Application of the Nano-Positioning System to the Analysis of Fluorescence Resonance Energy Transfer Networks
—Supplementary Material—

Adam Muschielok[†]

and

Jens Michaelis[*,†,‡]

jens.michaelis@uni-ulm.de, tel: +49 731 50 23050, fax: +49 731 50 23059

[*] corresponding author

[†] Ludwig-Maximilians-University Munich, Chemistry Department
Butenandtstrasse 11, 81377 Munich, Germany

[‡] Ulm University, Physics Department, Institute for Experimental Biophysics
Albert-Einstein-Allee 11, 89081 Ulm, Germany

1 Supplementary theory

1.1 The likelihood in the case of ambiguous axial depolarizations

Here, we treat the general case of ambiguous axial fluorescence depolarizations $\langle d_i^x \rangle$, which occur if the residual fluorescence anisotropies are small enough so that Equation (1) in the main text possesses two solutions. This happens if $r_{i,\infty}/r_{i,0} < 1/4$ and hence the sign of the axial depolarization $s_i := \mathrm{sgn}(\langle d_i^x \rangle)$ is ambiguous.[1]

As the axial depolarization is a physical property of each dye, we must account for each possible combination of axial depolarizations of all dyes in the likelihood. In that case Equation (9) in the main text is generalized to

$$p\left(\{E_{ij}\},\{A_{ij}\}|\left\{\boldsymbol{x}_i,\boldsymbol{\Omega}_i\right\},I\right) = \frac{1}{\prod_k S_k}\sum_{\{s_k\}}\prod_{ij\in M} L_{ij}^{s_i s_j}\left(\boldsymbol{x}_i,\boldsymbol{\Omega}_i,\boldsymbol{x}_j,\boldsymbol{\Omega}_j\right), \tag{1}$$

where the ambiguities of $\mathcal{E}_{ij}$ and $\mathcal{A}_{ij}$ and hence also of L_{ij} (Equation (10) in the main text) are symbolized by the superscript $(.)^{s_i s_j}$. $\{s_k\}$ denotes one possible combination of the signs s_k, while $\prod_k S_k$ is the number of all possible combinations calculated by setting $S_k = 2$ when $\langle d_k^x \rangle$ is ambiguous and otherwise $S_k = 1$.

Although it is not possible to factorize equation (1) completely, even a partial factorization should be considered to speed up the calculation.

1.2 Average TDM orientation and reference frame orientation priors

A flat prior in the average TDM orientation $\mathbf{\Omega}_i = (-\cos\theta_i, \phi_i)$ of the fluorophore i,

$$p\left(\mathbf{\Omega}_i|I\right) = p\left(-\cos\theta_i, \phi_i|I\right) = (2\pi)^{-1} \text{ for } -1 \leq -\cos\theta_i < 1 \text{ and } 0 \leq \phi_i < \pi, \qquad (2)$$

does not favor any particular orientation. θ_i and ϕ_i are the polar and azimuthal angle, respectively (main text, Figure 1a). This parametrization of $\mathbf{\Omega}_i$ was chosen because a flat prior in $-\cos\theta_i$ and ϕ_i is invariant under rotations of the laboratory reference frame.

In analogy, we assign a flat prior in the docking reference frame orientation $\mathbf{\Xi}_{(k)}$ parametrized by $\mathbf{\Xi}_{(k)} = (\rho_{(k)}, -\cos\theta_{(k)}, \phi_{(k)})$, where $\rho_{(k)}$, $\theta_{(k)}$ and $\phi_{(k)}$ are the angles of 3 subsequent rotations (main text, Figure 2b). The prior is constant within $-\pi \leq \rho_{(k)} < \pi$, $-\pi \leq \phi_{(k)} < \pi$ and $-1 \leq -\cos\theta_{(k)} < 1$. This prior is also invariant under rotations of the laboratory coordinate system and does not favor any particular orientation.

1.3 *Position - Förster distance* NPS as a special case of *position - orientation* NPS

Given the antenna position $\boldsymbol{x}_a$ and average TDM orientation $\mathbf{\Omega}_a$, as well as the positions $\{\boldsymbol{x}_i\}$ and average TDM orientations $\{\mathbf{\Omega}_i\}$ of an arbitrary number of satellites, the *position - Förster distance* NPS can be derived in two steps: First, we express the average TDM orientations in terms of the Förster distances R_{ai} (eqs. (2) and (3) in the main text) and use $\{R_{ai}\}$ as model parameters together with the satellite and antenna positions. Second, in the respective prior, $p\left(\boldsymbol{x}_a, \{\boldsymbol{x}_i\}, \{R_{ai}\}|I\right)$, the approximation of

independent Förster distances is made, which, in reality, are coupled by the average TDM orientation $\boldsymbol{\Omega}_a$ of the antenna dye. By using this approximation, the prior can be factorized,

$$p\left(\boldsymbol{x}_a, \{\boldsymbol{x}_i\}, \{R_{ai}\}|I\right) \approx p\left(\boldsymbol{x}_a|I\right) \prod_i p\left(\boldsymbol{x}_i|I\right) p\left(R_{ai}|I\right), \tag{3}$$

where the Förster distance prior is given by

$$p\left(R_{ai}|I\right) = \int d\boldsymbol{\Omega}_a \int d\boldsymbol{\Omega}_i \; \delta\left(R_{ai} - R\left(R_{ai}^{\mathrm{iso}}, \boldsymbol{\Omega}_a, \boldsymbol{\Omega}_i\right)\right) p\left(\boldsymbol{\Omega}_a|I\right) p\left(\boldsymbol{\Omega}_i|I\right). \tag{4}$$

In the above equation $\delta(.)$ denotes the Dirac point measure. $R(.)$ is the Förster distance as function of the isotropic Förster distance, the average TDM orientations (Equation (4) in the main text), and the average TDM orientation priors are defined as in the *position - orientation* NPS. The independent Förster distance approximation enables us to do the analysis almost entirely on an analytical basis[2] by simplifying the calculation of the marginal antenna position posterior (main text, Equation (8)).

2 Supplementary methods

2.1 Calculation of the marginal position density of arbitrary points

The marginal posterior PDF of the "laboratory" coordinates $\boldsymbol{q}$ of an arbitrary point Q having the coordinates $\boldsymbol{q}^{(k)}$ in the docked reference frame k is calculated as follows

from the marginal posterior PDF of reference frame origins $\boldsymbol{o}_{(k)}$ and orientations $\boldsymbol{\Xi}_{(k)}$:

$$p\left(\boldsymbol{q}|\{E_{ij}\},\{A_{ij}\},I\right) = \int d\boldsymbol{o}_{(k)} \int d\boldsymbol{\Xi}_{(k)}\ \boldsymbol{\delta}\left[\boldsymbol{q} - T_{\boldsymbol{o}_{(k)},\boldsymbol{\Xi}_{(k)}}\left(\boldsymbol{q}^{(k)}\right)\right] p\left(\boldsymbol{o}_{(k)},\boldsymbol{\Xi}_{(k)}|\{E_{ij}\},\{A_{ij}\},I\right), \quad (5)$$

where $\boldsymbol{\delta}(.)$ is the Dirac point measure in three dimensions and $T_{\boldsymbol{o}_{(k)},\boldsymbol{\Xi}_{(k)}}(.)$ is the coordinate transformation converting the coordinates of the reference frame k into the laboratory reference frame.

The marginal posterior PDF of reference frame origins and orientations is computed by integrating out all fluorophore positions and average TDM orientations, as well as the positions and orientations of other reference frames,

$$p\left(\boldsymbol{o}_{(k)},\boldsymbol{\Xi}_{(k)}|\{E_{ij}\},\{A_{ij}\},I\right) = \int d\{\boldsymbol{x}_i^{(m)},\boldsymbol{\Omega}_i^{(m)}\} \int d\{\boldsymbol{o}_{(m\neq k)},\boldsymbol{\Xi}_{(m\neq k)}\}\ p\left(\left\{\boldsymbol{x}_i^{(m)},\boldsymbol{\Omega}_i^{(m)}\right\},\{\boldsymbol{o}_{(m)},\boldsymbol{\Xi}_{(m)}\}|\{E_{ij}\},\{A_{ij}\},I\right). \quad (6)$$

In the *position - orientation* NPS, the PDF at the left-hand side of Equation (5) is computed approximately by creating a three-dimensional histogram of samples $\boldsymbol{q}_i$ from $p\left(\boldsymbol{q}|\{E_{ij}\},\{A_{ij}\},I\right)$. The samples $\boldsymbol{q}_i$ are computed by transforming the coordinates $\boldsymbol{q}^{(k)}$ of the point Q with $T_{\boldsymbol{o}_{(k),i},\boldsymbol{\Xi}_{(k),i}}(.)$, which is defined for each posterior sample i by the position $\boldsymbol{o}_{(k),i}$ and orientation $\boldsymbol{\Xi}_{(k),i}$ of the k^{th} reference frame.

2.2 Credible volumes

It is easy to show that for a differentiable probability density $p\left(\boldsymbol{x}\right)$ satisfying $\nabla_x p\left(\boldsymbol{x}\right) \neq \boldsymbol{0}$ at almost every position $\boldsymbol{x}$ satisfying the condition $p\left(\boldsymbol{x}\right) > 0$, the credible volume at

the level P is bounded by an iso-surface $S_P := \{\boldsymbol{x}|p(\boldsymbol{x}) = p_P\}$ at probability density p_P, so that

$$P = \int\limits_{p(\boldsymbol{x})\geq p_P} p(\boldsymbol{x}) d\boldsymbol{x}. \tag{7}$$

Instead of using $p(\boldsymbol{x})$ and p_P that still depend on the PDF, we utilize the *cumulative distribution function* $c(\boldsymbol{x})$ and define it as the probability accumulated in the parameter space region where the PDF exceeds the value $p(x)$,

$$c(\boldsymbol{x}) = \int\limits_{p(\boldsymbol{x}')\geq p(\boldsymbol{x})} p(\boldsymbol{x}') d\boldsymbol{x}'. \tag{8}$$

An iso-surface S_P of the cumulative distribution function at the level P, i.e.

$$S_P = \{\boldsymbol{x}|c(\boldsymbol{x}) = P\}, \tag{9}$$

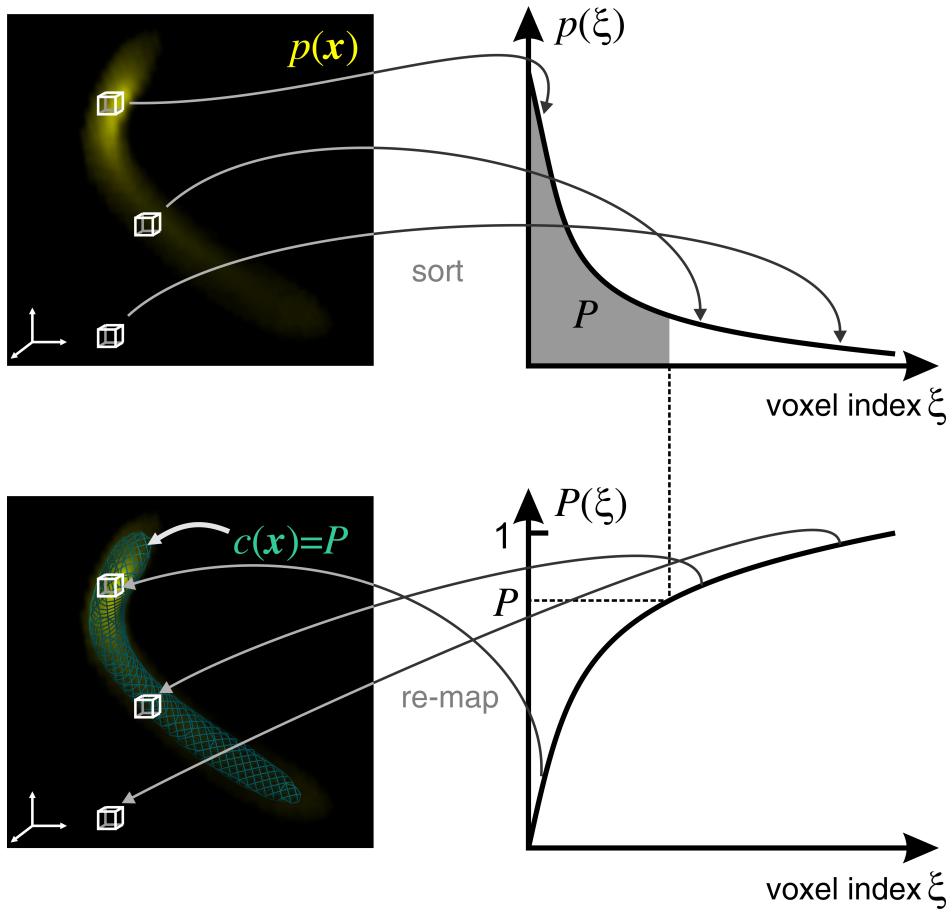
encloses the smallest volume containing the probability P, which is equivalent to the credible volume.

In order to display a credible volume we computed $c(\boldsymbol{x})$ of the probability density $p(\boldsymbol{x})$ at the supporting points $\boldsymbol{x} = \boldsymbol{x}_m$. $\boldsymbol{x}_m$ were spaced on a cubic grid with constant spacing δx, and the size of the spacing was chosen so that the displayed iso-surfaces were looking smooth. Here, δx varied between 0.5 and 14 Å for the displayed densities.

To compute $c(\boldsymbol{x}_m)$, the probability density $p(\boldsymbol{x}_m)$ was sorted in descending order, so that

$$p(\xi) \equiv p\left(\boldsymbol{x}_{m(\xi)}\right) \geq p\left(\boldsymbol{x}_{m(\xi+1)}\right), \tag{10}$$

where ξ and m are the voxel indices after and before the sorting, respectively, $m(\xi)$ is the sorting operation, and $p(\xi)$ denotes the sorted probability density. Thereafter, the



Suppl. Figure 1: Computation of the cumulative distribution function. The position space (left) and the sorted voxel space (right) are shown. The cumulative distribution function $c(\boldsymbol{x})$ used to display credible volumes is calculated from the position probability density $p\,(\boldsymbol{x})$ (yellow fog) by sorting, calculation of cumulative probabilities, and re-mapping of the cumulative probabilities into the position space. The sorted probability density, $p\,(\xi)$, and its cumulative distribution, $P(\xi)$, are shown on the right side as functions of the voxel index ξ. Three voxels in the position space are symbolized by white cubes. The cumulative distribution function $c(\boldsymbol{x})$ contoured at the level P (cyan mesh) is the surface of the credible volume corresponding to the probability P.

cumulative sorted probabilities, $P(\xi)$, defined as

$$P(\xi) = (\delta x)^3 \sum_{\xi'=1}^{\xi} p\,(\xi')\,, \tag{11}$$

were calculated and mapped back into the three-dimensional position space (Suppl. Figure 1).

2.3 Displaying of NPS calculations

Marginal posterior PDFs of fluorophore positions were saved in MRC/CCP4 format (referred to as MAPFORMAT in[3]) and displayed as "solid" (i.e. with an intensity proportional to the PDF) with the UCSF Chimera software package[4,5].

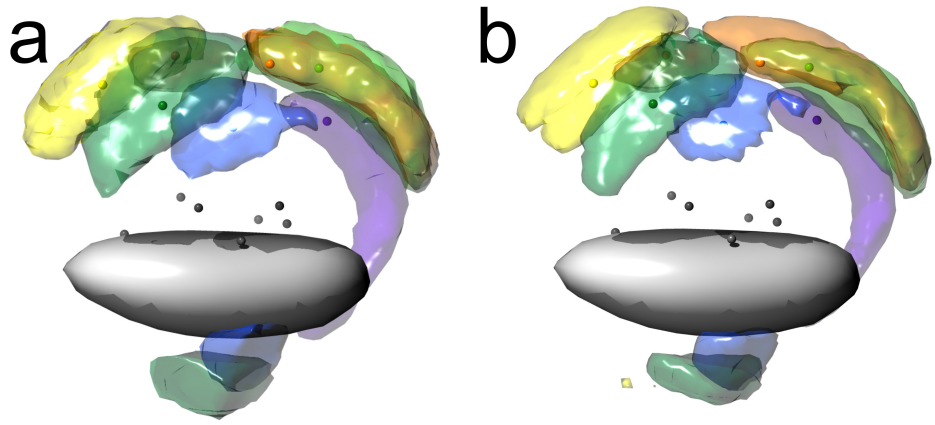
For better visualization of the PDFs, the credible volumes (Suppl. methods 2.2) at the level P were shown either as "surface" or "mesh" in Chimera. To this end we computed the cumulative distribution function $c(\boldsymbol{x}_m)$ of the marginal posterior of a fluorophore position or, alternatively, of the PDF of an arbitrary point $\boldsymbol{q}^{(k)}$ in a docked reference frame. We saved $c(\boldsymbol{x}_m)$ in the MRC/CCP4 format and displayed the credible volume at level P by showing an iso-surface of $c(\boldsymbol{x}_m)$ at the level P in Chimera.

2.4 Quantification of average TDM orientation estimates

The average TDM orientation was calculated directly from reweighted posterior samples obtained by nested sampling. In order to deal with spherical topology the average orientation of the TDM of the fluorophore i of each sample was represented as the 3-dimensional coordinates of the two intersections of the line defined by the average TDM orientation and the unit sphere. By construction, the center of mass of these points $\{\boldsymbol{s}_l\}$ is the origin. In order to determine the *mean average TDM orientation* $\langle\boldsymbol{\Omega}_i\rangle$ and the corresponding orientation uncertainty, we first computed the covariance matrix $C_i^{\mathrm{ori}} = \langle\boldsymbol{s}_l\boldsymbol{s}_l^{\mathrm{T}}\rangle$, diagonalized it, and obtained 3 principal directions and variances. We defined $\langle\boldsymbol{\Omega}_i\rangle$ as the principal direction with the largest variance, and obtained two standard deviations $\sigma_1^{\boldsymbol{\Omega}_i}$ and $\sigma_2^{\boldsymbol{\Omega}_i}$ in the remaining principal directions, which quantify the uncertainty of $\langle\boldsymbol{\Omega}_i\rangle$. In order to summarize the uncertainty of $\langle\boldsymbol{\Omega}_i\rangle$ in one number, we computed the geometric mean $\overline{\sigma^{\boldsymbol{\Omega}_i}} = (\sigma_1^{\boldsymbol{\Omega}_i}\sigma_2^{\boldsymbol{\Omega}_i})^{1/2}$. By averaging $\overline{\sigma^{\boldsymbol{\Omega}_i}}$ over all antennas

of the FRET network, we computed the *mean average TDM orientation uncertainty* $\langle\overline{\sigma^{\mathbf{\Omega}}}\rangle$, which was used to compare the different analysis scenarios based on the spread of the average TDM orientation estimates (Figure 5 in the main text).

3 Supplementary results



Suppl. Figure 2: Influence of combined FRET efficiency and anisotropy measurements in separate NPS. In separate analysis of the FRET network, the additional measurement of FRET anisotropy improved localization accuracy only slightly (b) when compared to the results obtained by separately analyzing FRET efficiency data only (a). The mean localization uncertainty $\langle\overline{\sigma^{x}}\rangle$ was 11.8 Å in (a) and 9.5 Å in (b).

4 Supplementary discussion

4.1 Förster distance priors in *position - Förster distance* NPS

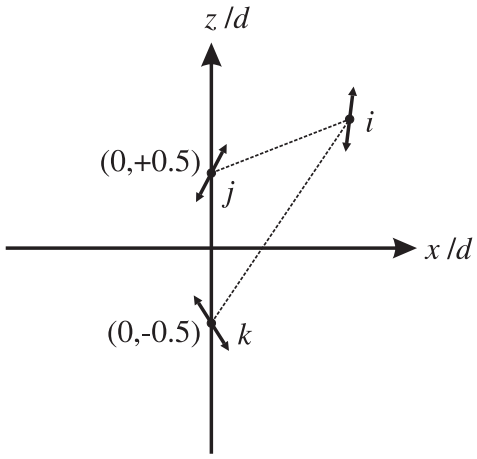
In the derivation of the *position - Förster distance* NPS[2] we assumed that the Förster distances of different satellite/antenna pairs are mutually independent and were thus able to calculate the marginal antenna position posterior almost entirely on an analytical basis. However, this assumption needs to be tested as the Förster distances R_{ij} and R_{ik} of two FRET efficiency measurements between the antenna i and the satellites j and k depend on the position and average TDM orientation of the antenna, and hence should be correlated.

To test our assumption we computed the normalized mutual information (NMI)[6] $\mathcal{I}(R_{ij}, R_{ik})$ of the distribution of Förster distances
$p\left(R_{ij}, R_{ik}|\boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{x}_k, I\right)$ of two FRET pairs that share one fluorophore (Suppl. Figure 3). The NMI is a measure of the statistical dependency of two random variables, here R_{ij} and R_{ik}, and ranges from 0 (completely independent) to 1 (completely dependent).

In more detail, the NMI $\mathcal{I}(X, Y)$ is the normalized dissimilarity of the joint distribution of $p\left(X, Y|I\right)$ compared to a reconstruction from the marginal distributions $p\left(X|I\right)p\left(Y|I\right)$. It is computed by means of the Kullback-Leibler divergence $D_{\mathrm{KL}}[.||.]$[7] and normalized with the geometric mean of the Shannon informations $H[.]$ of the marginal distributions.

Explicitly, the NMI for the Förster distances R_{ij} and R_{ik} of two FRET pairs with a common fluorophore i is given by

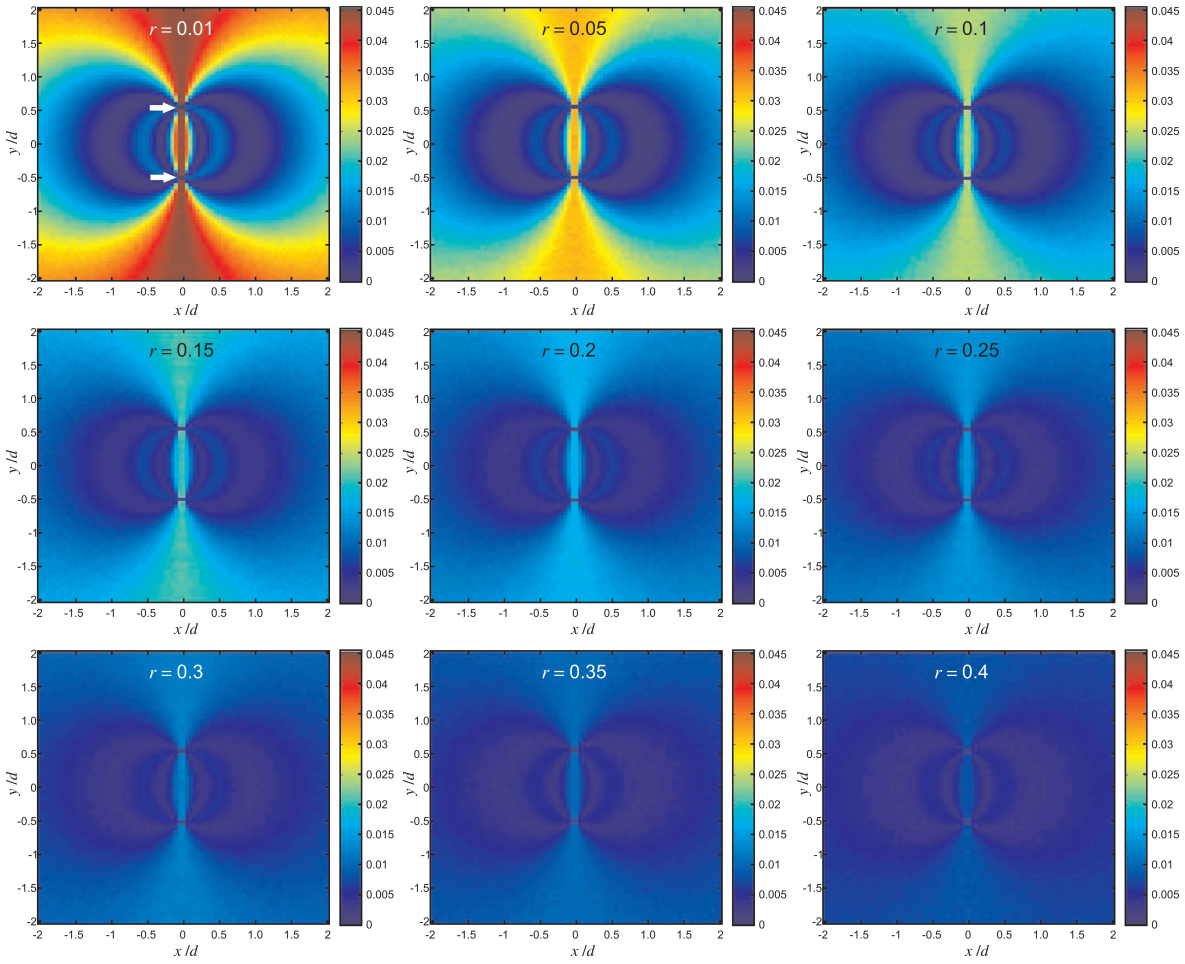
$$\mathcal{I}(R_{ij}, R_{ik}) = \frac{D_{\mathrm{KL}}\left[p\left(R_{ij}, R_{ik}|\boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{x}_k, I\right) \;||\; p\left(R_{ij}|\boldsymbol{x}_i, \boldsymbol{x}_j, I\right)p\left(R_{ik}|\boldsymbol{x}_i, \boldsymbol{x}_k, I\right)\right]}{\sqrt{H\left[p\left(R_{ij}|\boldsymbol{x}_i, \boldsymbol{x}_j, I\right)\right]H\left[p\left(R_{ik}|\boldsymbol{x}_i, \boldsymbol{x}_k, I\right)\right]}} \quad (12)$$



Suppl. Figure 3: Simulation of the joint Förster distance density. We investigated the Förster distances R_{ij} and R_{ik} of the FRET efficiencies measured between three fluorophores, i, j and k (double arrows) constituting two FRET pairs (FRET efficiency measurements are shown as dashed lines). Both Förster distances R_{ij} and R_{ik} are correlated by the position and average TDM orientation of the dye i. In the simulation it was assumed that the isotropic Förster distances of both FRET efficiency measurements are equal, $R_{ij}^{\text{iso}} = R_{ik}^{\text{iso}} = R^{\text{iso}}$. The positions of the fluorophores j and k was kept fixed and their distance was set to d, while the position of the third fluorophore, i, was varied in the (x, z)-plane. To obtain the density $p\left(R_{ij}, R_{ik}|\boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{x}_k, I\right)$ conditional only on the fluorophore positions, the density $p\left(R_{ij}, R_{ik}|\boldsymbol{x}_i, \boldsymbol{\Omega}_i, \boldsymbol{x}_j, \boldsymbol{\Omega}_j, \boldsymbol{x}_k, \boldsymbol{\Omega}_k, I\right)$, which is conditional on the average transition dipole moment orientations, was marginalized with a uniform prior in the orientations.

We computed $\mathcal{I}(R_{ij}, R_{ik})$ given various fluorophore positions $\boldsymbol{x}_i, \boldsymbol{x}_j$ and $\boldsymbol{x}_k$, fluorescence anisotropies $r_{\infty,i}, r_{\infty,j}$ and $r_{\infty,k}$, and under the assumption of an isotropic prior for the average transition dipole orientations of the fluorophores. We assumed further that the fluorescence anisotropies are equal, i.e. $r_{\infty,i} = r_{\infty,j} = r_{\infty,k} = r$. In the simulations, a maximum NMI of $5 \cdot 10^{-2}$ was achieved when fluorescence anisotropies were very small ($r = 0.01$) and orientation effects should be negligible (Suppl. Figure 4). For larger anisotropies ($r = 0.05, \ldots, 0.4$) the NMI ranged from 0.025 to 0.009 and indicated that Förster distances coupled by one fluorophore only are statistically almost independent as assumed in the *position - Förster distance* NPS (Suppl. Figure 4). This finding clearly breaks down when a complete FRET network is analyzed globally as we observed by

improvement of the localization accuracy in the globally analyzed data (see results & discussion in the main text).



Suppl. Figure 4: Normalized mutual information (NMI), $\mathcal{I}(R_{ij}, R_{ik})$, of two correlated Förster distances, R_{ij} and R_{ik}. The dependence of the NMI on the position $\boldsymbol{x}_i = (x, 0, z)$ of the fluorophore i is shown in each panel. The residual fluorescence anisotropies of all fluorophores were set to the same value r, i.e. $r_{\infty,i} = r_{\infty,j} = r_{\infty,k} = r$, which was varied from 0.01 to 0.4.

4.2 Coupling of NPS with FRET data pre-processing

In the context of NPS, pre-processing of FRET data one needs to extract FRET efficiencies and FRET anisotropies of molecular species from the measured data, and one must know the number of conformational states of the studied macromolecule. At the same time, also other important quantities like transition rates between different macromolecular conformations can be inferred by pre-processing, which cannot be analyzed by NPS, but nevertheless, this can help to determine the number of states. It is thus necessary to derive how the results of pre-processing approaches and NPS can be used together in order to determine the correct number of conformational states.

4.2.1 General derivation

In this section, we address formally how pre-processing of FRET data can be coupled to NPS. After having performed the measurements usually one is confronted with the problem of how many states are needed to explain the experimental data. This question is crucial as it usually will have a large impact on the biological interpretation of the experiment.

In Bayesian data analysis this problem is referred to as "model selection".[8 Ch.4] One can propose several models explaining the observations and compute the probability of each model given the measured data. Based thereon one can decide whether the data is sufficient to support mainly one of the initially proposed models, or whether more experiments should be performed in order to answer this question unambiguously.

In our case, the measured FRET data is described by using the model M, in which the macromolecule and the attached fluorophores can exist in K_M states. The model probability given the experimental data will be denoted as $p\left(M|\{O_{ij}(t)\}, I\right)$. Here we assume that the data is usually available as several time series $O_{ij}(t)$ of observables

measured for the fluorophore pairs ij. The observables are in our case FRET efficiency or/and FRET anisotropy, and by using Bayes theorem we can relate $p\left(M|\{O_{ij}(t)\},I\right)$ and the evidence $p\left(\{O_{ij}(t)\}|M,I\right)$,

$$p\left(M|\{O_{ij}(t)\},I\right) \propto p\left(\{O_{ij}(t)\}|M,I\right) p\left(M|I\right) = Z_M \; p\left(M|I\right), \tag{13}$$

where $p\left(M|I\right)$ is the prior probability of the model, $Z_M := p\left(\{O_{ij}(t)\}|M,I\right)$ is its evidence. Usually, $p\left(M|I\right)$ is assumed to be constant, assuming that initially each model is equally likely, so that the posterior probability of the model is proportional to the evidence Z_M, which can be calculated by integration of the posterior over the model parameters,

$$Z_M = \int d\{\boldsymbol{x}_{i,k}\} \int d\{\boldsymbol{\Omega}_{i,k}\} \sum_{\{s_{i,k}\}} p\left(\{O_{ij}(t)\}|\{s_i\},\{\boldsymbol{x}_{i,k}\},\{\boldsymbol{\Omega}_{i,k}\},M,I\right) \times p\left(\{s_{i,k}\},\{\boldsymbol{x}_{i,k}\},\{\boldsymbol{\Omega}_{i,k}\}|M,I\right). \tag{14}$$

In the above equation $\{s_{i,k}\}$ are the signs of the depolarizations and $\{\boldsymbol{x}_{i,k}\}$ and $\{\boldsymbol{\Omega}_{i,k}\}$ are the positions and average TDM orientations of the dyes in state k.

We will show now that the likelihood in equation 14 can be obtained from the result of commonly used data pre-processing procedures. We assume that the model used for pre-processing comprises the vector of additional hidden parameters $\boldsymbol{h}_M$ and the "true" values $\{\tilde{O}_{ij,k}\}$, i.e. the set of the observables (i.e. FRET efficiencies and FRET anisotropies) of all measured FRET pairs ij in all states $k = 1,\ldots,K_M$. The hidden parameters denote all model parameters except the NPS observables FRET efficiency and FRET anisotropy, like for example the widths of peaks in histograms or the transition rates of a hidden Markov model. With these definitions we introduce the pre-processing

parameters $\{\tilde{O}_{ij,k}\}$ and $\boldsymbol{h}_M$ into the likelihood $p\left(\{O_{ij}(t)\}|\{s_{i,k}\},\{\boldsymbol{x}_{i,k}\},\{\boldsymbol{\Omega}_{i,k}\},M,I\right)$ by marginalizing the joint likelihood of $\{O_{ij}(t)\}$ and the pre-processing parameters, and by subsequently applying Bayes' theorem,

$$\begin{aligned} p\left(\{O_{ij}(t)\}|\{s_{i,k}\},\{\boldsymbol{x}_{i,k}\},\{\boldsymbol{\Omega}_{i,k}\},M,I\right) = \\ \int d\{\tilde{O}_{ij,k}\}\int d\boldsymbol{h}_M p\left(\{O_{ij}(t)\},\{\tilde{O}_{ij,k}\},\boldsymbol{h}_M|\{s_{i,k}\},\{\boldsymbol{x}_{i,k}\},\{\boldsymbol{\Omega}_{i,k}\},M,I\right) = \\ \int d\{\tilde{O}_{ij,k}\}\int d\boldsymbol{h}_M p\left(\{O_{ij}(t)\}|\{\tilde{O}_{ij,k}\},\boldsymbol{h}_M,\{s_{i,k}\},\{\boldsymbol{x}_{i,k}\},\{\boldsymbol{\Omega}_{i,k}\},M,I\right) \\ \times p\left(\{\tilde{O}_{ij,k}\},\boldsymbol{h}_M|\{s_{i,k}\},\{\boldsymbol{x}_{i,k}\},\{\boldsymbol{\Omega}_{i,k}\},M,I\right). \quad (15) \end{aligned}$$

Since the data $\{O_{ij}(t)\}$ depends directly only on the "true" observables $\{\tilde{O}_{ij,k}\}$ (but not on the NPS parameters) we can cancel the NPS parameters in the third line of equation (15). If $\{\tilde{O}_{ij,k}\}$ and $\boldsymbol{h}_M$ are independent a priori, we can factorize the probability in the last line of equation (15), and if the hidden parameters do not depend on $\{s_{i,k}\}$, $\{\boldsymbol{x}_{i,k}\}$ and $\{\boldsymbol{\Omega}_{i,k}\}$ a priori, we can finally write

$$\begin{aligned} p\left(\{O_{ij}(t)\}|\{s_{i,k}\},\{\boldsymbol{x}_{i,k}\},\{\boldsymbol{\Omega}_{i,k}\},M,I\right) = \\ \int d\{\tilde{O}_{ij,k}\}\int d\boldsymbol{h}_M p\left(\{O_{ij}(t)\}|\{\tilde{O}_{ij,k}\},\boldsymbol{h}_M,M,I\right) \\ \times p\left(\boldsymbol{h}_M|M,I\right) p\left(\{\tilde{O}_{ij,k}\}|\{s_{i,k}\},\{\boldsymbol{x}_{i,k}\},\{\boldsymbol{\Omega}_{i,k}\},M,I\right). \quad (16) \end{aligned}$$

The term $p\left(\{\tilde{O}_{ij,k}\}|\{s_{i,k}\},\{\boldsymbol{x}_{i,k}\},\{\boldsymbol{\Omega}_{i,k}\},M,I\right)$ in the above equation can be identified with a product of Dirac δ distributions since the "true" value of the observables must be identical with the observables $\mathcal{O}_{ij,k}$ (i.e. $\mathcal{E}$ or/and $\mathcal{A}$) expected from the depolarization signs as well as from the fluorophore positions and orientations. We will

abbreviate this term by $\boldsymbol{\delta}(\{\tilde{O}_{ij,k}\} - \{\mathcal{O}_{ij,k}\})$, which is defined by

$$\boldsymbol{\delta}(\{\tilde{O}_{ij,k}\} - \{\mathcal{O}_{ij,k}\}) = \prod_{ij,k} \delta\left(\tilde{O}_{ij,k} - \mathcal{O}'_{ij,k}\right), \tag{17}$$

so that we get the following expression for the likelihood:

$$\begin{aligned} &p\left(\{O_{ij}(t)\}|\{s_{i,k}\}, \{\boldsymbol{x}_{i,k}\}, \{\boldsymbol{\Omega}_{i,k}\}, M, I\right) = \\ &\int d\{\tilde{O}_{ij,k}\} \int d\boldsymbol{h}_M p\left(\{O_{ij}(t)\}|\{\tilde{O}_{ij,k}\}, \boldsymbol{h}_M, M, I\right)\ p\left(\boldsymbol{h}_M|M, I\right)\ \boldsymbol{\delta}(\{\tilde{O}_{ij,k}\} - \{\mathcal{O}_{ij,k}\}). \end{aligned} \tag{18}$$

It is important to note that by doing so we implicitly assumed that we can uniquely assign each molecular state k to a set of "true" observables $\tilde{O}_{ij,k}$. This is for example the case if each state can be uniquely identified by the average fraction of the time a molecule spends in it, e.g. as in every experiment ij a small and a large fraction of molecules is observed. If this is not possible, as for instance the population of two or more states are very similar in all measured pairs of labeling sites ij, one must treat every possible assignment of the "true" observables $\tilde{O}_{ij,k}$ obtained in the pre-processing to the observables $\mathcal{O}_{ij,k}$ expected from NPS as a separate model.

Next, we discuss the special case of a pre-processing method, in which a flat prior $p\left(\tilde{O}_{ij,k}|M, I_{\mathrm{PP}}\right) =: \pi_M^{\mathrm{PP},\tilde{O}}$ is used in the "true" observables. In this case, we can use Bayes theorem and rewrite Equation 18 in order to introduce the posterior PDF of the data pre-processing problem, $p\left(\boldsymbol{h}_M, \{\tilde{O}_{ij,k}\}|\{O_{ij}(t)\}, M, I\right)$, together with the respective evidence, Z_M^{PP},

$$
\begin{aligned}
&p\left(\{O_{ij}(t)\}|\{s_{i,k}\},\{\boldsymbol{x}_{i,k}\},\{\boldsymbol{\Omega}_{i,k}\},M,I\right) = \\
&\quad \frac{Z_M^{\mathrm{PP}}}{\pi_M^{\mathrm{PP},\tilde{O}}} \int d\{\tilde{O}_{ij,k}\} \int d\boldsymbol{h}_M p\left(\boldsymbol{h}_M,\{\tilde{O}_{ij,k}\}|\{O_{ij}(t)\},M,I\right) \boldsymbol{\delta}(\{\tilde{O}_{ij,k}\}-\{\mathcal{O}_{ij,k}\}) = \\
&\qquad \frac{Z_M^{\mathrm{PP}}}{\pi_M^{\mathrm{PP},\tilde{O}}} \int d\{\tilde{O}_{ij,k}\} p\left(\{\tilde{O}_{ij,k}\}|\{O_{ij}(t)\},M,I\right) \boldsymbol{\delta}(\{\tilde{O}_{ij,k}\}-\{\mathcal{O}_{ij,k}\}) = \\
&\qquad\qquad \frac{Z_M^{\mathrm{PP}}}{\pi_M^{\mathrm{PP},\tilde{O}}}\, p\left(\{\tilde{O}_{ij,k}\}|\{O_{ij}(t)\},M,I\right)\Big|_{\{\tilde{O}_{ij,k}\}=\{\mathcal{O}_{ij,k}\}}. \quad (19)
\end{aligned}
$$

As we see in the above equation, the likelihood of the NPS problem can be obtained by weighting the posterior of the pre-processing problem by $Z_M^{\mathrm{PP}}/\pi_M^{\mathrm{PP},\tilde{O}}$. The likelihood derived here will replace the product over the likelihood factors in Equation (9) in the main text because there we did not yet treat the ambiguity introduced by the sign of the depolarizations. $p\left(\{\tilde{O}_{ij,k}\}|\{O_{ij}(t)\},M,I\right)\Big|_{\{\tilde{O}_{ij,k}\}=\{\mathcal{O}_{ij,k}\}}$ is also clearly multimodal, as it is invariant under arbitrary permutations of $(\mathcal{O}_{ij,1},\ldots,\mathcal{O}_{ij,K_M})$. This has to be accounted for by ab additional factor of $K_M!$ in the total evidence of the model M,[9] which is given by

$$
Z_M = \frac{Z_M^{\mathrm{PP}}}{\pi_M^{\mathrm{PP},\tilde{O}}}\; K_M!\; Z_M^{\mathrm{NPS}}. \quad (20)
$$

Z_M^{NPS} is the evidence of the NPS problem calculated (without accounting for the permutation symmetry) by using the function $p\left(\{\tilde{O}_{ij,k}\}|\{O_{ij}(t)\},M,I\right)\Big|_{\{\tilde{O}_{ij,k}\}=\{\mathcal{O}_{ij,k}\}}$ instead of the likelihood from Equation (9) in the main text. Using the likelihood above is important because in general there will be correlations between the observables of different states $\tilde{O}_{ij,k}$ for the same pair of labeling sites, which are not accounted for in equation (9) in the main text.

We stress again that equation (20) is valid for any pre-processing method working

with FRET efficiency and/or FRET anisotropy as model parameters and essentially assuming a flat prior in these observables. This applies particularly to least-squares histogram fits as well as to maximum likelihood estimates, and although these methods lack an explicit prior, they can be regarded as special cases of Bayesian data analysis. [8 Ch.3.5].

4.2.2 Application to least-squares histogram fitting

In the following, we apply equation (20) to the popular pre-processing of data by least-squares fitting of FRET histograms. First, the data $O_{ij}(t)$ obtained in a FRET experiment between the labeling sites i and j is binned and displayed as a histogram, so that the time dependence is lost and does not need to be analyzed. However, we will keep the time dependence in the notation to indicate that $O_{ij}(t)$ consists of many data points.

In the model M, each histogram of $O_{ij}(t)$ is described by $N_M^{O,ij}$ FRET species (i.e. peaks in the histogram). Each peak k in the histogram is usually parameterized by a center position (i.e. $\tilde{E}_{ij,k}$ or $\tilde{A}_{ij,k}$) and a parameter modeling the width of the peak (i.e. $w_{ij,k}^{\tilde{E}}$ or $w_{ij,k}^{\tilde{A}}$). We will abbreviate the center positions and widths by $[\tilde{O}_{ij,k}] := (\tilde{O}_{ij,1}, \ldots, \tilde{O}_{ij,N_M^{O,ij}})$ and $[w_{ij,k}^{\tilde{O}}] := (w_{ij,1}^{\tilde{O}}, \ldots, w_{ij,N_M^{O,ij}}^{\tilde{O}})$, respectively[1]. The widths $[w_{ij,k}^{\tilde{O}}]$, which are not processed directly in the NPS, are treated as hidden parameters and are part of the vector of hidden parameters, $\boldsymbol{h}_K$ (see equations (18) and (19)).

We emphasize that the number of FRET species $N_M^{O,ij}$ does not necessarily have to be equal to the number of states of the macromolecule, K_M, since it is possible that several macromolecule conformations lead to the same value of the observable $\tilde{O}_{ij,k}$ because the position and average TDM orientation of the fluorophores stays the same.

[1] Note that the brackets denote a set of observables or widths, which extends over all states k and in which the FRET pair ij stays the same.

It is also possible that more than K_M FRET species are observed if experimental artifacts need to be accounted for, e.g. donor-only labeled molecules, which causes typically a peak at 0 FRET efficiency. Similar to the peak widths we will consider the model parameters needed to explain experimental artifacts as part of the "hidden" parameters $\boldsymbol{h}_M$. In the notation they will therefore not appear as "true" observables $\{\tilde{O}_{ij,k}\}$ in the following.

Next, we will discuss least-squares fitting of the histograms of $O_{ij}(t)$ from a Bayesian point of view. The likelihood of a Bayesian calculation corresponding to a least-square fit is given by[8 Ch.3.5]

$$p\left(O_{ij}(t)|[\tilde{O}_{ij,k}],[w^{\tilde{O}}_{ij,k}],M,I\right) \propto \exp\left[-\frac{\chi^2_{O,ij}}{2}\right], \tag{21}$$

where $\chi^2_{O,ij}$ is the sum of the normalized errors squared, here given by

$$\chi^2_{O,ij} = \sum_b \left(\frac{H_b^{O,ij} - F_b^{O,ij}}{\epsilon_b^{O,ij}}\right)^2. \tag{22}$$

In the above equation, the number of counts in the bin b of the histogram of the data is given by $H_b^{O,ij}$, and $F_b^{O,ij}$ denotes the average number of counts expected from the relevant model parameters $[\tilde{O}_{ij,k}]$ and $[w^{\tilde{O}}_{ij,k}]$. As each H_b^{ij} should follow a Poissonian distribution, the expected error in each bin, $\epsilon_b^{O,ij}$, can be assumed to be approximately $\epsilon_b^{O,ij} \propto \sqrt{F_b^{O,ij}} \approx \sqrt{H_b^{O,ij}}$ if all $H_b^{O,ij} \gg 1$.[8 Ch.3.5]

By minimizing $\chi^2_{O,ij}$ and therefore by performing a least-square fit we maximize the likelihood in equation (21). If we assume a flat prior $p\left([\tilde{O}_{ij,k}],[w^{\tilde{O}}_{ij,k}]|M,I\right)$ in all fit parameters, the posterior $p\left([\tilde{O}_{ij,k}],[w^{\tilde{O}}_{ij,k}]|O_{ij}(t),M,I\right)$ will be proportional to the likelihood, and thus minimizing $\chi^2_{O,ij}$ corresponds to finding the maximum of the

posterior. This holds approximately even for non-uniform priors if the prior is varying slowly as a function of the parameters in the proximity of the likelihood maximum, so that the main variability of the posterior can be attributed to the likelihood and the prior can be approximated by its nearly constant value at the likelihood maximum.

In the following, we use a flat prior in the peak center parameters, $[\tilde{O}_{ij,k}]$, and a Jeffreys prior[10] in the peak width parameters $[w^{\tilde{O}}_{ij,k}]$,

$$p\left([\tilde{O}_{ij,k}],[w^{\tilde{O}}_{ij,k}]|M,I\right) \propto \left[\prod_{k=1}^{N_M^{O,ij}} \left(\tilde{O}_{\max}-\tilde{O}_{\min}\right) w^{\tilde{O}}_{ij,k}\right]^{-1}. \tag{23}$$

In the above equation, the possible range of the "true" observable values $(\tilde{O}_{\min},\tilde{O}_{\max})$ is given by $\tilde{E}_{\min}=0$ and $\tilde{E}_{\max}=1$ for FRET efficiencies, and by $\tilde{A}_{\min}=-1/5$ and $\tilde{A}_{\max}=2/5$ for FRET anisotropies.

The evidence $Z_M^{O,ij}$ is given by the product of prior and likelihood integrated over the model parameters,

$$Z_M^{O,ij} \propto \int d[\tilde{O}_{ij,k}] \int d[w^{\tilde{O}}_{ij,k}] \left[\left[\prod_{k=1}^{N_M^{O,ij}} \left(\tilde{O}_{\max}-\tilde{O}_{\min}\right) w^{\tilde{O}}_{ij,k}\right]^{-1} \times \exp\left[-\frac{\chi^2_{O,ij}}{2}\right]\right], \tag{24}$$

and by assuming that the prior is approximately constant in the proximity of the likelihood maximum located at $[\tilde{O}^*_{ij,k}]$ and $[w^{*\tilde{O}}_{ij,k}]$ we can take the prior out of the integral and get finally

$$Z_M^{O,ij} \tilde{\propto} \left[\prod_{k=1}^{N_M^{O,ij}} \left(\tilde{O}_{\max}-\tilde{O}_{\min}\right) w^{*\tilde{O}}_{ij,k}\right]^{-1} \times \int d[\tilde{O}_{ij,k}] \int d[w^{\tilde{O}}_{ij,k}] \ \exp\left[-\frac{\chi^2_{O,ij}}{2}\right]. \tag{25}$$

Oftentimes it might be possible to approximate the posterior and, given a constant prior, also the likelihood $\exp\left[-\chi^2_{O,ij}/2\right]$ by a multivariate normal distribution, which

is characterized by its amplitude calculated from χ^2 evaluated at the minimum, $\chi^{2,\min}_{O,ij}$, and by its width, which can be calculated from the covariance matrix $C_{O,ij}$ [8 Ch.3.2 and App.3]. In this case, the evidence is given by

$$Z_M^{O,ij} \tilde{\propto} \left[\prod_{k=1}^{N_M^{O,ij}} \left(\tilde{O}_{\max} - \tilde{O}_{\min} \right) w_{ij,k}^{*\tilde{O}} \right]^{-1} (2\pi)^{N_M^{O,ij}} \sqrt{\det C_{O,ij}} \exp\left[-\frac{\chi^{2,\min}_{O,ij}}{2} \right]. \tag{26}$$

As the pre-processing of each histogram is clearly done independently, the evidence of the complete pre-processing, Z_M^{PP}, can be calculated from the product of all $Z_M^{O,ij}$,

$$Z_M^{\mathrm{PP}} = \prod_{O,ij} Z_M^{O,ij}. \tag{27}$$

Since in equation (20) the pre-processing prior $\pi_M^{\mathrm{PP},\tilde{O}}$ for the "true" observables is flat within $(\tilde{O}_{\min}, \tilde{O}_{\max})$, we can write explicitly

$$\pi_M^{\mathrm{PP},\tilde{O}} = \prod_{O,ij} \left[\prod_{k=1}^{N_M^{O,ij}} \left(\tilde{O}_{\max} - \tilde{O}_{\min} \right) \right]^{-1}, \tag{28}$$

we can substitute equations (26), (27) and (28) into equation (20), so that we finally get

$$Z_M \tilde{\propto} \prod_{O,ij} \left[\left(\prod_{k=1}^{N_M^{O,ij}} w_{ij,k}^{*\tilde{O}} \right)^{-1} (2\pi)^{N_M^{O,ij}} \sqrt{\det C_{O,ij}} \exp\left[-\frac{\chi^{2,\min}_{O,ij}}{2} \right] \right] \times K_M!\; Z_M^{\mathrm{NPS}}. \tag{29}$$

We write this equation in a more compact form

$$Z_M \tilde{\propto} W^{-1} (2\pi)^{N_M} \sqrt{\det C} \exp\left[-\frac{\chi^2_{\min}}{2} \right] \times K_M!\; Z_M^{\mathrm{NPS}} \tag{30}$$

by introducing the following abbreviations:

$$W^{-1} = \prod_{O,ij} \prod_{k=1}^{N_M^{O,ij}} \left(w_{ij,k}^{*\tilde{O}} \right)^{-1} \tag{31}$$

$$N_M = \sum_{O,ij} N_M^{O,ij} \tag{32}$$

$$\det C = \prod_{O,ij} \det C_{O,ij} \tag{33}$$

$$\chi^2_{\min} = \sum_{O,ij} \chi^{2,\min}_{O,ij}, \tag{34}$$

i.e. W^{-1} is the product of the inverse widths of all peaks in all fits (both FRET efficiency and FRET anisotropy), N_M is the total number of peaks, $\det C$ is the product of all determinants of covariance matrices, and $\chi^2_{\min}$ is the total of all minimized χ^2 values obtained.

5 Supplementary Tables

fluorophore i	x_i/Å	y_i/Å	z_i/Å	θ_i/rad	ϕ_i/rad	$r_{\infty,i}$
sat. 1	36.7	12.0	27.2	3.01	0.78	0.31
sat. 2	23.2	46.9	3.4	1.53	0.37	0.31
sat. 3	45.7	31.3	3.6	1.11	3.00	0.15
sat. 4	8.5	36.3	44.9	0.46	0.07	0.32
sat. 5	46.9	3.0	2.0	1.44	2.54	0.24
sat. 6	5.0	5.3	48.9	2.69	0.17	0.22
sat. 7	4.4	56.4	18.6	1.16	0.12	0.29
sat. 8*	49.9	-27.3	30.2	2.52	0.73	0.24
sat. 9*	44.2	61.6	-24.7	0.73	1.99	0.21
ant. 1	56.5	49.7	75.6	1.52	0.91	0.31
ant. 2	89.7	47.0	36.6	1.42	2.26	0.18
ant. 3	65.1	5.7	82.0	2.55	1.29	0.31
ant. 4	72.6	77.4	28.4	0.61	2.07	0.27
ant. 5	81.1	2.1	54.1	2.43	2.05	0.27
ant. 6	86.6	11.1	27.0	1.04	2.29	0.23
ant. 7	49.8	78.9	17.0	0.47	1.69	0.22

*) used in extended FRET network only

Suppl. Table 1: Fluorophore data. Exact fluorophore positions $\boldsymbol{x}_i = (x_i, y_i, z_i)$ and average TDM orientations $\boldsymbol{\Omega}_i = (-\cos\theta_i, \phi_i)$ together with residual fluorescence anisotropy $r_{\infty,i}$.

R_{ij}^{iso}/Å	ant. 1	ant. 2	ant. 3	ant. 4	ant. 5	ant. 6	ant. 7
sat. 1	56.7	62.3	57.7	56.7	56.3	55.2	58.6
sat. 2	57.4	64.3	62.7	57.8	60.0	56.6	64.4
sat. 3	64.1	64.8	58.2	59.6	61.9	64.2	64.8
sat. 4	63.5	64.6	58.2	63.0	63.9	63.4	59.4
sat. 5	60.8	56.0	63.4	60.9	55.2	57.1	56.0
sat. 6	60.6	61.2	64.1	57.1	64.1	64.8	61.3
sat. 7	64.0	57.2	64.3	55.4	63.3	57.8	63.2
sat. 8	-	-	60.5*	-	61.6*	8.3*	-
sat. 9	-	63.2*	-	64.8*	-	-	55.2*

*) used in extended FRET network only

Suppl. Table 2: Isotropic Förster distances in Å.

E_{ij}	ant. 1	ant. 2	ant. 3	ant. 4	ant. 5	ant. 6	ant. 7
sat. 1	0.5516	0.1688	0.4556	0.1684	0.3280	0.3773	0.3356
sat. 2	0.0501	0.3696	0.0866	0.4617	0.1382	0.1544	0.9560
sat. 3	0.2306	0.5455	0.0642	0.5336	0.3407	0.6438	0.7262
sat. 4	0.8171	0.1009	0.1183	0.1647	0.0603	0.1334	0.2951
sat. 5	0.0623	0.2111	0.0709	0.0502	0.2089	0.7019	0.0946
sat. 6	0.1512	0.0289	0.3600	0.0286	0.2969	0.1339	0.0758
sat. 7	0.1187	0.1656	0.1430	0.0727	0.1313	0.1295	0.6309
sat. 8	-	-	0.1871*	-	0.6646*	0.2966*	-
sat. 9	-	0.2714*	-	0.4459*	-	-	0.9042*

*) used in extended FRET network only

Suppl. Table 3: FRET efficiencies.

A_{ij}	ant. 1	ant. 2	ant. 3	ant. 4	ant. 5	ant. 6	ant. 7
sat. 1	−0.1516	−0.1113	0.2124	0.1290	0.1173	−0.0363	0.1478
sat. 2	0.1888	−0.0839	−0.1131	−0.1453	−0.1392	−0.1036	−0.1211
sat. 3	−0.0505	0.0470	−0.0450	0.0379	−0.1011	0.0940	−0.0219
sat. 4	−0.1036	−0.1137	0.0494	0.0298	0.1305	−0.1158	0.1153
sat. 5	−0.1338	0.1825	−0.1327	0.0174	−0.0429	0.1660	−0.0541
sat. 6	−0.0992	−0.0628	0.1538	0.1243	0.0086	0.0308	0.0952
sat. 7	0.0480	−0.0490	−0.1419	−0.1342	−0.0278	−0.1066	−0.0778
sat. 8	-	-	0.0483*	-	0.1651*	−0.1189*	-
sat. 9	-	0.0637*	-	0.2335*	-	-	0.1840*

*) used in extended FRET network only

Suppl. Table 4: FRET anisotropies.

calculation	$\ln Z$
FRET efficiency only	8.3 ± 0.1
FRET efficiency only, extended network	9.5 ± 0.3
FRET efficiency and anisotropy	133.5 ± 0.4
FRET efficiency and anisotropy, finite sat. prior	125.0 ± 0.3

Suppl. Table 5: Evidences of global NPS calculations. The natural logarithm of the evidence ($\ln Z$) and its error listed for the global NPS calculations.

References

[1] Dale, R.; Eisinger, J.; Blumberg, W. *Biophys. J.* **1979**, *26*, 161–193.

[2] Muschielok, A.; Andrecka, J.; Jawhari, A.; Brückner, F.; Cramer, P.; Michaelis, J. *Nat. Methods* **2008**, *5*, 965–971.

[3] Crowther, R. A.; Henderson, R.; Smith, J. M. *J. Struct. Biol.* **1996**, *116*, 9–16.

[4] UCSF Chimera Software package. http://www.cgl.ucsf.edu/chimera, 2010; `http://www.cgl.ucsf.edu/chimera`.

[5] Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. *J. Comput. Chem.* **2004**, *25*, 1605–1612.

[6] Strehl, A.; Ghosh, J. *Journal of Machine Learning Research* **2002**, *3*, 583–617.

[7] Kullback, S.; Leibler, R. A. *The Annals of Mathematical Statistics* **1951**, *22*, 79–86.

[8] Sivia, D. *Data Analysis, 2nd edition*; Oxford University Press, USA, 2006.

[9] Sivia, D. S.; Carlile, C. J. *J. Chem. Phys.* **1992**, *96*, 170–178.

[10] Jeffreys, H. *Theory of probability*; Clarendon Press, Oxford, 1939.