory mechanism controlling the behavior of living cells. Characterization of these interactions, in particular the

measurement of protein affinities, is of interest for various applications including tissue engineering, drug discovery and development of predictive models of cell behavior. Although many methods have been developed to measure the binding affinities of interacting proteins, including *in vitro* assays [1–6], methods for quantitative local characterization of protein–protein binding in live cells still require improvement.

Fluorescence microscopy is the method of choice for direct visualization of proteins in native cellular environments [7–10], and recent developments in imaging techniques promise measurement of protein interactions with improved spatial and temporal resolution [5, 8, 11]. Protein–protein binding inside live cells can be visualized by fluorescence resonance energy transfer (FRET) [12]. FRET is the non-radiative transfer of fluorescence energy from an excited

Correspondence: Professor Jennifer J. Linderman, Department of Chemical Engineering, University of Michigan, Ann Arbor, MI 481092136, USA

E-mail: linderma@umich.edu

Fax: +1-734-764-0459

Abbreviations: CFP, cyan fluorescent protein; FRET, fluorescence (or forster) resonance energy transfer; PSF, point spread function; YFP, yellow fluorescent protein; 3-D FSR, 3-D FRET stoichiometry

fluorescent donor to a nearby lower energy fluorescent acceptor *via* dipole–dipole interactions. This process results in decreased emission of the donor and increased emission from the acceptor. The range over which FRET can occur is less than 10 nm and thus the appearance of FRET is indicative of spatial proximity of the two interacting proteins. The spatial proximity can be used to infer the association of donor and acceptor-labeled proteins [13]; however, independent verification of the interaction may be required. In cellular systems, FRET can be used to ascertain the binding of specific protein pairs by fluorescently labeling them with different variants of fluorescent proteins. Typical experimental data in form of large numbers of images of multiple cells make analyses difficult and time consuming. Although qualitative information on the binding affinity has been routinely inferred from the images, methods for *quantitative* characterization of protein interactions are needed.

Determining the values of key physical parameters characterizing protein–protein binding, *e.g.* the equilibrium dissociation constant K_d , from FRET experiments will require additional steps beyond image acquisition. Microscope images are blurred by the optical imaging process such that points within an image plane contain light from out-of-focus planes and adjacent points. Microscope blurring is characterized by the point spread function (PSF) of the microscope, which is the image of a single point source. This optical blurring limits the accuracy of intensity-based calculations. Confocal microscopes reduce blurring as compared with conventional wide-field microscopes, but significant optical distortion is still present. Image deconvolution algorithms can deblurr data from both confocal and wide-field microscopes [14, 15]; however, their impact on the estimation of concentrations from image intensities and hence K_d is not well understood. Further quantification of fluorescent images will need a calibration function to map image intensities to molecular concentrations [16]. In the case of FRET microscopy, image intensities need to be corrected for spectral overlap of the donor and acceptor emissions and possible direct excitation of the acceptor at the donor excitation wavelength. Also, an independent estimate of FRET efficiency (E) is needed to characterize protein binding. There are a number of algorithms available for measurement of FRET as well as estimation of apparent FRET efficiencies [17–24]. The presence of unlabeled proteins (endogenous, photobleached or misfolded) which can compete with labeled species for binding introduces an additional complication; there have been efforts to estimate the FRET efficiency in this case [25] but the impact of such unlabeled proteins in inferring K_d remains unknown.

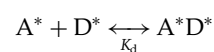
In this study, we demonstrate the feasibility of inferring local values of the apparent equilibrium disassociation constant (K_d) within a cell from FRET images using a synthetic data set. We use a simultaneous image deconvolution and spectral unmixing algorithm to accurately recover the concentration distribution of proteins [26] and study how the accuracy of the algorithm can aid in estimation of local

values of K_d . We investigate the impact of noise of the detection systems and the presence of unlabeled (*e.g.* endogenous) species or multiple binding affinities on the accuracy of K_d inference for both the wide-field microscope and also confocal microscope. Finally, we apply this algorithm to infer K_d from image data on binding of yellow fluorescent protein (YFP)-Rac and cyan fluorescent protein (CFP)-PBD in mammalian cells.

2 Materials and Methods

2.1 Reaction system

We consider the case of a bimolecular elementary reaction of labeled acceptor protein (A^*) and donor-labeled protein (D^*) tagged with variants of fluorescence protein appropriate for FRET to occur:

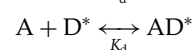
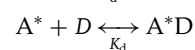
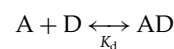


where A^*D^* is the acceptor–donor complex. Binding is quantified by the equilibrium dissociation constant (K_d) defined for a volume element v as

$$K_d = \frac{[A^*][D^*]}{[A^*D^*]} \quad (1)$$

where $[A^*]$, $[D^*]$ and $[A^*D^*]$ denote the concentrations of the labeled acceptor, donor and complexes in the volume under consideration. We assume that the continuum approximation holds within the volume, and hence K_d can be described by a mean value rather than its probabilistic equivalent.

In a general case where there are significant amounts of unlabeled acceptor (A) and/or donor (D) proteins present, binding reactions between labeled and unlabeled species or between two unlabeled species can also occur:



All reactions between A and D species, whether labeled or not, are assumed to have the same value of K_d and equilibrium relationships analogous to Eq. (1) can be written for each of these reactions.

2.2 FRET imaging experiment

The cell with tagged proteins is imaged using fluorescence microscopy to obtain images with intensities corresponding to the concentrations of the acceptor, donor and acceptor–donor complex proteins. In accordance with the nomenclature of Hoppe *et al.* [18], the following images of cellular contents are taken in a FRET experiment:

I_A : Image at acceptor excitation and acceptor emission (acceptor image)

I_D : Image at donor excitation and donor emission (donor image)

I_F : Image at donor excitation and acceptor emission (FRET image)

The images I_A , I_D and I_F can be acquired on a conventional wide-field microscope or a confocal microscope and they need to be analyzed further to gather information on the concentrations of the individual species.

2.3 3-D FRET stoichiometry reconstruction for improved local concentration estimates

We use the recently developed method termed 3-D FRET Stoichiometry (3-D FSR) [26] to take 3-D images corresponding to I_A , I_D and I_F and, by accounting for optical blurring due to the imaging process and donor–acceptor spectral overlap, obtain improved estimates for the concentrations of acceptor $[A^*]$, donor $[D^*]$ and donor–acceptor complex $[A^*D^*]$ in individual 3-D image pixels (voxels). Briefly, iterative maximum likelihood estimation is used for image deconvolution. An initial guess is convolved with the known PSF of the microscope and mixed as *per* the spectral overlap of donor–acceptor fluorophores to generate an estimate of the image which is then compared with the measured image to generate the next iterate. The optimization is allowed to proceed until a specified number (25) of iterations, the value of which is guided by our previous work. The algorithm corrects for spectral overlap using the spectral mixing model for FRET, which can be represented as

$$\begin{bmatrix} I_D \\ I_A \\ I_F \end{bmatrix} = c \begin{bmatrix} \xi & 0 & -\xi \\ 0 & \gamma/\alpha & 0 \\ \xi\beta & \gamma & 1 - \beta\xi \end{bmatrix} \times \begin{bmatrix} [D^*]_{\text{tot}} \\ [A^*]_{\text{tot}} \\ E[A^*D^*] \end{bmatrix} \quad (2)$$

This equation relates the images (I_A , I_D and I_F) to the concentrations of total (unbound+bound) labeled acceptor and donor ($[A^*]_{\text{tot}}$ and $[D^*]_{\text{tot}}$) and the concentration of the acceptor–donor complex times the FRET efficiency E ($E[A^*D^*]$). In absence of any spectral overlap, the matrix in Eq. (2) would be an identity matrix. The constants in the matrix are the characteristics of the microscope and fluorescent probes [19]. The matrix operation and parameters are equivalent to FRET stoichiometry [18], with the exception that ξ/γ has been replaced with ξ . The parameters used for our computations are in accordance with Hoppe *et al.* [19] and are $\xi = 0.2298$, $\alpha = 0.025$, $\beta = 0.7275$ and $\gamma = 0.0514$. There are a number of ways to independently estimate the FRET efficiency with individual merits and demerits [24]. For this study, we assume that E is known. The factor c is the calibration constant necessary to obtain absolute concentration values. Various approaches can be used to obtain c ; however, the estimation of c for local cellular

subcompartments remains challenging (*e.g.* [10, 16, 27]). For the *in silico* imaging described here, we assume that $c = 1$ intensity unit/ μM , while for the data on YFP-Rac2(V12) binding with CFP-PBD, we estimated the value of c from experiments as described in Section 2.6.

2.4 Computing K_d from image data

For a general case when both labeled and unlabeled molecules are present, the measured total labeled acceptor and donor concentrations include additional species:

$$[A^*]_{\text{tot}} = [A^*] + [A^*D] + [A^*D^*] \quad (3)$$

$$[D^*]_{\text{tot}} = [D^*] + [AD^*] + [A^*D^*] \quad (4)$$

Similarly, unlabeled total acceptor and donor concentrations can be expressed as

$$[A]_{\text{tot}} = [A] + [AD^*] + [AD] \quad (5)$$

$$[D]_{\text{tot}} = [D] + [A^*D] + [AD] \quad (6)$$

We define the variables r_A and r_D as the ratio of total labeled to unlabeled protein concentrations for acceptor and donor inside the cell respectively:

$$r_A = \frac{[A^*]_{\text{tot}}}{[A]_{\text{tot}}} \quad (7)$$

$$r_D = \frac{[D^*]_{\text{tot}}}{[D]_{\text{tot}}} \quad (8)$$

We assume that estimates of r_A and r_D are available from independent population-level experiments (*e.g.* Western blot analysis) or single cell experiments. Local values of r_A and r_D within particular 3-D volume elements (voxels) can vary due to the uncertainty in the measurement as well as fluctuations in the spatial distributions of the labeled and unlabeled species.

Algebraic manipulation of Eq. (1) (and analogous relationships with unlabeled species) along with Eqs. (3)–(8) gives

$$K_d = \frac{(r_A([D^*]_{\text{tot}} - [A^*D^*]) - [A^*D^*])(r_D([A^*]_{\text{tot}} - [A^*D^*]) - [A^*D^*])}{r_A r_D [A^*D^*]} \quad (9)$$

which allows a value for K_d to be calculated for each individual voxel. Individual voxel data can be combined to generate a probability distribution of inferred K_d (normalized by the total number of voxels in the original object) from the experimental data. The probability that the calculated value of K_d falls within a fraction f of a given value K_d^* is given by the area under the probability density distribution:

$$\Pr(K_d \in [K_d^* - fK_d^*, K_d^* + fK_d^*]) = \int_{K_d^* - fK_d^*}^{K_d^* + fK_d^*} p(K_d) dK_d \quad (10)$$

If the signal intensity in individual voxels is low, neighboring voxels may be binned to form elementary volume

compartments before calculation of K_d . Only those voxels (or compartments) where both labeled proteins A^* and D^* are present are useful in calculation of K_d and hence it is most efficient to perform calculations for only those compartments. Furthermore, blurring redistributes light to empty voxels, resulting in spurious estimates of K_d which are amplified by the non-linear form of Eq. (9). Thus we pre-select voxels for calculation of K_d based on the intensity/concentration of molecules by using a threshold criteria to identify compartments i with useful data:

$$([A^*]_i > \Gamma * \text{Max}[A^*]) \quad \text{and} \quad ([D^*]_i > \Gamma * \text{Max}[D^*]) \quad (11)$$

Γ is the threshold parameter and defines the minimum intensity value as a fraction of the maximum intensity value that should be present in the compartment for it to be used for the estimation of K_d . The intensity of the acceptor–donor complex is not included in the criteria to avoid selection biases based on the value of K_d .

2.5 Generation of synthetic test data

We generate synthetic images to test our methods for inferring K_d . Synthetic data generation consists of creating a cell object containing fluorescent molecules and complexes and then simulating the imaging process by convoluting the object with the PSF of a wide-field or confocal microscope.

First, a spherical cell object is generated using cubic 3-D volume pixels (voxels). Each voxel is randomly assigned discrete counts of labeled and unlabeled (if also present) acceptor and donor proteins chosen from a uniform distribution over a specified interval. The total numbers of labeled and unlabeled proteins in the object are determined from assumed values of the ratios r_A and r_D . As r_A and r_D are parameters which can be determined at best only on a whole cell-averaged basis, for individual voxels we assume that the ratio of labeled and unlabeled proteins is normally distributed with coefficient of variation σ_r . Using these sampled values of r_A and r_D and the total numbers of acceptor and donor proteins ($A+A^*$ and $D+D^*$) together with an assumed value of K_d , we calculate the numbers of complexes (AD , A^*D , AD^* , A^*D^*) in each voxel so as to satisfy Eqs. (3)–(8).

To simulate the imaging process, we mix the intensities on a voxel-by-voxel basis to simulate spectral mixing, using information on the spectral overlap of CFP and YFP to determine the mixed image as *per* the scheme outlined by Hoppe *et al.* [19]. In addition, we convolve this object with theoretical PSFs for either a wide-field or a confocal microscope. PSFs were generated in MATLAB 7.3 (MathWorks, USA) (Fig. 1). To simulate the spatial arrangement of multiple cells or compartments within a single cell, images of two or four spheres with smaller radii were created. Their radii were adjusted to keep the total volume constant. Each image is set to be equal to $100 \times 100 \times 100$ voxels with each voxel a cube of 60 nm side.

Any imaging process with a detection device has associated inherent noise; the presence of shot noise is unavoi-

able [28]. We simulate the shot noise in our images by using a Poisson-distributed detection noise model with variance and mean equal to the original intensity of the object. The S/N is here defined as the square root of the mean original intensity of the object, and we simulate various noise levels by changing the intensity of the original object.

2.6 Live cell FRET imaging

COS7 cells were transfected as in [26, 29] with previously described plasmids encoding YFP-Rac2(V12), CFP-PBD, CFP, YFP and YFP-CFP. The cells with linked YFP-CFP were used as positive control, and cells expressing free CFP and YFP molecules were used as negative control. All YFP molecules were actually monomeric citrine, containing the Q69M (pH desensitizing) and A206K (monomeric) mutations. PBD is the (p21 binding domain) from human PAK1. This domain provides an excellent test system because it has been demonstrated numerous times to interact exclusively with the small GTPases, Cdc42, Rac1 and Rac2 [30, 31].

Imaging was performed as described in [26]. Briefly, a novel high-speed microscope was used to collect 3-D FRET micro-

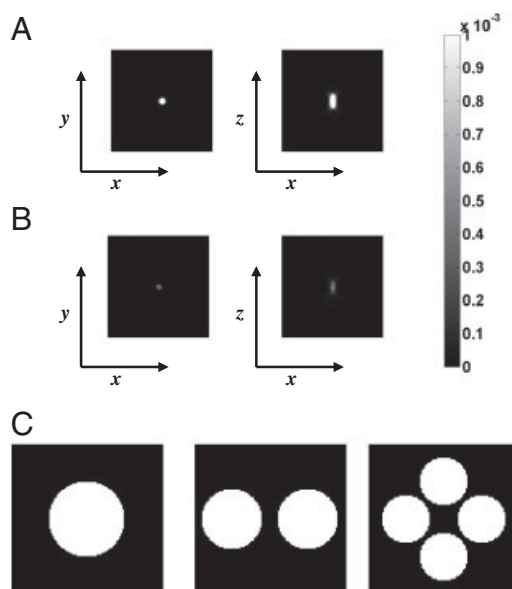


Figure 1. PSFs and synthetic test data. 3-D PSFs for a wide-field (A) and confocal microscope (B) were used to generate model data. The 3-D space had dimensions of $100 \times 100 \times 100$. The x - y slice is at midplane along z ($z = 50$), and the x - z slice is shown at $y = 50$. The theoretical PSFs were generated for emission wavelength = 530 nm. NA = 1.2 and voxel size = $60 \times 60 \times 60$ nm. Figures show grayscale-mapped images with a colormap scale of [0–0.001]. (C) Synthetic spherical cells contained acceptor–donor interactions. The diameter of the single large sphere (left) was 50 pixels, equivalent to $3 \mu\text{m}$ diameter. For images containing two and four spheres, the cell diameter was adjusted to conserve total volume. The spheres were centered in z -plane and were symmetrically arranged in the x - y plane.

scopy data by acquiring I_A , I_D and I_F images at each z -plane of a living cell. Acquisition of each 3-D data set took approximately 2.5 s. Estimates of $[D^*]$, $[A^*]$ and $[DA^*]$ were obtained by reconstruction of these data with 3-D FSR. The photo bleaching correction, as estimated from a representative experiment by the photon flux in each images I_A , I_D and I_F measured in successive z -planes, was found to be less than 2% in each signal, resulting in at the most 6% correction in the final value of K_d , and hence was neglected for this study.

To estimate the value of the calibration constant c , we imaged yeast cells expressing a chromosomal YFP fusion to Arf1 (Arf1-YFP) present at $\sim 20\,000$ copies/cell [32, 33]. Assuming that the cell is a $5\ \mu\text{m}$ diameter sphere, we can estimate an average number of 0.636 molecules *per* image voxel. Further, summing the 3-D intensity for whole yeast cells the intensity of a voxel on average was found to be about 0.2 units, and hence c is approximately 0.31 intensity units/molecule. The concentration of any species ($[s]_i$) in voxel i can then be computed from

$$[s]_i = \frac{I_{s,i}}{vcN_{\text{av}}} \quad (12)$$

where $I_{s,i}$ is the intensity corresponding to species s in the voxel, N_{av} is the Avogadro constant, and v is the volume of the voxel. The value of E for these data was taken as 0.37, estimated as *per* the previously published results [19].

3 Results and discussion

3.1 Impact of optical blurring in estimating protein concentrations

Synthetic images of a single sphere of uniform unit intensity were generated and then convolved with the PSF of either a wide-field or a confocal microscope to mimic imaging. In the absence of the optical distortion, we expect the image intensity distribution to be a single spike at unit intensity. Blurring, or optical distortion caused by optical imaging, disperses the intensity distribution and is more significant for the wide-field than the confocal microscope (Figs. 2A and B). Deconvolution of the measured images improves estimation of the local intensities (Figs. 2C and D), returning a peak intensity closer to the true intensity of the object, albeit with some distortions arising from the loss of information during imaging (these distortions can be seen by the shift in intensity for the reconstructed wide-field histogram (Fig. 2D) and multiple peaks in the reconstructed confocal histogram (Fig. 2C)).

Recovery of true voxel intensities is also affected by the shape and spatial arrangement of fluorophores. To investigate this, we repeated the convolution (imaging) and deconvolution steps above with a two- or four-sphere arrangement (Fig. 1C) while conserving the total volume of objects imaged. Figure 2E shows the fraction of pixels having intensity within 20% of the original object for the one-, two- and four-sphere systems. As spatial heterogeneity

is increased, the intensity histogram is more dispersed for multiple spheres than for a single larger sphere. This effect cannot be eliminated by deconvolution; however, the deconvolved images are significantly more accurate than the raw images. Deconvolution of acquired images can therefore improve the accuracy of measurement of local molecular concentrations by estimating the true intensity of the individual voxels. Deconvolution hence will be essential for estimation of K_d .

3.2 Inferring K_d from the image data

We next investigate the feasibility and accuracy of inferring K_d from measured image data. First, we considered a simple case in which all proteins under investigation are labeled and there is no measurement noise. We assumed uniform concentrations of acceptor, donor and acceptor–donor complex inside the 3-D volume of the test object. The test object was imaged by simulation using the 3-D FRET microscopy model [26] and these images are reconstructed by 3-D FSR to produce the corrected images shown in Figs. 3A and 4A for wide-field and confocal microscopes, respectively. Appropriate voxels were selected using the threshold criterion (Eq. (11)) and the corresponding K_d probability distribution was calculated using Eq. (9) (Figs. 3B and 4B). The fraction of voxels returning values of K_d within a specified fraction of the true value are shown in Figs. 3C and 4C. The dispersion in the probability distribution is a direct result of the optical distortion brought about by the imaging process. The distributions have a maximum near the true K_d for both the confocal and the wide-field microscopes, indicating that the proposed method is useful for identifying the local binding affinity from image data, although optical distortion can limit its accuracy. The effect of optical distortion on estimation of K_d increases with spatial heterogeneity (two and four sphere system; Figs. 3C and 4C). In all cases, the confocal microscope allows greater accuracy in measurement of K_d than the wide-field microscope.

Biological values of K_d are likely to vary widely. We repeated our inference procedure for various values of K_d and found that the shape of the probability distribution is unchanged. This is expected since the magnitude of intensities of the donor, acceptor and donor–acceptor complex images have no impact on deconvolution, and hence, in the absence of detection noise, while the absolute distortion does get scaled, the shape of the curve does not change (data not shown).

To investigate the effect of detection noise on inference of K_d , we modeled the image detection process with Poisson noise. The K_d probability distribution was calculated from 3-D FSR-reconstructed data with various noise levels, Figs. 5A–C. As expected, increasing noise disperses the probability distribution and limits the accuracy of our inference. To quantify the accuracy, we plot the probability of recovering the K_d within a specified fraction of the true value (Fig. 5D). As shown in the figure, even at low S/N , the algorithm can recover information on the true value of K_d .

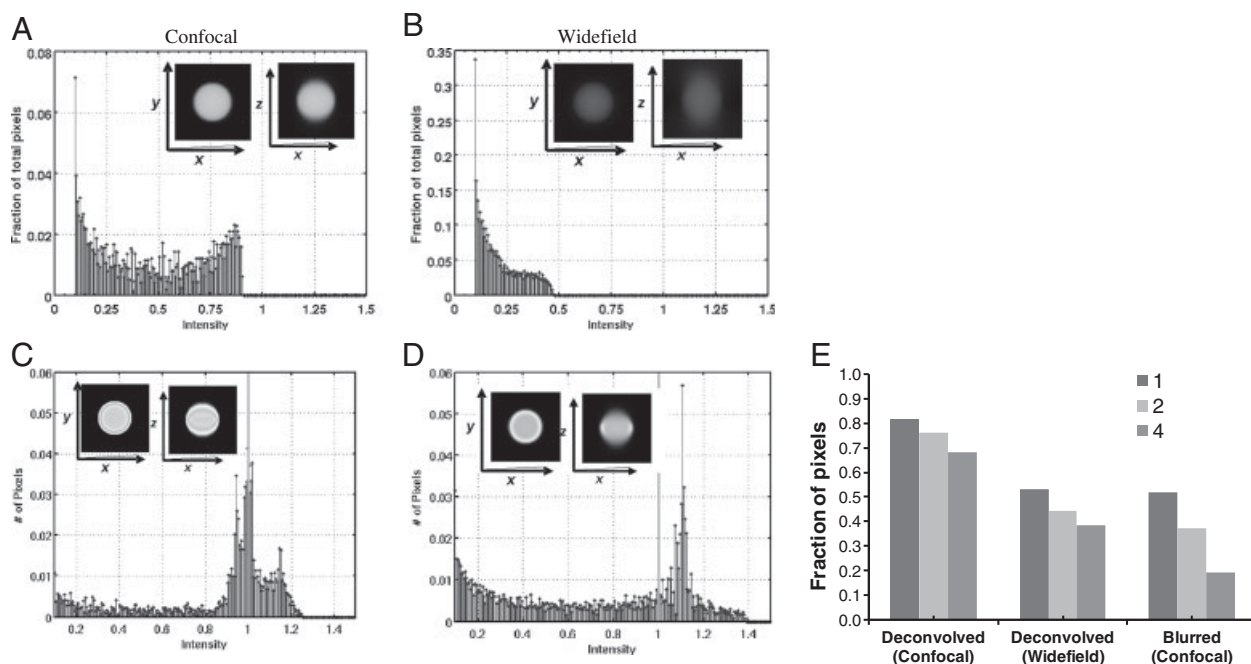


Figure 2. Deconvolution is essential for quantitative measurement of protein concentrations. The imaging process was simulated using a sphere of unit intensity and assuming a PSF for a confocal or wide-field microscope (Fig. 1). 3-D FSR was used to deconvolve images. Confocal image and intensity histogram are shown prior to (A) and subsequent to (C) deconvolution. The intensity colormap is set to [0, 1.5] for all images for comparison. Wide-field image and intensity histogram for the intensity interval [0.1, 1.5] are shown prior to (B) and subsequent to (D) deconvolution. Solid line indicates the true intensity distribution. (E) The fraction of voxels within $\pm 20\%$ of the true value ([0.8, 1.2]) is plotted for both the confocal and the wide-field microscope images, both before (blurred) and after deconvolution. The wide-field raw image before deconvolution has negligible voxels in the range, and hence does not appear in the histogram. Results for the two- and four-sphere case (see Fig. 1) are also shown.

3.3 Using thresholds to counter optical distortion and noise

Optical blurring can result in the assignment of low but non-zero intensity values to voxels that originally have no source of fluorescence. Low voxel intensities may also be the result of background noise. We use threshold criteria to exclude from calculations any voxels that have intensities lower than a fraction Γ of the maximum intensity of the acceptor and donor species in the deconvolved image data (Eq. 11). Figure 6 shows the effect of using thresholds on the K_d probability distribution for confocal microscopes. Comparing the distributions obtained using different threshold values (Figs. 6A–C), we can see that by limiting the calculations of K_d to voxels/compartments with a sufficient number of acceptor and donor proteins, one can improve the accuracy of K_d inference. This improvement is seen irrespective of the type of microscope and also for the case with larger spatial heterogeneity (multiple sphere system; data not shown).

Increasing the value of threshold parameter (Γ) will result in fewer compartments used for computation, and hence the

possible loss of meaningful data. On the other hand, keeping the threshold to a lower value will result in a broader distribution owing to the contribution of low-intensity voxels. Figure 6D shows the effect of increasing Γ on the mean value of the K_d probability distribution. The results indicate that there is a saturating effect of the threshold parameter above a critical value of the threshold. The optimal value of the threshold parameter will depend on the microscope and cellular system under investigation and could be found empirically, *e.g.* by constructing the curve similar to Fig. 6D.

3.4 Inferring K_d in the presence of multiple protein-binding states

Multiple values of K_d for a particular protein–protein pair within a single cell may be possible due to multiple protein states or cellular environments. To investigate the performance of our K_d inference algorithm for such a case, we generated synthetic data for a hypothetical case with two distinct binding constants occurring at different concentration ratios (1:1 and 3:1) and used our algorithm to obtain the K_d probability

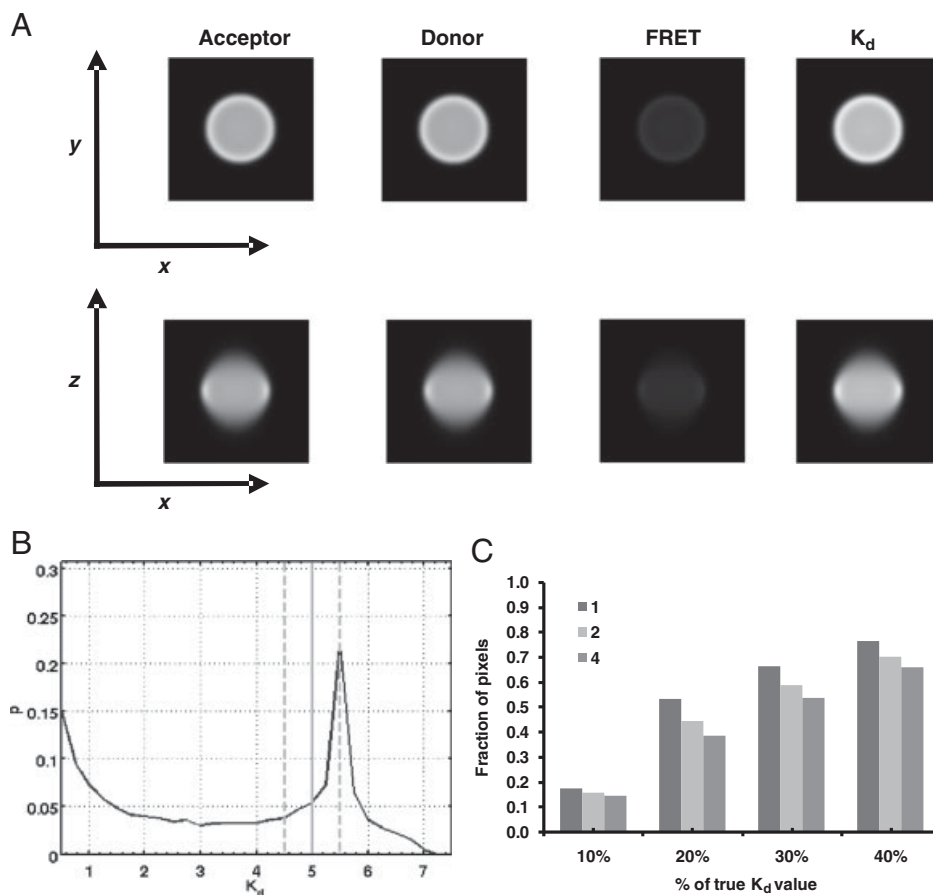


Figure 3. Inferring K_d from wide-field image data: Effect of optical distortion. (A) Acceptor [A], donor [D] and FRET image, [DA] following imaging (convolution) by a wide-field microscope, spectral un-mixing, and deconvolution of synthetic images. The calculated K_d image is also shown. (B) The K_d probability distribution for the sphere in (A) was calculated. The solid vertical line indicates the true value of K_d and the dotted vertical lines indicate the interval $\pm 10\%$. (C) To investigate the effect of spatial arrangement, calculations were also repeated for the two- and four-sphere arrangements of Fig. 1. The fraction of voxels with K_d within 10–40% of the true value are plotted for the one-, two- and four-sphere arrangements. The colormap is set to [0, 2] for all images, except the K_d image, where the colormap is set to [0, 7.5] and the units of K_d are expressed in μM .

distribution (Fig. 7). To quantify the relative concentrations of each binding state, we determined the fraction of the total voxels with a particular value of K_d . This will entail the calculation of area under the probability distribution. However, since the spread of the probability distribution depends on the absolute value of K_d , the correct approach is to normalize the area under the curve by the value of K_d . Mathematically, it is equivalent to calculating the area under the curve from a semi-log probability density distribution

$$A = \int \frac{pdK_d}{K_d} = \int pd \log(K_d) \quad (13)$$

Figure 7 shows that our algorithm can distinguish and correctly identify the existence of the two different binding states as indicated by the two distinct peaks for both confocal and wide-field microscopes. For the case where both the states are in equal concentrations, the ratio of the area under the first peak to the area under the second peak in the K_d probability distribution is 1.17 for the confocal (Fig. 7A) and 1.27 for the wide-field image (Fig. 7B), reasonably close to the true value of unity (Area was computed from the semi-log plot of the probability distribution with $f50.1$, 0.2 and 0.3 (Eq. (10)) and averaging the three results for each case). For the case where the concentration ratios were adjusted to 3:1 in the original image, we found the ratios of 3.22 for confocal (Fig. 7C) and

3.17 for wide-field image (Fig. 7D), again in agreement with the true number (3.33). Thus, our algorithm can successfully identify the relative concentrations of the two binding states. We note that the ability to distinguish two values of K_d increases as they become more different from each other, and if they are more spatially segregated.

3.5 Inferring K_d when unlabeled proteins are present

We extend our analysis to the case in which unlabeled proteins A and/or D are present and compete with labeled species A^* and D^* for binding. We now need independent measurements of the ratio of labeled to unlabeled proteins, r_A and r_D . (Eqs. (7) and (8)) to infer binding affinity. In the limit where r_A and r_D are very large, there are few unlabeled species present, and the system corresponds to the cases described in the earlier sections.

Figure 8 shows the performance of the inference procedure when unlabeled proteins are present. A random variation of the parameters r_A and r_D based on a normal distribution is superimposed to account for voxel-to-voxel variation in the number of labeled and unlabeled molecules arising from diffusion. Although the algorithm cannot recover the true

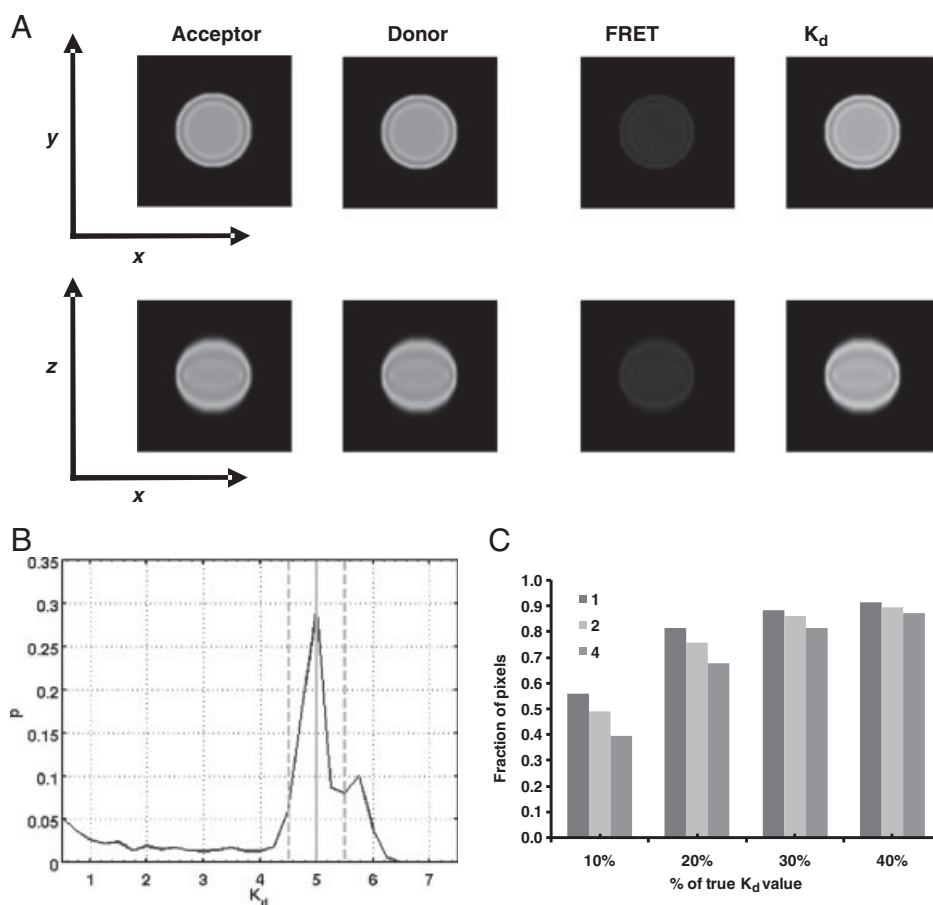


Figure 4. Inferring K_d from confocal image data: Effect of optical distortion. (A) Acceptor, donor and FRET images following imaging (convolution) by a confocal microscope, spectral un-mixing, and deconvolution of synthetic images. The calculated K_d image is also shown. (B) The K_d probability distribution for the sphere in (A) was calculated. The solid vertical line indicates the true value of K_d and the dotted vertical lines indicate the interval $\pm 10\%$. (C) To investigate the effect of spatial arrangement, calculations were also repeated for the two- and four-sphere arrangements of Fig. 1. The fraction of voxels with K_d within 10–40% of the true value is plotted for the one-, two- and four-sphere arrangements. The colormap is set to [0, 2] for all images, except the K_d image, where the colormap is set to [0, 7.5] and the units of K_d are expressed in μM .

value of K_d for the case when r_A and r_D are small (<0.05 ; data not shown), the spread of the distribution is not affected when unlabeled proteins are present and the distribution is similar to the case of no unlabeled species (Figs. 3 and 4). Figure 8 also shows the effect of these voxel-to-voxel variations on the inferred probability distribution of K_d . Variations in parameters r_A and r_D will affect the accuracy of the inferred K_d as per Eq. (9), and hence we expect the probability distribution to be broader for larger variation in r_A and r_D .

3.6 Application to cellular data on Rac-PBD binding

We applied our algorithm to FRET images obtained from imaging COS7 cells expressing the constitutively active mutant YFP-Rac2(V12) which binds to co-expressed CFP-PBD. Two negative controls, cells expressing free CFP and YFP-Rac2(V12) and cells with free over-expressed CFP and YFP, were used. Linked CFP and YFP molecules (CFP-YFP) were used as a positive control. Representative donor, acceptor and FRET images for all the four cases are shown in Fig. 9. We expect high affinity binding of CFP-PBD with YFP-Rac2(V12), (Fig. 9A), approaching the positive control case where CFP and YFP are linked

and expressed in the cell (Fig. 9D). In contrast, free CFP binds poorly to YFP-Rac2(V12) (Fig. 9B) or to free YFP (Fig. 9C).

Next, we calculated K_d probability distributions from the images, neglecting competition from unlabeled species under the assumption that the ectopically expressed proteins were in excess. Figure 9 also shows computed spatial distribution of protein-binding affinities (calculated as $1/K_d$ from the deconvolved image data) and Fig. 10 shows the computed probability distribution for all four cases. Increasing the value of the threshold parameter Γ from 0.1 to 0.3 did not significantly alter the location of the peak of the distribution (data not shown). We see a single, sharp peak in the probability density distribution curve corresponding to the real binding event of YFP-Rac2(V12) with CFP-PBD (Fig. 10A) at $K_d \sim 6 \mu\text{M}$, and a similar sharp peak (at higher affinity) for the linked CFP-YFP case (Fig. 10D) at $K_d \sim 1.4 \mu\text{M}$. Dispersion in the value of K_d as indicated by the spread of the distribution calculated from the experimental data (Figs. 10A and D) is on the order of 1 log, which is slightly higher than the dispersion in the value of K_d calculated from the synthetic data, even with noise, which is on the order of $1/2$ log (cf. Fig. 5). The greater dispersion shown in

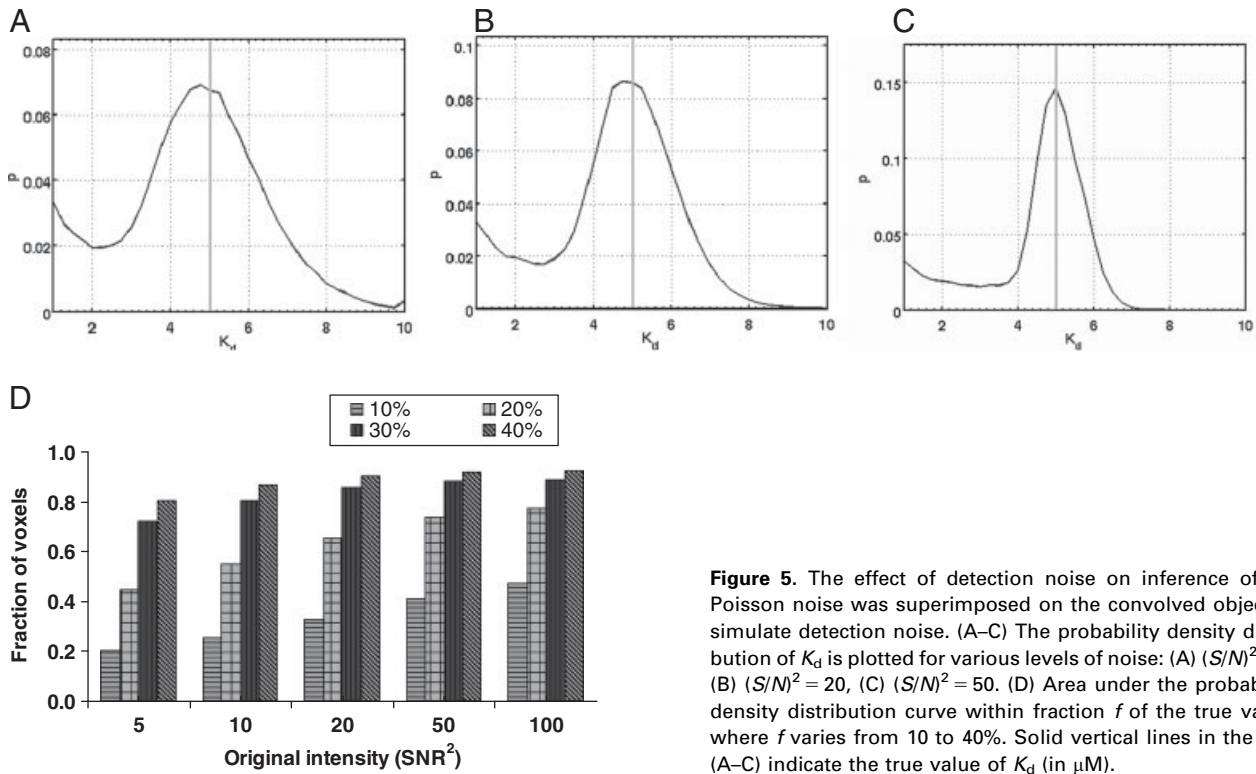


Figure 5. The effect of detection noise on inference of K_d . Poisson noise was superimposed on the convolved object to simulate detection noise. (A–C) The probability density distribution of K_d is plotted for various levels of noise: (A) $(S/N)^2 = 5$, (B) $(S/N)^2 = 20$, (C) $(S/N)^2 = 50$. (D) Area under the probability density distribution curve within fraction f of the true value, where f varies from 10 to 40%. Solid vertical lines in the plot (A–C) indicate the true value of K_d (in μM).

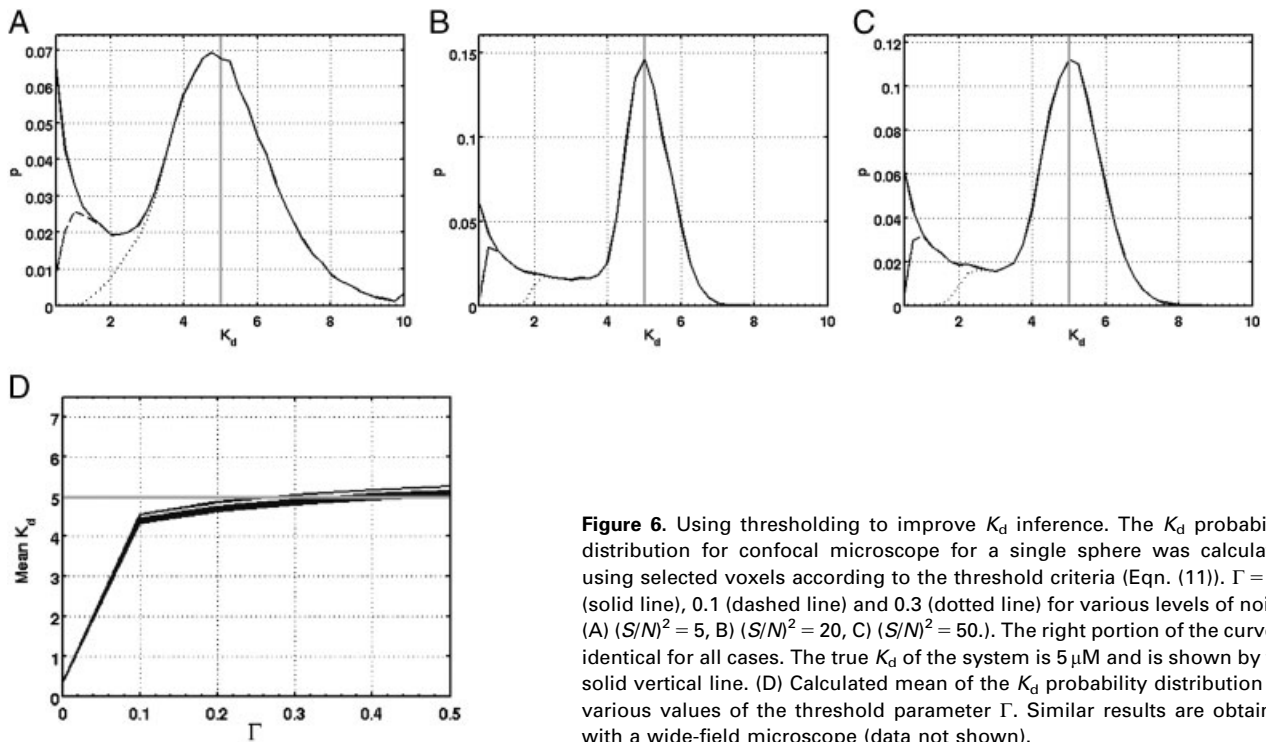


Figure 6. Using thresholding to improve K_d inference. The K_d probability distribution for confocal microscope for a single sphere was calculated using selected voxels according to the threshold criteria (Eqn. (11)). $\Gamma = 0.0$ (solid line), 0.1 (dashed line) and 0.3 (dotted line) for various levels of noise: (A) $(S/N)^2 = 5$, B) $(S/N)^2 = 20$, C) $(S/N)^2 = 50$. The right portion of the curve is identical for all cases. The true K_d of the system is 5 μM and is shown by the solid vertical line. (D) Calculated mean of the K_d probability distribution for various values of the threshold parameter Γ . Similar results are obtained with a wide-field microscope (data not shown).

Figs. 10A and D is due to the use of wide-field rather than confocal microscopy, the inclusion of all noise in the experimental situation rather than only shot noise as in

Fig. 5, and the smaller signaling volume in Figs. 10A and D (dispersion increases as object size decreases; not shown, but intuitive from Figs. 3 and 4).

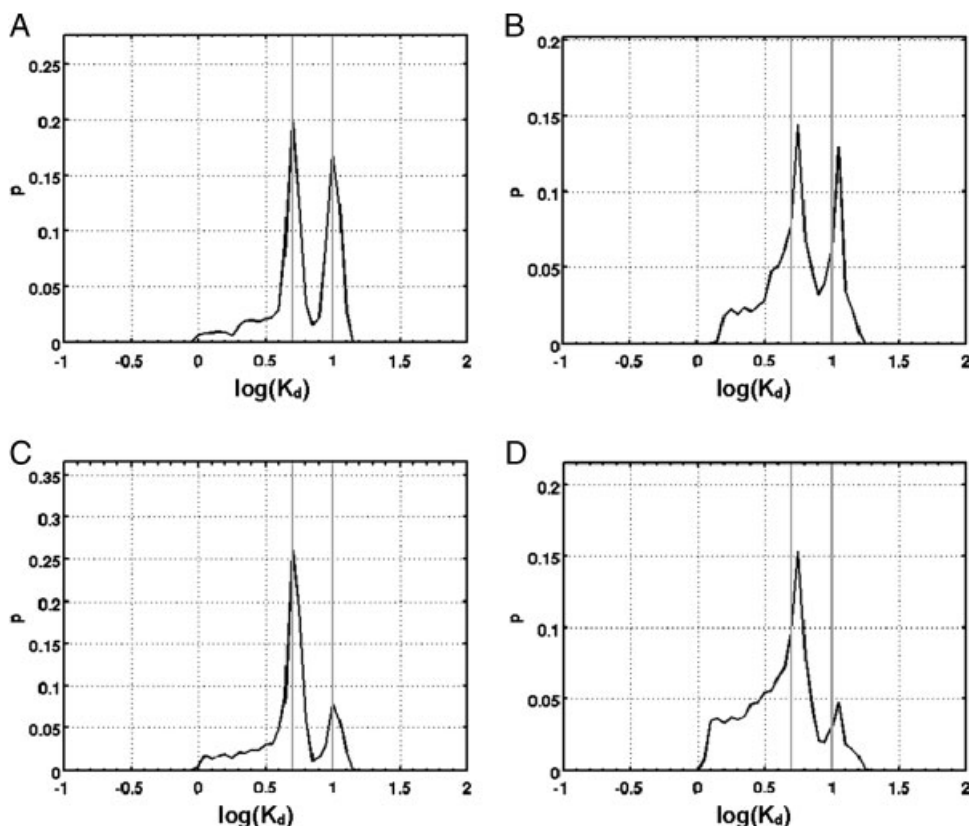


Figure 7. Inferring multiple values of K_d . Inferred K_d probability distributions are derived from synthetic data when the protein can bind with two possible values of K_d ($K_d = 5 \mu\text{M}$ and $K_d = 10 \mu\text{M}$). The four-sphere system shown in Fig. 1C is used for the computations. Distributions are shown for confocal (A) and wide-field (B) microscopes with synthetic data generated so that two of the four spheres have acceptors and donors binding with affinity $K_d = 5 \mu\text{M}$, and in the remaining two spheres proteins bind with affinity $K_d = 10 \mu\text{M}$. (C) and (D) show the corresponding distributions derived from confocal and wide-field microscopes respectively, when the three of the four spheres have data corresponding to high affinity binding state ($K_d = 5 \mu\text{M}$) and remaining one sphere has data corresponding to binding state $K_d = 10 \mu\text{M}$. True values of K_d are shown by solid vertical lines.

In the case of the free CFP controls (Figs. 10B and C), the peak location is about 2–5 orders of magnitude higher than for Figs. 10A and D, indicating poor binding as expected. We attribute the small but non-zero affinity with a broad distribution seen in the negative controls (Figs. 10B and C) to non-specific binding of the two molecules; the slight peak at very low affinity ($K_d \sim 10^7 \mu\text{M}$) is partly due to the logarithmic binning used to construct the probability distributions. We also note that computational errors associated with low-intensity FRET images can make the accurate detection of extremely low-affinity binding difficult. It is necessary to increase the number of iterations and the tolerance of the deconvolution algorithm, as done here, to prevent the appearance of spurious peaks in the probability distribution, especially for the negative controls.

The probability distribution for the binding of YFP-Rac2(V12) with CFP-PBD (Fig. 10A) can be used to compute a mean or cell-averaged value of K_d . The mean value of K_d as measured by our algorithm is $\sim 6 \mu\text{M}$, somewhat higher than the reported *in vitro* value of $0.2 \mu\text{M}$ [34]. The difference is likely at least in part real and due to significant differences between a cellular and *in vitro* environment. Inaccuracies in calibration and/or imaging and image processing may also contribute. However, our results clearly indicate at least 3–5 orders of magnitude difference between the binding affinities of the positive and

negative control, confirming that the algorithm can identify and also quantitatively distinguish the binding of Rac to PBD.

4 Concluding remarks

We have presented a method for inferring the local value of protein–protein equilibrium dissociation constant K_d from FRET microscopy imaging of cells. We have shown that deconvolution of both wide-field and confocal microscope image data is essential to inferring local molecular concentrations, and hence the value of K_d , and our algorithm can identify the existence of multiple binding states and their relative abundance. Using synthetic test data, we show that our algorithm can provide accurate values of K_d despite reasonable levels of noise and the presence of unlabeled proteins.

Our method builds on research done in the area of image deconvolution and spectral unmixing for measuring FRET efficiency [26] and the effect of free donors and acceptors [25] by quantifying the protein interactions *via* measurement of K_d . Our inference procedure utilizes the information from small volume elements of the cell (voxels), providing a distribution probability distribution for K_d and avoiding potential inaccuracies from averaging the signal from the whole cell [4]. A key strength of our algorithm is

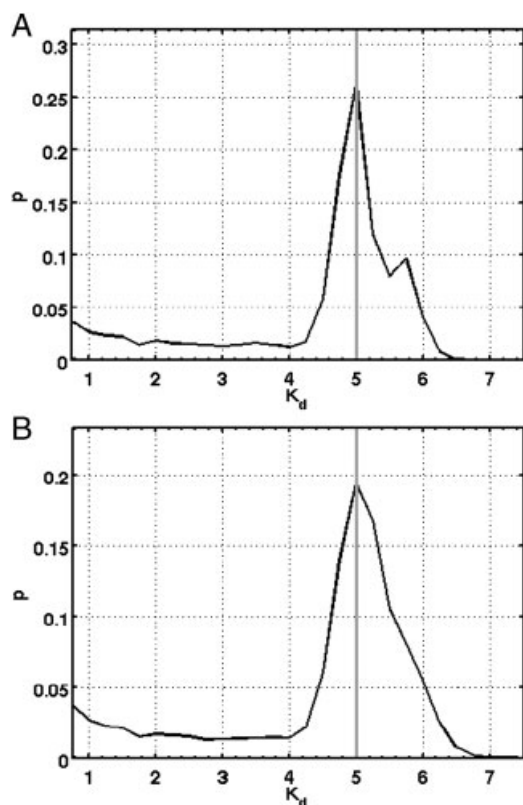


Figure 8. Effect of partial labeling of interacting proteins. The probability distribution of the inferred K_d is shown for $r_A = r_D = 1$. In the individual voxel the ratios were allowed to vary according to normal distribution with standard deviation 10% (A) and 30% (B). For low values of r_A and r_D (few labeled molecules), the algorithm will fail to recover the true values of K_d . (data not shown). The figures were simulated for a confocal microscope PSF, and the true value of K_d was set at $5 \mu\text{M}$ as shown by solid vertical lines.

the ability to quantify local protein interactions, and thus it can also be applied when there is protein sequestration or with non-cytosolic proteins. As imaging is conducted in a time frame that is much smaller than typical protein turnover times, the algorithm will not be affected by turnover. We have focused on intensity-based FRET measurements since they allow measurement of $[D]$, $[A]$ and $E[DA]$, which are not readily accessible by FLIM. However, one can imagine ways in which the analysis could be extended to other types of FRET experiments including FRET-FLIM-based measurements of protein interactions [35, 36].

The efficiency and accuracy of our approach can be affected by a number of factors, apart from the usual parameters affecting FRET microscopy. Difficulty in measuring the value of the calibration factor c is the primary limiting step in accurate determination of the local value of K_d . The presence of a significant number of unlabeled proteins adds the variances associated with the estimation of the local values of the ratio parameters r_A and r_D to the prediction of

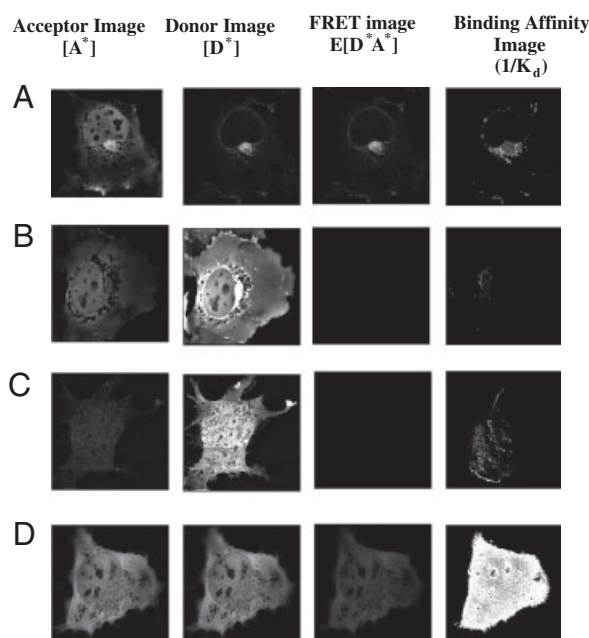


Figure 9. 3-D FSR imaging of mammalian cells. Representative images of total donor $[D^*]_{\text{tot}}$, total acceptor $[A^*]_{\text{tot}}$, donor acceptor complex $E[D^*A^*]$ and the spatial distribution of binding affinity $(1/K_d)$ after deconvolution and reconstruction from cells expressing different YFP and CFP constructs. (A) Images of cell expressing YFP-Rac2(V12) and CFP-PBD. (B) Images of cell expressing YFP-Rac2(V12) and free CFP (negative control). (C) Images of cell expressing free CFP and free YFP (negative control). (D) Images of cell expressing fused CFP-YFP (positive control). Images shown are x - y plane images at a representative z -plane and are grayscale with the same intensity map for the first three images on the same row ($[A^*]_{\text{tot}}$, $[D^*A^*]_{\text{tot}}$ and $E[D^*A^*]$) to enable visual evaluation of binding affinity. Affinity images $(1/K_d)$, were computed from these deconvolved images for each case. For the affinity images only, the colormap (grayscale) is set to $[0, 0.5]/\mu\text{m}$ for lines A and D and to $[0, 0.05]/\mu\text{m}$ for lines B and C to facilitate visual comparison.

the K_d value. Finally, our procedure involves considerable computation as compared with the use of cell-averaging methods.

While FRET imaging is not a direct measure of the molecular interaction, it is one of the better means of visualizing protein interactions in the native environment of the cell and hence the ability to use FRET data to quantify the protein–protein interactions at the subcellular scale is significant. Obtaining values of protein–protein-binding affinities may allow meaningful comparisons between the effects of different drugs or inhibitors, giving useful insights into the mechanisms of their action. In addition, quantitative values of protein–protein-binding affinities are important for reconstructing protein networks inside the cell. Analysis of FRET imaging data with the methods described here might be further extended to analyze time course image data for the kinetic parameters of protein–protein interactions.

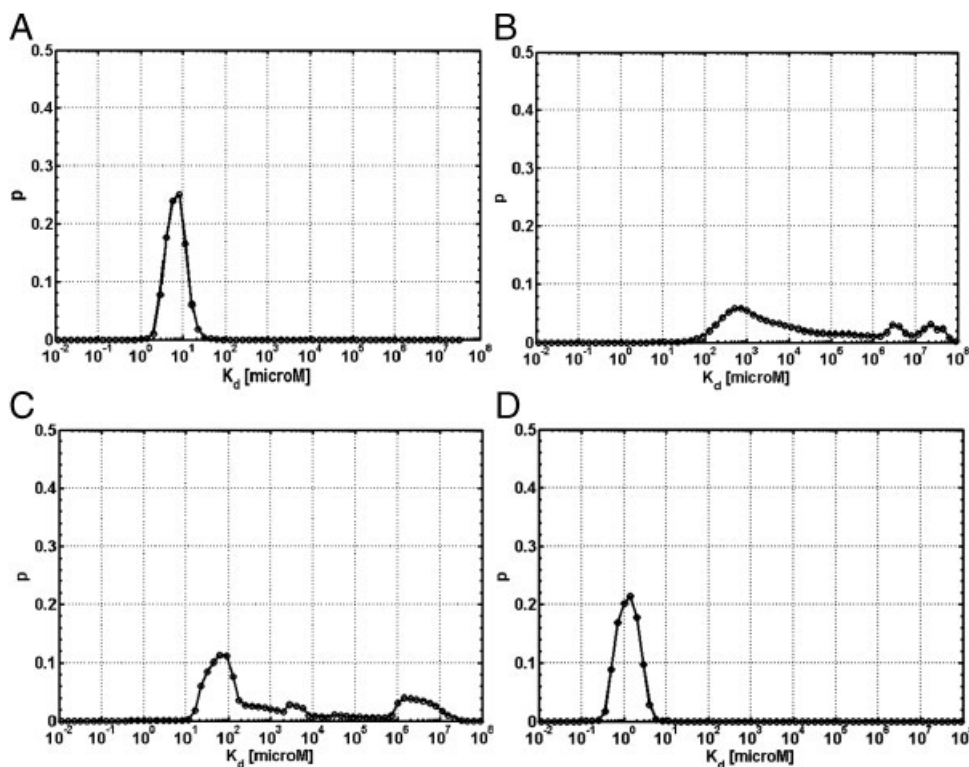


Figure 10. Inferring K_d from 3-D FSR images of mammalian cells. Inferred K_d probability distributions calculated for all four cases (Figs. 9A–D) are shown. (A) YFP-Rac2(V12)+CFP-PBD, (B) YFP-Rac2(V12)+CFP. (C) YFP+CFP. (D) CFP-YFP. Probability distributions were constructed using a histogram method with bins of size $0.15\ \mu\text{M}$ on the log scale. The curves shown for (A) and (B) are constructed from the values of three different experiments. The threshold value was fixed at $\Gamma = 0.1$. The mean value of K_d calculated from the distribution above based on calculations from individual pixels are 6.4, 1720, 129 and $1.19\ \mu\text{M}$, for the four cases, respectively. It should be noted that the distributions for (A–D) are not normal, and hence the mean values are not representative of the distribution.

Finally, FRET imaging can be performed with high 3-D resolution over time inside living cells. The terabytes of image data produced by these technologies will far outstrip human capacity to interpret, digest or analyze biochemical pathways. Thus, the development of analysis tools to infer key biophysical quantities from these image data, such as protein-binding affinities described here, will be essential.

This work was supported by US Army Research Laboratories and Research Office grant DAAD 19-03-1-0168. The authors thank the anonymous reviewers for their useful comments and suggestions.

The authors have declared no conflict of interest

5 References

- [1] Shoemaker, B. A., Panchenko, A. R., Deciphering protein-protein interactions. Part I. Experimental techniques and databases. *PLoS Comput. Biol.* 2007, **3**, e42.
- [2] Selbach, M., Mann, M., Protein interaction screening by quantitative immunoprecipitation combined with knock-down (QUICK). *Nat. Methods* 2006, **3**, 981–983.
- [3] Kerppola, T. K., Complementary methods for studies of protein interactions in living cells. *Nat. Methods* 2006, **3**, 969–971.
- [4] Chen, H., Puhl, H. L., III, Ikeda, S. R., Estimating protein-protein interaction affinity in living cells using quantitative Forster resonance energy transfer measurements. *J. Biomed. Opt.* 2007, **12**, 054011.
- [5] Piehler, J., New methodologies for measuring protein interactions *in vivo* and *in vitro*. *Curr. Opin. Struct. Biol.* 2005, **15**, 4–14.
- [6] You, X., Nguyen, A. W., Jabaiah, A., Sheff, M. A. *et al.*, Intracellular protein interaction mapping with FRET hybrids. *Proc. Natl. Acad. Sci. USA* 2006, **103**, 18458–18463.
- [7] Fricker, M., Runions, J., Moore, I., Quantitative fluorescence microscopy: from art to science. *Annu. Rev. Plant Biol.* 2006, **57**, 79–107.
- [8] Fernandez-Gonzalez, R., Munoz-Barrutia, A., Barcellos-Hoff, M. H., Ortiz-de-Solorzano, C., Quantitative *in vivo* microscopy: the return from the 'omics'. *Curr. Opin. Biotechnol.* 2006, **17**, 501–510.
- [9] Thaler, C., Koushik, S. V., Blank, P. S., Vogel, S. S., Quantitative multiphoton spectral imaging and its use for measuring resonance energy transfer. *Biophys. J.* 2005, **89**, 2736–2749.
- [10] Lippincott-Schwartz, J., Snapp, E., Kenworthy, A., Studying protein dynamics in living cells. *Nat. Rev. Mol. Cell. Biol.* 2001, **2**, 444–456.
- [11] Sako, Y., Imaging single molecules in living cells for systems biology. *Mol. Syst. Biol.* 2006, **2**, 56.
- [12] Kenworthy, A. K., Imaging protein-protein interactions using fluorescence resonance energy transfer microscopy. *Methods* 2001, **24**, 289–296.
- [13] Lakowicz, J., *Principles of Fluorescence Spectroscopy*, Plenum, New York 1999.
- [14] Swedlow, J. R., Quantitative fluorescence microscopy and image deconvolution. *Methods Cell. Biol.* 2007, **81**, 447–465.

- [15] McNally, J. G., Karpova, T., Cooper, J., Conchello, J. A., Three-dimensional imaging by deconvolution microscopy. *Methods* 1999, 19, 373–385.
- [16] Wu, J. Q., Pollard, T. D., Counting cytokinesis proteins globally and locally in fission yeast. *Science* 2005, 310, 310–314.
- [17] Gordon, G. W., Berry, G., Liang, X. H., Levine, B., Herman, B., Quantitative fluorescence resonance energy transfer measurements using fluorescence microscopy. *Biophys. J.* 1998, 74, 2702–2713.
- [18] Hoppe, A., Christensen, K., Swanson, J. A., Fluorescence resonance energy transfer-based stoichiometry in living cells. *Biophys. J.* 2002, 83, 3652–3664.
- [19] Hoppe, A., Anonymous *Imaging Cellular and Molecular Biological Functions*, 2007, pp. 157–181.
- [20] van Rheenen, J., Langeslag, M., Jalink, K., Correcting confocal acquisition to optimize imaging of fluorescence resonance energy transfer by sensitized emission. *Biophys. J.* 2004, 86, 2517–2529.
- [21] Chen, H., Puhl, H. L., III, Koushik, S. V., Vogel, S. S., Ikeda, S. R., Measurement of FRET efficiency and ratio of donor to acceptor concentration in living cells. *Biophys. J.* 2006, 91, L39–L41.
- [22] Raicu, V., Jansma, D. B., Miller, R. J., Friesen, J. D., Protein interaction quantified *in vivo* by spectrally resolved fluorescence resonance energy transfer. *Biochem. J.* 2005, 385, 265–277.
- [23] Chen, Y., Periasamy, A., Intensity range based quantitative FRET data analysis to localize protein molecules in live cell nuclei. *J. Fluoresc.* 2006, 16, 95–104.
- [24] Berney, C., Danuser, G., FRET or no FRET: a quantitative comparison. *Biophys. J.* 2003, 84, 3992–4010.
- [25] Wlodarczyk, J., Woehler, A., Kobe, F., Ponimaskin, E. *et al.*, Analysis of FRET signals in the presence of free donors and acceptors. *Biophys. J.* 2008, 94, 986–1000.
- [26] Hoppe, A. D., Shorte, S. L., Swanson, J. A., Heintzmann, R., 3D-FRET reconstruction microscopy for analysis of dynamic molecular interactions in live cells. *Biophys. J.* 2008, 13, 13.
- [27] Fink, C., Morgan, F., Loew, L. M., Intracellular fluorescent probe concentrations by confocal microscopy. *Biophys. J.* 1998, 75, 1648–1658.
- [28] Garini, Y., Young, I. T., McNamara, G., Spectral imaging: principles and applications. *Cytometry A* 2006, 69, 735–747.
- [29] Hoppe, A. D., Swanson, J. A., Cdc42, Rac1, and Rac2 display distinct patterns of activation during phagocytosis. *Mol. Biol. Cell* 2004, 15, 3509–3519.
- [30] Bokoch, G. M., Biology of the p21-activated kinases. *Annu. Rev. Biochem.* 2003, 72, 743–781.
- [31] DerMardirossian, C., Schnelzer, A., Bokoch, G. M., Phosphorylation of RhoGDI by Pak1 mediates dissociation of Rac GTPase. *Mol. Cell* 2004, 15, 117–127.
- [32] Huh, W. K., Falvo, J. V., Gerke, L. C., Carroll, A. S. *et al.*, Global analysis of protein localization in budding yeast. *Nature* 2003, 425, 686–691.
- [33] Ghaemmaghami, S., Huh, W. K., Bower, K., Howson, R. W. *et al.*, Global analysis of protein expression in yeast. *Nature* 2003, 425, 737–741.
- [34] Hoppe, A. D., Development of quantitative FRET microscopy for study of RHO GTPase and phosphoinositide signaling in phagocytosis. Ph.D. Thesis, 2003, pp. 116–118.
- [35] Lleres, D., Swift, S., Lamond, A. I., Detecting protein-protein interactions *in vivo* with FRET using multiphoton fluorescence lifetime imaging microscopy (FLIM). *Curr. Protoc. Cytom.* 2007, Chapter 12, Unit12.10.
- [36] Buranachai, C., Kamiyama, D., Chiba, A., Williams, B. D., Clegg, R. M., Rapid frequency-domain FLIM spinning disk confocal microscope: lifetime resolution, image improvement and wavelet analysis. *J. Fluoresc.* 2008, 18, 929–942.