THE JOURNAL OF PHYSICAL CHEMISTRY B

ARTICLE

pubs.acs.org/JPCB

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

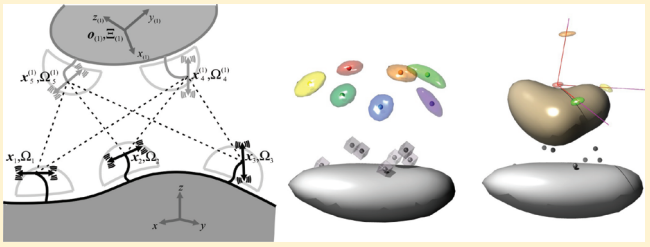
Adam Muschielok[†] and Jens Michaelis[*,†,‡]

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

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

Supporting Information

ABSTRACT: Single-molecule fluorescence resonance energy transfer (sm-FRET) has been recently applied to distance and position estimation in macromolecular complexes. Here, we generalize the previously published Nano-Positioning System (NPS), a probabilistic method to analyze data obtained in such experiments, which accounts for effects of restricted rotational freedom of fluorescent dyes, as well as for limited knowledge of the exact dye positions due to attachment via flexible linkers. In particular we show that global data analysis of complete FRET networks is beneficial and that the measurement of FRET anisotropies in addition to FRET efficiencies can be used to determine accurately both position and orientation of the dyes. This measurement scheme improves localization accuracy substantially, and we can show that the improvement is a consequence of the more precise information about the transition dipole moment orientation of the dyes obtained by FRET anisotropy measurements. We discuss also rigid body docking of different macromolecules by means of NPS, which can be used to study the structure of macromolecular complexes. Finally, we combine our approach with common FRET analysis methods to determine the number of states of a macromolecule.



INTRODUCTION

Long after its full quantum mechanical description by Förster[1] and some years after the first experiments on the single-molecule level,[2] fluorescence resonance energy transfer (FRET), often referred to as a "spectroscopic ruler",[3] has meanwhile evolved into a versatile tool in single-molecule biophysics.[4−7] FRET efficiency is often interpreted in terms of distances or distance changes,[3,8−13] and was used to determine the position of yet unlocalized residues in biological macromolecules[14−18] and to perform distance-restrained docking between several known structures constituting a macromolecular complex.[19−22] The common theme behind these experiments is that single-molecule FRET measurements can yield valuable information if standard approaches like X-ray crystallography or nuclear magnetic resonance spectroscopy fail. For example, we have recently localized yet unobserved parts of nucleic acids in the eukaryotic RNA polymerase II elongation complex.[16,23,24] To analyze the data of such FRET localization experiments, we have developed the Nano-Positioning System (NPS).[23] NPS relies on a partly known structure of a macromolecular complex, which serves as a platform onto which fluorescent dyes, referred to as satellites, are attached via flexible linkers. By analyzing sequential FRET efficiency measurements between several satellites and the so-called antenna fluorophore bound to a structurally unresolved region of the macromolecule, it is possible to calculate the antenna position and hence localize this region. In principle, the measurement process could also be accelerated by using one antenna and several switchable satellites in the same experiment.[25]

There are certain uncertainties that limit the resolution of FRET-based distance measurements and localization in general and of NPS in particular. These include the FRET efficiency measurement error and the limited knowledge of the exact dye positions caused by attachment to the macromolecule via flexible linkers.[17,21,26] Another important source of localization uncertainty originates from the dependence of FRET efficiency on the average transition dipole moment (TDM) orientations of the dyes, which occurs when the fluorophores are constrained in their orientation fluctuations. Indications for such uncertainties come from measurements of nonzero fluorescence anisotropies, which has been already discussed for pairs of fluorophores.[27−30] NPS faces all uncertainties above by applying Bayesian parameter estimation,[31,32] a general and powerful probabilistic data analysis technique.

Up to now, NPS was able to determine only the position of a single antenna based on FRET efficiency measurements to several satellites.[23] In the following, we will extend NPS to simultaneously calculate not only the position but also the orientation of all fluorophores in FRET networks containing an arbitrary number of antennas and satellites. Moreover, this new global analysis scheme benefits from the additional measurement of FRET anisotropy, i.e., the fluorescence anisotropy of the acceptor after a polarized excitation of the donor. This quantity

Received: June 27, 2011
Revised: August 30, 2011
Published: September 02, 2011

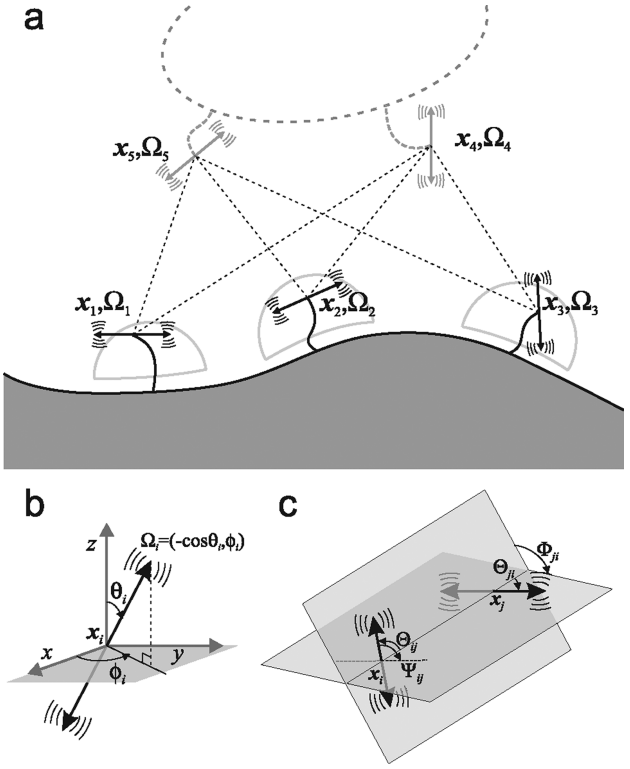


Figure 1. FRET network model. (a) Fluorophores (shown as double arrows) are attached via flexible linkers either to parts of the macromolecule with known structure (gray area) or to the yet unlocalized parts with unknown structure (gray dashed oval, top). Each dye is characterized by its spatial position $\boldsymbol{x}_i$ and the average orientation of the TDM $\boldsymbol{\Omega}_i$. For fluorophores attached to the known parts of the macromolecule (satellites) the accessible volume is indicated by a light gray contour around each satellite. The accessible volume for fluorophores attached to the unknown parts of the macromolecule (antennas) is the region outside the known structure. Measurements of FRET efficiency and/or FRET anisotropy are indicated by dashed lines between the fluorophores. (b) The fluorophore i is described by its transition dipole moment (TDM) at position $\boldsymbol{x}_i$ that fluctuates (indicated by motion lines) around its average orientation, shown as double arrow and described by the polar and azimuth angle θ_i and ϕ_i. The parametrization $\boldsymbol{\Omega}_i = (-\cos\theta_i\phi_i)$ is used, since a flat prior in $\boldsymbol{\Omega}_i$ is invariant under rotations of the coordinate system and does not favor any particular orientation. (c) A FRET pair consisting of two fluorophores i and j at positions $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$, respectively. The angles Θ_{ij} and Θ_{ji} between each average TDM orientation $\boldsymbol{\Omega}_{i/j}$ and the vector interconnecting the fluorophores, $(\boldsymbol{x}_j - \boldsymbol{x}_i)$, as well as the angle Φ_{ij} between the planes containing $\boldsymbol{\Omega}_{i/j}$ and $\boldsymbol{x}_j - \boldsymbol{x}_i$ are needed in the calculation of the dynamically averaged orientation factor $\langle\kappa_{ij}^2\rangle$. The FRET anisotropy depends directly on the angle Ψ_{ij} between the average TDM orientations (the dashed line has the same direction as the average TDM orientation of the dye j).

carries important information about the relative orientations of the average TDMs of the dyes.[27,30,33] Thereafter we demonstrate that NPS can be used to dock several macromolecules of known structure constituting a structurally unknown macromolecular complex. In this way, we can quantitatively determine the position and orientation of the macromolecules within a complex, which is an important question in structural biology. We finally show how NPS can be combined with common FRET analysis methods[9,13,14,16,34−39] to infer the number of conformational states of a macromolecule.

NPS THEORY

Here we first review the FRET theory applied in the NPS and derive thereafter 3 variants of NPS: the position−orientation NPS, the position−Förster distance NPS, and the NPS docking analysis scheme. All NPS variants are suited to study a network of fluorophores that are attached to a macromolecule (or a complex of several macromolecules) and act as donor/acceptor pairs in separate FRET measurements. Depending on whether the position of a dye is known or is to be determined it will be called satellite or antenna.

Physical Model of a FRET Network. We study a FRET network consisting of N dye molecules and several separately conducted FRET measurements between pairs of these dyes (Figure 1a). To describe the FRET measurements we use the theory derived by Dale et al.[28] for a single FRET pair, which will be described shortly.

In our model each fluorophore i (i = 1, ..., N) is located at a static position $\boldsymbol{x}_i$. It possesses a transition dipole moment (TDM), which performs rapid, constrained orientation fluctuations that are much faster than the time-scale of the fluorescence lifetime. The density of visited TDM orientations is assumed to be axially symmetric, and the average TDM orientation, i.e. the symmetry axis of the distribution, will be denoted by $\boldsymbol{\Omega}_i$ (Figure 1b). Further, we assume that there is no additional angular reorientation on a slower time-scale and that the orientation distributions of absorption and emission TDMs are identical even though there might be an angle between these TDMs resulting in a fundamental fluorescence anisotropy $r_{i,0} := \lim_{t\to 0} r_i(t)$ smaller than 2/5, where $r_i(t)$ is the time-resolved fluorescence anisotropy decay. $r_i(t)$ can be obtained from polarization-resolved measurements of fluorescence after a pulsed and linearly polarized excitation of the dye i. (See ref 33 and Chapter 11 in ref 40.) Then, the amount of orientation fluctuations is characterized by the dynamically averaged axial depolarization $\langle d_i^x\rangle$ determined from the residual fluorescence anisotropy $r_{i,\infty} := \lim_{t\to\infty} r_i(t)$ of an ensemble of fixed but randomly oriented macromolecules (Chapter 11 in ref 40) by solving

$$\langle d_i^x\rangle^2 = r_{i,\infty}/r_{i,0} \tag{1}$$

For a FRET pair consisting of fluorophores i and j (Figure 1c), Dale et al.[28] derived the dynamically averaged orientation factor $\langle\kappa_{ij}^2\rangle$

$$\begin{aligned}\langle\kappa_{ij}^2\rangle = {} & (\cos\Psi_{ij} - 3\cos\Theta_{ij}\cos\Theta_{ji})^2\langle d_i^x\rangle\langle d_j^x\rangle \\ & + (1/3 + \cos^2\Theta_{ji}\langle d_j^x\rangle)(1 - \langle d_i^x\rangle) \\ & + (1/3 + \cos^2\Theta_{ij}\langle d_i^x\rangle)(1 - \langle d_j^x\rangle)\end{aligned} \tag{2}$$

where Θ_{ij}, Θ_{ji}, and Ψ_{ij} are angles that describe the relative orientation of the fluorophores and can be calculated from $\boldsymbol{x}_i$, $\boldsymbol{x}_j$, $\boldsymbol{\Omega}_i$, and $\boldsymbol{\Omega}_j$. From $\langle\kappa_{ij}^2\rangle$ the expected Förster distance R_{ij} and the expected FRET efficiency $\mathcal{E}_{ij}$ can be calculated to be

$$R_{ij}^6 = R_{ij}^{\mathrm{iso}\,6}\frac{\langle\kappa_{ij}^2\rangle}{2/3} \tag{3}$$

$$\mathcal{E}_{ij} = (1 + |\boldsymbol{x}_i - \boldsymbol{x}_j|^6/R_{ij}^6)^{-1} \tag{4}$$

where R_{ij}^{iso} is the isotropic Förster distance for the FRET pair ij that defines the length scale of energy transfer in the isotropic dynamic averaging case, i.e. for $\langle\kappa_{ij}^2\rangle$ = 2/3. Experimentally, FRET efficiency is usually calculated from donor fluorescence

lifetime measurements or from the fluorescence signals of donor and acceptor (Chapter 13 in ref 40). The expected FRET anisotropy, $\mathcal{A}_{ij}$, which is defined as the fluorescence anisotropy of the acceptor after polarized excitation of the donor, depends on the angle Ψ_{ij} between $\mathbf{\Omega}_i$ and $\mathbf{\Omega}_j$ and determines the expected FRET anisotropy,[28]

$$\mathcal{A}_{ij} = \frac{2}{5}\langle d_i^x\rangle\langle d_j^x\rangle\left(\frac{3}{2}\cos^2\Psi_{ij} - \frac{1}{2}\right) \tag{5}$$

In total, our FRET network model is characterized by $5N$ parameters: the positions and average TDM orientations of each dye. In the following, we show how the measured data, which is FRET efficiency and, if possible, FRET anisotropy, is used in the analysis.

Position−Orientation NPS. We apply Bayesian parameter estimation to infer fluorophore positions and orientations from FRET efficiencies and optionally also from FRET anisotropies measured in a FRET network consisting of an arbitrary number of satellites and antennas. Let in the following $\{\boldsymbol{x}_i,\mathbf{\Omega}_i\} = (\boldsymbol{x}_1,\mathbf{\Omega}_1, ..., \boldsymbol{x}_N, \mathbf{\Omega}_N)$ be the $5N$-dimensional vector of fluorophore positions and orientations that describes the FRET network, and let $\{E_{ij}\}$ and $\{A_{ij}\}$ be all FRET efficiencies and FRET anisotropies measured between fluorophore pairs ij.

Posterior. The posterior probability density function (PDF) that represents the inferred information can be calculated with Bayes' theorem

$$p(\{\boldsymbol{x}_i,\mathbf{\Omega}_i\}|\{E_{ij}\},\{A_{ij}\},I) = p(\{\boldsymbol{x}_i,\mathbf{\Omega}_i\}|I)p(\{E_{ij}\},\{A_{ij}\}|\{\boldsymbol{x}_i,\mathbf{\Omega}_i\},I)/Z \tag{6}$$

Here, $p(.|.)$ denotes conditional PDFs, in particular, $p(\{\mathbf{x}_i,\mathbf{\Omega}_i\}|\{E_{ij}\},\{A_{ij}\},I)$ is the posterior PDF that represents the inferred information and is hence the result of the analysis.

The prior PDF $p(\{\boldsymbol{x}_i,\mathbf{\Omega}_i\}|I)$ is the state of knowledge before the experiment is analyzed, while the likelihood $p(\{E_{ij}\},\{A_{ij}\}|\{\boldsymbol{x}_i,\mathbf{\Omega}_i\},I)$ describes the expected data given particular dye positions and average TDM orientations. Z denotes the evidence that normalizes the product of prior and likelihood and which can be calculated as follows:

$$Z = \int \mathrm{d}\{\boldsymbol{x}_i,\mathbf{\Omega}_i\}\; p(\{\boldsymbol{x}_i,\mathbf{\Omega}_i\}|I)\;\; p(\{E_{ij}\},\{A_{ij}\}|\{\boldsymbol{x}_i,\mathbf{\Omega}_i\},I) \tag{7}$$

All PDFs are conditional on the background information I, which consists here of Förster theory, the assumptions that enter the model, and other information on the macromolecule and fluorophores. By integrating over all model parameters but $\boldsymbol{x}_i$ we compute the marginal position posterior PDF of the dye i

$$p(\boldsymbol{x}_i|\{E_{ij}\},\{A_{ij}\},I) = \int \mathrm{d}\{\boldsymbol{x}_{j\neq i},\mathbf{\Omega}_j\}\; p(\{\boldsymbol{x}_j,\mathbf{\Omega}_j\}|\{E_{ij}\},\{A_{ij}\},I) \tag{8}$$

which represents the available information on the position of the dye i. Analogously, $p(\mathbf{\Omega}_i|\{E_{ij}\},\{A_{ij}\},I)$, the marginal posterior PDF of the average TDM orientation $\mathbf{\Omega}_i$ can be calculated by integrating over all model parameters but $\mathbf{\Omega}_i$.

Likelihood. Since all FRET efficiencies and FRET anisotropies are measured independently, the likelihood consists of the product of contributions of each FRET pair.

$$p(\{E_{ij}\},\{A_{ij}\}|\{\boldsymbol{x}_i,\mathbf{\Omega}_i\},I) = \prod_{ij\,\in\,M} L_{ij}(\boldsymbol{x}_i,\mathbf{\Omega}_i,\boldsymbol{x}_j,\mathbf{\Omega}_j) \tag{9}$$

L_{ij} denotes the contribution of FRET pair ij as a function of the positions and orientations of the respective fluorophores, and M is the set of measured FRET pairs. Depending on whether FRET efficiency, FRET anisotropy or both were measured for a FRET pair, we assign L_{ij} as a uni- or bivariate Gaussian

$$L_{ij}(\boldsymbol{x}_i,\mathbf{\Omega}_i,\boldsymbol{x}_j,\mathbf{\Omega}_j) = \begin{cases} \mathcal{N}_{\Delta E_{ij}}(E_{ij}-\mathcal{E}_{ij}) & E_{ij}\text{ measured} \\ \mathcal{N}_{\Delta A_{ij}}(A_{ij}-\mathcal{A}_{ij}) & A_{ij}\text{ measured} \\ \mathcal{N}_{\Delta E_{ij}}(E_{ij}-\mathcal{E}_{ij})\mathcal{N}_{\Delta A_{ij}}(A_{ij}-\mathcal{A}_{ij}) & E_{ij}\text{ and }A_{ij}\text{ measured} \end{cases} \tag{10}$$

Here, $\mathcal{N}_\sigma(\cdot)$ is the normal distribution with standard deviation σ. $\mathcal{E}_{ij}$ and $\mathcal{A}_{ij}$ are the expected FRET efficiency and FRET anisotropy defined in eq 4 and 5, respectively, and ΔE_{ij} and ΔA_{ij} denote the standard measurement errors of FRET efficiency and FRET anisotropy, respectively.

For clarity in the main manuscript we have assumed in eq 9 that $\mathcal{E}_{ij}$ and $\mathcal{A}_{ij}$ have unique values, which is true if for all dyes $r_{i/j,\infty}/r_{i/j,0} > 1/4$ and hence all axial depolarizations are positive (eq 1). A complete treatment of the possible ambiguities caused by low residual fluorescence anisotropies is given in the Supporting Information (Suppl. theory 1.1).

Prior. The prior PDF describes the possible fluorophore positions and average TDM orientations in the absence of experimental data. In this way, the influence of the shape of the macromolecule and fluorophores (and if the attachment position is known, also the linker length; Figure 1a) can be encoded. We assume that the fluorophores cannot interact with each other in a way that would change their positions and/or average TDM orientations, hence the prior factorizes

$$p(\{\boldsymbol{x}_i,\mathbf{\Omega}_i\}|I) = \prod_i p(\boldsymbol{x}_i,\mathbf{\Omega}_i|I) \tag{11}$$

Usually, one would assess the accessible fluorophore positions and average TDM orientations from simulations wherever possible to reduce uncertainties in the localization.[14] As a simulation is not always possible, either because the local structure of the fluorophore attachment site is unknown or if the macromolecule is simply too large to simulate accurately, we use a more primitive approach here and neglect correlations of fluorophore position and average TDM orientation, which allows us to factorize the position and orientation prior

$$p(\boldsymbol{x}_i,\mathbf{\Omega}_i|I) = p(\boldsymbol{x}_i|I)p(\mathbf{\Omega}_i|I)\text{ for all dyes }i \tag{12}$$

Furthermore, without doing simulations, there is no reason to favor any specific orientation, hence we assign a flat (i.e., constant) prior in $\mathbf{\Omega}_i$ (Supporting Information, Suppl. methods 1.2), which is parametrized by $\mathbf{\Omega}_i = (-\cos\theta_i, \phi_i)$, θ_i and ϕ_i being the polar and azimuth angle of the average TDM orientation (Figure 1b). Moreover, we also assign a flat position prior that does not favor any particular position within the volume V_i accessible to each dye

$$p(\boldsymbol{x}_i|I) = \begin{cases} \left(\int_{\boldsymbol{x}_i\,\in\,V_i}\mathrm{d}\boldsymbol{x}_i\right)^{-1} & \text{for } \boldsymbol{x}_i \in V_i \\ 0 & \text{otherwise} \end{cases} \tag{13}$$

Position−Förster Distance NPS. We have previously published a simple NPS analysis scheme,[23] which we now denote as

position−Förster distance NPS. This NPS variant can be applied if a FRET network consists of several satellites and only one single antenna and if there is only FRET efficiency data available.

Position−Förster distance NPS is a special case of the position−orientation NPS, and allows to calculate only the marginal posterior of the antenna position, however, at a low computational cost. In short, while the antenna position $\boldsymbol{x}_a$, the satellite positions $\{\boldsymbol{x}_i\}$ and the respective dye position priors are the same as in the position−orientation NPS, now the Förster distances $\{R_{ai}\}$ take the role of the average TDM orientations as model parameters (Supporting Information, Suppl. theory 1.3). Though the Förster distance priors are computed based on the unknown average TDM orientations, the position−Förster distance NPS conceptually differs from the position−orientation NPS as we additionally assume that the Förster distances R_{ai} are mutually independent. In the Supporting Information, Suppl. discussion 4.1, we demonstrate that this approximation is usually applicable as the introduced error is negligible.

NPS Docking. Docking of several macromolecules of known structure can yield important structural information about the complex they form. Docking has been applied previously to FRET efficiency based constraints assuming an isotropic orientation factor $\kappa^2 = 2/3$.[19−22] In the following, we discuss NPS docking, which is capable of treating orientation effects and enables us to analyze both FRET efficiency and/or FRET anisotropy based on the FRET model previously introduced in this manuscript. Mathematically, NPS docking extends the position−orientation NPS framework by describing also the positions and orientations of structurally known constituents of a macromolecular complex, which are treated as rigid bodies.

Model and Likelihood. In order to dock M rigid constituents of the macromolecular complex by means of FRET we choose one of the parts of the complex as the "laboratory" reference frame and attach to every additional part $k = 1, ..., M$ a reference frame with the origin $\boldsymbol{o}_{(k)}$ and the orientation $\boldsymbol{\Xi}_{(k)} = (\rho_{(k)}, -\cos\theta_{(k)}, \phi_{(k)})$ (Figure 2a). $\boldsymbol{\Xi}_{(k)}$ is characterized by three subsequent rotations by the angles $\rho_{(k)}$, $\theta_{(k)}$ and $\phi_{(k)}$ (Figure 2b). A fluorophore i linked to the part k is now described by its position $\boldsymbol{x}_i^{(k)}$ and average transition dipole orientation $\boldsymbol{\Omega}_i^{(k)}$ relative to the reference frame k. Expected FRET efficiency and FRET anisotropy values are then computed after coordinate transformation of $\boldsymbol{x}_i^{(k)}$ and $\boldsymbol{\Omega}_i^{(k)}$ into the "laboratory" reference frame like in the position−orientation NPS. In this way, the model is then parametrized by the positions and orientations of all fluorophores and reference frames.

In the end the information of all fluorophores linked to the reference frame k is used to position and orient this reference frame with respect to the laboratory coordinate system.

Prior. We neglect possible interactions of fluorophores with other macromolecule parts than the one used for their attachment, and assign the fluorophore position and orientation prior exactly as described above. The prior PDF $p(\{\boldsymbol{o}_{(k)},\boldsymbol{\Xi}_{(k)}\}|I)$ of the additional model parameters will be correlated in general since the possible orientations of the macromolecules will depend on their positions. The parameter space of accessible positions and orientations can be determined by simulations; however, we neglect the correlations between $\boldsymbol{o}_{(k)}$ and $\boldsymbol{\Xi}_{(k)}$ here for simplicity

$$p(\{\boldsymbol{o}_{(k)},\boldsymbol{\Xi}_{(k)}\}|I) = \prod_k p(\boldsymbol{o}_{(k)}|I)p(\boldsymbol{\Xi}_{(k)}|I) \quad (14)$$

and set $p(\boldsymbol{o}_{(k)}|I)$ constant within an accessible volume and zero outside, in analogy to the fluorophore position priors, assuming

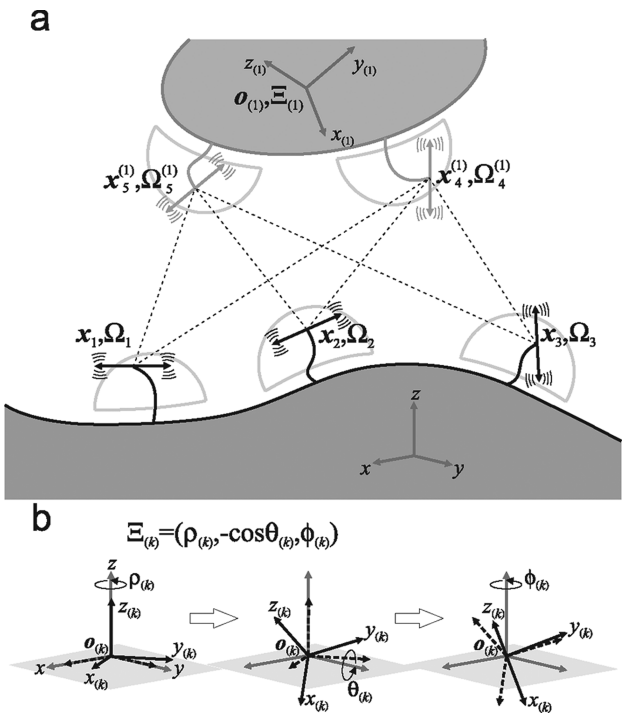


Figure 2. NPS docking. (a) The NPS can be used for docking different parts of a macromolecule. Here, the structure of the unlocalized part of the macromolecule is known (gray oval, top), and the volumes accessible to the antennas (gray contours around the antennas) are known relative to this structure. A reference frame $(x_{(k)},y_{(k)},z_{(k)})$ is attached to the unlocalized part k. In this way, the position and orientation of the unlocalized part is described by the origin $\boldsymbol{o}_{(k)}$ and orientation $\boldsymbol{\Xi}_{(k)}$ of the reference frame relative to the "laboratory" reference frame (x,y,z). (b) The frame of reference k used for docking is shown as axes $x_{(k)}$, $y_{(k)}$, and $z_{(k)}$. It is parametrized by the position of its origin, $\boldsymbol{o}_{(k)}$, and its orientation defined by the angles $\rho_{(k)}$, $\theta_{(k)}$ and $\phi_{(k)}$ of three consecutive rotations around the axes z, y, and z, respectively. Both, position and orientation, are given relative to the laboratory reference frame (x,y,z).

thereby that all accessible combinations of positions and orientations are equally likely. The prior for the reference frame orientation $\boldsymbol{\Xi}_{(k)}$ is flat and does not favor any particular orientation (Supporting Information, Suppl. methods 1.2). The complete prior, $p(\{\boldsymbol{x}_i^{(k)},\Omega_i^{(k)}\},\{\boldsymbol{o}_{(k)},\boldsymbol{\Xi}_{(k)}\}|I)$, is given by the product of eqs 14 and 11.

Posterior. The posterior PDF is computed as in the position−orientation NPS. In particular, the marginal posterior of any fluorophore position or average TDM orientation is calculated by integrating over all remaining parameters, which involve here also the reference frame positions and orientations. We are also able to compute 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 (Supporting Information, Suppl. methods, 2.1). By displaying this PDF, the position estimate of any atom in the docked structure can be visualized directly, of course under the assumption that this structure behaves as a rigid body.

METHODS

Simulated Data. For the simulation of data used in our calculations we considered typical conditions like the number and overall spacing of the dyes and a distribution of low,

intermediate and high FRET efficiencies observed in previous experiments.[23,24] In an actual experiment intermediate FRET efficiencies are to be preferred because in this range FRET is most sensitive to changes of distance and average TDM orientation of the dyes, although it is also possible to analyze low and high FRET efficiencies, which lead to less precise distance and position estimates. The analysis of very high FRET efficiencies should be avoided in practice since the dyes can be too close for the point-like dipole approximation of Förster theory[1] to hold. Even if Förster theory holds, the donor lifetime could become comparable to its rotation correlation time, so that the Dale–Eisinger–Blumberg theory[27,28] breaks down. However, since we did not do actual experiments it was not necessary to account for these complications here.

In our simulation, we chose seven satellite and seven antenna positions (Supporting Information, Suppl. Table 1) from a uniform distribution in the positive octant of the laboratory reference frame under the conditions that the distance to the origin is between 40 and 60 Å (satellites) and 90 and 110 Å (antennas). Furthermore, we restricted the minimum distance between fluorophores to 25 Å for two reasons: First, we were able to test several different antenna positions without changing the network, and second, one should place the satellites in several different positions to obtain independent information (i.e., FRET efficiency data) that is used to localize the antennas. Nevertheless, in general, the distances between two antennas or two satellites do not need to be limited as the dye molecules are usually not used simultaneously in the FRET measurement.

In order to form an extended FRET network two additional satellites were placed manually in positions that promised to reduce localization uncertainty. This is a typical procedure in an actual experiment to increase localization accuracy. In brief, we used our previous model, in which a FRET efficiency measurement between a satellite and an antenna fluorophore imposes a fuzzy distance restraint between the dyes[23] and looked for positions from which a FRET efficiency measurement could exclude many of the possible antenna positions estimated by analyzing the original FRET network.

Average fluorophore transition dipole orientations were chosen from an isotropic distribution without any constraint. Fluorescence anisotropies were randomly chosen from a uniform distribution between 0.15 and 0.32 (Supporting Information, Suppl. Table 1) motivated by values determined in previous experiments.[16,23,24] Similarly, isotropic Förster distances were chosen from a uniform distribution between 55 and 65 Å (Supporting Information, Suppl. Table 2), in order to mimic the spread of isotropic Förster distances in our previous observations.[16,23,24]

Using these values, the expected FRET efficiencies and FRET anisotropies were calculated (Supporting Information, Suppl. Tables 3,4) based on the model described above. The expected data was used directly without adding any noise that would otherwise simulate experimental measurement errors. In that way it was possible to calculate the posterior maximum directly and observe whether it was found by the analysis algorithm. For all FRET efficiency and FRET anisotropy data an uncertainty of $\Delta E_{ij} = 0.02$ and $\Delta A_{ij} = 0.01$ was assigned, respectively. The chosen ΔE_{ij} are typical values observed in our measurements,[23] while ΔA_{ij} was chosen to be approximately 2% of the possible range of FRET anisotropy values.

Accessible Volumes. In order to demonstrate the effect of satellite position uncertainty, we used two kinds of accessible volumes, either infinitely small volumes that result in position priors described by three-dimensional Dirac δ point measures, or cubes with 8 Å side length closely resembling the size of satellite position priors in previous experiments.[23,24] Both sorts of accessible volumes were centered at the respective simulated satellite positions. We neglected possible steric clashes of the antennas with the macromolecule and used cubes with 230 Å side length centered at the point $(x,y,z) = (85,85,85)$ Å as volumes accessible to each antenna.

Position–Orientation NPS and Position–Förster Distance NPS. The same simulated data set was analyzed in several ways, referred to as scenarios in the following. In the first scenario, the approximate analysis, the FRET network of 7 antennas and 7 satellites (7A/7S) was decomposed into subnetworks that consisted of FRET efficiency measurements from all satellites to a single antenna i and analyzed with the previously published position–Förster distance NPS software.[23] In order to mimic infinitely small accessible volumes of satellites we used isotropic normal distributions with a negligible standard deviation of 0.01 Å that were centered at the actual satellite positions.

All remaining scenarios were analyzed with the position-orientation NPS. In particular, in the separate analysis of the 7A/7S FRET network we studied the same subnetworks as in the approximate analysis in order to compare the fully numerical and the approximate analytical approaches. All investigated FRET networks consisted of FRET efficiency data, or alternatively, combined FRET efficiency and anisotropy data. The accessible volumes of the satellites were infinitely small in all scenarios except the last one, in which finite-sized accessible volumes were used (see preceding section "Accessible Volumes").

Posterior samples and the evidence Z were calculated with a custom implementation of the nested sampling algorithm[41] based on Markov chain Monte Carlo (MCMC) in C and MATLAB (The MathWorks). For all nested sampling calculations, we used 1×10^3 or 2×10^3 objects except for the global NPS analysis of the network with 7 antennas and 7 satellites, where 10×10^3 objects had to be used because the algorithm did not converge to the global posterior maximum with less objects. The posterior samples that resulted from nested sampling were reweighted (Chapter 9 in ref 32) to have a relative frequency proportional to the posterior PDF. Approximate marginal antenna position posterior PDFs were computed by binning the reweighted samples into a 3-dimensional histogram with bin sizes that varied from 0.5 to 8 Å dependent on the number and spatial spread of samples. The marginal dye position posteriors were used to compute credible volumes, the Bayesian counterpart of confidence regions (Supporting Information, Suppl. methods 2.2 and Suppl. Figure 1). Both marginal posteriors and credible volumes were displayed with Chimera[42,43] (Supporting Information, Suppl. methods 2.3). Marginal average TDM orientation posteriors and respective credible volumes were computed and displayed in a similar way with Matlab (The MathWorks).

Quantification of Antenna Position Estimates. Average antenna positions $\langle \boldsymbol{x}_i \rangle = \langle (x_i, y_i, z_i) \rangle$ were calculated from the approximate marginal posteriors. The localization uncertainty was computed by diagonalizing the covariance matrix $C_i = \langle (\boldsymbol{x}_i - \langle \boldsymbol{x}_i \rangle)(\boldsymbol{x}_i - \langle \boldsymbol{x}_i \rangle)^T \rangle$ as described earlier[23] resulting in three standard deviations $\sigma_1^{\boldsymbol{x}_i}$, $\sigma_2^{\boldsymbol{x}_i}$, and $\sigma_3^{\boldsymbol{x}_i}$ in the principal directions. In order to compare the different analysis schemes we calculated the average position uncertainty as follows: first, the geometric mean $\overline{\sigma^{\boldsymbol{x}_i}} = (\sigma_1^{\boldsymbol{x}_i} \sigma_2^{\boldsymbol{x}_i} \sigma_3^{\boldsymbol{x}_i})^{1/3} = (\det C_i)^{1/6}$ was computed for each antenna. The $\overline{\sigma^{\boldsymbol{x}_i}}$ obtained for each scenario were then averaged over all antennas of the FRET network to calculate the mean position

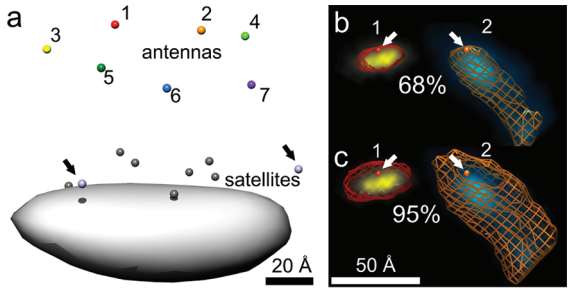


Figure 3. Actual fluorophore positions and position estimates. (a) Actual positions of the fluorophores used to calculate simulated data are shown as spheres relative to a hypothetical macromolecule (gray surface). The antennas, numbered 1−7, are shown in color, the satellites in gray. The additional satellites 8 and 9 used to extend the measurement network are highlighted by arrows. (b and c) The actual positions of antennas 1 and 2 are shown as red and orange spheres (see arrows), respectively. The corresponding marginal position densities of a separate analysis with the position−orientation NPS are shown in yellow (antenna 1) and blue (antenna 2). Credible volumes are shown as red (antenna 1) and orange (antenna 2) meshes for $P = 68\%$ (b) and $P = 95\%$ (c). To show whether the true positions lie within the credible volumes, only the part of the densities lying behind a plane containing both fluorophores is shown. In the case of antenna 1, both, the relatively small 68% and 95% credible volumes contain the true fluorophore position, whereas for antenna 2 the true position is outside of the 68% credible volume but inside the 95% volume.

uncertainty $\langle\overline{\sigma^{\mathbf{x}}}\rangle$. The mean average TDM orientation and the corresponding mean average TDM orientation uncertainty $\langle\sigma^{\overline{\mathbf{\Omega}}}\rangle$ was calculated similarly (Supporting Information, Suppl. methods 2.4).

NPS Docking. To demonstrate docking of two macromolecules, we used the same data as in the previous scenarios and analyzed it with the position−orientation NPS software, which is also capable to dock an arbitrary number of macromolecules. To this end, the satellites were located in the laboratory reference frame, while the positions and orientations of the antennas were now described relative to a mobile docking reference frame k = dock. The volume accessible to the docked reference frame origin was set to a cube of 300 Å centered at the position (x,y,z) = (100,100,100) Å. We analyzed measurements of (i) FRET efficiency in combination with infinitely small volumes accessible to satellites and antennas, (ii) the effect of additional FRET anisotropy measurements, and (iii) the effect of finite-sized position priors. For nested sampling 2×10^3 objects were used throughout.

In order to display position and orientation of the docked reference frame we computed $p(\boldsymbol{q}|\{E_{ij}\},I)$ and $p(\boldsymbol{q}|\{E_{ij}\},\{A_{ij}\},I)$ for $\boldsymbol{q}^{(\mathrm{dock})}$ = (0,0,0), (40,0,0), (0,40,0), and (0,0,40) Å. The resulting densities were treated in the same way like marginal dye position posteriors, i.e. they were displayed as 95% credible volumes, and the position uncertainty $\overline{\sigma^{\boldsymbol{q}}}$ of the point $\boldsymbol{q}$ was computed as $\overline{\sigma^{\boldsymbol{q}}} = \det(\langle(\boldsymbol{q} - \langle\boldsymbol{q}\rangle)(\boldsymbol{q} - \langle\boldsymbol{q}\rangle)^{\mathrm{T}}\rangle)^{1/6}$.

RESULTS AND DISCUSSION

Comparison of NPS Analysis Schemes. In order to compare different schemes for the analysis of FRET localization experiments, we simulated FRET data (Supporting Information, Suppl. Tables 2−4) based on a random FRET network (Figure 3a, Methods) consisting of 7 antennas and 7 satellites (7A/7S network). We analyzed the data with each of the NPS variants position−Förster distance NPS and position−orientation NPS. For both variants we calculated the marginalized antenna position posterior PDFs and displayed them as credible volumes showing directly the uncertainty that can be attributed to the inferred positions (see Methods). Before proceeding to the actual results we want to make the reader more familiar with the notion of credible volumes. In Figure 3b and c we plotted two exemplary marginal antenna position posterior PDFs and their credible volumes containing 68% and 95% probability. The surfaces of the 68% and 95% credible volumes correspond to 3-dimensional error bars sized 1 and 2 standard deviations, respectively.

We structured our comparison of NPS analysis schemes in several scenarios: In the first scenario, referred to as approximate analysis, we used the position−Förster distance NPS to analyze

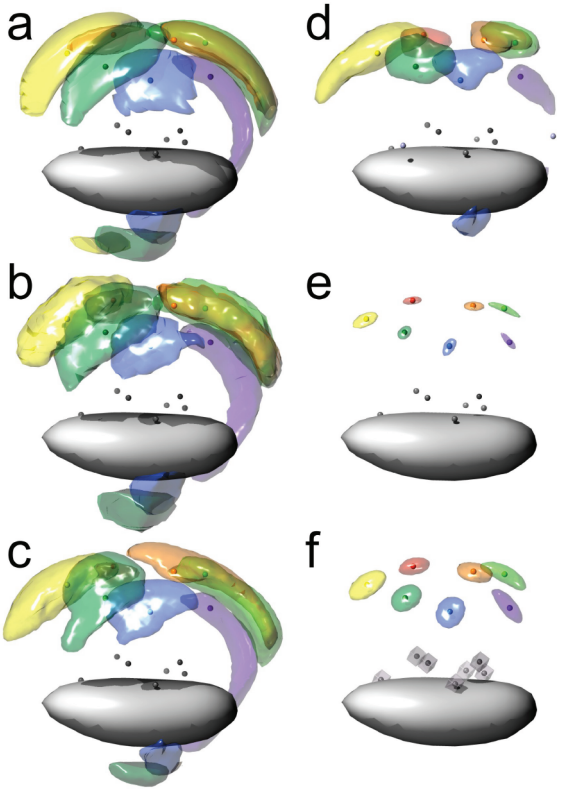


Figure 4. Estimates of antenna positions. 95% credible volumes of all antenna position estimates are shown as colored semitransparent surfaces, and the exact positions of all fluorophores are shown as spheres (color code see Figure 3) in the context of a hypothetical macromolecule (gray surface). Only small differences are visible between approximate (a) and separate analysis (b), supporting the approximations made to simplify the analysis in the position−Förster distance NPS and the correctness of the numerical approach. Globally analyzed data leads to an increase in localization accuracy; that is, the credible volumes are smaller when the same data is analyzed globally (c). When measurements to additional satellites are introduced (global analysis of the extended network) the localization accuracy improves substantially (d). When FRET anisotropy is measured in addition to FRET efficiency, the localization accuracy improves dramatically (e) and suggests that the unknown average TDM orientations are the main source of localization uncertainty. When finite-sized satellite position priors (uniform within a $(8\ \text{Å})^3$ cube, shown in gray) are used (f) in contrast to exact satellite positions (e) the accuracy decreases slightly but still accurate position information is obtained.

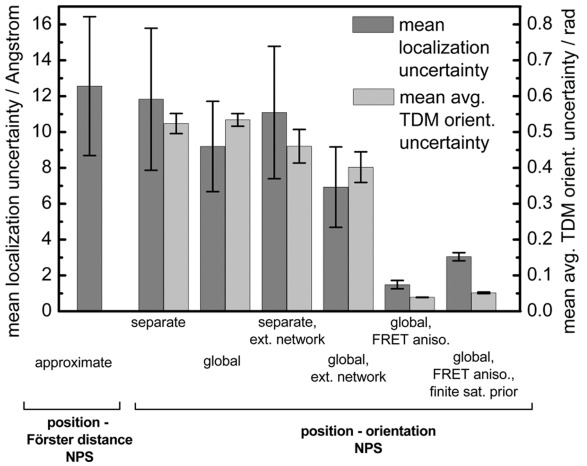


Figure 5. Mean uncertainties of localization and average TDM orientation. Mean localization uncertainty, $\langle\overline{\sigma^{x}}\rangle$, and mean average TDM orientation uncertainty, $\langle\overline{\sigma^{\Omega}}\rangle$, shown for different scenarios together with the corresponding standard deviation (whiskers) indicating the spread of uncertainties for different antennas.

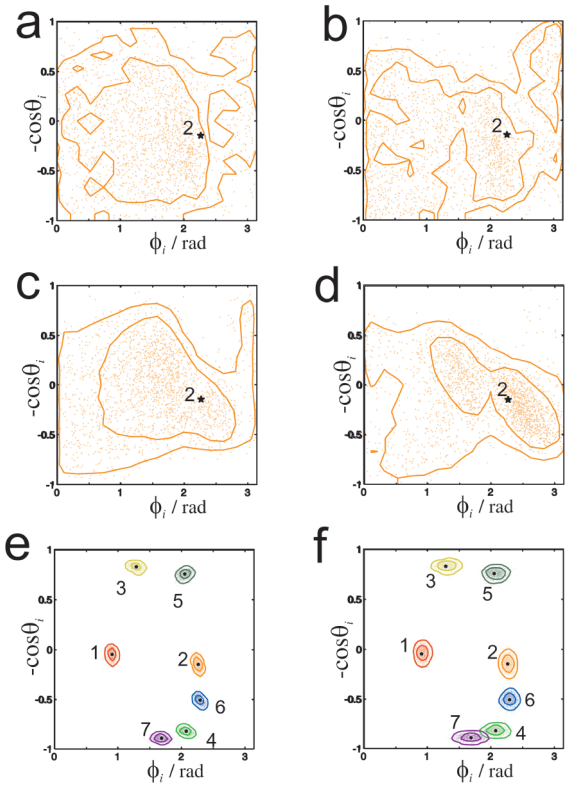


Figure 6. Estimates of average antenna TDM orientations. Scatter plots of marginal antenna average TDM orientations are shown together with 68% and 95% credible intervals (inner and outer countours) for antenna 2 (a–d) and all antennas (e and f). The true orientations are marked with black stars (a–d)/dots (e and f). In (a and b) and (c and d) the 7A/7S network and the extended network was analyzed, respectively. The accuracy of orientation is higher for global (b and d) than for separate analysis (a and c). Panels e and f show the estimated average TDM orientations in the presence of additional FRET anisotropy data. Finite-sized satellite position priors increase the orientation uncertainty only slightly (f) as compared to (e), where satellite positions were known with infinite accuracy, but still the orientation could be estimated accurately.

FRET efficiency measurements in the 7A/7S network (Figure 4a). The calculations were carried out almost entirely on an analytical basis but were restricted to the computation of the marginal position posterior PDF of a single antenna and included also the assumption of independent Förster distances. The same data was then analyzed with the position–orientation NPS, which will be referred to as separate analysis. In this analysis we performed computations completely numerically and could drop the assumption of independent Förster distances (Figure 4b). To study whether a global analysis of the data is advantageous as it exploits correlations between model parameters, we also analyzed the 7A/7S network at once, which we refer to as global analysis (Figure 4c).

In order to compare the different methods on a numerical basis, we computed the mean localization uncertainty $\langle\overline{\sigma^{x}}\rangle$, which quantifies the average size of the antenna position estimates (methods). Approximate and separate analysis applied to FRET efficiency only data give very similar results (Figures 4a, b) as indicated by $\langle\overline{\sigma^{x}}\rangle$ of 12.6 and 11.8 Å (Figure 5), which supports the validity of the approximations made in the position–Förster distance NPS. Global analysis yields in average smaller credible volumes (Figure 4c) and $\langle\overline{\sigma^{x}}\rangle$ of 11.1 Å (Figure 5) since correlations between Förster distances can be taken into account. The correlations are mediated not only by the antenna orientation like in the separate analysis but also by the satellite orientations.

While the above analysis yields quantitative localization of the satellites, oftentimes answering a biological question requires a higher accuracy. This can be done, for example, by introducing additional satellites into the FRET network at positions that promise to give informative data. Here, we chose to add two new satellites at promising positions (Figure 3a) and extended the FRET network to seven antennas and nine satellites (extended network). We analyzed this network globally with position–orientation NPS, referring to that scenario as extended network (Figure 4d). Although the introduction of additional satellites improved localization accuracy and $\langle\overline{\sigma^{x}}\rangle$ dropped to 6.9 Å (Figure 5), with every new satellite we introduced not only new data but also two more model parameters, namely the average TDM orientation. Unfortunately, by measuring only FRET efficiency one obtains very little direct information about these orientations, and moreover, it would be better to increase the amount of data while leaving the number of parameters the same. We therefore investigated how additional measurements of FRET anisotropy (the fluorescence anisotropy of the acceptor after polarized excitation of the donor) would help to reduce the localization uncertainty. As FRET anisotropy depends only on the angle between the average TDM orientations, while FRET efficiency depends on both, the fluorophore distance and the relative average TDM orientations, FRET anisotropy measurements are a source of independent information about the TDM orientations. After introducing additional FRET anisotropy data in the global NPS analysis of the 7A/7S network we observed drastic improvement of localization accuracy (Figure 4e) yielding $\langle\overline{\sigma^{x}}\rangle$ of only 1.5 Å (Figure 5). This suggests that the lack of orientation information is the main cause of localization uncertainty in the network studied here. We find also that the "true" antenna positions are located closer to the center of the credible volumes, which is expected in the absence of experimental noise. We observe this pronounced improvement of the localization

accuracy only if the FRET efficiency and FRET anisotropy data of the whole FRET network is analyzed globally, and find a much smaller effect when we analyze the network separately ($\langle\overline{\sigma^{\mathbf{x}}}\rangle$ = 9.5 Å, Supporting Information, Suppl. Figure 2).

Since we have now shown that with the position−orientation NPS we can accurately determine the antenna positions, we can release our initial constraint about perfectly known satellite positions and investigate how finite-sized volumes change the localization accuracies. We find that localization accuracy decreases (Figure 4f) and $\langle\overline{\sigma^{\mathbf{x}}}\rangle$ = 3.0 Å is twice as large as before (Figure 5). Thus it is desirable to have well-defined satellite positions, e.g., by using short linkers.

Up to now, we focused only on the marginal antenna position estimates, however, position−orientation NPS allows also to infer the average TDM orientations of each fluorophore. In analogy to the analysis of positions we define the mean average TDM orientation uncertainty, $\langle\overline{\sigma^{\mathbf{\Omega}}}\rangle$ (Supporting Information, Suppl. methods), which quantifies the spread of the average antenna TDM orientation estimates within each scenario.

We find that the improvement of the localization accuracy through the course of the studied scenarios is in most cases accompanied by a better estimation of the average TDM orientations as shown in Figures 6 and 5. While it is impossible to calculate the average TDM orientations precisely by separately analyzing FRET efficiency data as shown for a typical antenna in both the 7A/7S network (Figure 6a, $\langle\overline{\sigma^{\mathbf{\Omega}}}\rangle$ = 0.52 rad) and the extended (Figure 6b, $\langle\overline{\sigma^{\mathbf{\Omega}}}\rangle$ = 0.53 rad), global analysis allows already to compute rough estimates (Figure 6c,d, $\langle\overline{\sigma^{\mathbf{\Omega}}}\rangle$ = 0.46 and 0.40 rad, respectively). By additional measurement of FRET anisotropies, the average TDM orientations are accurately estimated (Figure 6e, $\langle\overline{\sigma^{\mathbf{\Omega}}}\rangle$ = 0.04 rad) even when the satellite position is not exactly known (Figure 6f, $\langle\overline{\sigma^{\mathbf{\Omega}}}\rangle$ = 0.05 rad). Thus, although the average TDM priors did not contain any information about orientation, and FRET anisotropy depends only on the relative angle between the average TDM orientations (eq 5), we could recover the orientation of average TDMs relative to the macromolecule.

NPS Docking. An important application of FRET measurements is to determine the position and orientation of proteins of known structure within large transient complexes. Therefore we explore here the capability of NPS docking to analyze FRET measurements between different macromolecules within a complex. Similar to the above localization of fluorophores we studied the influence of FRET efficiency only measurements, combined FRET efficiency and anisotropy measurements, and infinitely small as well as finite-sized fluorophore accessible volumes. While the inherently global analysis allows for docking of several components simultaneously, we applied it here to two components and used the same fluorophore positions, orientations and data as in the calculations before (Methods). In contrast to the previous section now all antennas were attached to a hypothetical yet unlocalized macromolecule. First, we studied how unknown fluorophore orientation limits the localization accuracy of the docked component and used infinitely small accessible volumes for both satellites and antennas, fixing hereby the relative positions of the antennas. We found that in this case FRET efficiency data was already sufficient for accurate docking (Figure 7a), which is reflected by the small localization uncertainty (see Methods) of the point at the origin of the docked frame of reference, $\overline{\sigma^{(0,0,0)}}$ = 1.1 Å. Additional FRET anisotropy

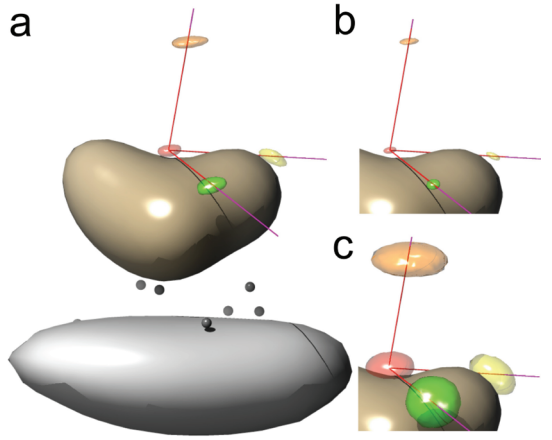


Figure 7. NPS Docking. A reference frame attached to a macromolecule (brown ocher surface) that was docked to another macromolecule (gray surface) is shown. The docking was done by analyzing either FRET efficiency only (a) or FRET efficiency together with FRET anisotropy (b and c) between satellites (gray spheres) and antennas (not shown). Calculation results are displayed as 95% credible volumes (colored semitransparent surfaces) of the origin (red) and three points (orange, yellow, and green) located on each coordinate axis 40 Å away from the origin. The docking remains well-defined even when uncertainties in the positions of satellites and antennas are introduced (c).

data improved docking (Figure 7b, $\overline{\sigma^{(0,0,0)}}$ = 0.6 Å). After introduction of finite-sized antenna and satellite accessible volumes we observed that in the analysis of FRET efficiency and FRET anisotropy data the localization uncertainty increased strongly ($\overline{\sigma^{(0,0,0)}}$ = 2.5 Å) (Figure 7c). This effect limiting the localization accuracy here can be easily understood as satellite and antenna attachment contribute both to the localization uncertainty. Especially in a docking application it is thus desirable that the dye positions are accurately known with respect to the macromolecule they are attached to.

Alternative FRET Localizaion Experiment Design. We have shown that the localization accuracy of FRET experiments can be limited to a large extent by the unknown average transition dipole orientations of the fluorophores even when fluorescence anisotropies are moderate. A common strategy to reduce orientation effects (and therefore fluorescence anisotropy) is the usage of long linkers for fluorophore attachment. While this strategy can still be a good choice when a change in FRET efficiency should be interpreted as distance change in just one FRET pair, we have shown that even when orientation information is inferred from several FRET measurements, long linkers can decrease the localization and docking accuracy (Figures 4e,f and 7). Moreover, long linkers can potentially lead to different conformations with measurably different photophysical properties, especially when the dyes can stick to the protein.[44] This, in turn, can be misinterpreted as conformational dynamics of the macromolecule and thus complicate the analysis.

We propose therefore to use fluorophores with short, and when possible, even bivalent linkers used in polarization microscopy,[4,45−47] when both FRET efficiency and FRET anisotropy can be measured in a sufficiently large FRET network.

In this way, the accuracy decrease caused by inaccurate satellite fluorophore positions will be minimized even though strong orientation effects will appear. As these orientation effects can be accounted for in the position−orientation NPS, FRET data

measured under these experimental conditions can be used to accurately determine both the position and orientation of fluorophores. Also the presence of dynamic or heterogeneous FRET efficiencies and anisotropies, which should be accounted for in future NPS models, can be then directly related to potentially interesting conformational changes or structural heterogeneity of the macromolecule. It would be also possible to apply NPS docking and precisely determine the orientation of a macromolecule within a complex by using only a few labeling sites if the average TDM orientation of satellites were known a priori. Measurements of the absolute orientation of the fluorophore[48] or macromolecule[46,47,49] requiring immobilization of the complex in a known orientation could be avoided in this way.

The benefit of shorter linkers and the measurement of FRET anisotropy in addition to FRET efficiency has been demonstrated recently by Sindbert et al.,[33] who measured distances between dyes attached to nucleic acids. There, the authors determined distances based on FRET efficiencies averaged over all possible TDM orientations and accessible dye positions and assumed that the dyes must be able to perform a slow, completely unconstrained rotational and translational motion. Due to this averaging, they were not able to recover fluorophore orientations and did account for orientational effects on the level of Förster distances, which is in a distant sense comparable to our position−Förster distance NPS. Though their approach seems to work for intermediate and long linkers they reported that it fails for short linkers as the dyes start to stick to the nucleic acids. In contrast, the theory discussed here should work well also if very short linkers are used or if the fluorophores are hindered in their rotation, like for example in our previous measurements of RNA polymerase II.[16,23,24] Although Sindbert et al.[33] used FRET anisotropy measurements to constrain the range of possible Förster distances, their analysis cannot account for correlations in the Förster distances of a FRET network, and is thus unable to achieve the localization accuracy presented here by globally analyzing FRET efficiency and FRET anisotropy data with position−orientation NPS.

Model Selection and Data Preprocessing. In all preceding analyses we assumed to know the number of possible states the macromolecule can exist in, and based thereon, we could construct a suitable model describing the data. For example, one could think of one antenna fluorophore located in two different positions depending on the conformation of the macromolecule, so that two antennas (instead of only one) should be used in the NPS analysis. However, in most cases the number of conformational states needs to be derived from the data. This can be easy if the state of the macromolecule can be controlled, e.g., by purification or by stabilization of a complex with a ligand, but it can be difficult to determine the number of states if control is not possible.

Choosing the number of states based on the experimental data is a special case of model selection, and it is crucial for the biological interpretation of the experiment. If single-molecule FRET (sm-FRET) is monitored over time, one can use hidden Markov models (HMM) and apply either maximum-likelihood estimation (MLE)[34,35] or Bayesian data analysis.[36] If time information is not available or extracting information about possible underlying dynamic processes is not necessary, popular methods to analyze sm-FRET data are fitting of histograms,[9,13,14,16,23,24] probability distribution analysis (PDA),[37,38] and proximity ratio histograms (PRH).[39] In most of the above approaches, the number of states is typically obtained by some heuristic rule based on the goodness of the fit or by visual inspection of the data. In Bayesian data analysis, however, the number of conformational states is determined by selecting the model M with K_M states which maximizes the evidence that is proportional to the posterior probability of the model M.[36]

From the point of view of NPS, the methods above, referred to as preprocessing methods in the following, are very important as they allow one to extract the FRET efficiencies and FRET anisotropies from the measured data. Furthermore, since up to now model selection was applied to study sm-FRET data measured for only one pair of labeling sites, NPS can yield structural information that could be very valuable to select the correct number of conformational states. For example, a state with a highly improbable fluorophore position (e.g., inside of the macromolecule) can be easily ruled out, something that is not possible with methods which cannot account for data of several different pairs of labeling sites simultaneously. It is thus necessary to combine preprocessing approaches with NPS and to obtain the posterior probability of a structural model.

In order to compare different models we derive how the evidence obtained in an NPS calculation can be used together with the results of sm-FRET data preprocessing, so that it is possible to determine the number of conformational states based on FRET measured between several different labeling sites. We assume that initially every model is equally probable and that a constant prior for the FRET observables (i.e., FRET efficiency and FRET anisotropy) is used in the preprocessing. We find that the model probability (full derivation given in the Supporting Information, Suppl. discussion 4.2), which is proportional to the evidence Z_M, is given by

$$p(M|\text{data}, I) \propto Z_M = \frac{Z_M^{\text{PP}}}{\pi_M^{\text{PP,O}}} K_M! Z_M^{\text{NPS}} \tag{15}$$

where Z_M^{PP} is the evidence calculated with preprocessing, $\pi_M^{\text{PP,O}}$ is the value of the constant prior in all FRET observables abbreviated by O, and K_M is the number of states in the model M. Z_M^{NPS} is the evidence obtained in a position−orientation NPS calculation (eq 7). The computed evidence for all global analysis schemes discussed here can be found in the Supporting Information, Suppl. Table 5.

Since least-squares-fitting FRET histograms is a popular way of extracting FRET efficiencies and FRET anisotropies, we also derive how the results of this preprocessing method can be used to approximately calculate the evidence Z_M and hence the probability of the model M. We assume that one or more distributions (e.g., Gaussians) are fitted to the histogram of FRET efficiencies (or FRET anisotropies) of each measured pair of dyes. Each distribution is characterized by a center FRET efficiency (or FRET anisotropy) value and a width parameter (e.g., standard deviation). In that case the evidence of the model is approximately given by

$$Z_M \approx \text{const} \cdot W^{-1} (2\pi)^{N_M} \sqrt{\det C} \exp\left[-\frac{{\chi_{\min}}^2}{2}\right] K_M! Z_M^{\text{NPS}} \tag{16}$$

where W^{-1} is the inverse of the product of the widths of all peaks in all fits (both FRET efficiency and FRET anisotropy), N_M is the total number of peaks fitted in all histograms, $\det C$ is the product of all determinants of covariance matrices, and ${\chi_{\min}}^2$ is the sum of all minimized χ^2 values obtained. The constant const is the same

for all models used to fit the same data set, thus Z_M can be used to select models based on their posterior probability. The full derivation of eq 16 is given in the Supporting Information, Suppl. discussion 4.2.2.

In order to be valid, eq 16 requires as a prerequisite that the center values obtained in the fit lie several standard errors away from the minimum and maximum possible values of the observables. For example, the center FRET efficiency E of a Gaussian determined within ΔE standard error should satisfy $a(\Delta E) < E < 1 - a(\Delta E)$, where a should be set to 2 or more. One must also make sure for each fit that the χ^2 surface in the proximity of the estimated parameters is approximately a quadratic function of the parameters.

A special commonly occurring application of eqs 15 and 16 is the comparison of models with the same number of conformational states K_M but with unknown assignment of FRET peaks to conformational states. In that case the probability of model M can be computed directly from Z_M since all remaining terms in eq 16 are constant and do not depend on the particular model.

CONCLUSION

We have extended NPS to the analysis of FRET networks with multiple satellites and antennas by dropping the approximations made previously[23] and analyzing the data on a fully numerical basis. We compared the new, more general variant (position−orientation NPS), to the previous approach (position−Förster distance NPS) and analyzed FRET efficiency networks containing only one antenna. Both NPS variants produced similar results indicating that the approximations made in the position−Förster distance NPS hold, and that the MCMC calculations work correctly.

By analyzing a FRET network consisting of several satellites and antennas, we demonstrated that the localization accuracy can be improved by global data analysis of a sufficiently large network. Further, we studied how the introduction of FRET anisotropy in addition to FRET efficiency data influences localization, and found that FRET anisotropy can be used to estimate the average TDM orientations of the dyes and thus eliminate an important source of uncertainty inherent to FRET localization methods that rely only on FRET efficiency measurements.

Based upon these findings, we propose to use short or even bivalent linkers rather than the commonly used long linkers for dye attachment. In this alternative FRET localization experiment design, one aims to achieve a well-defined fluorophore position and orientation, which can be determined by NPS, so that possible dynamics or heterogeneity of FRET can be directly attributed to the macromolecule.

We could also demonstrate docking using FRET measurements by parametrizing the position and orientation of an unlocalized but structurally known macromolecule part, describing antenna positions and average TDM orientations relative to this part, and analyzing the problem with position−orientation NPS.

Finally, we derive how common FRET data preprocessing techniques can be coupled to NPS. In this way it is possible to select structural models of a macromolecule or a macromolecular complex based on their probability and to determine the number of conformational states. We point out that the probability of the model is calculated from FRET data measured between several different labeling positions and allows to combine several FRET data sets in an consistent way, whereas up to now only the analysis of one pair of labeling positions was possible.

ASSOCIATED CONTENT

Supporting Information. Supplementary theory, supplementary methods, supplementary results, supplementary discussion, and supplementary tables. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: jens.michaelis@uni-ulm.de. Tel: +49 731 50 23050. Fax: +49 731 50 23059.

ACKNOWLEDGMENT

We are thankful to Malte Gersch and Moritz Haag for programming and testing of important parts of the position−orientation NPS software. We also acknowledge financial support by the European Research Council (starting investigator grant), Deutsche Forschungsgemeinschaft (SFB 646), the Center for Integrated Protein Science Munich, and the Nanosystems Initiative Munich.

REFERENCES

(1) Förster, T. *Ann. Phys-Leipzig* **1948**, *437*, 55–75.
(2) Ha, T.; Rasnik, I.; Cheng, W.; Babcock, H. P.; Gauss, G. H.; Lohman, T. M.; Chu, S. *Nature* **2002**, *419*, 638–641.
(3) Stryer, L.; Haugland, R. P. *Proc. Natl. Acad. Sci. U.S.A.* **1967**, *58*, 719–726.
(4) Peterman, E. J. G.; Sosa, H.; Moerner, W. E. *Annu. Rev. Phys. Chem.* **2004**, *55*, 79–96.
(5) Schuler, B.; Eaton, W. A. *Curr. Opin. Struct. Biol.* **2008**, *18*, 16–26.
(6) Roy, R.; Hohng, S.; Ha, T. *Nat. Methods* **2008**, *5*, 507–516.
(7) Bräuchle, C., Lamb, D. C., Michaelis, J., Eds.; *Single Particle Tracking and Single Molecule Energy Transfer*; Wiley-VCH: Weinheim, Germany, 2010.
(8) Mukhopadhyay, J.; Kapanidis, A. N.; Mekler, V.; Kortkhonjia, E.; Ebright, Y. W.; Ebright, R. H. *Cell* **2001**, *106*, 453–463.
(9) Kapanidis, A. N.; Margeat, E.; Laurence, T. A.; Doose, S.; Ho, S. O.; Mukhopadhyay, J.; Kortkhonjia, E.; Mekler, V.; Ebright, R. H.; Weiss, S. *Mol. Cell* **2005**, *20*, 347–356.
(10) Kapanidis, A. N.; Margeat, E.; Ho, S. O.; Kortkhonjia, E.; Weiss, S.; Ebright, R. H. *Science* **2006**, *314*, 1144–1147.
(11) Ermolenko, D. N.; Spiegel, P. C.; Majumdar, Z. K.; Hickerson, R. P.; Clegg, R. M.; Noller, H. F. *Nat. Struct. Mol. Biol.* **2007**, *14*, 493–497.
(12) Ermolenko, D. N.; Majumdar, Z. K.; Hickerson, R. P.; Spiegel, P. C.; Clegg, R. M.; Noller, H. F. *J. Mol. Biol.* **2007**, *370*, 530–540.
(13) Santoso, Y.; Joyce, C. M.; Potapova, O.; Reste, L. L.; Hohlbein, J.; Torella, J. P.; Grindley, N. D. F.; Kapanidis, A. N. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 715–720.
(14) Margittai, M.; Widengren, J.; Schweinberger, E.; Schröder, G. F.; Felekyan, S.; Haustein, E.; König, M.; Fasshauer, D.; Grubmüller, H.; Jahn, R.; Seidel, C. A. M. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 15516–15521.
(15) Rasnik, I.; Myong, S.; Cheng, W.; Lohman, T. M.; Ha, T. *J. Mol. Biol.* **2004**, *336*, 395–408.
(16) Andrecka, J.; Lewis, R.; Brückner, F.; Lehmann, E.; Cramer, P.; Michaelis, J. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 135–140.
(17) Woźniak, A. K.; Schröder, G. F.; Grubmüller, H.; Seidel, C. A. M.; Oesterhelt, F. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 18337–18342.
(18) Auclair, S. M.; Moses, J. P.; Musial-Siwek, M.; Kendall, D. A.; Oliver, D. B.; Mukerji, I. *Biochemistry.* **2010**, *49*, 782–792.
(19) Mekler, V.; Kortkhonjia, E.; Mukhopadhyay, J.; Knight, J.; Revyakin, A.; Kapanidis, A. N.; Niu, W.; Ebright, Y. W.; Levy, R.; Ebright, R. H. *Cell* **2002**, *108*, 599–614.

(20) Mukhopadhyay, J.; Sineva, E.; Knight, J.; Levy, R. M.; Ebright, R. H. *Mol. Cell* **2004**, *14*, 739–751.
(21) Knight, J. L.; Mekler, V.; Mukhopadhyay, J.; Ebright, R. H.; Levy, R. M. *Biophys. J.* **2005**, *88*, 925–938.
(22) Choi, U. B.; Strop, P.; Vrljic, M.; Chu, S.; Brunger, A. T.; Weninger, K. R. *Nat. Struct. Mol. Biol.* **2010**, *17*, 318–324.
(23) Muschielok, A.; Andrecka, J.; Jawhari, A.; Brückner, F.; Cramer, P.; Michaelis, J. *Nat. Methods* **2008**, *5*, 965–971.
(24) Andrecka, J.; Treutlein, B.; Arcusa, M. A. I.; Muschielok, A.; Lewis, R.; Cheung, A. C. M.; Cramer, P.; Michaelis, J. *Nucleic. Acids. Res.* **2009**, *37*, 5803–5809.
(25) Uphoff, S.; Holden, S. J.; Reste, L. L.; Periz, J.; van de Linde, S.; Heilemann, M.; Kapanidis, A. N. *Nat. Methods* **2010**, *7*, 831–836.
(26) Majumdar, Z. K.; Hickerson, R.; Noller, H. F.; Clegg, R. M. *J. Mol. Biol.* **2005**, *351*, 1123–1145.
(27) Dale, R. E.; Eisinger, J. *Biopolymers.* **1974**, *13*, 1573–1605.
(28) Dale, R.; Eisinger, J.; Blumberg, W. *Biophys. J.* **1979**, *26*, 161–193.
(29) Hillel, Z.; Wu, C. W. *Biochemistry* **1976**, *15*, 2105–2113.
(30) Ivanov, V.; Li, M.; Mizuuchi, K. *Biophys. J.* **2009**, *97*, 922–929.
(31) Laplace, P.-S. *Théorie analytique des probabilités, 1. suppl.*; Courcier Imprimeur: Paris, 1812.
(32) Sivia, D. *Data Analysis*, 2nd ed.; Oxford University Press: New York, 2006.
(33) Sindbert, S.; Kalinin, S.; Nguyen, H.; Kienzler, A.; Clima, L.; Bannwarth, W.; Appel, B.; Müller, S.; Seidel, C. A. M. *J. Am. Chem. Soc.* **2011**, *133*, 2463–2480.
(34) McKinney, S. A.; Joo, C.; Ha, T. *Biophys. J.* **2006**, *91*, 1941–1951.
(35) Liu, Y.; Park, J.; Dahmen, K. A.; Chemla, Y. R.; Ha, T. *J. Phys. Chem. B* **2010**, *114*, 5386–5403.
(36) Bronson, J. E.; Fei, J.; Hofman, J. M.;, R. L. G., Jr.; Wiggins, C. H. *Biophys. J.* **2009**, *97*, 3196–3205.
(37) Antonik, M.; Felekyan, S.; Gaiduk, A.; Seidel, C. A. M. *J. Phys. Chem. B* **2006**, *110*, 6970–6978.
(38) Kalinin, S.; Felekyan, S.; Antonik, M.; Seidel, C. A. M. *J. Phys. Chem. B* **2007**, *111*, 10253–10262.
(39) Nir, E.; Michalet, X.; Hamadani, K. M.; Laurence, T. A.; Neuhauser, D.; Kovchegov, Y.; Weiss, S. *J. Phys. Chem. B* **2006**, *110*, 22103–22124.
(40) Lakowicz, J. R. *Principles of Fluorescence Spectroscopy*, 3rd ed.; Springer: New York, 2006.
(41) Skilling, J. *Bayesian Anal.* **2006**, *1*, 833–860.
(42) UCSF Chimera Software package. http://www.cgl.ucsf.edu/chimera, 2010.
(43) 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.
(44) Eggeling, C.; Fries, J. R.; Brand, L.; Günther, R.; Seidel, C. A. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 1556–1561.
(45) Corrie, J. E.; Brandmeier, B. D.; Ferguson, R. E.; Trentham, D. R.; Kendrick-Jones, J.; Hopkins, S. C.; van der Heide, U. A.; Goldman, Y. E.; Sabido-David, C.; Dale, R. E.; Criddle, S.; Irving, M. *Nature* **1999**, *400*, 425–430.
(46) Peterman, E. J. G.; Sosa, H.; Goldstein, L. S.; Moerner, W. E. *Biophys. J.* **2001**, *81*, 2851–2863.
(47) Forkey, J. N.; Quinlan, M. E.; Goldman, Y. E. *Biophys. J.* **2005**, *89*, 1261–1271.
(48) Ha, T.; Laurence, T. A.; Chemla, D. S.; Weiss, S. *J. Phys. Chem. B* **1999**, *103*, 6839–6850.
(49) Rosenberg, S. A.; Quinlan, M. E.; Forkey, J. N.; Goldman, Y. E. *Acc. Chem. Res.* **2005**, *38*, 583–593.