Development and Application of a Quantitative Analysis Method for Fluorescence Resonance Energy Transfer Localization Experiments

Bayesian Inference of Macromolecule Structures

Adam Marek Muschielok

aus

Pyskowice / Peiskretscham, Polen

München 2011

Erklärung
Diese Dissertation wurde im Sinne von § 13 Abs. 3 der Promotionsordnung vom
29. Januar 1998 (in der Fassung der vierten Änderungssatzung vom 26. November 2004)
von Herrn Prof. Dr. Jens Michaelis betreut.

Ehrenwörtliche Versicherung
Diese Dissertation wurde selbständig, ohne unerlaubte Hilfe erarbeitet.

München, den 23. November 2010

Adam Muschielok

Abstract

This thesis focuses on the Bayesian data analysis of single-molecule fluorescence reso-
nance energy transfer (FRET) experiments carried out to infer structural information of
biological macromolecules and macromolecular complexes labeled with fluorophores.

In general, the measurement of FRET efficiencies allows to determine the distance
of two fluorophores quantitatively and thus acts as a molecular ruler on the nanometer
length scale. Moreover, it is possible to calculate the yet unknown position of one or more
fluorophores relative to each other by trilateration. In doing so, FRET efficiency measure-
ments between the fluorophore to be localized, called *antenna*, and several fluorophores at
known locations, called *satellites*, are used to constrain the position of the antenna. This,
in turn, can be used to answer questions in structural biochemistry, when the antennas are
attached for example to a yet unlocalized constituent of a macromolecular complex, while
the satellites are linked to positions known from e.g. X-ray crystallography experiments.

However, FRET efficiency depends not only on the fluorophore separation but also
on the orientation of the transition dipole moments of the fluorophores. This makes a
simple conversion of FRET efficiencies into distances inapplicable, as the orientations are
usually unknown. Although these orientation effects were known to be a major source
of uncertainty in FRET-based localization experiments, they have often been ignored or
argued away in the literature.

The main result of this thesis is a novel data analysis tool, which was developed in order
to account for both distance and orientation effects and thus allowing an accurate FRET-
based localization. This tool was called Nano-Positioning System (NPS) and applies
Bayesian data analysis to infer possible positions of fluorophores and optionally also the
positions and orientations of the subunits of a macromolecular complex.

The results of NPS can be readily displayed in the form of probability densities that
reflect the information about the position of a fluorophore or a particular subunit in
a macromolecular complex. The location, form and fuzziness of the densities impart
simultaneously the position and the uncertainty of localization. Shown together with an
already known macromolecule structure, these densities can then be interpreted in an
intuitive way.

NPS was successfully tested by localizing a fluorophore attached to a known position
in yeast RNA polymerase II (Pol II) elongation complexes (ECs). To demonstrate the
practical value of NPS, it was applied to study the influence of the transcription factor
IIB (TFIIB) on the position of the nascent RNA as well as to map the pathway of the
nontemplate and upstream DNA in yeast Pol II ECs. Furthermore, the position and ori-
entation of the TATA binding protein (TBP) in initial transcribing complexes of Pol II
were inferred. A deeper understanding of NPS was obtained by analyzing synthetic data.
Unknown fluorophore orientations were found to be indeed the major source of localiza-
tion uncertainty under commonly encountered experimental conditions. Synergy effects
emerging from the simultaneous analysis of a FRET network containing several antenna
and satellite fluorophores were observed to improve the accuracy of the inference. It is
proposed to use FRET anisotropy data in addition to the commonly measured FRET effi-
ciencies to calculate accurate fluorophore orientations and thus dramatically increase the
localization accuracy of NPS. Finally, general aspects of NPS are discussed and possible
future improvements of NPS are pointed out.

Contents

Contents

1 Introduction

Nowadays, many different experimental methods exist to study the structure and function of biologically relevant macromolecules like proteins, nucleic acids and protein complexes. They cover a wide range of applications from fundamental research to rational drug design in pharmaceutical industry.

This thesis focuses on an application of fluorescence to structural studies of macromolecules. The word "structure" will be used throughout in the sense of the three-dimensional alignment of subunits, e.g. amino acids in a protein. In order to put the method studied into a wider context, an overview of the most common techniques used in structural biology will be outlined.

The probably furthest developed standard methods for the determination of three-dimensional macromolecular structures are X-ray crystallography (Drenth, 1999), nuclear magnetic resonance (NMR) spectroscopy (Wüthrich, 1986) and single particle electron microscopy (EM) (van Heel et al., 2000; Frank, 2006). X-ray crystallography and single particle EM probe directly the positions of atoms, whereas NMR spectroscopy, at least when applied to large molecules, uses the distance-dependent coupling of nuclear spins to reconstruct a three-dimensional structure.

The structures obtained by NMR spectroscopy or X-ray crystallography have usually very high resolution, oftentimes even down to the size of atoms. Both are ensemble-based methods and rely on the presence of many identical molecules that contribute to the detected signal. However, those methods fail when the studied ensemble consists of molecules in too many conformations that each give rise to slightly different signals, which in turn cannot be resolved any more. Moreover, too large disorder or structural differences can even inhibit crystallization and thus make X-ray crystallography inapplicable and, in general, large macromolecules are difficult to solve with NMR spectroscopy because of anisotropy effects (Griswold and Dahlquist, 2002).

In this case, single particle EM can be used to collect data from single macromolecules separately. Since the scattering signal of single macromolecules is weak, many of them need to be observed in order to obtain a structure. In addition, the overall resolution is often lower compared to X-ray crystallography and NMR spectroscopy. However, because the data is acquired on a single-molecule level, one can resolve different conformation populations even when their structures differ much.

Conformation inhomogeneities, when observed, may be either static or dynamic, i.e. the macromolecules might remain in a conformation or show transitions between different states. These transitions might be important for the biological function, but it is not possible to study them in single particle EM since the sample is usually shock-frozen or fixated.

The method used for structural studies in this work is the so-called Nano-Positioning System (NPS) (Andrecka et al., 2008). It is based on single-molecule fluorescence resonance energy transfer (FRET) and is capable of obtaining time-resolved structural information from single biomolecules under native conditions. Single-molecule FRET can therefore be used to resolve different macromolecule conformations as well as the transitions between those states. In particular, NPS can be applied to localize fluorescent dye molecules that are attached to regions of a macromolecule with yet unknown structure. Although the data obtained is of lower resolution compared to X-ray crystallography and

NMR spectroscopy it can be a valuable source of information when the standard methods of structural biology fail or when independent measurements are needed to reduce the impact of artifacts.

The scope of this work is the development of a quantitative data analysis method for FRET localization experiments such as the NPS. The information about transitions between conformation states will not be analyzed here, but instead, it will be discussed how to combine NPS with methods that extract this kind of information.

After an introduction of some relevant physical phenomena and of the data analysis methods used in this work, a quantitative framework for the analysis of NPS data will be developed, which accounts for various error sources that accompany the underlying FRET measurements.

The thesis focuses mainly on the methodical but yet interesting aspects of the construction of a consistent physical model, the data analysis and the visualization of the analysis results. Bayesian inference will be applied to analyze the NPS data. It is a powerful and general method, which is used for example in the interpretation of mass spectra (Schwarz-Selinger et al., 2001) and was already applied to NMR spectroscopy structure determination (Rieping et al., 2005).

Later on, an application of NPS to the eukaryotic RNA polymerase II elongation complex will be shown. In particular, the effect of the transcription factor IIB on the position of the nascent RNA, the localization of the nontemplate DNA strand, and the preliminary analysis of the position and orientation of the TATA binding protein in the initial transcribing complex will be demonstrated.

Based on model calculations, a strategy will be proposed to increase the localization accuracy of NPS by the additional measurement of FRET anisotropy data and by the adequate design of the FRET network. Finally, possible improvements and extensions of NPS will be outlined.

2 Basics

2.1 Fluorescence

Fluorescence is the emission of light accompanied by the transition of a molecule from the first or a higher electronically excited singlet state $(S_1, S_2, \dots)$ to the electronic ground state S_0 (Michalet et al., 2003; Lakowicz, 2006, chapter 1). In the following sections, some basics of single-molecule fluorescence spectroscopy are explained.

2.1.1 Physical model of fluorescence

In most biological applications, the excited state is populated by absorption of light, as shown in form of a *Jabłoński diagram* in figure 2.1a. The fluorophore that exists initially in the electronic ground state with low vibrational excitation is transferred by the absorption of a photon with frequency ν_{exc} to higher vibrational states of the electronic excited states. Caused by collisions with the solvent molecules, the excitation relaxes typically on the picosecond time scale to the lowest vibrational state of the first electronic excited state that is metastable on a nanosecond time scale. From there, the excitation can decay into vibrationally excited states of the electronic ground state by emission of a photon of frequency ν_{em}, and relax further, again due to collisions, to the vibrational ground state. Alternatively the excitation can be quenched by other processes, and the transition to the ground state is called non-radiative.

The average ratio of emitted photons to the average number of photons absorbed is called quantum yield Q. It can be expressed in terms of rates, i.e. probabilities per time interval, for the radiative and non-radiative transitions to occur from the excited state to the ground state. When Γ is the rate to emit a photon and the total of non-radiative processes is summarized in the rate k_{nr} the quantum yield of a fluorophore (Lakowicz, 2006, chapter 1) is given by

$$Q = \frac{\Gamma}{\Gamma + k_{\mathrm{nr}}}. \tag{2.1.1}$$

The quantum yield not only quantifies the relative importance of radiative and non-radiative excitation decay channels, but it is also of practical impact when applied to energy transfer.

Since absorption and emission occurs almost always from the vibrational ground states, the emitted photon has usually a lower energy compared to the absorbed photon. This phenomenon is called *Stokes shift*, and consequently, the absorption and fluorescence spectra are also shifted (Lakowicz, 2006, chapter 1) (figure 2.1b). The Stokes shift is the reason why fluorescence based methods are widely used to study single molecules: it allows the spectral separation of the emitted from the excitation light and allows to observe the comparatively weak fluorescence of single molecules in non-fluorescent surroundings.

2.1.2 Time-dependent fluorescence decay

It is possible to measure the time dependence of the fluorescent light photon flux $I(t)$ after a pulsed excitation at time $t = 0$. For a solution of identical fluorophores with a
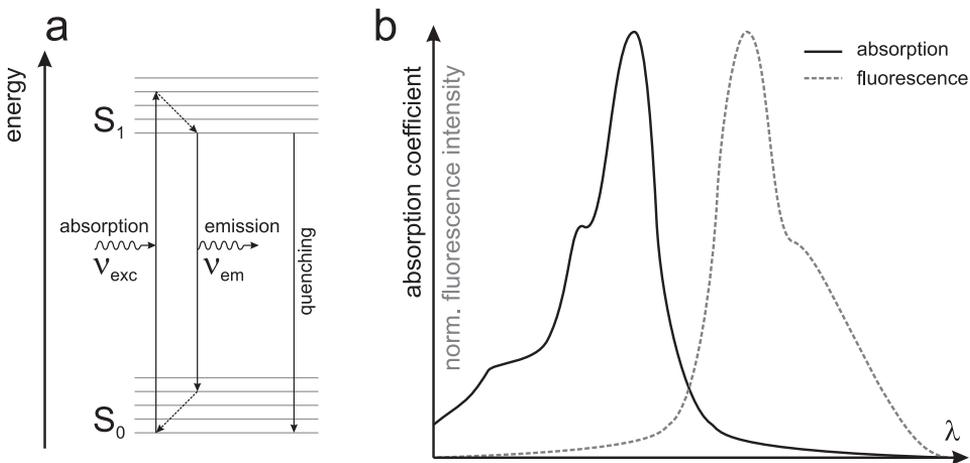
Figure 2.1: Scheme of fluorescence. (a) Jabłoński diagram of fluorescence. The energies of the singlet ground state (S_0) and the first electronic excited state (S_1) are shown together with their associated vibrational states. The fluorophore in S_0 absorbs a photon of the excitation light with frequency ν_{exc} and is elevated (arrow) to vibrationally excited states of S_1. From there the vibrational excitations relax rapidly on a picosecond time scale (dashed arrows). On a nanosecond time scale, the fluorophore in the lowest vibrational state of S_1 decays to vibrationally excited states in the electronic ground state and emits a photon of frequency ν_{em}, followed by a rapid relaxation to the vibrational ground state. Alternatively, the excitation can relax without emission of a photon into the electronic and vibrational ground state (quenching). Adapted from (Lakowicz, 2006, chapter 1). (b) Fluorescence spectra, sketch. Absorption coefficient and normalized fluorescence emission intensity are shown as function of the wavelength. Adapted from (Lakowicz, 2006, chapter 1).

single fluorescent transition the observed data can be described by an exponential decay,

$$I(t) \propto \exp\left(-\frac{t}{\tau}\right), \tag{2.1.2}$$

where $\tau = \Gamma^{-1}$ is the lifetime of the excited state S_1, which is termed often the fluorescence lifetime (Lakowicz, 2006). A non-exponential decay indicates a complex fluorescence transition scheme that might be caused e.g. by several conformations of the fluorophore exhibiting different lifetimes, or by a sequence of states populated after excitation but before the actual fluorescence occurs.

2.1.3 Fluorescence polarization

Another important property of light besides its frequency is the polarization state described by the polarization vector, $\hat{\boldsymbol{E}}$, i.e. the orientation of the electrical field component of light. Since absorption and fluorescence emission originate both in the interaction of the fluorophore with the electric field of light, the relative orientations of both, the fluorophore and the polarization vector, are of importance.

The probability for absorption of light with polarization $\hat{\boldsymbol{E}}_{ex}$ is proportional to $\cos^2(\angle(\hat{\boldsymbol{E}}_{ex}, \boldsymbol{\mu}_{abs}))$, the squared cosine of the angle between absorption transition dipole moment $\boldsymbol{\mu}_{abs}$ and $\hat{\boldsymbol{E}}_{ex}$. This means that fluorophores oriented with $\boldsymbol{\mu}_{abs}$ parallel to $\hat{\boldsymbol{E}}_{ex}$ are excited with highest probability, and fluorophores with $\boldsymbol{\mu}_{abs}$ perpendicular to $\hat{\boldsymbol{E}}_{ex}$ cannot
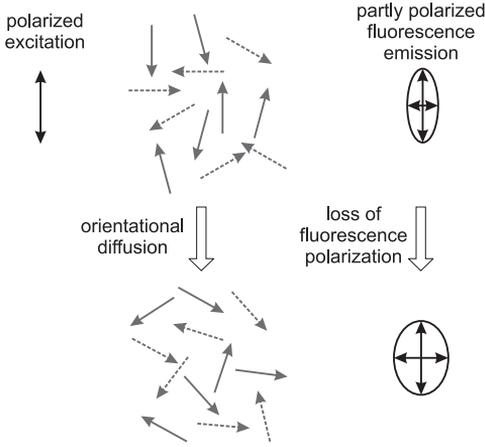
Figure 2.2: Fluorescence anisotropy. Randomly oriented, freely mobile fluorophores (gray arrows) are excited more likely by polarized light when their absorption transition dipole moments and the excitation polarization (double-headed arrow, left) make a small angle. The preferential orientation of excited fluorophores (solid gray arrows) is destroyed over time due to random orientation fluctuations. The amount of emitted fluorescent light is partly polarized (ellipses with arrows) immediately after excitation, and the degree of polarization decreases with time. Adapted from Lakowicz (2006, chapter 1).

be excited at all. Likewise, the probability to emit a photon with polarization $\hat{\boldsymbol{E}}_{\mathrm{em}}$ is proportional to $\cos^2(\angle(\hat{\boldsymbol{E}}_{\mathrm{em}}, \boldsymbol{\mu}_{\mathrm{em}}))$, the square of the cosine of the angle between emission transition dipole moment $\boldsymbol{\mu}_{\mathrm{em}}$ and $\hat{\boldsymbol{E}}_{\mathrm{em}}$. That, in turn, means that the fluorophore shows the typical transition dipole emission characteristics and emits light with preferential polarization along its emission transition dipole moment.

The polarization effects described above are called *photoselection* and *polarized emission*, respectively. By exploiting these effects it is possible to infer information about the angle between absorption and emission transition dipole moments and the orientational freedom of the fluorophores. An example of freely mobile fluorophores with parallel absorption and emission transition dipole moments is shown in figure 2.2. There, the excited fluorophores aligned mostly parallel to the excitation light polarization lose their preferential orientation with time by orientational diffusion while staying excited. Hence, the amount of polarization of the emitted light decays in time, an effect that will be discussed in more detail in the next section.

2.1.4 Fluorescence anisotropy

A quantity often used to describe the polarization effects introduced previously is the *fluorescence anisotropy r* defined as

$$r = \frac{I_{\parallel} - I_{\perp}}{I_{\parallel} + 2I_{\perp}}, \tag{2.1.3}$$

where $I_{\parallel}$ and $I_{\perp}$ denote the intensities of fluorescence light with polarization parallel and perpendicular to the polarization of the excitation light.

When a pulsed and polarized excitation light source is used, e.g. a pulsed laser, the time dependent fluorescence intensities $I_{\parallel}(t)$ and $I_{\perp}(t)$ can be measured, and in analogy to equation (2.1.3) the time dependent fluorescence anisotropy $r(t)$ is monitored. Here, t denotes the time after the pulsed excitation, and the fluorescence intensities parallel and perpendicular to the excitation light polarization can be expressed as

$$I_\parallel(t) = \frac{1}{3}I(t)\left[1 + 2r(t)\right]$$

$$I_\perp(t) = \frac{1}{3}I(t)\left[1 - r(t)\right]. \tag{2.1.4}$$

Usually, a decay of anisotropy is observed that can be related to the rotational diffusion of the fluorophores (Lakowicz, 2006, chapter 10). When the fluorophores can rotate without any constraints the fluorescence anisotropy approaches 0 for $t \to \infty$. When the fluorophore molecules are restricted in motion, e.g. because they are immobilized in a solid polymer matrix, $r(t)$ will approach a non-zero residual anisotropy, r_∞.

In the limit of $t \to 0$ the anisotropy approaches r_0, the so called fundamental or limiting anisotropy. For a solution of randomly oriented fluorophores it is related directly to the angle β between absorption and emission transition dipole moments by

$$r_0 = \frac{3\cos^2\beta - 1}{5}. \tag{2.1.5}$$

As a consequence of equation (2.1.5), $\frac{2}{5}$ and $-\frac{1}{5}$ are upper and lower bounds for the fundamental anisotropy under these conditions. Throughout this work it will be assumed that excitation and emission transition dipole moments are parallel and $r_0 = 2/5$.

The fluorescence anisotropy defined in equation (2.1.3) is called steady-state anisotropy, and it is a fluorescence weighted average of the time dependent anisotropy $r(t)$

$$r = \frac{\langle I(t)r(t)\rangle}{\langle I(t)\rangle}. \tag{2.1.6}$$

This can be derived easily when one considers that the steady-state intensities $I_\parallel$ and $I_\perp$ are time averages of the time dependent intensities $I_\parallel(t)$ and $I_\perp(t)$, respectively.

2.2 Fluorescence Resonance Energy Transfer (FRET)

When two fluorophores are in close proximity their electronic systems interact with each other directly without emitting and absorbing photons, and due to this coupling excitation energy is transferred in between the fluorophores. The farthest reaching interaction responsible for energy transfer is the dipolar coupling of the electronic systems, and the resulting energy transfer between fluorophores is termed Fluorescence Resonance Energy Transfer (FRET). The first correct quantum mechanical description of FRET originates back to Förster (1948), and the acronym is also often referred to as Förster Resonance Energy Transfer. In the Förster theory the fluorophores are treated as infinitely small point like dipoles, an approximation that holds well for fluorophore separations that are larger than the size of the fluorophore electron systems (Förster, 1948).

2.2.1 Physical model of FRET

A sketch of the FRET scheme is shown in figure 2.3. The fluorophore that is initially in the excited state is termed donor, and it can transfer its excitation to the other fluorophore, the acceptor, which is initially in the electronic ground state. Due to energy conservation one prerequisite for the observation of FRET is the overlap of the donor emission and acceptor absorption spectra, and that is also the reason for the term "resonance" in the acronym.
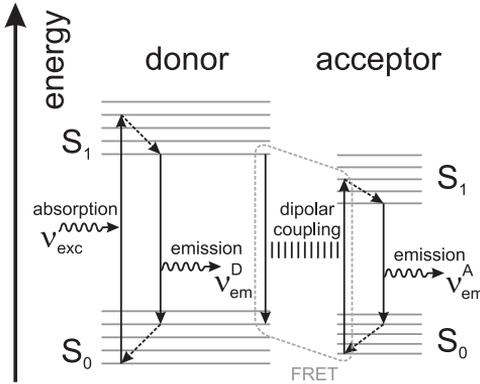
Figure 2.3: Jabłoński diagram of Fluorescence Resonance Energy Transfer (FRET). Additionally to the basic fluorescence scheme (see figure 2.1a) the excitation can travel from the donor to the acceptor fluorophore. This is caused by dipolar coupling of the electronic systems of donor and acceptor (lines between term systems). Effectively, the donor fluorescence is quenched and the acceptor is excited simultaneously. Finally, the acceptor relaxes to its ground state by emission of a photon of frequency $\nu_{\text{em}}^{\text{A}}$.

Depending on the strength of the coupling the excited donor can relax to its electronic ground state either by emission of a photon or by non-radiative energy transfer to the acceptor that would become fluorescent in the latter case. Likewise, as shown in figure 2.1 there might be also other processes that quench the fluorescence of donor and/or acceptor, and in an experiment one must account for these processes.

2.2.2 Definition of FRET efficiency

Similar to fluorescence, the energy transfer between two fluorophores can be described by a rate (Förster, 1948). In the absence of an acceptor the donor has the rate $k_{\text{D}} = 1/\tau_{\text{D}}$ of fluorescence emission, where τ_{D} is its fluorescence lifetime (see equation (2.1.2)). In the case of FRET the excitation is transferred to the acceptor with the rate k_{T} given by

$$k_{\text{T}} = k_{\text{D}} \frac{R^6}{d^6}. \tag{2.2.1}$$

d is the distance between the positions of the transition dipole moments of the fluorophores, and R is the characteristic length scale of the energy transfer, called Förster distance. The usual nomenclature for the Förster distance is R_0, but in this work this index will be omitted for better legibility. From equation (2.2.1) one recognizes immediately that the transfer rate increases the shorter the distance between the fluorophores.

A useful dimensionless parameter that quantifies the amount of energy transfer compared to the radiative emission of both donor and acceptor is the *FRET efficiency E* defined by

$$E = \frac{k_{\text{T}}}{k_{\text{D}} + k_{\text{T}}} = \frac{1}{1 + (d/R)^6}. \tag{2.2.2}$$

The range of FRET efficiency lies between 0 and 1. The lower limit is caused by an infinitely large separation of donor and acceptor or alternatively a Förster distance equal to zero. A FRET efficiency close to 1 must be handled with care since the assumptions of Förster theory might be violated.

The Förster distance is a function of different fluorophore properties like the overlap of the donor emission and acceptor absorption spectra, J, the quantum yield of the donor, Q_{D}, the transition moment orientations relative to the interconnecting line between the fluorophore positions that enter the orientation factor, κ^2, as well as the refractive index

Figure 2.4: Relative fluorophore orientations in FRET. The donor and acceptor fluorophores are located at the positions x_{D} and x_{A}, respectively. Their transition dipole moments (arrows) lie in the gray planes. The relative orientation of the transition dipole moments is defined by three angles. θ_{D} and θ_{A} are the angles between the donor and acceptor transition dipole moments and the line drawn between the fluorophores. ψ_{AD} is the angle between the transition dipole moments of donor and acceptor.
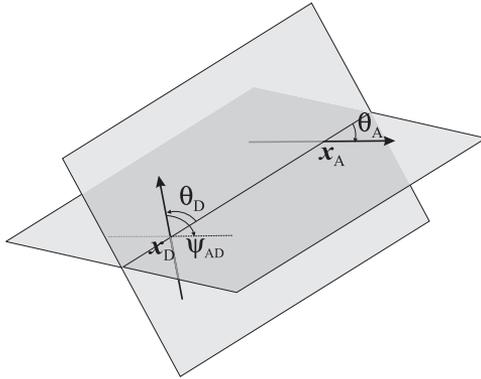
n of the medium surrounding the fluorophores:

$$R^6 = 8.79 \cdot 10^{-5} \frac{\text{\AA}^6 \, \text{cm} \, \text{M}}{\text{nm}^4} \cdot \frac{\kappa^2 Q_{\mathrm{D}} J}{n^4 N}. \tag{2.2.3}$$

N denotes the Avogadro number, and the spectral overlap J is calculated from the spectral density of the donor emission fluorescence intensity, $f_{\mathrm{D}}(\lambda)$, and the absorption spectrum (i.e. the absorption coefficient of the acceptor as function of the excitation wavelength), $\epsilon_{\mathrm{A}}(\lambda)$, by

$$J = \frac{\int_0^\infty d\lambda f_{\mathrm{D}}(\lambda)\epsilon_{\mathrm{A}}(\lambda)\lambda^4}{\int_0^\infty d\lambda f_{\mathrm{D}}(\lambda)} \tag{2.2.4}$$

The wavelength λ must be measured in nm, the fluorescence $f_{\mathrm{D}}(\lambda)$ can have arbitrary units (but must be a density normalized to unity area), and $\epsilon_{\mathrm{A}}(\lambda)$, the absorption coefficient, must be given in the units of $\mathrm{M}^{-1}\mathrm{cm}^{-1}$ in order to get the spectral overlap J in units of $\mathrm{M}^{-1}\mathrm{cm}^{-1}\mathrm{nm}^4$ and be compatible with equation (2.2.3).

For donor and acceptor fluorophores with a fixed relative orientation the orientation factor κ^2 is given by

$$\kappa^2 = \left(\cos\psi_{\mathrm{AD}} - 3\cos\theta_{\mathrm{D}}\cos\theta_{\mathrm{A}}\right)^2, \tag{2.2.5}$$

where ψ_{AD} is the angle between the transition dipole moments of the fluorophores, and $\cos\theta_{\mathrm{D}}$ and $\cos\theta_{\mathrm{A}}$ are the angles between the transition dipole moments of donor and acceptor and the line interconnecting the fluorophores (see figure 2.4).

When both donor and acceptor fluorophores are free to rotate and undergo fast reorientation compared to the fluorescence lifetime of the donor in the presence of the acceptor the orientation effects vanish and κ^2 adopts it isotropic value of $2/3$ irrespective of the relative position of the fluorophores. The Förster distance under these conditions will be called isotropic Förster distance and is abbreviated as R^{iso}. It is related to the Förster distance by

$$R = R^{\mathrm{iso}} \sqrt[6]{\frac{3}{2}\kappa^2}. \tag{2.2.6}$$

Fluorophore pairs chosen to optimize energy transfer have isotropic Förster distances usually between 20 and 90 Å (Lakowicz, 2006, chapter 13), which is a suitable length for studies of interaction of proteins and protein complexes.

2.2.3 Definition of FRET anisotropy

As described in the previous section, the energy transfer in between donor and acceptor is strongly dependent on the transition dipole moment orientations. Moreover, in the case of polarized excitation the polarization properties of the light emitted by donor and acceptor differ since the transfer process acts to reduce the degree of polarization. This effect is called *transfer depolarization* and it influences the *FRET anisotropy* r_{T}. The latter is defined as the fluorescence anisotropy of the acceptor fluorophore after the polarized excitation of the donor (equation (2.1.3)).

According to Dale et al. (1979) a randomly oriented ensemble of donors and acceptors fixed in space with the constant angle ψ_{AD} between their transition dipole moments shows the FRET anisotropy of

$$r_{\mathrm{T}} = \frac{2}{5} \cdot \frac{3\cos^2\psi_{\mathrm{AD}} - 1}{2}. \tag{2.2.7}$$

The FRET anisotropy carries thus information about the angle between donor and acceptor transition dipole moments. The limiting values are like for limiting anisotropies $-\frac{1}{5}$ and $\frac{2}{5}$. r_{T} vanishes when the angle between transition dipole moments is equal to the magic angle $\psi_{\mathrm{AD}} = \arccos(\frac{1}{\sqrt{3}}) \approx 54.7°$.

2.2.4 Measurement of FRET efficiency

In the experiments conducted to collect the data analyzed in this work the fluorescence of donor and acceptor fluorophores is monitored simultaneously. Since for the calculation of FRET efficiency it is sufficient to measure the number of photons per time, i.e. the photon flux, emitted by each type of fluorophore irrespective of the photon frequency, there is no need to measure a detailed spectrum. Instead, the photon flux is measured by detecting the photons within a spectral range selected by suitable optical filters. To this end, the fluorophores must be chosen so as to separate spectrally their fluorescent light.

Let $F_{\mathrm{D}}(t_i)$ and $F_{\mathrm{A}}(t_i)$ be the detector signals of donor and acceptor fluorescence acquired at the time t_i. $F_{\mathrm{D}}(t_i)$ and $F_{\mathrm{A}}(t_i)$ are proportional to the number of photons emitted by the donor and acceptor during the acquisition time of the detectors, respectively. The proportionality constant is determined experimentally since it depends on the fluorophore type, i.e. the emission spectrum, the optical instrumentation and the detectors.

Though there are excellent and sophisticated ways to extract the FRET efficiency from the measured fluorescence signals (Antonik et al., 2006; Nir et al., 2006; Liu et al., 2010b) the easiest method (Roy et al., 2008) will be presented here to demonstrate how E can be obtained.

Given perfect spectral separation of the donor and acceptor fluorescence channels and assuming that only the donor is excited[1], the fluorescence signals F_{D} and F_{A} are proportional on average to the rate of donor fluorescence, k_{D}, and the rate of energy transfer, k_{T}, respectively. Using equation (2.2.2) the measured FRET efficiency, E_{measured}, is obtained by

$$E_{\mathrm{measured}} = \frac{F_{\mathrm{A}}}{F_{\mathrm{A}} + \gamma F_{\mathrm{D}}}, \tag{2.2.8}$$

where the correction factor γ compensates for the different quantum yields of donor and acceptor as well as for the different detection efficiencies mentioned above.

---

[1]Often, the donor and acceptor fluorescence detection channels overlap, which is called spectral crosstalk. Furthermore, direct excitation of the acceptor can occur due to a non-zero absorption coefficient at the wavelength of the donor excitation light. Fortunately, these effects can be easily accounted for (see subsection 5.1.3, equation (5.1.1)).

Having seen how the FRET efficiency can obtained by fluorescence measurement the next section treats the theory of the particular scenario that occurs typically in biological applications in which the donor and acceptor fluorophores are bound to a macromolecule.

2.3 FRET in labeled macromolecules

Because of the strong dependence of FRET efficiency on the separation of donor and acceptor, and also due to the suitable range of distances, FRET is a popular tool to monitor distance changes (Clegg, 2002; Schuler and Eaton, 2008; Santoso et al., 2010) and measure absolute distances between directly interacting biomolecules like proteins, nucleic acids and macromolecular complexes (Mukhopadhyay et al., 2004; Knight et al., 2005; Andrecka et al., 2008). Stryer and Haugland (1967) used even the terminology of the "spectroscopic ruler" in respect to FRET.

To measure distances by FRET the macromolecules have to be labeled with donor and acceptor fluorophores, which are usually aromatic compounds that have absorption and fluorescence spectra in the visible and near-IR range. Consequently, their fluorescence can be detected efficiently with the currently available detectors like avalanche photo diodes, photomultipliers and highly sensitive charged coupled device (CCD) cameras.

Absolutely necessary for the application discussed in this work is the site-specific labeling, i.e. donor and acceptor fluorophores are covalently attached to two different positions on the macromolecule known at least from its primary structure. A popular labeling approach for proteins is the covalent attachment of fluorophore derivates that contain a maleimide group, which can be cross-linked with the thiol group of cysteins (Michalet et al., 2003). Site-specific labeling is then usually achieved by creating single-cystein mutants of the proteins of interest. Other labeling approaches include labeling with reactive unnatural amino-acids (Kapanidis and Weiss, 2002) and the labeling of lysines, but the latter approach is mostly unspecific because lysine is a very common amino acid in proteins.

In order to preserve the biological function a direct attachment of the fluorophore to the macromolecule is avoided and instead flexible linkers that have the length of several covalent bonds are used to tether the fluorophore. This section gives an introduction into the effects on fluorescence anisotropy and FRET caused by the flexible attachment of the fluorophores.

Furthermore, it will be assumed throughout this work that the macromolecule has a defined tertiary structure and does not show continuous structure fluctuations, which can be tested by increasing the time resolution of the instrument.

2.3.1 Fluorescence anisotropy of a macromolecule-bound fluorophore

When compared to direct immobilization of a fluorophore, one expects from fluorophores immobilized via flexible linkers that the orientational and spatial mobilities are increased but are still smaller when compared to free fluorophores.

While it is probably difficult to measure the spatial mobility of the fluorophore at length scales smaller than the linker, a common means to quantify the amount of orientational freedom is the measurement of the fluorescence anisotropies of donor and acceptor fluorophores attached to the macromolecules (Lakowicz, 2006, chapter 11). As shown in figure 2.5, the fluorophore is reoriented due to thermal fluctuations during the time it stays in the excited state and finally emits a photon at a different transition dipole moment orientation.
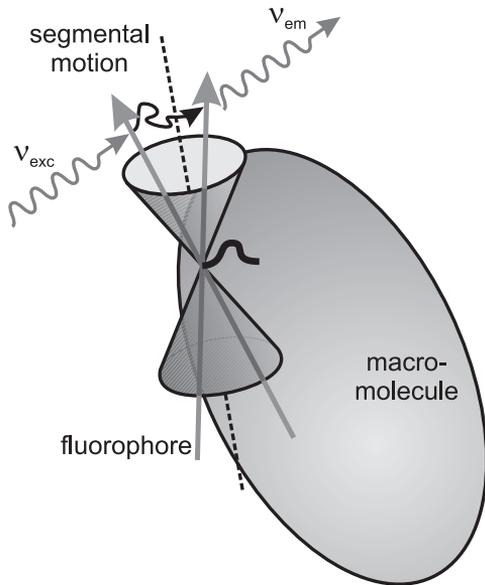
Figure 2.5: Standard model of a macromolecule-bound fluorophore exhibiting constrained segmental motion. The transition dipole moment of a fluorophore (arrow in the cone) attached to a large macromolecule is shown at the moments of absorption of a photon with frequency ν_{exc} and emission of a photon with frequency ν_{em}. In between absorption and emission the fluorophore undergoes a constrained segmental motion relative to the macromolecule. An axially symmetric orientation constraint imposed by the macromolecule is shown here as a cone, its axis (dashed line) pointing in the direction of the average transition dipole moment. On a short time scale the decay of fluorescence anisotropy is mainly caused by the segmental motion. On a longer time scale also the reorientation of the macromolecule will contribute to the decay. Adapted from Dale et al. (1979).

When there is a difference in the time scales of segmental motion of the attached fluorophore, ρ_f, and macromolecule rotation, ρ_p, then the time-resolved anisotropy will have an approximately bi-exponential decay (Lakowicz, 2006, chapter 11),

$$r(t) = (r_0 - r_\infty) \exp\left(-\left(\rho_f^{-1} + \rho_p^{-1}\right) t\right) + r_\infty \exp\left(-\rho_p^{-1} t\right). \qquad (2.3.1)$$

The slow decay on the time scale ρ_p is attributed to the overall rotational motion of the macromolecule, whereas the fast decay on the time scale $(\rho_f^{-1} + \rho_p^{-1})^{-1}$ is attributed to the reorientation of both the fluorophore and the macromolecule. r_∞ is the residual anisotropy in the reference frame of the macromolecule, i.e. as if all macromolecules were immobilized but would have random orientations. It quantifies the amount of orientational freedom of the fluorophore relative to the macromolecule.

In the limit $\rho_f \ll \tau \ll \rho_p$, where τ is the fluorescence lifetime, the steady-state fluorescence anisotropy, r, will approximately coincide with the residual anisotropy in the reference frame of the macromolecule as a consequence of equation (2.1.6), and the steady-state r can thus be used instead of r_∞.

2.3.2 Influence of segmental motion on the FRET efficiency and anisotropy

The segmental motion of the fluorophores introduced above does also have influence on FRET efficiency and anisotropy as pointed out by Dale and Eisinger (1974) and Dale et al. (1979). Their model describes the orientation factor κ^2 and the FRET anisotropy r_T in the intermediate region between completely immobile donor and acceptor molecules (equations (2.2.5) and (2.2.7)) and the complete orientation averaging limit resulting in $\kappa^2 = \frac{2}{3}$ and $r_T = 0$.

The proposed model is based on the assumption that both donor and acceptor fluorophores have fixed positions during the energy transfer but undergo a fast reorientation constrained by the macromolecule. In that case the relevant transitions, namely, the
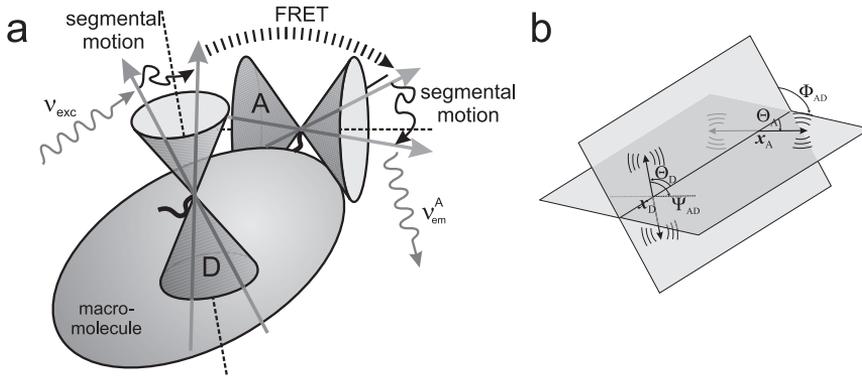
Figure 2.6: Model of FRET for fluorophores exhibiting constrained segmental motion. (a) FRET between flexibly attached fluorophores. The donor, symbolized by the direction of its transition dipole moment (arrows in the cone on the left), absorbs a photon of frequency ν_{exc} and undergoes a constrained segmental motion before the excitation is transferred to the acceptor (arrows in the cone on the right). Likewise does the acceptor before the photon of frequency ν_{em}^A is emitted. In the theory of Dale et al. (1979) both constraints are required to result in axially symmetric distributions of donor and acceptor transition dipole moment orientations, shown as cones labeled with a "D" and "A", respectively. The average orientations of the transition dipole moments are shown as dashed lines. The segmental motions of the fluorophores as well as the energy transfer are depolarizing steps. (b) Angle definitions in a FRET pair of flexibly attached donor and acceptor. A FRET pair is shown, the donor fluorophore at position x_D and the acceptor fluorophore at position x_A. The relative average orientations of the transition dipole moments are described by the angles Θ_D and Θ_A to the interconnecting line between the fluorophores, and the angle Φ_{AD} between the planes that contain the average transition dipole moment orientations and the interconnecting line. Ψ_{AD} is the angle between the average transition dipole orientations. Adapted from Dale et al. (1979).

absorption of a photon by the donor, the energy transfer to the acceptor and the final emission of a photon by the acceptor occur from different independent transition dipole moment orientations (figure 2.6a). More precisely, the reorientation time scales of both fluorophores must be much shorter than the lifetime of the donor in the presence of the acceptor, so that the fluorophores sample the possible orientations many times before the transfer takes place in average.

When these assumptions apply, i.e. the so called *dynamic orientational averaging limit* is fulfilled, the quickly fluctuating transfer rate can be replaced by its time average, and equations (2.2.1) and (2.2.2) are replaced by

$$\langle k_T \rangle = k_D \frac{\langle R^6 \rangle}{d^6} \text{ and} \tag{2.3.2}$$

$$E = \frac{1}{1 + d^6 / \langle R^6 \rangle}, \tag{2.3.3}$$

with the dynamically averaged Förster distance $\langle R^6 \rangle$ proportional to the corresponding average orientation factor $\langle \kappa^2 \rangle$,

$$\langle R^6 \rangle = \left(R^{iso} \right)^6 \frac{3}{2} \langle \kappa^2 \rangle. \tag{2.3.4}$$

Under the additional assumption that the orientation distributions of the fluorophores

are axially symmetric but remarkably without need of a more detailed model, Dale et al. (1979) derived the average value of the orientation factor as a function of the orientations of the symmetry axes, i.e. the average transition dipole moment orientations of donor and acceptor,

$$
\begin{aligned}
\left\langle \kappa^2 \right\rangle =\ & (\cos \Psi_{\mathrm{AD}} - 3 \cos \Theta_{\mathrm{A}} \cos \Theta_{\mathrm{D}})^2 \left\langle d_{\mathrm{A}}^x \right\rangle \left\langle d_{\mathrm{D}}^x \right\rangle \\
& + \left(1/3 + \cos^2 \Theta_{\mathrm{D}} \left\langle d_{\mathrm{D}}^x \right\rangle \right) \left(1 - \left\langle d_{\mathrm{A}}^x \right\rangle \right) \\
& + \left(1/3 + \cos^2 \Theta_{\mathrm{A}} \left\langle d_{\mathrm{A}}^x \right\rangle \right) \left(1 - \left\langle d_{\mathrm{D}}^x \right\rangle \right).
\end{aligned}
\tag{2.3.5}
$$

Here, Ψ_{AD}, Θ_{A} and Θ_{D} are the angles that define the relative orientation of the average donor and acceptor transition dipole moments (see figure 2.6b), and $\left\langle d_{\mathrm{D}}^x \right\rangle$ and $\left\langle d_{\mathrm{A}}^x \right\rangle$ are the average axial depolarizations, which are related to the residual fluorescence anisotropies of donor and acceptor by

$$
\left\langle d_{\mathrm{D}}^x \right\rangle^2 = \frac{r_{\mathrm{D},\infty}}{r_{\mathrm{D},0}} \quad \text{and} \quad \left\langle d_{\mathrm{A}}^x \right\rangle^2 = \frac{r_{\mathrm{A},\infty}}{r_{\mathrm{A},0}},
\tag{2.3.6}
$$

which can be written as

$$
\left\langle d_{\mathrm{D}}^x \right\rangle^2 = \frac{5}{2} r_{\mathrm{D}} \quad \text{and} \quad \left\langle d_{\mathrm{A}}^x \right\rangle^2 = \frac{5}{2} r_{\mathrm{A}}
\tag{2.3.7}
$$

if the segmental motion of the donor and acceptor is much faster than the fluorescence lifetime, and if the absorption and emission transition dipole moments of both donor and acceptor are parallel. It is worth mentioning that the axial depolarizations have an ambiguous sign when $r_{\mathrm{D/A},\infty}/r_{\mathrm{D/A},0} < 1/4$. Alternatively, the cosine of Ψ_{AD} can be expressed by the angles Θ_{A}, Θ_{D} and Φ_{AD}, the angle between the planes defined by the line interconnecting the fluorophores centers and each of the average transition dipole moments (figure 2.6b),

$$
\cos \Psi_{\mathrm{AD}} = \sin \Theta_{\mathrm{D}} \sin \Theta_{\mathrm{A}} \cos \Phi_{\mathrm{AD}} + \cos \Theta_{\mathrm{D}} \cos \Theta_{\mathrm{A}}.
\tag{2.3.8}
$$

Dale et al. (1979) derived also an expression for the transfer depolarization in the dynamic orientational averaging limit, $\left\langle d_{\mathrm{T}} \right\rangle$,

$$
\left\langle d_{\mathrm{T}} \right\rangle = \left\langle d_{\mathrm{D}}^x \right\rangle \frac{1}{2} \left(3 \cos^2 \Psi_{\mathrm{AD}} - 1 \right) \left\langle d_{\mathrm{A}}^x \right\rangle,
\tag{2.3.9}
$$

which is related to the average FRET anisotropy by

$$
\left\langle r_{\mathrm{T}} \right\rangle = \frac{2}{5} \left\langle d_{\mathrm{T}} \right\rangle.
\tag{2.3.10}
$$

In this way, the measurement of the fluorescence anisotropy allows for the characterization of the orientational freedom of the fluorophores, and measurement of the FRET anisotropy quantifies the angle between the average orientation of the transition dipole moments. That, in turn, reduces the inevitable uncertainties of the Förster distance due to the unknown angles Θ_{A} and Θ_{D}, and can be used to increase the accuracy of distance measurements as proposed by Dale et al. (1979) and Ivanov et al. (2009). However, these authors discuss only the minimum and maximum values κ^2 can take. In contrast, in this work the resulting uncertainties in position and distance will be treated in a more elegant way with probabilistic data analysis.

2.4 Probabilistic data analysis

Probabilistic data analysis, also called Bayesian data analysis after Reverend Thomas Bayes' theorem (Bayes, 1763) that is central to probability calculus, is based upon the mathematical representation of the degree of belief, or plausibility, as a probability or probability density. This interpretation was already used by Pierre-Simon de Laplace to infer the mass of Saturn (Laplace, 1812, premier supplément) and differs fundamentally from the nowadays common definition of probability as the relative frequency of an event taken to the limit of an infinite number of events.

Based on the excellent introduction to Bayesian data analysis by Jaynes (2003) and Sivia (2006), in this section, it will be explained how information about several competing hypotheses, e.g. the parameter values quantifying a model, can be expressed as probability, and how probability calculus can be used to merge different sources of information. That is, in turn, the essence of data analysis and will be applied later on in this work to the analysis of FRET data.

2.4.1 Probability calculus

Let X be a proposition or statement, and the probability, i.e. belief that X is true, will be written as $p\left(X|I\right)$. The first axiom of probability calculus is that by specifying $p\left(X|I\right)$ the probability $p\left(\overline{X}|I\right)$ that X is false is already defined implicitly, meaning that

$$p\left(X|I\right) + p\left(\overline{X}|I\right) = 1, \tag{2.4.1}$$

with $p\left(\text{false}\right) = 0$ and $p\left(\text{true}\right) = 1$ defining the certainty that something is either false or true, respectively. I denotes the *background information*, i.e. the state of knowledge of the observer, and the probabilities are *conditional* to I as symbolized by the vertical bar "|".

Equation (2.4.1) can be generalized to continuous variables, resulting in normalization of the probability density $p\left(X|I\right)$,

$$\int dX p\left(X|I\right) = 1. \tag{2.4.2}$$

The integration domain will be the whole range where X is defined (e.g. $-\infty < X < \infty$). This notation will be kept throughout this work. Equation (2.4.2) states that a probability density is normalized in all variables written left of the vertical bar. The converse is not true, and in general probability densities are not normalized in the condition variables right of the vertical bar.

When Y is a second proposition, the joint probability that X and Y are true, abbreviated as $p\left(X,Y|I\right)$ can be calculated by multiplying the probability that Y is true with the probability that X is true given Y is true,

$$p\left(X,Y|I\right) = p\left(X|Y,I\right) p\left(Y|I\right). \tag{2.4.3}$$

The product rule above is the second axiom of probability calculus, and Bayes' theorem (equation (2.4.4)) can be directly derived from it when the expansion is done with X and Y interchanged,

$$p\left(Y|X,I\right) = \frac{p\left(X|Y,I\right) p\left(Y|I\right)}{p\left(X|I\right)}. \tag{2.4.4}$$

Another useful corollary of equations (2.4.3) and (2.4.2) is the marginalization rule,

$$\int dY\, p\,(X,Y|I) = p\,(X|I)\,,\tag{2.4.5}$$

which works as well when the roles of X and Y are interchanged,

$$\int dX\, p\,(X,Y|I) = p\,(Y|I)\,.\tag{2.4.6}$$

It should be stressed again for clarity that a conditional probability density is not necessarily normalized in the condition variable, and consequently, the integration over a variable in the condition does not yield unity,

$$\int dY\, p\,(X|Y,I) \neq 1\,.\tag{2.4.7}$$

In the next section, it will be shown how to use Bayes' theorem and the marginalization rule in order to analyze experimental data.

2.4.2 Bayesian parameter estimation

Let the data d quantify the outcome of an experiment, and θ be the parameters that are part of a model that describes the measurements. For example, the measurement could be the acquisition of a signal and the parameter would be the "true" amplitude of the signal.

Usually, there will be no unique mapping between the model parameters and the data, for instance due to measurement errors, or because the data is inherently noisy. A common example for the latter case is measuring the light intensity by counting the number of photons. In such experiments, the number of photons (data) detected in a time interval is connected by a probability distribution, the Poisson distribution, to the light intensity, which is the average number of photons (model parameter) that determines the shape of the distribution.

The density $p\,(d|\theta,I)$ of the probability distribution that describes the expectation of possible data that could be produced in the experiment for given model parameters is called *likelihood* in Bayesian data analysis. The likelihood is of course conditional on the background information I that subsumes all information relevant to the experiment.

In order to extract information about the parameters in the light of the measured data, one must apply Bayes' theorem and calculate

$$p\,(\theta|d,I) = \frac{p\,(d|\theta,I)\,p\,(\theta|I)}{p\,(d|I)}\,.\tag{2.4.8}$$

$p\,(\theta|d,I)$ is called *posterior* in Bayesian data analysis, and is calculated by multiplying the likelihood with the *prior* $p\,(\theta|I)$ and dividing by the *evidence* $p\,(d|I)$.

The evidence $p\,(d|I)$, abbreviated often as Z, is a constant that normalizes the product of prior and likelihood. It becomes clear that the evidence is a constant since the data d does not vary once it is measured. Furthermore, when $p\,(d|I)$ is constant and the posterior $p\,(\theta|d,I)$ is proportional to the product of prior and likelihood, normalization is necessary since the likelihood is normalized only as a function of the data but not as function of the parameters. Consequently, the evidence can be computed by integrating the product of

likelihood and prior,

$$p\left(d|I\right) = \int d\theta \; p\left(d|\theta, I\right) p\left(\theta|I\right). \tag{2.4.9}$$

Though the model parameters do not appear in the evidence explicitly, this does not mean that the evidence solely depends on the data. That, in turn, can be recognized by a closer look on the background information I, since it contains implicitly all information about the theory applied to the parameter estimation problem, i.e. the parametrization and the functional form of the likelihood and the prior. The evidence will be important later on when different models that could describe the same data will be compared.

The prior $p\left(\theta|I\right)$ encodes the information about the parameters θ without knowledge of the experimental results, e.g. in the state of knowledge before the experiment was done. If, for instance, there is information about the parameters from an old measurement it can be encoded in the prior.

Before focusing on the assignment of priors, consider the example mentioned in the beginning of this section. Assume that the signal s measured in the experiment can deviate from the "true" amplitude a of the signal, and that the difference $s - a$ is known to be normal distributed with the standard deviation σ, the measurement error. In that case the likelihood can be written as

$$p\left(s|a, I\right) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(s-a)^2}{2\sigma^2}\right]. \tag{2.4.10}$$

If it is known, for instance from an old measurement, that the "true" amplitude has some approximate value, $a \approx a_0 \pm \sigma_0$, this can be expressed in the prior,

$$p\left(a|I\right) = \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left[-\frac{(a-a_0)^2}{2\sigma_0^2}\right]. \tag{2.4.11}$$

One has to keep in mind that equation (2.4.11) is just the observers limited knowledge about a. Even though this knowledge can be very vague (i.e. when σ_0 is large) according to the "theory" in that example the parameter a can have only one particular value in "reality", and therefore it is referred to as the "true" amplitude.

The equations (2.4.10) and (2.4.11) can be used together with Bayes' theorem (2.4.8) and the definition of the evidence (2.4.9) to compute the posterior. In this example, it is possible to write down the posterior in closed form,

$$p\left(a|s, I\right) = \frac{1}{\sqrt{2\pi}\overline{\sigma}} \exp\left[-\frac{(a-\overline{a})^2}{2\overline{\sigma}^2}\right], \tag{2.4.12}$$

with $\overline{a}$ being the weighted mean of a_0 and the data s,

$$\overline{a} = \frac{\frac{a_0}{\sigma_0^2} + \frac{s}{\sigma^2}}{\frac{1}{\sigma_0^2} + \frac{1}{\sigma^2}}, \tag{2.4.13}$$

and $\overline{\sigma}$ being the propagated errors,

$$\overline{\sigma} = \left(\frac{1}{\sigma_0^2} + \frac{1}{\sigma^2}\right)^{-1/2}. \tag{2.4.14}$$

The example is also shown graphically in figure 2.7. As a consequence of equation (2.4.14), the width $\overline{\sigma}$ of the posterior is smaller than the width σ_0 of the prior, and that in fact, the prior information contributes to the posterior maximum position as one single properly

weighted data point. The latter finding remains true even when several independent measurements are performed (Dose, 2002).

Generally speaking, the more localized the probability density of a variable, i.e. the narrower the density, the more precise is the information about the variable.

After having demonstrated how the prior information gets updated to the posterior with the information present in the measured data, the example will be extended in order to show how marginalization can be applied.

Assume that the measured signal s consists of two contributions, the "true" amplitude a and a "true" background b. The latter is measured with a reference measurement that results in the data r. The likelihood $p(s, r|a, b, I)$ will be again normal distributed in the measured signal s and the reference r, so that it has the maximum at $s = a + b$ and $r = b$. After defining the prior $p(a, b|I)$ and using Bayes' theorem, the posterior $p(a, b, |s, r, I)$ can be computed and is shown in figure 2.8.

While the posterior contains the full information available on the model parameters a and b, only the signal amplitude a might be of practical importance. Nevertheless, the accuracy of the reference measurement will influence the uncertainties of the amplitude a, because a and b contribute both to the measured signal s. In classical data analysis this problem is solved by propagation of errors. Probabilistic data analysis has the inherent property of propagating the limited information about the parameters if the corresponding marginal posterior densities are computed. In the example, this means that if b is not of importance it can be removed from the posterior by integration over all possible values, a process referred to as *marginalization* (see equation (2.4.5)),

$$p(a|s, r, I) = \int db \ p(a, b|s, r, I).$$ (2.4.15)

What was basically done here is the consideration of all possible values of the background b for a fixed amplitude a and weighting them by their probability. The result is the *marginal* posterior density $p(a|s, r, I)$. One can do the same for the other parameter a to get $p(b|s, r, I)$ as shown in figure 2.8.

Now, after having seen how less interesting model parameters can be marginalized, the assignment of priors will be discussed. This topic becomes increasingly important if the information content of the data is very low.
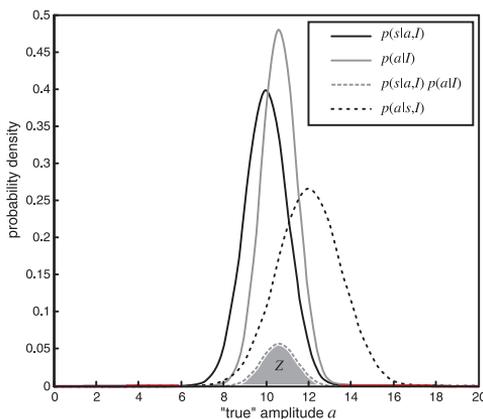
Figure 2.7: Example of parameter estimation. All quantities are shown as functions of the "true" amplitude a. The posterior $p(a|s, I)$ (solid gray line), here the probability density of the "true" amplitude a given the measured signal s is shown together with the likelihood $p(s|a, I)$ (solid black line), the prior $p(a|I)$ (dashed black line) and the product of prior and likelihood, $p(s|a, I) p(a|I)$ (dashed gray line). The posterior is calculated by dividing the product of likelihood and prior by the evidence Z (gray area). $s = 10$, $\sigma = 1$, $a_0 = 12$ and $\sigma_0 = 1.5$ was used in the drawing.
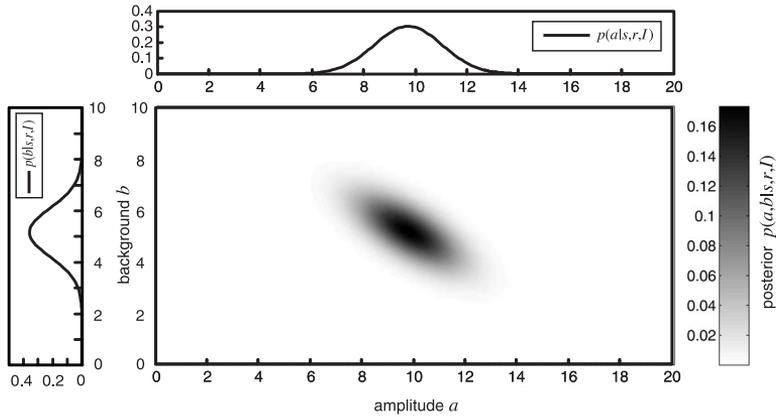
Figure 2.8: Example of marginalization. The shape of the posterior $p(a, b|s, r, I)$ shown in gray scale, indicates that the signal amplitude a and the background b are correlated. The marginal posterior densities $p(a|s, r, I)$ and $p(b|s, r, I)$ (top and left side, solid lines) are shown in the top and left panels.

2.4.3 Assignment of priors

Jaynes (1963, 1968) generalized the principle of maximum entropy to continuous variables. A prior $p(\theta|I)$ that reflects the complete lack of information about the parameter θ can be therefor found by maximizing the entropy S,

$$S = -\int d\theta \; p(\theta|I) \ln \left[\frac{p(\theta|I)}{m(\theta)} \right] \tag{2.4.16}$$

under the constraint of normalization

$$\int d\theta \; p(\theta|I) = 1. \tag{2.4.17}$$

This is solved usually with the method of Lagrange multipliers (Bronstein et al., 2000). In equation (2.4.16), $m(\theta)$ denotes a function (in fact a Lebesgue measure) that ensures that the entropy S is invariant under changes of parametrization, since it transforms in the same way like $p(\theta|I)$.

Beside the normalization, also other constraints that represent some piece of information can be accounted for. A constraint could be, for example, a known average value or variance of θ, but for this work only non-informative priors are of importance, which are defined on the basis of maximum entropy and normalization alone.

A maximum entropy prior has the property that given the constraints a minimum of information will be encoded. For example, when $m(\theta)$ is constant, no particular value of θ would be preferred. It can be shown in general that in the case of complete ignorance, i.e. lack of any other information besides the normalization property in equation (2.4.17), a maximum entropy prior $p(\theta|I)$ is proportional to the measure $m(\theta)$ (Sivia, 2006, chapter 5).

At first, it seems that the problem of prior assignment is just deferred to the assignment of the measure $m(\theta)$. The solution to this was proposed again by Jaynes (1963), and according to him, $m(\theta)$ should be chosen, so that a non-informative prior (and hence

$m(\theta)$) is invariant under a transformation group of the parameter space that reflects the fundamental invariances of the inference problem. In the following, this principle will be discussed with the help of two particular examples, a location and a scale invariance, occurring often in data analysis.

Location prior

Imagine for example that two observers A and B measure the location of an object in two different coordinate systems shifted by the distance α, so that B's location θ' can be calculated from A's θ by

$$\theta' = \theta + \alpha. \tag{2.4.18}$$

Equation (2.4.18), a translation, represents a transformation group since α can have any arbitrary value. Let A and B each assign a prior of the form

$$p\left(\theta|I\right) = f(\theta) \quad \text{and} \quad p\left(\theta'|I\right) = g(\theta'), \tag{2.4.19}$$

where f and g are normalized, positive-definite functions. When on the one hand A and B are both completely ignorant about the object's location the prior probabilities reflect the same state of knowledge and should transform according to

$$p\left(\theta|I\right)d\theta = p\left(\theta'|I\right)d\theta'. \tag{2.4.20}$$

On the other hand, since neither A's nor B's reference frame is to be preferred, both A and B should assign a prior of the same functional form in θ and θ', respectively, so that $f = g$, and finally

$$f(\theta)d\theta = f(\theta')d\theta'. \tag{2.4.21}$$

After inserting equation (2.4.18) and as $d\theta' = d(\theta + \alpha) = d\theta$, equation (2.4.21) reads

$$f(\theta) = f(\theta + \alpha), \tag{2.4.22}$$

which is satisfied by choosing

$$f(\theta) = constant. \tag{2.4.23}$$

Hence, the constant prior $p\left(\theta|I\right)$ is invariant under arbitrary translations of the frame of reference. Since $p\left(\theta|I\right)$ does not contain any structure it is often referred to as a *flat prior*. In contrast, in the next example a prior will be discussed, which is not flat but yet uninformative.

Jeffreys prior

Assume now that θ is a quantity that describes a scale, i.e. a positive observable, for example the length of an object. Imagine two observers A and B who want to measure the length of an object that is denoted by θ and θ' in the respective coordinate systems. Since A and B can use completely different units, e.g. the metric and the imperial unit systems, the observables θ and θ' are related by the following scale transformation,

$$\theta' = \beta\theta, \tag{2.4.24}$$

where β is a factor that converts the units of A into the units of B. Now, the argument to derive a prior for θ follows the same steps as before. Again, both observers assign a prior of the general form stated in equation (2.4.19), and when both are in the same state knowledge, their prior probabilities interconvert as shown in equation (2.4.20). Assume now that A and B are in a state of complete ignorance, then they should assign a prior

of the same functional form, which results in equation (2.4.21). Now, after substituting equation (2.4.24) into equation (2.4.21), and since $d\theta' = d(\beta\theta) = \beta d\theta$, the function $f(\theta)$ has to satisfy

$$f(\theta)d\theta = f(\beta\theta)\beta d\theta, \tag{2.4.25}$$

which is solved by choosing

$$f(\theta) = \frac{constant}{\theta}. \tag{2.4.26}$$

The prior $p(\theta|I) = constant/\theta$ was proposed by and named after Sir Harold Jeffreys (1939). It is invariant under arbitrary scale transformations and assigns the same probability to every order of magnitude. The latter is shown easily when one considers a flat prior in the logarithm of θ, i.e. $p(\ln(\theta)|I) = constant$, which is equivalent to Jeffreys prior, since

$$p(\ln(\theta)|I) \; d(\ln(\theta)) \; = \; constant \cdot \frac{1}{\theta}d\theta \; = \; p(\theta|I)\,d\theta. \tag{2.4.27}$$

Jeffreys prior should be used when there is no information about the scale of a parameter. Although it is not flat and assigns more probability to small values of θ compared to large values, it encodes only the information that θ is a scale variable.

An application of Jeffreys prior to an example from physical chemistry is shown in the following. Imagine two spectroscopists who want to infer the spectral position of a peak. One of them prefers to work with wavelengths and tries to estimate λ, whereas the other favors wavenumbers and would like to estimate $\tilde{\nu}$. Both quantities are related by $\lambda = 1/\tilde{\nu}$.

It is obvious that λ and $\tilde{\nu}$ are both scale variables since they are positive and describe the scale of the wavelength of light and the scale of the energy of a transition. A Jeffreys prior in the wavelength λ transforms into an equivalent Jeffreys prior in the wavenumber as follows,

$$p(\lambda|I)\,d\lambda \; = \; \frac{constant}{\lambda}d\lambda \; = \; constant \cdot \tilde{\nu}\left|\frac{d\lambda}{d\tilde{\nu}}\right|d\tilde{\nu} \; = \tag{2.4.28}$$

$$= constant \cdot \tilde{\nu}\left|-\frac{1}{\tilde{\nu}^2}\right|d\tilde{\nu} \; = \; \frac{constant}{\tilde{\nu}}d\tilde{\nu} \; = \; p(\tilde{\nu}|I)\,d\tilde{\nu}. \tag{2.4.29}$$

The choice of different priors, in particular a flat prior, would lead to inconsistencies in the prior knowledge of both spectroscopists.

In conclusion, the principle of maximum entropy together with the transformation group invariance of the measure $m(\theta)$ lead to a unique and objective definition of the prior in both the presence and absence of additional information (Jaynes, 1968).

Now, after having laid the foundations for the solution of parameter estimation problems, the next section deals with the visualization of probability densities, as well as with the calculation of an estimate and an error margin from the posterior information.

2.4.4 Characterization of continuous probability distributions

Although in Bayesian data analysis the information has the form of a probability density, which exactly reflects the fuzziness of the knowledge, often it is difficult to display the density adequately. This is especially true when the density is defined in more than one dimension, and therefore it is convenient to subsume the information in a few numbers that state the position of the probability maximum and the width of the density. Alternatively, one can use other easy-to-display means of visualization that are not as fuzzy as a density plot or scatter plot.
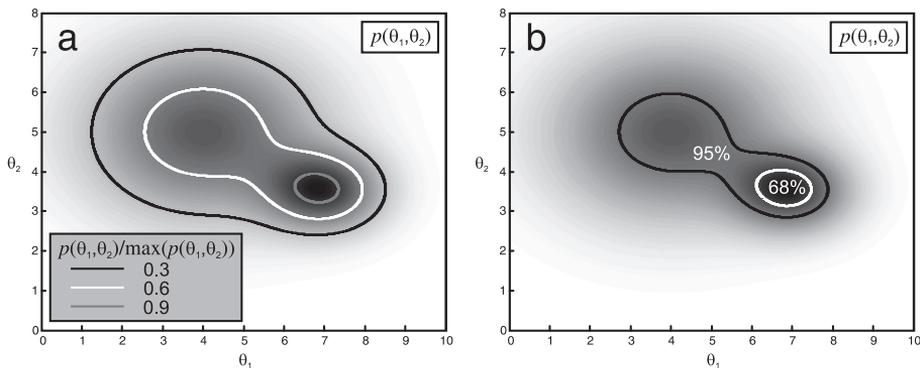
Figure 2.9: A two-dimensional probability density $p(\theta_1, \theta_2)$ (grayscale) shown with iso-lines at 30%, 60% and 90% of the maximum density (a) and credible intervals containing 68% and 95% probability (b).

Visualization by iso-surfaces

In two or three parameter space dimensions one can usually depict a smooth and continuous probability density $p(\theta)$ by a set of iso-surfaces S_i that contour the density at a few levels p_i, e.g. at the 90%, 60% and 30% fractions of the probability density maximum,

$$S_i = \{\theta | p(\theta) = p_i\}. \tag{2.4.30}$$

The iso-surface will be a line when the probability density is two-dimensional (figure 2.9a), and it will be a surface when it is three-dimensional.

Instead of contouring $p(\theta)$ at some fraction of the maximum one can display iso-surfaces that are related to the amount of probability P_j contained inside the volume enclosed by the surface, e.g. 68% or 95% (figure 2.9b),

$$S_{P_j} = \{\theta | p(\theta) = p_j, \int_{p(\theta) > p_j} d\theta \; p(\theta) = P_j\}. \tag{2.4.31}$$

Drawn for a posterior probability density $p(\theta | \text{data}, I)$ the volumes enclosed by these iso-surfaces are called *credible intervals* and are similar to *confidence intervals* in classical statistics. Credible intervals are the smallest volumes in the parameter space that contour the given probability. Later on in this work three-dimensional credible intervals will be called credible volumes.

Note that iso-surfaces and credible intervals are only means to display a probability density, and that the complete information is contained in the density itself.

Specification of uncertainties

Sometimes it is necessary to compress the information even more by stating only a few numbers. Usually, the position of the maximum of the probability density, $\theta_{\max}$, is specified together with an uncertainty. This is equivalent to an expansion of $\ln p(\theta)$ up to

second order,

$$\ln p\left(\theta\right) \approx const + \frac{1}{2}\left(\theta - \theta_{\max}\right)^{\mathrm{T}}\left[\frac{\partial^2 \ln p\left(\theta\right)}{\partial\theta^2}\right]_{\theta=\theta_{\max}}\left(\theta - \theta_{\max}\right), \qquad (2.4.32)$$

where the linear term has vanished since $\ln p\left(\theta\right)$ is expanded at the maximum $\theta_{\max}$, and $(.)^{\mathrm{T}}$ denotes the transposition operation. Equation (2.4.32) is equivalent to a Gaussian approximation of $p\left(\theta\right)$. The uncertainty is computed from the second derivative matrix of $\ln p\left(\theta\right)$, also called the Hessian matrix $\mathcal{H} = \partial^2 \ln p\left(\theta\right)/\partial\theta^2$ evaluated at $\theta_{\max}$. When it is not important to report correlations of the different parameter space dimensions, the uncertainty of the ith component θ_i can be stated in the form $\theta_i = \theta_{i,\max} \pm \Delta\theta_i$. The uncertainties $\Delta\theta_i$ are calculated from the diagonal elements of the inverse Hessian matrix,

$$\Delta\theta_i = \sqrt{-\left(\mathcal{H}^{-1}\right)_{ii}}. \qquad (2.4.33)$$

They are equivalent to the standard deviations of the marginalized Gaussian approximation of $p\left(\theta\right)$ at its maximum. If correlations are important it is necessary to write down the complete matrix $-\mathcal{H}^{-1}$ instead of stating only the marginal errors $\Delta\theta_i$.

Since the approximation in equation (2.4.32) is valid only when there is one symmetric peak in the posterior, the above shorthand notation cannot be applied if the probability distribution is highly asymmetric or even has more than one maximum. When a quantity and an uncertainty are given one should always keep in mind that this is a short-hand notation for a Gaussian approximation of the probability density maximum.

An alternative way to describe a probability density is stating its expectation value $\langle\theta\rangle$,

$$\langle\theta\rangle = \int d\theta\ \theta p\left(\theta\right), \qquad (2.4.34)$$

together with the uncertainty given by its covariance matrix C

$$C = \int d\theta\left(\theta - \langle\theta\rangle\right)^{\mathrm{T}}\left(\theta - \langle\theta\rangle\right)p\left(\theta\right). \qquad (2.4.35)$$

The uncertainty (standard deviation) in the parameter space dimension i is obtained from the covariance by computing the square root of the diagonal elements of C.

Note that in the case that $p\left(\theta\right)$ is exactly a multivariate Gaussian the covariance matrix will be equal to the negative of the inverse Hessian matrix, $C = -H^{-1}$.

After this useful section the focus will be on the comparison of different models that describe the same data.

2.4.5 Bayesian model selection

When there are different competing models or theories for the same data, Bayesian reasoning can be used to calculate the probabilities that a certain model is valid. For example, the models might differ in the functional form of the data, e.g. when a spectrum is measured different line shapes might be possible. Because of broadening of the spectral lines it also might not be obvious how many lines contribute to the measured spectrum. In the following section it will be shown using an example, how probabilistic data analysis can be used to compare different models.

Imagine for example that the observable y is measured as a function of some experimentally controllable quantity x, so that the data $\{y_i\} = y_1, y_2, \ldots, y_N$ is available for the values $\{x_i\} = x_1, x_2, \ldots, x_N$. Assume that there are two models, A and B, to which

different functions $f_A(x)$ and $f_B(x)$ are underlying to the data apart from the unknown deviations (i.e. measurement errors) $\{\epsilon_i\} = \epsilon_1, \epsilon_2, \ldots, \epsilon_N$,

$$y_i = f_{A/B}(x_i) + \epsilon_i. \tag{2.4.36}$$

Say that in the model A the data does not show any dependence on x,

$$y = f_A(x) = c, \tag{2.4.37}$$

where c, the "true" value of the observable y, is a model parameter. In the competing model B the data is a linear function of x,

$$y = f_B(x) = a + bx, \tag{2.4.38}$$

where a and b are the offset and slope of the line that describes y as a function of x.

Assume now that though the exact values of the experimental errors $\{\epsilon_i\}$ are not known the distribution of $\{\epsilon_i\}$ is known, so that the likelihoods $p\left(\{y_i\}|c, A, I\right)$ and $p\left(\{y_i\}|a, b, B, I\right)$ can be stated for the models A and B, respectively. In addition, the model parameter priors $p\left(c|A, I\right)$ and $p\left(a, b|B, I\right)$ of both models A and B must have been assigned. In order to infer whether model A or B is more plausible given the measured data $\{y_i\}$, the probabilities $p\left(A|\{y_i\}, I\right)$ and $p\left(B|\{y_i\}, I\right)$ have to be computed. After using Bayes' theorem the probabilities of each model are

$$p\left(A|\{y_i\}, I\right) = \frac{p\left(\{y_i\}|A, I\right) p\left(A|I\right)}{p\left(\{y_i\}|I\right)} \quad \text{and}$$

$$p\left(B|\{y_i\}, I\right) = \frac{p\left(\{y_i\}|B, I\right) p\left(B|I\right)}{p\left(\{y_i\}|I\right)}, \tag{2.4.39}$$

with the total evidence $p\left(\{y_i\}|I\right)$ being a constant. The probabilities $p\left(A|I\right)$ and $p\left(B|I\right)$ are prior probabilities of the models A and B, which encode whether one model is preferred over the other. They should not be confused with the priors of the model parameters. When there is no reason to prefer one of the models then both probabilities $p\left(A|I\right)$ and $p\left(B|I\right)$ should be equal.

The evidences of each model, $p\left(\{y_i\}|A, I\right)$ and $p\left(\{y_i\}|B, I\right)$, must be computed as in equation (2.4.9) from likelihood times prior integrated over the parameters,

$$p\left(\{y_i\}|A, I\right) = \int dc \, p\left(\{y_i\}|c, A, I\right) p\left(c|A, I\right)$$

$$p\left(\{y_i\}|B, I\right) = \int da \int db \, p\left(\{y_i\}|a, b, B, I\right) p\left(a, b|B, I\right). \tag{2.4.40}$$

According to equations (2.4.39) and (2.4.40) the probability of a model rises the larger the evidence of that particular model.

In order to get rid of the constant probability $p\left(\{y_i\}|I\right)$, the so called *odds ratio* can be computed,

$$\frac{p\left(A|\{y_i\}, I\right)}{p\left(B|\{y_i\}, I\right)} = \frac{p\left(\{y_i\}|A, I\right)}{p\left(\{y_i\}|B, I\right)} \cdot \frac{p\left(A|I\right)}{p\left(B|I\right)}. \tag{2.4.41}$$

An odds ratio > 1 would support model A over model B and vice versa. If, as in this example, the models have a different number of parameters and thus differ in their complexity, it can be shown with the help of equation (2.4.41) that usually the model with less parameters (here model A) should be preferred when both models explain the

data equally likely.[2]

Of course, also more than two models can be compared at once, but that changes merely the normalization probability in the right hand side of equation (2.4.39) and the finding that the probability of a model is proportional to $p\left(\text{data}|\text{model}, I\right) \cdot p\left(\text{model}|I\right)$ remains unaffected.

In the next section some challenges will be discussed that are often encountered in "real life" applications of probabilistic data analysis. They usually appear when one is confronted with marginalization of a posterior distribution or the calculation of the evidence of a model.

2.4.6 Application to real problems

When probabilistic data analysis is applied to real problems, oftentimes the evidence as well as marginal posterior distributions cannot be computed analytically any more.

Also the evaluation of the posterior on a regularly spaced grid, which could be used as an approximation in order to compute the evidence by "brute force" methods becomes inapplicable when the parameter space has too many dimensions: in that case, the number of supporting points and posterior evaluations will become too large and the problem is infeasible to calculate in a reasonable time.

Fortunately, it is often not necessary to know the posterior value on every point of a grid, as many of the evaluations would yield a probability density close to zero, since most of the posterior probability is often concentrated in a small region of the parameter space. This can be used to tackle the problem of representation and the calculation of the evidence by concentrating on a computationally feasible number of supporting points, so called *samples*, which are placed in parameter space regions of substantially high posterior probability density.

In the next section, some techniques used to obtain samples from probability densities will be discussed.

---

[2]Model selection based on the model complexity is often called of *Occam's razor*. This common practice is reasonable when initially none of the models is preferred, i.e. prior probabilities of the models are approximately equal, $p\left(A|I\right) \approx p\left(B|I\right)$, and hence drop out of equation (2.4.41). For simplicity it will be assumed that in both models there is a parametrization, so that the prior of the model parameters is constant and the posterior is proportional to the likelihood. When both models describe the data equally well then both likelihoods will have similar optimal values that will be the main contribution to the evidence of the model. In the more complex model the prior probability of parameter values close to the likelihood maximum will be usually smaller compared to the more simple model because the prior (more precisely its support) will occupy a larger parameter space volume. In other words it will be less probable to have parameter values that would explain the data in the complex model compared to the simple one because it is easier to "detune" the complex model.
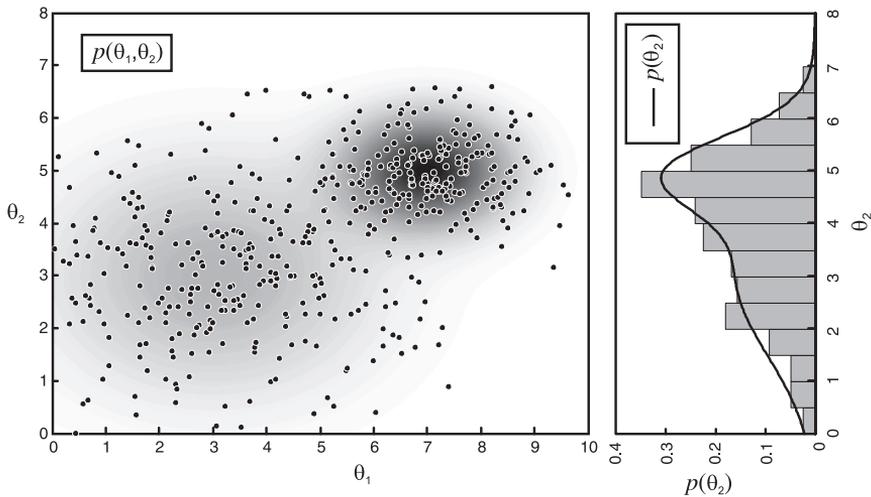
Figure 2.10: Samples representation of a probability density. A probability density, here an example in two dimensions, $p(\{\theta_i\}) = p(\theta_1, \theta_2)$ (gray scale, left), can be approximately represented by a set of samples $\{\theta_i\}_j$ (dots) that are drawn with the probability $p(\{\theta_i\}_j)$. Approximate marginal densities can be computed by creating normalized (area = 1) histograms in the parameter space dimensions of interest and neglecting the other dimensions (right).

2.5 Sampling from probability densities

When marginal posterior densities have to be calculated in order to project a probability density onto a few relevant parameter space dimensions, or when the evidence of a model needs to be computed, it is often infeasible to do this by analytic techniques. The "brute force" numerical approximation of these integrals as done by evaluating the product of prior and likelihood on a grid and subsequently using quadrature is applicable only when the number of parameter space dimensions to integrate over is small as it becomes inefficient in high dimensions. The problem of the integration of probability densities in high dimensional parameter spaces has been already addressed in the field of statistical physics. A popular approach is the use of Monte Carlo techniques that are based on the representation of a probability density by a set of samples.

The probability to draw a sample at the parameter space position $\{\theta_i\} = (\theta_1, \theta_2, \ldots, \theta_N)$ from the probability density $p(\{\theta_i\})$ is given by $p(\{\theta_i\})d\{\theta_i\}$, where $d\{\theta_i\}$ is the infinitesimal parameter space region around the position $\{\theta_i\}$. When many of such samples are drawn, the density of the samples in the parameter space divided by the number of samples will approximate the probability density $p(\{\theta_i\})$ (see figure 2.10).

By using these samples one can calculate for example approximate average values of an arbitrary function $f(\{\theta_i\})$,

$$\langle f \rangle = \int d\{\theta_i\} \, f(\{\theta_i\})p(\{\theta_i\})$$

$$\approx \frac{1}{M} \sum_{j=1}^{M} f(\{\theta_i\}_j). \tag{2.5.1}$$

Here, $\{\theta_i\}_j$ denotes the jth sample of M samples in total, which were drawn with the probability $p(\{\theta_i\}_j)dp\{\theta_i\}$. Also approximate marginal densities can be constructed from such a set of samples by creating histograms normalized to area $= 1$ of a subset of parameters.

In the next sections two methods will be discussed that were used to solve the inference problems appearing in this work. First, the well known technique of Markov chain Monte Carlo and the Metropolis algorithm will be explained. Thereafter, the nested sampling algorithm will be introduced that will be used later on to compute model evidences and to produce also sets of samples that approximate the posterior.

2.5.1 Markov chain Monte Carlo

A popular technique to draw samples from high dimensional probability distributions is Markov chain Monte Carlo (MCMC) that has its origin in statistical physics (Metropolis and Ulam, 1949). A good introduction is found for example in Bishop (2006, chapter 11) and Neal (1993). The MCMC method is often based on a random walk in the parameter space of the probability density $p(\{\theta_i\})$ to be sampled. In analogy to classical physics, one can imagine a particle that is subject to thermal fluctuations but trapped in a potential landscape $U(\{\theta_i\}) = -\ln[p(\{\theta_i\})]$ that is constructed from the probability density $p(\{\theta_i\})$. When the particle is in equilibrium, it will explore the whole potential landscape driven by the fluctuations. In the limit of an infinitely long random walk the distribution of positions $\{\theta_i\}_t$ occupied will be equal to the Boltzmann distribution that is of course exactly $p(\{\theta_i\})$.

Definition of a Markov chain

The random walk is realized in a way that the position $\{\theta_i\}_t$ occupied at the discrete time t depends only on the preceding position $\{\theta_i\}_{t-1}$ but not positions $\{\theta_i\}_{t-2}, \ldots, \{\theta_i\}_0$ further in the past, which makes the positions $\{\theta_i\}_t$ with $t = 0, \ldots, N$ a first order *Markov chain*. The probability $p(\{\theta_i\}_t|\{\theta_i\}_{t-1})$ to obtain a position $\{\theta_i\}_t$ based on the preceding position $\{\theta_i\}_t$ is called the *transition probability* and will be denoted by $T_t(\{\theta_i\}_t, \{\theta_i\}_{t-1})$ in the following. When the transition probabilities do not depend directly on t, i.e. when $T_t = T_{t'} \equiv T$ for any two times t and t' then the Markov chain is called *homogeneous*.

In order for a homogeneous Markov chain to converge to the equilibrium distribution $p(\{\theta_i\})$ it is sufficient (but not necessary) that the latter is invariant or stationary with respect to the Markov chain, i.e.

$$p(\{\theta_i\}) = \int d\{\theta_i\}' T(\{\theta_i\}, \{\theta_i\}') p(\{\theta_i\}'). \qquad (2.5.2)$$

This property can be achieved by choosing the transition probabilities to satisfy the condition of *detailed balance*,

$$T(\{\theta_i\}', \{\theta_i\}) \, p(\{\theta_i\}) = T(\{\theta_i\}, \{\theta_i\}') \, p(\{\theta_i\}'), \qquad (2.5.3)$$

which requires that the flux of probability from $\{\theta_i\}$ to $\{\theta_i\}'$ cancels the reverse flux from $\{\theta_i\}'$ to $\{\theta_i\}$, which is similar to what happens in a chemical reaction in equilibrium. A Markov chain is called *reversible* when it satisfies detailed balance.

The Metropolis algorithm

Metropolis et al. (1953) proposed a way to construct a reversible Markov chain with the desired stationary probability distribution $p(\{\theta_i\})$ as follows: First, for a given po-

sition $\{\theta_i\}_t$ at time t a possible next position $\{\theta_i\}_{t+1}$ is proposed with the probability $p_{\text{prop}}(\{\theta_i\}_{t+1}|\{\theta_i\}_t)$. The only requirement on the *proposal probability* is that it is symmetric with respect to the two positions,

$$p_{\text{prop}}(\{\theta_i\}_{t+1}|\{\theta_i\}_t) = p_{\text{prop}}(\{\theta_i\}_t|\{\theta_i\}_{t+1}). \tag{2.5.4}$$

The proposed position can be then accepted (A) or rejected ($\overline{A}$) with the probability

$$p(A|\{\theta_i\}_{t+1}, \{\theta_i\}_t) = \min\left(1, \frac{p(\{\theta_i\}_{t+1})}{p(\{\theta_i\}_t)}\right). \tag{2.5.5}$$

When the proposed position is accepted it becomes the "current" position in the random walk. In the case of a rejection the random walk remains in the old position and the new position $\{\theta_i\}_{t+1}$ is set to the old position $\{\theta_i\}_t$.

It is easy to show that the Markov chain is reversible when equations (2.5.4) and (2.5.5) hold. Because only the ratio of the probability densities $p(\{\theta_i\}_{t+1})/p(\{\theta_i\}_t)$ appears in equation (2.5.5) one might replace $p(\{\theta_i\})$ with a density $\tilde{p}(\{\theta_i\})$ that is not normalized but proportional to $p(\{\theta_i\})$ without changing the sampling. This property is especially useful for Bayesian parameter estimation since the posterior is known often only up to a normalizing constant (the evidence), provided the prior and likelihood can be evaluated.

Convergence consideration

A second important property needed for a robust MCMC is the convergence of the Markov chain to the probability distribution $p(\{\theta_i\})$ irrespectively of the initial position $\{\theta_i\}_0$. This property is called *ergodicity*, and it can be shown (Neal, 1993) that a homogeneous Markov chain will be ergodic already under weak restrictions on the transition probabilities and the invariant distribution (Bishop, 2006, chapter 12).

Of practical importance is the time it takes the chain to converge to $p(\{\theta_i\})$ since for shorter times the chain will not be in equilibrium and the produced samples should not be used. Even long after convergence the samples will still be highly correlated, and when independent samples are needed, the samples in between have to be discarded.

Monte Carlo Moves

Usually the proposal probability is a normal distribution centered at $\{\theta_i\}_t$, but it is also possible to use other distributions that satisfy the symmetry condition in equation (2.5.4). Such proposal probability distributions are called *moves*. One can even combine several different moves, by for example deciding at random with a constant probability p_m, which move m will be used to propose the next position, so that

$$p_{\text{prop}}(\{\theta_i\}'|\{\theta_i\}) = \sum_m p_m \, p_{\text{prop},m}(\{\theta_i\}'|\{\theta_i\}), \tag{2.5.6}$$

where $p_{\text{prop},m}(\{\theta_i\}'|\{\theta_i\})$ is the proposal probability of the mth move. That strategy is useful to reduce the time of convergence and the time in between independent samples. An example of different moves would be the proposal probabilities $p_{\text{prop},m}(\{\theta_i\}'|\{\theta_i\})$ that change only the mth component of the parameter vector $\{\theta_i\}$. The application of these moves would be then equivalent to deciding, in which dimension to move by choosing m proportional to p_m, and then proposing a new position $\{\theta_i\}' = (\theta_1, \ldots, \theta'_m, \ldots, \theta_N)$. Another example would be the mixing of moves with different step sizes.

After this short introduction to Markov chain Monte Carlo for sampling a probability

density $p(\{\theta_i\})$, in the next subsection a Monte Carlo method will be addressed that can be used to integrate a probability density.

2.5.2 Nested sampling

Nested sampling was developed by Skilling (2006) and can be used to compute the evidence in probabilistic data analysis. Up to now, it was mainly applied in the fields of astrophysics and cosmology (Shaw et al., 2007; Feroz et al., 2010). It can be useful in problems in which phase changes occur, i.e. when there are one or more small high-probability density regions in the parameter space that contain a substantial amount of the total probability, while other, much larger regions have low probability densities. Such problems are difficult to solve with other Monte Carlo methods like simulated annealing, since the high-probability density regions must be found, which might be difficult when the Markov chain is located initially in a large low probability density regions (Skilling, 2006).

In contrast, nested sampling is based on the compression of a set of samples referred to as *objects*. The algorithm starts with an initial set of objects drawn from the prior and proceeds by repeatedly replacing the object with the smallest likelihood value by a new object drawn from the prior that has a higher likelihood. In this way, the objects are compressed into parameter space regions that have high likelihood values.

Unlike the Metropolis algorithm, Nested sampling is not a technique to obtain samples from a probability density directly. Instead, it is a protocol that uses samples from the constrained prior drawn by other sampling methods to compute the evidence. However, one can reuse the samples produced during nested sampling to get a set of posterior samples as shown at the end of this section.

In the following, the abstract concept of the *sorted likelihood* function will be introduced first. This concept is important to explain how the evidence can be computed with nested sampling, which will be shown thereafter.

Sorted likelihood

To understand the principle of nested sampling consider the prior $p(\{\theta_i\}|I) \equiv \pi(\{\theta_i\})$ and the likelihood $p(\mathrm{data}|\{\theta_i\}, I) \equiv L(\{\theta_i\})$. Imagine now that the parameter space can be partitioned into small volume elements δV_k, in which the likelihood has the approximate value L_k (figure 2.11a).

The amount of prior probability inside these volumes is equal to

$$\delta\pi_k = \int\limits_{\{\theta_i\}\in\delta V_k} d\{\theta_i\}\; \pi(\{\theta_i\}), \qquad (2.5.7)$$

and by adjusting the size of the individual volume elements δV_k the corresponding prior probability elements $\delta\pi_k$ can be made equally sized, $\delta\pi_k \equiv \delta\pi$.
In this notation, the evidence Z (equation (2.4.9)) is approximately

$$Z \approx \sum_k L_k\delta\pi_k = \sum_k L_k\delta\pi. \qquad (2.5.8)$$

A natural order of the prior probability elements is given by the corresponding likelihood values L_k, and one could require an ordering by descending likelihood, $L_k \geq L_{k+1}$. Similarly, the whole prior volume can be sorted like this when the limit to infinitely small prior probability elements, $\delta\pi \to d\pi$, is taken. In this case, the *sorted prior mass* ξ is formally defined as the prior probability accumulated in the parameter space volume where the
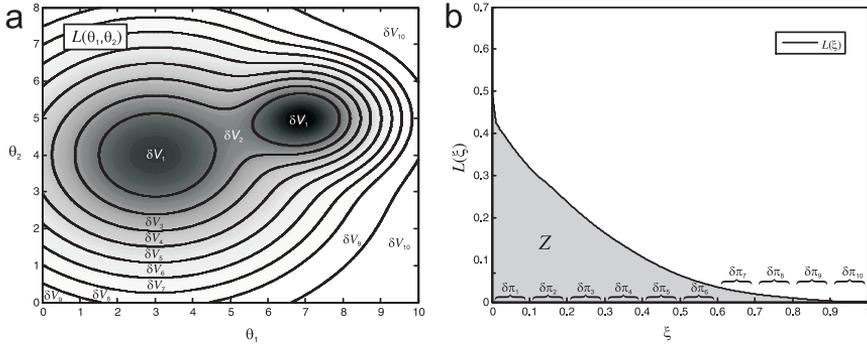
Figure 2.11: Sorting the likelihood. (a) Segmentation of the parameter space. The likelihood $L(\{\theta_i\})$ is shown as a function of the parameters (gray scale). The parameter space can be subdivided into small regions δV_k with approximately equal likelihood (stripes in between black lines) containing an equal amount of prior probability $\delta \pi_k = \delta \pi$. In the two-dimensional example shown here there are 10 regions δV_k. For simplicity the prior was taken to be constant. (b) The sorted likelihood function, $L(\xi)$, is plotted as a function of the sorted prior mass ξ. The prior probabilities π_k contained in the parameter space regions δV_k are shown on the $\xi - $ axis. Each $\delta \pi_k$ corresponds to the volume δV_k in (a). The area (gray) under the sorted likelihood function is equal to the evidence Z.

likelihood exceeds a given value λ,

$$\xi(\lambda) = \int\limits_{L(\{\theta_i\}) > \lambda} d\{\theta_i\} \pi(\{\theta_i\}). \qquad (2.5.9)$$

Note that this is only a formal definition of the sorted prior mass, and that $\xi(\lambda)$ will be computed based on an estimate later on.

The element of sorted prior mass, $d\xi$, is then equal to the element of prior probability defined before, $d\pi$. The likelihood can be regarded as a function of the sorted prior mass ξ by defining the *sorted likelihood* function as follows,

$$L(\xi) \equiv \lambda. \qquad (2.5.10)$$

By comparison with equation (2.5.8) in the limit of infinitesimally small prior probability elements it is obvious that the evidence Z is simply the area below the sorted likelihood function $L(\xi)$,

$$Z = \int d\xi L(\xi), \qquad (2.5.11)$$

as shown in figure 2.11b.

The nested sampling algorithm

Nested sampling is a Monte Carlo technique that uses the sorted likelihood function to compute the evidence. It is based upon a collection of M independent objects that are uniformly distributed within the interval $[0, \xi^*)$. The uniform sampling in the constrained interval of the sorted prior mass is equivalent to the sampling of the prior $\pi(\{\theta_i\})$ subject to the likelihood constraint $L(\{\theta_i\}) > L(\xi^*) \equiv L^*$.

Initially, at iteration $j = 0$, ξ^* is set to 1, and, consequently, there is no likelihood
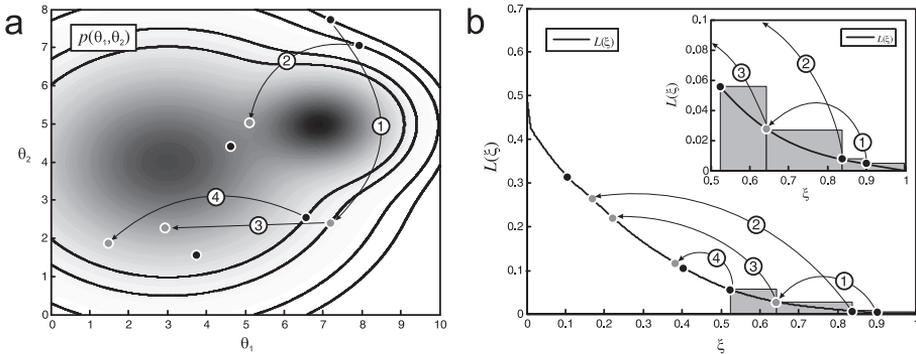
Figure 2.12: Principle of nested sampling. The nested sampling scheme displayed in the parameter space (a) and the sorted likelihood function $L(\xi)$ (b). The first four iterations are shown. Objects from the collection are shown as black and gray dots. The arrows indicate the replacement of the object with the lowest likelihood by a new object with a higher likelihood drawn from the constrained prior (gray dots). The respective constraints are shown as black lines in the left panel. The first four weights $w_j = L_j \cdot (\xi_{j-1} - \xi_j)$ are shown in the right panel as gray rectangles (magnified view in the inset).

constraint for which reason all objects are sampled from the unconstrained prior. At each following iteration j the object with the lowest likelihood value is identified and the corresponding likelihood and parameter values are saved in the list of samples as $\{\theta_i\}_j$ and $L_j = L(\{\theta_i\}_j)$. The likelihood constraint is then set to $L^* = L_j$, and the object with the lowest likelihood is replaced by another one sampled from the prior subject to the updated constraint. Consequently, the M objects become more and more constrained to regions of high likelihood as the algorithm advances (figure 2.12a).

When observed on the axis of the sorted prior mass, the M objects become compressed towards $\xi = 0$ and leave behind a sequence of discarded samples at the positions ξ_j with likelihoods L_j. If the algorithm proceeded far enough to the left of the ξ-axis one could calculate the evidence approximately from these samples by adding the rectangular areas w_j in between successive samples (figure 2.12b),

$$Z \approx \sum_j w_j = \sum_j L_j \cdot (\xi_{j-1} - \xi_j), \tag{2.5.12}$$

In that way, each sample ξ_j is attributed a weight $w_j = L_j \cdot (\xi_{j-1} - \xi_j)$.

Unfortunately, the exact position of the samples on the ξ − axis is not known. Instead, one can estimate the average position and use it in equation (2.5.12). To do so, consider first the *shrinking ratio* $t_j = \xi_j / \xi_{j-1}$ of two successive sorted prior masses ξ_j and ξ_{j-1}. As stated by Skilling (2006), it is distributed as

$$p(t_j) = M t_j^{M-1}. \tag{2.5.13}$$

The above equation can be shown easily: Assume that the M objects are placed one at a time and with uniform probability in the interval $[0, \xi_{j-1})$. First, the outermost object (i.e. the object with largest ξ) is chosen out of M objects in total, which can be done in M realizations, which explains the first term in the right hand side of equation (2.5.13). However, the outermost object must be placed at the position ξ_j to be consistent with the definition of the shrinking ratio. The remaining objects have to be placed then in the

interval $[0, \xi_j)$, which is done with the probability of $\xi_j / \xi_{j-1} = t_j$ for each of the $M - 1$ objects, which explains the remaining term t_j^{M-1}.

The mean and standard deviation of the natural logarithm of the shrinking ratios, $\ln(t_j)$, are consequently (Skilling, 2006)

$$\langle \ln t_j \rangle = -1/M, \tag{2.5.14}$$
$$\Delta \ln t_j = 1/M.$$

The position of the samples on the ξ axis can be expressed as the product of successive shrinking ratios,

$$\xi_j = \prod_{k=1}^{j} t_k, \tag{2.5.15}$$

in which each shrinking ratio is independently distributed according to equation (2.5.13). Hence, the average and standard deviation of $\ln \xi_j$ is

$$\langle \ln \xi_j \rangle = -j/M, \tag{2.5.16}$$
$$\Delta \ln \xi_j = \sqrt{j}/M.$$

When these estimates are used instead of the unknown exact positions in equation (2.5.12) the further approximated evidence is given by the sample weights (rectangular areas in figure 2.12b),

$$Z \approx \sum_j L_j \cdot [\exp\left(\langle \ln \xi_{j-1}\rangle\right) - \exp\left(\langle \ln \xi_j \rangle\right)] = \sum_j L_j \cdot \left[\exp\left(-\frac{j-1}{M}\right) - \exp\left(-\frac{j}{M}\right)\right]. \tag{2.5.17}$$

An example of this approximation is shown in figure 2.13a. Usually, the weights will be very small at the beginning and at the end of the nested sampling run and will have a maximum in between (figure 2.13b). This happens because at first the vanishingly small likelihood values will dominate the widths of the rectangles. As the likelihood rises rapidly the weights will increase though the widths $\xi_j - \xi_{j-1}$ will shrink. At the end, when all objects are crowded close to the maximum of the sorted likelihood function the extremely small widths of the rectangles will dominate the high likelihood values that cannot rise as much as in the beginning, and the weights w_j will be small again.

Because the actual values of the sample positions ξ_j deviate from their average, the evidence is accompanied by an uncertainty inherent to the Monte Carlo approach. Since the logarithm of the evidence rather than the evidence itself is distributed symmetrically it is more convenient to state the mean and standard deviation of $\ln Z$,

$$\langle \ln Z \rangle \approx \ln\left(\sum_j w_j\right), \tag{2.5.18}$$

$$\Delta \ln Z \approx \sqrt{\frac{H}{M}}.$$

Here, the logarithmic uncertainty $\Delta \ln Z$ is approximately given by the number of objects M and the information or negative entropy, H, of the posterior relative to the prior,

$$H \approx \sum_j \frac{w_j}{Z} \ln\left(\frac{L_j}{Z}\right). \tag{2.5.19}$$
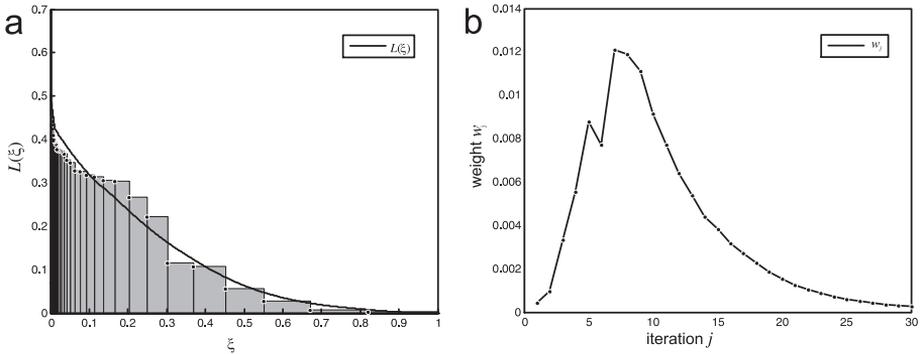
Figure 2.13: Estimated sample weights. (a) The estimated sample weights w_j (gray rectangles) add up to the approximate evidence Z. The deviations of the sample positions (dots) from the exact values on the sorted prior mass axis ξ are visible as horizontal shifts off the sorted likelihood function $L(\xi)$. (b) Estimated weights w_j as function of the iteration j.

The nested sampling algorithm should be terminated when a substantial fraction of the evidence Z was acquired, i.e. when the collection of objects passed most of the area under the sorted likelihood function. Since the main contribution to Z occurs usually where $\xi \approx \exp(-H)$ (Skilling, 2006), the algorithm may be terminated when the iteration number significantly exceeds MH, as the sample produced at iteration $j = MH$ is approximately at the position $\ln \xi_{j=MH} \approx -j/M|_{j=MH} = -H$.

Parallelization

When nested sampling is run multiple times, for instance on different computers, one can merge the resulting samples in a straight forward way. Assume that the samples $\{\theta_i\}_j^{(r)}$ with corresponding likelihoods $L_j^{(r)}$ and (hypothetical) positions on the sorted prior mass axis $\xi_j^{(r)}$ were obtained in separate nested sampler runs r, each run having a collection M_r objects. These samples can be ordered by ascending likelihood and basically treated as the result of only one nested sampling run with a total of $M = \sum_r M_r$ objects (Skilling, 2006).

Reweighting the samples

Because the samples $\{\theta_i\}_j$ acquired during nested sampling have different weights w_j, they cannot be used to display the posterior or its marginal distributions directly. In order to obtain *equally weighted samples*, i.e. samples drawn from the posterior, one can apply *staircase sampling* (Sivia, 2006, chapter 9), which is similar to importance sampling but requires to draw only one random number and ensures that a sample cannot be drawn twice.

First a cumulant "staircase" S_k is constructed,

$$S_k = \sum_{j=1}^{k} \frac{w_j}{w_{\max}}, \tag{2.5.20}$$

where $w_{\max} = \max_j w_j$ is the largest weight of the nested samples. The largest step in the staircase will have thus the height 1. The staircase is now used to choose a subset
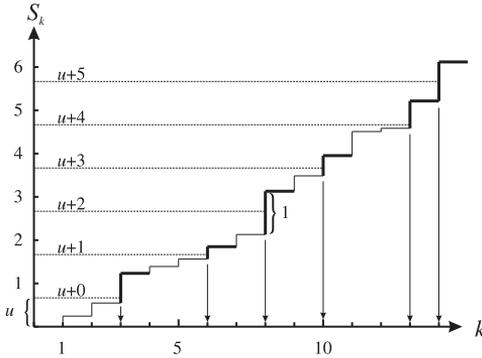
Figure 2.14: Staircase sampling. The cumulant staircase, S_k, is plotted versus the sample number, k. A sample is selected when its "stair" is "hit" by a horizontal line at the offset $u + n$, where u is a random number uniformly distributed in $[0, 1)$, and n is integer. The selected samples are indicated by arrows, and the corresponding "stairs" are drawn with a thicker line. In the drawing, the sample $k = 8$ has the largest weight, and therefore a "stair" of height 1. Independently of the random number drawn it will be selected always.

of nested samples $\{\theta_i\}_n^{\mathrm{eq}}$ that have equal weight. To this end, first, a random number u uniformly distributed in $[0, 1)$ is drawn. Then, for each integer number $n = 0, 1, \ldots$, the sample is selected that has the smallest index, k, and, in addition, possesses a step S_k in between $u + n$ and $u + n + 1$ in the stair case (figure 2.14),

$$\{\theta_i\}_j^{\mathrm{eq}} = \{\theta_i\}_{k_{\mathrm{min},j}} \text{ where } k_{\mathrm{min},j} = \min_{u+j-1 < S_k < u+j} (k) \quad . \tag{2.5.21}$$

The resulting samples $\{\theta_i\}_n^{\mathrm{eq}}$ are then equivalent to samples directly taken from the posterior and have equal weights, however, many of the original samples are discarded in this way.

Final remarks

Nested sampling requires that the objects in the collection are *independently* sampled from the constrained prior. These samples could be produced with any standard sampling method. However, when Markov chain Monte Carlo is used to generate the samples it might take a long time until a random walk leads to sufficient independence. Moreover, since the shape of the constrained prior is changing constantly the proposal probabilities used in the MCMC must be adjusted during the progress of nested sampling. The same applies to the correlation time that will vary depending on how difficult it is to sample the constrained prior, and it must be evaluated "on the fly" as well. In total, care must be taken when nested sampling is set up based on MCMC, and one should always check whether the results are plausible.

After these final remarks on nested sampling the next chapter will focus on the inference problem of FRET-based fluorophore localization relative to a macromolecule.

3 Development of the Nano-Positioning System

The idea of trilateration, i.e. the localization of an object in space from distances measured between the object and three other known positions (figure 3.1), is conceptually simple and has been applied often to the analysis of experiments in which FRET was used to obtain distance restraints between fluorophores attached to biological macromolecules (Mukhopadhyay et al., 2001; Mekler et al., 2002; Rasnik et al., 2004; Kapanidis et al., 2005, 2006; Santoso et al., 2010; Ermolenko et al., 2007a,b; Margittai et al., 2003; Woźniak et al., 2008; Andrecka et al., 2008). Determining the position of a fluorophore can provide valuable information that might solve biochemical questions when the fluorophore to be localized is attached to a part of the macromolecule that cannot be observed by standard methods like X-ray crystallography or NMR spectroscopy.

The basic idea behind such FRET localization experiments is the following: As a prerequisite the macromolecule of interest must be specifically labeled with donor and acceptor fluorophore pairs. The positions of some of the labeling sites will be known relative to the partially defined structure of the macromolecule. The fluorophores at these labeling sites will be called satellites (SATs) in this work, those attached to the remaining labeling sites, which are not present in the structure, will be called antennas (ANTs). The objective is now to localize the ANTs by using the distances inferred from separate FRET measurements between pairs of ANTs and SATs. Each ANT-SAT pair must of course consist of a donor and an acceptor fluorophore, and usually a fluorophore will be either donor or acceptor, unless three types of fluorophores are used. In that case, one of the fluorophores could act as both donor and acceptor depending on its FRET partner. Because of the obvious similarity to the Global Positioning System (GPS), namely the localization of the ANT, this method will be referred to as "Nano-Positioning System" (NPS).

Usually, one assumes that the FRET efficiency is directly related to the distance of two fluorophores as described in chapter 2.2. Further, as soon as the Förster distance is characterized, the problem of localization is equivalent to the minimization of a scoring function by adjusting the positions of the fluorophores. In the most simple case of three FRET measurements and exactly known SAT positions, the global minimum of the scoring
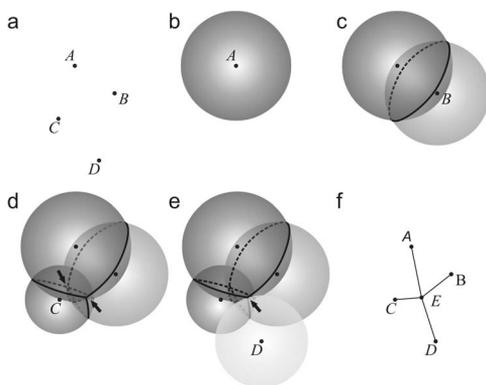
Figure 3.1: Trilateration basics. (a) The positions of four points in space, A, B, C and D are known, the position of a fifth point E (not shown) is unknown. (b) When the distance $\overline{AE}$ is known the position of the point E is limited to a sphere with radius $\overline{AE}$ centered at A. (c) A second distance $\overline{BE}$ constraints E to be on the intersection of the two spheres. (d) A third distance $\overline{CE}$ leaves only two possible positions for E (arrows). (e) With a fourth distance $\overline{DE}$ all four spheres intersect in exactly one point, E. (f) By construction, point E satisfies all four distance constraints.
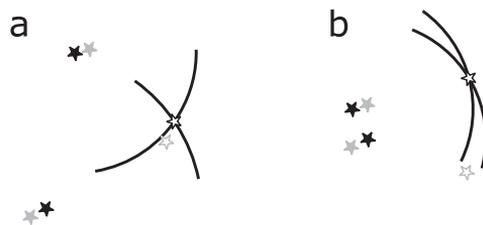
Figure 3.2: Effects of satellite placement on the localization uncertainty. Two-dimensional example. The "true" fluorophore positions are shown in gray, those based on the available information in black. SATs and ANTs are shown as filled and open stars, respectively. Position constraints of the ANT are partly shown (black circle segments). (a) The SATs differ much in position and the constraints are almost perpendicular at their intersection. The error of the inferred ANT position is on the same scale as the uncertainties of the SAT positions. (b) The SATs are very close and the constraints intersect at a small angle. The error of the inferred ANT position is much larger than the uncertainties of the SAT positions.

function can be calculated analytically by trilateration from the SAT positions as shown by Andrecka et al. (2008). With more than three FRET measurements and also more realistic approaches that account for the attachment of the fluorophores via a linker, the analysis becomes much more sophisticated. Unfortunately, in this case the physical model can get obscured by the complicated data processing, which often implicates assumptions not clearly stated by the authors (Margittai et al., 2003; Knight et al., 2005; Choi et al., 2010). The reader might hence be confused or not aware of the pitfalls of the method applied. Furthermore, it is difficult to understand the results in detail and judge whether the taken approach is justified or not.

In this work, instead of using such ad-hoc methods for the analysis of FRET localization experiments, the different error sources that contribute to the localization uncertainty will be identified first, and a physical model that produces the observed data will be stated. Then, Bayesian parameter estimation will be applied to the problem, and by doing so, one is forced (by probability calculus) to state clearly the approximations made in the inference process.

The first and most obvious error source is the measurement error of the FRET efficiency. It consists of both, systematical and statistical contributions. In the following, it will be assumed that both contributions were estimated correctly and were combined in the total FRET efficiency error, which is assumed to be normal distributed (section 5.1.3).

Another source of uncertainty is the limited information on the SAT positions of the fluorophores introduced by the attachment to the macromolecule via flexible linkers. In addition, the placement of the SATs is of great importance to the final localization uncertainty. For instance, the measurements between an ANT and two SATs that are attached at completely different parts of the macromolecule should contribute more information compared to a scenario, in which the SATs are attached at almost the same position (see figure 3.2).

The last and most important source of uncertainty treated here are the orientation effects in FRET. Often, the reported fluorescence anisotropies of fluorophores attached to macromolecules exceed 0.2 (Rasnik et al., 2004; Andrecka et al., 2008; Andrecka, 2009; Roy et al., 2008), and, depending on the average orientation of the transition dipole moments of donor and acceptor, the Förster distance may vary considerably (section 2.3.2). Unfortunately, the fluorophore orientations are not known, and this in turn precludes the direct relation between FRET efficiency and distance. Although these complications are

known in general, many authors still work with the isotropic dynamic averaging model, even when the fluorescence anisotropies are moderate or high, and thereby completely neglect a major source of uncertainty (Rasnik et al., 2004; Andrecka et al., 2008; Chen et al., 2009; Choi et al., 2010).

The contribution of this work to the field of FRET-based structural studies of biomolecules is to estimate the influence of these orientation effects on the localization uncertainty that goes beyond a maximum error estimation proposed by Dale et al. (1979) and Ivanov et al. (2009).

Three closely related models were developed by the author, which all base upon the model proposed by Dale et al. (1979) (see section 2.3.2), but differ in the parametrization and the kind of data that can be analyzed.

The first model can be used to compute the ANT position only and is restricted to the analysis of FRET efficiency measurements in between several SATs and a single ANT. As the spatial positions of the fluorophores and the Förster distances of each individual FRET efficiency measurement are used as parameters, the model is called "*position - Förster distance* model".

The second model is more general: it can be used to analyze networks of fluorescent dyes with FRET efficiency only or both, FRET efficiency and anisotropy, measured between arbitrary numbers of SATs and ANTs. This analysis can infer the positions and average transition dipole orientations of any of the fluorophores in the network. Since these quantities are used as parameters the method is hence called "*position - orientation* model".

Finally, in the third model, called "*position - orientation* model with docking", the *position - orientation* model has been extended in order to determine the positions and orientations of the subunits comprising a macromolecular complex.

3.1 The position - Förster distance model

This section focuses on the NPS parametrized in the fluorophore positions and Förster distances as described by Muschielok et al. (2008).

3.1.1 Model assumptions and parametrization

In the *position - Förster distance* NPS model a system of donor and acceptor fluorophores attached to a macromolecule is treated in terms of fluorophore positions and Förster distances. The method is limited to the analysis of N_{sat} satellite fluorophores and only one antenna fluorophore (see figure 3.3). An experiment performed with several ANTs must be separated into multiple problems with only one ANT, which have to be analyzed separately. The fluorophore positions are assumed to be fixed relative to the macromolecule, meaning that the position fluctuations are much smaller than the separation of the fluorophores. As a consequence, the fluorophores and linkers must be in a stable conformation within the time resolution of the experiment. The same applies to the macromolecule, which is modeled as a rigid object. Furthermore, the macromolecular structure is assumed to be known from, e.g. , X-ray crystallography or NMR spectroscopy studies.

For each FRET pair that consists of the ANT at position $\boldsymbol{x}_a$ and the ith SAT ($i = 1,2,\ldots,N_{\text{sat}}$) at position $\boldsymbol{x}_i$, one can calculate the expected FRET efficiency $\mathcal{E}_{ai}$ (equation (2.3.3)) given that the dynamically (but not isotropically) averaged Förster distance R_{ai} is known,

$$\mathcal{E}_{ai}\left(\boldsymbol{x}_a, \boldsymbol{x}_i, R_{ai}\right) = \frac{1}{1 + \left(\left|\boldsymbol{x}_a - \boldsymbol{x}_i\right|/R_{ai}\right)^6}. \tag{3.1.1}$$
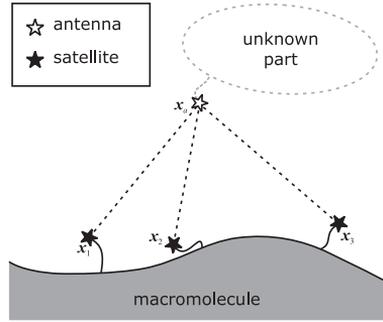
Figure 3.3: Schematic drawing of the *position*-*Förster distance* model. The antenna fluorophore at position $\boldsymbol{x}_a$ is attached to a part of the macromolecule with unknown position. Several satellite fluorophores at positions $\boldsymbol{x}_i$ are attached to the macromolecule by carbon chain linkers. The sequentially measured FRET efficiencies between the satellites and the antenna are shown as dotted lines.

Note that, in order to improve legibility, a different notation is used here for the dynamically averaged Förster distance. The notations R_{ai} in equation (3.1.1) and $\left\langle R^6 \right\rangle$ in equation (2.3.4) are related by $R_{ai} \equiv \sqrt[6]{\left\langle R^6 \right\rangle}$.

3.1.2 Likelihood

The likelihood $p(\{E_{ai}\}|\boldsymbol{x}_a, \{\boldsymbol{x}_i\}, \{R_{ai}\}, I)$ describes how well the FRET efficiencies $\{E_{ai}\} = E_{a1}, E_{a2}, \ldots, E_{aN_\mathrm{sat}}$ measured between the ANT and each SAT, can be reproduced by the model with fluorophore positions $\boldsymbol{x}_a, \{\boldsymbol{x}_i\} = \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{N_\mathrm{sat}}$ and Förster distances $\{R_{ai}\} = R_{a1}, R_{a2}, \ldots, R_{aN_\mathrm{sat}}$. I denotes the background information, i.e. the FRET model used. Under the assumption that the measurements are independent, the likelihood can be factorized, and each likelihood factor $p\left(E_{ai}|\boldsymbol{x}_a, \boldsymbol{x}_i, R_{ai}, I\right)$ describes one measurement,

$$p(\{E_{ai}\}|\boldsymbol{x}_a, \{\boldsymbol{x}_i\}, \{R_{ai}\}, I) = \prod_i p(E_{ai}|\boldsymbol{x}_a, \boldsymbol{x}_i, R_{ai}, I). \qquad (3.1.2)$$

When the E_{ai} are normal distributed around the expected FRET efficiency $\mathcal{E}_{ai}$, the corresponding likelihood factor can be written as

$$p(E_{ai}|\boldsymbol{x}_a, \boldsymbol{x}_i, R_{ai}, I) = \frac{1}{\Delta E_{ai}\sqrt{2\pi}} \exp\left\{ -\frac{\left[\mathcal{E}_{ai}\left(\boldsymbol{x}_a, \boldsymbol{x}_i, R_{ai}\right) - E_{ai}\right]^2}{2\left(\Delta E_{ai}\right)^2} \right\}$$
$$= L_i\left(|\boldsymbol{x}_a - \boldsymbol{x}_i|/R_{ai}\right). \qquad (3.1.3)$$

When regarded as a function of the parameters and as indicated by the abbreviation $L_i\left(|\boldsymbol{x}_a - \boldsymbol{x}_i,|/R_{ai}\right)$ the likelihood factor in equation (3.1.3) depends only on the ratio of the ANT-SAT distance and the Förster distance, $|\boldsymbol{x}_a - \boldsymbol{x}_i|/R_{ai}$. The measurement error of the FRET efficiency is denoted by ΔE_{ai} and is implicitly contained in the background information I.

3.1.3 Prior

The prior $p(\boldsymbol{x}_a, \{\boldsymbol{x}_i\}, \{R_{ai}\}|I)$ can be factorized when the positions of each dye and the Förster distances of each measurement are independent, so that

$$p(\boldsymbol{x}_a, \{\boldsymbol{x}_i\}, \{R_{ai}\}|I) = p(\boldsymbol{x}_a|I) \prod_i \left[p(\boldsymbol{x}_i|I)p(R_{ai}|I)\right]. \qquad (3.1.4)$$

At this level of approximation, it is obvious that the ANT position and the SAT posi-

tions are mutually independent. However, the independence of the fluorophore positions and the Förster distances is not evident since these quantities can be calculated from the fluorophore positions and the average transition dipole moment orientations. The same criticism applies to the independence of different Förster distances R_{ai} and R_{aj}, as they are correlated by the common position and average transition dipole moment orientation of the ANT fluorophore.

In the following, the mutual independence of the fluorophore positions and Förster distances will be assumed, but it should be considered as an approximation. The meaning of this approach will become clear later on in section 3.2.3, where the prior assignment in the *position - orientation* model is introduced. The discussion of the validity of this approximation will be postponed until section 6.1.3.

Fluorophore positions

The priors of the antenna position, $p(\boldsymbol{x}_a|I)$, and satellite position, $p(\boldsymbol{x}_i|I)$, are assumed to be constant inside the volume V_a and V_i accessible to the ANT and SAT fluorophore, respectively. As there is no reason to prefer a particular position, this flat location prior is invariant under infinitesimal translations (subsection 2.4.3). Outside of the accessible volume the prior is set to 0, i.e.

$$p\left(\boldsymbol{x}_{a/i}|I\right) = \begin{cases} \left(\int_{\boldsymbol{x}_{a/i} \in V_{a/i}} d\boldsymbol{x}_{a/i}\right)^{-1} & \text{for } \boldsymbol{x}_{a/i} \in V_{a/i} \\ 0 & \text{otherwise} \end{cases} \tag{3.1.5}$$

The accessible volume is limited by steric clashes of the fluorophore with the macromolecule, and, when present, the attachment by a linker that prevents complete detachment from the macromolecule. In this work, the accessible volumes are computed with a crude model that treats the fluorophore as a ball with known diameter and the linker as a flexible rod with known length and diameter (see figure 3.4a and section 5.3.1 for details), but one could also use more realistic models.

For computational purposes, each SAT position prior is approximated by a superposition of isotropic 3-dimensional Gaussian kernels $S_{ij}(\boldsymbol{x}_i)$ (see figure 3.4b),

$$p(\boldsymbol{x}_i|I) \approx \sum_j w_{ij} S_{ij}(\boldsymbol{x}_i). \tag{3.1.6}$$

Each Gaussian is normalized to unit volume, and the weights w_{ij} of each kernel sum up to 1 ($\sum_j w_{ij} = 1$), so that $p(\boldsymbol{x}_i|I)$ is normalized (see section 5.4 for details). The ANT position prior is not approximated in this manner.

Förster distances

The Förster distance priors base upon equations (2.3.4) and (2.3.5) as explained in section 2.3.2. The idea is that, given the experimentally measured isotropic Förster distance and the fluorescence anisotropies of donor and acceptor (section 5.1.4), one can compute the dynamically averaged Förster distance, when the relative orientations of the average transition dipole moments and the vector interconnecting the fluorophores are known. As a second prerequisite, the distribution of transition dipole moment orientations, which are sampled while the fluorophores are in the excited state, must be axially symmetric in order to apply the FRET model of Dale et al. (1979). Since neither the orientation of the average transition dipole moments nor the interconnecting vector between the
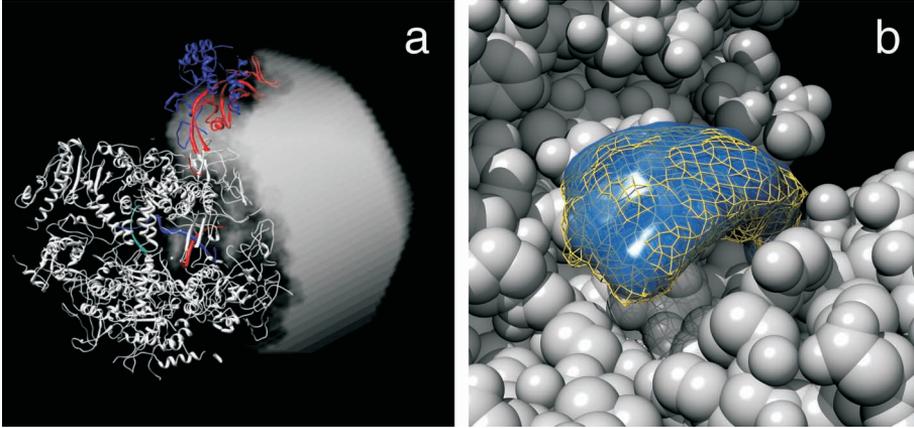
Figure 3.4: (a) A typical antenna fluorophore position prior. The accessible volume (gray cloud) is bounded by the macromolecule (gray, red and blue ribbons). Other constraints, such as the attachment of the ANT by a long flexible linker molecule, can be used in addition. (b) The accessible volume of a typical satellite fluorophore is shown as yellow mesh. The approximation of the SAT position prior $p\left(\boldsymbol{x}_i|I\right)$ consists of several superimposed Gaussian kernels $S_{ik}(\boldsymbol{x}_i)$ and its 68% credible interval is shown as blue surface. The protein atoms are shown as gray spheres.

fluorophores is known, each possible orientation should contribute with the same weight to the Förster distance distribution.

In this work, Monte Carlo simulations were used to compute samples from the Förster distance distribution (section 5.3.2). The histograms of these samples were then approximated by superpositions of Gaussians (see figure 3.5), which are, in turn, approximations of the Förster distance priors $p\left(R_{ai}|I\right)$ given by

$$p(R_{ai}|I) \approx \sum_k v_{ik}Q_{ik}(R_{ai}). \tag{3.1.7}$$

In order to simplify the subsequent calculation, the functions $Q_{ik}(R_{ai})$ are pseudo-Gaussian kernels that are normalized to unity area. Each kernel has an approximately Gaussian form and a zero-crossing at $R_{ai} = 0$ and consists of a sum of two Gaussians with a positive and negative amplitude (see section 5.4 for details). The weights v_{ik} sum up to 1 ($\sum_k v_{ik} = 1$).

3.1.4 Posterior

According to Bayes' theorem (see section 2.4), the posterior can be computed from the likelihood and the prior. With the approximations used in the assignment of the prior, the not yet normalized posterior factorizes into

$$p(\boldsymbol{x}_a, \{\boldsymbol{x}_i\}, \{R_{ai}\}|\{E_{ai}\}, I) \;\propto\; p(\boldsymbol{x}_a|I)\prod_i p(E_{ai}|\boldsymbol{x}_a, \boldsymbol{x}_i, R_{ai}, I)p(\boldsymbol{x}_i|I)p(R_{ai}|I). \tag{3.1.8}$$
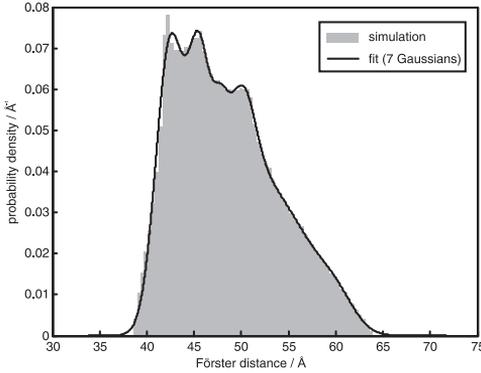
Figure 3.5: A typical Förster distance distribution. $1 \cdot 10^6$ samples from the Förster distance distribution were computed (fluorescence anisotropies $r_a = 0.2$, $r_i = 0.3$, isotropic Förster distance $R_{ai}^{\text{iso}} = 50\text{Å}$) and a histogram of the simulated Förster distances was created (gray area). The approximate Förster distance prior $p(R_{ai}|I)$ (black line) used for further computation consists of a superposition of 7 pseudo-Gaussian kernels $Q_{ik}(R_{ai})$ fitted to the histogram.

In order to determine the estimate of the ANT position, the posterior is marginalized by integrating over all SAT positions and Förster distances,

$$p(\boldsymbol{x}_a|\{E_{ai}\}, I) = \int d\{\boldsymbol{x}_i\} \int d\{R_{ai}\} \, p(\boldsymbol{x}_a, \{\boldsymbol{x}_i\}, \{R_{ai}\}|\{E_{ai}\}, I) \qquad (3.1.9)$$

Further, the marginal posterior can be written as

$$p(\boldsymbol{x}_a|\{E_{ai}\}, I) \propto p(\boldsymbol{x}_a|I) \prod_i \int d\boldsymbol{x}_i \int dR_{ai} \, L_i(|\boldsymbol{x}_a - \boldsymbol{x}_i|/R_{ai})p(\boldsymbol{x}_i|I)p(R_{ai}|I)$$

$$= p(\boldsymbol{x}_a|I) \prod_i \sum_j \int d\boldsymbol{x}_i \int dR_{ai} \, L_i(|\boldsymbol{x}_a - \boldsymbol{x}_i|/R_{ai})w_{ij}S_{ij}(\boldsymbol{x}_i) \, p(R_{ai}|I)$$

$$= p(\boldsymbol{x}_a|I) \prod_i \sum_j w_{ij}K_{ij}(\boldsymbol{x}_a). \qquad (3.1.10)$$

The functions $K_{ij}(\boldsymbol{x}_a)$ are defined as

$$K_{ij}(\boldsymbol{x}_a) = \int d\boldsymbol{x}_i \int dR_{ai} \, L_i(|\boldsymbol{x}_a - \boldsymbol{x}_i|/R_{ai})S_{ij}(\boldsymbol{x}_i) \, p(R_{ai}|I), \qquad (3.1.11)$$

and represent fuzzy spherical shells centered at the position of the jth Gaussian kernel of the ith SAT, $S_{ij}(\boldsymbol{x}_i)$. Details of the computation of $K_{ij}(\boldsymbol{x}_a)$ are described in section 5.4. The fuzzy shells have approximately the radius $R_{ai}^{\max} \cdot (1 - 1/E_{ai})^{1/6}$, where $R_{ai}^{\max}$ is the maximum of $p(R_{ai}|I)$. Consequently, $K_{ij}(\boldsymbol{x}_a)$ is the contribution of the FRET efficiency measurement between ANT and the ith SAT, given the SAT position is described by the Gaussian kernel $S_{ij}(\boldsymbol{x}_i)$.

The sum $\sum_j w_{ij}K_{ij}(\boldsymbol{x}_a)$ in equation (3.1.10) superimposes the fuzzy sphere shells and represents therefore the total contribution of the FRET efficiency measurement between the ANT and the ith SAT. When the approximate radii of the fuzzy shells $K_{ij}(\boldsymbol{x}_a)$ are smaller than the spread of the center positions of the Gaussian kernels S_{ij}, also the sum $\sum_j w_{ij}K_{ij}(\boldsymbol{x}_a)$ will be approximately a fuzzy sphere shell. The product of the measurement contributions from different SATs with the ANT prior results then in the marginal ANT position posterior (see figure 3.6).

The composition of the marginal antenna position posterior is reasonable, since the widths of the contribution of the ith measurement comprise the uncertainty of the ith satellite position, the imprecisely known Förster distance, as well as the error of the
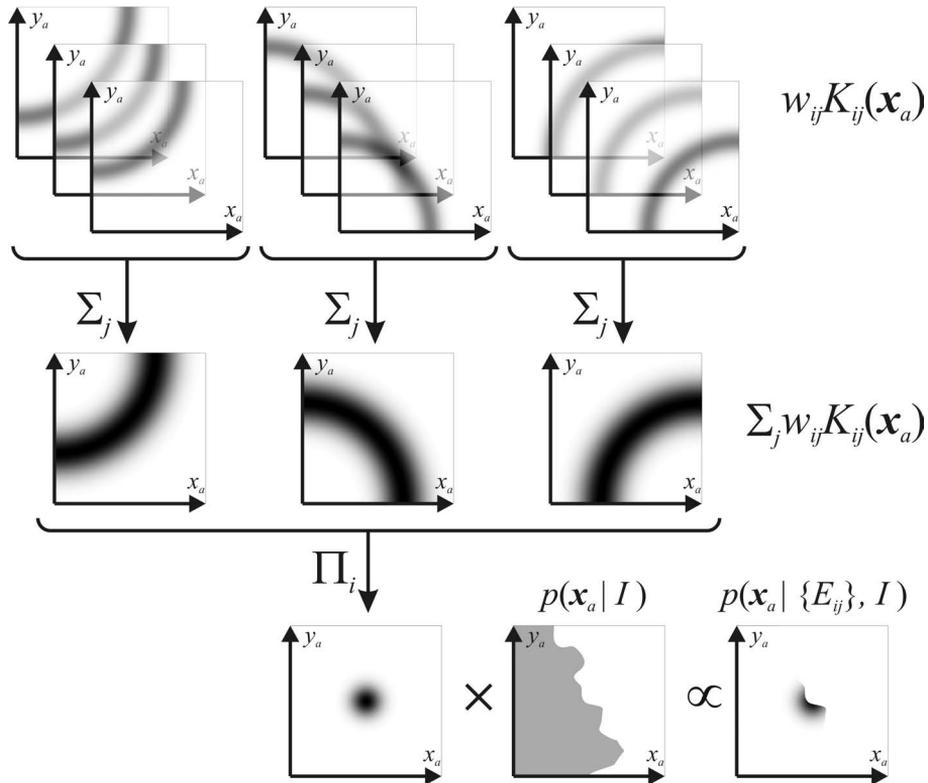
Figure 3.6: The marginal ANT position posterior, $p(\boldsymbol{x}_a|\{E_{ai}\}, I)$ (bottom right), consists of a product of the ANT position prior, $p(\boldsymbol{x}_a|I)$, and contributions from each FRET measurement, $\sum_j w_{ij} K_{ij}(\boldsymbol{x}_a)$. These sums are linear combinations of fuzzy spheres (here shown schematically as fuzzy circles) centered around the Gaussian kernels $S_{ij}(\boldsymbol{x}_i)$ the SAT position prior, $p(\boldsymbol{x}_i|I)$, is consisting of.
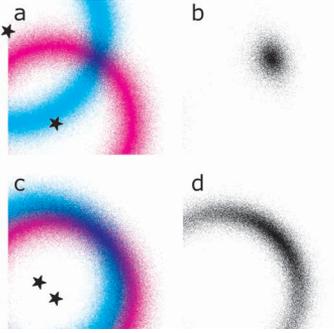
Figure 3.7: Effect of the satellite placement on the inferred antenna fluorophore position. The schematic drawing depicts the contributions of two FRET efficiency measurements shown as fuzzy circles. When the SAT positions (stars) are substantially different and the contributions of each measurement, $\sum_j w_{ij} K_{ij}(\boldsymbol{x}_a)$ (cyan and magenta densities), intersect approximately perpendicularly (a), the localization uncertainty (gray density) will be smaller (b) compared to a scenario, in which the intersection angles are small (c, d).

corresponding FRET efficiency measurement. Also, the effect of the satellite fluorophore attachment site on the inferred position is reflected in a plausible way, as shown in figure 3.7.

3.1.5 Evidence

Only three out of $4 \times N_{\text{sat}}$ parameter space dimensions remain after marginalization of the posterior. Hence, the evidence Z (equation (2.4.9)), used to normalize the product of likelihood and prior (equation (3.1.8)), can be computed easily, as soon as the unnormalized marginal ANT position posterior $p\left(\boldsymbol{x}_a|I\right) \prod_i \sum_j w_{ij} K_{ij}(\boldsymbol{x}_a)$ has been evaluated on a reasonable amount of supporting points, $\boldsymbol{x}_a^{(k)}$. In this work, a lattice with grid spacing δx is used, and the evidence Z is given by

$$Z = \int d\boldsymbol{x}_a \; p\left(\boldsymbol{x}_a|I\right) \prod_i \sum_j w_{ij} K_{ij}(\boldsymbol{x}_a) \approx \delta x^3 \sum_k p\left(\boldsymbol{x}_a^{(k)}|I\right) \prod_i \sum_j w_{ij} K_{ij}(\boldsymbol{x}_a^{(k)}).$$
(3.1.12)

After the *position - Förster distance* NPS model has been explained the focus of the next section focuses is the more general *position - orientation* NPS model. There, the restriction to FRET networks with only one antenna fluorophore will be lifted.

3.2 The position - orientation model

In this section, the NPS analysis problem will be generalized to several antenna fluorophores. To this end, a more fundamental set of parameters will be used, namely the positions and average transition dipole moment orientations of each fluorophore. Furthermore, additionally to the FRET efficiencies, a new kind of observable is introduced, the FRET anisotropy. Because of these generalizations, there are no symmetries that one could use to compute the marginal antenna position posterior analytically, and it is necessary to switch to fully numerical methods. Therefore, only the formal definition of the data analysis problem will be given here, while the computation details are described in section 5.5.

3.2.1 Model assumptions and parametrization

As before, in the *position - Förster distance* model (section 3.1.1), each fluorophore in the FRET network is located at a fixed position. The position of the ith fluorophore is denoted

Figure 3.8: In the *position - orientation* NPS model, the fluorophores (double arrows) are described by their position $\boldsymbol{x}_i$ and average transition dipole moment orientation $\boldsymbol{\Omega}_i$. The axially symmetric fluctuations around the average orientation are indicated by motion lines. Here, three SATs attached to a known macromolecule structure (bottom) and two ANTs attached to an unknown macromolecule part (top) are shown. Separately measured FRET efficiencies and/or FRET anisotropies are indicated by dashed lines between the fluorophores.
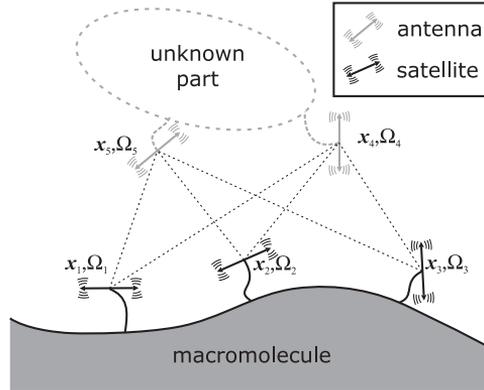
by $\boldsymbol{x}_i$, irrespective of whether it is a SAT or ANT.

Exactly like in the assignment of the Förster distance priors in the *position - Förster distance* model (section 3.1.3), the transition dipole moments of each fluorophore are assumed to sample axially symmetric distributions. The symmetry axes, i.e. the average transition dipole moment orientations of each fluorophore, will be denoted by $\boldsymbol{\Omega}_i$ and are assumed to be fixed. In order to meet the conditions of dynamical averaging (section 2.3.2), the reorientation of the transition dipole moments must be much faster compared to the times the fluorophores stay in the excited state.

The FRET network (see figure 3.8) is then defined by the parameters $\{\boldsymbol{x}_i, \boldsymbol{\Omega}_i\} = (\boldsymbol{x}_1, \boldsymbol{\Omega}_1), \ldots, (\boldsymbol{x}_N, \boldsymbol{\Omega}_N)$, where N is the number of fluorophores in the network.

Given that the average axial depolarizations $\langle d_i^x \rangle$ are known for each fluorophore from measurement of fluorescence anisotropies (see equation (2.3.6)), one can calculate the expected FRET efficiencies and FRET anisotropies. The expected FRET efficiency, $\mathcal{E}_{ij}(\boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{\Omega}_i, \boldsymbol{\Omega}_j)$ (see equation (3.1.1)), is defined by

$$\mathcal{E}_{ij}\left(\boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{\Omega}_i, \boldsymbol{\Omega}_j\right) = \frac{1}{1 + \left(\left|\boldsymbol{x}_i - \boldsymbol{x}_j\right| / R_{ij}\left(\boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{\Omega}_i, \boldsymbol{\Omega}_j\right)\right)^6}, \qquad (3.2.1)$$

where the dynamically averaged Förster distance, $R_{ij}\left(\boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{\Omega}_i, \boldsymbol{\Omega}_j\right)$, is computed directly from the isotropic Förster distance R_{ij}^{iso} for the FRET pair ij and from the angles Θ_{ij}, Θ_{ji} and Ψ_{ij} that describe the orientation of the fluorophores relative to each other (figure 3.9). These angles are, in turn, functions of the model parameters $\boldsymbol{x}_i$, $\boldsymbol{x}_j$, $\boldsymbol{\Omega}_i$ and $\boldsymbol{\Omega}_j$.

The expected FRET anisotropy (see equations (2.3.10) and (2.3.9)) is given by

$$\mathcal{A}_{ij}(\boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{\Omega}_i, \boldsymbol{\Omega}_j) = \frac{2}{5} \langle d_i^x \rangle \langle d_j^x \rangle \left\{ \frac{3}{2} \cos^2 \left[\Psi_{ij}\left(\boldsymbol{\Omega}_i, \boldsymbol{\Omega}_j\right)\right] - \frac{1}{2} \right\}, \qquad (3.2.2)$$

where $\langle d_i^x \rangle$ and $\langle d_j^x \rangle$ are the axial depolarizations of the fluorophores i and j, respectively[1].

---

[1] Note that $\mathcal{E}_{ij}$ and $\mathcal{A}_{ij}$ can take up to 4 and 2 different values, respectively, depending on the measured fluorescence anisotropies $r_{\infty,i}$ and $r_{\infty,j}$. This is due to the ambiguities caused by the possibly negative sign of the average axial depolarizations when $0 \leq r_{\infty,i}/r_{0i} \leq 1/4$ and/or $0 \leq r_{\infty,j}/r_{0j} \leq 1/4$ (equation (2.3.6)).

Figure 3.9: A FRET pair consisting of two fluorophores i and j at positions $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$, respectively. The dynamically averaged Förster distance R_{ij} is calculated by using the angles Θ_{ij}, Θ_{ji} and Ψ_{ij}. Θ_{ij} and Θ_{ji} are the angles between the preferred transition dipole orientations, $\boldsymbol{\Omega}_{i/j}$, and the vector interconnecting the fluorophores, $(\boldsymbol{x}_j - \boldsymbol{x}_i)$. Ψ_{ij} is the angle between the preferred fluorophore orientations $\boldsymbol{\Omega}_i$ and $\boldsymbol{\Omega}_j$. Alternatively, Ψ_{ij} can be expressed as a function of Θ_{ij}, Θ_{ji} and Φ_{ij}, the angle between the planes defined by the orientations $\boldsymbol{\Omega}_{i/j}$ and the vector $(\boldsymbol{x}_j - \boldsymbol{x}_i)$ can be used to compute R_{ij}. Unlike the FRET efficiency, the FRET anisotropy depends only on Ψ_{ij} and is independent of the direction of the vector interconnecting the fluorophores.
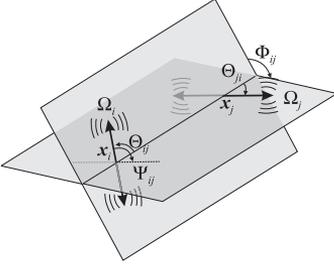
3.2.2 Likelihood

Since FRET efficiencies and/or FRET anisotropies are measured independently, the likelihood is the product of contributions from each FRET pair,

$$p\left(\{E_{ij}\}, \{A_{ij}\}|\{\boldsymbol{x}_i, \boldsymbol{\Omega}_i\}, I\right) = \prod_{ij \in M} L_{ij}\left(\boldsymbol{x}_i, \boldsymbol{\Omega}_i, \boldsymbol{x}_j, \boldsymbol{\Omega}_j\right). \quad (3.2.3)$$

L_{ij} denotes the contribution of the FRET pair ij as a function of the positions and average transition dipole moment 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, L_{ij} is assigned to be an uni- or bivariate Gaussian

$$L_{ij}\left(\boldsymbol{x}_i, \boldsymbol{\Omega}_i, \boldsymbol{x}_j, \boldsymbol{\Omega}_j\right) = \begin{cases} \mathcal{N}_{\Delta E_{ij}}\left(E_{ij} - \mathcal{E}_{ij}\right) & E_{ij} \text{ measured} \\ \mathcal{N}_{\Delta A_{ij}}\left(A_{ij} - \mathcal{A}_{ij}\right) & A_{ij} \text{ measured} \\ \mathcal{N}_{\Delta E_{ij}}\left(E_{ij} - \mathcal{E}_{ij}\right) \cdot \mathcal{N}_{\Delta A_{ij}}\left(A_{ij} - \mathcal{A}_{ij}\right) & E_{ij} \text{ and } A_{ij} \text{ measured.} \end{cases}$$
$$(3.2.4)$$

Here, $\mathcal{N}_\sigma(.)$ is the normal distribution of standard deviation σ centered at 0. ΔE_{ij} and ΔA_{ij} denote the measurement uncertainties of FRET efficiency and FRET anisotropy, respectively. The expected observables $\mathcal{E}_{ij}$ and $\mathcal{A}_{ij}$ of course depend on $\boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{\Omega}_i$ and $\boldsymbol{\Omega}_j$, as already mentioned in section 3.2.1.

When the average axial fluorescence depolarizations have ambiguous signs, $s_i = \text{sgn}(\langle d_i^x \rangle)$, the likelihood in equation (3.2.3) must be averaged over all combinations of $\{s_i\}$. This is equivalent to the marginalization over all average axial depolarizations with the signs $s_i = 1$ and $s_i = -1$ being equally probable. Consequently, equation (3.2.3) is generalized to

$$p\left(\{E_{ij}\}, \{A_{ij}\}|\{\boldsymbol{x}_i, \boldsymbol{\Omega}_i\}, I\right) = \frac{1}{\prod_k S_k} \sum_{\{s_k\}} \prod_{ij \in M} L_{ij}^{(s_i s_j)}\left(\boldsymbol{x}_i, \boldsymbol{\Omega}_i, \boldsymbol{x}_j, \boldsymbol{\Omega}_j\right), \quad (3.2.5)$$

where the dependencies on the signs of the average axial fluorescence depolarizations is denoted by the indices $(s_i s_j)$. $S_k = 2$ when $\langle d_k^x \rangle$ is ambiguous, otherwise $S_k = 1$. Hence, $\prod_k S_k$ is the number of combinations of $\{s_k\}$.

Figure 3.10: The fluorophore i is described by its transition dipole moment at position $\boldsymbol{x}_i$ that fluctuates (indicated by motion lines) around its preferred orientation $\boldsymbol{\Omega}_i$ (double arrow). $\boldsymbol{\Omega}_i$ is described by the polar and azimuth angle, θ_i and ϕ_i, relative to the frame of reference of the macromolecule, (x, y, z). In the analysis, 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 of the average transition dipole moment.
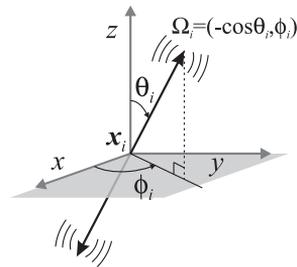
3.2.3 Prior

Similarly than in the *position - Förster distance* model, all correlations between the model parameters, here the positions and average transition dipole moment orientations, are neglected. Correlations between positions and orientations of different fluorophores might be caused either by direct interactions of the fluorophores with each other or by a change of the macromolecule structure induced by the labeling. However, here it will be assumed that these correlations are negligible and thus the prior can be factorized and written as

$$p\left(\{\boldsymbol{x}_i, \boldsymbol{\Omega}_i\}|I\right) = \prod_i p\left(\boldsymbol{x}_i, \boldsymbol{\Omega}_i|I\right). \tag{3.2.6}$$

It is further assumed that the position and the average transition dipole orientation of a fluorophore are independent, and consequently, that the prior for the fluorophore i factorizes as well, so that

$$p\left(\boldsymbol{x}_i, \boldsymbol{\Omega}_i|I\right) = p\left(\boldsymbol{x}_i|I\right) p\left(\boldsymbol{\Omega}_i|I\right). \tag{3.2.7}$$

The actual meaning of equation (3.2.7) is the lack (or ignorance) of any detailed knowledge about the fluorophore structure, its attachment to the macromolecule and the interactions of macromolecule and fluorophore. For example, there could be certain stable conformations of the fluorophore-linker-macromolecule construct that introduce correlations to $\boldsymbol{x}_i$ and $\boldsymbol{\Omega}_i$. Also, fluorophore positions in close proximity to the macromolecule surface are accessible only when the fluorophore aligns parallel to the surface.

These and other scenarios are ignored here due to the difficulties of obtaining reliable quantitative information for large macromolecular system by simulations. Instead, a flat position prior $p\left(\boldsymbol{x}_i|I\right)$ will be assigned inside the volume accessible to the fluorophores, which will be computed with the same crude model as in the *position - Förster distance* NPS model (sections 3.1.3 and 5.3.1).

In order to assign a prior in the average transition dipole moment orientation, $p\left(\boldsymbol{\Omega}_i|I\right)$, the parametrization of $\boldsymbol{\Omega}_i$ needs to be clarified first. $\boldsymbol{\Omega}_i$ will be parametrized by $-\cos\theta_i$ and ϕ_i, where θ_i and ϕ_i are the polar and azimuth angle relative to the frame of reference of the macromolecule, (x, y, z) (see figure 3.10). A flat prior is assigned in this parametrization, and the prior of each orientation can be written as

$$\begin{aligned} p\left(\boldsymbol{\Omega}_i|I\right) &= p\left(-\cos\theta_i, \phi_i|I\right) = (2\pi)^{-1} \\ &\text{for } -1 \leq -\cos\theta_i < 1 \text{ and } 0 \leq \phi_i < \pi. \end{aligned} \tag{3.2.8}$$

$p\left(\boldsymbol{\Omega}_i|I\right)$ is invariant under arbitrary rotations of the coordinate system and, consequently, does not favor any particular orientation of the average transition dipole moments.

3.2.4 Posterior and evidence

Since the inference problem stated here is far more complex than the one of the *position - Förster distance* NPS model, the computation of the posterior and the evidence must be performed fully numerically. In this work, Monte Carlo methods are used to this end, and the reader is referred to sections 2.5 and 5.5 for details. The computation result will be a set of samples that are used to represent and visualize the posterior $p\left(\{\boldsymbol{x}_i, \boldsymbol{\Omega}_i\} | \{E_{ij}\}, \{A_{ij}\}, I\right)$.

In the next section, the *position - orientation* NPS model will be extended to model in addition the macromolecule part the antenna fluorophores are attached to.

3.3 The position - orientation model with docking

Now, the *position - orientation* model introduced in the last section will be extended in order to perform "remote docking", i.e. the inference of relative positions and orientations of two or more macromolecules within a complex. This model is of interest when the structures of the components of the complex are known, but there is no structure of the whole complex, and when FRET efficiency or/and anisotropy can be measured between FRET pairs distributed over different components. Determining the component locations is to be preferred over simple localization of fluorophores, since finally one is interested in the properties of the biochemically important object and not in the properties of the probes.

3.3.1 Model assumptions, parametrization and likelihood

In order to dock the parts of a macromolecular complex by means of FRET, a frame of reference is attached to each constituent. One of the components of the complex is chosen to be the reference, and its coordinate system $(x_{(0)}, y_{(0)}, z_{(0)})$ will be called "root". $(x_{(0)}, y_{(0)}, z_{(0)})$ is in fact identical with the coordinate system used in the other NPS models. The relative alignment of the remaining N_{mol} components to the root component is then described by the origins $\boldsymbol{o}_{(k)}$ and orientations $\boldsymbol{\Xi}_{(k)}$ of the frames of reference $(x_{(k)}, y_{(k)}, z_{(k)})$ (figure 3.11), where $k = 1, \ldots, N_{\mathrm{mol}}$ enumerates the component.

In addition to the assumptions made in the *position - orientation* model (section 3.2.1), the parts of the macromolecular complex are regarded as rigid bodies that have a fixed position and orientation relative to each other. In that case, the fluorophore i linked to the k_i^{th} component is uniquely described by its position $\boldsymbol{x}_i^{(k_i)}$ and average transition dipole orientation $\boldsymbol{\Omega}_i^{(k_i)}$ relative to the reference frame k_i. With the known reference frame position and orientation, the fluorophore position $\boldsymbol{x}_i^{(0)}$ and orientation $\boldsymbol{\Omega}_i^{(0)}$ in the root reference frame can be calculated by a simple coordinate transformation.

The expected FRET efficiency and FRET anisotropy values, $\mathcal{E}_{ij}$ and $\mathcal{A}_{ij}$, are then computed from $\boldsymbol{x}_i^{(0)}$ and $\boldsymbol{\Omega}_i^{(0)}$ according to the equations (3.1.1) and (3.2.1), used already in the *position - orientation* model (section 3.2). The likelihood is computed exactly like in the *position - orientation* model using $\boldsymbol{x}_i^{(0)}$ and $\boldsymbol{\Omega}_i^{(0)}$, the fluorophore positions and average transition dipole orientations in the root reference frame.

Figure 3.11: NPS with docking. The structure of several macromolecules constituting a complex is known, but their relative position and orientation is not. The "root" reference frame is assigned to one of the macromolecules (here macromolecule 0), and the other macromolecules (here macromolecule 1) are docked to the root macromolecule by means of separately measured FRET efficiencies or/and anisotropies (dashed lines). The position and orientation of the kth docked macromolecule ($k \geq 1$) is parametrized by the origin $\boldsymbol{o}_{(k)}$ and the orientation $\boldsymbol{\Xi}_{(k)}$ of an attached reference frame $(x_{(k)}, y_{(k)}, z_{(k)})$. The fluorophores (double arrows) are described as in the *position - orientation* model by the position $\boldsymbol{x}_i^{(k_i)}$ and orientation $\boldsymbol{\Omega}_i^{(k_i)}$ of their average transition dipole moment. However, these parameters are given now relative to the reference frame k_i of the macromolecule the fluorophore is attached to (indicated by the superscript $^{(k_i)}$).
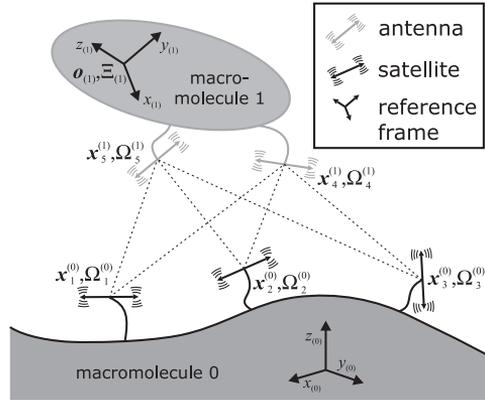
3.3.2 Prior

Like in the *position - orientation* model, the prior is assumed to factorize, i.e. the positions and orientations of the fluorophores ($\{\boldsymbol{x}_i^{(k_i)}, \boldsymbol{\Omega}_i^{(k_i)}\}$) and the reference frames ($\{\boldsymbol{o}_{(k)}, \boldsymbol{\Xi}_{(k)}\}$) must be independent,

$$p\left(\{\boldsymbol{x}_i^{(k_i)}, \boldsymbol{\Omega}_i^{(k_i)}\}, \{\boldsymbol{o}_{(k)}, \boldsymbol{\Xi}_{(k)}\} | I\right) = \prod_i \left[p\left(\boldsymbol{x}_i^{(k_i)} | I\right) p\left(\boldsymbol{\Omega}_i^{(k_i)} | I\right) \right] \prod_k \left[p\left(\boldsymbol{o}_{(k)} | I\right) p\left(\boldsymbol{\Xi}_{(k)} | I\right) \right]. \tag{3.3.1}$$

This implies, of course, the same assumptions as described in section 3.2.3. Additionally, to enable a factorization as in equation (3.3.1), it must be assumed that neither the different parts of the macromolecular complex do interfere with each other by steric clashes, nor that a fluorophore attached to the component k can interact with another component $l \neq k$. For some situations, these simplifications might be unrealistic and one should account for such effects in future NPS versions.

Similar to the position and average transition dipole orientation priors of the fluorophores, the position prior of the reference frame origin, $p\left(\boldsymbol{o}_{(k)} | I\right)$, is constant within a volume $V_{(k)}$ and zero outside,

$$p\left(\boldsymbol{o}_{(k)} | I\right) = \begin{cases} \left(\int_{\boldsymbol{o}_{(k)} \in V_{(k)}} d\boldsymbol{o}_{(k)}\right)^{-1} & \text{for } \boldsymbol{o}_{(k)} \in V_{(k)} \\ 0 & \text{otherwise} \end{cases} \tag{3.3.2}$$

The volume $V_{(k)}$ must be chosen large enough to allow a wide range of positions around the "root" macromolecule.

The orientation of the kth reference frame is parametrized by $\boldsymbol{\Xi}_{(k)} =$
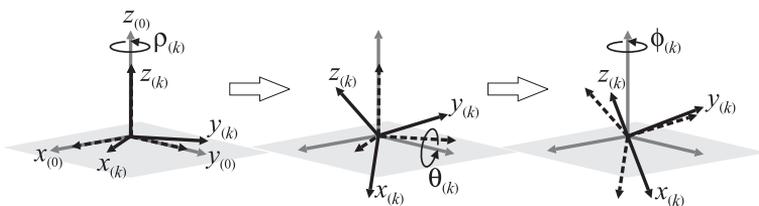
Figure 3.12: Parametrization of the reference frame orientation. The coordinate axes $x_{(k)}$, $y_{(k)}$ and $z_{(k)}$ of the kth reference frame (black arrows) are shown relative to the root reference frame, $(x_{(0)}, y_{(0)}, z_{(0)})$ (gray arrows). The orientation $\boldsymbol{\Xi}_{(k)}$ is defined by the angles $\rho_{(k)}$, $\theta_{(k)}$ and $\phi_{(k)}$ of three consecutive rotations (curved arrows) around the axes $z_{(0)}$, $y_{(0)}$ and $z_{(0)}$, respectively. The state of the kth reference frame before each rotation is shown as dashed arrows.

$(\rho_{(k)}, -\cos\theta_{(k)}, \phi_{(k)})$, where $\rho_{(k)}$, $\theta_{(k)}$ and $\phi_{(k)}$ are the angles of three subsequent rotations (figure 3.12). The orientation prior is assigned to be constant as well, as given by

$$p\left(\boldsymbol{\Xi}_{(k)}|I\right) = p\left(\rho_{(k)}, -\cos\theta_{(k)}, \phi_{(k)}|I\right) = (8\pi^2)^{-1}$$
$$\text{for } -\pi \leq \rho_{(k)} < \pi, \ -1 \leq -\cos\theta_{(k)} < 1, \text{ and } -\pi \leq \phi_{(k)} < \pi. \tag{3.3.3}$$

This prior is invariant under infinitesimal translations and rotations of the root reference frame and satisfies the requirements of a maximum entropy prior (section 2.4.3).

3.3.3 Posterior and evidence

The posterior and the evidence are calculated with the same methods like in the *position-orientation* NPS model (section 3.2.4). Additionally to the marginal positions and orientations of the fluorophores, the distributions of the reference frame origins and orientations can be calculated.

Furthermore, another kind of marginalization is possible that can be used to compute the estimated position of an arbitrary atom or any other point in the reference frame of the docked macromolecule. Let the point Q (e.g. an atom) be located at an arbitrary, fixed position $\boldsymbol{q}^{(k)}$ in the reference frame k. One can compute the posterior estimate of the position $\boldsymbol{q}^{(l)}$ of this point in any other arbitrary reference frame l. This is done by calculating the coordinates $\boldsymbol{q}^{(l)}$ from $\boldsymbol{q}^{(k)}$ and the origin and orientation of the kth reference frame contained in each posterior sample.

In this way, an estimate of the center and orientation of the docked macromolecule can be calculated by choosing the point to be at the center of the macromolecule and at a position far from the center, respectively.

After this final remark on the details of NPS analysis, in the next chapter, a biologically important macromolecular complex will be introduced briefly. It is the eukaryotic RNA polymerase II elongation complex, a macromolecular complex to which NPS was already successfully applied to.

4 The eukaryotic RNA polymerase II

The genetic information contained in the deoxyribonucleic acid (DNA) of most cells of any living organism needs to be read out to produce proteins.For most proteins, this flow of information is indirect with the DNA being *transcribed* into ribonucleic acid (RNA), which is then *translated* into proteins (Thieffry and Sarkar, 1998), as proposed by Francis Crick in 1958 (Crick, 1958) and revised to greater generality later (Crick, 1970).

The transcription of the DNA by synthesis of an complementary RNA chain is carried out by DNA dependent RNA polymerases (Alberts et al., 2008, chapter 6). These multisubunit enzymes catalyze the polymerization of nucleoside triphosphates (NTPs), the building blocks of RNA, in a highly conserved fashion (Steitz, 1998). Without going into detail, in eukaryotic cells different RNA polymerases exist, which produce RNAs that play different roles in cellular processes. One of them, the *RNA polymerase II* (Pol II), was studied in this work by NPS. It produces mainly the so called precursor messenger RNA (pre-mRNA), which is further modified to become messenger RNA (mRNA) and codes for the sequence of proteins.

4.1 Structure and function

Pol II has a molecular weight of approximately $514\,\mathrm{kDa}$ and is approximately $13\,\mathrm{nm}$ in diameter. It consists of 12 protein subunits (Rpb1, ..., Rpb12) and comprises a 10 subunit core, as well as two peripheral protein subunits, the Rpb4/7 heterodimer (Cramer et al., 2008).

First structures of Pol II were obtained by electron microscopy (Darst et al., 1991) and X-ray crystallography (Fu et al., 1999), followed by a structures of higher resolution that show the 10 subunit core (Cramer et al., 2001). Nowadays, crystal structures have been observed that show the complete 12 subunit Pol II in a complex with DNA, RNA and NTPs (Kettenberger et al., 2004), called the *elongation complex*. A detailed review of current research in the field of eukaryotic RNA polymerases was given by Cramer et al. (2008).

The elongation complex is characterized by Pol II bound to one strand of a partially melted (i.e. opened) DNA double helix, the template DNA strand (see figure 4.1). The melted DNA region without base pairing between the template and the complementary strand, the so called nontemplate DNA, is called *transcription bubble*. In this region, the synthesized RNA strand is base-paired with the template DNA, and aligned in a way that an incoming NTP can bind at the enzymatically active site of Pol II. When the NTP is able to base pair with the corresponding base in the template DNA it gets covalently bound to the RNA strand under formation of the byproduct pyrophosphate. After incorporation of the new nucleotide the polymerase shifts by one base, so that the next unpaired template DNA base becomes aligned at its active site. A few bases away from the active site the DNA-RNA hybrid gets separated. In this way the template DNA can hybridize again with its complementary strand, the nontemplate DNA, after it has exited Pol II. The NTPs and the RNA enter and exit Pol II through the NTP entry channel and the RNA exit channel, respectively, while the template and nontemplate DNA slide through and above the enzyme during the elongation process.
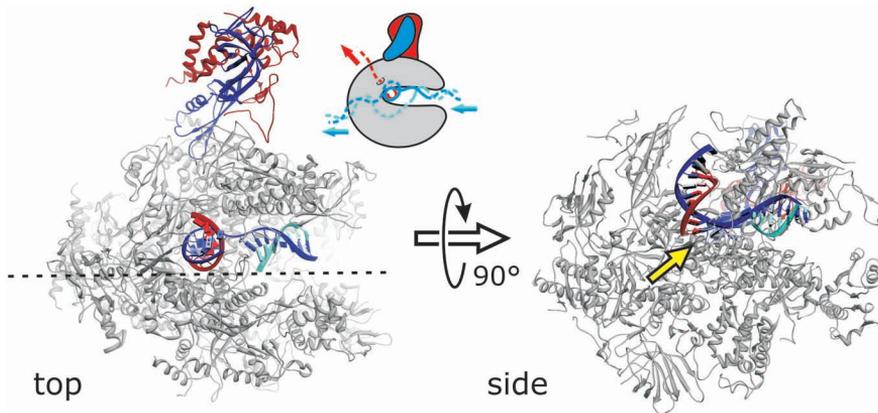
Figure 4.1: Crystal structure of the RNA polymerase II elongation complex. Two standard views, the top and the side view of the PDB structure 1Y1W (Kettenberger et al., 2004) are shown. The Pol II core, Rpb4 and Rpb7 are shown as gray, red and blue ribbons, respectively. The nucleic acids are colored red (RNA), blue (template DNA) and cyan (nontemplate DNA). In the side view panel the NTP incorporation site is marked by an arrow, and the part of Pol II in front of the nucleic acids was cut away as indicated by the dashed line in the front view panel. The inset shows Pol II schematically in the top view and in the same color code as the crystal structures. Nucleic acids resolved and unresolved in the crystal structure are shown as solid and dashed lines, respectively. The motion of the nucleic acids during transcription is indicated by arrows.

The DNA ahead of the Pol II active site, when viewed in the transcription direction, is called *downstream* DNA. Its nucleotides are usually numbered with positive integers, the active site being at register $+1$. The DNA lying behind the Pol II active site is called *upstream* DNA and is numbered with negative integers that start at -1 directly behind the active site. Thus, there is no nucleotide "0" in this system. The RNA is numbered by positive integers, beginning at $+1$ directly behind the active site, so that the RNA nucleotide at register $+n$ will usually be able to form a base pair with the template DNA nucleotide at register $-n$.

In the cell, Pol II does not work on its own. Instead, it is embedded in a network of dependencies with other proteins, which initiate and regulate the transcription process in eukaryotes (Hahn, 2004; Cramer, 2007; Alberts et al., 2008, chapter 6). The proteins most closely interacting with Pol II are called general transcription factors. The transcription factor IIB (TFIIB), for example, accurately places the Pol II at the transcription start site and plays a crucial role in transcription initiation, whereas the TATA binding protein (TBP), a subunit of the transcription factor IID (TFIID), is required to recognize the TATA box, a specific DNA sequence found in front of many transcription start sites. The general transcription factor IIF (TFIIF) functions in both transcription initiation and elongation and is needed for recruitment of other transcription factors (Eichner et al., 2010, and references therein) that constitute the so-called pre-initiation complex.

Until today, only few crystal structures exist that show Pol II in assembly with transcription factors (Bushnell et al., 2004; Kettenberger et al., 2003, 2004; Kostrewa et al., 2009), and even in these studies only fragments of the transcription factors have been observed. However, such structures shed light upon the molecular mechanism of transcription initiation and regulation.

4.2 Questions addressed

Although many interesting structural details have been observed already by X-ray crystallography, the crystal structures of Pol II elongation complexes lack nevertheless the density of the nascent RNA after it has left Pol II, the density of the single stranded nontemplate DNA in the transcription bubble, and the density of the upstream DNA.

Therefore, two questions arise: Do the unobserved but yet present RNA and DNA regions have a well defined position? And if there is a well defined position, where are the nucleic acids located? Since the term "well-defined" depends on the method used to observe the position (see introduction), it will be assumed that the nucleic acids and Pol II are rigid enough to be observed in single-molecule FRET experiments as one or only a few populations, which can be assigned to a biochemically reasonable state of the elongation complex.

The position of the nascent RNA chain in the transcribing Pol II complex is of general interest since the RNA is modified before it gets transported to the ribosome, and Pol II together with other associated proteins could promote or initiate this post-processing. In addition, the knowledge of the nontemplate DNA position has important implications not only for elongation but also for understanding transcription initiation and completes the picture of the Pol II elongation complex.

To answer these questions, the Nano-Positioning System was applied to the Pol II elongation complex. The size of the complex makes it perfectly suitable for FRET studies. Moreover, site-specific fluorescent labeling, which would be extremely difficult in a macromolecular complex of this size is feasible since the elongation complex can be reconstituted from its constituents. Instead of labeling the 10 subunit core, the separately expressed Rpb4/7 as well as commercially available nucleic acids serve as a platform to attach satellite and antenna fluorophores.

In an earlier work, the trilateration-based "ancestor" of the NPS has been used to determine the maximum probable positions of antenna fluorophores attached to the 5'-end of RNAs with lengths varying between 17 and 35 nucleotides (Andrecka et al., 2008). There, the observation 26 and 29 nucleotide long RNAs suggested the binding of the 5' RNA end to the Pol II core at the so called dock domain. This finding apparently contradicted studies in which RNA of comparable length could be covalently cross-linked to the Rpb4/7 heterodimer (Újvári and Luse, 2006). However, it was difficult to judge whether there was a true contradiction, or whether both experiments were not comparable. The latter was likely since first, the elongation complexes used by Andrecka et al. (2008) did not contain any TFIIB, and second because no meaningful uncertainties were stated for the position of the 5' RNA end.

Apart from these questions, there is ongoing research on the structure of the initial transcribing complex (ITC). The ITC consists of Pol II and the general initiation factors, and is the state of the transcription machinery just before Pol II starts to synthesize RNA continuously (Young et al., 2002; Murakami and Darst, 2003). In the ITC, Pol II is still close to the transcription start site and synthesizes short RNA fragments, a process called abortive initiation (Carpousis and Gralla, 1980; Grachev and Zaychikov, 1980; Munson and Reznikoff, 1981; Kapanidis et al., 2006; Revyakin et al., 2006).

The transition from the ITC to the elongation complex is still unclear. Besides the position of TFIIB in the ITC, which is crucial to understand in detail how Pol II transits from the ITC to the elongation complex, also the position of TBP and the upstream DNA is important, since TFIIB binds Pol II and TBP located at the TATA box (Kostrewa et al., 2009; Liu et al., 2010a). The analysis of preliminary FRET data will be shown that reveals the position and orientation of TBP and the TATA box relative to Pol II (Treutlein et al., unpublished) (section 6.6).

The next chapter allows to have a closer look on some experimental details and on the data analysis needed to perform and interpret the experiments outlined above. The focus is clearly on the analysis part since this work focuses on the NPS as a general method for structural studies of macromolecules, which is based upon FRET efficiency and additional FRET anisotropy data that could be obtained also by other means than those described in the following.

The reader interested mainly in the general results may skip the next chapter and is encouraged to continue reading in chapter 6 instead.

5 Materials and methods

5.1 Experimental methods, data acquisition and data pre-processing

In this section, the experimental methods for measuring the data analyzed in this work are described briefly.

Sample preparation, all FRET efficiency and fluorescence anisotropy measurements analyzed in this work, as well as data pre-processing were done by Joanna Andrecka, Barbara Treutlein, Julia Nagy and Maria Angeles Izquierdo Arcusa. Design of labeling sites on proteins and nucleic acid oligomers, as well as labeling of proteins were done by Joanna Andrecka and Barbara Treutlein.

All experiments were performed in the absence of NTPs, so that the Pol II complexes were stalled and not transcribing.

5.1.1 Macromolecular complexes

This section describes the labeling and the reconstituted macromolecular complexes used in the FRET measurements analyzed in this work.

Fluorescent labeling

Nucleic acid oligomers labeled with Alexa Fluor dyes 647, 555, and tetra-methyl rhodamine (TMR) were purchased from IBA and Biomers (Andrecka, 2009). Site-specific labeling of proteins was achieved with maleimide chemistry. Here, single cystein mutants of Rpb4 and Rpb7 subunits of eukaryotic RNA polymerase II that were expressed in *E. coli* were labeled with Alexa Fluor 647 maleimide (Invitrogen / Molecular Probes) (Andrecka, 2009). Single cystein mutants of the *S. cerevisiae* TATA binding protein were labeled site-specifically with cyanine dye Cy3 maleimide (GE Healthcare) and Alexa Fluor 647 maleimide (Invitrogen / Molecular Probes). The labeling sites are listed in table 5.1.

Reconstituted Pol II complexes

Pol II elongation complexes were assembled from nucleic acids and the Pol II enzyme following a protocol of Armache et al. (2003) as described by Andrecka et al. (2008). Depending on the FRET pair measured in the experiment, some of above the components were labeled with fluorophores.

The complexes used for inference of fluorophore positions attached to the RNA consisted of nucleic acid scaffolds, the 10 subunit Pol II core purified from *S. cerevisiae* (provided by Patrick Cramer), and the Rpb4/7 heterodimer. The latter was either expressed in *E. coli* or purified from *S. cerevisiae* (both in collaboration with Patrick Cramer). The nucleic acid scaffolds were assembled from a 41 nucleotides (nt) long template DNA, a 41 nt long biotinylated nontemplate DNA and a 29 nt RNA. The DNA possessed a 11 nt long mismatched region, which formed an artificial transcription bubble and did readily bind to the Pol II enzyme without help of other factors. The RNA was designed to bind the template strand in the middle of the mismatched region over a length of 8 nt, and thus mimicked the nascent RNA strand and stabilized the complex.

5 Materials and methods

abbrev.	residue	chain	fluorophore	linker len.	PDB-ID	project
tDNA(-10)	T28	templ. DNA	Alexa 647	12 C	1Y1W	1,2
tDNA(-10)	T28	templ. DNA	TMR	12 C	1Y1W	3
tDNA(+3)	T16	templ. DNA	Alexa 647	12 C	1Y1W	1,2
tDNA(+3)	T16	templ. DNA	TMR	12 C	1Y1W	3
tDNA(+7)	T12	templ. DNA	Alexa 647	12 C	1Y1W	3
tDNA(+7)	T12	templ. DNA	TMR	12 C	1Y1W	3
tDNA(+9)	T10	templ. DNA	Alexa 647	12 C	1Y1W	1,2
tDNA(+12)	T7	templ. DNA	Alexa 647	12 C	2VUM[‡]	3
RNA(+1)	3'-end	RNA	Alexa 647	6 C	1Y1W	2
RNA(+1)	3'-end	RNA	Alexa 555	6 C	1Y1W	1
RNA(+1)	3'-end	RNA	TMR	6 C	1Y1W	1,3
RNA(+4)	U14	RNA	Alexa 647	12 C	1Y1W	2
RNA(+10)	5'-end	RNA	Alexa 647	6 C	1Y1W	2
Rpb7(C150)	C150	Rpb7	Alexa 647	6 C[†]	1Y1W	1,2,3
Rpb7(C94)	C94	Rpb7	Alexa 647	6 C[†]	1Y1W	1
Rpb4(S73C)	S73C	Rpb4	Alexa 647	6 C[†]	1Y1W	1,2
Rpb7(S16C)	S16C	Rpb7	Alexa 647	6 C[†]	1Y1W	1
RNA(+29)	5'-end	RNA	TMR	6 C	-	1
ntDNA(+1)	T26	nontempl. DNA	TMR	12 C	-	2
ntDNA(+1)	T26	nontempl. DNA	Alexa 555	12 C	-	2
ntDNA(-2)	T24	nontempl. DNA	TMR	12 C	-	2
ntDNA(-2)	T24	nontempl. DNA	Alexa 555	12 C	-	2
ntDNA(-4)	T22	nontempl. DNA	TMR	12 C	-	2
ntDNA(-4)	T22	nontempl. DNA	Alexa 555	12 C	-	2
ntDNA(-7)	T19	nontempl. DNA	TMR	12 C	-	2
ntDNA(-7)	T19	nontempl. DNA	Alexa 555	12 C	-	2
ntDNA(-12)	T14	nontempl. DNA	TMR	12 C	-	2
ntDNA(-12)	T14	nontempl. DNA	Alexa 555	12 C	-	2
ntDNA(-15)	T11	nontempl. DNA	TMR	12 C	-	2
ntDNA(-15)	T11	nontempl. DNA	Alexa 555	12 C	-	2
ntDNA(-18)	T8	nontempl. DNA	TMR	12 C	-	2
ntDNA(-18)	T8	nontempl. DNA	Alexa 555	12 C	-	2
ntDNA(-20)	T7	nontempl. DNA	TMR	12 C	1BNA[+]	3
ntDNA(-30)	T24	nontempl. DNA	TMR	12 C	1VOL[*]	3
ntDNA(-37)	T17	nontempl. DNA	TMR	12 C	1VOL[*]	3
TBP(S159C)	S159C	TBP	Alexa 647	6 C	1VOL[*]	3
TBP(S159C)	S159C	TBP	Cy3	6 C	1VOL[*]	3

Table 5.1: Labeling sites and attached fluorescent dyes used in the measurements. The length of the linker (linker len.) is given as number of backbone atoms. projects: (1) RNA / TFIIB, (2) nontemplate DNA, (3) initial transcribing complex (ITC)
[+] position known relative to a segment of double strand DNA in B-form (Drew et al., 1981, PDB-ID: 1BNA) but unknown relative to the Pol II elongation complex (Kettenberger et al., 2004, PDB-ID: 1Y1W)
[*] position known relative to the TBP/DNA complex (Nikolov et al., 1995, PDB-ID: 1VOL) but unknown relative to 1Y1W
[†] comprises a 2-carbohydrate linker of the fluorophore and the maleimid group
[‡] the structure 2VUM (Brückner and Cramer, 2008) was matched to 1Y1W

56

Elongation complexes with TFIIB from *S. cerevisiae* were prepared as described in (Muschielok et al., 2008), either by directly adding TFIIB during the measurement, or by preincubating the nucleic acid scaffolds with TFIIB.

Elongation complexes used for inference of fluorophore positions attached to the nontemplate DNA were assembled as described above, but possessed a 17 nt long RNA.

Initial transcribing complexes were assembled similarly and purified thereafter by size exclusion chromatography, but contained different nucleic acid scaffolds, and, in addition, the transcription factor IIF and the TBP, both from *S. cerevisiae* (Treutlein et al., unpublished). The nucleic acid scaffolds consisted of a 63 nt long template DNA, a 61 nt long biotinylated nontemplate DNA, and a 7 nt RNA. The DNA double strand was partially mismatched over a region of 18 nt, in which the RNA could bind to the template DNA strand. The first thymine of the TATA box was located at register -37, i.e. 36 nt from the base pairing with the last ribonucleotide at the RNA 3'-end. When labeled, TBP was a full length construct expressed in *E. coli* (in collaboration with Patrick Cramer, the plasmid was provided by Michael Meisterernst). Unlabeled TBP was a deletion mutant that lacked the first 60 N-terminal amino acids.

5.1.2 Isotropic Förster distance measurements

The isotropic Förster distances R^{iso} for all constructs are listed in table 2, 5 and 7 (appendix II) and were determined as described by Andrecka (2009), following a standard procedure (Vamosi et al., 1996). In short, the donor quantum yield was determined from absorption and emission spectra of donor samples versus an ethanol solution of Rhodamine 101 as standard. The isotropic Förster distances were calculated with the PhotochemCAD software (Lindsey, 2008) from donor emission spectra and acceptor absorption spectra, assuming the refractive index $n = 1.3492$ and $\kappa^2 = 2/3$ (see section 2.2.2). All fluorescence spectra were measured in bulk samples using a steady state fluorescence spectrometer (Edinburgh Instruments F900) and for all absorption spectra, a UV-VIS spectrophotometer (Varian Cary 50) was used.

5.1.3 FRET efficiency measurements

FRET efficiency data was collected with a custom-built single-molecule TIRF microscope (Lewis, 2009; Andrecka, 2009) and is listed in tables 1, 4 and 6 (appendix II). In short, the reconstituted complexes were immobilized to the surface of a quartz glass slide functionalized with biotinylated polyethyleneglycol and incubated with neutravidin, which was bound to the biotin on the nontemplate DNA. Time traces of donor and acceptor fluorescence signals acquired for the immobilized complexes were recorded with an acquisition time of 100 ms. The traces were averaged with a ten-frame sliding average filter and the FRET efficiencies $E(t)$ were calculated taking into account instrumental systematic errors as well as heterogeneities between different molecules observed:

$$E(t) = \frac{F_{\mathrm{A}}(t) - \beta F_{\mathrm{D}}(t)}{F_{\mathrm{A}}(t) + \gamma F_{\mathrm{D}}(t)}. \qquad (5.1.1)$$

$F_{\mathrm{D}}(t)$ and $F_{\mathrm{A}}(t)$ denote the time trace of the acquired background-corrected donor and acceptor fluorescence signals, respectively. The factors β and γ are used to correct for leakage of donor fluorescence into the acceptor channel, different detection efficiencies, as well as quantum yields that may vary from molecule to molecule. They are determined

Figure 5.1: Typical FRET efficiency histogram. FRET efficiency histograms were computed by binning the acquired and smoothed FRET efficiencies of well-defined FRET pairs. One ore more Gaussians were fitted to the histograms in order to obtain the centers of the peaks (arrows) and the statistical uncertainties of the center estimates. The uncertainties obtained from the fit and the estimated systematic errors were used as data in the Bayesian analysis.
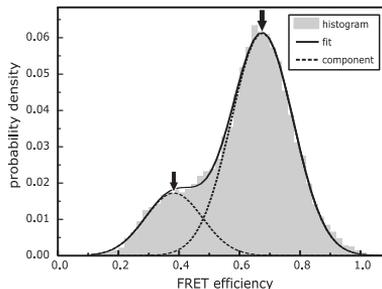
for each molecule as follows,

$$\beta = \frac{\langle F_A \rangle_{D\cancel{A}}}{\langle F_D \rangle_{D\cancel{A}}} \tag{5.1.2}$$

$$\gamma = \frac{\langle F_A \rangle_{DA} - \langle F_A \rangle_{D\cancel{A}}}{\langle F_D \rangle_{D\cancel{A}} - \langle F_D \rangle_{DA}}. \tag{5.1.3}$$

The averages $\langle . \rangle_{DA}$ and $\langle . \rangle_{D\cancel{A}}$ are taken before donor and acceptor bleaching, and before donor bleaching but after acceptor bleaching, respectively. Hence, only molecules exhibiting acceptor bleaching prior to donor bleaching were analyzed.

Average FRET efficiencies E_{ij} between the fluorescent dyes i and j were obtained by least-square fitting a Gaussian to the histograms of the measured FRET efficiencies. When more than one FRET efficiency population was visible in the histogram, multiple Gaussians were used for fitting, each Gaussian representing a population of molecules (see figure 5.1).

For each average FRET efficiency E_{ij}, its uncertainty ΔE_{ij} was computed from the fit error and the systematic error assuming both error sources to be independent. The systematic error was estimated from the standard deviation of the average FRET efficiencies obtained by fitting data acquired on different days.

Pre-processed in this way, the average FRET efficiencies E_{ij} and errors ΔE_{ij} that belong to the same state of the Pol II complex were used as data points in the Bayesian analysis.

5.1.4 Fluorescence anisotropy measurements

Fluorescence anisotropies were acquired as described by Andrecka (2009) with a steady state fluorescence spectrometer (Edinburgh Instruments F900) and are listed in table 3 (appendix II). In short, the instrumental corrections were accounted for, and anisotropy values were calculated as the spectral average over a small region around the maximum of the fluorescence emission spectrum. The measurements were performed on site-specifically labeled complexes with either donor or acceptor. Due to the slow rotational diffusion of the Pol II macromolecular complexes, the measured fluorescence anisotropies were attributed completely to the segmental motion of the fluorophores.

5.1.5 FRET anisotropy data

In this work, only artificial (i.e. simulated) transfer anisotropy data was used (see section 5.3.4), but in principle transfer anisotropies can be acquired in a similar way like fluorescence anisotropies from the polarization resolved acceptor fluorophore emission intensity

after polarized donor excitation. However, it would be best to measure FRET anisotropies like FRET efficiencies in single-molecule experiments to avoid systematic errors, which might occur when several FRET populations are present in the sample.

5.2 Nano-Positioning System software

In order to solve the Nano-Positioning System parameter estimation problem, utility programs and graphical user interfaces (GUIs) were developed in MATLAB (The Math-Works).

Separate GUIs were developed for the NPS, based on the *position - Förster distance* model and the *position - orientation* model, both without and with optional docking of rigid bodies. The GUIs allow for a user-friendly operation of the utility programs as well as for input of FRET measurement data and a convenient visualization of the parameter estimation results.

The utility programs were used to define model parameter priors, to perform the parameter estimation calculations, to visualize the calculation results and to check them for consistency, and finally, to export marginalized position posteriors for visualization in molecular viewing software (see section 5.3.5).

5.2.1 Position - Förster distance model

The parameter estimation calculations based on the *position - Förster distance* model were done according to sections 3.1 and 5.4. The information flow and coarse structure of the software developed to compute the marginal antenna position posterior is shown in figure 5.2.

In short, `FRETnps.m` was used to define the FRET network (i.e. for measurement data input and prior input), to simulate the antenna position prior, to perform the calculations, and to manage the FRET network data as well as the resulting marginal antenna position posterior. The data was stored in the global structure `FRETnpsDat` in memory, as well as in files. In the latter case, two types of files were generated, the *measurement* and the *posterior files*. The measurement files were written and loaded by `FRETnpsGUI.m` and contained the definition of the FRET network and the prior knowledge, i.e. the FRET efficiency with measurement error, the Förster distance and satellite position prior (both as approximations), the antenna position prior (as a binary grid), as well as the positions and element names of the atoms in the PDB file used to compute the antenna position prior. The posterior files were written by `FRETnpsGUI.m` and contained the marginal antenna position posterior in addition to the FRET network definition. The posterior files could be processed further by utility functions.

The `FRETnpsTools.m` GUI was used to start `FRETnps.m` and other utility functions and GUIs. More precisely, the GUIs `ModelSatellitePosition.m` and `ModelFoersterradiusPrior.m` were used to model satellite position priors and Förster distance priors, respectively. `ExportDensity.m` was used to save the marginal antenna position posterior densities in `.MRC` and `.XPLOR` formats, which could be read by molecular viewing software (see section 5.3.5).

The utility program `ConsistencyCheck.m` was used to check the consistency of the analyzed data (see section 5.3.6). Both `ExportDensity.m` and `ConsistencyCheck.m` processed posterior files.
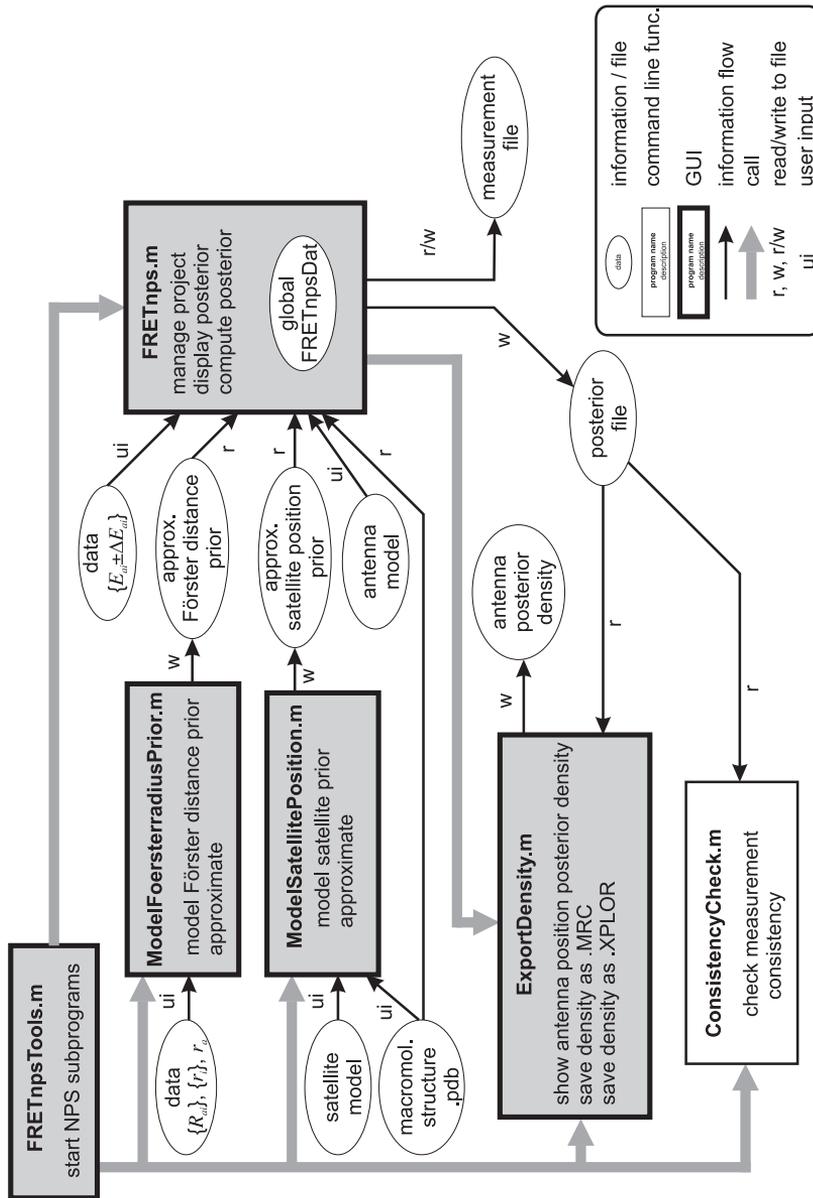
Figure 5.2: *Position - Förster distance* model NPS software overview

Figure 5.3: *Position - orientation* model NPS software overview

5.2.2 Position‑orientation model with optional docking

The parameter estimation calculations based on the *position‑orientation* model with optional docking were done according to sections 3.2 and 3.3. The information flow and coarse structure of the software developed to this purpose is shown in figure 5.3.

In short, `FRETnpsGUI.m` was used to manage the definitions and results of NPS problems, as well as to call the GUIs `FRETnpsEditgeometry.m`, `FRETnpsCalculation.m` and `FRETnpsViewresults.m`, which were used to define the FRET network, perform the calculations and visualize the results. The data was stored in the global structure `FRETnpsDat` in memory, as well as in project files, which could be saved and loaded by `FRETnpsGUI.m`.

`FRETnpsEditgeometry.m` was used to define the FRET network, i.e. to specify the measurement data, to load or edit the priors of fluorophore parameters, and, optionally, as well of reference frame parameters. In addition, it is also possible to introduce distance constraints between fluorophore positions.

The priors of the satellite fluorophore positions were simulated with `ModelSatellitePosition.m` and its sub-functions as described in section 5.2.1. They were manually converted into the box collection prior format (BCP, see section 5.5.3), which could be opened by `FRETnpsEditgeometry.m`. The priors for the parameters of docking reference frames were edited manually with `EditBCP.m`.

The calculations underlying the inference process were based on nested sampling and Metropolis sampling of the posterior. They were performed with the general sampling engine described in section 5.5 and called from the GUI `FRETnpsCalculation.m`.

`FRETnpsViewresults.m` was used to display the results of the calculations. The export of marginal densities of fluorophore positions and of points defined in a docking reference frame was done in the GUI `ExportDensity.m`.

5.3 Simulation and visualization

This section describes how simulations were carried out that were relevant for the assignment of priors, the generation of artificial data for test calculations, and for visualization and testgin of the Nano-Positioning System analysis results.

5.3.1 Calculation of accessible volumes

Fluorophore positions were assumed to be limited by two constraints, steric clashes with the macromolecule, and the maximum length of the linker used for the attachment. A clash was defined by Van der Waals contact of either dye or linker (if present) with the macromolecule.

In order to calculate the accessible volume, i.e. the positions that satisfy the constraints, the fluorophore was approximated with a sphere of diameter D_{fl}. In this work, D_{fl} was set to 7Å for all fluorophores. If the fluorophore was attached to the macromolecule via a flexible linker, e.g. a carbon chain linker, RNA or DNA, the linker was modeled to have the thickness D_{li} and the length L_{li} (see figure 5.4a). The linker lengths used in the calculations were the lengths of the stretched carbohydrate chain in the case of short carbohydrate linkers, and 80% of the contour length of nucleic acid linkers (see table 5.2). The linker and fluorophore diameters were set to the respective minimum diameter in a space-filling representation. In this way, fluorophore positions were allowed that were accessible only at certain linker segment or fluorophore orientations.

The structures of the macromolecules were taken from the Protein Data Bank (PDB format). The labeling sites were C6 in the case of DNA, and O3 in the case of RNA on the respective nucleotides. The labeling sites on the protein were the sulfur atoms of
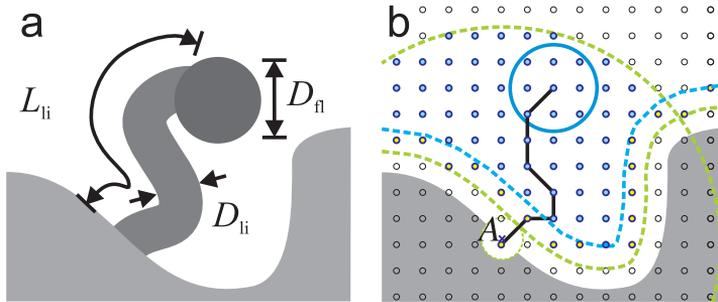
Figure 5.4: Computation of accessible volumes. (a) Coarse model used to compute the accessible volume for a fluorophore. In this model the fluorophore is treated as a sphere of diameter D_{fl}. It is attached to the macro-molecule (light gray area, bottom) by a flexible linker of length L_{li} and diameter D_{li}.
(b) Accessible volumes were computed on a lattice (small gray circles). The lattice points in which a sphere of diameter D_{li} could be placed without causing steric clashes with the macromolecule (small yellow and blue filled circles above the bottom green dashed line). Points directly around the attachment site A were allowed as well (inside the green dashed circle). From these points a sub-set was selected (in between the green dashed lines) that could be connected by a line of at most $L_{li} + D_{fl}/2$ length that started at A and could connect allowed points that are within third order neighborhood only. From these lattice points those were assigned as accessible to the fluorophore (small blue filled circles) which allowed to place a sphere of diameter D_{fl} (big blue open circle) without clash with the macromolecule.

the respective cysteins. If the cystein was introduced by the mutation of a serine, the coordinates of the respective oxygen atom were used in the simulations instead.

In order to calculate the volumes accessible to the satellite fluorophores, a cubic lattice inside a box was used. The box was centered around the attachment point and had the edge lengths $2L_{li} + D_{fl}$. The lattice spacing was equal to $\frac{1}{5}$ of the smallest of the lengths appearing in the model (D_{fl}, D_{li} or L_{li}) by default. In the case of antenna fluorophores, the box was chosen manually to incorporate the macromolecule and a reasonable amount of extra space around it, and a lattice spacing of either 1 or 2 Å was used.

In the initial step of the calculation, all lattice points were chosen which allowed to place a sphere with diameter D_{li} without clashing with the macromolecule. When the attachment of the fluorophore was not considered, D_{fl} was used instead of D_{li}, and the algorithm terminated after the first step, so that all lattice points accessible to the sphere were treated as accessible to the fluorophore.

When a linker was simulated, all points were allowed in addition, which were located inside the sphere with diameter D_{li} centered at the attachment point of the linker to the macromolecule. This was done in order to allow the linker to "escape" the macromolecule interior. From these lattice points those were chosen that could be connected to the lattice point located closest to the attachment point by a line with length smaller than $L_{li} + D_{fl}/2$. The line had to connect neighboring and allowed lattice points only, and was able to fold back on itself (see figure 5.4b). Neighboring lattice points were defined as all points within a third order neighborhood in three dimensions, i.e. at a maximum distance of $\sqrt{3}$ of the lattice spacing. Finally, for all points that were reached by the linker, it was checked whether there was enough space in the surroundings. To this end a sphere of diameter D_{fl} had to fit into the position without causing a clash with the macromolecule. All lattice points that have survived the selection steps were treated as accessible to the fluorophore.

labeling site	fluorophore	attachment position/Å			L_{li}/Å
		x	y	z	
tDNA(-10)[*]	ALX647,TMR	77.9	35.8	-16.4	13
tDNA(+3)[*]	ALX647,TMR	95.7	59.8	-15.3	13
tDNA(+7)[*]	ALX647,TMR	93.2	79.8	-23.4	13
tDNA(+9)[*]	ALX647	93.2	79.8	-23.4	13
tDNA(+12)[*]	ALX647	94.9	87.8	-31.5	13
RNA(+1)[*]	ALX647,ALX555,TMR	105.0	45.7	8.9	7
RNA(+4)[*]	ALX647	92.0	41.7	0.5	13
RNA(+10)[*]	ALX647	85.4	29.1	-15.5	7
Rpb7(C150)[*]	ALX647	117.3	30.0	-64.5	7
Rpb7(C94)[*]	ALX647	118.1	7.3	-74.5	7
Rpb4(S73C)[*]	ALX647	147.4	18.5	-40.5	7
Rpb7(S16C)[*]	ALX647	133.8	19.3	-27.0	7
ntDNA(-20)[+]	TMR	2.7	1.5	0.41	13
ntDNA(-30)[†]	TMR	2.2	18.8	-10.3	13
ntDNA(-37)[†]	TMR	12.0	1.8	-8.5	13
TBP(S159C)[†,‡]	ALX647,CY3	-4.3	-4.4	8.1	23

Table 5.2: Attachment position coordinates (x, y, z) relative to the X-ray structure model, as well as linker lengths L_{li} used to simulate accessible volumes at different labeling sites. Linker diameters $D_{\mathrm{li}} = 4.5$Å as well as fluorophore diameters $D_{\mathrm{fl}} = 7$Å were used in all simulations. Fluorophore abbreviations: ALX647: Alexa 647, ALX555: Alexa 555, CY3: Cyanine 3, TMR: Tetra-methyl rhodamine

[*] relative to Pol II elongation complex (Kettenberger et al., 2004, PDB-ID: 1Y1W)

[+] relative to a B-DNA double strand (Drew et al., 1981, PDB-ID: 1BNA)

[†] relative to TBP/DNA/TFIIB complex (Nikolov et al., 1995, PDB-ID: 1VOL). TFIIB was deleted, the DNA was extended with B-DNA at the 3' and 5'-ends, and the structure was moved to the origin. The attachment points of NT-DNA(-30), NT-DNA(-37) and TBP(S159C) are identical with the positions of the C6 atom of guanine 24, the C6 atom of thymine 17 (both in the nontemplate DNA strand), and the Cα atom of Gln164 (in the TBP), respectively.

[‡] Residues 158-163 in the TBP chain were deleted from the structure but added to the linker because there was no secondary structure element. The attachment point was Gln164, the closest amino acid that shows secondary structure.

5.3.2 Monte Carlo simulation of Förster distances

Förster distances of single FRET pairs

For the calculation of the Förster distance prior in the *position - Förster distance* NPS model, $p\left(R_{ij}|I\right)$, the assumption of independent and completely unknown average transition dipole moment orientations was made.

To simulate $p\left(R_{ij}|I\right)$, random relative orientations of the average transition dipole moment were calculated that correspond to uniformly distributed average transition dipole moment orientations for each fluorophore,

$$\Theta_{ij} = \arccos\left[\text{rand}(0,1)\right], \tag{5.3.1}$$

$$\Theta_{ji} = \arccos\left[\text{rand}(0,1)\right], \tag{5.3.2}$$

$$\Phi_{ij} = 2\pi \cdot \text{rand}(0,1), \tag{5.3.3}$$

where $\text{rand}(0,1)$ denotes uniform pseudo random numbers between 0 and 1. Θ_{ij} and Θ_{ji} are the angles between the average transition dipole orientations of the fluorophores i and j to the line interconnecting the fluorophores, and Φ_{ij} is the angle between the planes defined by each average orientation and the interconnecting line (see figure 3.9).

Θ_{ij}, Θ_{ji} and Φ_{ij}, the isotropic Förster distance R_{ij}^{iso}, as well as the average axial depolarizations $\left\langle d_i^x \right\rangle$ and $\left\langle d_j^x \right\rangle$ calculated from fluorescence anisotropies were used to compute the dynamically averaged Förster distance $R_{ij} = \sqrt[6]{\left\langle R_{ij}^6 \right\rangle}$ (see section 3.2.1). When the axial depolarizations had ambiguous signs, two or four Förster distances were computed for a single set of angles Θ_{ij}, Θ_{ji} and Φ_{ij}. The histogram of the simulated Förster distances was calculated and normalized to unity area, and was used as an approximation of the prior $p\left(R_{ij}|I\right)$.

Förster distances of correlated FRET pairs

Simulations of the joint probability distribution of two dynamically averaged Förster distances, R_{ij} and R_{ik} were carried out in a similar way like the simulations in the section above. The studied FRET pairs, ij and ik have the fluorophore i in common. Since the Förster distances are correlated by the average transition dipole moment orientation and the position of the fluorophore i, the probability distribution $p\left(R_{ij}, R_{ik}|\boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{x}_k, I\right)$ will depend on the relative position of fluorophore i in respect to fluorophores j and k.

In the simulation, independent uniformly distributed orientations of the average transition dipole moments of all fluorophores were assumed. Consequently, the orientations given by the azimuth angles θ_i, θ_j, θ_k and the polar angles ϕ_i, ϕ_j, ϕ_k (see figure 3.9) were simulated as

$$\theta_{i/j/k} = \arccos\left[2 \cdot \text{rand}(0,1) - 1\right], \tag{5.3.4}$$

$$\phi_{i/j/k} = 2\pi \cdot \text{rand}(0,1). \tag{5.3.5}$$

In addition, the isotropic Förster distances of both FRET pairs were assumed to be equal, i.e. $R_{ij}^{\text{iso}} = R_{ik}^{\text{iso}} = R^{\text{iso}}$. This is possible as the only reason that causes a finitely broad Förster distance distribution in the *Position - orientation* NPS model are orientation effects, which are independent of the isotropic Förster distance as a length scale. Without loss of generality, the fluorophores j and k were placed on the $z-$axis in a distance of R^{iso}. The position $\boldsymbol{x}_i$ of the fluorophore i was scanned with a step size of $0.05 \cdot R^{\text{iso}}$ in a plane, which contained the $z-$axis.

The angles $\theta_{i/j/k}$, $\phi_{i/j/k}$, and the fluorophore positions $\boldsymbol{x}_{i/j/k}$ were used to compute

the angles Θ_{ij}, Θ_{ji}, Ψ_{ij} and Θ_{ik}, Θ_{ki}, Ψ_{ik}, which describe the relative orientations of the average transition dipole moments (see figure 3.9). These angles together with the average axial depolarizations $\langle d_i^x \rangle$, $\langle d_j^x \rangle$ and $\langle d_k^x \rangle$ were used to calculate the pair of dynamically averaged Förster distances R_{ij} and R_{ik} at the particular positions of the fluorophore i. When the sign of an average axial depolarization was ambiguous, it was chosen to be either positive or negative with equal probability.

The Förster distance pairs (R_{ij}, R_{ik}) were binned into two-dimensional histograms at each position $\boldsymbol{x}_i$. Subsequently normalized to unity volume, these histograms were approximations of the joint probability distribution $p(R_{ij}, R_{ik}|\boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{x}_k, I)$.

5.3.3 Measuring statistical dependence

A good method to quantify the statistical dependence of two random variables, X and Y, is the normalized mutual information (NMI), $\mathcal{I}(X, Y)$ (Strehl and Ghosh, 2002). This quantity is the normalized dissimilarity of the joint distribution of $p(X, Y|I)$ compared to a reconstruction from the marginal distributions $p(X|I)p(Y|I)$. $\mathcal{I}(X, Y)$ is computed by means of the Kullback-Leibler divergence, D_{KL}, (Kullback and Leibler, 1951) and normalized with the geometric mean of the Shannon informations H of the marginal distributions.

Explicitly, in this work the NMI was used to compute the amount of statistical dependence of the Förster distances R_{ij} and R_{ik} of two FRET pairs with a common fluorophore i (see section 5.3.2 for simulation details). It is given by

$$\mathcal{I}(R_{ij}, R_{ik}) = \frac{D_{\mathrm{KL}}\left[p(R_{ij}, R_{ik}|\boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{x}_k, I) \;||\; p(R_{ij}|I)p(R_{ik}|I)\right]}{\sqrt{H[p(R_{ij}|I)]H[p(R_{ik}|I)]}}, \tag{5.3.6}$$

where the Kullback-Leibler divergence, D_{KL}, of two probability densities $p(A)$ and $p(B)$ is defined as

$$D_{\mathrm{KL}}\left[p(A) \;||\; p(B)\right] = \int dA \; p(A) \ln\left[\frac{p(A)}{p(B)}\right], \tag{5.3.7}$$

and the Shannon information H of a probability density $p(A)$ is given by

$$H[p(A)] = \int dA \; p(A) \ln\left[p(A)\right], \tag{5.3.8}$$

which is basically the negative of the Shannon-Jaynes entropy (equation (2.4.16)) with a uniform measure.

5.3.4 Artificial data for test calculations

For test calculations an artificial data set was generated. Initially, random uniformly distributed fluorophore positions were created within a reasonable volume, but under the condition that the fluorophores are separated by at least 25 Å. The volume was bounded by a cone (half opening angle $60°$) and two spheres centered at the cone tip. The spheres had a radius of 40 and 60 Å for the satellites, and 90 and 110 Å for the antennas. In addition to the fluorophores created above, two more satellite fluorophores were added manually to the network in positions that promised to reduce localization uncertainty. The average transition dipole moment orientations were randomly chosen from a uniform distribution in the solid angle without any constraint. Residual fluorescence anisotropies were randomly chosen from a uniform distribution between 0.15 and 0.32. In this range all experimental fluorescence anisotropies were observed (see table 3,

appendix II). The fluorophore data (positions, average transition dipole orientations and fluorescence anisotropies) are listed in table 8 (appendix II).

Isotropic Förster distances were chosen from a uniform distribution between 55 Å and 65 Å (see table 10, appendix II).

Using these values, the expected FRET efficiencies and anisotropies (tables 9 and 11) were calculated based on the *position - orientation* model (see 3.2, appendix). The expected data was used directly without adding any noise that is used usually to model experimental noise. This was done since it is necessary to know the maximum posterior solution independently of the numerical computation in order to test the convergence of the algorithm.

For FRET efficiency and anisotropy data an error of 0.02 and 0.01 was assumed, respectively.

5.3.5 Probability density visualization

Marginal posterior probability densities of fluorophore positions were computed in two ways. Either, the marginalized posterior was evaluated on a grid (*position - Förster distance* NPS model, see section 5.4), or, normalized histograms of fluorophore positions were calculated from the posterior samples (*position - orientation* NPS model). In both ways, the marginal posterior was represented by an approximate probability density value assigned to each voxel in space.

Densities

The three-dimensional marginal posterior densities of fluorophore positions were saved as files in XPLOR or MRC/CCP4 map format (the latter referred to as MAPFORMAT by Crowther et al. (1996)). The files were loaded with the molecular viewing software UCSF Chimera (Chi, 2010; Pettersen et al., 2004) and displayed as "solid", so that low probability densities were shown more transparent than high probability densities.

Credible volumes

In order to display three-dimensional credible volumes, the probability density $p(\boldsymbol{x})$ was transformed into the function, $c(\boldsymbol{x})$, which has the property that an iso-surface S_P at the level P, i.e.

$$S_P = \{\boldsymbol{x}|c(\boldsymbol{x}) = P\}, \tag{5.3.9}$$

encloses the credible volume that corresponds to the probability P (see section 2.4.4). The function $c(\boldsymbol{x})$ was evaluated at the supporting points $\boldsymbol{x} = \boldsymbol{x}_m$ ($m = 1, \ldots, N_x \cdot N_y \cdot N_z$) of the probability density $p(\boldsymbol{x})$, and saved in MRC/CCP4 format. The supporting points were spaced on a cubic grid with spacing δx consisting of $N_x \cdot N_y \cdot N_z$ lattice sites. The actual size of the spacing was chosen, so that the iso-surfaces were looking smooth and varied between 0.5 and 14 Å for the densities displayed in this work. The MRC/CCP4 files were loaded in UCSF Chimera and used to visualize the credible volumes for an arbitrary probability P. To this end, iso-surfaces of $c(\boldsymbol{x})$ were displayed as "surface" or "mesh".

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

$$p(\xi) \equiv p\left(\boldsymbol{x}_{m(\xi)}\right) \geq p\left(\boldsymbol{x}_{m(\xi+1)}\right), \tag{5.3.10}$$

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

Figure 5.5: Visualization of credible volumes. The position space (left) and the sorted voxel space (right) are shown. The function $c(\boldsymbol{x})$ used to display credible volumes is calculated from the marginal 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 function $c(\boldsymbol{x})$ contoured at the level P (cyan mesh) is the surface of the credible volume corresponding to the probability P.
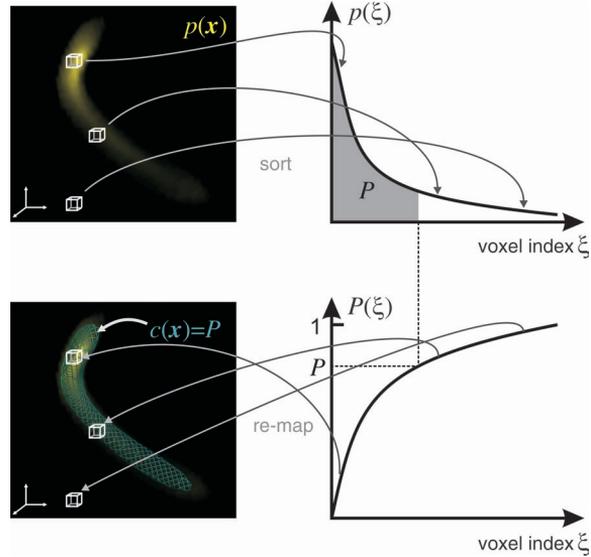
cumulative sorted probabilities, $P(\xi)$, defined as

$$P(\xi) = (\delta x)^3 \sum_{\xi'=1}^{\xi} p\left(\xi'\right), \tag{5.3.11}$$

were calculated and mapped back into the three-dimensional position space (figure 5.5).

5.3.6 NPS measurement consistency check

Depending on the NPS model used, the consistency of measurements was checked in different ways, which will be shown in the following.

Position - Förster distance model

In the *position - Förster distance* model, the consistency of FRET efficiency measurements was checked by comparison of the posterior, $p(\boldsymbol{x}_a|\{E_{ai}\}, I)$, with the contributions of each measurement between satellite i and the antenna a, $\sum_j w_{ij} K_{ij}(\boldsymbol{x}_a)$. To this end, the *measurement profiles*, $\mathcal{P}_i(d)$, of the the densities $\sum_j w_{ij} K_{ij}(\boldsymbol{x}_a)$ and the *posterior profiles*, $\mathcal{P}_{i,\text{post}}(d)$, of the posterior density $p(\boldsymbol{x}_a|\{E_{ai}\}, I)$ were compared. The profiles were calculated by integrating the density that lies in the distance d from the average SAT position, $\langle \boldsymbol{x}_i \rangle$, and reweighting it with the factor $1/d^2$, i.e.

$$\mathcal{P}_{i/i,\text{post}}(d) = \frac{d^{-2} \int d\boldsymbol{x}_a \delta\left((|\boldsymbol{x}_a - \langle \boldsymbol{x}_i \rangle| - d)\,\mathcal{D}(\boldsymbol{x_a})\right)}{\int d\boldsymbol{x}_a \mathcal{D}(\boldsymbol{x_a})}, \tag{5.3.12}$$

where $\mathcal{D}$ is the density to be projected, i.e. $\mathcal{D} = \sum_j w_{ij} K_{ij}(\boldsymbol{x}_a)$ or $\mathcal{D} = p\left(\boldsymbol{x}_a | \{E_{ai}\}, I\right)$. $\delta(.)$ denotes the Dirac delta distribution. The average SAT position, $\langle \boldsymbol{x}_i \rangle$, is defined as

$$\langle \boldsymbol{x}_i \rangle = \int d\boldsymbol{x}_i \ \boldsymbol{x}_i p\left(\boldsymbol{x}_i | I\right). \tag{5.3.13}$$

Profiles, which were calculated relative to the same satellite i, were compared visually. A large overlap of $\mathcal{P}_i(d)$ and $\mathcal{P}_{i,\mathrm{post}}(d)$ suggested that the measurements were consistent, whereas a small overlap suggested inconsistencies.

Position - orientation model with optional docking

In the *position - orientation* model with optional docking, the consistency of measurements was checked by comparing the experimental data with the expected observables FRET efficiency, $\mathcal{E}_{mn}$, and FRET anisotropy, $\mathcal{A}_{mn}$, which were computed from posterior samples. This was done by plotting normalized histograms of the expected observables together with the corresponding marginal likelihoods defined by

$$p\left(E_{mn} | \{\boldsymbol{x}_{ij}\}, \{\boldsymbol{\Omega}_{ij}\}, I\right) = \int d\{E_{kl \neq mn}\} \int d\{A_{kl}\} \ p\left(\{E_{kl}\}, \{A_{kl}\} | \{\boldsymbol{x}_{ij}\}, \{\boldsymbol{\Omega}_{ij}\}, I\right) \tag{5.3.14}$$

and

$$p\left(A_{mn} | \{\boldsymbol{x}_{ij}\}, \{\boldsymbol{\Omega}_{ij}\}, I\right) = \int d\{E_{kl}\} \int d\{A_{kl \neq mn}\} \ p\left(\{E_{kl}\}, \{A_{kl}\} | \{\boldsymbol{x}_{ij}\}, \{\boldsymbol{\Omega}_{ij}\}, I\right). \tag{5.3.15}$$

When the FRET measurements are independent, the expressions above can be simplified, so that only the sums of the likelihood factors (see equations (3.2.4) and (3.2.5)) remain, i.e.

$$p\left(E_{mn} | \{\boldsymbol{x}_{ij}\}, \{\boldsymbol{\Omega}_{ij}\}, I\right) = \frac{1}{S_m S_n} \sum_{s_m s_n} \mathcal{N}_{\Delta E_{mn}}(E_{mn} - \mathcal{E}_{mn}^{(s_m s_n)}) \tag{5.3.16}$$

and

$$p\left(A_{mn} | \{\boldsymbol{x}_{ij}\}, \{\boldsymbol{\Omega}_{ij}\}, I\right) = \frac{1}{S_m S_n} \sum_{s_m s_n} \mathcal{N}_{\Delta A_{mn}}(A_{mn} - \mathcal{A}_{mn}^{(s_m s_n)}). \tag{5.3.17}$$

Here, $\mathcal{N}_\sigma(.)$ denotes the normal distribution with σ standard deviation, which is equal either to the measurement uncertainty ΔE_{mn} and ΔA_{mn} of FRET efficiency and FRET anisotropy, respectively. The index $(s_m s_n)$ indicates the ambiguities of the expected FRET efficiency and FRET anisotropy caused by the ambiguous signs of the average axial depolarizations, whereas S_k denotes, whether there is an ambiguity ($S_k = 2$) or not ($S_k = 1$) (see section 3.2.2).

Like in the previous section, a large overlap of the histograms and the marginal likelihoods suggested consistent measurements, whereas a small overlap suggested inconsistencies or errors in the computation.

5.3.7 Characterization of fluorophore positions and orientations

In all NPS models, average fluorophore positions, $\langle \boldsymbol{x}_i \rangle$, were calculated from the approximate marginal posterior evaluated on a cubic grid of points in space, as described in section 5.3.5. The localization accuracy was computed from the covariance matrix $C_i^{\mathrm{pos}} = \left\langle (\boldsymbol{x}_i - \langle \boldsymbol{x}_i \rangle)(\boldsymbol{x}_i - \langle \boldsymbol{x}_i \rangle)^{\mathrm{T}} \right\rangle$ by diagonalization (see equation (2.4.35)). This resulted 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

summarize the localization accuracy in one number the geometric mean $(\sigma_1^{\boldsymbol{x}_i} \sigma_2^{\boldsymbol{x}_i} \sigma_3^{\boldsymbol{x}_i})^{1/3} = \det(C_i^{\mathrm{pos}})^{1/6}$ was computed, which is equivalent to the 6th root of the determinant of C_i^{pos}.

In the *position - orientation* model, the average fluorophore orientation was calculated directly from posterior samples. In order to deal with the spherical topology correctly, the average transition dipole moment orientation from each sample was represented by two points on the unit sphere. These points, $\{\boldsymbol{s}_l\}$, were the intersections of the unit sphere and the line defined by the average transition dipole moment orientation. By construction, the center of mass of all points $\{\boldsymbol{s}_l\}$ was the origin.

Like in the case of fluorophore positions, the covariance matrix $C_i^{\mathrm{ori}} = \left\langle \boldsymbol{s}_l \boldsymbol{s}_l^{\mathrm{T}} \right\rangle$ was computed and diagonalized, which resulted in three principal directions and variances. The *mean estimated average transition dipole moment orientation* $\langle \boldsymbol{\Omega}_i \rangle$ was defined as the principal direction with the largest variance, and the *orientation accuracy* as the standard deviations $\sigma_1^{\boldsymbol{\Omega}_i}$ and $\sigma_2^{\boldsymbol{\Omega}_i}$ in the remaining principal directions. In order to summarize the orientation accuracy in one number, the geometric mean $(\sigma_1^{\boldsymbol{\Omega}_i} \sigma_2^{\boldsymbol{\Omega}_i})^{1/2}$ was computed.

5.4 Computation of the marginal antenna position posterior in the position - Förster distance model

In this section, the computation of the the marginal posterior of the antenna fluorophore position in the *position - Förster distance* model (Muschielok et al., 2008, supplementary material) is described in detail. The variables used here were explained in section 3.1.

All integrals were computed either by hand or using the computer algebra software Maple (Maplesoft, Waterloo Maple Inc). Integrals that could not be computed in a closed form were integrated numerically using adaptive quadrature that is implemented in the MATLAB function `quadgk` (Gander and Gautschi, 2000).

As stated in section 3.1, the marginal ANT position posterior was written as

$$p(\boldsymbol{x}_a | \{E_{ai}\}, I) \propto p(\boldsymbol{x}_a | I) \prod_i \sum_j w_{ij} K_{ij}(\boldsymbol{x}_a), \qquad (5.4.1)$$

with the functions $K_{ij}(\boldsymbol{x}_a)$ defined by

$$K_{ij}(\boldsymbol{x}_a) = \int d\boldsymbol{x}_i \int dR_{ai} \; L_i(|\boldsymbol{x}_a - \boldsymbol{x}_i| / R_{ai}) S_{ij}(\boldsymbol{x}_i) \, p(R_{ai} | I), \qquad (5.4.2)$$

where $L_i(|\boldsymbol{x}_a - \boldsymbol{x}_i| / R_{ai})$ is the likelihood factor, and the other two terms originate from the prior. The likelihood factor is given by

$$L_i\left(|\boldsymbol{x}_a - \boldsymbol{x}_i| / R_{ai}\right) = \frac{\exp\left\{ - \left[\mathcal{E}_{ai}\left(\boldsymbol{x}_a, \boldsymbol{x}_i, R_{ai}\right) - E_{ai}\right]^2 / \left[2\left(\Delta E_{ai}\right)^2\right] \right\}}{\Delta E_{ai}\sqrt{2\pi}}. \qquad (5.4.3)$$

The functions $S_{ij}(\boldsymbol{x}_i)$ in equation (5.4.2) are Gaussian kernels that approximate the position prior of the ith satellite, $p(\boldsymbol{x}_i | I)$, as follows,

$$p(\boldsymbol{x}_i | I) \approx \sum_j w_{ij} S_{ij}(\boldsymbol{x}_i), \qquad (5.4.4)$$

$$S_{ij}(\boldsymbol{x}_i) = \left(\sqrt{2\pi}\Delta s_{ij}\right)^{-3} \exp\left(-\frac{\left|\boldsymbol{x}_i - \langle\boldsymbol{s}\rangle_{ij}\right|^2}{2\left(\Delta s_{ij}\right)^2} \right). \qquad (5.4.5)$$

The centers $\langle \boldsymbol{s} \rangle_{ij}$, widths Δs_{ij} and weights w_{ij} were obtained from a least-square fit of the position prior to the normalized accessible volume V_i (see sections 3.1.3 and 5.3.1). To save computation time in the last calculation step later on, all widths were kept at an equal value during the fit, i.e. $\Delta s_{ij} = \Delta s$.

The Förster distance prior is approximated similarly with the functions $Q_{ik}(R_{ai})$,

$$p(R_{ai}|I) \approx \sum_k v_{ik} Q_{ik}(R_{ai}),$$

$$Q_{ik}(R_{ai}) = \left[\sqrt{2\pi} \Delta \tilde{q}_{ik} \operatorname{erf}\left(\frac{\tilde{q}_{ik}}{\sqrt{2}\Delta \tilde{q}_{ik}} \right) \right]^{-1} \left\{ \exp\left[-\frac{(R_{ai} - \tilde{q}_{ik})^2}{2(\Delta \tilde{q}_{ik})^2} \right] - \exp\left[-\frac{(R_{ai} + \tilde{q}_{ik})^2}{2(\Delta \tilde{q}_{ik})^2} \right] \right\},$$
$$(5.4.6)$$

where $\tilde{q}_{ik}$ and $\Delta \tilde{q}_{ik}$ are parameters related to the center and width of the pseudo-Gaussians $Q_{ik}(R_{ai})$. $\tilde{q}_{ik}$ and $\Delta \tilde{q}_{ik}$ were obtained from a least square fit to the simulated normalized histograms of Förster distances (see section 5.3.2).

In the following paragraphs, the computation of the four-dimensional (three dimensions for $\boldsymbol{x}_i$ and one for R_{ai}) integrals in equation (5.4.2) will be discussed in detail. Maple was used to integrate over three out of 4 dimensions analytically. To this end, spherical coordinates centered at $\boldsymbol{x}_a = (x_a^x, x_a^y, x_a^z)$ were used to substitute the satellite position $\boldsymbol{x}_i = (x_i^x, x_i^y, x_i^z)$, i.e. $(x_i^x - x_a^x, x_i^y - x_a^y, x_i^z - x_a^z) \to (r_i, \theta_i, \phi_i)$. The integration over the angular variables yielded the following expression:

$$K_{ij}(\boldsymbol{x}_a) = \sum_k v_{ik} \int_0^\infty \int_0^\infty L_i(r_i/R_{ai}) \frac{r_i}{\sqrt{2\pi}\Delta s_{ij} d_{ij}}$$
$$\left[\exp\left(-\frac{(r_i - d_{ij})^2}{2(\Delta s_{ij})^2} \right) - \exp\left(-\frac{(r_i + d_{ij})^2}{2(\Delta s_{ij})^2} \right) \right] Q_{ik}(R_{ai}) dR_{ai} dr_i \qquad (5.4.7)$$

where $d_{ij} = \left| \boldsymbol{x}_a - \langle \boldsymbol{s} \rangle_{ij} \right|$. Since $K_{ij}(\boldsymbol{x}_a)$ depends not directly on $\boldsymbol{x}_a$ but on the distance d_{ij} only, it possesses spherical symmetry around the point $\langle \boldsymbol{s} \rangle_{ij}$. Consequently, $K_{ij}(\boldsymbol{x})$ was calculated by evaluating its radial profile $F_{ij}(d_{ij})$,

$$K_{ij}(\boldsymbol{x}_a) = F_{ij}(d_{ij}) = \sum_k v_{ik} F_{ijk}(d_{ij}), \qquad (5.4.8)$$

where $F_{ijk}(d_{ij})$ is the contribution of the kth kernel of the Förster distance prior to the radial profile $F_{ij}(d_{ij})$.

In order to calculate $F_{ijk}(d_{ij})$, equation (5.4.6) was substituted into equation (5.4.7). The integration variables were transformed, $(r_i, R_{ai}) \to (r_i, \rho_i) = (r_i, R_i/r_i)$, and integration over r_i yielded the lengthy expression

$$F_{ijk}(d_{ij}) = \left[\sqrt{2\pi} d_{ij} \operatorname{erf}\left(\tilde{q}_{ik}/\sqrt{2}\Delta \tilde{q}_{ik} \right) \right]^{-1}$$
$$\int_0^\infty \frac{L_i(1/\rho_i)}{\left[(\Delta \tilde{q}_{ik})^2 + \rho_i^2 (\Delta s_{ij})^2 \right]^{5/2}} \left\{ \left[\left(\tilde{q}_{ik}^2 + (\Delta \tilde{q}_{ik})^2 \right) (\Delta s_{ij})^4 \rho_i^2 + \right. \right.$$
$$\left. \left(d_{ij}^2 + (\Delta s_{ij})^2 \right) (\Delta \tilde{q}_{ik})^4 \right] X_- + 2 \left[d_{ij} (\Delta \tilde{q}_{ik})^2 \tilde{q}_{ik} \rho_i (\Delta s_{ij})^2 \right] X_+ \right\} d\rho_i, \quad (5.4.9)$$

where

$$X_\pm = \exp\left[-\frac{\frac{1}{2}(d_{ij}\rho_i - \tilde{q}_{ik})^2}{(\Delta\tilde{q}_{ik})^2 + \rho_i^2(\Delta s_{ij})^2}\right] \pm \exp\left[-\frac{\frac{1}{2}(d_{ij}\rho_i + \tilde{q}_{ik})^2}{(\Delta\tilde{q}_{ik})^2 + \rho_i^2(\Delta s_{ij})^2}\right] \tag{5.4.10}$$

The radial profiles $F_{ijk}(d_{ij})$ were evaluated in several hundred supporting points on the d_{ij}-axis by solving the integral in equation (5.4.9) numerically with adaptive quadrature.

In the next step, the radial profile $F_{ij}(d_{ij})$ of the fuzzy sphere (equation (5.4.8)) was calculated. The tree-dimensional fuzzy spheres, $K_{ij}(\boldsymbol{x}_a)$, were evaluated on a cubic lattice (either 1 Å or 2 Å spacing) by interpolation of their radial profiles, $F_{ij}(d_{ij})$.

Finally, the fuzzy spheres that contributed to the same FRET measurement, i.e. $K_{ij}(\boldsymbol{x}_a)$ for a fixed i, were weighted, added and multiplied with the antenna fluorophore position prior (equation (5.4.1)). In this way, the not yet normalized antenna position posterior was calculated. The evidence, Z, was computed approximately by trapezoidal integration of the sampled density with the MATLAB function `trapz`, so that the marginal antenna position posterior was given by

$$p(\boldsymbol{x}_a|\{E_{ai}\}, I) = Z^{-1}p(\boldsymbol{x}_a|I)\prod_i\sum_j w_{ij}K_{ij}(\boldsymbol{x}_a). \tag{5.4.11}$$

5.5 Implementation and development of a sampling engine for Bayesian data analysis

In contrast to the previous section, the complexity of the parameter estimation problem in the *position - orientation* NPS model makes it necessary to use a fully numerical approach.

To this sake, two sampling algorithms were implemented in MATLAB (The MathWorks) and C: The nested sampling algorithm (section 2.5.2) based on Markov-chain Monte Carlo was used to search high probability regions of the posterior in the multi-dimensional parameter space, to compute the evidence and a small number[1] of posterior samples. More posterior samples were drawn by using the Metropolis Monte Carlo algorithm (section 2.5.1). The implementation of the algorithms, called *sampling engine*, was kept general in order to solve arbitrary finite-dimensional parameter estimation problems with a continuous parameter space. The sampling engine was adapted to the NPS parameter estimation by implementing a custom prior and likelihood.

In this section the requirements of the sampling engine will be discussed, followed by a description of the implementation and the adaptation to the NPS analysis.

5.5.1 Requirements

In the case of NPS, which was here a moderate-dimensional (≈ 50 dimensional) parameter estimation problem, the posterior has a complicated shape because of strong correlations between fluorophore positions and average transition dipole orientations caused by the likelihood. Probably, the posterior does not even possess a unique global maximum since, in most cases, the number of parameters exceeds the number of data (see section 6.7.4 for a detailed discussion). In addition, the posterior probability values of the samples are usually spread over many orders of magnitude and must be handled in a way that keeps rounding errors small.

---

[1]The number of posterior samples computed by reweighting samples obtained by nested sampling was small (here: 5 to 20 times less) compared to the number of samples from nested sampling (here: $25 \cdot 10^3$ to $170 \cdot 10^3$).

When MCMC is used to compute samples, the Markov chain must be able to maneuver efficiently in the parameter space constrained by either the posterior directly, as in the case of the Metropolis algorithm, or by a likelihood constraint, as in nested sampling. Since the likelihood constraint changes gradually, the step sizes of local moves must be adapted as well. In order to ensure ergodicity and to produce independent samples, the chain must be also able to move long distances between high probability regions, which was achieved by a non-local move.

Since the prior is flat but has at the same time a rather difficult form caused by the complicated shape of the volumes accessible to the fluorophores, the Markov chain must check frequently whether it stays within the prior. To this end an efficient access to the prior information was required.

In the following, the most important details of the implementation will be discussed.

5.5.2 Working with small numbers

When the Markov chain samples parameter space regions with strongly varying likelihood values, care must be taken in order to avoid rounding errors due to the finite representation of numbers on a computer. To this end, instead of using directly the likelihood values, their logarithms were utilized in the implementation of the sampling engine. This common approach (Sivia, 2006, chapter 9) increases the dynamic range at cost of precision. When two numbers a and b were added, rounding errors were kept small by adding the logarithms of a and b (Sivia, 2006, chapter 9),

$$\ln(a+b) = \begin{cases} \ln(a) + \ln\left\{1 + \exp\left[\ln(b) - \ln(a)\right]\right\} & a > b \\ \ln(b) + \ln\left\{\exp\left[\ln(a) - \ln(b)\right] + 1\right\} & a \leq b. \end{cases} \tag{5.5.1}$$

5.5.3 Implementation of prior information

Let $\{\theta_i\} = (\theta_1, \theta_2, \ldots, \theta_N)$ be the vector of model parameters in the $N-$ dimensional parameter space. The prior is assumed to factorize into M independent priors, each factor covering a subspace P_k of the parameter space,

$$p\left(\{\theta_i\}|I\right) = \prod_k p\left(\{\theta_{i \in P_k}\}|I\right). \tag{5.5.2}$$

Now, each factor in the equation above is a sum of non-overlapping boxes B_j with uniform probability density p_j,

$$p\left(\{\theta_{i \in P_k}\}|I\right) = \begin{cases} p_j & \{\theta_{i \in P_k}\} \in B_j \\ 0 & \text{otherwise.} \end{cases} \tag{5.5.3}$$

Each box j is defined by its lower and upper corner, $\{\theta_{i \in P}^{\text{L},j}\}$ and $\{\theta_{i \in P}^{\text{U},j}\}$, respectively. The prior factors are called *box collection priors* (BCPs), and a schematic example is shown in figure 5.6. In this way, non-uniform priors can be described by assigning different probability densities p_j to each box. Though this can be used in future, it was not needed in the calculations performed in this work, as all priors were flat.

For fast access to the prior information, i.e. for the calculation of $p\left(\{\theta_i\}|I\right)$, one needs to determine the boxes (one in each subspace) that contain the point in parameter space. Since most of the moves are expected to be local, it is checked first whether $\{\theta_{i \in P_k}\}$ is located in the boxes that contained its last position.

When the box did not contain the point, a search was performed that used lists of

Figure 5.6: Schematic sketch of a box collection prior. Four boxes ($B_1 \ldots B_4$, solid lines) are shown with different probability densities indicated as different gray scales. The total upper ($\theta_i^{\mathrm{U,tot}}$) and lower ($\theta_i^{\mathrm{L,tot}}$) limits of the collection of boxes are shown as dashed lines.
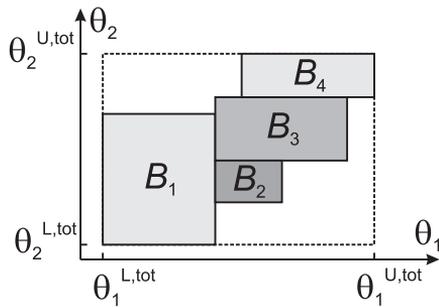
boxes sorted according to their lower and upper bounds. The sorted lists allow a fast determination of boxes that could potentially contain the point, and intersection of the sets of possible candidates yields the box that contains $\{\theta_{i \in P_k}\}$. When the resulting set is empty the point is "outside" of the prior, i.e. $p\left(\{\theta_{i \in P_k}\}|I\right) = 0$.

5.5.4 Markov chain Monte Carlo

The MCMC comprises of two sorts of moves, a local and a non-local move. In order to maintain detailed balance, at each step of the random walk it is determined at random with probability p_{NL} whether the local or the global move is used. In the calculations, the probability to choose the non-local move was set much smaller than the probability to choose the local move, i.e. $p_{\mathrm{NL}} \ll 1$, since the non-local move is computationally expensive.

Local Monte-Carlo move

The move that ensures local mobility of the Markov chain alters the current position of the Markov chain in all parameter space dimensions simultaneously. More precisely, the move from the position $\{\theta_i\}_t$ at time t to the position $\{\theta_i\}_{t+1}$ at the next time increment is done by adding or subtracting exponentially distributed random numbers to each component of the parameter vector, so that the probability to propose $\{\theta_i\}_{t+1}$ as the next position of the chain is

$$p_{\mathrm{prop}}(\{\theta_i\}_{t+1}|\{\theta_i\}_t) = \prod_i \frac{1}{2s_i} \exp\left(-\frac{|\theta_{i,t+1} - \theta_{i,t}|}{s_i}\right). \qquad (5.5.4)$$

s_i is the length scale of the move in the parameter space dimension i. Figure 5.7 shows a two-dimensional example. The move in equation (5.5.4) is similar to the physical diffusion of a particle since a random walk is performed, but instead of normal distributed jumps the computationally cheaper exponential distribution is used.

The proposed position $\{\theta_i\}_{t+1}$ is then accepted (A) or rejected ($\overline{A}$) in order to maintain detailed balance. The acceptance probability differs depending on whether samples from the posterior (during Metropolis sampling) or from the constrained prior (during nested sampling) are drawn. Checking of the acceptance is done in two consecutive steps by applying the acceptance criterion of Metropolis to the prior and to the likelihood separately. In the first step, the position can be hence accepted (A_π) or rejected ($\overline{A}_\pi$) due to the
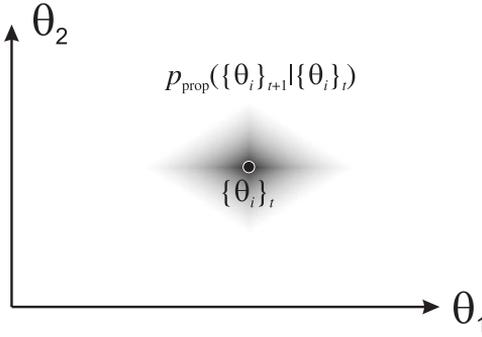
Figure 5.7: Local move proposal. Projection of the probability density function $p_{\text{prop}}(\{\theta_i\}_{t+1}|\{\theta_i\}_t)$ onto two parameter space dimensions, θ_1 and θ_2. The gray scale is proportional to the probability density of the position $\{\theta_i\}_{t+1}$ proposed during a local move from the current position $\{\theta_i\}_t$ (black dot in the center).

prior $\pi(\{\theta_i\}) := p(\{\theta_i\}|I)$,

$$p(A_\pi|\{\theta_i\}_{t+1}, \{\theta_i\}_t) = \begin{cases} 1 & \text{if } \pi(\{\theta_i\}_{t+1}) \geq \pi(\{\theta_i\}_t) \\ \pi(\{\theta_i\}_{t+1})/\pi(\{\theta_i\}_t) & \text{otherwise.} \end{cases} \qquad (5.5.5)$$

This ensures that the chain stays within the parameter space occupied by the prior and would yield prior samples when applied alone. In the second step, if the proposed position was accepted due to the prior, the likelihood L is evaluated at $\{\theta_i\}_{t+1}$. Again, the position can then be accepted (A_L) or rejected ($\overline{A}_L$), but now due to the likelihood. In the case of sampling the posterior, the new position is accepted with the probability

$$p(A_L|\{\theta_i\}_{t+1}, \{\theta_i\}_t) = \begin{cases} 1 & \text{if } L(\{\theta_i\}_{t+1}) \geq L(\{\theta_i\}_t) \\ L(\{\theta_i\}_{t+1})/L(\{\theta_i\}_t) & \text{otherwise.} \end{cases} \qquad (5.5.6)$$

In the case of nested sampling, the prior constrained by $L(\{\theta_i\}) > L^*$ is sampled, which is done by the following acceptance criterion:

$$p(A_L|\{\theta_i\}_{t+1}, \{\theta_i\}_t) = \begin{cases} 1 & \text{if } L(\{\theta_i\}_{t+1}) > L^* \\ 0 & \text{otherwise.} \end{cases} \qquad (5.5.7)$$

Finally, the "total" acceptance probability is given by

$$p(A|\{\theta_i\}_{t+1}, \{\theta_i\}_t) = p(A_\pi|\{\theta_i\}_{t+1}, \{\theta_i\}_t)\, p(A_L|\{\theta_i\}_{t+1}, \{\theta_i\}_t) \qquad (5.5.8)$$

and the transition probability, i.e. the probability to move from $\{\theta_i\}_t$ to $\{\theta_i\}_{t+1}$ is

$$p(\{\theta_i\}_{t+1}|\{\theta_i\}_t) = p_{\text{prop}}(\{\theta_i\}_{t+1}|\{\theta_i\}_t)\, p(A|\{\theta_i\}_{t+1}, \{\theta_i\}_t). \qquad (5.5.9)$$

It is clear, that when s_i is constant during the random walk the move maintains detailed balance since $p_{\text{prop}}(\{\theta_i\}_{t+1}|\{\theta_i\}_t) = p_{\text{prop}}(\{\theta_i\}_t|\{\theta_i\}_{t+1})$, and in the case of posterior sampling the ratio of transition probabilities in between two positions is

$$\frac{p(\{\theta_i\}_{t+1}|\{\theta_i\}_t)}{p(\{\theta_i\}_t|\{\theta_i\}_{t+1})} = \frac{\pi(\{\theta_i\}_{t+1})L(\{\theta_i\}_{t+1})}{\pi(\{\theta_i\}_t)L(\{\theta_i\}_t)}, \qquad (5.5.10)$$

which proves that the local move generates samples proportional to the posterior. In the case of nested sampling the ratio of transition probabilities is

$$\frac{p(\{\theta_i\}_{t+1}|\{\theta_i\}_t)}{p(\{\theta_i\}_t|\{\theta_i\}_{t+1})} = \frac{\pi(\{\theta_i\}_{t+1})}{\pi(\{\theta_i\}_t)}, \quad \text{as long as } L(\{\theta_i\}_t), L(\{\theta_i\}_{t+1}) < L^*, \qquad (5.5.11)$$

i.e. the local move generates samples from the constrained prior.

Often during the random walk the proposed position will be in a region where the prior is zero, which results in a rejection and reduces the efficiency of the sampling algorithm. This behavior is especially problematic in high dimensional parameter spaces since the probability that the move was too large in at least one parameter space direction rises with the dimensionality. One can, in this case, try to "recycle" the random numbers as proposed by Sivia (2006, chapter 9). To this end, different kind of "boundary conditions" were applied to the initially proposed position, $\{\theta_i\}_{t+1}$, before the acceptance was checked. The objective was to re-map positions outside of the box that encloses all prior probability into the interior of the box. As long as the re-mapping is symmetrical[2], this procedure maintains detailed balance and is especially effective when the prior consists of only one box.

The enclosing box is defined by

$$\theta_i^{\mathrm{L,tot}} < \theta_i < \theta_i^{\mathrm{U,tot}} \tag{5.5.12}$$

$$\theta_i^{\mathrm{L,tot}} = \min\{\theta_i | \pi(\{\theta_i\}) > 0\} \tag{5.5.13}$$

$$\theta_i^{\mathrm{U,tot}} = \max\{\theta_i | \pi(\{\theta_i\}) > 0\}, \tag{5.5.14}$$

where $\theta_i^{\mathrm{L,tot}}$ and $\theta_i^{\mathrm{U,tot}}$ are the lower and upper limits of the prior, respectively. The size of the enclosing box in each parameter space dimension is, Δ_i, is given by

$$\Delta_i = \theta_i^{\mathrm{U,tot}} - \theta_i^{\mathrm{L,tot}}. \tag{5.5.15}$$

The most simple boundary conditions are the reflecting

$$\theta_i \leftarrow \begin{cases} \theta_i^{\mathrm{U,tot}} - \left(\theta_i - \theta_i^{\mathrm{U,tot}}\right) \mod \Delta_i & \text{if } \theta_i > \theta_i^{\mathrm{U,tot}} \\ \theta_i^{\mathrm{L,tot}} + \left(\theta_i^{\mathrm{L,tot}} - \theta_i\right) \mod \Delta_i & \text{if } \theta_i < \theta_i^{\mathrm{L,tot}} \\ \theta_i & \text{otherwise} \end{cases} \tag{5.5.16}$$

and the periodic boundary condition

$$\theta_i \leftarrow \begin{cases} \theta_i^{\mathrm{L,tot}} + \left(\theta_i - \theta_i^{\mathrm{L,tot}}\right) \mod \Delta_i & \text{if } \theta_i > \theta_i^{\mathrm{U,tot}} \\ \theta_i^{\mathrm{U,tot}} - \left(\theta_i^{\mathrm{U,tot}} - \theta_i\right) \mod \Delta_i & \text{if } \theta_i < \theta_i^{\mathrm{L,tot}} \\ \theta_i & \text{otherwise.} \end{cases} \tag{5.5.17}$$

The latter is useful when the parameter space bends back onto itself and the likelihood and prior are periodic (see figure 5.8a). In this way a maximum close to the limits $\theta_i^{\mathrm{U,tot}}$ and $\theta_i^{\mathrm{L,tot}}$ is only apparently separated as the chain can easily transit in between the ends of the box.

Also more complicated boundary conditions can be used that mirror the parameter space coordinate i when the coordinate j was re-mapped by periodic or reflecting boundary conditions, i.e.

$$\theta_i \leftarrow \begin{cases} \theta_i^{\mathrm{L,tot}} + \theta_i^{\mathrm{U,tot}} - \theta_i & \text{if } \theta_j \text{ was remapped} \\ \theta_i & \text{otherwise.} \end{cases} \tag{5.5.18}$$

This boundary condition should be used in a Möbius tape or Klein bottle topology (figure 5.8b).

---

[2]If a move from the initial position $\{\theta_i\}_t$ to the proposed position $\{\theta_i\}_{t+1}$ is re-mapped to $\{\theta_i\}'_{t+1}$, the reverse move from $\{\theta_i\}'_{t+1}$ to $\{\theta_i\}'_{t+1} - \{\theta_i\}_{t+1} + \{\theta_i\}_t$ must be re-mapped to $\{\theta_i\}_t$.
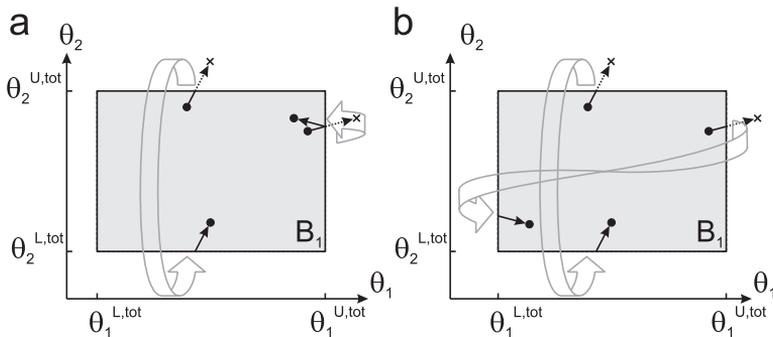
Figure 5.8: Boundary conditions examples. Local moves (black arrows) from position $\{\theta_i\}_t$ (dot) to an invalid position $\{\theta_i\}_{t+1}$ (cross) outside the box containing all the prior probability are re-mapped (gray arrows) in order to lie inside the box. This defines different topologies, and two examples are shown. The parameter space in (a) has the topology of a ring: reflective boundary conditions in coordinate 1 and periodic boundary conditions in coordinate 2. In (b) it has the topology of a Klein bottle: periodic boundary conditions in coordinate 1 and 2, but coordinate 2 is also mirrored after re-mapping coordinate 1. The box collection prior consists here only of one box, B_1, hence the total extension of the prior given by the lower and upper limits, is identical to B_1.

Non-local Monte-Carlo move

The non-local move was designed to enable the Markov chain to jump in between the neighborhoods of two of the N equilibrated samples, called reference points. The underlying idea is, that in the proximity of the reference points there is parameter space with sufficiently high probability density, and this, in turn, allows to move the Markov chain from the current position $\{\theta_i\}_t$ near the "source" reference position, $\{r_i\}_s$, to some other position $\{\theta_i\}_{t+1}$ close to the "destination" reference position, $\{r_i\}_d$. In the case of nested sampling, the reference points were the set of active objects, which intrinsically satisfy the likelihood constraint. In the case of posterior sampling, the reference points were a small set of posterior samples. Here, the equally weighted samples of a nested sampling run were used for this purpose.

The move comprises of five steps (see figure 5.9). In the first step, the source reference point, $\{r_i\}_s$, is chosen from the set of reference points $\{r_i\}_k$ with the probability

$$p_{\text{ref}}(\{r_i\}_s | \{\theta_i\}_t) = \frac{\exp\left(-\frac{1}{2} \|\{r_i\}_s - \{\theta_i\}_t\|^2_{\{s_i\}}\right)}{\sum_k \exp\left(-\frac{1}{2} \|\{r_i\}_k - \{\theta_i\}_t\|^2_{\{s_i\}}\right)}. \tag{5.5.19}$$

Here $\|.\|_{\{s_i\}}$ denotes the Euclidean distance that is scaled by the length scales of the local move,

$$\|\{\theta_i\}\|_{\{s_i\}} := \left(\sum_i |\theta_i/s_i|^2\right)^{1/2}. \tag{5.5.20}$$

It is therefore most probable to choose the reference point, which has the smallest normalized distance to the current position.

In the second step, the destination reference point, $\{r_i\}_d$, is chosen with the probability
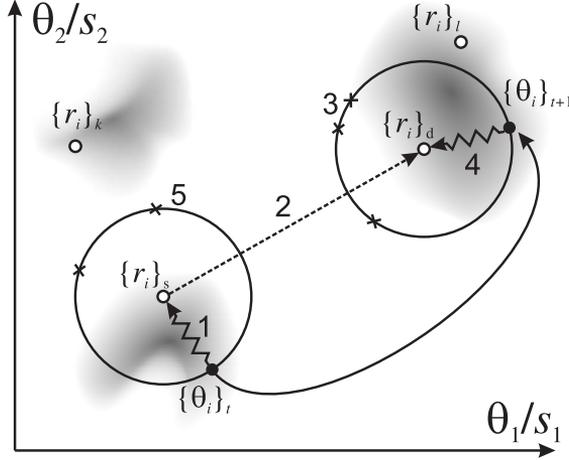
Figure 5.9: A non-local move from $\{\theta_i\}_t$ to $\{\theta_i\}_{t+1}$ (bent solid arrow) comprises of five steps. (1) Choose from the set of equilibrated samples a "source" reference point $\{r_i\}_s$ that is in proximity to the current position $\{\theta_i\}_t$ (zig-zag arrow). (2) Choose from the remaining equilibrated samples a "destination" reference point $\{r_i\}_d$ (dashed arrow). (3) Try at maximum M times to find an accepted position on the hyper-sphere (circle) around the destination reference point. Not accepted positions are shown as crosses, the accepted position as black a dot. Reject the whole move if no accepted position was found. (4) Verify that $\{r_i\}_d$ is in proximity of the accepted position $\{\theta_i\}_{t+1}$ (zig-zag arrow), reject the move if it is not. (5) Choose M_{rej}, a uniformly distributed integer from 0 to $M-1$, and satisfy reversibility of the non-local move by forcing exactly M_{rej} successive rejections (crosses) from proposed positions on a hyper-sphere (circle) around the source reference point. If M_{rej} successive rejections cannot be produced reject the whole move.

density

$$p_{\mathrm{dst}}(\{r_i\}_d|\{r_i\}_s) = \frac{1}{N-1}\sum_{k\neq s}\delta(\{r_i\}_d - \{r_i\}_k), \qquad (5.5.21)$$

where $\delta(.)$ is the Dirac-delta distribution. The probability density in equation (5.5.21) is simply generated by choosing the coordinates of one of the remaining reference points.

In the third step, the parameter space normalized by s_i is probed on the hypersphere with center $\{r_i\}_d$ and radius $\|\{\theta_i\}_t - \{r_i\}_s\|_{\{s_i\}}$ in order to find an accepted position $\{\theta_i\}_{t+1}$. To this end, random numbers $\{u_i\}^{\mathrm{sphere}}$ are generated that are uniformly distributed on a hypersphere with radius 1 using the algorithm described by Muller (1956, 1959). The proposed position is set to $\theta_{i,t+1} = r_{i,t} + s_i u_i^{\mathrm{sphere}} \cdot \|\{\theta_i\}_t - \{r_i\}_s\|_{\{s_i\}}$, and therefore

$$p_{\mathrm{prop}}(\{\theta_i\}_{t+1}|\{r_i\}_d, \{r_i\}_s, \{\theta_i\}_t) = \begin{cases} const. & \text{if } \|\{r_i\}_d - \{\theta_i\}_{t+1}\| = \|\{r_i\}_s - \{\theta_i\}_t\| \\ 0 & \text{otherwise} \end{cases}$$

$$(5.5.22)$$

The acceptance probability is the same as in equation (5.5.8). The probing is repeated until an accepted position is found, but not more than M times in total. The whole move is rejected when no accepted position is found, so that the probability density to find a

valid position $\{\theta_i\}_{t+1}$ is equal to

$$p\left(\{\theta_i\}_{t+1}, A | \{r_i\}_\mathrm{d}, \{r_i\}_\mathrm{s}, \{\theta_i\}_t\right) =$$

$$p_\mathrm{prop}(\{\theta_i\}_{t+1} | \{r_i\}_\mathrm{d}, \{r_i\}_\mathrm{s}, \{\theta_i\}_t) p\left(A | \{\theta_i\}_{t+1}, \{\theta_i\}_t\right) \sum_{n=0}^{M-1} \mathcal{R}^n(\{r_i\}_\mathrm{d}, \{r_i\}_\mathrm{s}), \quad (5.5.23)$$

where $\mathcal{R}(\{r_i\}_\mathrm{d}, \{r_i\}_\mathrm{s})$ denotes the average probability to reject a proposed position,

$$\mathcal{R}(\{r_i\}_\mathrm{d}, \{r_i\}_\mathrm{s}) = \int d\{\theta_i\}'\, p_\mathrm{prop}(\{\theta_i\}' | \{r_i\}_\mathrm{d}, \{r_i\}_\mathrm{s}, \{\theta_i\}_t) \left(1 - p\left(A | \{\theta_i\}', \{\theta_i\}_t\right)\right),$$

(5.5.24)

and where the nth powers of the average rejection probability are denoted by $\mathcal{R}^n(\{r_i\}_\mathrm{d}, \{r_i\}_\mathrm{s}) \equiv [\mathcal{R}(\{r_i\}_\mathrm{d}, \{r_i\}_\mathrm{s})]^n$.

Now, the fourth and fifth step are needed to maintain detailed balance that would be severely violated by the repeated probing and by the choice of the source reference position. In the fourth step, it is verified whether $\{\theta_i\}_{t+1}$ is in the proximity of the destination reference point by accepting the move with probability $p_\mathrm{ref,acc}$, which is given as

$$p_\mathrm{ref,acc} = p_\mathrm{ref}(\{r_i\}_\mathrm{d} | \{\theta_i\}_{t+1}). \quad (5.5.25)$$

In the fifth step, a term is constructed that acts as a "counterbalance" of the sum in equation (5.5.23). This is done by forcing M_rej successive rejections during the probing of the parameter space on a hypersphere in analogy to step three, but centered at the source reference position instead. M_rej is a uniformly distributed integer from 0 to $M - 1$, each number being chosen with the probability $1/M$. Hence, the probability to pass step five is

$$p_\mathrm{pass}(\{r_i\}_\mathrm{d}, \{r_i\}_\mathrm{s}, \{\theta_i\}_{t+1}) = \frac{1}{M} \sum_{n=0}^{M-1} \mathcal{R}^n(\{r_i\}_\mathrm{s}, \{r_i\}_\mathrm{d}), \quad (5.5.26)$$

where $\mathcal{R}(\{r_i\}_\mathrm{s}, \{r_i\}_\mathrm{d})$ is the average rejection probability at the source reference point,

$$\mathcal{R}(\{r_i\}_\mathrm{s}, \{r_i\}_\mathrm{d}) = \int d\{\theta_i\}'\, p_\mathrm{prop}(\{\theta_i\}' | \{r_i\}_\mathrm{s}, \{r_i\}_\mathrm{d}, \{\theta_i\}_{t+1}) \left(1 - p\left(A | \{\theta_i\}', \{\theta_i\}_{t+1}\right)\right).$$

(5.5.27)

Finally, the probability to accept the complete non-local move is the product of the probabilities in the equations (5.5.19), (5.5.21), (5.5.23), (5.5.25) and (5.5.26) marginalized over source and destination reference positions,

$$p\left(\{\theta_i\}_{t+1} | \{\theta_i\}_t\right) = p\left(A | \{\theta_i\}_{t+1}, \{\theta_i\}_t\right) \times$$

$$\sum_{k,l,k\neq l} p_\mathrm{ref}(\{r_i\}_k | \{\theta_i\}_t) \frac{1}{N-1} p_\mathrm{ref}(\{r_i\}_l | \{\theta_i\}_{t+1}) \times$$

$$p_\mathrm{prop}(\{\theta_i\}_{t+1} | \{r_i\}_l, \{r_i\}_k, \{\theta_i\}_t) \times$$

$$\left[\sum_{n=0}^{M-1} \mathcal{R}^n(\{r_i\}_l, \{r_i\}_k)\right] \frac{1}{M} \left[\sum_{n=0}^{M-1} \mathcal{R}^n(\{r_i\}_k, \{r_i\}_l)\right]. \quad (5.5.28)$$

The acceptance probability was taken out of the sum over k and l since it does not depend on the reference points. The sum over k and l is invariant under exchange of $\{\theta_i\}_{t+1}$ and $\{\theta_i\}_t$ because of two facts: First, k and l can be exchanged due to the summation over all reference point pairs. Second, each line in equation (5.5.28), including the proposal
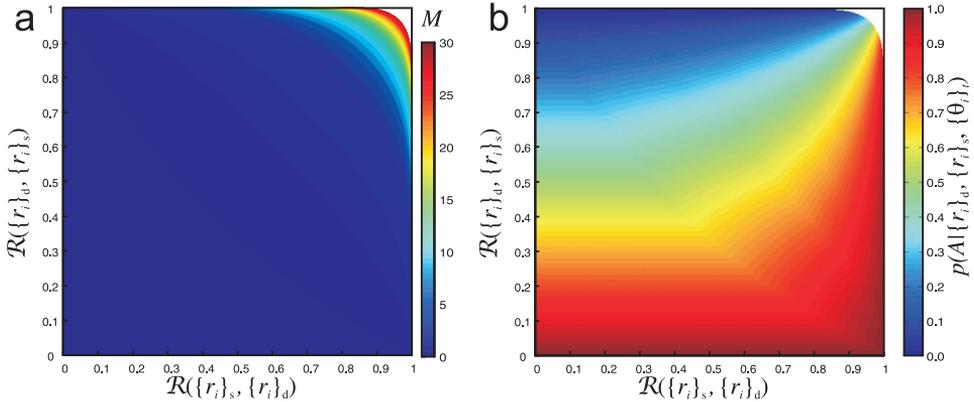
Figure 5.10: (a) The number of probing retries, M, that maximizes the probability of the non-local move given source and destination reference points. The optimal M is plotted as a function of the average rejection probabilities at source and destination, $\mathcal{R}(\{r_i\}_s, \{r_i\}_d)$ and $\mathcal{R}(\{r_i\}_d, \{r_i\}_s)$. The higher the average rejection probabilities the higher is the optimal M. Values up to $M = 30$ were computed. In the white area in the plot the optimal $M > 30$.
(b) Probability to accept the non-local move in average for an optimal number of probing retries M as function of $\mathcal{R}(\{r_i\}_s, \{r_i\}_d)$ and $\mathcal{R}(\{r_i\}_d, \{r_i\}_s)$.

probability in particular, is invariant under simultaneous exchange of $\{\theta_i\}_{t+1} \leftrightarrow \{\theta_i\}_t$ and $k \leftrightarrow l$ (note that the average rejection probabilities depend on $\{\theta_i\}_t$ and $\{\theta_i\}_{t+1}$).

When the ratio $p\left(\{\theta_i\}_{t+1}|\{\theta_i\}_t\right)/p\left(\{\theta_i\}_t|\{\theta_i\}_{t+1}\right)$ is computed, the result is the same as in equations (5.5.10) and (5.5.11) since everything besides the acceptance probabilities drops out, and finally detailed balance is maintained by the non-local move.

In total, the probability to perform a non-local move is increased by the repeated probing, which is especially useful when the average rejection probabilities are close to 1. Nevertheless, one has to select a suitable number of retries, M, because too small M might not work well when the average rejection probabilities are high, and vice versa, which can be concluded from the steps three and five, respectively. It is therefore of interest to see how M influences the maximum probability to perform a non-local move on average. The average acceptance probability for given reference points is given by

$$P_{\langle\mathrm{acc}\rangle}(\{r_i\}_d, \{r_i\}_s) = \int d\{\theta_i\}' \, p_{\mathrm{prop}}(\{\theta_i\}'|\{r_i\}_d, \{r_i\}_s, \{\theta_i\}_t) p\left(A|\{\theta_i\}', \{\theta_i\}_t\right) \times$$

$$\left[\sum_{n=0}^{M-1} \mathcal{R}^n(\{r_i\}_d, \{r_i\}_s)\right] \frac{1}{M} \left[\sum_{n=0}^{M-1} \mathcal{R}^n(\{r_i\}_s, \{r_i\}_d)\right]$$

$$= (1 - \mathcal{R}^n(\{r_i\}_d, \{r_i\}_s)) \left[\sum_{n=0}^{M-1} \mathcal{R}^n(\{r_i\}_d, \{r_i\}_s)\right] \frac{1}{M} \left[\sum_{n=0}^{M-1} \mathcal{R}^n(\{r_i\}_s, \{r_i\}_d)\right] \quad (5.5.29)$$

In figure 5.10 the optimal M and the optimal average acceptance probability are shown as functions of the average rejection probabilities. Since the latter are not known in advance, M has to be chosen by other means. This leads to the topic of the next section, the tuning of the Monte Carlo to a particular data analysis problem.

5.5.5 Automatic adjustment of Monte Carlo parameter settings

For a successful computation, the Monte Carlo parameter settings, like the length of the random walk needed to acquire an independent sample, the ratio of acceptances to rejections, as well as the probability and number of retries of the non-local move had to be set to meaningful values. As discussed briefly in section 2.5.2, these values must be tuned in a way, so that the sampling algorithm is efficient. This is especially needed during nested sampling since the shape of the constrained prior changes during the progress of the algorithm.

Nested sampling

In every nested sampling iteration, the length scales of the local move, s_i, were adjusted *before* the actual random walk was performed, during which s_i had to be kept constant. The tuning was done with a modified version of the algorithm proposed by Skilling in Sivia (2006, chapter 9). In brief, the aim was to tune s_i in a way that a given ratio $\alpha_{\mathrm{A/R}}$ of acceptances to rejections is achieved. To this end, several short random walks, which contained local moves only, were started from randomly chosen *equilibrated samples* in parameter space. The equilibrated samples were essentially the current objects that satisfied the likelihood constraint L^*. The local move length scales s_i were decreased when the actual acceptance/rejection ratio was smaller than the set value $\alpha_{\mathrm{A/R}}$, and increased when it was larger. In contrast to Skilling's original algorithm, the modified version (algorithm 1) is able to treat each parameter space dimension i separately. Algorithm 1 makes use of the random numbers u_i, which were used to compute the size $\theta_{i+1} - \theta_i$ of the proposed step (even if it was rejected),

$$u_i = 1 - \exp(-\left|\theta_{i,t+1} - \theta_i\right|/s_i). \qquad (5.5.30)$$

u_i serves as a measure of the step size in the parameter space direction i, and approaches 0 and 1 at infinitely small and large steps, respectively. The amount of increase and decrease of s_i was proportional to u_i and was reduced during the initial random walk in order to stabilize s_i. The fluctuations of s_i between different nested sampling iterations were further damped by applying exponential smoothing.

The ratio of acceptances to rejections, $\alpha_{\mathrm{A/R}}$, was initially set by the user. It was adapted only in the parallel version of nested sampling (see section 5.5.6), otherwise $\alpha_{\mathrm{A/R}}$ stayed constant.

The length of the random walk, t_c, in between two sufficiently independent samples is initially set by the user. It is adjusted every 10 nested sampling iterations, so that the average correlation $g(\tau)$ after t_c time increments decays below a user-specified threshold c,

$$g(\tau) = \frac{\left\langle \left(\theta_{i,t} - \langle\theta_{i,t}\rangle_t\right)\left(\theta_{i,t+\tau} - \langle\theta_{i,t}\rangle_t\right)\right\rangle_t}{\left\langle \left(\theta_{i,t} - \langle\theta_{i,t}\rangle_t\right)^2\right\rangle_t} \le c \quad \text{for } \tau > t_c \text{ and all } i \qquad (5.5.31)$$

Here, the averages $\langle.\rangle_t$ are taken only over the time increments in which the argument is defined. In particular, $\langle\theta_i\rangle$ denotes the average position computed over the whole Markov chain.

Instead of calculating the autocorrelation functions in each parameter space dimension and fitting them with exponential decays, the averaged power spectra were computed and fitted with a separate Lorentz function each. This procedure is based on Parseval's theorem and has the advantage that the fit can be done purely analytically as described

---

Algorithm 1 Adjust length scales

// *initialize length scale adjustment*
$s_i = s_{i,\mathrm{prev}}$ // *start with length scale from previous MCMC or a user-defined value*
// *initialize acceptance and rejection counters in each parameter space dimension*
$n_i^{\mathrm{A}} = 0, \quad n_i^{\mathrm{R}} = 0$
choose random equilibrated sample θ_i as starting position
// *adjust length scale*
repeat
   // *move to another equilibrium position with probability p_{repo}:*
   if $\mathrm{rand}(0,1) < p_{\mathrm{repo}}$ **then**
      replace θ_i with new randomly chosen equilibrated position
   end if
   propose next position and check acceptance
   if move accepted **then**
      $n_i^{\mathrm{A}} \leftarrow n_i^{\mathrm{A}} + 2u_i$ // *increase acceptance counter*
   else
      $n_i^{\mathrm{R}} \leftarrow n_i^{\mathrm{R}} + 2u_i$ // *increase rejection counter*
   end if
   // *adjust length scales using adjustment sensitivity σ:*
   if $n_i^{\mathrm{A}} < \alpha_{\mathrm{A/R}} n_i^{\mathrm{R}}$ **then**
      // *decrease s_i when there are too many rejections*
      $s_i \leftarrow s_i \exp\left(-u_i\sigma / \left(1 + n_i^{\mathrm{A}} + n_i^{\mathrm{R}}\right)\right)$
   else if $n_i^{\mathrm{A}} > \alpha_{\mathrm{A/R}} n_i^{\mathrm{R}}$ **then**
      // *increase s_i when there are too many acceptances*
      $s_i \leftarrow s_i \exp\left(u_i\sigma / \left(1 + n_i^{\mathrm{A}} + n_i^{\mathrm{R}}\right)\right)$
   end if
until $(n_i^{\mathrm{A}} + n_i^{\mathrm{R}}) \geq N_{\mathrm{adj}}$ for all i
// *damp fluctuations in s_i by exponential smoothing $\ln s_i$ with decay rate $1 - \iota$:*
$s_i \leftarrow s_i^{\iota} s_{i,\mathrm{prev}}^{1-\iota}$

---

by Berg-Sørensen and Flyvbjerg (2004) without the need to use a time-consuming non-linear least squares algorithms. Even though the random walk does not satisfy the model assumptions underlying the fit, namely that it is a confined diffusive process with normal distributed step sizes $\theta_{i,t+1} - \theta_{i,t}$, the procedure was used to infer an average correlation decay.

Using the exponential decay model, the smallest length of the random walk, t_c', was calculated, which satisfied the inequalities (5.5.31) in every parameter space dimension. The random walk length used in the next 10 nested sampling iterations was calculated from the estimated length t_c' and the current value t_c as follows:

$$t_c \leftarrow \begin{cases} t_c' & \text{when } t_c < t_c' \\ \max\left(\lceil (t_c' + t_c)/2 \rceil, 20\right) & \text{when } t_c > t_c'. \end{cases} \tag{5.5.32}$$

Another important parameter is the probability to try a non-local move, p_{NL}. It was initially set to the 1/1000 fraction of the maximum, $p_{\mathrm{NL}}^{\max}$, which was set by the user. During nested sampling, it was adjusted after every iteration, based on two quantities obtained from the random walk.

The first quantity is the decay of correlation in between non-local moves. It was determined in a similar way like the correlation time t_c', which was explained above. To

this end, the average positions in equation (5.5.31), were replaced by "local" averages $\langle\theta_{i,t}\rangle_{t,t_k\leq t<t_{k+1}}$ computed in between two successive non-local moves at time t_k and t_{k+1}, i.e.

$$\langle\theta_{i,t}\rangle_{t,t_k\leq t<t_{k+1}} = \frac{1}{t_{k+1}-t_k}\sum_{t=t_k}^{t_{k+1}-1}\theta_{i,t}. \tag{5.5.33}$$

The average power spectra were thus computed from the random walk with subtracted local average, $\theta_t - \langle\theta_{i,t}\rangle_{t,t_k\leq t<t_{k+1}}$. To this end, the random walk with subtracted local average was split in 10 segments of equal length, and a power spectrum was computed for each of the segments. Finally, the power spectra were averaged and fitted with a Lorentz function each. From the fits, the average correlation decay was determined.

The second quantity used in the adjustment of p_{NL} is the ratio of "far" to "near" diffusion events, which characterizes the random walk in between two successive accepted non-local moves. Let the non-local moves have the source and destination reference points $\{r_i\}_{s1}$, $\{r_i\}_{d1}$, and $\{r_i\}_{s2}$, $\{r_i\}_{d2}$. When $\{r_i\}_{d1} \neq \{r_i\}_{s2}$ the Markov chain was able to escape the proximity of the destination reference position $\{r_i\}_{d1}$ by local moves only, and this is categorized as a *far diffusion event*. In the opposite case, when $\{r_i\}_{d1} = \{r_i\}_{s2}$, the successive non-local moves end and start at the same reference point, which is called a *near diffusion event*.

The non-local move probability, p_{NL}, is adjusted after each nested sampling iteration, as listed in algorithm 2. In short, by changing p_{NL} the actual far/near diffusion event ratio is controlled, so that it is close to the user-specified ratio of far to near diffusion events, $\alpha_{F/N}$. The adjustment is applied only if the correlation in between non-local moves is above the maximum correlation, c_{NL}, also specified by the user.

Algorithm 2 Adjust non-local move probability

---

if (correlation in between non-local moves did not decay on average below c_{NL}) **then**

    if (number of far diffusion events $< \alpha_{F/N} \times$ number of near diffusion events) **then**

        $p_{NL} \leftarrow \exp\left(\log\left(p_{NL}\right)\times\exp\left(0.009+0.001\times\text{rand}(0,1)\right)\right)$ // *decrease p_{NL}*

    end if

else

    // *correlation decayed on average*

    $p_{NL} \leftarrow \exp\left(\log\left(p_{NL}\right)\times\exp\left(-0.009-0.001\times\text{rand}(0,1)\right)\right)$ // *increase p_{NL}*

    $p_{NL} \leftarrow \min\left(p_{NL}, p_{NL}^{\max}\right)$ // *and stay below*

    upper limit

end if

The idea behind sampling at the ratio $\alpha_{F/N}$ (that should be set > 1) is that even when the correlation between non-local moves does not decay below c_{NL}, the parameter space in between the reference points is sampled sufficiently well. When p_{NL} would be too high and $\alpha_{F/N} \ll 1$, a global correlation decay could be achieved by jumps in between positions close to the reference points, without ever sampling the parameter space in between them properly.

The number of retries, M, to find an accepted position in the non-local move was also adjusted every iteration. Hereby the numbers of "forward" and "backward" rejections were used, i.e. the number of rejections due to failure to find an accepted position in the proximity of the destination reference point (step 3 in figure 5.9) and the number of rejections due to failure to generate a given number of rejections in the proximity of the source reference point (step 5 in figure 5.9), respectively. When the number of forward rejections was larger/less than the number of backward rejections, M was increased/decreased by

1. The minimum value for M was 1, the maximum $M_{\max}$ could be set by the user.

Posterior sampling

During posterior sampling, $\alpha_{\mathrm{A/R}}$ was set to the average value used to calculate the nested samples with 90-100% of the maximum weight. The length scales of the local move, s_i, were adjusted with algorithm 1 once before the actual sampling took place. A subset of the equally weighted samples computed from the previous nested sampling run was used as both, the equilibrated samples needed in the adjustment, and the reference points used in the non-local move during sampling. Also the procedures described to adjust t_c, p_{NL} and M were applied before the actual sampling. They were run simultaneously and repeated until t_c and p_{NL} were stable.

5.5.6 Parallelization

The computation of samples was parallelized on up to 8 threads, which were executed on a personal computer (PC). The PCs used to perform the calculations were equipped with various processors (Intel Core2 Quad 3 GHz, Intel Core i7 2.8 GHz, AMD Phenom II X4 3.39 GHz). In the case of posterior sampling, the parallelization was straight forward after the Monte Carlo parameters have been set since the code execution was completely independent.

In the case of nested sampling, however, the Message Passing Interface (MPI) functionality of the MATLAB parallel computing toolbox was used to set up communication in between the threads, also called "labs" in MATLAB. Every 10 iterations, the labs exchanged the newly generated active objects, which were included into the set of reference points of each lab. The current Monte Carlo parameter settings were exchanged and compared as well in between labs. The settings (including the local move length scales) of the lab exhibiting the fastest decay of correlation were then used by all labs for the next 10 iterations after a small random alteration. This simple genetic algorithm optimized the Monte Carlo parameter settings during the nested sampling run.

Sample weights and evidence were computed from the samples generated in each lab as described in section 2.5.2 by using the sorted list of all samples.

5.5.7 Settings used for NPS analysis

When the sampling engine was used for NPS analysis, each prior factor in equation (5.5.2) contained the position and orientation parameters of either one fluorophore or one docking reference frame. The prior information of the five fluorophore parameters $(\boldsymbol{x}_i, \boldsymbol{\Omega}_i)$ respectively 6 reference frame parameters $(\boldsymbol{o}_k, \boldsymbol{\Xi}_k)$ is thus encoded in a box collection prior. In this way, it is possible to express correlations between positions and orientations, even though this feature was not exploited in the current work.

The boundary conditions were reflecting in all spatial variables, i.e. fluorophore positions, $\boldsymbol{x}_i$, and reference frame positions, $\boldsymbol{o}_k$. The orientations of the average fluorophore transition dipole moments had the following Klein bottle-like boundary conditions

$$-\cos(\theta_i) \leftarrow \begin{cases} +\cos(\theta_i) & \text{if } \phi_i < 0 \text{ or } \phi_i > \pi \\ -\cos(\theta_i) & \text{otherwise} \end{cases} \tag{5.5.34}$$

$$\phi_i \leftarrow \begin{cases} (\phi_i - \pi) \mod \pi & \text{if } \phi_i > \pi \\ \phi_i \mod \pi & \text{if } \phi_i < 0 \\ \phi_i & \text{otherwise} \end{cases} \tag{5.5.35}$$

$$- \cos(\theta_i) \leftarrow \begin{cases} -1 + (-\cos(\theta_i) - 1) \mod 2 & \text{if } -\cos(\theta_i) > 1 \\ 1 - (1 + \cos(\theta_i)) \mod 2 & \text{if } -\cos(\theta_i) < -1 \\ -\cos(\theta_i) & \text{otherwise.} \end{cases} \quad (5.5.36)$$

In the angular parameters $\rho_{(k)}$ and $\phi_{(k)}$ used for reference frame orientation, periodic boundary conditions were applied, in the parameter $-\cos(\theta_{(k)})$ it was reflective boundary conditions. Unfortunately these are not the correct boundary conditions since a remapping of $-\cos(\theta_{(k)})$ should be accompanied by a shift of both $\rho_{(k)}$ and $\phi_{(k)}$ by π. Since both the correct and incorrect re-mapping would occur when $-\cos(\theta_{(k)}) \approx \pm 1$ the orientation of the $x_{(k)}$ and $y_{(k)}$ axes of the reference frame would not change much in both cases. Remapping with the incorrect boundary conditions will thus not lead to a completely different reference frame orientation.

The Monte Carlo parameter settings used in the calculations are listed in table 5.3. Different numbers of active objects were used, ranging from 120 objects in the most simple calculations that involved only one antenna and one satellite, up to 10000 objects in the case of a calculation with 7 antennas and 7 satellites.

Nested sampling was terminated when all of the following conditions were met: First, the average sample weights had to decrease, which guaranteed that the likelihood was rising slower than the shrinking of the constrained prior mass. Second, the product of the number of objects in the collection, the acquired information H (equation (2.5.19)) and a user defined factor between 1.5 and 2 had to exceed the number of nested sampling iterations.

parameter name	abbrev.	value
initial local move length scale	s_i	$0.1 \times (\theta_i^{\mathrm{U,tot}} - \theta^{\mathrm{L,tot}})$
repositioning probability	p_{repo}	0.5
adjustment sensitivity	σ	0.9
adjustment influence	ι	0.1
initial acceptance/rejection ratio	$\alpha_{\mathrm{A/R}}$	$0.3 \ldots 1.0$
correlation decay	c	0.05
initial random walk length	t_c	200000
number of adjustment steps	N_{adj}	1000
maximum non-local move probability	p_{NL}	0.1
correlation decay between non-local moves	c_{NL}	0.05
initial number of retries	M	1
maximum number of retries	M_{max}	100
far/near diffusion ratio	$\alpha_{\mathrm{F/N}}$	3

Table 5.3: Monte Carlo parameter settings used in the NPS analysis.

5.5.8 Implementation

The sampling engine is called via two main sampling functions, `NestedSampler.m` and `MetropolisSampler.m`, or alternatively via their parallel versions `ParNestedSampler.m` and
`ParMetropolisSampler.m` (figure 5.11). `NestedSampler.m`/`ParNestedSampler.m` perform nested sampling, whereas `MetropolisSampler.m`/`ParMetropolisSampler.m` are used to sample the posterior.

The posterior is encoded in the likelihood (specified in `loglikelihood.cpp`), the prior (specified in box collection prior format), and the data. The user-specified function `loglikelihood()`, which is located in `loglikelihood.cpp`, must return the natural logarithm of the likelihood given the position in parameter space and the data that is passed internally through the calling functions. In this work `loglikelihood.cpp` was adapted to the NPS parameter estimation problem by implementing the likelihood of the *position - orientation* model with optional docking (equation (3.2.3)). The calling functions of `loglikelihood.cpp` are `seedfun.cpp` and `explfun.cpp` that compute samples from the unconstrained prior and produce a new sample from a given starting position by performing a random walk, respectively. They can be accessed from MATLAB by the wrapper functions, `seedfunction.cpp` and `explorationfunction.cpp`.

Before executing the sampling functions, the C source code must be compiled into `.mex` functions that are executable in MATLAB (The MathWorks). The compilation was done with `buildmex.m`, which called the MATLAB MEX compiler. Since nested sampling and Metropolis posterior sampling both use Markov chain Monte Carlo, the same source code was used to compile the executables `explorationfunctionNS.mex` and `explorationfunctionMS.mex`, which were called during nested sampling and posterior sampling, respectively. The only difference between the two functions is the acceptance criterion used during the random walk, which was switched by a preprocessor flag at time of compilation. `seedfunction.mex` used the box collection prior to draw samples from the unconstrained prior and was called by `NestedSampler.m` in order to compute the initial set of samples together with their log-likelihood values. From the samples acquired in the nested sampling run, posterior samples were computed by `EquallyWeightedSamples.m` and passed to `MetropolisSampler.m`, where they were used then as reference positions in the non-local move. `NestedSampler.m` and `MetropolisSampler.m` generated temporary files, which were used to recover the calculation results when one of the programs did not terminate properly.

After the above discussion of the methodical details, the next chapter will focus on the main results of the NPS analysis.
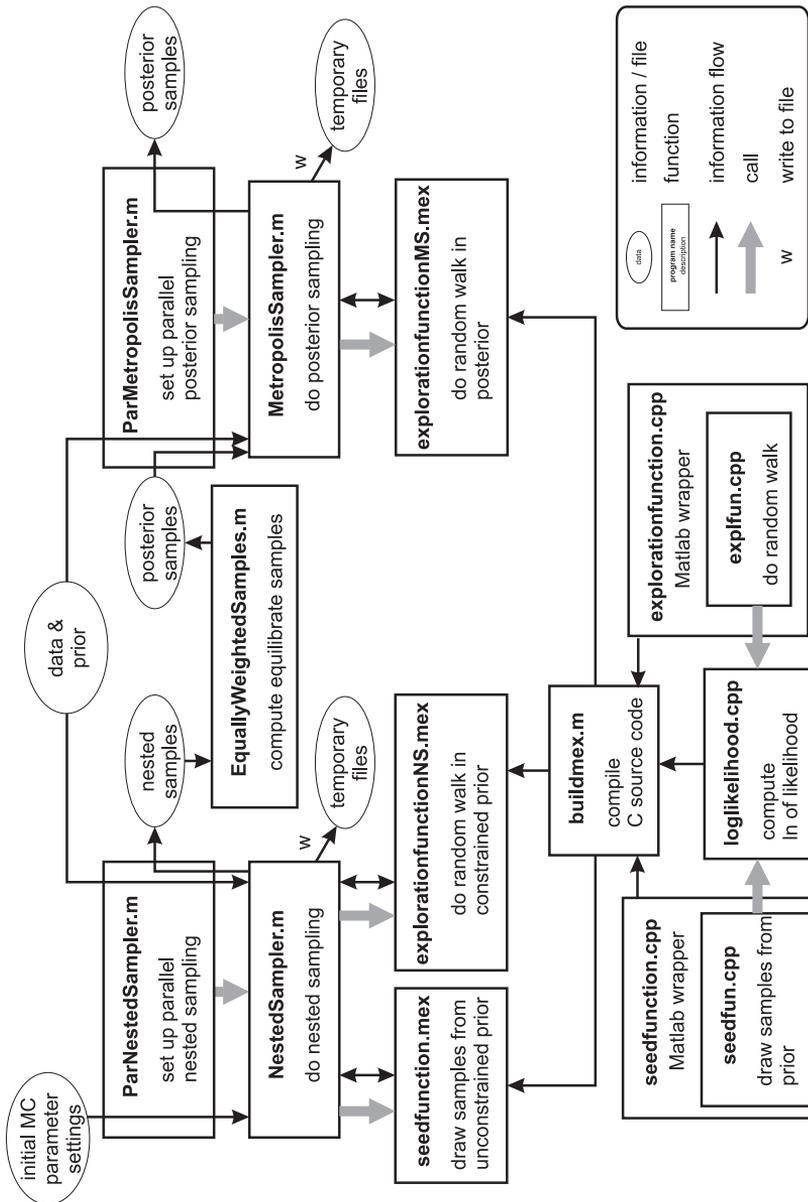
Figure 5.11: Sampling engine software overview

6 Results and discussion

This chapter consists of the main results accumulated during the development of the NPS. Simulations and applications to the Pol II system will be presented in an alternating order. Starting from simulations of Förster distance distributions, more and more complex scenarios will be addressed. Each simulation and application will be discussed immediately, followed by a general discussion at the end of the chapter.

6.1 Simulation I – Förster distances

In the first part of this section, the effect of fluorescence anisotropies on the simulated Förster distance priors will be studied. The justification of the independence of Förster distances, which is required in the *position - Förster distance* NPS model, will be motivated in the second part.

6.1.1 Effects of fluorescence anisotropy on the Förster distance distribution

To study the impact of fluorophore orientation constraints on the Förster distance, the Förster distance priors $p(R_{ai}|I)$ of a FRET pair consisting of an antenna a and a satellite i were calculated as described in section 5.3.2. The computed densities are shown in figures 6.1 and 6.2 for different residual fluorescence anisotropies of donor and acceptor. The priors consist of a peak centered roughly at the isotropic Förster distance and exhibit one and in certain cases two maxima. The highest maximum lies below the isotropic Förster distance when both fluorescence anisotropies are low, and above when both are high. The width of the densities is rising with the fluorescence anisotropies, as one can readily see in figures 6.1 and 6.2. This observation which is supported by the standard deviations of $R_{ai}/R_{ai}^{\mathrm{iso}}$ (table 6.1). Even when the fluorescence anisotropy of one of the fluorophores is small, the corresponding Förster distance distribution can be broadened if the other residual fluorescence anisotropy is high.

6.1.2 Discussion

It is evident from the above simulations that the effect of fluorescence anisotropies on the Förster distance distribution is strong. Even at small residual fluorescence anisotropies of $r_{\infty,i} = r_{\infty,a} = 0.1$ the standard deviation of $p(R_{ai}|I)$ is above 5% of the isotropic Förster

$r_{\infty,i}$ \ $r_{\infty,a}$	0.01	0.1	0.2	0.3	0.4
0.01	1.7%	4.1%	5.4%	6.6%	7.7%
0.1		6.3%	7.8%	9.6%	12.0%
0.2			9.0%	10.8%	12.9%
0.3				13.1%	16.0%
0.4					23.9%

Table 6.1: Relative Förster distance standard deviations are shown for different ANT and SAT fluorescence anisotropy values, $r_{\infty,a}$ and $r_{\infty,i}$, respectively. They were calculated by dividing the standard deviations of the simulated Förster distance priors, $p(R_{ai}|I)$, by the corresponding isotropic Förster distance R_{ai}^{iso}.
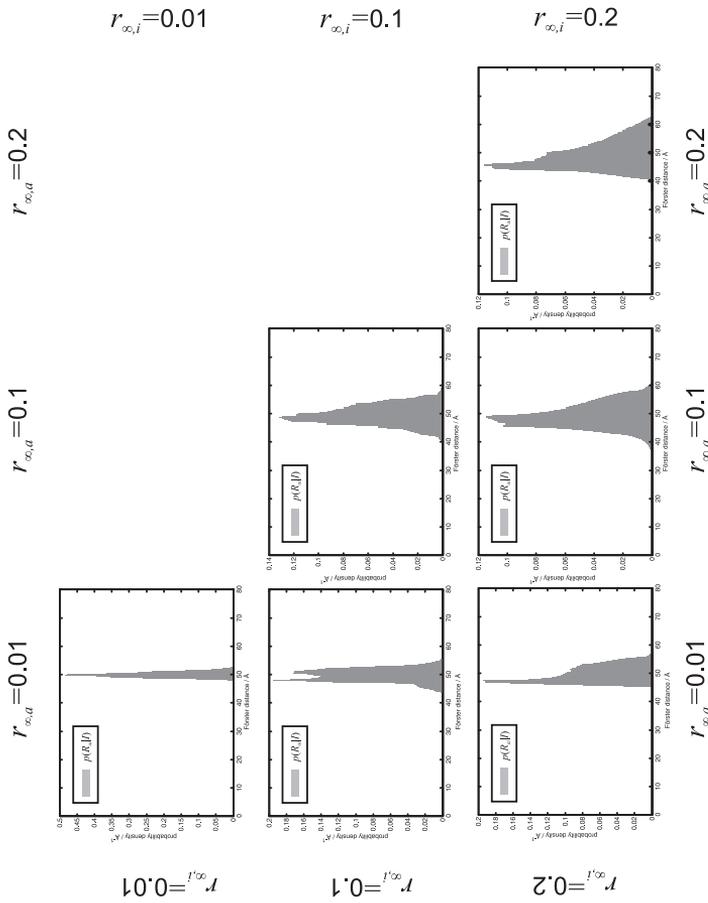
Figure 6.1: Simulated Förster distances, part I. The second part with legend is depicted in figure 6.2.
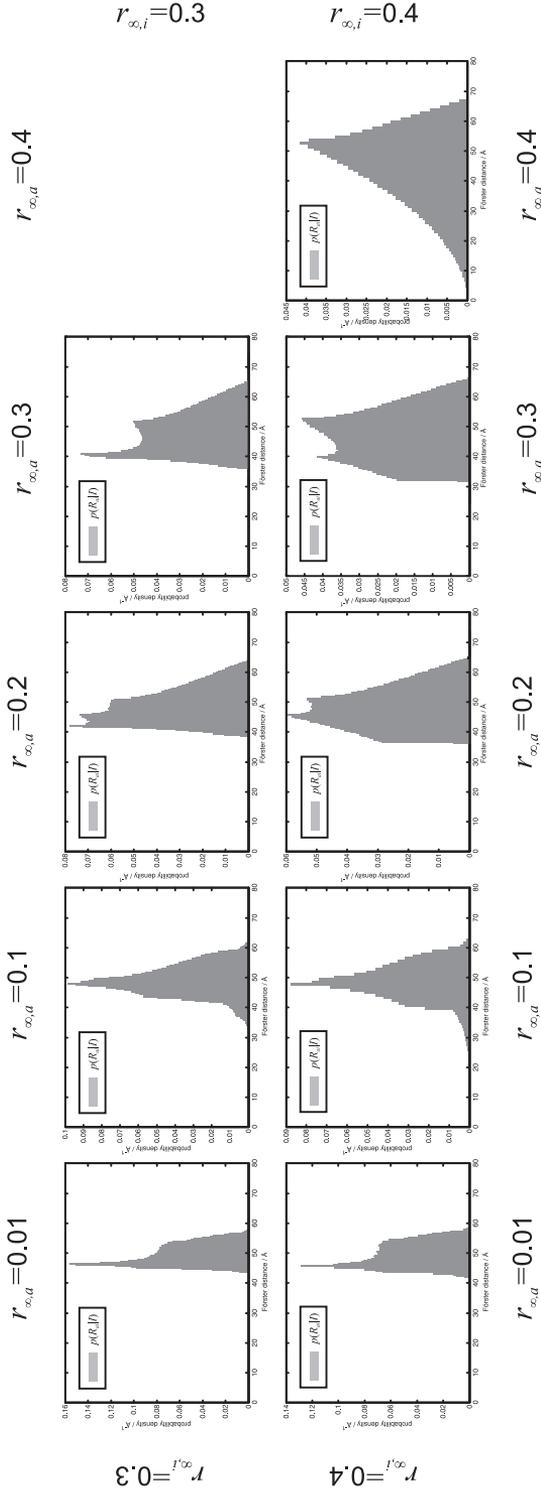
Figure 6.2: Simulated Förster distances, part II. The first part is shown in figure 6.1. The prior probability densities were simulated with residual fluorescence anisotropy values of antenna ($r_{\infty,a}$) and satellite fluorophores ($r_{\infty,i}$) ranging from 0.01 to 0.4, and an isotropic Förster distance of $R^{\mathrm{iso}} = 50$Å. Even when one of the fluorescence anisotropies is small (0.01), there is substantial broadening of the Förster distance prior $p\left(R_{ij}|I\right)$. The largest broadening is observed when the fluorophores have fixed orientations, i.e. their fluorescence anisotropies are 0.4.

distance. For $R^{\text{iso}} = 50\,\text{Å}$, this would result in an approximately $10\,\text{Å}$ wide 95% credible interval of $p\left(R_{ai}|I\right)$. When $r_i = r_a = 0.2$, this value would be even $18\,\text{Å}$. Since the Förster distance is proportional to the inferred distances, the uncertainty in R_{ai} has a similar effect like the uncertainties caused by the linkers (table 5.2). Both sources of uncertainty will often have comparable magnitudes, and both should be accounted for, since they can be easily scaled up when the satellite fluorophores are placed in a disadvantageous way.

In particular, using the "rule of thumb" (Roy et al., 2008, and references therein) of setting the orientation factor to its isotropic value when the fluorescence anisotropies are below 0.2, can result in a severe underestimation of the uncertainties in the inferred distances.

As a remark, one should stress that the information contained in the Förster distance prior will be used in full detail in the NPS. The maximum error estimate proposed by others (Dale et al., 1979; Ivanov et al., 2009) and used for instance by Kaiser et al. (2006) would decrease the information content of the inferred position estimate. When a flat prior in the Förster distance between the maximum and minimum values of R_{ia} were used instead of the prior shown here, one would state to have information about the average transition dipole moment orientations. In that way, the unlikely cases of extremely good and bad coupling would be assigned more weight in the calculation. Thus, one would overestimate the scenario of the average transition dipole moments being parallel to the line between donor and acceptor, which corresponds to good coupling. Also the perpendicular orientations of the average transition dipole moments of donor and acceptor, which correspond to bad coupling, would be considered more important than they actually are. Finally, the non-informative prior as computed in section 5.3.2 should be used, unless there is reliable information about the average transition dipole moment orientations.

Having demonstrated the influence of fluorescence anisotropies on the Förster distance distribution of two fluorophores, in the next section the correlations of Förster distances will be studied.

6.1.3 Förster distance correlations

An assumption made in the *position - Förster distance* NPS model is the mutual independence of the Förster distances R_{ij} and R_{ik} of the FRET pairs ij and ik, which "share" the antenna fluorophore i. This assumption is crucial for the factorization of the posterior (equation (3.1.8)). However, it is clearly an approximation, since both Förster distances depend on the position and average transition dipole moment orientation of the common fluorophore i. A correlation is thus expected, yet it was not proven that the approximation holds.

To test the validity range of the above approximation, simulations of the Förster distance distribution of two FRET pairs were carried out (section 5.3.2). As described above, the FRET pairs comprised the fluorophores i, j and k, and were coupled by the common fluorophore i (figure 6.3). Assuming ignorance about the average transition dipole moment orientations $\boldsymbol{\Omega}_i$, $\boldsymbol{\Omega}_j$ and $\boldsymbol{\Omega}_k$, the density $p\left(R_{ij}, R_{ik}|\boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{x}_k, I\right)$ was computed for given fluorophore positions.

Now, to quantify the degree of independence of the two Förster distances, the normalized mutual information (NMI), $\mathcal{I}(R_{ij}, R_{ik})$, (Strehl and Ghosh, 2002) of the density $p\left(R_{ij}, R_{ik}|\boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{x}_k, I\right)$ was computed (section 5.3.3). This quantity 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). Without loosing generality, it was assumed that the isotropic Förster distances are equal, $R_{ij}^{\text{iso}} = R_{ik}^{\text{iso}} = R^{\text{iso}}$, and that the fluorophores j and k are separated by the distance $d = |\boldsymbol{x}_j - \boldsymbol{x}_k|$. The residual fluo-

Figure 6.3: Simulation of correlated Förster distance distributions. The joint Förster distance density $p\left(R_{ij}, R_{ik} | \boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{x}_k, I\right)$ between three fluorophores, i, j and k (double arrows) is correlated by the orientation and position of the fluorophore i. The FRET measurements corresponding to the Förster distances are indicated by dashed lines. It was assumed that the isotropic Förster distances of both FRET efficiency measurements are equal, $R_{ij}^{\mathrm{iso}} = R_{ik}^{\mathrm{iso}} = R^{\mathrm{iso}}$. The positions of two fluorophores j and k was fixed, 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.
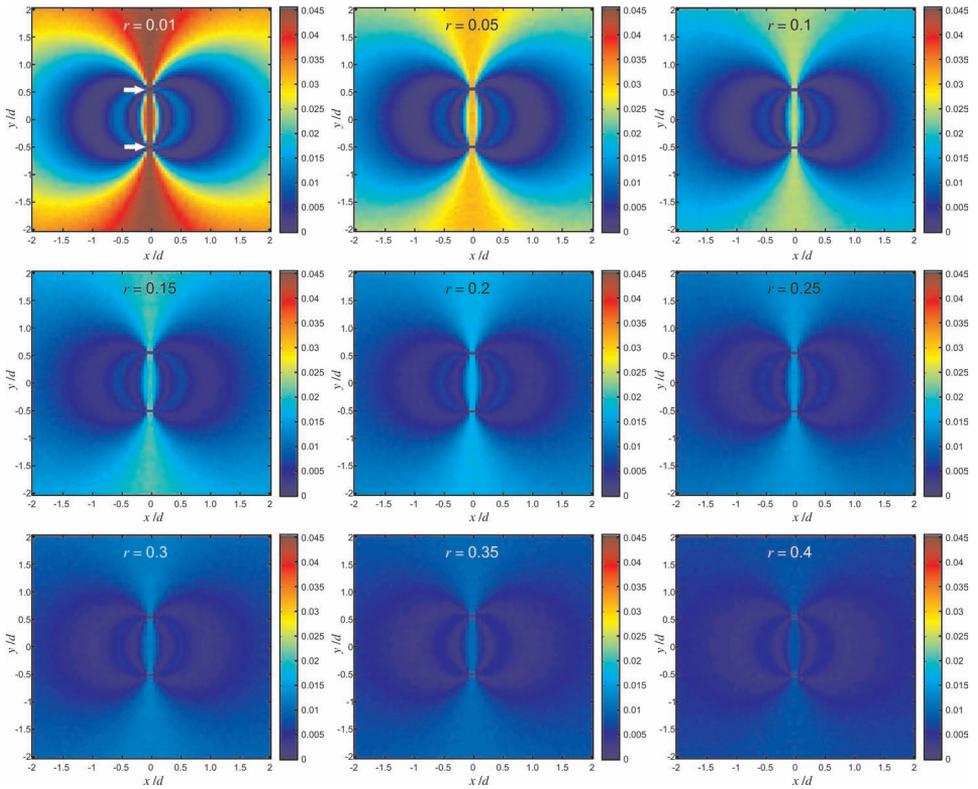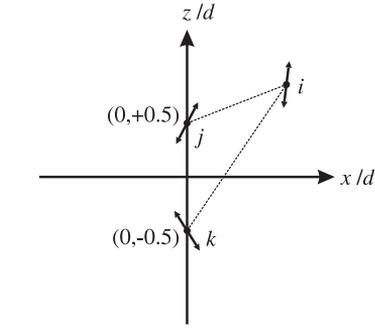
Figure 6.4: Normalized mutual information (NMI), $\mathcal{I}(R_{ij}, R_{ik})$, of two correlated Förster distances, R_{ij} and R_{ik}. The dependence of the NMI on the position $\boldsymbol{x}_i = (x, 0, z)$ of the fluorophore i is shown in each subfigure. The residual fluorescence anisotropies of all 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.

rescence anisotropies of each fluorophore were set to the same value, r. The position $\boldsymbol{x}_i$ of the fluorophore i was scanned in the (x, z)-plane within the limits $-2d < x < 2d$ and $-2d < z < 2d$, and at each position the NMI was plotted for different values of r (see figure 6.4).

The highest NMI values were observed when all fluorophores were aligned along a common axis. Independent of the fluorophore position, along this axis the NMI was constant and ranged from 0.054 for the extremely small residual fluorescence anisotropy $r = 1 \cdot 10^{-5}$ to 0.0085 for $r = 0.4$. For moderate fluorescence anisotropy values of $r = 0.1$, the NMI was 0.025 and on average (over the range of fluorophore positions x_i) it was much smaller than these maximum values.

6.1.4 Discussion

The small NMI values obtained in the simulations indicate that the correlations of the Förster distances introduced by a common antenna fluorophore are indeed negligible. This can be readily seen in figure 6.5, where the correlated joint densities of Förster distances, $p\left(R_{ij}, R_{ik} | \boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{x}_k, I\right)$, and the uncorrelated reconstructions from marginal densities, $p\left(R_{ij} | I\right) p\left(R_{ik} | I\right)$, are shown.

Judging from the maximum NMI values (figure 6.4), the approximation of independent Förster distances holds best for high residual fluorescence anisotropies and is still reasonable when the residual fluorescence anisotropies are small and orientation effects are less important. Thus, the approach of independent Förster distances used in the *position - Förster distance* NPS model is put on a firm footing.

The approximate independence of the Förster distances will break down when more information on the average transition dipole moment orientations is available. This applies either when an informative fluorophore orientation prior exists, or when there is data, like FRET anisotropy, that couples the fluorophore orientations tightly.

In the next section, two applications of the NPS based on the *position - Förster distance* model will be shown.
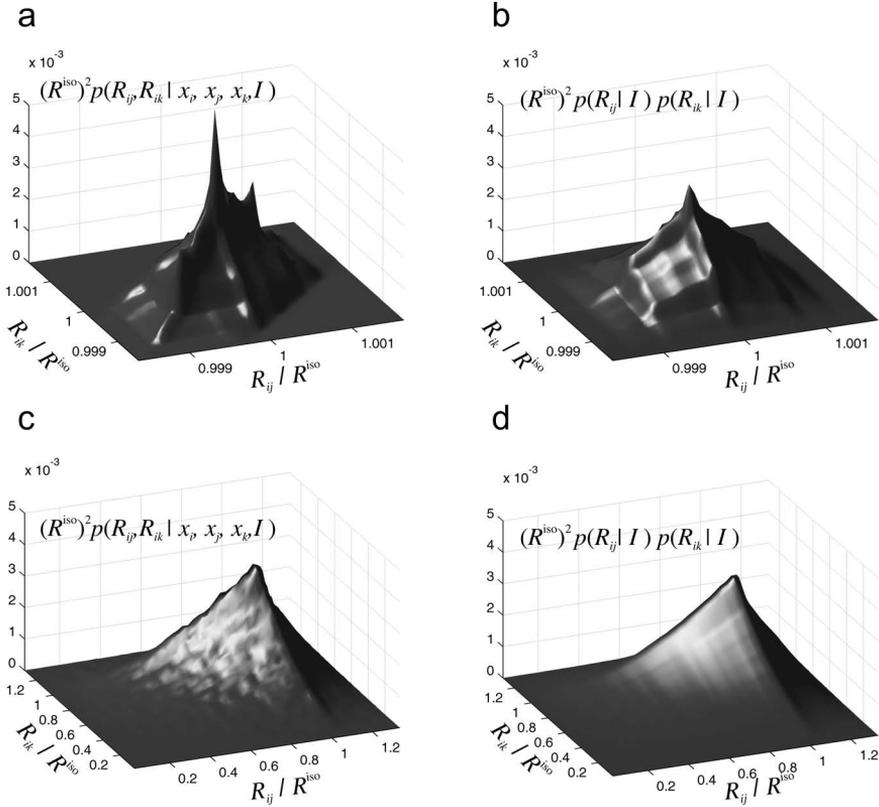
Figure 6.5: Joint and reconstructed Förster distance distributions. In panel (a) and (c), the correlated joint Förster distance densities $p\left(R_{ij}, R_{ik} | \boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{x}_k, I\right)$ are displayed, while the panels (b) and (d) show the uncorrelated Förster distance densities $p\left(R_{ij} | I\right) p\left(R_{ik} | I\right)$ reconstructed from the marginal probability densities. In the simulation the fluorophores were aligned on a common axis. The densities are shown for two fluorescence anisotropies $r = 1 \cdot 10^{-5}$ (a, b) and $r = 0.4$ (c,d). In the case of $r = 1 \cdot 10^{-5}$, the bulk of the joint probability density is described well by the reconstruction, while only fine features like the ridge at $R_{ij}/R^{\text{iso}} = R_{ik}/R^{\text{iso}}$ and the missing density on the anti-diagonal $R_{ij}/R^{\text{iso}} + R_{ik}/R^{\text{iso}} = 1$ cannot be recovered. Note also that the absolute width of the distribution is extremely small (only 1/1000 of the isotropic Förster distance). For high fluorescence anisotropies, the joint distribution and the reconstructed distribution are almost identical.

6.2 Application I − RNA in the Pol II elongation complex

In the following, the localization capability of the NPS will be demonstrated with experimental data. The *position - Förster distance* model was used to determine positions of fluorophores attached to the 3' and 5'-end of a 29 nucleotides long RNA within the Pol II elongation complex (Muschielok et al., 2008). Since all experiments were performed in the absence of NTPs, the elongation complexes were stalled and not able to transcribe.

6.2.1 NPS inference of the 3'-RNA end position

The position of the antenna fluorophore attached to the 3'-RNA end was determined to serve as a proof of principle for NPS, since this antenna position is known from the crystal structure. The analysis is based on experimental data measured by Joanna Andrecka and listed in the appendix II in tables 1 (FRET efficiencies), 2 (isotropic Förster distances) and 3 (fluorescence anisotropies).

Alexa 647 was used as satellite fluorophore and was attached at five labeling sites. Three of them (tDNA(-10), tDNA(+3), tDNA(+9)) were located on the template DNA, the other two (Rpb7(S16C), Rpb7(C150)) were situated on the Rpb4/7 heterodimer. Depending on the measurement, Alexa 555 or TMR was used as antenna fluorophore, which was attached to the 3'-RNA end (RNA(+1)). Both, Alexa 555 and TMR, were assumed to be located at the same position. The measurements are summarized in figure 6.6, and the approximated satellite position priors are shown in figure 6.7. The antenna position prior was a cuboid of several hundred Ångström side length and did not contain any information about the Pol II elongation complex besides its gross position. In particular, the ANT was not excluded from the interior of Pol II, since the accuracy of NPS should be tested.

The marginal ANT position posterior, as estimated with the *position - Förster distance* NPS, is shown in figure 6.8. As a control, the possible positions of the antenna fluorophore were determined based on the model used to compute the accessible volumes of the satellites. The marginal posterior has a maximum at $(x, y, z) = (117, 46, -2)$ Å in the coordinate system of the crystal structure 1Y1W (Kettenberger et al., 2004), and standard deviations in the principal directions of $4.6, 5.4$ and 9.5 Å. It is consistent with the control, which has its center-of-mass at $(x, y, z) = (110, 53, 9)$ Å.

The influence of the number of measurements was studied as well (figure 6.9). When only the SATs on the DNA are used, the marginal posterior resembles a ring. With a fourth measurement from Rpb7(S16C), the marginal posterior separates into two modes connected by a thin thread of probability. Finally, the measurement from Rpb7(C150) selects one of the modes.

To verify the consistency of the data, i.e. whether some FRET efficiency measurements

Figure 6.6: 3'-RNA end FRET network. The position of the antenna fluorophore attached to the 3'-end of RNA (position RNA(+1)) was determined from 5 FRET efficiencies (dotted lines) to satellite fluorophores attached to Pol II and the template DNA. Alexa 647 (ALX647) was used as satellite throughout, while TMR and Alexa 555 (ALX555) were used as antenna. TMR and ALX555 were assumed to occupy the same position.
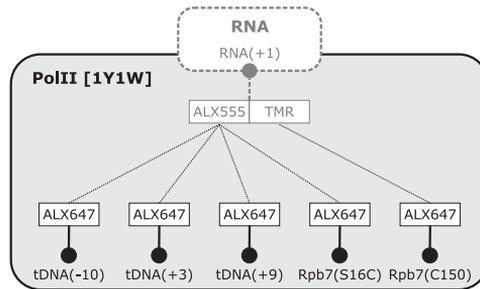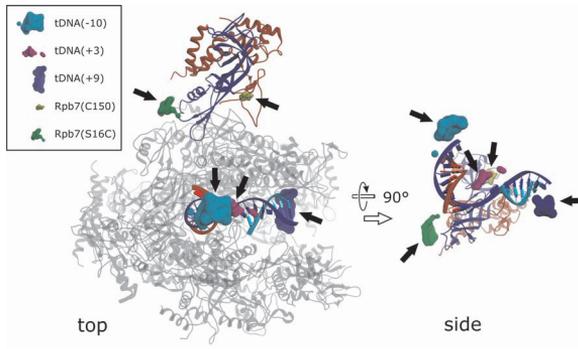
Figure 6.7: 3'-RNA satellite position priors. The 68% credible volumes of the satellite position priors used in the inference of the 3'-RNA end position are highlighted by arrows. The Pol II elongation complex is shown in the top and side view. The core (gray ribbons), Rpb4 (red ribbons), Rpb7 (blue ribbons), template DNA (blue), nontemplate DNA (cyan) and RNA (red) are displayed. The Pol II core was left out in the side view for clarity.
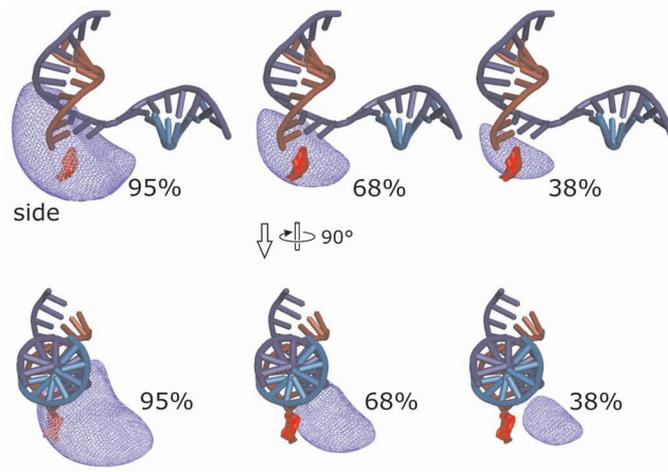
Figure 6.8: Inference of the 3' RNA end position. The 95%, 68% and 38% credible volumes (blue mesh) of the marginal antenna position posterior are shown in the context with the nucleic acids (cartoons, template DNA: dark blue, nontemplate DNA: bright blue, RNA: red). The upper panels show the standard side view, the view in the bottom panels is rotated by 90°. The red surface is the volume accessible to the antenna fluorophore, in which the attachment and the steric constraints were accounted for.

Figure 6.9: 3' RNA end position inference with different number of data. (a,b,c) show the side view, (d,e,f) the bottom view. The 68% credible volumes of the antenna position are shown as red mesh. In (a,d) only three measurements were used (tDNA(-10), tDNA(+3), tDNA(+9)), whereas in (b,e) Rpb7(S16C) and in (c,f) Rpb7(C150) were added consecutively.
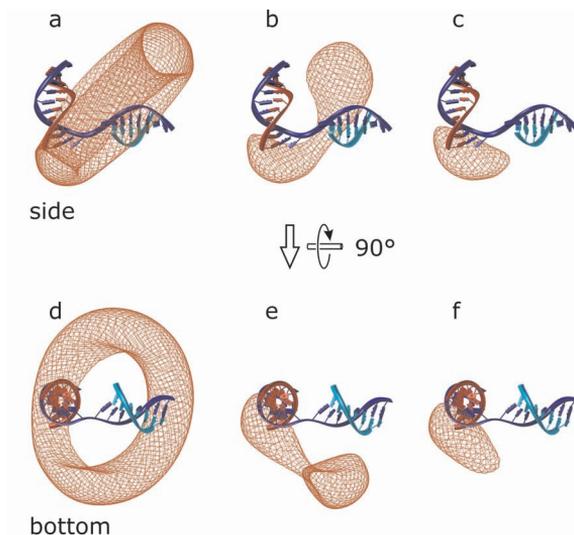
Figure 6.10: Schematic computation of posterior and measurement profiles. The posterior (density, left) and a measurement contribution (density, right) are projected (curved arrows) onto a semi-infinite line. Each probability element at the distance d from the center of mass of the satellite position prior (solid circles) is thus accumulated, and contributes to the posterior profile and measurement profile, respectively (bottom).
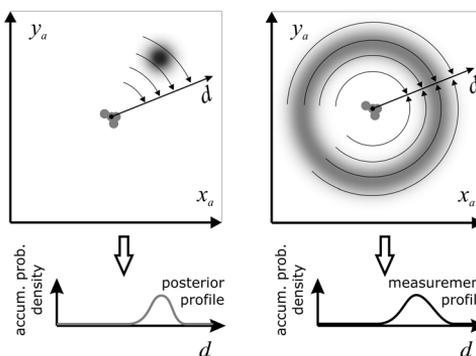
were "outliers", the overlap of the marginal posterior, $p\left(\boldsymbol{x}_a | \{E_{ai}\}, I\right)$, and the contribution of each measurement, $\sum_j w_{ij} K_{ij}(\boldsymbol{x}_a)$, were compared. To this end, *posterior profiles* and *measurement profiles* were computed (section 5.3.6). These are projections of either the marginal posterior or the contribution of a measurement onto a semi-infinite line, which starts at the center-of-mass of the satellite position posterior (see figure 6.10). A vanishing overlap between the profiles signalizes an inconsistency of the measured data. In the case of the ANT attached to the 3' RNA end, the data was consistent with the model, as can be seen from figure 6.11.

6.2.2 Discussion

As readily seen from figure 6.8, the position of the antenna fluorophore attached to the 3'-end of RNA was correctly inferred by five FRET efficiency measurements. The localization uncertainty, which ranges from 4% to 7% of the size of Pol II, proves NPS a versatile tool to answer structural questions on the 10 Å length scale.

It is important to stress that the shape of the marginal antenna position posterior does not imply any mobility, since the fluorophore positions do not vary over time in the model. Instead, the uncertainties originate in the limited information about the satellite position
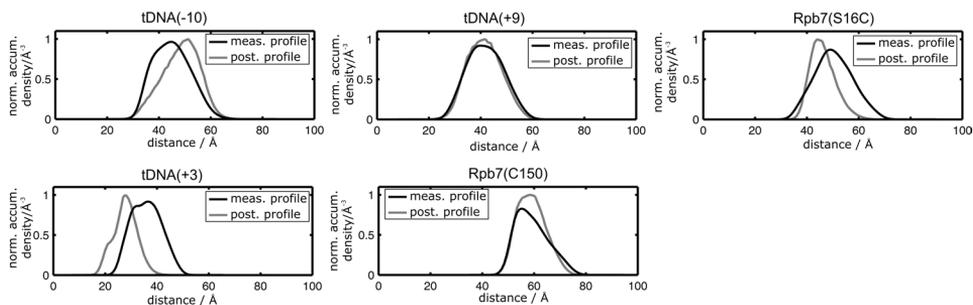
Figure 6.11: 3' RNA end position inference, posterior and measurement profiles. The well overlapping normalized measurement profiles (black) and posterior profiles (gray) indicate that there is no data inconsistent with the model.

(see accessible volumes in figure 6.7), the widths of the Förster distance distributions, and the number and quality of the measurements available. As depicted in figure 6.9, the estimated antenna position changes its shape dramatically, depending on how many and which measurements are analyzed.

The fact that also the satellite positions are important for the uncertainty of the position determination becomes evident in panel (a) and (d) of figure 6.9, where three measurements from the satellites tDNA(-10), tDNA(+3) and tDNA(+9) were analyzed. The observed shape of the posterior does not differ much from what one would expect when only two measurements of tDNA(-10) and tDNA(+9) were analyzed. In fact, the position prior of tDNA(+3) is located close to the center of the ring and cannot contribute much new information (compare figures 6.9a and 6.7, side view). The other measurements from Rpb7(S16C) and Rpb7(C150) are more informative and concentrate the posterior further, first into two high-probability regions (additional measurement from Rpb7(S16C), figure 6.7b,e) and then into a single maximum (all measurements, figure 6.7c,f).

Finally, since the posterior overlapped well with the contributions of each measurement, the data did not show any inconsistencies. In this particular case, the inference is based on five FRET efficiency measurements, a fairly small number of data when compared to the minimum number of four measurements needed to create a marginal antenna position posterior closest to a point-like object (see introduction of chapter 3). The presence of inconsistencies would hence allow only to tell that at least one of the measurements is not consistent with the rest. However, it would be impossible to say which one it is (when it is only one), since every combination of four measurements yields a different "intersection point", which is inconsistent with the remaining measurement.

For the more general situation in which more distance measurements are used, it is important to check always the consistency since it can reveal gross errors in the acquired data or in the analysis assumptions.

After having demonstrated the capability of NPS to localize a fluorophore the first biological application will be presented and discussed in the next two sections.

6.2.3 NPS inference of the 5' RNA end position

The position of the antenna fluorophore attached to the 5'-end of a 29 nucleotide long RNA was determined in order to clarify the contradictions between Andrecka et al. (2008) and Újvári and Luse (2006). Újvári and Luse could covalently cross-link the RNA to the
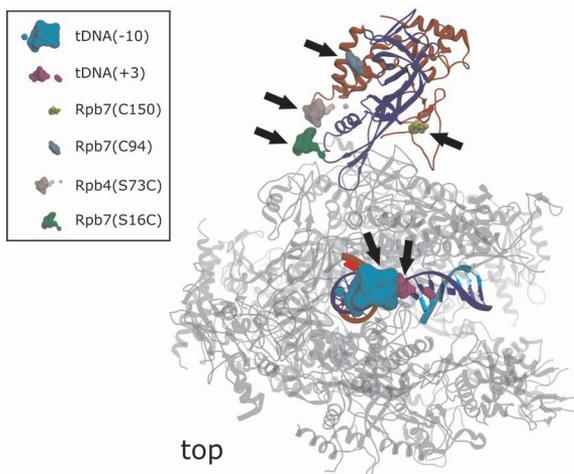
Figure 6.12: The satellite position priors used for inference of the 5' RNA end position are shown as 68% credible volumes in the top view of Pol II (PDB-ID: 1Y1W). The elongation complex is colored according to figure 6.7.

Rpb7 subunit, whereas Andrecka et al., who used the "ancestor" of NPS, observed the RNA at a position distant to Rpb4/7. Since TFIIB was present in the experiments of Újvári and Luse but was not used by Andrecka et al., the experiments analyzed in this work were performed both, in the absence and presence of TFIIB. Like before, all data were measured by Joanna Andrecka and are listed in the appendix II in tables 1 (FRET efficiencies), 2 (isotropic Förster distances) and 3 (fluorescence anisotropies).

In short, Alexa 647 was used as satellite fluorophore and was attached to 6 labeling sites. Two of them (tDNA(-10), tDNA(+3)) were located at the template DNA, the other four (Rpb7(S16C), Rpb4(S73C), Rpb7(C94), Rpb(C150)) were on the Rpb4/7 heterodimer. TMR was used as antenna fluorophore, which was attached to the 5' RNA end (RNA(+29)). The approximated satellite position priors are shown in figure 6.12, and the measurements are summarized in figure 6.13.

In the absence of TFIIB, all measurements resulted in single FRET efficiency populations, with the exception of Rpb4(S73C), which always showed two populations. The latter were attributed to different conformational states of the linker, since the accessible volume of Alexa 647 attached to Rpb4(S73C) was fragmented in two separate regions. Hence, both Rpb4(S73C) FRET efficiencies were used in the inference, effectively as two satellites with a prior of exactly the same form.

In the presence of TFIIB, all FRET efficiency measurements resulted in two populations. The major population ($\sim 80\%$ of molecules) had the same FRET efficiency as measured without TFIIB, and the smaller population ($\sim 20\%$ of molecules) was attributed to Pol II elongation complexes that had TFIIB bound. The recorded FRET efficiency traces did not show dynamic switching between the FRET efficiency populations within the duration of the experiments ($\sim 10\,\mathrm{s}$). The satellite Rpb4(S73C) was not measured because of possible complications in the FRET efficiency assignment, since four populations were expected. Control experiments verified that TFIIB was capable of binding to the elongation complexes (Muschielok et al., 2008, supplementary material).

As before, the antenna position prior was a cuboid of several hundred Ångström side length, but now the antenna was excluded from the Pol II elongation complex interior. In addition, the accessible volume was constrained by the length of the RNA, which was treated as a flexible chain attached to the last ribonucleotide observed in the crystal structure 1Y1W (see figure 3.4a, page 40).
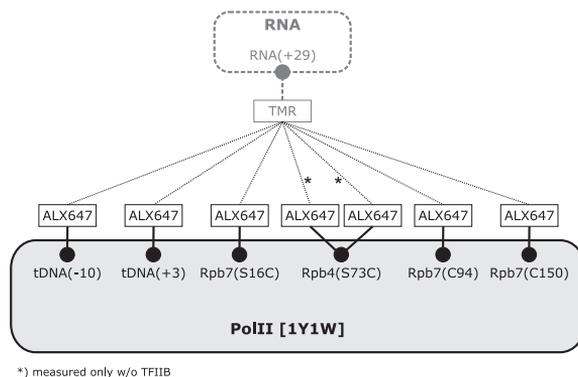
Figure 6.13: Scheme of the FRET network used to infer the 5' RNA end position, both without and with TFIIB. The FRET efficiency measurement from Rpb4(S73C) denoted by stars resulted in two populations already without TFIIB, which were interpreted as two different conformations of the satellite fluorophore linker. Rpb4(S73C) was not used in the experiments with TFIIB as the assignment of four expected peaks in the FRET efficiency histogram would have been too difficult.

The inferred position of the antenna fluorophore in the presence of TFIIB can be seen in figure 6.14. The posterior maximum is located far from the Rpb4/7 heterodimer at $(x, y, z) = (99, -9, -14)$Å with standard deviations in the principal directions of 3, 6 and 9 Å. In the presence of TFIIB, the posterior density is shifted towards Rpb4/7, its maximum is located close to the interface of Pol II core and Rpb4/7 at $(x, y, z) = (138, 15, -23)$Å, with standard deviations in the principal directions of 4, 6 and 15 Å (figure 6.14). A stereo-image of the inferred ANT positions is depicted in figure 6.15.

The measurements did not show any inconsistencies as one can see from the good overlap of posterior and measurement profiles (figures 6.16 and 6.17).

6.2.4 Discussion

It can be concluded from the experiments that the 5'-end of the 29 nucleotide long RNA adapts a position even further away from Rpb4/7 than former measurements by Andrecka et al. (2008) suggested. Covalent cross-linking of RNA and Rpb7 can be ruled out from this position. The deviation of the previously determined position from the result of NPS can be ascribed to different factors. First, in the former analysis only a fraction of the data presented here was analyzed, since only measurements from satellites attached to tDNA(-10), tDNA(+3) and Rpb7(C150) were available. Second, the attachment of satellite fluorophores via carbon chain linkers was not accounted for, but instead, the attachment sites at the protein and the DNA, respectively, were used as satellite positions. Finally, there were no error estimates given and the former calculations were based upon the complete isotropic dynamical averaging model (orientation factor $\kappa^2 = 2/3$).

In the presence of TFIIB, however, the 5'-end of the RNA is most probably located close to the Rpb4/7 heterodimer. Covalent cross-linking of RNA and Rpb7 is thus possible in this position and it is in agreement with Újvári and Luse (2006). The displacement of the nascent RNA by TFIIB is plausible, since the amino-terminal domain of TFIIB binds to the so called dock domain on the Pol II surface (Bushnell et al., 2004). The dock domain is close to the position of the RNA 5'-end in the absence of TFIIB (figure 6.15), and it lies on the path of the nascent RNA (Andrecka et al., 2008; Andrecka, 2009). Therefore,
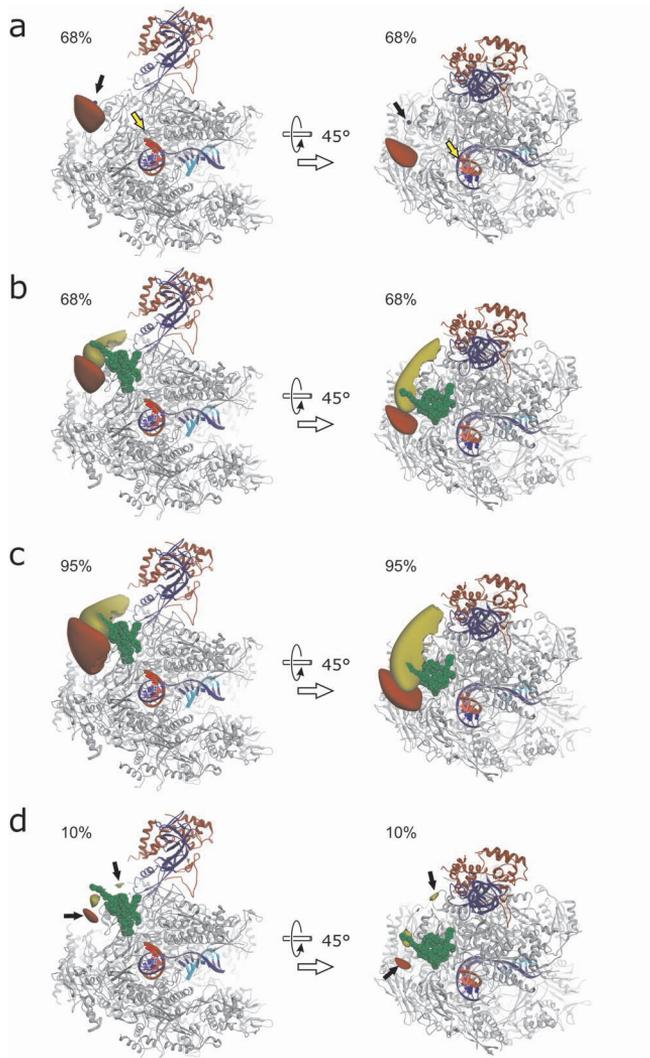
Figure 6.14: 5' RNA end position inference. The Pol II elongation complex (Kettenberger et al., 2004, PDB-ID: 1Y1W) is shown in the top view (left) and a rotated view (right). The colors are according to figure 6.7. (a) shows the 68% credible volume of the inferred ANT position in the absence of TFIIB (red surface) together with the position inferred by Andrecka et al. (2008) (small blue sphere, black arrow). The last base of RNA observed in the X-ray structure is marked with a yellow arrow. In addition, the amino-terminal domain of TFIIB (Bushnell et al., 2004, PDB-ID: 1R5U) is shown as green spheres in (b) together with the 68% credible volume of the inferred ANT position in the presence of TFIIB (yellow surface). (c) and (d) show 95% and 10% credible volumes. In (d) the location of the marginal posterior maxima are marked by arrows.
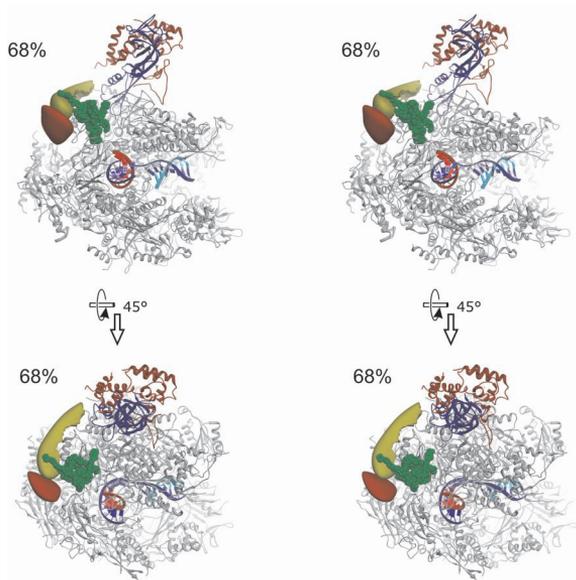
Figure 6.15: Stereo image of the inferred 5' RNA end position. The top view (top) and a rotated view (bottom) of the Pol II elongation complex (Kettenberger et al., 2004, PDB-ID: 1Y1W) is shown. The colors are according to figure 6.7. In addition, the amino-terminal domain of TFIIB (Bushnell et al., 2004, PDB-ID: 1R5U) is shown as green spheres. The 68% credible volumes of the marginal antenna position posterior are shown as red surface (absence of TFIIB) and yellow surface (with TFIIB).
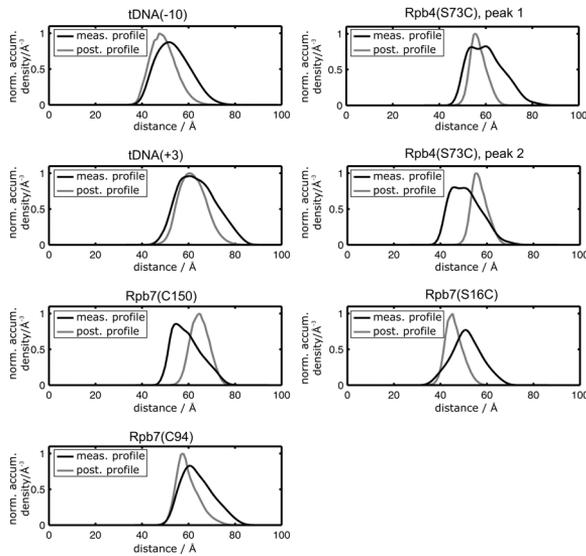
Figure 6.16: Posterior profiles and measurement profiles for the inference of the 5' RNA end position in the absence of TFIIB. In all measurements, there is substantial overlap of the profiles, i.e. there is no inconsistent data.
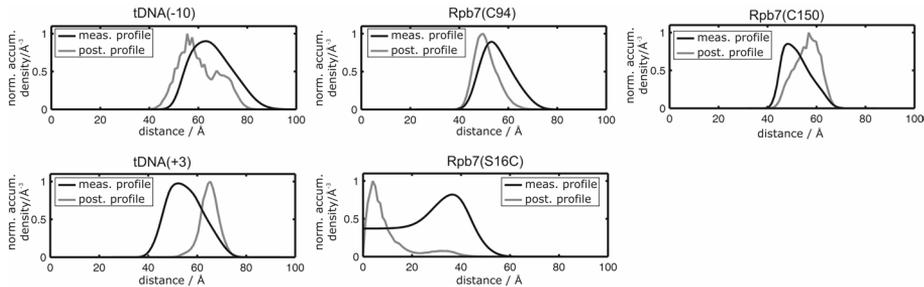
Figure 6.17: Posterior profiles and measurement profiles for the inference of the 5' RNA end position in the presence of TFIIB. In all measurements, there is substantial overlap of the profiles, i.e. there is no inconsistent data.

TFIIB and the nascent RNA compete for the binding site on the dock domain. The contradiction between Andrecka et al. and Újvári and Luse was thus caused by different experimental conditions.

Although the posterior maximum is located close to Rpb4/7, it would be desirable to shrink the uncertainty even more by including more FRET efficiency measurements between new SATs and the ANT, so that the elongated shape of the posterior becomes smaller. This would however be experimentally difficult for the following reasons. First, additional labeling sites on Rpb4/7 would be either too close to the present marginal posterior with the effect that only a maximum distance to a satellite can be measured, or at a position where the posterior density lies at roughly the same distance from the satellite. The same effect can be seen in the measurement profile of Rpb7(S16C) (figure 6.17), which stays at a high probability level even at small distances. Second, new labeling sites on the nucleic acids do not improve the localization accuracy either, since they are located roughly in the same plane as the already present satellites. Third, labeling of the Pol II core at a suitable position is extremely difficult with the techniques used here.

Finally, one can speculate that TFIIB remains bound to the elongation complex during elongation of the RNA and thus helps to direct the RNA towards the Rpb4/7 heterodimer.

After demonstrating that the Nano-Positioning System based on the *position - Förster distance* NPS model is a versatile tool to handle structural biology questions, its refined version, the *position - orientation* NPS model, will be motivated and compared to the *position - Förster distance* model in the next section.

6.3 Simulation II – Improvement of localization accuracy

In the first part of this section, simulations will be shown that demonstrate the minimum localization accuracy of FRET measurements if one only considers the uncertainties caused by the unknown orientation of the average transition dipole moments.

In the second part, test calculations performed with the *position - orientation* NPS model will be shown. The *position - orientation* model will be compared to the *position - Förster distance* model based of these calculations. The improvement of the localization accuracy expected from the *position - orientation* model will be demonstrated, and the influence of the data on the localization accuracy will be discussed. In addition, the performance of the Monte Carlo algorithm utilized in the calculations will be presented.
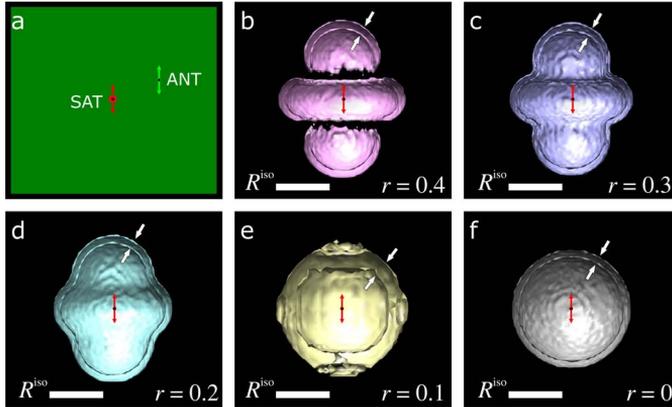
Figure 6.18: The effect of fluorescence anisotropy on the localization given known fluorophore orientations. (a) The satellite and antenna average transition dipole orientations (red and green double arrows) are known to be aligned along the $z-$axis. The satellite position prior (red) is concentrated on a point in the center of the coordinate system, whereas the antenna position prior is flat in a cube around the satellite. The panels (b-f) show 90% credible volumes of the inferred antenna position density for varying residual fluorescence anisotropies $r =$0.4, 0.3, 0.2, 0.1 and 0, respectively. The anterior part of the density was removed in order to show the thickness of the credible volumes (white arrows).

6.3.1 Effects of anisotropy and orientation on the localization uncertainty

To demonstrate the minimum localization uncertainty caused only by the FRET efficiency measurement error, an analysis with one simulated FRET pair was carried out with the *position - orientation* NPS model. The FRET pair consisted of a satellite fluorophore, s, and and antenna fluorophore, a, of equal residual fluorescence anisotropies, $r_{\infty,s} = r_{\infty,a} = r$. A FRET efficiency value of $E_{as} = 0.5 \pm 0.02$ was used as the only data. Different informative average transition dipole moment orientation priors were used in the simulations, in contrast to the non-informative priors presented in section 3.2.3. Additionally, the satellite fluorophore possessed an infinitely small accessible volume to rule out that the satellite position uncertainty propagates into the localization error. The antenna position prior was flat within a cube of several isotropic Förster distances side length.

In the first calculation, the average transition dipole moments of both fluorophores were known to be oriented along the $z-$axis of the coordinate system (figure 6.18a), which is equivalent to the prior

$$p\left(\mathbf{\Omega}_{s/a}|I\right) \propto \frac{1}{\pi}\delta(\cos\theta_{s/a} - 1), \tag{6.3.1}$$

where $\delta(.)$ denotes the Dirac-delta distribution.

Now, the fluorescence anisotropies were varied from 0.4 to 0 to demonstrate the effect on the shape of the inferred antenna positions. The calculation was performed using nested sampling with 120 objects, followed by Metropolis sampling of the posterior, which resulted in $1 \cdot 10^5$ independent samples. The 90% credible intervals of the marginal antenna position posterior are shown in figure 6.18b-f.

The computed densities exhibit shell-like structures that are highly anisotropic when $r = 0.4$ (figure 6.18b) and gradually approaches the form of a sphere when the fluorescence anisotropy drops to 0 (figure 6.18f). In the latter case, the density has sphere symmetry,

whereas in the scenarios with $r > 0$ the density belongs to the $C_{\infty h}$ point group, i.e. has rotation symmetry along the $z - $axis and mirror symmetry at the $(x, y) - $plane (figure 6.18b-e). The thickness of the 90% credible interval is the same for fluorescence anisotropies $r > 0.1$ and $r = 0$. When $r = 0.1$, the thickness is larger.

In a second calculation, the residual fluorescence anisotropy was kept constant at $r = 0.4$, and the orientation of the antenna average transition dipole moment was rotated gradually to be perpendicular to the antenna average transition dipole orientation. This corresponds to the following prior:

$$p\left(\mathbf{\Omega}_a | I\right) = \delta(\cos\theta_a - \cos\theta_{a,0})\delta(\phi_a), \qquad (6.3.2)$$

where $\theta_{a,0}$ takes the values $\theta_{a,0} = \pi/6, \pi/3, \pi/2$. As in the first calculation, nested sampling with 120 objects was used, followed by Metropolis sampling of the posterior, which resulted in $1 \cdot 10^5$ independent samples. As a result, the inferred antenna position density transformed gradually into a four-lobed shape that belongs to the D_{4h} point group (figure 6.19a-d).

When less informative orientation priors are used, e.g. when the antenna is known to be oriented in the $(x, y) - $plane,

$$p\left(\mathbf{\Omega}_a | I\right) = \frac{1}{\pi}\delta(\cos\theta_a), \qquad (6.3.3)$$

or when no information is present, which is the most realistic case, the thickness of the antenna credible volume increases (figure 6.19e,f). When there is also less knowledge about the satellite average transition dipole orientation, the thickness increases even more (figure 6.19g), and finally sphere-symmetry is regained (figure 6.19h).

6.3.2 Discussion

In the first calculation, the fluorescence anisotropy was varied, the average transition dipole orientations were known, and the satellite position was known with high accuracy. The simulation indicates that the localization limited by the FRET efficiency measurement error is roughly independent from the fluorescence anisotropy as long as $r > 0.1$. Anisotropy values below 0.1 are accompanied by an increased localization uncertainty caused by the ambiguity of the sign of the average axial depolarizations (equation (2.3.6)). This increase in uncertainty is reduced again when the fluorescence anisotropy approaches 0.

Though the shapes of the posterior are far less symmetrical compared to a sphere, they could be used to infer fluorophore positions equally well as in the case of vanishing fluorescence anisotropy. Yet, this would be applicable only when the absolute orientation of the average transition dipole moments would be known.

The second simulation shows that, given high fluorescence anisotropy values, the localization uncertainty increases dramatically, which has been already shown in section 6.1.1. Now, the estimated antenna position posterior densities can be interpreted as an average over a continuum of shapes similar to those shown in figure 6.19a-d. The smaller the fluorescence anisotropy, the more resemble these shapes a sphere (figure 6.18b-f), and the more the averaged density is concentrated on the surface of a sphere.

Beside the fluorophore positions, the *position - orientation* NPS model allows to infer the average transition dipole moment orientations from the data. That information, in turn, can improve the localization accuracy, as shown above. In the following, the conditions will be tested that allow to reduce the localization uncertainty.
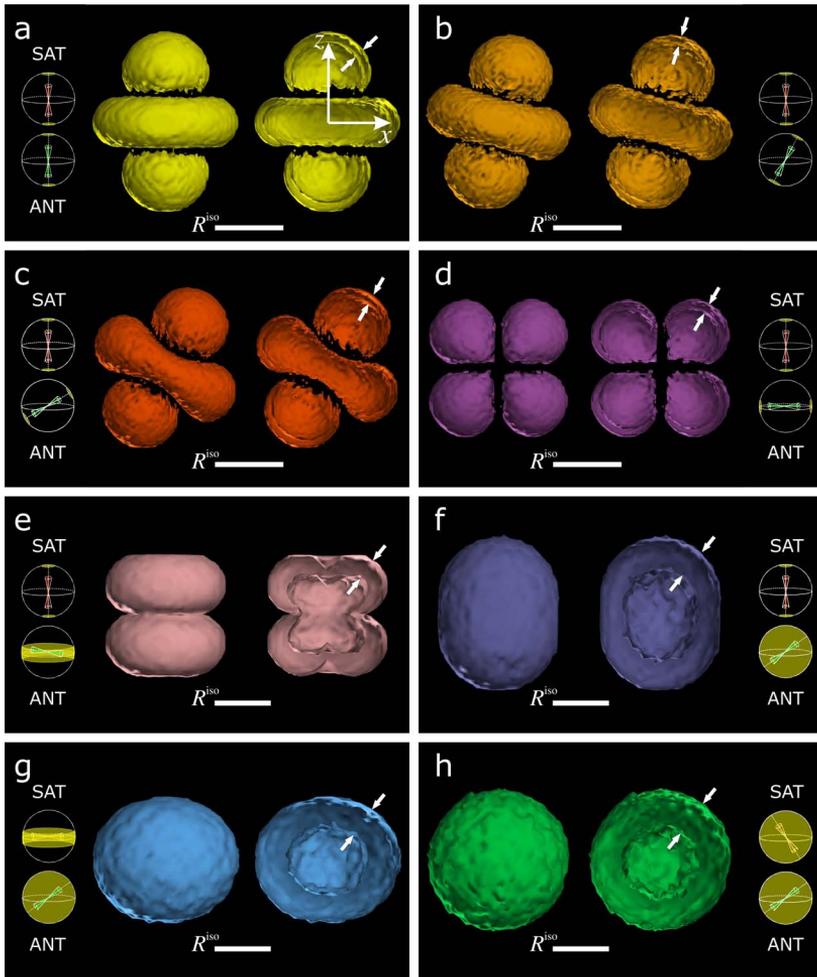
Figure 6.19: Effect of average transition dipole moment orientation prior on the estimated ANT position. The surface of the 90% credible volume of the ANT position is shown as exterior view and as a section (left and right parts of each panel, respectively). In the section view the shell-like structure can be seen. The estimate in (a) has axial symmetry and was already shown in figure 6.18b. The fluorescence anisotropies of SAT and ANT were 0.4 throughout. In (a-d) the average transition dipole orientations of both fluorophores are known, and the angle between them is changed (see insets, yellow: possible average transition dipole moment orientations). In (e-f) less informative priors are used. In particular, only the satellite orientation is known accurately in (e,f), while the antenna is either known to be oriented in the (x, y) − plane (e) or its orientation is completely unknown (f). In the subfigures (g,h) the orientation information is decreased for the satellite as well. The thickness of the 90% credible volumes is indicated by two arrows.

6.3.3 Position‑orientation NPS model test calculations

Here, more test calculations performed with the *position‑orientation* NPS model will be presented. The objective was to determine under which condition the improvement of localization accuracy expected from the *position‑orientation* model analysis can be observed, i.e. when the independent Förster distance approximation breaks down. This is of interest, since the calculation with the *position‑orientation* model takes much more computational time compared to the calculation with the *position‑Förster distance* model. In addition, the effect of including FRET anisotropy data in the inference process will be examined. Finally, the numerical performance of the implementation of nested sampling will be briefly discussed.

FRET network structure

The studied FRET network consisted of 7 antennas (ANT 1-7) and 7 satellites (SAT 1-7), which is similar to the network used to infer the nontemplate DNA position in the Pol II elongation complex (section 6.4). In some of the calculations, two additional satellites (SAT 8 and 9) were used to increase the amount of data and hence to improve the localization accuracy, which is also often done in experiments when the resulting antenna position densities are too large to be accurately localized. The fluorophore positions (figure 6.20) and average transition dipole moment orientations are listed in table 8 (appendix II).

All test calculations were based on synthetic data generated as described in section 5.3.4. The data is listed in the appendix II in tables 8 (fluorescence anisotropies), 9 (FRET efficiencies), 10 (isotropic Förster distances) and 11 (FRET anisotropies). Briefly, 49 FRET measurements (FRET efficiencies as well as FRET anisotropies) between ANT 1-7 and SAT 1-7, as well as 6 measurements between ANT 2-7 and SAT 8 and 9 were simulated. The residual fluorescence anisotropies of all fluorophores were in the range between 0.15 and 0.32, which consistent with the fluorescence anisotropies observed in the experiments with Pol II.

In all but the last analysis of the synthetic data the size of the accessible volumes of the satellites was vanishingly small to study orientation effects only, while the volumes accessible to the antennas were flat within a large cube around the satellites. In contrast, in the last analysis, finite-sized satellite position priors were used that are comparable in size with those simulated in the Pol II experiments. The average transition dipole moment orientation priors of all fluorophores were flat and uninformative (equation (3.2.8)).

Figure 6.20: Fluorophore positions (synthetic data). The positions of the fluorophores used to compute synthetic data are shown as spheres relative to a hypothetical macromolecule (gray surface). The ANTs, numbered 1-7, are shown in color, the SATs, numbered 1-7, in gray. The additional SATs 8 and 9 used to extend the measurement network are shown in blueish-gray and are highlighted by arrows.
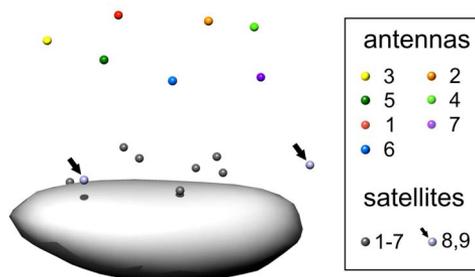
Antenna positions inferred with the position‑orientation NPS

To compare the *position‑Förster distance* NPS model with the *position‑orientation* NPS model, the FRET efficiency data was separated into seven subsets of measurements between each single ANT fluorophore and the SATs 1-7. Then, in the scenario referred to as *separate analysis*, each data subset was analyzed with both models. The analysis resulted in very similar marginal antenna position posteriors (figure 6.21a,b). In particular, the position densities of all but ANT 1 and 6 were elongated, and many had a second probability maximum at the "wrong" side of the hypothetical macromolecule[1]. Beside the large position uncertainty, most of the antenna positions were estimated well, except ANT 2. Its simulated position was rather far from the estimated density, since it was located on the surface of the 95% credible interval (*position‑Förster distance* model) or was found even outside of it (*position‑orientation* model).

In the next scenario, the *global analysis*, all 49 measurements between ANT 1-7 and SAT 1-7 were analyzed at once with the *position‑orientation* model (figure 6.21c). As a result, some of the estimated antenna position densities were smaller as compared to the separate analysis, and all second maxima contained now less probability than before. Also the simulated position of ANT 2 was found inside of the 95% credible interval.

The separate and global analysis with the *position‑orientation* model were repeated after including the additional measurements to the satellites SAT 8 and 9 into the analyzed data set. These scenarios will be referred to as *extended network*. All marginal antenna position densities did shrink when compared to the previous FRET network with only 7 satellites (figure 6.21d,e), and the difference between separate and global analysis was more pronounced. Yet, the simulated positions of many antennas were still at asymmetrical positions in the posterior density.

When FRET anisotropy data was used in addition to FRET efficiency data (section 3.2.2), an overwhelming decrease in localization uncertainty was observed even when the smaller FRET network was analyzed (figure 6.21f). Now, all estimated densities were nicely centered around the simulated antenna positions.

In the last calculations, the influence of the imprecisely known satellite positions was studied. To this end, the satellite position priors were changed to be flat within small cubes of 8 Å side length centered around the simulated position (figure 6.21g). The inferred antenna position densities spread slightly, but the localization uncertainty was small when compared to the uncertainties obtained from FRET efficiency data alone, and the estimated densities stayed nicely centered around the simulated antenna positions.

Average transition dipole moment orientations inferred with the position‑orientation NPS

When the *position‑orientation* model was used, there is more information contained in the analysis results than just the fluorophore positions shown above. The samples drawn from the posterior carry information about the average transition dipole moment orientations, which will be studied next.

The 68% and 95% credible intervals of the marginal average transition dipole moment orientation of a typical antenna fluorophore are shown in figure 6.22a to give an impression of the inference accuracy of separate analysis. Note that all positions $-\cos\theta_i = \pm 1$ are physically equivalent, and that the parameter space at $\phi_i = 0$ and $\phi_i = \pi$ is connected by $(-\cos\theta_i, \phi_i = 0) \equiv (+\cos\theta_i, \phi_i = \pi)$. This is caused by the fact that flipping the direction of any average transition dipole moment does not change the likelihood value.

---

[1]In a real application, such ambiguities can be often removed by considering the space the antenna can reach, which is easily encoded in the corresponding position prior
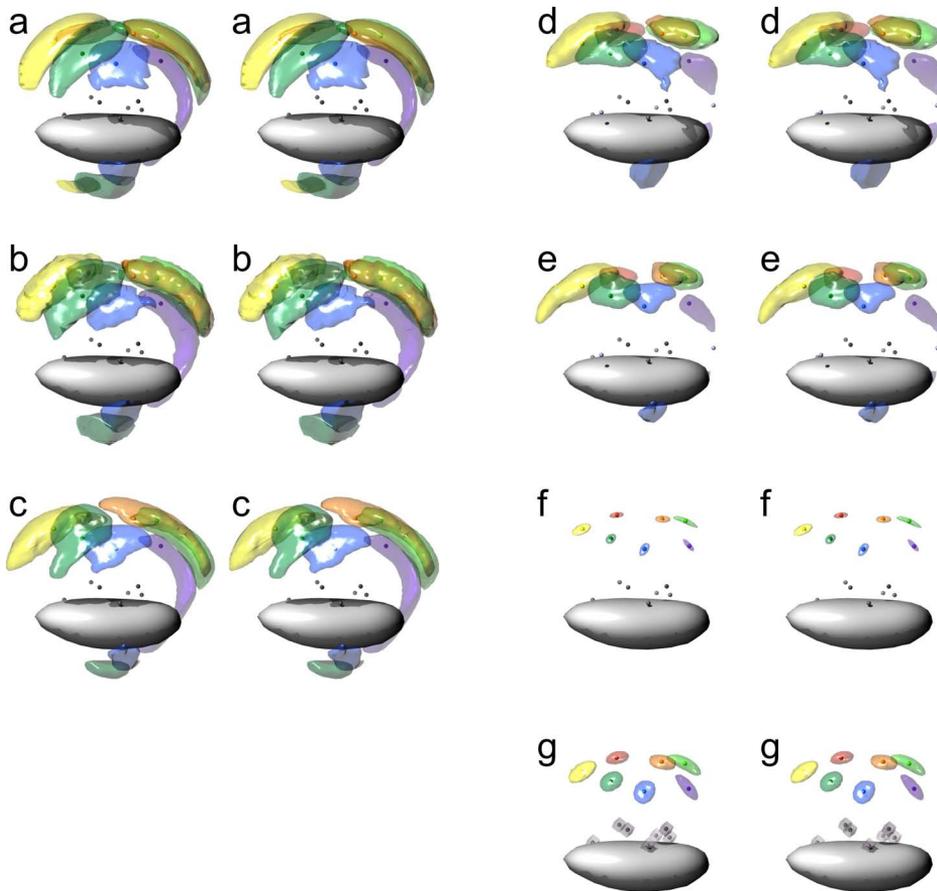
Figure 6.21: Comparison of analysis methods. Stereo images of 95% credible volumes of the marginal antenna position priors are shown as surfaces and were colored in the same way like the antennas in figure 6.20. In (a) and (b) the results of the separate analysis are shown, computed with the *position - Förster distance* model and the *position - orientation* model, respectively. The global analysis of the small FRET network is shown in (c), whereas the separate and global analysis of the extended network is depicted in (d) and (e), respectively. In (f) and (g), the smaller network was analyzed like in (a-c), but FRET anisotropy data was used in addition to FRET efficiency data. (a-f) were computed with infinitely small satellite position priors, while finite-sized accessible volumes were used in (g).

While the separate analysis of the extended network decreased the uncertainties only slightly (figure 6.22b), global analysis improved the average transition dipole moment orientation accuracy stronger (figure 6.23a,b). Compared to the accuracy of localization (see marginal antenna position posteriors, figure 6.21), the accuracy of the inferred average transition dipole moment orientations was much lower when only FRET efficiency data was analyzed, as the marginal posterior density was very diffuse.

In the extended network the presence of more data improved not only the inference of fluorophore positions but also the accuracy to determine average transition dipole moment (figure 6.23c,d). The analysis of FRET anisotropy data had an even stronger effect (figure 6.23e,f), so that all average transition dipole moment orientations were computed correctly. The reason therefor was the small error of FRET anisotropy measurements ($\Delta A_{ij} = 0.01$), which was sufficient to detect orientation effects expected to be in the order of $(r_{i,\infty} r_{j,\infty})^{1/2}$ (equations (2.3.10), (2.3.9) and (2.3.7)). The introduction of finite-sized satellite position priors increased the orientation uncertainty only slightly (figure 6.23g,h).

Overall quantification of uncertainties and evidences

In order to summarize the overall quality of the inference, the relative localization and orientation uncertainties were computed (figure 6.24). To this end, the approximate localization and orientation accuracies (section 5.3.7) were normalized by the accuracy of the separate analysis and averaged over all ANTs.

The *position - Förster distance* NPS model has a slightly larger average localization uncertainty when compared to the separate analysis using the *position - orientation* NPS model. The other average localization uncertainties follow the trend that was described in words in the previous two subsections. In particular, the uncertainty drops by almost an order of magnitude when FRET anisotropy data is analyzed in addition to FRET efficiency data.

The same applies to the orientation uncertainty. Both accuracies increase upon in-
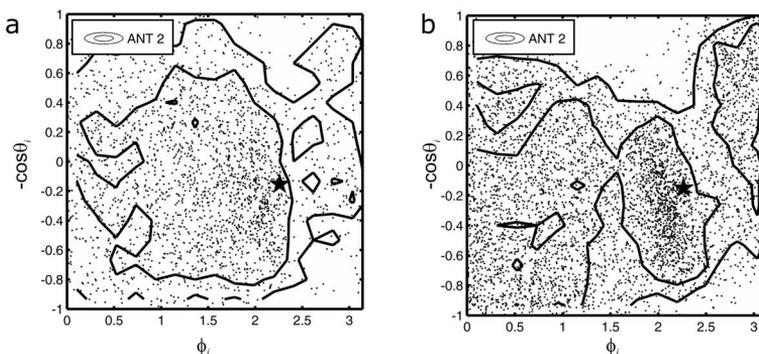
Figure 6.22: Typical marginal average transition dipole moment orientation posterior densities computed by *separate analysis*. The density of a typical antenna fluorophore, here ANT 2, is shown as scatter plot and credible intervals. 68% and 95% correspond to the inner and outer lines, respectively. For comparison, the simulated average transition dipole moment orientation underlying the data is displayed as a star. In (a) the measurements of the satellites SAT 1-7 were used to infer the orientation, whereas (b) was computed based on SAT 1-9 (extended FRET network).
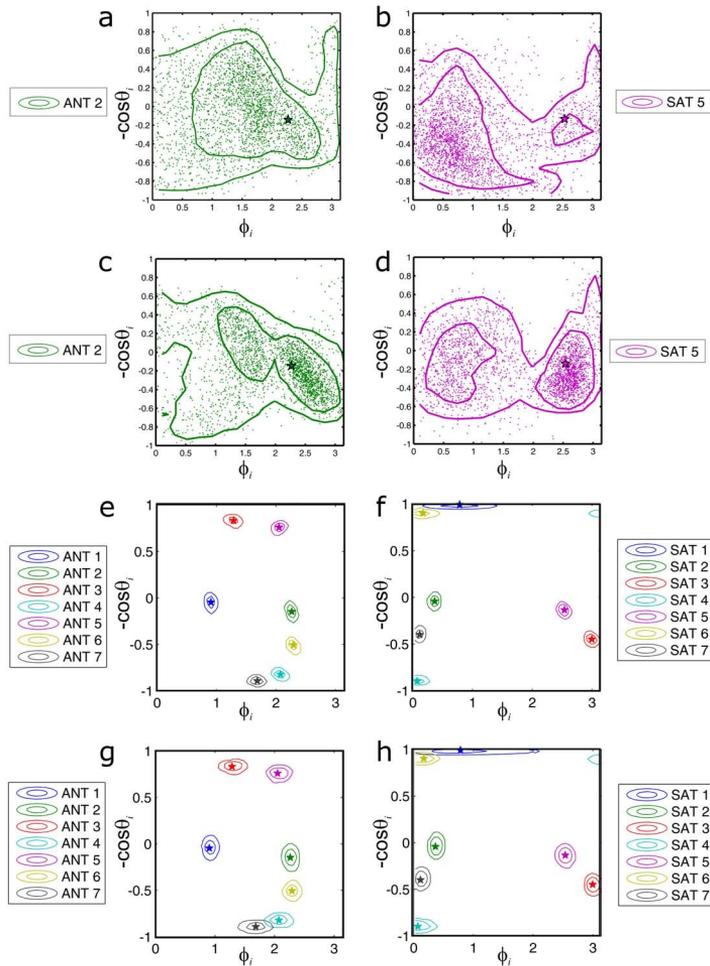
Figure 6.23: Marginal average transition dipole moment orientations in the *global analysis*. 68% and 95% credible intervals (inner and outer lines, respectively) of the marginal average transition dipole moment orientation posterior of antennas (left column) and satellites (right column) computed by global analysis are shown. For comparison, the simulated orientations underlying the data are displayed as stars. The following scenarios are shown: (a,b) FRET efficiency data only, (c,d) extended network, FRET efficiency data only, (e,f) FRET efficiency and anisotropy data, (g,h) FRET efficiency and anisotropy data with finite-sized SAT position priors. In the presence of FRET anisotropy data (e-h), the fluorophore orientations are inferred with high accuracy, while the accuracy is lower when only FRET efficiency data is available (a-d). For clarity reasons the density of only one fluorophore is displayed both as credible intervals and scatter plot in the latter scenario. In panels (a-d), the assignment of the fluorophore orientations is better in the extended FRET network (c,d) than in the smaller FRET network with 7 ANTs and 7 SATs (a,b).
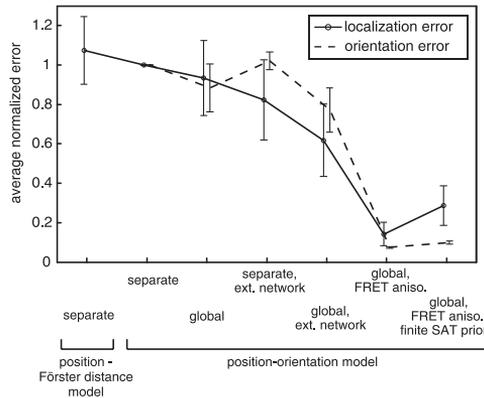
Figure 6.24: Average relative uncertainties. The position and average transition dipole moment orientation uncertainty of each antenna was divided by the respective uncertainty value obtained in the separate analysis. Averaging of these relative uncertainties over the ANTs resulted in the average relative uncertainty, which are shown together with the standard deviation (whiskers) for each scenario. The average relative uncertainties of positions and orientations behave similarly. The *position - Förster distance* model shows a slightly higher position uncertainty compared to the separate analysis with the *position - orientation* model. Compared to separate analysis, global analysis has a stronger impact when more data is present (extended network). When FRET anisotropy is analyzed, the uncertainties decrease drastically to values limited probably by data quality, whereas the analysis of FRET efficiency w/o FRET anisotropy suffers from lack of orientation information. Finite-sized SAT position priors have a stronger effect on the position error than on the orientation error.

troduction of finite-sized SAT position priors, but the effect is small compared to the accuracy gain caused by additional analysis of FRET anisotropy measurements.

Supplemental separate analyses of FRET efficiency data only were performed with finite-sized satellite position priors (data not shown). This had only a minor effect on the average normalized localization and orientation uncertainties, which increased by only 11% and 2.6%, respectively.

In principle, the evidence, Z, calculated by nested sampling (equation (2.5.19)) can be used in model selection (see sections 2.4.5 and 6.7.5). Evidences of various global analyses are listed in table 6.2 together with the information gain, H (equation (2.5.19)), which reflects the degree of dissimilarity between posterior and prior.

To determine whether nested sampling could find the likelihood maximum, the largest sampled likelihood values were compared to the maximum possible likelihood values (table 6.2). The latter were known in the test calculations, but are usually not known when experimental data is used. In general, nested sampling converged closer to the likelihood maximum when FRET efficiency data was analyzed together with FRET anisotropy data.

Performance of the sampling engine

The calculations based on the *position - orientation* model shown above were performed with nested sampling based on Markov chain Monte Carlo (sections 2.5.1 and 5.5). The number of objects used in the different scenarios is listed in table 6.3 together with the number of free parameters. A detailed listing of Monte Carlo settings can be found in section 5.5.7. In order to study the performance of the sampling engine (section 5.5), the nested sampling used in the global analysis of FRET efficiency and anisotropy data

calculation	$\ln Z$	H/bit	$\ln L_{\max}^{\mathrm{sampl}}$	$\ln L_{\max}$	$L_{\max}/L_{\max}^{\mathrm{sampl}}$
FRET efficiency only	8.3 ± 0.1	164	144.2	146.7	12.2
FRET efficiency only, extended network	9.5 ± 0.3	185	162.5	164.6	8.2
FRET efficiency and anisotropy	133.5 ± 0.4	244	326.5	327.3	2.2
FRET efficiency and anisotropy, finite SAT prior	125.0 ± 0.3	253	327.1	327.3	1.2

Table 6.2: Evidences, information and likelihoods of global NPS calculations. The natural logarithm of the evidence ($\ln Z$), the information (H), the natural logarithms of the maximum likelihood achieved by nested sampling ($\ln L_{\max}^{\mathrm{sampl}}$), the analytically obtained likelihood ($\ln L_{\max}$) as well as their ratio are listed for the global NPS calculations.

scenario	sep., E	glob., E	sep., E, ext. netw.	glob., E, ext. netw.	glob., $E\&A$	glob., $E\&A$, fin. sat. pos.
objects	500^*, $1000^\dagger$, $2000^\ddagger$	10000	1000	2000	1000	1000
num. of free parameters	19	49	21	53	49	70

$*$) ANT 1-4 and 7 $\qquad$ $\dagger$) ANT 6 $\qquad$ $\ddagger$) ANT 5

Table 6.3: Number of sampling objects and free parameters in the different scenarios

with finite-sized satellite priors will be discussed in detail. This particular calculation was chosen since it has the most free parameters.

During the progress of nested sampling, the objects were compressed more and more to parameter space regions of high likelihood values. This is demonstrated for two fluorophores (ANT 1 and SAT 5) in figure 6.25 showing the marginal position of the samples (i.e. objects with the lowest likelihood) produced in each nested sampling iteration. Already in an early state of the calculation (iteration $\sim 25 \cdot 10^3$), the ANT positions were roughly determined. They were refined as soon as the correct average transition dipole orientations of all fluorophores were found (iteration $\sim 100 \cdot 10^3$). The SAT positions do not show any noticeable gain in position accuracy since their priors were already very informative.

Each iteration of the nested sampling algorithm a new sample was drawn from the constrained prior by Markov chain Monte Carlo. Two types of moves were used in the Monte Carlo, a local move that ensured the short-range mobility of the Markov chain, and a non-local move that should allow the chain to move between several not interconnected posterior modes. As the shape of the constrained prior changed dramatically during the progress of nested sampling, the Monte Carlo control parameters like for example the length of the Markov chain (also called length of the random walk in the following) needed to be adjusted to guarantee an efficient exploration of the constrained prior (section 5.5.5). This will be shown in the following.

The adaptation of the local move length scales s_i is shown in figure 6.26 for the parameters describing ANT 1 and SAT 5. After the initial collapse of ANT positions, the local move length scales shrink approximately exponentially in all parameters.

Other quantities beside the local move length scales were recorded as well during the course of nested sampling. The logarithm of the likelihood approaches asymptotically the maximum, while the sample weights reach a maximum around iteration $j_{\max} = 175 \cdot 10^3$ and decrease thereafter until the termination criterion is satisfied, here at iteration
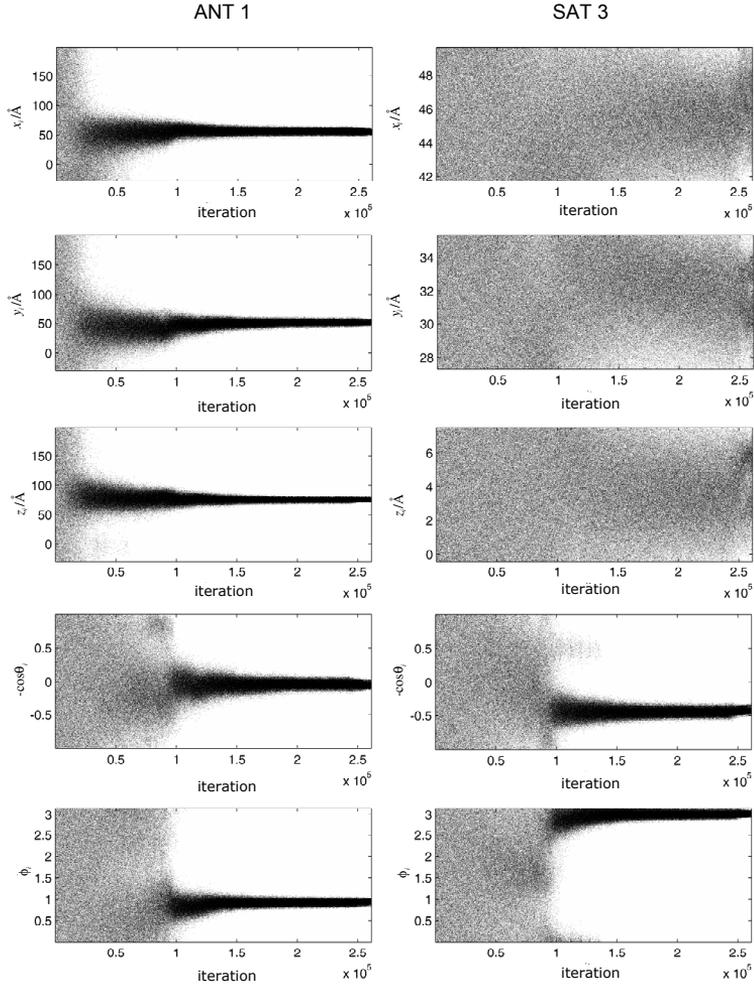
Figure 6.25: Samples produced during nested sampling. The samples produced in each nested sampling iteration are shown as dots for a selected set of coordinates in the multi-dimensional parameter space. These coordinates are the position and orientation of ANT 1 (left column) and SAT 5 (right column). The collection of objects used in nested sampling is compressed in two phases. In the first phase (iteration $\sim 25 \cdot 10^3$), the rough ANT position is determined. In the second phase (iteration $\sim 100 \cdot 10^3$), the average fluorophore transition dipole moment orientations of both ANT and SAT are found, and the ANT position is refined.
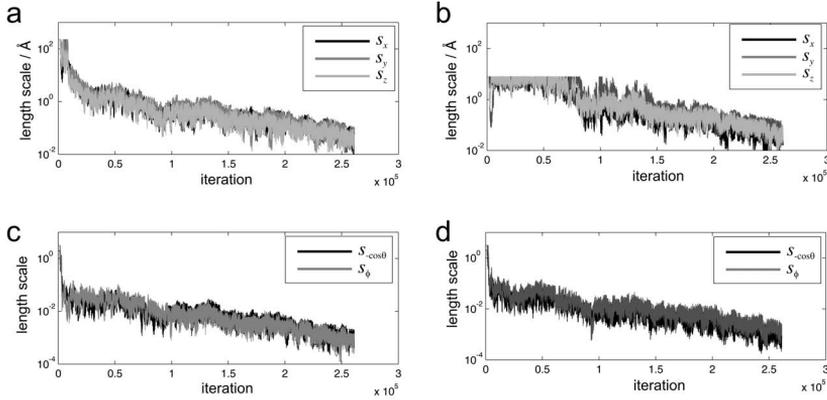
Figure 6.26: Adaptation of local move length scales. The local move length scales, s_i, are shown for a subset of parameters, namely the positions (a, b) and orientations (c, d) of ANT 1 (a, c) and SAT 5 (b, d).

$261 \cdot 10^3$ (figure 6.27a,b). This was in agreement with the chosen termination criterion $j > 1.5 \times MH \approx j_{max}$, where j is the iteration number, M the number of objects (here $M = 1000$), and H the information in nats (subsection 2.5.2 page 31). The length of the random walk needed to create an approximately independent sample from the constrained prior was adapted (section 5.5.5) and is shown in figure 6.27c. Also the ratio of accepted and rejected local moves was adapted to minimize the length of the random walk (figure 6.27d).

Non-local moves were used to improve the exploration of the constrained prior. The probability to use a non-local move as well as the number of retries, which is a parameter of the move (section 5.5.4), were found almost always at their maximum values of 0.1 and 100, respectively (traces not shown). As an adjustment to smaller values has been observed in other calculations (not shown), the adaptation of these Monte Carlo parameters apparently depends on the inference problem. The average number of successful non-local moves and their average efficiency was very low here (figure 6.27e,f), but in certain regions several successful moves per trace were typically observed (figure 6.28).
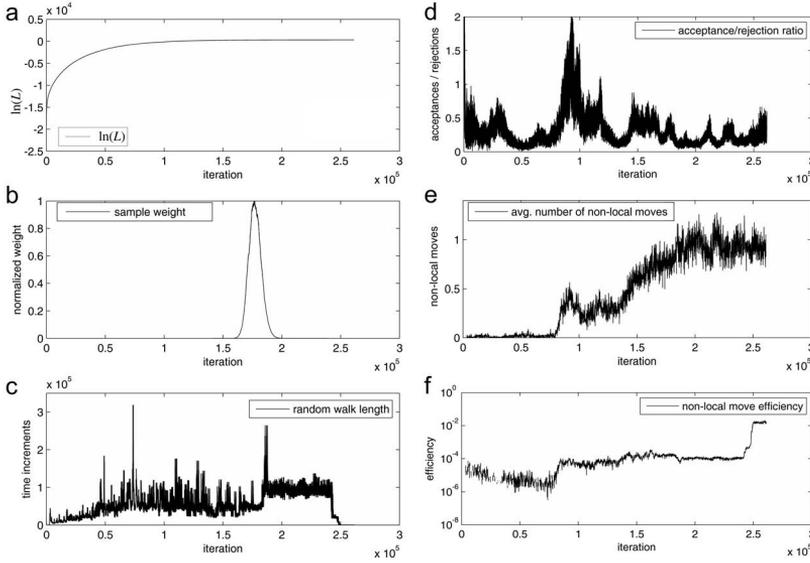
Figure 6.27: Nested sampling and Monte Carlo quantities. The logarithm of the likelihood (a) and the weights (b) of the samples produced during nested sampling are shown as functions of the nested sampling iteration. Several other Monte Carlo parameters are shown in the other subfigures: (c) Length of the random walk needed to compute an approximately independent sample. (d) Ratio of local move acceptances and rejections averaged over 100 iterations. (e) Number of non-local moves per random walk averaged over 100 iterations. (f) Efficiency of non-local moves averaged over 100 iterations.
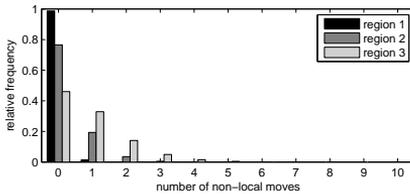
Figure 6.28: Relative frequency of successful non-local moves per random walk. Data from three different regions during nested sampling is shown: iteration 1–$70 \cdot 10^3$ (region 1), iteration $80 \cdot 10^3$–$130 \cdot 10^3$ (region 2), iteration $150 \cdot 10^3$–$230 \cdot 10^3$ (region 3).

6.3.4 Discussion

Inference results

The comparison of both NPS models on the basis of the separate analysis of FRET efficiency data supports the approximations used in the *position - Förster distance* NPS model further. In contrast to the separate analysis, more accurate position and orientation estimates are achieved when the data is analyzed globally. This is obvious, since the measurements of FRET efficiency between a particular antenna and a set of satellites can be used to estimate the positions and average transition dipole orientations of the satellites, and might even introduce correlations between different satellites. In the spirit of Bayesian data analysis, this information should be used as prior in the next analysis. In fact, this was not done in the separate analysis, as the same uninformative prior was used each time. In contrast, the global analysis can profit from the synergy effects because all data is analyzed at once.

It is even possible to compute rough estimates of the average transition dipole orientations by globally analyzing FRET efficiency data, as can be seen in the extended FRET network scenario. Yet, in general, the fluorophore positions were determined better than the average transition dipole moment orientations. One reason therefor might be the different sensitivity of the FRET efficiency on variations of fluorophore distances and orientations in the range of the fluorescence anisotropies analyzed. Another reason is definitely the main difference between position and orientation priors: while informative priors were used for all satellite positions (informative in the sense of a localized density), the average transition dipole moment orientation priors were always completely uninformative.

A trivial result of the test calculations was that the fluorophore positions and their average transition dipole moment orientations are determined more precisely when more data is analyzed. At a second glance, this result is amazing from the point of view of traditional data analysis, since for small FRET networks, it might be that each new satellite introduces more unknown model parameters (especially average transition dipole moment orientations) than data. This "paradox" will be discussed in more detail later in section 6.7.4.

The most striking result was that FRET anisotropy data analyzed in addition to FRET efficiency data drastically reduces the localization uncertainty and the uncertainty in the estimated average transition dipole moment orientations. The fact that the localization uncertainty is only slightly increased when finite-sized satellite position priors are used proves that, under the chosen conditions, the major source of localization uncertainty is indeed the a priori unknown average transition dipole moment orientation. One could argue that the FRET anisotropy measurement errors of 0.01 used here are overly optimistic, since it is difficult to measure FRET anisotropy accurately when the FRET efficiency is low ($\sim 20\%$ of the simulated FRET efficiencies were smaller than 0.1). Nonetheless, one would try to place satellites in a way, so that moderate FRET efficiencies are observed in a realistic experiment, and thus FRET anisotropy is observable as well.

Eventually, it follows from the definition of the likelihood (equation (3.2.5)) and the FRET anisotropy (equations (2.3.10), (2.3.9) and (2.3.7)) that the ability to determine the average transition dipole moment orientations depends on whether the FRET anisotropy measurement error is notably smaller than the range in which the FRET anisotropies are expected to vary. This range is given by 3/2 of the geometric mean of the respective fluorescence anisotropies and in the calculations shown here it exceeds the FRET anisotropy measurement error by an order of magnitude.

Although the measurement of FRET anisotropy is promising, one should not forget

that when only few informative measurements were present, e.g. only the FRET efficiency data of a small network was analyzed, the simulated antenna positions were often found at low probability density values of the marginal posterior, and many of the densities were elongated. This finding was even more pronounced in the average transition dipole moment orientations, which might have several reasons. One reason (probably easiest to fight but most difficult to prove) is the convergence of the nested sampler. Here, it was possible to check how far the sampling had proceeded, since the maximum possible posterior value was known. As the prior was flat and the experimental errors that are usually modeled by the width of the likelihood were not simulated, the posterior maximum was located exactly at the simulated fluorophore positions and average transition dipole moment orientations. Thus, the maximum possible likelihood value, $L_{\max}$, was a suitable indicator for the progress of nested sampling.

The worst convergence, indicated by the largest ratio of the maximum possible likelihood and the maximum sampled likelihood, $L_{\max}^{\mathrm{sampl}}$, was found indeed in the case of globally analyzed FRET efficiency data ($L_{\max}/L_{\max}^{\mathrm{sampl}} = 12.2$, table 6.2), followed by the globally analyzed extended network ($L_{\max}/L_{\max}^{\mathrm{sampl}} = 8.2$). In the presence of FRET anisotropy data, the ratio was only 2.2, and when the finite satellite position priors were used, it was only 1.2. This might indicate that, when FRET anisotropy data helps to fix the average transition dipole orientations, the posterior has a shape that can be sampled more easily.

Yet, even convergence does not guarantee that the "true" position will have a high probability density in the marginal posterior. Another reason for the peripheral location of the simulated positions and average transition dipole moment orientations might be a strange shape of the posterior. After marginalization, the posterior maximum might be masked by many suboptimal solutions, especially when they are asymmetrically distributed around the maximum. This effect will be discussed in detail in section 6.7.3.

At last, there is evidence that supports a strange shape of the posterior, which is that even in the *position - Förster distance* model calculations the simulated positions were found often at small marginal posterior values. Since the *position - Förster distance* model works without any complicated sampling scheme and both models produce nearly the same results, there is no convergence problem in the separate analysis, but instead, the posterior must have a peculiar shape. Of course, this argument does not rule out convergence problems in the global analysis, but it seems that a straight-forward interpretation of the marginal position densities is problematic.

Since these effects were notably smaller in the test calculations when more data was available, one should, from the practical point of view, shape the marginal antenna position posteriors as compact as possible by measurements to additional satellites placed in adequate positions. In that case, the posterior is easier to sample, which will improve the convergence properties and speed up the calculation. Otherwise, when it is not feasible to introduce suitable satellites and the antenna position densities have an elongated shape, one should be careful with the interpretation of the posterior densities.

Sampling

The sampling engine was able to perform nested sampling based on Markov chain Monte Carlo, which converged close to the maximum of the posterior. As discussed above, the convergence was better when more data was available, especially when FRET anisotropy data was analyzed. The result of nested sampling was a set of samples with individual weights as well as the evidence, which can be used in model selection. The latter possibility is discussed in detail in section 6.7.5.

The compression of the objects (i.e. samples from the constrained prior) used in nested

sampling can be deduced from the plots of samples produced during the course of the algorithm. Though, strictly speaking, the sample produced at iteration j originates from the surface of the constrained prior and not from its bulk, such plots are useful to monitor the sampling process[2].

During the sampling process, the local move length scales were adapted successfully to achieve a specific ratio of accepted and rejected moves. Both the local move length scales and the acceptance/rejection (A/R) ratio are correlated quantities, since a decrease of the local move length scale results in more accepted local moves and vice versa. The A/R ratio, in turn, was adjusted in order to sample the constrained prior in an efficient way. The length of the random walk needed to produce an independent sample from the constrained prior was dynamically adapted during the course of nested sampling. The higher the random walk length, the more difficult it is to maneuver within the constrained prior.

From the behavior of the A/R ratio and the local move length scales, one can get a qualitative impression of the average local structure and the global structure of the constrained posterior density. For example, a high A/R ratio would occur when a density is compact on the length scale of the local move and occupies only a small fraction of parameter space volume on larger length scales. A further increase of local move length scale and therefore a lower number of accepted moves would yield more rejections, which, in turn, must be compensated by an increased random walk length. A decrease of local move length scale accompanied by an increase of the A/R ratio, would result in a slower exploration of the density, and thus be suboptimal.

In contrast, a low A/R ratio and therefore a large local move length scale corresponds to a rather scattered constrained prior density. In this case, it is efficient to put up with a larger number of rejections, as it is eventually possible to reach some distant probability region in a single local move. A decrease or increase of the A/R ratio would again increase the random walk length.

In the example calculation, the constrained prior density was compacted in two phases. In the first phase, the antenna positions condensed (iteration $\sim 25 \cdot 10^3$), which was preceded by a decrease of local move length scales in these parameters and accompanied by a decrease of the A/R ratio. After a close examination, there is evidence for an earlier compaction in the average transition dipole orientation parameters. This process is not visible in the plots of samples, but can be observed in the decrease of the local move length scales (iteration $< 10 \cdot 10^3$).

In the second phase, the average transition dipole orientation parameters become fixed, and the antenna positions are refined (iteration $\sim 100 \cdot 10^3$). This change is preceded by the shrinking of the local move length scales in all parameters (iterations $\sim 50 \cdot 10^3$ – $100 \cdot 10^3$) and visible in the fine structure of the produced samples (figure 6.25).

The non-local move was intended to speed up sampling by jumping in between the neighborhoods of two reference points. One of the reference points was chosen close to the current position of the random walk, whereas the other reference point was randomly chosen from the collection of nested sampling objects (section 5.5.4). Unfortunately, the non-local move was found to have only a very low efficiency (figure 6.27f). The reason therefor might be the infamous "curse of dimensionality". As the collection of objects might be sparsely distributed in the parameter space, the distance of the current position to its closest reference point might actually be very large, while the current position is still regarded to be "close" to the reference point. In high dimensional spaces, the amount

---

[2]When one is interested in the nested sampling objects at a specific iteration j, those can be approximately recovered by drawing from the samples $\{\theta_i\}_{k \geq j}$ with probability proportional to $e^{-(k-j)/M}$, where k is the index of the sample, and M is the total number of objects.

of constrained prior density in such a large neighborhood can be vanishingly small, and it might not be feasible to find these regions. This is supported by the increase of the non-local move efficiency (figure 6.27f) during nested sampling as follows: since the collection of objects gets more and more compressed during the progress of nested sampling, the size of the neighborhood of an object shrinks, which, in turn, might cause the observed effect.

It is questionable, however, whether the non-local move presented here is necessary, since the move is computationally expensive. Furthermore, the starting point of the random walk is chosen from the set of nested sampling objects, which might produce sufficiently independent new samples even though the random walk does not explore the whole constrained prior. Yet, this question should be checked in a separate study, and there might be a way to improve the efficiency of the move.

Now, after the theoretical discussion of the benefits and risks of global data analysis with the *position‑orientation* NPS model, as well as the possible improvement of commonly performed FRET localization experiments by introduction of FRET anisotropy measurements, a real FRET efficiency network will be analyzed by *position‑orientation* NPS in the next section.

6.4 Application II – Nontemplate DNA in the elongation complex

Only a small part of the nontemplate DNA was observed in the Pol II elongation complex crystal structures (chapter 4), while the major part located in the transcription bubble and the upstream DNA were missing. In the following, the *position‑orientation* NPS model will be used to localize the positions of fluorophores attached to different sites on the nontemplate strand. The resulting antenna fluorophore position densities will be compared to previously published results (Andrecka et al., 2009), which were obtained by *position‑Förster distance* model analysis and were used to build a model of the complete Pol II elongation complex. As in section 6.2, the experiments were performed in the absence of NTPs, hence the Pol II complex was always stalled.

6.4.1 NPS inference

The analysis is based on experimental data measured by Joanna Andrecka, Barbara Treutlein, Julia Nagy and Maria Angeles Izquierdo Arcusa.

FRET network structure and prior assumptions

The satellite fluorophores were attached on the template DNA, the RNA and the Rpb4/7 heterodimer to the positions tDNA(-10), tDNA(+3), tDNA(+9), RNA(+1), RNA(+4), RNA(+10), Rpb7(C150) and Rpb4(S73C). The antennas were placed on the nontemplate DNA at positions ntDNA(+1), ntDNA(-2), ntDNA(-4), ntDNA(-7), ntDNA(-12), ntDNA(-15) and ntDNA(-18). Almost every satellite-antenna combination was measured (figure 6.29). The data is listed in the appendix II in tables 3 (fluorescence anisotropies), 4 (FRET efficiencies) and 5 (isotropic Förster distances).

While Alexa 647 was used as satellite fluorophore throughout, the antenna fluorophores were either Alexa 555 or TMR. In order to stay in the distance range in which FRET efficiency measurements are sensitive, TMR was used when longer distances between antenna and satellite were expected, while Alexa 555 was used for shorter separations. In the *position‑Förster distance* NPS calculations it was assumed that TMR and Alexa 555 had the identical position when attached to the same labeling site, while both average
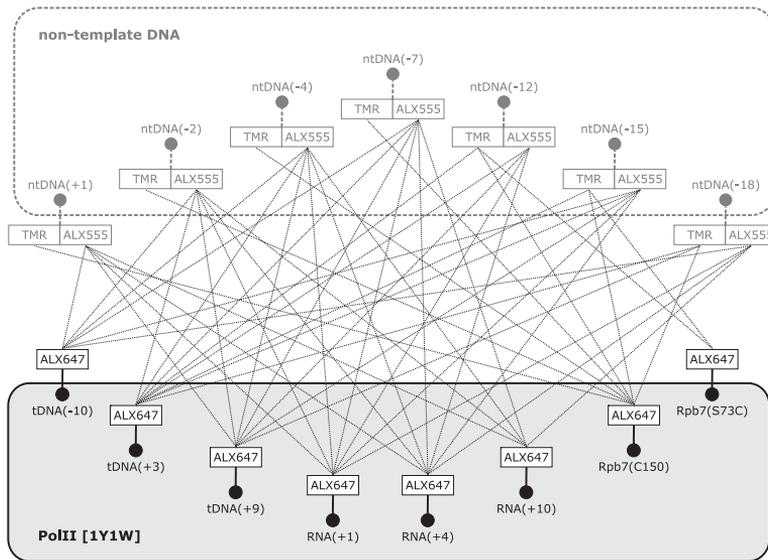
Figure 6.29: Nontemplate DNA FRET network. The positions of various antenna fluorophores attached to the nontemplate DNA was determined from 49 FRET efficiencies (dotted lines) to satellite fluorophores attached to Pol II. Alexa 555 (ALX555) and TMR were used as antennas, while Alexa 647 (ALX647) was used as satellite throughout. TMR and ALX555 were assumed to occupy the same position.

transition dipole moment orientation and position were assumed to be identical in the analysis with the *position - orientation* model.

The average transition dipole moment orientation priors of all fluorophores were flat and completely uninformative. The satellite position priors were flat within the accessible volumes shown in figure 6.30.

In the analysis based on the *position - Förster distance* model described by Andrecka et al. (2009), the antenna position prior was flat within the accessible volume bounded by the known Pol II structure. The minimum allowed distance between the antenna position and the protein was 2.5 Å in the case of ntDNA(+1), ntDNA(-2), ntDNA(-4) and ntDNA(-7), and 5 Å in the case of ntDNA(-12), ntDNA(-15) and ntDNA(-18). The antenna fluorophores ntDNA(-15) and ntDNA(-18) were bounded additionally by the attachment via the nontemplate DNA strand, which was treated effectively as a flexible linker. The linker attached at the last known nontemplate DNA nucleotide was 35 Å (ntDNA(-15)) and 45 Å long (ntDNA(-18)). Different FRET efficiency measurement uncertainties, ΔE_{ij}, were used depending on the measured FRET efficiency. $\Delta E_{ij} = 0.02$ was assumed when the measured FRET efficiency was smaller than 0.9, and a twice as large uncertainty was used otherwise.

The *position - orientation* model analysis was based upon slightly different antenna position priors. The minimal distance between the antenna fluorophore and the protein was 2.5 Å, and was the only constraint besides a large box around the Pol II structure. All measurement uncertainties were set to the same value of $\Delta E_{ij} = 0.02$ irrespective of the measured FRET efficiency. In order to compare the two NPS models on equal footing, an additional calculation based on the *position - Förster distance* model was set up with the same prior and FRET efficiency measurement uncertainties.
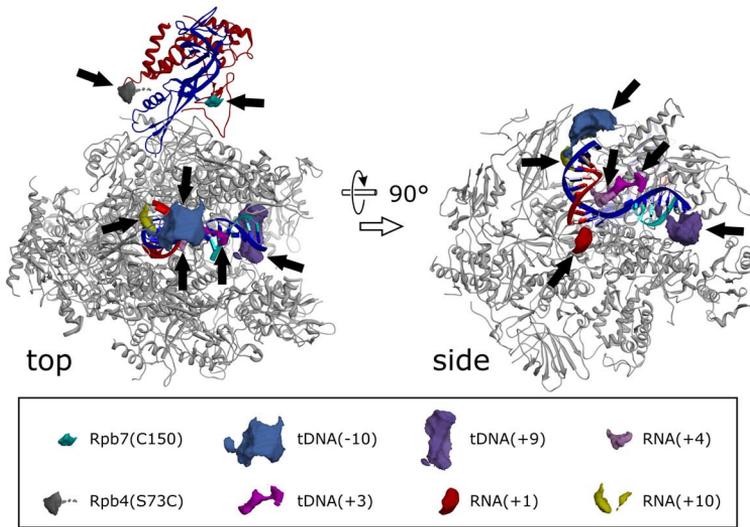
Figure 6.30: Nontemplate DNA, satellite position priors as published in Andrecka et al. (2009). The accessible volumes of the satellite fluorophores (indicated by arrows) are shown in top view (left) and side view (right) of the Pol II elongation complex. The position priors were flat within these volumes.

Inference with the position - Förster distance model

The marginal antenna position posteriors computed with the *position - Förster distance* NPS model were used to build a molecular model of the nontemplate DNA and the upstream DNA (Andrecka et al., 2009) (figure 6.31a). The model was manually constructed by Alan Cheung, who tried to maximize the overlap between the inferred antenna position densities and the predicted accessible volumes of the antenna fluorophores bound to the nontemplate DNA (figure 6.31b,c). At the same time, he tried to maintain the "molecular biologist's common sense", e.g. to take into account the attraction between electrical charges on the Pol II surface and the nontemplate DNA backbone. At the end of the manual modeling process, an energy minimization was performed.

Inference with the position - orientation model

The same FRET efficiency data was analyzed globally with the *position - orientation* NPS model by nested sampling the 75-dimensional posterior with a collection of 1000 objects. The analysis resulted in compact position densities even without encoding the constraints imposed by the attachment via the nontemplate DNA to Pol II in the prior (figure 6.32a,b). The orientation uncertainty (data not shown) was comparable in magnitude to the situation of the simulated 7 satellite / 7 antenna FRET efficiency network shown before (figure 6.23a,b). Yet, some inconsistencies were found in the measured data by comparing the marginal likelihood with the density of expected FRET efficiencies computed from the posterior (section 5.3.6, figure 6.33). This control is equivalent to identifying outliers in the data from unexpectedly large deviations of a data point from the fitted function. It can be interpreted in the same way like the measurement and posterior profiles in section 6.2. Here, two out of 49 measurements, namely the
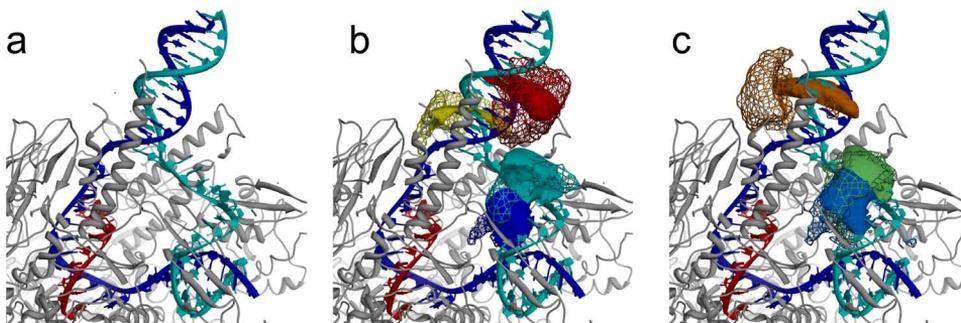
Figure 6.31: Nontemplate DNA, *position - Förster distance* model inference. In (a) the molecular model of the nontemplate and upstream DNA is shown, which is based on the marginal antenna position posterior densities inferred by the *position - Förster distance* NPS model. The nontemplate DNA, the template DNA and the RNA are shown in cyan, blue and red, respectively. (b) shows 68% credible intervals of the position densities (surfaces) of the antennas attached at ntDNA(+1) (blue), ntDNA(-4) (cyan), ntDNA(-12) (yellow) and ntDNA(-18) (red). The simulated accessible volumes of the fluorophores attached to the nontemplate DNA are displayed as meshes in the same colors. In (c) the analogous situation is shown for the antennas attached at ntDNA(-2) (corn flower blue), ntDNA(-7) (green) and ntDNA(-15) (orange).

measurements of ntDNA(-18)-TMR to tDNA(+3)-ALX647 and ntDNA(-18)-ALX555 to tDNA(+3)-ALX647 exhibited the strongest inconsistencies as indicated by a vanishing overlap between the marginal likelihood and the expected FRET efficiency density (figure 6.33e,f).

In contrast to the global *position - orientation* model calculation, an independent *position - Förster distance* model analysis with the same prior assumptions resulted in elongated (ntDNA(-12)) and in two cases (ntDNA(-15) and ntDNA(-18)) also bimodal marginal position posterior densities of the antennas (figure 6.32c,d). The remaining estimated antenna positions ntDNA(+1), ntDNA(-2), ntDNA(-4) and ntDNA(-7) are closer to the downstream DNA, similar to the original *position - Förster distance* NPS model calculation (figure 6.31b,c).
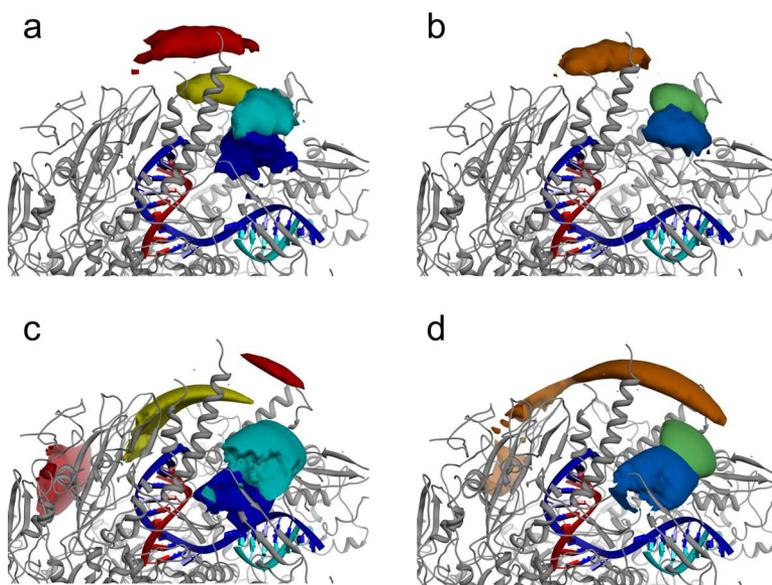
Figure 6.32: Nontemplate DNA, *position - orientation* model inference. The 68% credible volumes of the marginal antenna position posteriors are shown as surfaces. In (a) and (b) the *position - orientation* model results are displayed, while below (c,d) the corresponding *position - Förster distance* model calculations performed with the same prior assumptions are shown. Subfigures (a,c) depict the credible volumes of the antennas attached to the sites ntDNA(+1) (blue), ntDNA(-4) (cyan), ntDNA(-12) (yellow) and ntDNA(-18) (red), while subfigures (b,d) show the antennas attached to ntDNA(-2) (corn flower blue), ntDNA(-7) (green) and ntDNA(-15) (orange). The template DNA (blue), the nontemplate DNA (cyan) and the RNA (red) observed in the 1Y1W crystal structure are shown for better orientation.

_
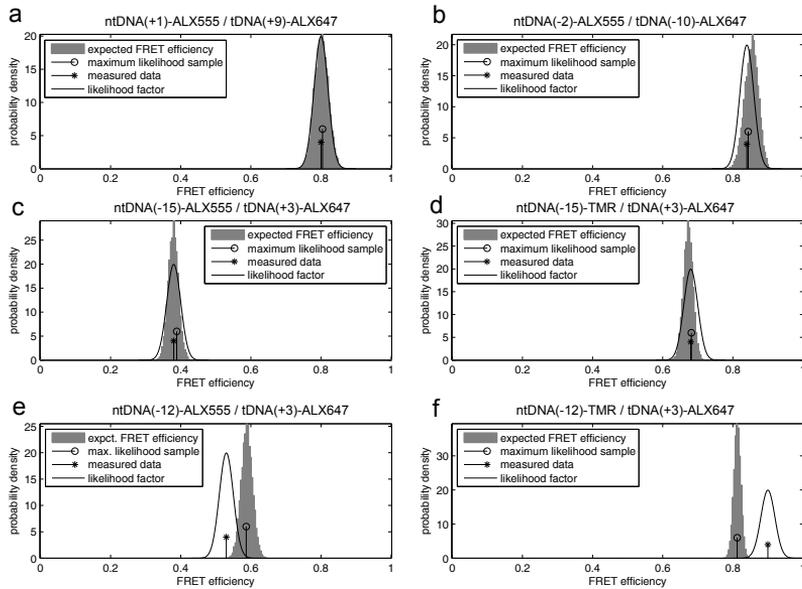
Figure 6.33: Nontemplate DNA consistency check. The marginal likelihood (solid line) and the expected FRET efficiency density computed from the posterior samples (gray) are plotted together with the expected FRET efficiency of the maximum likelihood sample and the measured FRET efficiency. For most of the measurements, the expected FRET efficiency distribution is almost identical with the marginal likelihood, like in the case of ntDNA(+1) / tDNA(+9) and ntDNA(-2) / tDNA(-10) (a, b). For some ANT-SAT attachment site pairs, FRET efficiencies were measured twice with different donor fluorophores (TMR and Alexa 555). Those measurements produced a slightly more peaked expected FRET efficiency density similar to the densities shown in (c) and (d). Only in the case of the FRET efficiency measurements between ntDNA(-18) and tDNA(+3) the data was noticeably incompatible (e, f).

6.4.2 Discussion

The model built on the basis of the *position - Förster distance* model NPS results (Andrecka et al., 2009) is consistent with published biochemical, biophysical and genetic data. The angle between the upstream and downstream DNA double strands of $\sim 60°$[3] is in agreement with atomic force microscopy data, which revealed an angle of $88 \pm 37°$ (Rees et al., 1993). Since the bases -5 to -10 of the nontemplate DNA are exposed, the model accounts for the observed cleavage of the upstream nontemplate DNA nucleotides (up to position -5) by micrococcal nuclease in the bacterial elongation complex (Andrecka et al., 2009; experiments by Wang and Landick, 1997). The model is also consistent with site-specific DNA-protein cross-linking within the bacterial elongation complex (Korzheva et al., 2000; Naryshkin et al., 2000). However, Korzheva et al. (2000) do not make any statement about protein-DNA cross-links in the nontemplate DNA region -4 to -14, which are predicted by the NPS based model.

Yet, the densities of the antennas attached at ntDNA(-12), ntDNA(-15) and ntDNA(-18) shown in figure 6.31b,c were strongly influenced by the prior, as verified by test calculations without additional constraints (figure 6.32c,d). Hence, the antenna positions on the upstream part of the nontemplate DNA were determined primarily by the estimated maximum distance between the antenna fluorophores and the last nontemplate DNA base visible in the structure. These distances, however, were arbitrarily set to 15 Å (ntDNA(-15)) and 45 Å (ntDNA(-18)) (Andrecka et al., 2009), which corresponds to roughly 80% of the DNA backbone lenght (Andrecka et al., 2009). The angle between the upstream and downstream DNA double strands might thus be imprecise.

The global analysis of the FRET efficiency data from Andrecka et al. (2009) with the *position - orientation* NPS model, however, increased the localization accuracy reasonably without the aid of a highly informative antenna position prior. The resulting marginal antenna position posteriors suggest that the angle between the upstream and downstream double strand DNA is larger than in the original model (figure 6.34a,b) and might be in better agreement with the published data (Rees et al., 1993). Also the bend in the single strand region of the nontemplate DNA is presumably located further away from the downstream DNA double strand (figure 6.34c-f), but it is difficult to judge whether more bases would then be susceptible to cleavage by nucleases.

However, the data contained inconsistencies, the most peculiar being the measurements between the positions ntDNA(-12) and tDNA(+3). Recent experiments performed by Wolfgang Kügel and Anders Barth shed light on the possible reason for this puzzle (Kügel and Barth, 2010, unpublished data). They used multiparameter fluorescence detection (MFD) (Eggeling et al., 2001) to study fluorescently labeled Pol II elongation complexes prepared as described by Andrecka et al. (2009). MFD is a single molecule technique that enables to measure various properties of fluorescent molecules in solution like time-resolved fluorescence anisotropy and fluorescence lifetime, FRET efficiency and anisotropy, as well as the translational and rotational diffusion. Kügel and Barth (2010) could distinguish two FRET efficiency populations in the samples. Based on the translational diffusion of the labeled molecules, the populations were identified as nucleic acids bound to Pol II and as free nucleic acids without Pol II. Astonishingly, in the case of measurements between the positions ntDNA(-12)-ALX555 and tDNA(+3)-ALX647, as well as the positions ntDNA(-18)-ALX555 and tDNA(-10)-ALX647, only a minor part of the labeled nucleic acids was bound to Pol II. Most of the detected molecules were free nucleic acids and exhibited a FRET efficiency comparable to the values reported by Andrecka et al. (2009).

Upon closer inspection, also the fluorescence anisotropy data has inconsistencies. For

---

[3]In the original publication a rough estimate of $\sim 80°$ was given, but a more accurate analysis resulted in the angle of $61°$.
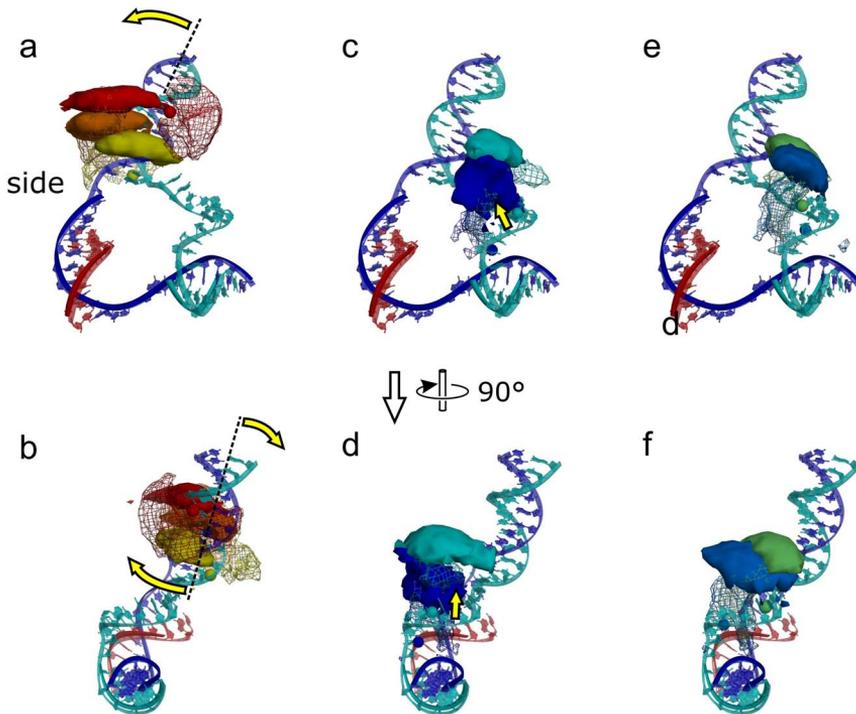
Figure 6.34: Nontemplate DNA, proposed changes in the model. The 68% credible volumes of the marginal antenna position posterior (surfaces) analyzed globally with the *position - orientation* model are shown together with the simulated antenna positions (meshes) based on the molecular model (transparent cartoons) previously proposed by Andrecka et al. (2009). The color code is as in figure 6.32, and the nucleic acids observed in the crystal structure 1Y1W are shown as well (opaque cartoons). The global analysis suggests that the upstream DNA double strand must be tilted to form a larger angle with the downstream DNA double strand (a, b), and that the single-stranded nontemplate DNA should be located further away from the downstream part of the template DNA (c-f).

instance, the anisotropies of Alexa 647 attached to the positions RNA(+1) and RNA(+4) are very small ($r = 0.15$) when compared to the anisotropies measured for e.g. tDNA(+3) ($r = 0.32$), a site located at similar depth inside the Pol II complex. A simple explanation is that either the nucleic acid scaffold consisting of RNA and DNA cannot bind efficiently to Pol II when the RNA is labeled with a bulky fluorophore that is intended to be incorporated deeply in the protein. Another reason might be a hindered formation of the DNA-RNA hybrid when RNA is labeled in the basepaired region of only 7 nucleotides length. Alternatively, the RNA could be stripped off when the nucleic acid scaffold binds to Pol II. It is thus questionable, whether the fluorescence anisotropies measured in bulk experiments represent the fluorescence anisotropies of the FRET samples, since they might be influenced by the signal of free nucleic acid scaffolds or even free RNA. While the fluorescence anisotropies determine the overall localization accuracy and have only a minor effect on the position of the Förster distance distribution in the *position - Förster distance* model (sections 6.1.1 and 6.1.2), the global analysis with the *position - orientation* model could suffer from systematical errors, since all model parameters are strongly correlated by the measurements.

Finally, the fluorescence lifetime of Alexa 555 is short (0.3 ns) when compared to the lifetime of Alexa 647 (1.0 ns) (Invitrogen / Molecular Probes, 2010) and TMR (up to 3.6 ns) (Eggeling et al., 1998). Thus, it is difficult to determine the residual fluorescence anisotropy of Alexa 555 even when time-dependent single-photon counting experiments are performed, and the steady state values used here might be systematically too high.

Hence, the model proposed by Andrecka et al. (2009) as well as the global analysis must be regarded as preliminary. New data should be acquired with a fluorophore that possesses a longer fluorescence lifetime than Alexa 555 and, at the same time, has a similar isotropic Förster distance when used as donor in a FRET pair with Alexa 647. In general, it would be desirable to increase experimental control. To this end, the fluorescence anisotropies should be acquired in single-molecule experiments. Moreover, weakly bound components of the Pol II EC could be labeled with an additional fluorophore that does not influence the other fluorophores constituting the FRET pair, so that correctly assembled molecules could be recognized by the presence of all fluorophores.

Up to now, only the positions and orientations of fluorophores were determined by NPS. In the next section, test calculations will be presented showing how two macromolecules can be "docked" by FRET measurements.

6.5 Simulation III – FRET-assisted docking

In this section, a test calculation will show the ability to dock macromolecules with the *position - orientation* NPS model. To this end, the artificial data used in section 6.3.3 was reanalyzed with different prior information.

The prior information of the satellite fluorophores was not changed, and they were assumed to be bound to known positions of a macromolecule, M0, with known structure[4]. The antenna fluorophores, however, were likewise assumed to be attached to known positions on a second macromolecule, M1, that was unlocalized yet (figure 6.35). As before, the average transition dipole moment orientations were not known in the analysis, and hence flat uninformative priors were assigned in these parameters.

The position and orientation of the reference frame, $(x_{(1)}, y_{(1)}, z_{(1)})$, that represents the second macromolecule was assigned an uninformative prior. The orientation prior was

---

[4]The macromolecule is only hypothetical, since its presence was not accounted for in the calculations. Its true purpose is to serve as a fiducial mark in the figures.
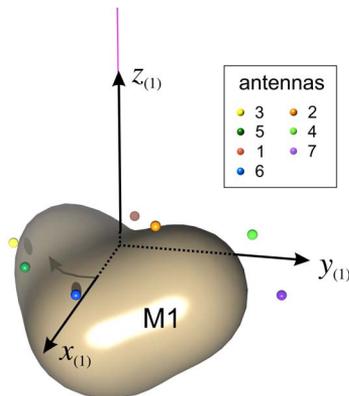
Figure 6.35: Antenna fluorophore positions relative to a second macromolecule. A reference frame $(x_{(1)}, y_{(1)}, z_{(1)})$ is attached to the macromolecule M1 (brown surface).

flat as stated in equation (3.3.3), and the position prior was flat in a large cube of 300 Å side length.

Three scenarios (I, II, III) were studied. All were based on the artificial data of the 7 satellite / 7 antenna FRET network (section 6.3.3). In the scenarios I and II, the positions of the fluorophores relative to the respective macromolecules were known with infinite accuracy. In scenario III, the position priors of all fluorophores were uniform within cubes of 8 Å side length centered at the simulated positions in the reference frames of the respective macromolecules. While only FRET efficiency data was used in scenario I, FRET anisotropies were analyzed in addition to FRET efficiencies in the other two scenarios.

Nested sampling with a collection of $2 \cdot 10^3$ objects was used to compute posterior samples in all three scenarios, which contained 34 (scenario I and II) and 76 (scenario III) model parameters. In the following, the results of the calculations will be shown and discussed.

6.5.1 FRET-assisted docking of two macromolecules

The inference results were visualized by 95% credible volumes of the marginal position of four points in the reference frame of the docked macromolecule M1. One point was the origin of the docked reference frame, and the other three points were located on the axes 60 Å away from the origin. The marginal position posterior density of the point in the origin was used as an estimate of the localization uncertainty, while the densities of the other points served as a measure of the orientation accuracy (figure 6.36). The smallest localization uncertainty was observed in scenario II, while the largest was found in scenario III.

To get an impression of the orientation accuracy of the fluorophores, the marginal posterior densities of the average transition dipole moment orientations were calculated. The corresponding 68% and 95% credible intervals are displayed for the satellite fluorophores in figure 6.37. In the case of antenna fluorophores, the orientation accuracies were of similar size and are not shown here. The calculations with the largest and smallest orientation accuracy were scenario I and II, respectively.

The maximum sampled likelihood, $L_{\max}^{\text{sampl}}$, was closest to the maximum possible likelihood in scenario II ($\ln L_{\max}^{\text{sampl}} = 326.7$), followed by scenario I ($\ln L_{\max}^{\text{sampl}} = 145.1$) and III ($\ln L_{\max}^{\text{sampl}} = 324.6$).
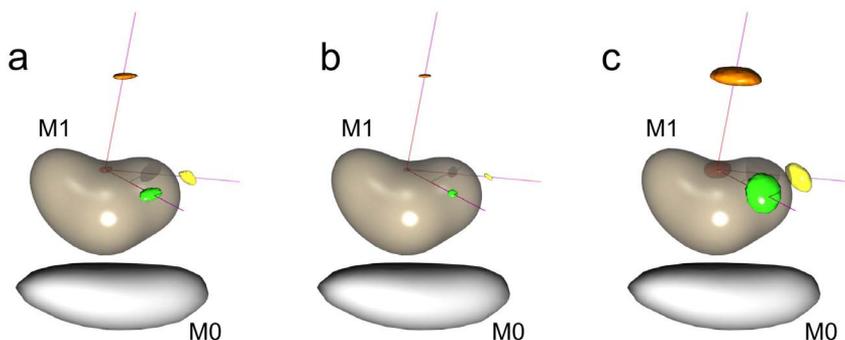
Figure 6.36: FRET docking. The docking of two macromolecules, M0 and M1, is shown for the scenarios I, II and III (a, b, and c). The marginal position posterior densities of four points are displayed as 95% credible volumes: red, within the semi-transparent surface of M1: origin of the reference frame $(x_{(1)}, y_{(1)}, z_{(1)})$ of the docked molecule M1; green, yellow and orange: points located at the x, y and z-axis and separated 60 Å from the origin. The tripod is shown at the simulated position and orientation.
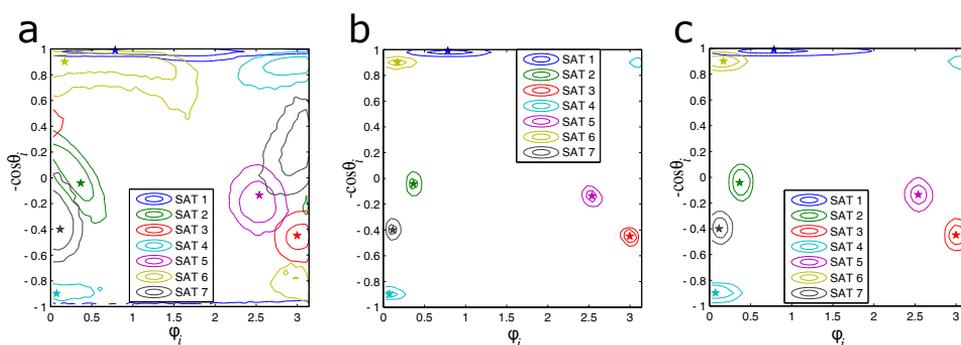
Figure 6.37: Average transition dipole moment orientations. The marginal posterior densities of the average satellite transition dipole moment orientations are shown as 68% and 95% credible intervals. The scenarios I, II and III are shown in (a), (b) and (c), respectively, and the simulated orientations are displayed as stars.

6.5.2 Discussion

In all three scenarios, the position and orientation of the docked frame of reference was estimated correctly. The highest localization and orientation accuracy was achieved in the presence of FRET efficiency and anisotropy data with precisely known relative fluorophore positions (scenario II).

The performed test calculations were based on the same data as the global analysis of FRET efficiency only data and combined FRET efficiency and anisotropy data in the 7 satellites / 7 antenna network in section 6.3.3. The corresponding scenarios will be hence compared.

The localization accuracies of the docked reference frame origin were always higher than the localization accuracies of the satellite positions in the corresponding calculations in section 6.3.3. This is not surprising, since the antenna positions were highly correlated by the attachment to the docked molecule, while each position was independent without this information.

One would expect that the introduction of uncertainty in the satellite and antenna fluorophore positions should have a larger effect on the position of the docked reference frame origin than on the positions of the antennas computed in section 6.3.3 (figure 6.21f,g). The uncertainty would apply twice in the case of docking, once on molecule M0 and a second time on the molecule M1. This effect can be seen by comparing figure 6.36b and c. Yet, on the absolute scale, the effect is compensated here by the still strong correlation of antenna positions, and finally, the origin of the docked reference frame is determined with a higher accuracy than any antenna position in the analysis without docking.

Even when only FRET efficiency data was available (scenario I), the correct average transition dipole moment orientations were almost perfectly recovered. Supposedly, the reason were the correlations between the antenna positions and orientations introduced by the measured FRET efficiencies and the information about the relative antenna and satellite positions.

The differences of the maximum sampled likelihood and maximum possible likelihood (table 6.2, page 114) were comparable to those determined in the calculations in section 6.3.3. This suggests that the sampling of the posterior was of a similar difficulty level when compared to the global analyses without docking. Further, one can infer that in the latter case it was indeed the lack of orientation information that caused the strongly asymmetric marginal antenna position posterior densities (figure 6.21), because most of the marginal densities are now nicely centered at the simulated average transition dipole moments.

Although the docking of only two macromolecules was shown here, the analysis method is capable of docking several macromolecules simultaneously. Furthermore, the posterior marginalized in the position and orientation of each docked macromolecule at the same time, can be used as a restraint in other docking applications. Effectively this would be done by a 6-dimensional docking potential (the negative of the natural logarithm of the marginal posterior) that acts on the position and orientation coordinates of the docked component. In that way, the complete information about the position and orientation of each component of the macromolecular complex is preserved. In contrast, the computation of marginal position posterior densities of arbitrary points in the docking reference frame, as presented here, destroys the correlations between position and orientation. In this spirit, the practice to determine the position and orientation of docked macromolecules from the marginal position posterior densities of arbitrary points in the docking reference frame should be replaced by a different method that allows to utilize the full information content.

In general, the docking of macromolecules as demonstrated in the test calculations above shifts the focus of NPS from determining fluorophore positions to the structurally more relevant positions and orientations of macromolecules relative to each other. In the next section, this framework will be applied to determine the gross structure of the initial transcribing complex of Pol II.

6.6 Application III - Position and orientation of TBP in the initial transcribing complex

This section deals with the coarse positioning of the TATA binding protein (TBP) within the initial transcribing complex (ITC), which is important for the detailed understanding of the initiation phase of eukaryotic transcription.

6.6.1 NPS inference

The analysis is based on experimental data measured by Barbara Treutlein, Joanna Andrecka and Monika Holzner. The measurements were carried out with stalled Pol II initial transcribing complexes (section 5.1.1). The analysis should be regarded as preliminary, as the ITC project is still subject of current research.

FRET Network structure and prior assumptions

The satellite fluorophores were attached on the template DNA and the Rpb4/7 heterodimer to the positions tDNA(-10), tDNA(+3), tDNA(+7), tDNA(+12), and Rpb7(C150). The antennas were placed on the nontemplate DNA at both ends of the TATA box at positions ntDNA(-30) and ntDNA(-37), as well as on TBP at the position TBP(S159C). It was assumed that these labeling sites are located on a rigid macromolecule, the TBP/TATA subcomplex, consisting of TBP and the bent DNA double strand in the region of the TATA box. An additional antenna fluorophore was placed at position ntDNA(-20) on the nontemplate DNA strand, which was assumed to be base paired with the template DNA and thus part of a locally rigid upstream DNA double strand segment.

The TBP/TATA subcomplex and the upstream DNA segment were docked to Pol II by means of 20 FRET efficiency measurements in total, which were acquired with various fluorophores (Alexa 647, TMR and Cy3) (figure 6.38). The data is listed in the appendix
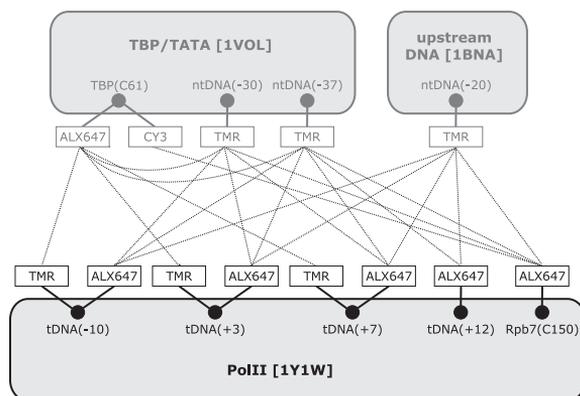
Figure 6.38: Initial transcribing complex FRET network. 20 FRET efficiencies (dotted lines) were measured between fluorophores attached to Pol II, the TBP/TATA subcomplex and the upstream DNA. Fluorophores attached to the same labeling site were treated as completely independent, i.e. were described by separate positions and average transition dipole moment orientations.
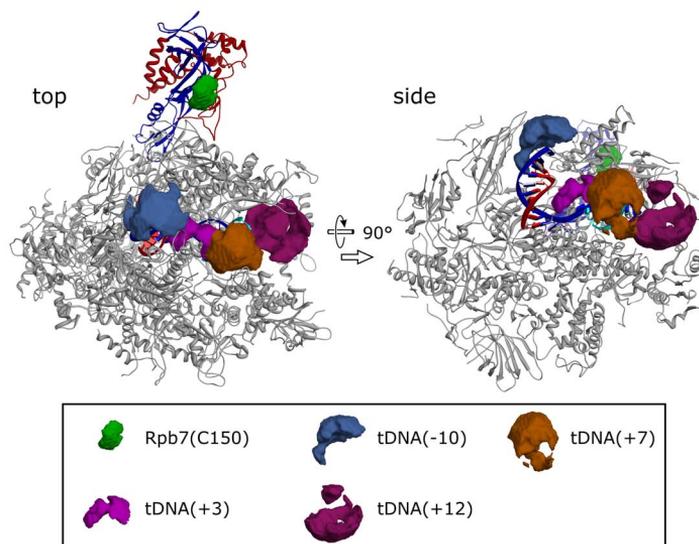
Figure 6.39: Satellite position priors. The volumes accessible to the satellite fluorophores attached to Pol II are shown in the top (left) and side view (right). The fluorophore position priors were flat within these volumes.

II in tables 3 (fluorescence anisotropies), 6 (FRET efficiencies) and 7 (isotropic Förster distances).

The volumes accessible to the fluorophores were computed based on the Pol II elongation complex structure (Kettenberger et al., 2004, PDB-ID: 1Y1W) (figure 6.39), the structure of TBP bound to DNA with the TATA sequence (Nikolov et al., 1995, PDB-ID: 1VOL) (figure 6.40a), and a structure of double-strand DNA in B-form (Drew et al., 1981, PDB-ID: 1BNA) (figure 6.40b). It was hence assumed that the structure of the Pol II enzyme is identical in the elongation and initial transcribing complexes. This applies in particular to the position of the Rpb4/7 heterodimer and the template DNA strand. Further, the local structure of the DNA bent by TBP at the TATA sequence was assumed to be unchanged by possible interactions with the other ITC components. Also the local structure of the upstream DNA was assumed to be double stranded and not bent within two lengths of the carbon chain linker used for attachment of the fluorophore at the position ntDNA(-20) (figure 6.40b). In this way, it was possible to compute the marginal position density of the attachment atom of ntDNA(-20).

In contrast to the calculations in the sections 6.2 and 6.4, each fluorophore was described by a separate set of parameters even when different fluorophores were attached to the same labeling site. Completely uninformative priors were used in the average transition dipole moment parameters of the fluorophores, and the fluorophore position priors were flat within the accessible volumes.

The docked macromolecule parts were described by the positions and orientations of reference frames (figure 6.40). The reference frame position priors were flat within large boxes, which contained the region that was accessible to the upstream DNA and the TBP/TATA subcomplex[5], while the orientation priors were completely uninformative.

---

[5]A supplemental analysis with a position prior that enclosed the whole Pol II resulted in a second
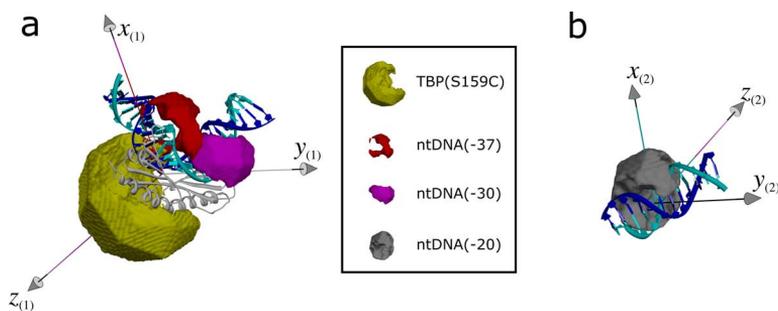
Figure 6.40: ITC, antenna position priors. The accessible volumes of the antenna fluorophores are shown for the TBP/TATA subcomplex (a) and the upstream DNA segment (b). The fluorophore position priors were flat within these volumes. The frames of reference, $(x_{(k)}, y_{(k)}, z_{(k)})$ ($k \in \{1, 2\}$), attached to the rigid macromolecule parts are shown as well. The accessible volume of TBP(S159C) was computed by using an unusually long linker, since the fluorophore is attached to the end of an amino acid sequence lacking secondary structure in the crystal. It was hence assumed that this sequence might be disordered and therefore at a different position in reality, which elongated the linker effectively.

Inference results and preliminary modeling

The inference of th 77 model parameters was performed with nested sampling using 1000 objects. The position of the TBP/TATA subcomplex was visualized by displaying the marginal position posterior density of the Cα carbon atom of isoleucin 168 (I168) in the TBP structure, which is located at the interface of TBP and DNA in close proximity to the bent DNA. To estimate the position of the upstream DNA at register -20, the marginal position posterior density of the attachment site ntDNA(-20) was computed (figure 6.41a).

To find the approximate orientation of the TBP/TATA subcomplex, the marginal position posterior densities of three points located far from I168 were computed. Two points were situated approximately on the axes defined by the DNA double strands upstream and downstream of the TATA box sequence, while the third point was placed on the opposite side of TBP, above the bent DNA. Since the marginal densities of these points were very large and hence the orientation uncertainty was high, the densities were used only as a guide to orient the TBP/TATA subcomplex (figure 6.41b). Finally, an existing biochemical model of the minimal pre-initiation complex (Kostrewa et al., 2009) was used as a rough guideline for the orientation of the DNA in between the TATA box and Pol II.

The orientation of the DNA segment between the transcription bubble and the TATA box was not evaluated, since it contained only one attachment site.

Inconsistent FRET efficiency data was not observed, as confirmed by comparison of each measurement's marginal likelihood factors and the distribution of the respective expected FRET efficiencies (not shown).

6.6.2 Discussion

The position of the TBP/TATA subcomplex was localized by FRET efficiency measurements above the cleft of Pol II with an accuracy of 9, 17 and 18 Å (standard deviation

---

probability density maximum, which could not be reached by the TBP/TATA subcomplex and the upstream DNA at register -20 when the complexes were assembled correctly.
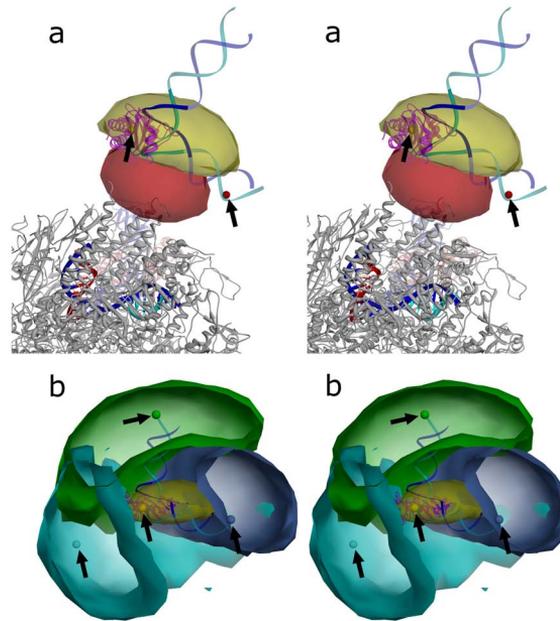
Figure 6.41: A preliminary ITC model. Stereo images of the preliminary ITC model. In (a), the approximate position of the TBP/TATA subcomplex (TBP: magenta ribbons, template DNA: blue, nontemplate DNA: cyan cartoons) located above the Pol II (1Y1W PDB structure) is shown. The model was built by considering the density of I168 (68% credible volume, yellow) shown together with the density of ntDNA(-20) (68% credible volume, red). Both positions are shown as spheres in the structure and marked by arrows. The position of the ntDNA(-20) density indicates that the DNA strand between TBP and Pol II must be bent, as it is expected from the approximate exit site of the upstream template DNA (blue cartoon inside of Pol II). In (b), a detailed view of the 68% credible volumes of of the I168 position and three other points in the TBP/TATA reference frame is shown. Each point is displayed as a sphere in the respective color of the credible volume surface and marked by an arrow. The TBP/TATA subcomplex is shown as magenta ribbons (TBP, barely visible) and blue/cyan cartoons. The front part of the surfaces are not shown.
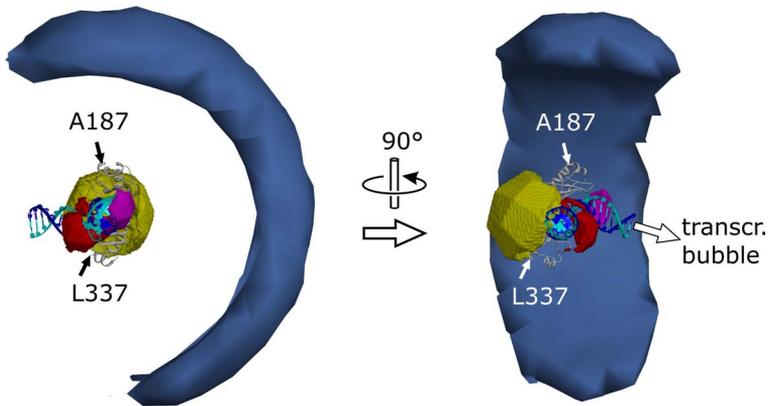
Figure 6.42: Pol II center of mass position in the TBP/TATA reference frame. The 68% credible volume of the marginal position posterior density of the Pol II center of mass $((x_{(0)}, y_{(0)}, z_{(0)}) = (107, 52, -2)$ Å in the coordinate system of 1Y1W) is shown (blue surface). The density is banana-shaped since the fluorophores attached to TBP and the DNA (accessible volumes are shown) constitute only a small trilateration basis in that direction. Measurements to additional labels at the positions TBP(A187) or TBP(L337) (arrows) would improve the localization accuracy of the Pol II center of mass, and therefore also the orientation accuracy of the TBP/TATA subcomplex relative to Pol II. The direction along the DNA towards the transcription bubble is shown by an arrow.

of the I168 density in the principal directions). The marginal position density of the upstream DNA segment labeled at the site ntDNA(-20) indicates that the DNA between TBP and Pol II is bent towards Pol II. Moreover, a straight DNA double strand between TBP and Pol II would clearly disagree with the inferred orientation of the TBP/TATA subcomplex (figure 6.40b, blue density), given the angle of DNA bent by TBP is the same in the ITC and the crystal structure 1VOL.

The DNA could be melted in addition, which is expected since the template DNA strand must reach its exit site separated from the nontemplate DNA in order to form the transcription bubble. The simultaneous melting and bending during transcription initiation, also called *scrunching*, was originally proposed in the bacterial RNA polymerase system (Kapanidis et al., 2006; Revyakin et al., 2006). However, the size of the melted region in the ITC is still unclear and subject of current research. This information will be important for both building a correct biochemical model and for the NPS analysis, since it is not known whether the DNA at position ntDNA(-20) is really double stranded as assumed.

Although the position of the TBP/TATA subcomplex was determined well, its orientation was less accurate. It would be possible to impove the orientation accuracy by analyzing FRET efficiency measurements between the satellites on the Pol II elongation complex and additional antenna fluorophores on TBP. This becomes evident when one switches to the reference frame of TBP/TATA and displays the marginal posterior position density of the Pol II center of mass (figure 6.42). The density is very elongated, and would shrink if additional labels were attached to the aminoacids A187 or L337 on TBP (sequence given as in 1VOL), which is equivalent to an improved orientation accuracy of TBP/TATA in the Pol II frame of reference.

In general, it was difficult to orient the TBP/TATA subcomplex by using the marginal

position densities of points located far from the subcomplex center. It is hence desirable to develop a better method to visualize the orientation of a rigid macromolecule, as discussed already in section 6.5.2.

TFIIB might play this role, since it interacts with DNA and TBP (Nikolov et al., 1995), but this should be proven by direct localization of TFIIB in the ITC.

After having shown the preliminary analysis of docking the TBP/TATA subcomplex to PolII, the next section will focus on some general aspects of NPS.

6.7 General discussion

In the following, NPS will be compared to similar FRET-based methods in the literature, and its basic assumptions will be examined closely. After discussing the impact of fluorophore attachment on their suitability for structural studies of macromolecules, some possible misinterpretations of NPS results will be pointed out. Thereafter, the design of FRET networks will be discussed in regard to keep localization uncertainty small. Finally, a way will be presented how NPS can be linked to current FRET data preprocessing techniques.

6.7.1 NPS and other FRET analysis methods

NPS will be now compared to other data analysis methods, which have been either applied to FRET experiments or were discussed theoretically. The studies will be examined mainly with respect to aspects important for the localization of fluorophores: the FRET network structure, the fluorophore- and FRET model on a single-molecule level, the inter-relationship between the FRET model and the measured data, as well as the presentation of the inference results.

FRET network structure

Two main categories of experiments exist. In the first, a single donor-acceptor distance is inferred in a qualitative (Mukhopadhyay et al., 2001; Kapanidis et al., 2006; Santoso et al., 2010) or quantitative way (Stryer and Haugland, 1967; Kapanidis et al., 2005; Kaiser et al., 2006; Ermolenko et al., 2007a,b), whereas in the second category multiple FRET measurements are used to compute the unknown position of one or more fluorophores (Margittai et al., 2003; Medintz et al., 2004; Rasnik et al., 2004; Sun et al., 2006; Radman-Livaja et al., 2005; Woźniak et al., 2008; Chen et al., 2009; Auclair et al., 2010) or to dock molecules (Mekler et al., 2002; Mukhopadhyay et al., 2004; Knight et al., 2005).

NPS belongs clearly to the second category, yet, it is also possible to extract distributions of distances in the *position - orientation* NPS model. The distances can be inferred either between arbitrary fluorophores, or, when macromolecules are docked, between any two points located in different reference frames. The calculation of distances is basically a kind of marginalization, since the information is already present in the posterior, represented by a set of samples. In the same way, NPS can be used to compute other geometric quantities, for example the angles between the average transition dipole moments of two fluorophores.

Fluorophore model and FRET model

Although the basics of FRET at large donor-acceptor distances is covered already by the quantum mechanical theory derived by Förster (1948) for immobile fluorophores, a

plethora of models has been developed thereafter to analyze experimental data. The reason is probably not the disagreement with the effects that can alter the theoretically expected FRET efficiency and anisotropy, which are basically the static and dynamic inhomogeneities in positions and orientations of the fluorophores (Schröder et al., 2005; Cherny et al., 2009; Auclair et al., 2010), but it is rather the lack of simple ways to characterize these inhomogeneities. Basically, there are two groups of effects that complicate the analysis of FRET data. One is the dynamics intrinsic to a single macromolecule, for example fluctuations of the macromolecule and the attached fluorophores. The other group of effects is extrinsic, as it is caused by the measurement process, for example the counting noise of detectors or the averaging over an ensemble of macromolecules. Here, an overview of the models will be given that describe effects intrinsic to the observed system, whereas the extrinsic factors will be discussed in the next part.

In most studies, a fixed fluorophore position and complete dynamic orientational averaging ($\kappa^2 = 2/3$) is assumed, whereas some authors account for the fluorophore attachment via carbon chain linkers (Mekler et al., 2002; Margittai et al., 2003; Medintz et al., 2004; Rasnik et al., 2004; Mukhopadhyay et al., 2004; Lee et al., 2005; Nir et al., 2006; Woźniak et al., 2008; Auclair et al., 2010), and again others ignore these effects or argue that they are not of importance for the analysis (Stryer and Haugland, 1967; Radman-Livaja et al., 2005; Knight et al., 2005; Kaiser et al., 2006; Sun et al., 2006; Ermolenko et al., 2007a,b; Andrecka et al., 2008; Chen et al., 2009; Auclair et al., 2010). In some cases, the modeling of linkers is kept open (Schröder and Grubmüller, 2004).

One should point out the work of Woźniak et al. (2008), who studied FRET pairs bound to a short DNA double strand at known positions, and simulated the volumes and orientations accessible to the fluorophores with molecular dynamics simulations. They compared different FRET models in detail, namely complete dynamic averaging in orientation and position, complete dynamic averaging in the orientation while the position is static, and FRET of an ensemble of completely static fluorophore orientations and positions. They observed low residual anisotropies of all fluorophores, and used the complete dynamic orientational averaging combined with a static position model, which described the measured FRET efficiency data best.

In the applications of NPS shown in this work, the assumption of $\kappa^2 = 2/3$ is clearly not applicable, since many fluorophores are located deep inside Pol II and have a very restricted orientational mobility, as proven by high fluorescence anisotropies. This is accounted for in NPS analysis in contrast to other published FRET localization methods. However, common to NPS and the literature cited above is the model of fluorophore positions, which are static on the time scale of fluorescence lifetime or on even larger time scales.

Although the treatment of anisotropy effects in NPS is based on the work of Dale et al. (1979), these authors computed only the maximum and minimum values of the Förster distances (Dale and Eisinger, 1974; Dale et al., 1979). This calculation was repeated by Ivanov et al. in more detail in a recent publication 2009. In contrast, Hillel and Wu (1976) proposed a probabilistic treatment of average transition dipole moment orientations, which shares some similarities with the model used in NPS. One of their results is the probability density of donor-acceptor distances, similar to the density obtained by NPS when only one FRET pair is analyzed. In particular, the transition dipole moment was assumed to be freely mobile on the surface of a cone, which is a more explicit model of the dynamically averaged fluorophore distributions when compared to the work of Dale et al. (1979) and to NPS. The opening angle of the cone was determined from the fluorescence anisotropies of the fluorophores like done by Dale and Eisinger (1974). The approach of Hillel and Wu (1976) can be interpreted in the Bayesian data analysis frame-

work as follows: In contrast to the NPS that uses a Gaussian likelihood and a flat prior in the antenna fluorophore position, a flat prior in the distance and a flat likelihood between an experimentally measured maximum and minimum FRET efficiency was used. Dale et al. (1979) explicitly reject the probabilistic treatment of Hillel and Wu, an opinion that is probably based on a misinterpretation of the distance distributions, combined with a criticism of the uninformative fluorophore orientation prior, which has a strong influence on the inference. However, 100% credible intervals of a Bayesian distance estimate would correspond to the reported maximum and minimum error margins proposed by Dale et al. (1979) and Ivanov et al. (2009).

In a theoretical investigation, which is based on molecular dynamics simulations, the correlations between the orientation factor κ^2 and the donor-acceptor distance were studied (VanBeek et al., 2007). The authors report that the fluorophore positions and orientations are correlated, and hence result in a correlation between κ^2 and the inter-fluorophore distance. These correlations, when averaged over the lifetime of the fluorophore excited states, can affect the energy transfer. Such behavior of fluorophores is supported by the combined experimental and theoretical study of fluorescence anisotropy decay of a fluorophore on a protein (Schröder et al., 2005), in which the fluorophore was observed to adapt two distinct conformations with different preferred transition dipole moment orientations.

NPS at its current state cannot account for such correlations, which is effectively dynamic averaging over the combined parameter space of fluorophore orientations and positions. Yet, as long as the fluorescence lifetime is high enough and the position fluctuations of the fluorophores are small compared to their separation, NPS does account correctly for constrained orientation effects observed in a complete FRET network, which is not possible with any of the methods presented above.

Data processing and the likelihood function

Driven by thermal fluctuations, the fluorophore positions and orientations discussed above are known to vary on different time scales. Fast fluctuations are usually caused by the attachment of the fluorophores via flexible carbon chain linkers (Kinosita et al., 1982; Eggeling et al., 2001), while slower dynamics is expected to be caused by conformational changes of the macromolecules (Schröder et al., 2005). Fluctuations much faster than the donor and acceptor fluorescence lifetimes are averaged over dynamically, and cannot be detected directly in FRET. However, slower dynamics can be detected in an experiment with sufficiently high time resolution (Nir et al., 2006; Kalinin et al., 2010b; Andrecka et al., 2008).

In addition, it is likely that inhomogeneities of photophysical properties like the acceptor quantum yield contribute considerably to the observed FRET efficiency broadening (McCann et al., 2010; Kalinin et al., 2010a).

The inhomogeneities described above are averaged over in bulk experiments like those conducted by Stryer and Haugland (1967); Mukhopadhyay et al. (2001); Mekler et al. (2002); Mukhopadhyay et al. (2004); Radman-Livaja et al. (2005); Knight et al. (2005); Lee et al. (2005); Kaiser et al. (2006); Sun et al. (2006); Ermolenko et al. (2007a,b) and Auclair et al. (2010). This averaging, in combination with the negligible contribution of photon counting noise to the measured FRET efficiency, is probably the reason why in bulk studies the measured FRET efficiency is usually directly translated into an average distance of fluorophores when the Förster distance is known. In few cases, a generous error is provided for this distance (Knight et al., 2005) or the corresponding FRET efficiency (Mekler et al., 2002; Mukhopadhyay et al., 2004; Kapanidis et al., 2005), which might be large enough to account for orientation effects.

In contrast to bulk techniques, single molecule measurements are affected by substantial contributions of photon counting noise, often called *shot noise*. Fortunately, this noise source can be quantified reliably and there are techniques like photon distribution analysis (PDA) (Antonik et al., 2006; Kalinin et al., 2007, 2010a) or the similar proximity ratio histograms (PRH) (Nir et al., 2006) that are able to extract from the measured FRET efficiency distribution the amount of inhomogeneous broadening exceeding shot noise. Yet, the development of methods that interconnect structural parameters like the average fluorophore position with the measured data is subject to current research, as presented by Woźniak et al. (2008) and Kalinin et al. (2010b) in a first approach. In general, it is desirable to derive a likelihood function that depends on the parameters, which physically describe the ensemble of observed molecules, and thus close the gap between observation and the molecular model of FRET.

Since such first-principles likelihood functions are still missing, a phenomenological data pre-processing approach is applied in NPS. It consists of the determination of the centers of Gaussian peaks in FRET efficiency histograms calculated from single-molecule intensity traces (section 5.1.3). This popular method is used by many other authors (Margittai et al., 2003; Medintz et al., 2004; Kapanidis et al., 2005, 2006; Margeat et al., 2006; Andrecka et al., 2008; Chen et al., 2009; Santoso et al., 2010), but as soon as a likelihood derived from first principles is available, one should use this more precise method. The method of using center FRET efficiencies is risky, since inhomogeneities are known to change the shape of FRET efficiency histograms and thus cause apparent discrepancies between average structural parameters, for example the average distance (Cherny et al., 2009). In particular, the distance of average fluorophore positions does not necessarily coincide with the distance corresponding to the maximum of the observed FRET efficiency (Churchman et al., 2006) for the following reasons. First, in the presence of static position inhomogeneities, the distance of the average fluorophore positions is smaller than the average over all distances, and second, FRET efficiency depends nonlinearly on the fluorophore distance, which has impact on the actual shape of the FRET efficiency histogram. However, these effects become less important on an absolute scale when the position inhomogeneity is small compared to fluorophore separation.

In conclusion, the data pre-processing method used in NPS works well when the FRET efficiency histograms are narrow. In that case, only small inhomogeneities are present and the differences between the maximum FRET efficiency derived from the data and the FRET efficiency corresponding to the distance of the average donor and acceptor positions can be accounted for in the FRET measurement error.

Visualization of results

The results of data analysis should be presented in a compact and comprehensible way together with the accuracy of the inference. In the following, different ways to visualize the information inferred by FRET localization experiments will be discussed.

Some authors depict the best fitting antenna fluorophore position as a sphere (Rasnik et al., 2004; Andrecka et al., 2008; Chen et al., 2009), which is indeed a very compact and intuitive way to display the experimental results. Yet, the position is often not accompanied by an error margin, and even with the uncertainty given as a number, the depiction of a single position might give the misleading impression of a precise localization.

Authors that use algorithms applied in NMR structure determination usually display several structures of the molecule or the measured position in the FRET network (Radman-Livaja et al., 2005; Sun et al., 2006). In other studies, scatter plots are used to display an ensemble of the positions of atoms or of fluorophores explaining the observed data (Margittai et al., 2003; Medintz et al., 2004; Knight et al., 2005). Both ways of

visualization are difficult to interpret in a Bayesian sense, since the conformations or positions are found with an algorithm that maximizes a FRET efficiency restraint based on a scoring function, and the conformations are not drawn from any particular probability distribution. Instead, the scatter might represent local maxima of the scoring function without indicating their probability. It is rather the performance of the maximum search than an uncertainty, which is shown in that way. Besides our own work (Muschielok et al., 2008; Andrecka et al., 2009) only Woźniak et al. (2008) report accurate uncertainties of the structural parameters.

In contrast to most studies discussed above, the fluorophore position densities computed with NPS have a well defined meaning as a marginalized posterior probability density, which can be easily displayed as credible volumes, and will be discussed in more detail later (see "credible volumes" in section 6.7.3). However, it might be advantageous to show several molecular models that are built into the same marginal densities like commonly done for NMR structures, since up to now, models built into NPS results did not contain any error margins. In that way, the uncertainty of the inferred positions could be expressed in the final model.

After having compared NPS to present FRET localization techniques, a possible way to improve the resolution of such methods will be discussed, and the concept of an "ideal" FRET state will be developed.

6.7.2 Well-defined FRET states as tools for structural biology

It is evident from the preceding discussion that FRET-based structural studies are hampered by molecular properties, which are difficult to control in the experiment. Therefore, the question arises, which properties the donor and acceptor fluorophores should possess in order to be suitable to be used in a FRET localization study.

In general, the fluorophores contributing to a well-defined FRET state used to analyze distances and orientations in macromolecules should be accurately characterized by experiments and simulations. In particular, the chemical structure of the fluorophores should be known in order to be able to simulate accessible volumes and orientations. Another important requirement is a single photophysical and conformational state of the fluorophore attached to the macromolecule. More than one state can obviously complicate the interpretation, since different conformational states of the fluorophores might be confused with different conformations of the macromolecule, which is usually not known a priori. Finally, the FRET efficiency should be in a sensitive range.

In its present form, NPS requires fluorophores in static positions that exhibit constrained, axially symmetric orientational fluctuations faster than the donor and acceptor fluorescence lifetimes. The fluorophores must thus have either long fluorescence lifetimes or short rotation correlation times. While photophysical artifacts like quenching by metal ions (Chung et al., 2010) cannot be ruled out in general, at least the effect of quenching in different environments caused by different fluorophore conformations on the macromolecule (Eggeling et al., 1998) could be reduced by an attachment by short linkers and/or the usage of fluorophores like Alexa 647 that can stick to a hydrophobic protein surface (Loman et al., 2010). This would, in addition, shrink the volume accessible to the fluorophore and thus decrease the influence of this source of uncertainty.

Unfortunately, probably the residual fluorescence anisotropy would be increased as well, which results in higher localization uncertainties. Yet, NPS is able to account for these uncertainties and is even able to compensate them by analyzing additional FRET anisotropy data as shown in the test calculations in section 6.3.3. Moreover, a pronounced orientational constraint would probably reduce the rotation correlation time scale of the

fluorophore in the reference frame of the macromolecule, since a smaller solid angle would be sampled by the transition dipole moment.

The angular and spatial immobilization of fluorophores to the macromolecule could be further improved by a bivalent attachment via two short linkers (Forkey et al., 2000; Peterman et al., 2001), but this would make site-specific labeling more complicated as more than one amino acid must be mutated.

In summary, the requirements of the Dale-Eisinger-Blumberg theory could be satisfied better than it is done nowadays, if long carbon chain linkers were used mainly to reduce anisotropy effects in FRET. By using short linkers one could reduce possible dynamic averaging over an ensemble of fluorophore positions, which is not accounted for in the theory. Also correlations between fluorophore orientations and fluorophore positions (VanBeek et al., 2007) are a complication, which should be avoided in experiments by a more rigid immobilization of the fluorophores to the macromolecule.

Although one might argue that the Dale-Eisinger-Blumberg theory is also valid when the fluorescence lifetime is shorter than the translational diffusion time in the accessible volume, the distribution of fluorophore positions is not known a priori. For example, the fluorophore might stick to several positions on the surface of the macromolecule, or it might be located far from the surface. The distribution of fluorophore positions, however, does influence the observed distribution of FRET efficiencies. The same criticism applies to a static average of FRET efficiencies that would have to be applied when the time resolution of the experiment is not high enough to resolve dynamic inhomogeneities on a time scale longer than the fluorescence lifetime.

In this spirit it would be desirable to attach the fluorophores in a well defined position and orientation. However, even when the fluorophores are rigidly attached the development of an accurate likelihood function derived from first principles is needed, since there might be still static and dynamic structural inhomogeneities of the observed macromolecules. The fluorescence lifetimes that are on the order of a couple of nanoseconds for fluorophores used typically in single-molecule FRET experiments could thus enable to study structural changes in the macromolecules that occur on a longer time scale. In future NPS experiments, this information could be then interpreted without the complications caused by long linkers and give detailed insight in the conformational dynamics of the studied macromolecules in three dimensions.

Finally, the analysis of well-defined FRET states promises to improve the resolution of FRET-based structural studies substantially.

6.7.3 Interpretation of NPS results

The following section deals with the interpretation of the marginal posterior probability densities computed with the NPS. The possible effects of invalid prior assumptions and data outliers will be treated in the first part, whereas in the second part, the interpretation of the posterior based on credible volumes will be discussed, since it was used often in this work.

Invalid prior and data

In any experiment, one should make sure that the assumed conditions apply, and that the acquired data does not contain any outliers or systematic errors. In the case of NPS, besides the failure of Dale-Eisinger-Blumberg theory discussed already in section 6.7.2, several other sources of systematic errors are possible.

A possible error source that affects only the prior assumptions is an incorrect crystal structure. One or more satellite position priors would then be shifted away from their

"true" position, and probably also the shape of the "true" and simulated accessible volume would differ. It is difficult to predict how strong such erroneous satellite priors would influence the inferred antenna positions, since the effect depends on the FRET network structure and the number of satellites.

Even when the structures of the individual macromolecules in the complex are correct, all complexes or just a fraction might be assembled in a wrong way. While the latter case results in a possible misassignment of peaks in FRET efficiency or FRET anisotropy histograms, invalid data is definitely acquired in the first case. In any of the scenarios above, the result of the inference might not reflect a real position of the antenna fluorophore. Similarly, like in the case of incorrectly assumed satellite positions, it is hard to predict the final localization accuracy.

Incorrectly assembled complexes are an even greater problem in bulk measurements, since they cannot be easily detected. In particular, such errors might affect the bulk fluorescence anisotropies when fluorophores exhibit different anisotropies in different complexes or complex constituents. Time-resolved fluorescence anisotropies should be hence determined in single-molecule measurements, if possible.

Presumably, systematic errors will have in general a larger impact when the FRET network is analyzed globally. At least in the case of FRET efficiencies and anisotropies, such errors should be easier to detect with the consistency check proposed (section 5.3.6), since it is more likely to produce a contradiction in the result. Yet, the consistency check will work only, when there is enough valid data and when a sufficient amount of satellite positions is assigned correctly (section 6.2.2). These conditions could be established by control measurements between different satellite attachment sites, which has other positive effects and will be discussed later (section 6.7.4).

One could also try to account for outliers and invalid prior assumptions in the likelihood and prior assignment. Instead of being purely Gaussian, each likelihood factor could be composed of a constant "background" and a Gaussian. The "background" would model possible outliers in the data, while the Gaussian would describe a correct measurement. Similarly, the satellite position priors could be assigned to be less informative and in that way, incorrect satellite positions could be modeled.

Yet, since the influence of the prior was found to be strong in all experiments shown in this work (sections 6.2, 6.4 and 6.6), a less informative satellite position prior and likelihood would lead to an even less peaked marginal antenna position posterior. This would express the general distrust in the data and the prior assumptions rather than help to detect and eliminate inconsistencies automatically. However, when more informative data is available, such modifications of the prior and the likelihood might be useful.

In addition, the prior assumptions and the measured data should be verified also by other methods than FRET, if possible.

Credible volumes

Even when the data and prior assumptions are correct, one has to be cautious when a probability density is interpreted in terms of credible intervals. This applies to all marginal posterior distributions, and in particular to the marginal fluorophore position posterior densities computed with NPS.

A credible interval is defined as the smallest volume that contains a certain probability, the credibility level (see section 2.4.4). When one would analyze many random FRET networks, and when the influence of the prior on the shape of the posterior would be negligible, the frequency of the "true" parameters being inside the volume should approximately correspond to the credibility level.
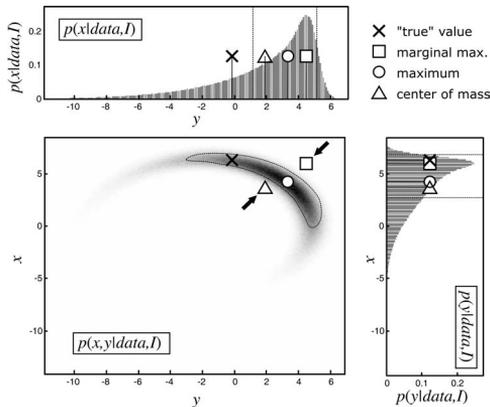
Figure 6.43: Misunderstandings in marginal posterior interpretation. In the case of a strange shaped posterior, $p(x, y|data, I)$ (grayscale), the "true" set of parameters (cross) might lie far away from its maximum (circle) but may still have reasonable probability density. The positions corresponding to the maxima of the marginalized density (squares) and the average values of x and y (triangles) are indicated by arrows and might have very low probability values. Credible intervals, like the 68% credible interval shown here (dashed lines), are just a way to display the posterior and should not be overinterpreted.

On the other hand, one could define an infinite number of other credible volumes (i.e. not the smallest one), and still, the above argument would be valid. Because of that, one has to be comfortable with the plain interpretation of the surfaces of credible volumes as a useful way to display iso-levels of the posterior. It is the complete posterior probability distribution that is, by definition, the result of a Bayesian parameter estimation.

Nevertheless, the credible volumes were used in order to be comparable with standard methods of data analysis. For example, a 68% and 95% credible interval corresponds to an error margin of one and two standard deviations, respectively, when the posterior is normal distributed.

In the same way, one should be aware of pitfalls that might occur when the proximity of the "true" parameters to the maximum of a posterior probability density is interpreted, especially when the posterior density was marginalized. In figure 6.43, a peculiar case is shown, in which the "true" parameters exhibit a fairly high posterior probability density and lie within the 68% credible volume, while being outside of the 68% credible interval of one of the marginalized posterior densities. Due to the special shape of the posterior also the point defined by the maxima of the marginal posterior densities as well as the center of mass of the posterior exhibit very low posterior density values and cannot be used to summarize the posterior in a few numbers. It is likely that exactly this effect is observed in the test calculations of section 6.3.3, in which only FRET efficiency data was analyzed. There, the simulated antenna positions were found at asymmetrical locations in the marginal posterior.

However, if there is enough informative data to concentrate the posterior density in a single spot in the parameter space, the posterior might be approximated well by a multivariate Gaussian. In that case, the effect discussed above will not occur, since the posterior maximum coincides with the point defined by the maxima of the marginalized posteriors, which are also described approximately by Gaussians.

As discussed in the previous part, it is desirable that the posterior probability is well localized in the parameter space. In the next section, a discussion of the FRET network architecture will be given in respect thereof.

N_{sat} \ N_{ant}	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	-2	-4	-6	-8	-10	-12	-14	-16	-18	-20	-22	-24	-26	-28	-30
2	-1	-2	-3	-4	-5	-6	-7	-8	-9	-10	-11	-12	-13	-14	-15
3	0*	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
5	2	4	6	8	10	12	14	16	18	20	22	24	26	28	30
6	3	6	9	12	15	18	21	24	27	30	33	36	39	42	45
7	4	8	12	16	20	24	28	32	36	40	44	48	52	56	60
8	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75
9	6	12	18	24	30	36	42	48	54	60	66	72	78	84	90
10	7	14	21	28	35	42	49	56	63	70	77	84	91	98	105
11	8	16	24	32	40	48	56	64	72	80	88	96	104	112	120
12	9	18	27	36	45	54	63	72	81	90	99	108	117	126	135
13	10	20	30	40	50	60	70	80	90	100	110	120	130	140	150
14	11	22	33	44	55	66	77	88	99	110	121	132	143	154	165
15	12	24	36	48	60	72	84	96	108	120	132	144	156	168	180

Table 6.4: Determinacy $N_{\text{obs}} - N_{\text{unk}}$ for FRET efficiency-only networks with fluorescence anisotropies equal 0 in dependence of number of satellites N_{sat} and number of antennas N_{ant}. Negative and zero determinacies are printed on a dark gray and light gray background, respectively. The minimal overdetermined networks is denoted by a star.

6.7.4 Optimal FRET network design

Besides the profound understanding of the measurement process and a realistic model for FRET between fluorophores attached to a macromolecule (section 6.7.2), the information content of the data must not be neglected. Although it is possible in Bayesian data analysis to account for nuisance parameters, like for instance the unknown average orientation of fluorophore transition dipole moments or the satellite fluorophore positions, one should be cautious when there is only few data available or when the data is not very informative. In general, the inference depends not only on the data but also on the background information, which might be dominant in such cases, causing the problem to be *ill-defined*. It is thus desirable to design experiments capable of inferring fluorophore positions and average transition dipole moment orientations based mainly on the data. A critical aspect to accomplish this is the FRET network architecture, which will be discussed in the following.

Minimal well-defined FRET networks

It is interesting to ask, how many measured FRET efficiencies and FRET anisotropies are necessary in order to have a *well-defined* inference problem.

It will be assumed in the following that the measurement error and the size of the satellite position priors are negligibly small, which is an approximation to the real experiment. In the best case of all residual fluorescence anisotropies are larger than 0.1, the likelihood will consist of a product of likelihood factors, one for each data point (equations (3.2.3) and (3.2.4)). Each likelihood factor, no matter whether it represents FRET efficiency or anisotropy data, defines in general a hyper-surface[6] in the parameter space of fluorophore positions and average transition dipole moment orientations. All parameter space points that are part of this hyper-surface will produce the same FRET efficiency or anisotropy value equal to the measured data. The region of the parameter space consistent with all measurements is hence the intersection of all hyper-surfaces.

Now, a *well-defined FRET network* consists of enough measurements, so that all hyper-surfaces intersect in exactly a single point of dimensionality 0. In the absence of mea-

---

[6]When the parameter space has N dimensions, in general, the hyper-surface is of dimensionality $N-1$.

N_sat \ N_ant	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	-6	-10	-14	-18	-22	-26	-30	-34	-38	-42	-46	-50	-54	-58	-62
2	-7	-10	-13	-16	-19	-22	-25	-28	-31	-34	-37	-40	-43	-46	-49
3	-8	-10	-12	-14	-16	-18	-20	-22	-24	-26	-28	-30	-32	-34	-36
4	-9	-10	-11	-12	-13	-14	-15	-16	-17	-18	-19	-20	-21	-22	-23
5	-10	-10	-10	-10	-10	-10	-10	-10	-10	-10	-10	-10	-10	-10	-10
6	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3
7	-12	-10	-8	-6	-4	-2	0*	2	4	6	8	10	12	14	16
8	-13	-10	-7	-4	-1	2*	5	8	11	14	17	20	23	26	29
9	-14	-10	-6	-2	2*	6	10	14	18	22	26	30	34	38	42
10	-15	-10	-5	0*	5	10	15	20	25	30	35	40	45	50	55
11	-16	-10	-4	2	8	14	20	26	32	38	44	50	56	62	68
12	-17	-10	-3	4	11	18	25	32	39	46	53	60	67	74	81
13	-18	-10	-2	6	14	22	30	38	46	54	62	70	78	86	94
14	-19	-10	-1	8	17	26	35	44	53	62	71	80	89	98	107
15	-20	-10	0	10	20	30	40	50	60	70	80	90	100	110	120

Table 6.5: Determinacy $N_\text{obs} - N_\text{unk}$ for FRET efficiency-only networks in dependence of number of satellites, N_sat, and number of antennas N_ant. Negative and zero determinacies are printed on a dark gray and light gray background, respectively. Minimal overdetermined networks are denoted by a star.

surement errors and when the satellite positions are known exactly, this point will correspond to the "true" positions and average transition dipole moment orientations of all fluorophores. When the measurements are not redundant, the intersection of all hypersurfaces has the dimensionality of either 0 or $N_\text{unk} - N_\text{obs}$, whichever is larger. N_unk denotes the number of unknown parameters, and N_obs is the number of measurements. Yet, due to symmetries, the intersection could consist of several points, even though its dimensionality is 0. In a well-defined FRET network, the number of data must be hence at least equal to the number of unknowns.

In the following, typical FRET network architectures will be discussed in respect of their determinacy, i.e. the number of unknowns subtracted from the number of observables, $N_\text{obs} - N_\text{unk}$. It will be assumed, that all possible satellite-antenna pairs are measured. A positive determinacy indicates that the FRET network is probably well-defined, and the FRET network will be called *overdetermined*. A negative determinacy suggests a possibly ill-defined inference problem, and the FRET network is termed *underdetermined*. A FRET network of determinacy zero will be treated here as overdetermined for simplicity.

If fluorescence anisotropies were 0, then the orientation factor κ^2 would be 2/3 for all fluorophores, and there would be no orientation effects at all. It would be hence unnecessary to define average transition dipole moment orientations. The number of unknowns in a FRET network consisting of N_sat satellite fluorophores and N_ant antenna fluorophores would be simply given by $N_\text{unk} = 3N_\text{ant}$, since only the antenna positions have to be determined. If all possible FRET efficiencies were measured, the number of observables were $N_\text{obs} = N_\text{ant} N_\text{sat}$. The resulting determinacy of networks with different N_sat and N_ant is shown in in table 6.4. It is obvious that no matter how many antenna fluorophores there are, three or more satellites are enough to calculate the position of the antenna, given every possible satellite-antenna combination is measured.

In the opposite case, when florescence anisotropies of all fluorophores were non-zero, the number of unknowns is given by $N_\text{unk} = 5N_\text{ant} + 2N_\text{sat}$, since in addition to the antenna positions, also the average transition dipole moment orientations of all antennas and satellites have to be determined. The determinacy of such a network has a completely different behavior when compared to the previous scenario without anisotropy effects (Table 6.5). In particular, the introduction of more satellites leads to a decrease of determinacy when only a single antenna fluorophore is in the FRET network, as for each new measurement

N_{sat} \ N_{ant}	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	-5	-8	-11	-14	-17	-20	-23	-26	-29	-32	-35	-38	-41	-44	-47
2	-5	-6	-7	-8	-9	-10	-11	-12	-13	-14	-15	-16	-17	-18	-19
3	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9
4	-5	-2	1*	4	7	10	13	16	19	22	25	28	31	34	37
5	-5	0*	5	10	15	20	25	30	35	40	45	50	55	60	65
6	-5	2	9	16	23	30	37	44	51	58	65	72	79	86	93
7	-5	4	13	22	31	40	49	58	67	76	85	94	103	112	121
8	-5	6	17	28	39	50	61	72	83	94	105	116	127	138	149
9	-5	8	21	34	47	60	73	86	99	112	125	138	151	164	177
10	-5	10	25	40	55	70	85	100	115	130	145	160	175	190	205
11	-5	12	29	46	63	80	97	114	131	148	165	182	199	216	233
12	-5	14	33	52	71	90	109	128	147	166	185	204	223	242	261
13	-5	16	37	58	79	100	121	142	163	184	205	226	247	268	289
14	-5	18	41	64	87	110	133	156	179	202	225	248	271	294	317
15	-5	20	45	70	95	120	145	170	195	220	245	270	295	320	345

Table 6.6: Determinacy $N_{\text{obs}} - N_{\text{unk}}$ for FRET networks with measured FRET efficiency and anisotropy in dependence of number of satellites, N_{sat}, and number of antennas, N_{ant}. Negative and zero determinacies are printed on a dark gray and light gray background, respectively. Minimal overdetermined networks are denoted by a star.

two unknown satellite average transition dipole moment orientations are introduced, while the number of data increases only by one (see table 6.5, first column).

Often, the number of labeling sites is limited, or it is hard to introduce additional labels. In this case, it is reasonable to choose the experimental conditions, so that the FRET network is well defined and the number of fluorophores is minimized. Such FRET networks will be called *minimal overdetermined FRET networks*. When only FRET efficiency data is measured and the fluorescence anisotropies are non-zero, a network with 14 fluorophores in total is minimal when $N_{\text{sat}} = 7 \ldots 10$ and $N_{\text{ant}} = 7 \ldots 4$.

When FRET anisotropy data is measured in addition, the size of such a well-defined FRET network will shrink, since there are twice as much observables as compared to networks in which only FRET efficiency data is available (see table 6.6). In that case, the minimal overdetermined FRET network has only 7 fluorophores in total, which are realized in two cases with $N_{\text{sat}} = 4$ / $N_{\text{ant}} = 3$ and $N_{\text{sat}} = 5$ / $N_{\text{ant}} = 2$ fluorophores.

In an experiment, which comprises of N_{dock} macromolecules that are docked to the macromolecule 0 by means of FRET, all satellites and antennas are located at precisely known positions on the respective macromolecules. The number of unknowns is then $N_{\text{unk}} = 6N_{\text{dock}} + 2(N_{\text{sat}} + N_{\text{ant}})$, since there are 6 additional unknowns originating in the position and orientation of each docked macromolecule, but only two unknowns per fluorophore, namely the angles defining the average transition dipole moment orientations. When there is only one macromolecule with unknown position and orientation (i.e. $N_{\text{dock}} = 1$), then the minimum number of fluorophores needed for a well-defined network is 11 ($N_{\text{sat}} = 4 \ldots 7$ and $N_{\text{ant}} = 7 \ldots 4$) and 6 ($N_{\text{sat}} = 3$ and $N_{\text{ant}} = 3$), respectively. The determinacy is listed in table 6.7 for the docking via FRET efficiency data only, and in table 6.8 when FRET efficiencies and anisotropies are measured.

It is obvious that the influence of the orientation effects will continuously decrease the smaller the fluorescence anisotropy is, and that there should be a smooth transition between tables 6.4, 6.5 and 6.6. In that case, a definition of the importance of the spatial and angular parameters is necessary, and the errors of the data should be accounted for as well.

Still, it is interesting to compute the determinacy of the FRET networks analyzed in sections 6.1–6.6. To this end, the residual fluorescence anisotropy was assumed to have a strong influence on the measured FRET efficiency, and all measurements between two

N_{sat} \\ N_{ant}	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	-9	-10	-11	-12	-13	-14	-15	-16	-17	-18	-19	-20	-21	-22	-23
2	-10	-10	-10	-10	-10	-10	-10	-10	-10	-10	-10	-10	-10	-10	-10
3	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3
4	-12	-10	-8	-6	-4	-2	0*	2	4	6	8	10	12	14	16
5	-13	-10	-7	-4	-1	2*	5	8	11	14	17	20	23	26	29
6	-14	-10	-6	-2	2*	6	10	14	18	22	26	30	34	38	42
7	-15	-10	-5	0*	5	10	15	20	25	30	35	40	45	50	55
8	-16	-10	-4	2	8	14	20	26	32	38	44	50	56	62	68
9	-17	-10	-3	4	11	18	25	32	39	46	53	60	67	74	81
10	-18	-10	-2	6	14	22	30	38	46	54	62	70	78	86	94
11	-19	-10	-1	8	17	26	35	44	53	62	71	80	89	98	107
12	-20	-10	0	10	20	30	40	50	60	70	80	90	100	110	120
13	-21	-10	1	12	23	34	45	56	67	78	89	100	111	122	133
14	-22	-10	2	14	26	38	50	62	74	86	98	110	122	134	146
15	-23	-10	3	16	29	42	55	68	81	94	107	120	133	146	159

Table 6.7: Determinacy $N_{\mathrm{obs}} - N_{\mathrm{unk}}$ for FRET efficiency-assisted docking of one macromolecule in dependence of number of satellites, N_{sat} (on macromolecule 0), and number of antennas, N_{ant} (on macromolecule 1). Negative and zero determinacies are printed on a dark gray and light gray background, respectively. Minimal overdetermined networks are denoted by a star.

N_{sat} \\ N_{ant}	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8
2	-8	-6	-4	-2	0	2	4	6	8	10	12	14	16	18	20
3	-8	-4	0*	4	8	12	16	20	24	28	32	36	40	44	48
4	-8	-2	4	10	16	22	28	34	40	46	52	58	64	70	76
5	-8	0	8	16	24	32	40	48	56	64	72	80	88	96	104
6	-8	2	12	22	32	42	52	62	72	82	92	102	112	122	132
7	-8	4	16	28	40	52	64	76	88	100	112	124	136	148	160
8	-8	6	20	34	48	62	76	90	104	118	132	146	160	174	188
9	-8	8	24	40	56	72	88	104	120	136	152	168	184	200	216
10	-8	10	28	46	64	82	100	118	136	154	172	190	208	226	244
12	-8	12	32	52	72	92	112	132	152	172	192	212	232	252	272
13	-8	14	36	58	80	102	124	146	168	190	212	234	256	278	300
14	-8	16	40	64	88	112	136	160	184	208	232	256	280	304	328
15	-8	18	44	70	96	122	148	174	200	226	252	278	304	330	356
16	-8	20	48	76	104	132	160	188	216	244	272	300	328	356	384

Table 6.8: Determinacy $N_{\mathrm{obs}} - N_{\mathrm{unk}}$ for FRET efficiency and anisotropy-assisted docking of one macromolecule in dependence of number of satellites, N_{sat} (on macromolecule 0), and number of antennas, N_{ant} (on macromolecule 1). Negative and zero determinacies are printed on a dark gray and light gray background, respectively. Minimal overdetermined networks are denoted by a star.

labeling sites were treated as only one data point. The computed determinacies are shown in table 6.9.

All experimentally studied FRET networks have a negative determinacy, while the networks studied in the test calculations exhibiting small localization uncertainties have positive determinacies. The network used in the nontemplate DNA study has the highest determinacy (-4) of all experimentally studied networks. However, as shown in simulations and applications in this thesis, it is possible to determine the positions of antenna fluorophores even when the determinacy of the FRET network is negative, though these estimates are usually not limited by the measurement accuracy but rather by insufficient information on the average transition dipole moment orientations. In this case, the inferred positions depend to a certain degree on prior information and will be correct as long as the prior assumptions are true.

Although the FRET network determinacy quantifies whether the positions and average transition dipole moment orientations of all fluorophores can be determined, it cannot be ruled out that the position and orientation of a particular fluorophore can be determined accurately. It could be even possible in this way to determine both the position and average transition dipole moment orientation of a single antenna by measuring FRET efficiencies to many satellite fluorophores. The inference would be then based on rather "statistical" reasoning but might be correct nevertheless. In this context it would be interesting to study the influence the network structure on the inferred antenna fluorophore positions and average transition dipole moment orientations in a systematic and quantitative way.

In the above discussion, FRET efficiency and FRET anisotropy data was treated equally, but it is clear that it would never be possible to infer antenna positions from FRET anisotropy measurements alone. However, since the FRET efficiency depends on the average transition dipole moment orientations, FRET anisotropy can be used to improve the localization accuracy, as shown already in section 6.3.3. In certain cases, it might be possible that some marginal antenna position posteriors are more compact even though the FRET network has a determinacy smaller than zero. The opposite case might occur as well by chance, and a FRET network might be ill-defined although its determinacy is positive.

Hence, when it is really of interest whether the posterior is concentrated in one small region of parameter space, one should check this individually. For this analysis, a more general definition of dimensionality will be needed, since the posterior is a probability density that is spread in all parameter space dimensions. Furthermore, such a generalized dimensionality must be inferred from the available posterior samples.

While a positive determinacy cannot guarantee that a FRET network is well-defined, it is yet a good indicator for a well-defined FRET network, as shown above. In the next part, it will be demonstrated, how the determinacy can be increased without the introduction of additional labeling sites.

Measurement platforms

Since the unknown average transition dipole moment orientation of the fluorophores involved in the measurement increases the localization uncertainty drastically, a measurement platform with known average transition dipole moment orientations is advantageous. This can be established by measuring FRET efficiency or/and anisotropy between known positions on a macromolecule.

To this end, FRET should be measured twice between all possible combinations ij of labeling sites on a macromolecule with known structure, once with the donor and acceptor

experimental network	N_{sat}	N_{ant}	N_{dock}	N_{obs}	determinacy
RNA(+1)	5	1	0	5	-10
RNA(+29) without TFIIB	7	1	0	7	-12
RNA(+29) with TFIIB	5	1	0	5	-10
ntDNA	8	7	0	47	-4
ITC	8	5	2	20	-18

test calculation	N_{sat}	N_{ant}	N_{dock}	N_{obs}	determinacy
separate	7	1	0	7	-12
separate, extended network	8	1	0	8	-13
global	7	7	0	49/37*	0/-12*
global, extended network	9	7	0	55/43*	+2/-10*
global, FRET anisotropy	7	7	0	98/86*	+49/37*
docking	7	7	1	49/37*	+15/+3*
docking, FRET anisotropy	7	7	1	98/86*	+64/+52*

Table 6.9: Determinacies of the analyzed FRET networks. The values marked with a * were obtained by ignoring measurements with FRET efficiency smaller than 0.1.

attached to the sites i and j, respectively, and the other time with interchanged donor and acceptor fluorophores. When the satellite position uncertainty is negligible and when there are N_{sat} labeling sites in total, the number of unknowns is $N_{\text{unk}} = 4N_{\text{sat}}$, and the number of observables measured in between the satellites is $N_{\text{obs}} = N_{\text{sat}}(N_{\text{sat}} - 1)$ when either FRET efficiency or anisotropy (E xor A) is measured and $N_{\text{obs}} = 2N_{\text{sat}}(N_{\text{sat}} - 1)$ when both FRET efficiency and anisotropy (E and A) is measured. One should be able to determine the average orientations of all fluorophores within the measurement platform when $N_{\text{obs}} > N_{\text{unk}}$ and thus when more than 5 (E xor A[7]) and 3 (E and A) labeling sites are used (figure 6.44).

Note that several types of fluorophores can be used in the final network, which consists of a donor and acceptor fluorophore at each satellite labeling site and one or more antenna fluorophores. For example, the antennas could be donors that are matched well to the acceptors on the macromolecule. Consequently, the additional fluorophores used to set up the measurement platform must be donors as well, but it is not necessary that they have the same chemical structure like the antenna donors. In that way, it should be possible to optimize the Förster distances between the acceptors and both types of donor fluorophores.

In addition, the measurements between known positions in a structure can be used to control whether the macromolecule structure is correct.

After the discussion of an optimal FRET network architecture, which is relevant for the analysis and planning of any complex FRET localization experiment, the next section will focus on the pre-processing of the FRET efficiency data, and how it can be coupled to the NPS in its present form.

---

[7] When FRET anisotropy is the only data, the average transition dipole moment orientations can be inferred only relative to each other. The absolute orientation in the reference frame of the macromolecule is not accessible by such measurements, unless the absolute orientation of at least one average transition dipole moment is known.

Figure 6.44: Example of a measurement platform. A donor and an acceptor fluorophore are attached to each of four satellite labeling sites (site 1 . . . 4). Each of the 12 possible FRET efficiency and anisotropy measurements between the satellites and 4 additional measurements to an antenna fluorophore (here a donor) are performed (dashed lines). The determinacy of this network is $N_{\text{obs}} - N_{\text{unk}} = +11$.
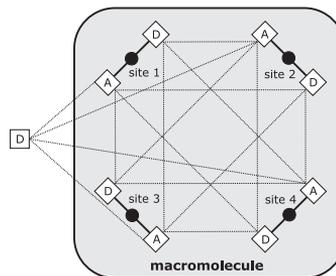
6.7.5 Data pre-processing and model selection

As shown in chapter 3, the NPS is based upon pre-processed FRET data, i.e. average FRET efficiencies and anisotropies, which depend on the distance between average donor and acceptor positions and the angle between the average transition dipole moment orientations. It is left open, how the pre-processing accomplishes to describe the "raw" FRET data "sufficiently well" by a yet unknown number of macromolecule states each with unknown average FRET efficiency and anisotropy.

In several recent publications the FRET efficiency time traces (McKinney et al., 2006) or the underlying donor and acceptor fluorescence signals (Liu et al., 2010b; Bronson et al., 2009) were analyzed. These authors addressed the questions how to model the macromolecule states, how many states to assign, and tried to quantify the above term "sufficiently well". Hidden-Markov-modeling (HMM) was used in these studies to characterize the average FRET efficiency of each state and in order to extract the transitions rates between the states. The data was analyzed with maximum-likelihood estimation (McKinney et al., 2006; Liu et al., 2010b) or with Bayesian data analysis (Bronson et al., 2009).

Other popular methods to extract the average FRET efficiency of different states without modeling the underlying dynamics explicitly are Gaussian fits of FRET efficiency histograms, photon distribution analysis (PDA), and proximity ratio histograms (PRH) (see section 6.7.1 for references).

Typically, the number of states is obtained by some heuristic rule based on the goodness of the fit. In Bayesian data analysis, there are better ways to deal with the selection of models according to their probability, which was done by Bronson et al. (2009). They calculated the probability of a model with a specific number of states given the experimental data, which is proportional to the evidence obtained in the pre-processing analysis.

However, seen from the point of view of Bayesian parameter estimation, all pre-processing methods listed above assign a flat prior in the parameters that describe a FRET state, which is in particular the FRET efficiency E. This choice seems to be rather arbitrary, since one could work as well with a flat prior in other parametrizations, e.g. $\ln(E)$ instead of E. However, a flat prior in E represents a different state of information than a flat prior in $\ln(E)$, which is inconsistent when absolutely no information about the FRET efficiency of a state is available.

As shown in section 2.4.3, a natural parametrization can be obtained from transformation invariances, and a maximum entropy prior in this parametrization truly expresses the experimenter's ignorance. This was done for NPS (section 3.2.1) and resulted in flat priors in the fluorophore positions and average transition dipole moment orientations.

Conversely, this means that the shape of the prior in the parameters that describe a FRET state, i.e. the FRET efficiencies, will not be flat in the pre-processing analysis. One can therefore reason that a non-flat prior could change the probabilities of concurring

models with different number of states. This, in turn, would have implications on the biochemical interpretation of the experiments.

As a consequence, one should couple NPS with a pre-processing method to infer the number of states of a macromolecule. This can be done by modifying the likelihood of the NPS analysis with the results of any pre-processing method that utilizes a flat prior in FRET efficiency and anisotropy, as will be shown in the following.

Coupling of NPS and FRET data preprocessing methods

In the following, it is assumed that the macromolecule and the attached fluorophores are described by a model, M, with K_M states in total[8]. A particular macromolecule state k is quantified by the positions $\boldsymbol{x}_{i;k}$ and the average transition dipole orientations $\boldsymbol{\Omega}_{i;k}$ of each fluorophore i, as well as by the potentially ambiguous signs of the average axial depolarizations, $s_{i;k}$. In the following, these structural parameters of NPS will be abbreviated by $q_{i;k}^{\mathrm{NPS}} := (\boldsymbol{x}_{i;k}, \boldsymbol{\Omega}_{i;k}, s_{i;k})$, and $\{q_{i;k}^{\mathrm{NPS}}\}$ will be the structural parameters of all fluorophores in all possible macromolecule states. The observables measured for the fluorophore pairs ij, for example the FRET efficiency, are available as time series, i.e. as FRET efficiency time traces, and will be denoted by $O_{ij}(t)$. The ensemble of the time traces of all measured molecules and FRET pairs ij will be denoted by $\{O_{ij}(t)\}$.

Now, in order to infer the structural parameters of the model M, one must compute the posterior probability density $p\left(\{q_{i;k}^{\mathrm{NPS}}\}|\{O_{ij}(t)\}, M, I\right)$, which is done with Bayes' theorem,

$$p\left(\{q_{i;k}^{\mathrm{NPS}}\}|\{O_{ij}(t)\}, M, I\right) = \frac{1}{Z_M}\ p\left(\{O_{ij}(t)\}|\{q_{i;k}^{\mathrm{NPS}}\}, M, I\right)\ p\left(\{q_{i;k}^{\mathrm{NPS}}\}|M, I\right), \quad (6.7.1)$$

where Z_M is the evidence of the model M, the likelihood of the time series data is denoted by $p\left(\{O_{ij}(t)\}|\{q_{i;k}^{\mathrm{NPS}}\}, M, I\right)$, and $p\left(\{q_{i;k}^{\mathrm{NPS}}\}|M, I\right)$ is the prior. Note that the likelihood depends on the "raw data", and not on pre-processed observables, in contrast to equation (3.2.5), and that the prior $p\left(\{q_{i;k}^{\mathrm{NPS}}\}|M, I\right)$ can be split into a contribution of the depolarization signs, $s_{i;k}$, and a contribution of fluorophore positions, $\boldsymbol{x}_{i;k}$, and average transition dipole moment orientations, $\boldsymbol{\Omega}_{i;k}$,

$$p\left(\{q_{i;k}^{\mathrm{NPS}}\}|M, I\right) = \frac{1}{\prod_{l;k} S_{l;k}}\ p\left(\{\boldsymbol{x}_{i;k}, \boldsymbol{\Omega}_{i;k}\}|M, I\right). \quad (6.7.2)$$

Here, $\prod_{l;k} S_{l;k}$ is the number of combinations of possible average axial depolarization signs (similar to equation (3.2.5)) and $p\left(\{\boldsymbol{x}_{i;k}, \boldsymbol{\Omega}_{i;k}\}|M, I\right)$ is the regular NPS prior (section 3.2.3).

The most probable macromolecule model can be found by maximizing the probability of the model M given the experimental data $\{O_{ij}(t)\}$. This probability can be related to the evidence Z_M of the model, again by using Bayes' theorem,

$$p\left(M|\{O_{ij}(t)\}, I\right) \propto p\left(\{O_{ij}(t)\}|M, I\right) p\left(M|I\right) = Z_M\ p\left(M|I\right), \quad (6.7.3)$$

where $p\left(M|I\right)$ denotes the prior probability of the model M with K_M states.

As shown in the appendix, one can rewrite the likelihood $p\left(\{O_{ij}(t)\}|\{q_{i;k}^{\mathrm{NPS}}\}, M, I\right)$ in equation (6.7.1) into an expression that contains the posterior of a completely independent data pre-processing analysis, e.g. HMM, PDA, PRH, or the simple fits of FRET efficiency histograms. When the applied method is not based on Bayesian parameter estimation, the analogue of a posterior density must be found (Sivia, 2006, chapter 3.5). The pre-

---

[8]Note that there might exist several different models with the same number of states.

processing must compute the estimates of the "true" values of the observables, $\{\tilde{O}_{ij;k}\}$, that characterize the states $k = 1 \ldots K_M$, and it must utilize a flat prior in $\{\tilde{O}_{ij;k}\}$. The coupling between NPS and the pre-processing method turns out to be

$$p\left(\{O_{ij}(t)\}|\{q_{i;k}^{\text{NPS}}\}, M, I\right) = \frac{Z_M^{\text{PP}}}{\pi_M^{\text{PP}}} \, p\left(\{\tilde{O}_{ij;k}\}|\{O_{ij}(t)\}, M, I\right)\Big|_{\tilde{O}_{ij;k}=\mathcal{O}_{ij;k}\left(q_{i;k}^{\text{NPS}}, q_{j;k}^{\text{NPS}}\right)},$$
(6.7.4)

where Z_M^{PP} is the evidence of the pre-processing inference, π_M^{PP} denotes the value of the flat prior in the "true" observables $\{\tilde{O}_{ij;k}\}$, i.e. $\pi_M^{\text{PP}} = p\left(\{\tilde{O}_{ij;k}\}|M, I\right)$, and $p\left(\{\tilde{O}_{ij;k}\}|\{O_{ij}(t)\}, M, I\right)$ is the posterior of the pre-processing analysis. The observables $\mathcal{O}_{ij;k}\left(q_{i;k}^{\text{NPS}}, q_{j;k}^{\text{NPS}}\right)$ expected from NPS (i.e. $\mathcal{E}_{ij;k}$ or $\mathcal{A}_{ij;k}$) are functions of the positions, average transition dipole moment orientations and average axial depolarization signs of the fluorophores i and j. $\mathcal{O}_{ij;k}\left(q_{i;k}^{\text{NPS}}, q_{j;k}^{\text{NPS}}\right)$ is substituted for the "true" observables $\tilde{O}_{ij;k}$ in the pre-processing posterior.

As a consequence of equations 6.7.1 and 6.7.4, the task to identify a fixed number of states is left to the pre-processing algorithm, whereas the computation of possible fluorophore positions and average transition dipole moment orientations is done by NPS. The posterior obtained by the pre-processing thus plays the role of the likelihood in the NPS calculation.

The probability of the model M can be then computed as follows:

$$p\left(M|\{O_{ij}(t)\}, I\right) \propto \frac{Z_M^{\text{PP}}}{\pi_M^{\text{PP}}} Z_M^{\text{NPS}} K_M! \cdot p\left(M|I\right).$$
(6.7.5)

Z_M^{NPS} is the evidence of the NPS problem, calculated with the posterior of the data pre-processing, $p\left(\{\tilde{O}_{ij;k}\}|\{O_{ij}(t)\}, M, I\right)\Big|_{\tilde{O}_{ij;k}=\mathcal{O}_{ij;k}}$, substituted for $\prod_{ij \in M} L_{ij}^{(s_i s_j)}\left(\boldsymbol{x}_i, \boldsymbol{\Omega}_i, \boldsymbol{x}_j, \boldsymbol{\Omega}_j\right)$ in equation (3.2.5). The factor $K_M!$ accounts for the multimodality of $p\left(\{\tilde{O}_{ij;k}\}|\{O_{ij}(t)\}, M, I\right)\Big|_{\tilde{O}_{ij;k}=\mathcal{O}_{ij;k}\left(q_{i;k}^{\text{NPS}}, q_{j;k}^{\text{NPS}}\right)}$ as this function is invariant under permutations of $\sigma_{ij;k}$ for different conformations k[9].

It might be legitimate to approximate the posterior of the pre-processing analysis by a multivariate normal distribution (see section 2.4.4) that is characterized by $\{O_{ij;k}\}$, the best estimate of the "true" observables, and a corresponding covariance matrix, $\Delta^2 O_{ij;k,lm;n}$, describing the correlations between $\tilde{O}_{ij;k}$ and $\tilde{O}_{lm;n}$. When the covariance matrix is diagonal, the standard deviations $\Delta O_{ij;k}$ of each estimated observable $O_{ij;k}$ can be obtained directly by taking the square root of the diagonal elements, i.e. $\Delta O_{ij;k} = \sqrt{\Delta^2 O_{ij;k,ij;k}}$. The pre-processed data $O_{ij;k} \pm \Delta O_{ij;k}$ can be used in the NPS calculation as already described in section 3.2.2. Note that the coupling of NPS and a data pre-processing method was used in all analyses presented in this thesis, since fits of FRET efficiency histograms were used to extract the center FRET efficiencies and their standard deviations. When the covariance matrix is not diagonal, the likelihood in equation (3.2.5) must be modified in order to support correlated FRET efficiencies and anisotropies.

---

[9]NPS in its present form requires a unique assignment of observables to conformations and thus cannot account for this permutation symmetry.

Discussion

As shown above, NPS can be coupled to arbitrary FRET efficiency and anisotropy pre-processing methods, which was proven formally in the appendix. In short, the posterior probability distribution computed in a pre-processing analysis like PDA, PRH or HMM, plays the role of the likelihood in the NPS calculation. The evidence obtained by NPS modifies the posterior probability of a model, which, in turn, can have consequences in Bayesian model selection.

However, the question arises what kind of prior one should choose in the NPS analysis, since usually one would expect rather small changes in the macromolecule structure, e.g. movements of domains. That would correlate the a priori expected positions and average transition dipole moment orientations of the same fluorophore in different states of the macromolecule. It might be necessary to introduce a different parametrization to describe the positions of the same fluorophores in different states of the macromolecule. Such a parametrization could, for instance, model possible conformational changes as rotations about "hinges" in the protein, for example similar to mechanistic models of bent DNA (Woźniak et al., 2008).

Finally, it still has to be verified, whether NPS is useful to determine the number of states of a macromolecule. It might be possible that the influence of the evidence calculated with NPS is overwhelmed by the evidence obtained by the pre-processing method. This could happen, when, for example dynamic information is analyzed with a HMM method, since the transitions between different states might be very informative themselves.

After this concluding section, the results of the thesis will be summarized in the following chapter.

7 Summary and outlook

This thesis focused on the data analysis of fluorescence resonance energy transfer (FRET) experiments carried out to infer structural information of biological macromolecules. The purpose of such experiments is to determine the yet unknown position of one or more fluorophores, called *antennas*, based on FRET measurements between the antennas and the *satellite fluorophores*, which are attached to known positions on the macromolecule determined by high-resolution structure methods. Alternatively, the relative position and orientation of several rigid components constituting a macromolecular complex can be computed, which is referred to as FRET-assisted docking. In this case, the structure of all components must be known, and the fluorophores must be attached to known positions on the components.

As it is easier to resolve structural heterogeneities in single-molecule experiments than with standard ensemble-based methods like X-ray crystallography or NMR spectroscopy, single-molecule FRET-based localization techniques can be used whenever high-resolution ensemble methods fail due to inhomogeneity effects.

However, since FRET is caused by dipolar coupling of the electronic systems of the fluorophores, inevitable orientation effects complicate the straight-forward interpretation of FRET efficiencies in terms of distances, and therefore also hamper the localization of fluorophores based on a simple trilateration scheme. In order to account for these effects, Bayesian data analysis has been applied to resolve the challenging task of FRET-based localization of fluorophores attached to biological macromolecules, as well as to FRET-assisted docking of several macromolecules.

This was accomplished by developing the Nano-Positioning System (NPS), a model to describe FRET between pairs of several partially oriented fluorophores attached to a macromolecule (chapter 3). In the model, the fluorophores are fixed in their position and exhibit constrained, fast, and axially symmetric orientation fluctuations. The FRET efficiency depends on the distance of the fluorophores and on the length scale of energy transfer, the so-called Förster distance. The latter can be calculated from the relative orientation of the average transition dipole moments of the fluorophores and from experimentally accessible observables, among them fluorescence anisotropies of the fluorophores, which quantify the magnitude of the orientation effects.

The challenge in the data analysis was to infer the antenna positions even though none of the average transition dipole moments was known. However, even the lack of information can be mathematically represented in Bayesian data analysis, and this state of knowledge is encoded by assigning an adequate maximum entropy prior. The prior also contains information about possible fluorophore positions calculated by taking into account the constraints imposed by the known macromolecule structure. Thereafter, the prior is combined with the likelihood that is the probability to generate the measured data by a specific configuration of fluorophores, in order to yield the posterior probability density, which is the result of the inference. Thus, it is possible to combine structural information obtained by high-resolution ensemble methods like X-ray crystallography or NMR spectroscopy with FRET data measured in single-molecule experiments.

The results of NPS can be readily displayed as marginal posterior probability densities, which reflect the information about the position of a fluorophore. The location, form and fuzziness of the densities impart simultaneously the position of the fluorophore and the

uncertainty of localization. The densities can also be shown as a set of credible volumes, which are three-dimensional credible intervals, and the Bayesian equivalent of confidence intervals. By displaying them along with a high resolution structure it is possible to perceive the available information in an intuitive way.

The major advantage of NPS over related analysis methods is the possibility to account for orientation effects, which are known to be a source of uncertainty in FRET-based localization experiments but have often been ignored or argued away in the literature so far. NPS also accounts for other uncertainties like FRET measurement errors as well as the positions of fluorophores known only to a limited extent due to the attachment via carbon chain linkers.

The NPS model was realized in two versions that differ in their parametrization. The first is called *position - Förster distance* NPS model and uses the fluorophore positions and the Förster distances as model parameters (section 3.1). This model is capable of inferring the unknown position of one antenna fluorophore by analyzing FRET efficiency measurements between the antenna and an arbitrary number of satellite fluorophores. In fact, the *position - Förster distance* NPS model is a special case and an approximation of the more general second version, the *position - orientation* NPS model, which uses the fluorophore positions and the average transition dipole moment orientations as model parameters (section 3.2). The *position - orientation* NPS model is able to infer the unknown positions and average transition dipole orientations of an arbitrary number of satellites and antennas. Moreover, the *position - orientation* NPS model was applied to the docking of several macromolecules by the additional modeling of the position and orientation of the docked macromolecule (section 3.3). The approach is also capable of processing an arbitrary number of FRET efficiency measurements and potentially also FRET anisotropy data, which carries information about the angle between the average transition dipole moment orientations of the fluorophores constituting a FRET pair.

Inference calculations of both the *position - Förster distance* NPS model and the *position - orientation* NPS model can be carried out in separate custom-written Matlab software packages (section 5.2). The software packages contain graphical user interfaces and can be operated by users with only basic knowledge of NPS. Marginal position densities of fluorophores as well as the densities of arbitrary positions in the reference frame of a docked macromolecule can be easily exported to common electron density map formats, which can be loaded into molecular viewing software (subsection 5.3.5).

The Nano-Positioning System was applied to study the eukaryotic RNA Polymerase II (Pol II) elongation complex. Though many structural details are already known from high resolution X-ray crystallography studies, a large part of the nascent RNA, the nontemplate DNA and the upstream part of the template DNA could not be observed in the crystal structure (chapter 4).

NPS was successfully tested by inferring the position of a fluorophore attached to the 3'-end of the RNA (section 6.2.1). This position was already known to be located close to the active site of Pol II from the crystal structure and NPS was able to reproduce this position.

In the first project, the position of a fluorophore attached to the 5'-end of a nascent 29 nucleotides long RNA was inferred from FRET efficiency data acquired both in the presence and absence of the transcription factor IIB (TFIIB) (section 6.2.3). Interestingly, it could be shown that the RNA can be displaced by TFIIB from a position close to the dock domain on Pol II towards the Rpb4/7 heterodimer (Muschielok et al., 2008). In general, NPS can be used in this way to monitor conformational changes of macromolecules that are initiated by factors or ligands.

In the second project, FRET efficiency measurements were used to compute the position of fluorophores attached to the nontemplate DNA in the region of the transcription bubble and to the upstream part of the DNA (section 6.4). The resulting position densities were used to build a model of the complete Pol II elongation complex (Andrecka et al., 2009). However, some of the data was found to contain strong systematic errors, which might change the inferred fluorophore positions. Since the errors were caused by insufficient control of the elongation complex integrity, these experiments and the analysis should be repeated under improved experimental conditions.

In the third project, FRET efficiency data was used to infer the position and the approximate orientation of the TBP/TATA subcomplex in the initial transcribing complex (ITC) of Pol II. In that way, FRET-assisted docking was demonstrated (section 6.6).

Beside the analysis of experimental data also test calculations were performed. In section 6.1, simulations of Förster distance distributions of one FRET pair have shown that orientation effects may contribute substantially to the localization uncertainty even if fluorescence anisotropies are low. This was found to be contrary to the common belief that orientation effects are negligible in the case of fluorescence anisotropies below 0.2. Furthermore, the assumption of statistically independent Förster distance distributions made in the *position - Förster distance* NPS model was confirmed in the case of missing information about the average transition dipole moment orientations.

Test calculations with the *position - orientation* NPS model revealed that irrespective of the fluorescence anisotropies the localization accuracy remains high when the average transition dipole moment orientations of both fluorophores constituting a FRET pair is known (section 6.3). Furthermore, it was demonstrated how the localization uncertainty increases upon loss of average transition dipole moment orientation information.

The *position - Förster distance* NPS model and the *position - orientation* NPS model were compared by analyzing artificial FRET efficiency data of a simulated FRET network (section 6.6). In particular, the *position - orientation* NPS model provided results similar to those of the *position - Förster distance* NPS model when the artificial data was analyzed separately, i.e. only measurements between all satellites and a single antenna were analyzed. In contrast, the simultaneous inference of all fluorophore positions and orientations, called global analysis, was found to produce more accurate position and average transition dipole moment orientation estimates than the separate analysis. This effect was stronger when the number of informative FRET measurements was increased by adding more satellite fluorophores to the FRET network.

An overwhelming improvement in localization and orientation accuracy was observed after the analysis of combined FRET efficiency and FRET anisotropy data. This calculation ultimately proved that one of the major sources of uncertainty in FRET localization experiments is the unknown orientation of the average transition dipole moments, and at the same time pointed out how these orientations can be inferred.

At the end of the results chapter (section 6.7.1), NPS was compared to other FRET localization techniques. Thereafter, the potential advantages of a more rigid fluorophore attachment were discussed (section 6.7.2). As a result, the usage of short or even bivalent linkers in order to attach fluorophores would possibly allow to create single, well-defined photophysical states, which can be described well by the theory. Although this would increase orientation effects in the first place, the measurement of both FRET efficiency and anisotropy data in a FRET network should, in total, allow an improved localization of fluorophores, as shown in the test calculations.

Possible consequences of systematic effects like incorrect prior assumptions or outliers

in the data were discussed. Common mistakes in the interpretation of marginal position posterior densities like confusing the posterior maximum with the maximum of the marginal posterior were pointed out, which is especially important when the inference is based only on FRET efficiency data. Furthermore, it was shown how FRET networks can be optimized in order to increase the number of data beyond the number of unknown model parameters, while the number of labeling sites is kept small.

Finally, a way was proposed how simple FRET efficiency histogram fitting and advanced FRET data pre-processing methods like photon distribution analysis (PDA), proximity ratio histograms (PRH) or hidden Markov modeling (HMM) can be linked to the Nano-Positioning system to obtain evidence values, which can be used in Bayesian model selection.

To carry out the data analysis with the *position - orientation* NPS model, a custom Markov chain Monte Carlo - based implementation of nested sampling was written in Matlab and C (section 5.5). The performance of this software called *sampling engine* was demonstrated and discussed for a typical NPS calculation (subsections 6.3.3 and 6.3.4) and was found to be well suited for NPS inference problems.

Since the sampling engine was developed to be applicable to general Bayesian inference calculations with a continuous moderate-dimensional parameter space, it is currently used in two other projects that require a thorough data analysis, as described in the following.

Applications and improvements of the sampling engine

The sampling engine developed to solve Bayesian parameter estimation problems (section 5.5) is currently used to analyze data from two further projects realized by doctorate students in the lab.

In one project, Wolfgang Kügel uses the sampling engine to globally analyze fluorescence correlation spectroscopy (FCS, Magde et al., 1972; 1974) data of fluorescently labeled DNA hairpins obtained in a Pulsed Interleaved Excitation (PIE, Müller et al., 2005) scheme . This data analysis problem is challenging, since many of the 20 model parameters are highly correlated and a precise estimation of uncertainties is desired. Of special interest are the marginal densities of rates that model the dynamic transitions between the open and closed state of a DNA hairpin in this particular case. In general, this method is capable of inferring the conformational dynamics of biological macromolecules.

In the other project conducted by Michael Budde the sampling engine is used to analyze force spectroscopy (Janshoff et al., 2000) data, which consists of rupture times of polydimethylsiloxane (PDMS) polymers, which were acquired in atomic force microscope (AFM, Binnig et al., 1986) experiments at different constant holding forces. The objective is to characterize the rupture process of a single silicon-oxygen bond within the scope of transition state theory. In particular, the free energy barrier between the bound and ruptured state is quantified in terms of its height and its position on the reaction coordinate, as well as by the rupture rate at zero force. As the data was acquired at many different holding forces it is not possible to construct a histogram of rupture times at a single force that contains enough data points to satisfy the conditions of a least squares fit. Moreover, it is not clear whether all bonds in a polymer can break independently or whether only one bond at the end of the polymer ruptures due to different boundary conditions. Consequently, a model selection should be applied to this problem optimally achieved by Bayesian data analysis. In general, the Bayesian approach, which can be applied as well to dynamic force spectroscopy experiments, promises to develop into a useful tool for characterization of chemical bonds.

However, the calculations can take a long time on a multicore personal computer (sometimes several weeks for large NPS problems), and therefore the sampling engine has to be improved.

In order to maneuver efficiently through the parameter space one could use Hybrid Monte Carlo (HMC) that is also more specifically and descriptively called Hamiltonian Monte Carlo (Duane et al., 1987; Neal, 2010). This method utilizes Hamiltonian mechanics to sample efficiently the parameter space. However, this method is not compatible with nested sampling, and for a calculation of the evidence it will be necessary to use other algorithms instead, for example simulated annealing (Kirkpatrick et al., 1983; Neal, 1993). This could be combined with a Metropolis Coupled Markov chain Monte Carlo (MCMCM) scheme, also called replica exchange (Swendsen and Wang, 1986), which should enable even faster exploration of a multimodal posterior density as demonstrated by Rieping et al. (2005). One should think as well of parallelizing the calculations to a computer cluster or to graphical processing units (GPUs). In either case, it will be necessary to rewrite the software in a different programming language than Matlab that is both generally and financially more suitable for parallel computing.

Speeding up present *position - orientation* NPS model calculations with an improved sampling engine will certainly make NPS data analysis more comfortable. However, the improvements pointed out above will definitely be necessary to make calculations based on more complex future NPS models feasible at all. Some of these models as well as ways to improve NPS will be outlined in the next section.

Possible NPS improvements and extensions

As already pointed out in the discussion (chapter 6) there is plenty space for future improvements of the Nano-Positioning System. A few specific ideas will be described in the following.

The most promising improvement will be the additional use of FRET anisotropy data acquired in single-molecule experiments, since a considerable increase of localization accuracy is expected, as was shown in section 6.3.3. However, the inferred positions and average transition dipole moment orientations have to be tested on a macromolecular system with well known structure, in order to check whether the *position - orientation* NPS model is valid. Beside sufficiently large proteins of known structure, a possible test system could be built from nucleic acid scaffolds by using the method of DNA origami (Rothemund, 2006).

A minor improvement of the NPS framework can be achieved by implementing a likelihood function that is able to process correlated FRET efficiency and FRET anisotropy data. Such correlations can occur when a FRET efficiency histogram contains partially overlapping peaks and the positions of all FRET efficiency states have to be inferred at once. Another minor enhancement is the implementation of correct boundary conditions for the reference frame orientations in a docking application. Similarly, a better parametrization of the average transition dipole moment orientations could make the posterior easier to sample, since the currently used mapping of a sphere surface to a rectangle distorts the average transition dipole moment orientations almost parallel to the $z - axis$ severely. The current parametrization has the effect that the local move length scales in the polar angle can strongly depend on the azimuthal angle, which could be improved by a different choice of parameters optimized for a random walk on a sphere.

More important and still feasible will be the improvement of data pre-processing. The donor and acceptor fluorescence signals used to calculate the FRET efficiency data in this work were computed by subtracting signal and local background counts of an EM-CCD

camera. The signal and background counts were obtained by binning the readout of a couple of pixels on the camera that contained mainly signal and background, respectively. However, one could optimize this procedure and reduce the noise in the acquired signals by fitting the correct point spread function (Mortensen et al., 2010, supplementary material) to the diffraction limited spot of the fluorophore (Holden and Kapanidis, 2009).

Another major improvement would be the parameterization of the average axial depolarizations $\langle d_i^x \rangle$ and the isotropic Förster distances R_{ij}^{iso}. In the current NPS version both are assumed to be known with infinite accuracy, but the experimentally determined values are obviously limited in precision. While a log-normal distribution will probably be suitable as a prior for the isotropic Förster distance, it will be more difficult to assign a correct prior for the average axial depolarizations.

An important long-term objective should be the porting of the NPS software to a free programming language, since Matlab is proprietary. NPS could then be applied and developed by more researchers.

It is necessary but simultaneously also challenging to extend the theory of Dale, Eisinger, and Blumberg (1979), who assumed parallel absorption and emission dipole moments of the fluorophores. However, this assumption holds only for certain fluorophores, since in reality, is exists a finite angle between those dipole moments. Most fluorophores exhibit thus fundamental fluorescence anisotropies below 0.4 (Lakowicz, 2006, chapter 10), although the deviation from parallel transition dipole moments can be very small for certain fluorophores (Faucon and Lakowicz, 1987; Sanborn et al., 2007). It will be particularly important to account for this in order to model FRET between fluorophores with highly restricted orientational movement, since the expected FRET efficiency depends on the relative orientation of the donor emission and acceptor absorption dipole moments, while the expected FRET anisotropy is a function of the relative orientation of the donor absorption and acceptor emission dipole moments.

Another important but ambitious refinement of the NPS model will be the use of FRET histogram widths to model position and orientation fluctuations of the fluorophores and the macromolecule. A physically reasonable parametrization of these widths will be needed and should then be used in an improved NPS model, which could be similar to the model of Schröder and Grubmüller (2004), and moreover would account for constrained fluorophore orientations. Thus the simultaneous monitoring of fluorescence anisotropy (Schröder et al., 2005), FRET efficiency and FRET anisotropy might help to distinguish position fluctuations from orientation fluctuations.

Even more appreciable is the development of a "first principles" likelihood function that accounts for extrinsic effects caused by the measurement process, for example shot noise, and likewise for intrinsic fluorophore and macromolecule fluctuations, which can occur on different time scales (see section 6.7.1). The latter objective is already subject of current research based on an isotropic FRET model assuming $\kappa^2 = 2/3$ (Savol and Chennubhotla, 2010), while the measurement process has already been put on a firm footing in photon counting experiments (Antonik et al., 2006; Nir et al., 2006) and TIRF-based experiments using an EM-CCD camera (Mortensen et al., 2010). Finally, both approaches should be combined and approximated to a level treatable with current computational methods.

As the measurement of all possible pairs of fluorophore locations in a large FRET network is time-consuming, a shorter measurement time is desirable. To this end, an approach to speed up the acquisition of FRET data was developed recently (Uphoff et al., 2010). By using a single site-specifically attached donor fluorophore and several switchable acceptors located at arbitrary positions on the same macromolecule, it was possible to measure FRET efficiencies that belong to several donor-acceptor pairs. NPS could be adapted to such experiments by simultaneous probabilistic assignment of the observed

FRET populations to pairs of fluorophores.

In order to reduce the number of unknowns in the NPS model, it is desirable to use more detailed prior information. To this end, one could simulate possible positions and orientations of the average transition dipole moments of the fluorophores. A first step in that direction was realized recently in the lab: in the scope of his bachelor thesis Moritz Haag used a space filling representation of a protein and attached a fluorophore to compute possible orientations and positions of the transition dipole moment (Haag, 2009). In order to use such simulations, one has to identify possible average transition dipole moment orientations that are consistent with both the simulated transition dipole moment orientations and with the measured fluorescence anisotropies. This still needs to be done, which would allow to encode orientation information in the prior. In this way, the information on the partial dependence of the fluorophore positions and average transition dipole moment orientations could be used.

However, this approach would fail completely if no macromolecular structure is available at all, and it would lead to incorrect results when the structure is incorrect. Thus one could try to infer fluorophore positions based only on FRET measurements, which could already be done with the current *position - orientation* NPS model. Yet, a convenient marginalization and representation of the inferred results has still to be found as there is no reference like in the case when a high resolution structure is available. Furthermore, it will be extremely important to measure a large and consistent FRET network, and to acquire as much data as possible, for example by measuring FRET efficiencies and anisotropies.

The control of experimental conditions like the correct assembly of complexes is crucial when the subunits are interacting weakly or only transiently. Alternating laser excitation (ALEX) (Lee et al., 2005) is a common way to control the presence of the acceptor in a FRET pair by its direct excitation. This technique has been used in some cases to acquire the data analyzed in this thesis. However, the ALEX excitation scheme might not be sufficient to control experiments similar to the RNA 5'-end localization presented here, as the presence of an additional factor (here TFIIB) might cause a very small and therefore unnoticed FRET change. In this case, the labeling of the additional factor with a third fluorophore emitting in a different spectral range without disturbing the donor and acceptor would increase the experimental control.

Also FRET experiments utilizing three or more fluorophores simultaneously (Hohng et al., 2004) can potentially increase experimental control and also yield data that contains valuable information on correlated conformational motions. However, in the case of a stable static conformation, the analysis of such experiments would not lead to an information gain compared to conventional FRET performed for each FRET pair separately.

A challenging but also very profitable endeavor would be the construction of a hybrid data analysis approach that combines NPS results with those obtained by low-resolution methods for structure determination. Reasonable candidates are for example chemical cross-linking experiments, small-angle X-ray scattering (SAXS) and pulsed dipolar electron spin resonance (ESR), which have all been applied to study the structure of biological macromolecules. While the proximity of residues can be determined by chemical cross-linking (Sinz, 2006) SAXS can be used to infer the approximate shape of macromolecules in solution (Koch et al., 2003) and ESR provides, similar to FRET, long range distance restraints (Bhatnagar et al., 2007).

A good approach would be to use already present modeling software, for example the Integrated Modeling Platform (IMP) developed in the lab of Andrej Sali (Alber et al., 2007a,b, 2008). This software package can analyze commonly applied biochemical and structural data obtained by various methods, and it should be possible to implement

marginal NPS results as a constraint in IMP (Lasker and Sali, 2010).

Furthermore, the prediction of protein and protein complex structures from their amino-acid sequence could profit from results obtained by NPS. This approach has been already documented with mass spectrometry and SAXS data (D'Abramo et al., 2009), and possibly it could be extended to use FRET-derived constrains as well.

Finally, the author hopes that this thesis could convince the reader of the usefulness of modern FRET-based localization methods as well as of the elegance of probabilistic data analysis.

List of abbreviations and variables

abbreviation	explanation
A, $A(t)$	measured FRET anisotropy
A_{ai}, A_{ij}	FRET anisotropy *measured* between fluorophores a/i, and i/j, resp.
$\mathcal{A}_{ai}$, $\mathcal{A}_{ij}$	FRET anisotropy *expected* between fluorophores a/i, and i/j, resp.
ALX	Alexa Fluor®
ANT	antenna fluorophore
BCP	box collection prior
CY3	cyanine 3
d^x, d_i^x	axial depolarization, axial depolarization of fluorophore i
$\langle d_i^x \rangle$	average axial depolarization, axial depolarization of fluorophore i
DNA	deoxyribonucleic acid
E, $E(t)$	measured FRET efficiency
E_{ai}, E_{ij}	FRET efficiency *measured* between fluorophores a/i, and i/j, resp.
$\mathcal{E}_{ai}$, $\mathcal{E}_{ij}$	FRET efficiency *expected* between fluorophores a/i, and i/j, resp.
$F_{\mathrm{D}}(t)$, $F_{\mathrm{A}}(t)$	fluorescence signal of donor (D) and acceptor (A) fluorophores
FRET	Fluorescence Resonance Energy Transfer
GUI	graphical user interface
κ^2	(orientation factor)^2
MCMC	Markov-Chain Monte Carlo
nt	nucleotide
Pol II	eukaryotic RNA polymerase II
r_i	steady-state fluorescence anisotropy of fluorophore i
$r_{\infty,i}$	residual fluorescence anisotropy of fluorophore i
R	Förster distance
R_{ai}, R_{ij}	Förster distance between the fluorophores a/i, and i/j, respectively
R^{iso}, R_{ai}^{iso}, R_{ij}^{iso}	isotropic Förster distance (assuming $\kappa^2 = 2/3$)
RNA	ribonucleic acid
Rpb4/7	heterodimer of the Pol II subunits Rpb4 and Rpb7
SAT	satellite fluorophore
TBP	TATA binding protein
TFII[X]	general transcription factor [X] for RNA polymerase II
TMR	tetramethylrhodamine
$\boldsymbol{x}$	arbitrary spatial position (e.g. of a fluorophore)
$\boldsymbol{x}_i$	spatial position of the ith fluorophore
$\boldsymbol{\Omega}_i$	average transition dipole moment orientation of the ith fluorophore

Appendix I: Coupling of data pre-processing to NPS

It will be shown here that the likelihood in equation (6.7.1) can be obtained from the result of commonly used data pre-processing procedures.

The "raw" data $\{O_{ij}\}(t)$, i.e. time traces of the observables like FRET efficiency and anisotropy acquired for the FRET pairs ij, will be described with a hierarchical model M that consists of a low-level part, which pre-processes the original data, and a high-level NPS part that uses the result of pre-processing as data. The model M assumes that the macromolecule is in one of K_M possible conformational states.

The pre-processing part of the model describes the time traces by a set of "true" values $\{\tilde{O}_{ij;k}\}$ of the observables in the state $k = 1 \ldots K_M$, and by some auxiliary parameters $\boldsymbol{h}_M$. The latter are used to account for properties of the data that are not directly related to FRET, for example the dynamics observed between different states. In the NPS part, the parameters $q_{i;k}^{\mathrm{NPS}} = (\boldsymbol{x}_{i;k}, \boldsymbol{\Omega}_{i;k})$ are used to model structural properties of the FRET network, i.e. the positions and average transition dipole moment orientations, as well as the signs of the average axial depolarizations (see section 6.7.5 for a definition).

The likelihood in equation (6.7.1) is the probability of time traces given the set of NPS parameters, $\{q_{i;k}^{\mathrm{NPS}}\}$, and can be written as

$$
p\left(\{O_{ij}(t)\}|\{q_{i;k}^{\mathrm{NPS}}\}, M, I\right) = \int d\{\tilde{O}_{ij;k}\}\; p\left(\{O_{ij}(t)\}, \{\tilde{O}_{ij;k}\}|\{q_{i;k}^{\mathrm{NPS}}\}, M, I\right) =
$$

$$
\int d\{\tilde{O}_{ij;k}\}\; p\left(\{O_{ij}(t)\}|\{\tilde{O}_{ij;k}\}, \{q_{i;k}^{\mathrm{NPS}}\}, M, I\right)\; p\left(\{\tilde{O}_{ij;k}\}|\{q_{i;k}^{\mathrm{NPS}}\}, M, I\right) =
$$

$$
\int d\{\tilde{O}_{ij;k}\}\; p\left(\{O_{ij}(t)\}|\{\tilde{O}_{ij;k}\}, M, I\right)\; \boldsymbol{\delta}(\{\tilde{O}_{ij;k} - \mathcal{O}_{ij;k}\}),
$$

In the first line, the "true" observable values were introduced, and Bayes' theorem was applied to move them into the condition. Since the time traces do not depend directly on the NPS parameters, they can be canceled from the condition in the second line. In the third line, the term $p\left(\{\tilde{O}_{ij;k}\}|\{q_{i;k}^{\mathrm{NPS}}\}, M, I\right)$ is identified with a product of Dirac-delta distributions,

$$
\boldsymbol{\delta}(\{\tilde{O}_{ij;k} - \mathcal{O}_{ij;k}\}) := \prod_{ij;k} \delta\left(\tilde{O}_{ij;k} - \mathcal{O}_{ij;k}\left(q_{i;k}^{\mathrm{NPS}}, q_{j;k}^{\mathrm{NPS}}\right)\right).
$$

This is obvious, since the "true" observable values $\{\tilde{O}_{ij;k}\}$ and the observables $\mathcal{O}_{ij;k}$ (i.e. $\mathcal{E}_{ij;k}$ or/and $\mathcal{A}_{ij;k}$) expected from the NPS parameters must be the same.

The auxiliary parameters $\boldsymbol{h}_M$ are introduced in the same way like the "true" observables,

$$
p\left(\{O_{ij}(t)\}|\{q_{i;k}^{\mathrm{NPS}}\}, M, I\right) =
$$

$$
\int d\{\tilde{O}_{ij;k}\} \int d\boldsymbol{h}_M\; p\left(\{O_{ij}(t)\}|\boldsymbol{h}_M, \{\tilde{O}_{ij;k}\}, M, I\right) p\left(\boldsymbol{h}_M|M, I\right) \boldsymbol{\delta}(\{\tilde{O}_{ij;k} - \mathcal{O}_{ij;k}\}).
$$

Here, it was assumed that the prior of the auxiliary parameters does not depend on the "true" observables, i.e. that $p\left(\boldsymbol{h}_M|M, I\right) = p\left(\boldsymbol{h}_M|\{\tilde{O}_{ij;k}\}, M, I\right)$.

The probability $p\left(\{O_{ij}(t)\}|\boldsymbol{h}_M, \{\tilde{O}_{ij;k}\}, M, I\right)$ is essentially the likelihood of the

data pre-processing, which is related to the posterior of the pre-processing, $p\left(\boldsymbol{h}_M, \{\tilde{O}_{ij;k}\}|\{O_{ij}(t)\}, M, I\right)$, by Bayes' theorem,

$$Z_M^{\mathrm{PP}} \, p\left(\boldsymbol{h}_M, \{\tilde{O}_{ij;k}\}|\{O_{ij}(t)\}, M, I\right) =$$
$$p\left(\tilde{O}_{ij;k}|M, I\right) \, p\left(\boldsymbol{h}_M|M, I\right) \, p\left(\{O_{ij}(t)\}|\boldsymbol{h}_M, \{\tilde{O}_{ij;k}\}, M, I\right).$$

Here, Z_M^{PP} denotes the evidence of the pre-processing, and $p\left(\tilde{O}_{ij;k}|M, I\right)$ is the prior of the "true" observables. From the two above equations, one obtains

$$p\left(\{O_{ij}(t)\}|\{q_{i;k}^{\mathrm{NPS}}\}, M, I\right) =$$
$$\int d\{\tilde{O}_{ij;k}\} \int d\boldsymbol{h}_M \, \frac{Z_M^{\mathrm{PP}}}{p\left(\tilde{O}_{ij;k}|M, I\right)} p\left(\boldsymbol{h}_M, \{\tilde{O}_{ij;k}\}|\{O_{ij}(t)\}, M, I\right) \boldsymbol{\delta}\left(\{\tilde{O}_{ij;k} - \mathcal{O}_{ij;k}\}\right) =$$
$$Z_M^{\mathrm{PP}} \int d\{\tilde{O}_{ij;k}\} \, \frac{1}{p\left(\tilde{O}_{ij;k}|M, I\right)} \, p\left(\{\tilde{O}_{ij;k}\}|\{O_{ij}(t)\}, M, I\right) \boldsymbol{\delta}\left(\{\tilde{O}_{ij;k} - \mathcal{O}_{ij;k}\}\right).$$

When the data pre-processing uses a constant prior in the "true" observables,

$$const = p\left(\tilde{O}_{ij;k}|M, I_{\mathrm{PP}}\right) =: \pi_M^{\mathrm{PP}},$$

which applies in particular to maximum likelihood approaches and least squares fits (Sivia, 2006, chapter 3.5). The prior can be taken out of the integral and the following equation results:

$$p\left(\{O_{ij}(t)\}|\{q_{i;k}^{\mathrm{NPS}}\}, M, I\right) =$$
$$\frac{Z_M^{\mathrm{PP}}}{\pi_M^{\mathrm{PP}}} \int d\{\tilde{O}_{ij;k}\} p\left(\{\tilde{O}_{ij;k}\}|\{O_{ij}(t)\}, M, I\right) \boldsymbol{\delta}\left(\{\tilde{O}_{ij;k} - \mathcal{O}_{ij;k}\}\right) =$$
$$\frac{Z_M^{\mathrm{PP}}}{\pi_M^{\mathrm{PP}}} \, p\left(\{\tilde{O}_{ij;k}\}|\{O_{ij}(t)\}, M, I\right)\Big|_{\tilde{O}_{ij;k} = \mathcal{O}_{ij;k}\left(q_{i;k}^{\mathrm{NPS}}, q_{j;k}^{\mathrm{NPS}}\right)}.$$

One easily recognizes that the likelihood of the NPS problem is obtained by weighting the posterior of the pre-processing problem with $Z_M^{\mathrm{PP}}/\pi_M^{\mathrm{PP}}$ and substituting the observables expected from NPS, $\mathcal{O}_{ij;k}$, for the "true" observables $\tilde{O}_{ij;k}$.

Appendix II: Data

1 Measured data

$E_{ai} \pm \Delta E_{ai}$	RNA(+1) -ALX555	RNA(+1) -TMR	RNA(+29) -TMR, w/o TFIIB	RNA(+29) -TMR, w/ TFIIB
tDNA(-10)-ALX647	0.507 ± 0.0035	–	0.673 ± 0.006	0.382 ± 0.008
tDNA(+3)-ALX647	0.800 ± 0.0014	–	0.422 ± 0.007	0.630 ± 0.027
tDNA(+9)-ALX647	0.610 ± 0.0040	–	–	–
Rpb7(C150)-ALX647	–	0.415 ± 0.0039	0.604 ± 0.006	0.770 ± 0.010
Rpb7(C94)-ALX647	–	–	0.522 ± 0.011	0.712 ± 0.043
Rpb4(S73C)-ALX647	–	–	$0.529 \pm 0.021/$ $0.752 \pm 0.029^*$	–
Rpb7(S16C)-ALX647	39.4 ± 0.38	–	0.736 ± 0.007	0.936 ± 0.041

*) two peaks were observed and attributed to different positions of the satellite
Rpb4(S73C)-ALX647.

Table 1: Measured FRET efficiencies (RNA, taken from (Muschielok et al., 2008))

$R_{ai}^{\mathrm{iso}}/\text{Å}$	RNA(+1) -ALX555	RNA(+1) -TMR	RNA(+29) -TMR, w/o TFIIB	RNA(+29) -TMR, w/ TFIIB
tDNA(-10)-ALX647	47	–	62	62
tDNA(+3)-ALX647	47	–	62	62
tDNA(+9)-ALX647	47	–	–	–
Rpb7(C150)-ALX647	–	57	65	65
Rpb7(C94)-ALX647	–	–	66	66
Rpb4(S73C)-ALX647	–	–	63	–
Rpb7(S16C)-ALX647	48	–	62	62

Table 2: Measured isotropic Förster distances (RNA, taken from (Muschielok et al., 2008))

Appendix II: Data

attachment site / fluorophore	fluoresc. anisotropy[*]
tDNA(-10)-ALX647	$0.30^{(1,2)}/0.28^{(3)}$
tDNA(-10)-TMR	$0.31^{(3)}$
tDNA(+3)-ALX647	$0.32^{(1,2)}/0.27^{(3)}$
tDNA(+3)-TMR	$0.30^{(3)}$
tDNA(+7)-ALX647	$0.26^{(3)}$
tDNA(+7)-TMR	$0.28^{(3)}$
tDNA(+9)-ALX647	$0.31^{(1,2)}$
tDNA(+12)-ALX647	$0.19^{(3)}$
RNA(+1)-ALX647	$0.15^{(2)}$
RNA(+1)-ALX555	$0.31^{(1)}$
RNA(+1)-TMR	$0.21^{(1)}$
RNA(+4)-ALX647	$0.15^{(2)}$
RNA(+10)-ALX647	$0.20^{(2)}$
Rpb7(C150)-ALX647	$0.23^{(1,2,3)}$
Rpb7(C94)-ALX647	$0.22^{(1)}$
Rpb4(S73C)-ALX647	$0.27^{(1,2)}$
Rpb7(S16C)-ALX647	$0.20^{(1)}$
RNA(+29)-TMR w/o TFIIB	$0.21^{(1)}$
RNA(+29)-TMR w/ TFIIB	$0.21^{(1)}$
ntDNA(+1)-TMR	$0.19^{(2)}$
ntDNA(+1)-ALX555	$0.29^{(2)}$
ntDNA(-2)-TMR	$0.19^{(2)}$
ntDNA(-2)-ALX555	$0.31^{(2)}$
ntDNA(-4)-TMR	$0.21^{(2)}$
ntDNA(-4)-ALX555	$0.30^{(2)}$
ntDNA(-7)-TMR	$0.18^{(2)}$
ntDNA(-7)-ALX555	$0.29^{(2)}$
ntDNA(-12)-TMR	$0.18^{(2)}$
ntDNA(-12)-ALX555	$0.27^{(2)}$
ntDNA(-15)-TMR	$0.16^{(2)}$
ntDNA(-15)-ALX555	$0.31^{(2)}$
ntDNA(-18)-TMR	$0.17^{(2)}$
ntDNA(-18)-ALX555	$0.27^{(2)}$
ntDNA(-20)-TMR	$0.24^{(3)}$
ntDNA(-30)-TMR	$0.29^{(3)}$
ntDNA(-37)-TMR	$0.26^{(3)}$
TBP(S159C)-ALX647	$0.14^{(3)}$
TBP(S159C)-CY3	$0.22^{(3)}$

(1) used in the RNA/TFIIB project (Muschielok et al., 2008)
(2) used in the nontemplate DNA project (Andrecka et al., 2009)
(3) used in the initial transcribing complex project (Treutlein et al., unpublished)
(∗) steady state anisotropy

Table 3: Measured fluorescence anisotropies

E_{ai}	tDNA (-10) -ALX647	tDNA (+3) -ALX647	tDNA (+9) -ALX647	RNA (+1) -ALX647	RNA (+4) -ALX647	RNA (+10) -ALX647	Rpb7 (C150) -ALX647	Rpb4 (S73C) -ALX647
ntDNA(+1)-TMR	–	–	–	–	–	–	0.32	–
ntDNA(+1)-ALX555	0.86	–	0.80	0.80	0.96	0.73	–	–
ntDNA(-2)-TMR	–	–	–	–	–	–	0.35	–
ntDNA(-2)-ALX555	0.84	0.91	0.75	0.64	0.90	0.76	–	–
ntDNA(-4)-TMR	–	–	–	–	–	–	0.36	–
ntDNA(-4)-ALX555	0.78	0.88	0.61	0.46	0.81	0.57	–	–
ntDNA(-7)-TMR	–	–	–	–	–	–	0.54	–
ntDNA(-7)-ALX555	0.80	0.83	0.57	0.40	0.78	0.60	–	–
ntDNA(-12)-TMR	–	0.90	–	–	–	–	0.57	0.32
ntDNA(-12)-ALX555	–	0.53	0.46	0.45	0.86	–	–	–
ntDNA(-15)-TMR	–	0.68	–	–	–	–	0.48	0.29
ntDNA(-15)-ALX555	0.89	0.38	0.30	0.33	0.63	–	–	–
ntDNA(-18)-TMR	–	0.40^*/ 0.60	0.37^*/ 0.71	–	–	–	0.31^*/ 0.39	–
ntDNA(-18)-ALX555	0.68^*/ 0.35	–	–	0.27^*/ 0.56	0.51^*/ 0.35	0.63^*/ 0.31	–	–

Andrecka et al. (2009) used $\Delta E_{ij} = 0.02$ for all FRET pairs with $E_{ij} < 0.9$, and $\Delta E_{ij} = 0.04$ for all FRET pairs with $E_{ij} \geq 0.9$.
In the comparison of *position - Förster distance* model and *position - orientation* model, $\Delta E_{ij} = 0.02$ was used for all FRET efficiency measurements.
∗) main peak used in the inference

Table 4: Measured FRET efficiencies (nontemplate DNA, taken from (Andrecka et al., 2009))

$R_{ai}^{\mathrm{iso}}/\text{Å}$	tDNA (-10) -ALX647	tDNA (+3) -ALX647	tDNA (+9) -ALX647	RNA (+1) -ALX647	RNA (+4) -ALX647	RNA (+10) -ALX647	Rpb7 (C150) -ALX647	Rpb4 (S73C) -ALX647
ntDNA(+1)-TMR	–	–	–	–	–	–	64	–
ntDNA(+1)-ALX555	54	–	54	54	54	53	–	–
ntDNA(-2)-TMR	–	–	–	–	–	–	62	–
ntDNA(-2)-ALX555	50	50	50	50	50	49	–	–
ntDNA(-4)-TMR	–	–	–	–	–	–	64	–
ntDNA(-4)-ALX555	49	49	49	49	49	48	–	–
ntDNA(-7)-TMR	–	–	–	–	–	–	64	–
ntDNA(-7)-ALX555	46	46	46	46	46	45	–	–
ntDNA(-12)-TMR	–	59	–	–	–	–	62	64
ntDNA(-12)-ALX555	–	49	49	49	49	–	–	–
ntDNA(-15)-TMR	–	60	–	–	–	–	63	64
ntDNA(-15)-ALX555	49	49	49	49	49	–	–	–
ntDNA(-18)-TMR	–	58	58	–	–	–	62	–
ntDNA(-18)-ALX555	48	–	–	48	48	48	–	–

Table 5: Measured isotropic Förster distances (nontemplate DNA, taken from (Andrecka et al., 2009))

E_{ij}	ntDNA (-20) -TMR	ntDNA (-30) -TMR	ntDNA (-37) -TMR	TBP (C61) -CY3	TBP (C61) -ALX647
tDNA(-10)-TMR	–	–	–	–	0.18
tDNA(-10)-ALX647	0.84	0.37	0.24	–	–
tDNA(+3)-TMR	–	–	–	–	0.20
tDNA(+3)-ALX647	0.58	0.32	0.11	–	–
tDNA(+7)-TMR	–	–	–	–	0.08
tDNA(+7)-ALX647	0.58	0.26	0.11	–	–
tDNA(+12)-ALX647	0.36	0.22	–	–	–
Rpb7(C150)-ALX647	0.34	0.34	0.12	0.018	–
TBP(S159C)-ALX647	–	0.71	0.77	–	undef.

$\Delta E_{ij} = 0.02$ for all FRET pairs.

Table 6: Measured FRET efficiencies (initial transcribing complex, taken from (Treutlein et al., unpublished))

R_{ij}^{iso}/Å	ntDNA (-20) -TMR	ntDNA (-30) -TMR	ntDNA (-37) -TMR	TBP (C61) -CY3	TBP (C61) -ALX647
tDNA(-10)-TMR	–	–	–	–	58
tDNA(-10)-ALX647	56	60	59	–	–
tDNA(+3)-TMR	–	–	–	–	60
tDNA(+3)-ALX647	56	60	59	–	–
tDNA(+7)-TMR	–	–	–	–	60
tDNA(+7)-ALX647	57	61	60	–	–
tDNA(+12)-ALX647	57	61	–	–	–
Rpb7(C150)-ALX647	60	65	63	47	–
TBP(S159C)-ALX647	–	60	57	–	undef.

Table 7: Measured isotropic Förster distances (initial transcribing complex, taken from (Treutlein et al., unpublished))

2 Artificial data

fluorophore i	$x_i/\text{Å}$	$y_i/\text{Å}$	$z_i/\text{Å}$	θ_i/rad	ϕ_i/rad	r_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

Table 8: Fluorophore data

Exact fluorophore positions $\boldsymbol{x}_i = (x_i, y_i, z_i)$ and average transition dipole moment orientations $\boldsymbol{\Omega}_i = (-\cos\theta_i, \phi_i)$ together with fluorescence anisotropy r_i.

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

$\Delta E_{ij} = 0.02$ for all FRET pairs.

Table 9: Simulated FRET efficiencies

$R_{ij}^{\mathrm{iso}}/\text{Å}$	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

Table 10: Simulated isotropic Förster distances

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

$\Delta A_{ij} = 0.01$ for all FRET pairs.

Table 11: Simulated FRET anisotropies

Bibliography

UCSF Chimera software package. http://www.cgl.ucsf.edu/chimera, 2010. URL http://www.cgl.ucsf.edu/chimera.

Frank Alber, Svetlana Dokudovskaya, Liesbeth M Veenhoff, Wenzhu Zhang, Julia Kipper, Damien Devos, Adisetyantari Suprapto, Orit Karni-Schmidt, Rosemary Williams, Brian T Chait, Michael P Rout, and Andrej Sali. Determining the architectures of macromolecular assemblies. *Nature*, 450(7170):683–694, Nov 2007a. doi: 10.1038/nature06404. URL http://dx.doi.org/10.1038/nature06404.

Frank Alber, Svetlana Dokudovskaya, Liesbeth M Veenhoff, Wenzhu Zhang, Julia Kipper, Damien Devos, Adisetyantari Suprapto, Orit Karni-Schmidt, Rosemary Williams, Brian T Chait, Andrej Sali, and Michael P Rout. The molecular architecture of the nuclear pore complex. *Nature*, 450(7170):695–701, Nov 2007b. doi: 10.1038/nature06405. URL http://dx.doi.org/10.1038/nature06405.

Frank Alber, Friedrich Förster, Dmitry Korkin, Maya Topf, and Andrej Sali. Integrating diverse data for structure determination of macromolecular assemblies. *Annual Review Of Biochemistry*, 77:443–477, 2008. doi: 10.1146/annurev.biochem.77.060407.135530. URL http://dx.doi.org/10.1146/annurev.biochem.77.060407.135530.

Bruce Alberts, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, and Peter Walter. *Molecular Biology of the Cell*. Garland Science, 5th edition, 2008.

Joanna Andrecka. *Single molecule fluorescence studies of the RNA polymerase II elongation complex*. PhD thesis, Ludwig-Maximilians-Universität München, 2009.

Joanna Andrecka, Robert Lewis, Florian Brückner, Elisabeth Lehmann, Patrick Cramer, and Jens Michaelis. Single-molecule tracking of mRNA exiting from RNA polymerase II. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 105(1):135–140, 2008. ISSN 1091-6490. doi: 10.1073/pnas.0703815105. URL http://www.ncbi.nlm.nih.gov/pubmed/18162559. PMID: 18162559.

Joanna Andrecka, Barbara Treutlein, Maria Angeles Izquierdo Arcusa, Adam Muschielok, Robert Lewis, Alan C M Cheung, Patrick Cramer, and Jens Michaelis. Nano positioning system reveals the course of upstream and nontemplate DNA within the RNA polymerase II elongation complex. *Nucleic Acids Research*, 37(17):5803–5809, September 2009. ISSN 1362-4962. doi: 10.1093/nar/gkp601. URL http://www.ncbi.nlm.nih.gov/pubmed/19620213. PMID: 19620213.

Matthew Antonik, Suren Felekyan, Alexander Gaiduk, and Claus A M Seidel. Separating structural heterogeneities from stochastic variations in fluorescence resonance energy transfer distributions via photon distribution analysis. *Journal Of Physical Chemistry B*, 110(13):6970–6978, Apr 2006. doi: 10.1021/jp057257+. URL http://dx.doi.org/10.1021/jp057257+.

Karim-Jean Armache, Hubert Kettenberger, and Patrick Cramer. Architecture of initiation-competent 12-subunit RNA polymerase II. *Proceedings Of The National*

Academy Of Sciences Of The United States Of America, 100(12):6964–6968, Jun 2003. doi: 10.1073/pnas.1030608100. URL http://dx.doi.org/10.1073/pnas.1030608100.

Sarah M Auclair, Julia P Moses, Monika Musial-Siwek, Debra A Kendall, Donald B Oliver, and Ishita Mukerji. Mapping of the signal peptide-binding domain of Escherichia coli SecA using Förster resonance energy transfer. *Biochemistry*, 49(4):782–792, Feb 2010. doi: 10.1021/bi901446r. URL http://dx.doi.org/10.1021/bi901446r.

Thomas Bayes. An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London*, 53:330–418, 1763.

Kirstine Berg-Sørensen and Henrik Flyvbjerg. Power spectrum analysis for optical tweezers. *Review Of Scientific Instruments*, 75(3):594–612, 2004.

Jaya Bhatnagar, Jack H Freed, and Brian R Crane. Rigid body refinement of protein complexes with long-range distance restraints from pulsed dipolar ESR. *Methods Enzymol*, 423:117–133, 2007. doi: 10.1016/S0076-6879(07)23004-6. URL http://dx.doi.org/10.1016/S0076-6879(07)23004-6.

G. Binnig, C. F. Quate, and Ch. Gerber. Atomic force microscope. *Phys. Rev. Lett.*, 56(9):930–933, March 1986. URL http://link.aps.org/doi/10.1103/PhysRevLett.56.930.

Christopher M. Bishop. *Pattern recognition and machine learning*. Springer, 2006.

Jonathan E Bronson, Jingyi Fei, Jake M Hofman, Ruben L Gonzalez Jr., and Chris H Wiggins. Learning rates and states from biophysical time series: A Bayesian approach to model selection and single-molecule FRET data. *Biophysical Journal*, 97(12):3196–3205, 2009. ISSN 00063495. doi: 10.1016/j.bpj.2009.09.031. URL http://www.cell.com/biophysj/abstract/S0006-3495(09)01513-6.

Ilja N Bronstein, Konstantin A Semendjajew, Gerhard Musiol, and Heiner Mühlig. *Taschenbuch der Mathematik*. Harri Deutsch, 5th revised and extended edition, 2000.

Florian Brückner and Patrick Cramer. Structural basis of transcription inhibition by alpha-amanitin and implications for RNA polymerase II translocation. *Nature Structural and Molecular Biology*, 15(8):811–818, Aug 2008. doi: 10.1038/nsmb.1458. URL http://dx.doi.org/10.1038/nsmb.1458.

David A Bushnell, Kenneth D Westover, Ralph E Davis, and Roger D Kornberg. Structural basis of transcription: an RNA polymerase II-TFIIB cocrystal at 4.5 Angstroms. *Science*, 303(5660):983–988, Feb 2004. doi: 10.1126/science.1090838. URL http://dx.doi.org/10.1126/science.1090838.

A J Carpousis and J D Gralla. Cycling of ribonucleic acid polymerase to produce oligonucleotides during initiation in vitro at the lac UV5 promoter. *Biochemistry*, 19(14):3245–3253, Jul 1980.

Chin-Yu Chen, Chia-Chi Chang, Chi-Fu Yen, Michael T.-K. Chiu, and Wei-Hau Chang. Mapping RNA exit channel on transcribing RNA polymerase II by FRET analysis. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 106(1):127–132, 2009. doi: 10.1073/pnas.0811689106. URL http://www.pnas.org/content/106/1/127.abstract.

Dmitry I Cherny, Ian C Eperon, and Clive R Bagshaw. Probing complexes with single fluorophores: factors contributing to dispersion of FRET in DNA/RNA duplexes. *European Biophysics Journal*, 38(4):395–405, Apr 2009. doi: 10.1007/s00249-008-0383-z. URL http://dx.doi.org/10.1007/s00249-008-0383-z.

Ucheor B Choi, Pavel Strop, Marija Vrljic, Steven Chu, Axel T Brunger, and Keith R Weninger. Single-molecule FRET-derived model of the synaptotagmin 1-SNARE fusion complex. *Nature Structural and Molecular Biology*, 17(3):318–324, Mar 2010. doi: 10.1038/nsmb.1763. URL http://dx.doi.org/10.1038/nsmb.1763.

Hoi Sung Chung, John M Louis, and William A Eaton. Distinguishing between protein dynamics and dye photophysics in single-molecule FRET experiments. *Biophysical Journal*, 98(4):696–706, Feb 2010. doi: 10.1016/j.bpj.2009.12.4322. URL http://dx.doi.org/10.1016/j.bpj.2009.12.4322.

L. Stirling Churchman, Henrik Flyvbjerg, and James A Spudich. A non-Gaussian distribution quantifies distances measured with fluorescence localization techniques. *Biophysical Journal*, 90(2):668–671, Jan 2006. doi: 10.1529/biophysj.105.065599. URL http://dx.doi.org/10.1529/biophysj.105.065599.

Robert M Clegg. FRET tells us about proximities, distances, orientations and dynamic properties. *Journal Of Biotechnology*, 82(3):177–179, Jan 2002.

P. Cramer, D. A. Bushnell, and R. D. Kornberg. Structural basis of transcription: RNA polymerase II at 2.8 Angstrom resolution. *Science*, 292(5523):1863–1876, Jun 2001. doi: 10.1126/science.1059493. URL http://dx.doi.org/10.1126/science.1059493.

P. Cramer, K-J. Armache, S. Baumli, S. Benkert, F. Brückner, C. Buchen, G. E. Damsma, S. Dengl, S. R. Geiger, A. J. Jasiak, A. Jawhari, S. Jennebach, T. Kamenski, H. Kettenberger, C-D. Kuhn, E. Lehmann, K. Leike, J. F. Sydow, and A. Vannini. Structure of eukaryotic RNA polymerases. *Annual Review of Biophysics*, 37:337–352, 2008. doi: 10.1146/annurev.biophys.37.032807.130008. URL http://dx.doi.org/10.1146/annurev.biophys.37.032807.130008.

Patrick Cramer. Finding the right spot to start transcription. *Nature Structural and Molecular Biology*, 14(8):686–687, Aug 2007. doi: 10.1038/nsmb0807-686. URL http://dx.doi.org/10.1038/nsmb0807-686.

Francis Crick. Central dogma of molecular biology. *Nature*, 227(5258):561–563, August 1970. URL http://dx.doi.org/10.1038/227561a0.

Francis H. Crick. The biological replication of macromolecules. In *Symp. Soc. Exp. Biol.*, volume XII, 1958.

R. A. Crowther, R. Henderson, and J. M. Smith. MRC image processing programs. *Journal Of Structural Biology*, 116(1):9–16, 1996. doi: 10.1006/jsbi.1996.0003. URL http://dx.doi.org/10.1006/jsbi.1996.0003.

Marco D'Abramo, Tim Meyer, Pau Bernado, Carles Pons, Juan Fernandez Recio, and Modesto Orozco. On the use of low-resolution data to improve structure prediction of proteins and protein complexes. *Journal Of Chemical Theory And Computation*, 5(11): 3129–3137, NOV 2009. ISSN 1549-9618. doi: 10.1021/ct900305m.

Bibliography

R.E. Dale, J. Eisinger, and W.E. Blumberg. The orientational freedom of molecular probes. The orientation factor in intramolecular energy transfer. *Biophysical Journal*, 26:161–193, 1979. with errata.

Robert E Dale and Josef Eisinger. Intramolecular distances determined by energy transfer. Dependence on orientational freedom of donor and acceptor. *Biopolymers*, 13:1573–1605, 1974.

S. A. Darst, A. M. Edwards, E. W. Kubalek, and R. D. Kornberg. Three-dimensional structure of yeast RNA polymerase II at 16 A resolution. *Cell*, 66(1):121–128, Jul 1991.

Volker Dose. Bayes in five days. Lecture notes from a ten hour tutorial on Bayesian analysis given at the international Max-Planck research school on bounded plasmas, Greifswald. Technical report, Max-Planck-Institut für Plasmaphysik, 2002.

Volker Dose. Die Bayes'sche Variante. *Physik Journal*, 4:67–72, 2005.

J. Drenth. *Principles of Protein X-Ray Crystallography*. Springer, New York, 1999.

H. R. Drew, R. M. Wing, T. Takano, C. Broka, S. Tanaka, K. Itakura, and R. E. Dickerson. Structure of a B-DNA dodecamer: conformation and dynamics. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 78(4):2179–2183, Apr 1981.

Simon Duane, A.D. Kennedy, Brian J. Pendleton, and Duncan Roweth. Hybrid Monte Carlo. *Physics Letters B*, 195:216–222, 1987.

C. Eggeling, J. R. Fries, L. Brand, R. Günther, and C. A. Seidel. Monitoring conformational dynamics of a single molecule by selective fluorescence spectroscopy. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 95(4):1556–1561, Feb 1998.

C. Eggeling, S. Berger, L. Brand, J. R. Fries, J. Schaffer, A. Volkmer, and C. A. Seidel. Data registration and selective single-molecule analysis using multi-parameter fluorescence detection. *Journal Of Biotechnology*, 86(3):163–180, Apr 2001.

Jesse Eichner, Hung-Ta Chen, Linda Warfield, and Steven Hahn. Position of the general transcription factor tfiif within the rna polymerase ii transcription preinitiation complex. *EMBO J*, 29(4):706–716, Feb 2010. doi: 10.1038/emboj.2009.386. URL http://dx.doi.org/10.1038/emboj.2009.386.

Dmitri N Ermolenko, Zigurts K Majumdar, Robyn P Hickerson, P. Clint Spiegel, Robert M Clegg, and Harry F Noller. Observation of intersubunit movement of the ribosome in solution using FRET. *Journal Of Molecular Biology*, 370(3):530–540, Jul 2007a. doi: 10.1016/j.jmb.2007.04.042. URL http://dx.doi.org/10.1016/j.jmb.2007.04.042.

Dmitri N Ermolenko, P. Clint Spiegel, Zigurts K Majumdar, Robyn P Hickerson, Robert M Clegg, and Harry F Noller. The antibiotic viomycin traps the ribosome in an intermediate state of translocation. *Nature Structural and Molecular Biology*, 14(6):493–497, Jun 2007b. doi: 10.1038/nsmb1243. URL http://dx.doi.org/10.1038/nsmb1243.

J. F. Faucon and J. R. Lakowicz. Anisotropy decay of diphenylhexatriene in melittin-phospholipid complexes by multifrequency phase-modulation fluorometry. *Archives Of Biochemistry And Biophysics*, 252(1):245–258, Jan 1987.

F. Feroz, J. R. Gair, P. Graff, M. P. Hobson, and A. Lasenby. Classifying LISA gravitational wave burst signals using Bayesian evidence. *Classical And Quantum Gravity*, 27 (7):075010, April 2010.

J. N. Forkey, M. E. Quinlan, and Y. E. Goldman. Protein structural dynamics by single-molecule fluorescence polarization. *Progress in Biophysics and Molecular Biology*, 74 (1-2):1–35, 2000.

Theodor Förster. Zwischenmolekulare Energiewanderung und Fluoreszenz. *Annalen Der Physik*, 437:55–75, 1948.

J. Frank. *Three-dimensional Electron Microscopy of Macromolecular Assemblies*. Oxford University Press, Oxford, UK, 2006.

J. Fu, A. L. Gnatt, D. A. Bushnell, G. J. Jensen, N. E. Thompson, R. R. Burgess, P. R. David, and R. D. Kornberg. Yeast RNA polymerase II at 5 A resolution. *Cell*, 98(6): 799–810, Sep 1999.

W. Gander and W. Gautschi. Adaptive quadrature - revisited. *Bit*, 40:84–101, 2000.

M. A. Grachev and E. F. Zaychikov. Initiation by Escherichia coli RNA-polymerase: transformation of abortive to productive complex. *Febs Letters*, 115(1):23–26, Jun 1980.

Ian J Griswold and Frederick W Dahlquist. Bigger is better: megadalton protein NMR in solution. *Nature Structural Biology*, 9(8):567–568, Aug 2002. doi: 10.1038/nsb0802-567. URL http://dx.doi.org/10.1038/nsb0802-567.

Moritz Paul Haag. *Monte-Carlo-Simulation von Linker-gebundenen Farbstoffen an Proteinen*. Bachelor thesis, Ludwig-Maximilians Universität München, 2009.

Steven Hahn. Structure and mechanism of the RNA polymerase II transcription machinery. *Nature Structural and Molecular Biology*, 11(5):394–403, May 2004. doi: 10.1038/nsmb763. URL http://dx.doi.org/10.1038/nsmb763.

Z. Hillel and C. W. Wu. Statistical interpretation of fluorescence energy transfer measurements in macromolecular systems. *Biochemistry*, 15(10):2105–2113, May 1976.

Sungchul Hohng, Chirlmin Joo, and Taekjip Ha. Single-molecule three-color fret. *Biophys J*, 87(2):1328–1337, Aug 2004. doi: 10.1529/biophysj.104.043935. URL http://dx.doi.org/10.1529/biophysj.104.043935.

Seamus J Holden and Achillefs Kapanidis. Inference of the fluorescence signal of an immobilized fluorophore by fitting a 2D-Gaussian with a fixed center position to an EM-CCD camera movie increases signal/noise (S/N) ratio by a factor of approximately 2 compared to a binning procedure., March 2009. personal communication.

Invitrogen / Molecular Probes. Fluorescence quantum yields (qy) and lifetimes (τ) for Alexa Fluor dyes. Technical report, Invitrogen / Molecular Probes, 2010. URL http://www.invitrogen.com/site/us/en/home/References/Molecular-Probes-The-Handbook/tables/Fluorescence-quantum-yields-and-lifetimes-for-Alexa-Fluor-dyes.html.

Bibliography

Vassili Ivanov, Min Li, and Kiyoshi Mizuuchi. Impact of emission anisotropy on fluorescence spectroscopy and FRET distance measurements. *Biophysical Journal*, 97(3): 922–929, Aug 2009. doi: 10.1016/j.bpj.2009.05.025. URL http://dx.doi.org/10.1016/j.bpj.2009.05.025.

Andreas Janshoff, Marcus Neitzert, York Oberdörfer, and Harald Fuchs. Force spectroscopy of molecular systems-single molecule spectroscopy of polymers and biomolecules. *Angew Chem Int Ed Engl*, 39(18):3212–3237, Sep 2000.

Edwin Thompson Jaynes. *Statistical Physics*, chapter "Information Theory and Statistical Mechanics", pages 181–218. W. A. Benjamin, Inc., 1963.

Edwin Thompson Jaynes. Prior probabilities. *IEEE Transactions On System Science and Cybernetics*, 4:227–241, 1968.

Edwin Thompson Jaynes. *Probability Theory: The Logic of Science*. Cambridge University Press, 2003.

Harold Jeffreys. *Theory of probability*. Clarendon Press, Oxford, 1939.

Christian M Kaiser, Hung-Chun Chang, Vishwas R Agashe, Sathish K Lakshmipathy, Stephanie A Etchells, Manajit Hayer-Hartl, F. Ulrich Hartl, and José M Barral. Real-time observation of trigger factor function on translating ribosomes. *Nature*, 444(7118): 455–460, Nov 2006. doi: 10.1038/nature05225. URL http://dx.doi.org/10.1038/nature05225.

Stanislav Kalinin, Suren Felekyan, Matthew Antonik, and Claus A M Seidel. Probability distribution analysis of single-molecule fluorescence anisotropy and resonance energy transfer. *Journal Of Physical Chemistry B*, 111(34):10253–10262, Aug 2007. doi: 10.1021/jp072293p. URL http://dx.doi.org/10.1021/jp072293p.

Stanislav Kalinin, Evangelos Sisamakis, Steven W Magennis, Suren Felekyan, and Claus A M Seidel. On the origin of broadening of single-molecule FRET efficiency distributions beyond shot noise limits. *Journal Of Physical Chemistry B*, 114(18):6197–6206, May 2010a. doi: 10.1021/jp100025v. URL http://dx.doi.org/10.1021/jp100025v.

Stanislav Kalinin, Alessandro Valeri, Matthew Antonik, Suren Felekyan, and Claus A M Seidel. Detection of structural dynamics by FRET: a photon distribution and fluorescence lifetime analysis of systems with multiple states. *Journal Of Physical Chemistry B*, 114(23):7983–7995, Jun 2010b. doi: 10.1021/jp102156t. URL http://dx.doi.org/10.1021/jp102156t.

Achillefs N Kapanidis and Shimon Weiss. Fluorescent probes and bioconjugation chemistries for single-molecule fluorescence analysis of biomolecules. *Journal Of Chemical Physics*, 117(24):10953–10964, DEC 2002. ISSN 0021-9606. doi: {10.1063/1.1521158}.

Achillefs N Kapanidis, Emmanuel Margeat, Ted A Laurence, Sören Doose, Sam On Ho, Jayanta Mukhopadhyay, Ekaterine Kortkhonjia, Vladimir Mekler, Richard H Ebright, and Shimon Weiss. Retention of transcription initiation factor sigma70 in transcription elongation: single-molecule analysis. *Molecular Cell*, 20(3):347–356, Nov 2005. doi: 10.1016/j.molcel.2005.10.012. URL http://dx.doi.org/10.1016/j.molcel.2005.10.012.

Achillefs N Kapanidis, Emmanuel Margeat, Sam On Ho, Ekaterine Kortkhonjia, Shimon Weiss, and Richard H Ebright. Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. *Science*, 314(5802):1144–1147, Nov 2006. doi: 10.1126/science.1131399. URL http://dx.doi.org/10.1126/science.1131399.

Hubert Kettenberger, Karim-Jean Armache, and Patrick Cramer. Architecture of the RNA polymerase II-TFIIS complex and implications for mRNA cleavage. *Cell*, 114(3): 347–357, Aug 2003.

Hubert Kettenberger, Karim-Jean Armache, and Patrick Cramer. Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. *Molecular Cell*, 16(6):955–965, Dec 2004. doi: 10.1016/j.molcel.2004.11.040. URL http://dx.doi.org/10.1016/j.molcel.2004.11.040.

K. Kinosita, A. Ikegami, and S. Kawato. On the wobbling-in-cone analysis of fluorescence anisotropy decay. *Biophysical Journal*, 37(2):461–464, Feb 1982. doi: 10.1016/S0006-3495(82)84692-4. URL http://dx.doi.org/10.1016/S0006-3495(82)84692-4.

S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, May 1983. doi: 10.1126/science.220.4598.671. URL http://dx.doi.org/10.1126/science.220.4598.671.

Jennifer L Knight, Vladimir Mekler, Jayanta Mukhopadhyay, Richard H Ebright, and Ronald M Levy. Distance-restrained docking of rifampicin and rifamycin SV to RNA polymerase using systematic FRET measurements: developing benchmarks of model quality and reliability. *Biophysical Journal*, 88(2):925–938, Feb 2005. doi: 10.1529/biophysj.104.050187. URL http://dx.doi.org/10.1529/biophysj.104.050187.

Michel H Koch, Patrice Vachette, and Dmitri I Svergun. Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution. *Quarterly Reviews Of Biophysics*, 36(2):147–227, May 2003.

N. Korzheva, A. Mustaev, M. Kozlov, A. Malhotra, V. Nikiforov, A. Goldfarb, and S. A. Darst. A structural model of transcription elongation. *Science*, 289(5479):619–625, Jul 2000.

Dirk Kostrewa, Mirijam E Zeller, Karim-Jean Armache, Martin Seizl, Kristin Leike, Michael Thomm, and Patrick Cramer. RNA polymerase II-TFIIB structure and mechanism of transcription initiation. *Nature*, 462(7271):323–330, Nov 2009. doi: 10.1038/nature08548. URL http://dx.doi.org/10.1038/nature08548.

Wolfgang Kügel and Anders Barth. personal communication, 2010.

S. Kullback and R. A. Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86, 1951. ISSN 00034851. URL http://www.jstor.org/stable/2236703.

Joseph R. Lakowicz. *Principles of Fluorescence Spectroscopy, 3rd edition*. Springer, 2006.

Pierre-Simon Laplace. *Théorie analytique des probabilités*. Courcier Imprimeur, Paris, 1812.

Keren Lasker and Andrej Sali. Arbitrary geometry restraints, for example also a combined position and orientation restraint, could be implemented as a C++ class in the Integrated Modeling Platform (IMP)., Feb. 2010. personal communication.

Bibliography

Nam Ki Lee, Achillefs N Kapanidis, You Wang, Xavier Michalet, Jayanta Mukhopadhyay, Richard H Ebright, and Shimon Weiss. Accurate FRET measurements within single diffusing biomolecules using alternating-laser excitation. *Biophysical Journal*, 88(4): 2939–2953, Apr 2005. doi: 10.1529/biophysj.104.054114. URL http://dx.doi.org/10.1529/biophysj.104.054114.

Robert Alexander Lewis. *Untersuchung und Aufklärung des molekularen Mechanismus von Enzymen der SWI2/SNF2-Familie mit Hilfe der Einzelmolekül-Fluoreszenzmikroskopie*. PhD thesis, Ludwig-Maximilians-Universität München, 2009.

Jonathan S. Lindsey. Photochemcad software. http://www.photochemcad.com/, 2008. URL http://www.photochemcad.com/.

Xin Liu, David A Bushnell, Dong Wang, Guillermo Calero, and Roger D Kornberg. Structure of an RNA polymerase II-TFIIB complex and the transcription initiation mechanism. *Science*, 327(5962):206–209, Jan 2010a. doi: 10.1126/science.1182015. URL http://dx.doi.org/10.1126/science.1182015.

Yang Liu, Jeehae Park, Karin A Dahmen, Yann R Chemla, and Taekjip Ha. A comparative study of multivariate and univariate hidden Markov modelings in time-binned single-molecule FRET data analysis. *Journal Of Physical Chemistry B*, 114(16):5386–5403, Apr 2010b. doi: 10.1021/jp9057669. URL http://dx.doi.org/10.1021/jp9057669.

Anastasia Loman, Ingo Gregor, Christina Stutz, Markus Mund, and Jörg Enderlein. Measuring rotational diffusion of macromolecules by fluorescence correlation spectroscopy. *Photochem Photobiol Sci*, 9(5):627–636, May 2010. doi: 10.1039/b9pp00029a. URL http://dx.doi.org/10.1039/b9pp00029a.

Douglas Magde, Elliot Elson, and Watt W. Webb. Thermodynamic fluctuations in a reacting system – measurement by fluorescence correlation spectroscopy. *Phys. Rev. Lett.*, 29(11):705–708, September 1972. URL http://link.aps.org/doi/10.1103/PhysRevLett.29.705.

Douglas Magde, Elliot Elson, and Watt W. Webb. Fluorescence correlation spectroscopy. II. An experimental realization. *Biopolymers*, 13(1):29–61, Jan 1974. doi: 10.1002/bip. 1974.360130103. URL http://dx.doi.org/10.1002/bip.1974.360130103.

Emmanuel Margeat, Achillefs N Kapanidis, Philip Tinnefeld, You Wang, Jayanta Mukhopadhyay, Richard H Ebright, and Shimon Weiss. Direct observation of abortive initiation and promoter escape within single immobilized transcription complexes. *Biophysical Journal*, 90(4):1419–1431, Feb 2006. doi: 10.1529/biophysj.105.069252. URL http://dx.doi.org/10.1529/biophysj.105.069252.

M. Margittai, J. Widengren, E. Schweinberger, G. F. Schröder, S. Felekyan, E. Haustein, M. König, D. Fasshauer, H. Grubmüller, R. Jahn, and C. A M Seidel. Single-molecule fluorescence resonance energy transfer reveals a dynamic equilibrium between closed and open conformations of syntaxin 1. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 100(26):15516–15521, Dec 2003. doi: 10.1073/pnas. 2331232100. URL http://dx.doi.org/10.1073/pnas.2331232100.

James J. McCann, Ucheor B. Choi, Liqiang Zheng, Keith Weninger, and Mark E. Bowen. Optimizing methods to recover absolute FRET efficiency from immobilized single molecules. *Biophysical Journal*, 99(3):961–970, August 2010. ISSN 0006-3495. URL http://linkinghub.elsevier.com/retrieve/pii/S000634951000562X.

Sean A McKinney, Chirlmin Joo, and Taekjip Ha. Analysis of single-molecule FRET trajectories using hidden Markov modeling. *Biophysical Journal*, 91(5):1941–1951, Sep 2006. doi: 10.1529/biophysj.106.082487. URL http://dx.doi.org/10.1529/biophysj.106.082487.

I. L. Medintz, J. H. Konnert, A. R. Clapp, I. Stanish, M. E. Twigg, H. Mattoussi, J. M. Mauro, and J. R. Deschamps. A fluorescence resonance energy transfer-derived structure of a quantum dot-protein bioconjugate nanoassembly. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 101(26):9612–9617, Jun 2004. doi: 10.1073/pnas.0403343101. URL http://dx.doi.org/10.1073/pnas.0403343101.

Vladimir Mekler, Ekaterine Kortkhonjia, Jayanta Mukhopadhyay, Jennifer Knight, Andrei Revyakin, Achillefs N Kapanidis, Wei Niu, Yon W Ebright, Ronald Levy, and Richard H Ebright. Structural organization of bacterial RNA polymerase holoenzyme and the RNA polymerase-promoter open complex. *Cell*, 108(5):599–614, Mar 2002.

Nicholas Metropolis and S. Ulam. The Monte Carlo method. *Journal Of The American Statistical Association*, 44(247):335–341, Sep 1949.

Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. Equation of state calculations by fast computing machines. *Journal Of Chemical Physics*, 21(6):1087–1092, June 1953. URL http://link.aip.org/link/?JCP/21/1087/1.

Xavier Michalet, Achillefs N Kapanidis, Ted Laurence, Fabien Pinaud, Sören Doose, Malte Pflughoefft, and Shimon Weiss. The power and prospects of fluorescence microscopies and spectroscopies. *Annual Review of Biophysics and Biomolecular Structure*, 32:161–182, 2003. doi: 10.1146/annurev.biophys.32.110601.142525. URL http://dx.doi.org/10.1146/annurev.biophys.32.110601.142525.

Kim I Mortensen, L. Stirling Churchman, James A Spudich, and Henrik Flyvbjerg. Optimized localization analysis for single-molecule tracking and super-resolution microscopy. *Nature Methods*, 7(5):377–381, May 2010. doi: 10.1038/nmeth.1447. URL http://dx.doi.org/10.1038/nmeth.1447.

Jayanta Mukhopadhyay, Achillefs N. Kapanidis, Vladimir Mekler, Ekaterine Kortkhonjia, Yon W. Ebright, and Richard H. Ebright. Translocation of $\sigma70$ with RNA polymerase during transcription: Fluorescence resonance energy transfer assay for movement relative to DNA. *Cell*, 106(4):453–463, August 2001. ISSN 0092-8674. URL http://linkinghub.elsevier.com/retrieve/pii/S0092867401004640.

Jayanta Mukhopadhyay, Elena Sineva, Jennifer Knight, Ronald M Levy, and Richard H Ebright. Antibacterial peptide microcin J25 inhibits transcription by binding within and obstructing the RNA polymerase secondary channel. *Molecular Cell*, 14(6):739–751, Jun 2004. doi: 10.1016/j.molcel.2004.06.010. URL http://dx.doi.org/10.1016/j.molcel.2004.06.010.

Barbara K Müller, Evgeny Zaychikov, Christoph Bräuchle, and Don C Lamb. Pulsed interleaved excitation. *Biophysical Journal*, 89(5):3508–3522, Nov 2005. doi: 10.1529/biophysj.105.064766. URL http://dx.doi.org/10.1529/biophysj.105.064766.

Mervin E. Muller. Some continuous Monte Carlo methods for the Dirichlet problem. *Annals Of Mathematical Statistics*, 27:569–589, 1956.

Bibliography

Mervin E. Muller. A note on a method for generating points uniformly on *n*-dimensional spheres. *Communications of the ACM*, 2(4):19–20, 1959. ISSN 0001-0782. doi: http://doi.acm.org/10.1145/377939.377946.

L. M. Munson and W. S. Reznikoff. Abortive initiation and long ribonucleic acid synthesis. *Biochemistry*, 20(8):2081–2085, Apr 1981.

Katsuhiko S Murakami and Seth A Darst. Bacterial RNA polymerases: the wholo story. *Current Opinion in Structural Biology*, 13(1):31–39, Feb 2003.

Adam Muschielok, Joanna Andrecka, Anass Jawhari, Florian Brückner, Patrick Cramer, and Jens Michaelis. A nano-positioning system for macromolecular structural analysis. *Nature Methods*, 5(11):965–971, November 2008. ISSN 1548-7105. doi: 10.1038/nmeth.1259. URL http://www.ncbi.nlm.nih.gov/pubmed/18849988. PMID: 18849988.

N. Naryshkin, A. Revyakin, Y. Kim, V. Mekler, and R. H. Ebright. Structural organization of the RNA polymerase-promoter open complex. *Cell*, 101(6):601–611, Jun 2000.

Radford M. Neal. Probabilistic inference using Markov chain Monte Carlo methods. Technical report, Department of Computer Science, University of Toronto, 1993.

Radford M. Neal. *Handbook of Markov Chain Monte Carlo*, chapter MCMC using Hamiltonian dynamics. Chapman & Hall / CRC Press, 2010.

D. B. Nikolov, H. Chen, E. D. Halay, A. A. Usheva, K. Hisatake, D. K. Lee, R. G. Roeder, and S. K. Burley. Crystal structure of a TFIIB-TBP-TATA-element ternary complex. *Nature*, 377(6545):119–128, Sep 1995. doi: 10.1038/377119a0. URL http://dx.doi.org/10.1038/377119a0.

Eyal Nir, Xavier Michalet, Kambiz M Hamadani, Ted A Laurence, Daniel Neuhauser, Yevgeniy Kovchegov, and Shimon Weiss. Shot-noise limited single-molecule FRET histograms: comparison between theory and experiments. *Journal Of Physical Chemistry B*, 110(44):22103–22124, Nov 2006. doi: 10.1021/jp063483n. URL http://dx.doi.org/10.1021/jp063483n.

Erwin J. G. Peterman, H. Sosa, L. S. Goldstein, and W. E. Moerner. Polarized fluorescence microscopy of individual and many kinesin motors bound to axonemal microtubules. *Biophysical Journal*, 81(5):2851–2863, Nov 2001. doi: 10.1016/S0006-3495(01)75926-7. URL http://dx.doi.org/10.1016/S0006-3495(01)75926-7.

Eric F Pettersen, Thomas D Goddard, Conrad C Huang, Gregory S Couch, Daniel M Greenblatt, Elaine C Meng, and Thomas E Ferrin. UCSF chimera–a visualization system for exploratory research and analysis. *Journal Of Computational Chemistry*, 25(13):1605–1612, Oct 2004. doi: 10.1002/jcc.20084. URL http://dx.doi.org/10.1002/jcc.20084.

Marta Radman-Livaja, Tapan Biswas, Dale Mierke, and Arthur Landy. Architecture of recombination intermediates visualized by in-gel FRET of lambda integrase-Holliday junction-arm DNA complexes. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 102(11):3913–3920, Mar 2005. doi: 10.1073/pnas.0500844102. URL http://dx.doi.org/10.1073/pnas.0500844102.

Ivan Rasnik, Sua Myong, Wei Cheng, Timothy M Lohman, and Taekjip Ha. DNA-binding orientation and domain conformation of the E. coli rep helicase monomer bound to a partial duplex junction: single-molecule studies of fluorescently labeled enzymes. *Journal Of Molecular Biology*, 336(2):395–408, Feb 2004.

W. A. Rees, R. W. Keller, J. P. Vesenka, G. Yang, and C. Bustamante. Evidence of DNA bending in transcription complexes imaged by scanning force microscopy. *Science*, 260 (5114):1646–1649, Jun 1993.

Andrey Revyakin, Chenyu Liu, Richard H Ebright, and Terence R Strick. Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. *Science*, 314(5802):1139–1143, Nov 2006. doi: 10.1126/science.1131398. URL http://dx.doi.org/10.1126/science.1131398.

Wolfgang Rieping, Michael Habeck, and Michael Nilges. Inferential structure determination. *Science*, 309(5732):303–306, Jul 2005. doi: 10.1126/science.1110428. URL http://dx.doi.org/10.1126/science.1110428.

Paul W K Rothemund. Folding dna to create nanoscale shapes and patterns. *Nature*, 440 (7082):297–302, Mar 2006. doi: 10.1038/nature04586. URL http://dx.doi.org/10.1038/nature04586.

Rahul Roy, Sungchul Hohng, and Taekjip Ha. A practical guide to single-molecule FRET. *Nature Methods*, 5(6):507–516, Jun 2008. doi: 10.1038/nmeth.1208. URL http://dx.doi.org/10.1038/nmeth.1208.

Matthew E Sanborn, Brian K Connolly, Kaushik Gurunathan, and Marcia Levitus. Fluorescence properties and photophysics of the sulfoindocyanine Cy3 linked covalently to DNA. *Journal Of Physical Chemistry B*, 111(37):11064–11074, Sep 2007. doi: 10.1021/jp072912u. URL http://dx.doi.org/10.1021/jp072912u.

Yusdi Santoso, Catherine M Joyce, Olga Potapova, Ludovic Le Reste, Johannes Hohlbein, Joseph P Torella, Nigel D F Grindley, and Achillefs N Kapanidis. Conformational transitions in DNA polymerase I revealed by single-molecule FRET. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 107(2):715–720, Jan 2010. doi: 10.1073/pnas.0910909107. URL http://dx.doi.org/10.1073/pnas.0910909107.

Andrej Savol and Chakra Chennubhotla. Protein trajectory reconstruction strategies from FRET-derived distance constraints. *Biophysical Journal*, 98(3, Supplement 1):445a–445a, January 2010. ISSN 0006-3495. URL http://www.sciencedirect.com/science/article/B94RW-4Y7W930-2X5/2/65ed8c2a0736ba78d4379559c106768f. Poster presentation.

Gunnar F. Schröder and Helmut Grubmüller. FRETsg: Biomolecular structure model building from multiple FRET experiments. *Computer Physics Communications*, 158(3):150–157, April 2004. ISSN 0010-4655. URL http://www.sciencedirect.com/science/article/B6TJ5-4BYR8SD-1/2/2c1b38ecf64c6cd96490a9b83a099fb1.

Gunnar F Schröder, Ulrike Alexiev, and Helmut Grubmüller. Simulation of fluorescence anisotropy experiments: probing protein dynamics. *Biophysical Journal*, 89(6):3757–3770, Dec 2005. doi: 10.1529/biophysj.105.069500. URL http://dx.doi.org/10.1529/biophysj.105.069500.

Bibliography

Benjamin Schuler and William A Eaton. Protein folding studied by single-molecule FRET. *Curr Opin Struct Biol*, 18(1):16–26, Feb 2008. doi: 10.1016/j.sbi.2007.12.003. URL http://dx.doi.org/10.1016/j.sbi.2007.12.003.

T. Schwarz-Selinger, R. Preuss, V. Dose, and W. von der Linden. Analysis of multicomponent mass spectra applying Bayesian probability theory. *Journal of Mass Spectrometry*, 36(8):866–874, 2001. ISSN 1096-9888. URL http://dx.doi.org/10.1002/jms.187.

J. R. Shaw, M. Bridges, and M. P. Hobson. Nested sampling for general Bayesian computation. *Monthly Notices Of The Royal Astronomical Society*, 378(4):1365–1370, July 2007.

Andrea Sinz. Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein-protein interactions. *Mass Spectrometry Reviews*, 25 (4):663–682, 2006. doi: 10.1002/mas.20082. URL http://dx.doi.org/10.1002/mas.20082.

Devinderjit Sivia. *Data Analysis, 2nd edition*. Oxford University Press, USA, 2006.

John Skilling. Nested sampling for general Bayesian computation. *Bayesian Analysis*, 1: 833–860, 2006. doi: 10.1214/06-BA127.

T. A. Steitz. A mechanism for all polymerases. *Nature*, 391:231–232, 1998.

Alexander Strehl and Joydeep Ghosh. Cluster ensembles - a knowledge reuse framework for combining multiple partitions. *Journal of Machine Learning Research*, 3:583–617, 2002. doi: 10.1162/153244303321897735.

Lubert Stryer and Richard P. Haugland. Energy transfer: a spectroscopic ruler. *Proceedings Of The National Academy Of Sciences Of The United States Of America-Physical Sciences*, 58:719–726, 1967.

Xingmin Sun, Dale F Mierke, Tapan Biswas, Sang Yeol Lee, Arthur Landy, and Marta Radman-Livaja. Architecture of the 99 bp DNA-six-protein regulatory complex of the lambda att site. *Molecular Cell*, 24(4):569–580, Nov 2006. doi: 10.1016/j.molcel.2006.10.006. URL http://dx.doi.org/10.1016/j.molcel.2006.10.006.

Robert H. Swendsen and Jian-Sheng Wang. Replica Monte Carlo simulation of spinglasses. *Physical Review Letters*, 57(21):2607–2609, November 1986. URL http://link.aps.org/doi/10.1103/PhysRevLett.57.2607.

D. Thieffry and S. Sarkar. Forty years under the central dogma. *Trends In Biochemical Sciences*, 23(8):312–316, Aug 1998.

Barbara Treutlein, Adam Muschielok, Dirk Kostrewa, Patrick Cramer, and Jens Michaelis. Position of TBP and the TATA box within the RNA polymerase II initial transcribing complex. unpublished.

Andrea Újvári and Donal S Luse. RNA emerging from the active site of RNA polymerase II interacts with the Rpb7 subunit. *Nat Struct Mol Biol*, 13(1):49–54, Jan 2006. doi: 10.1038/nsmb1026. URL http://dx.doi.org/10.1038/nsmb1026.

Stephan Uphoff, Seamus J Holden, Ludovic Le Reste, Javier Periz, Sebastian van de Linde, Mike Heilemann, and Achillefs N Kapanidis. Monitoring multiple distances within a single molecule using switchable FRET. *Nature Methods*, 7(10):831–836, Oct 2010. doi: 10.1038/nmeth.1502. URL http://dx.doi.org/10.1038/nmeth.1502.

G. Vamosi, C. Gohlke, and R.M. Clegg. Fluorescence characteristics of 5-carboxytetramethylrhodamine linked covalently to the 5' end of oligonucleotides: multiple conformers of single-stranded and double-stranded dye-DNA complexes. *Biophysical Journal*, 71(2):972–994, August 1996. ISSN 0006-3495. URL http://www.sciencedirect.com/science/article/B94RW-4V9K7CV-1M/2/07b7c86ee87c8ce7120e7bcfa2d05f3a.

M. van Heel, B. Gowen, R. Matadeen, E. V. Orlova, R. Finn, T. Pape, D. Cohen, H. Stark, R. Schmidt, M. Schatz, and A. Patwardhan. Single-particle electron cryo-microscopy: towards atomic resolution. *Quarterly Reviews Of Biophysics*, 33(4):307–369, Nov 2000.

Darren B VanBeek, Matthew C Zwier, Justin M Shorb, and Brent P Krueger. Fretting about FRET: correlation between kappa and R. *Biophysical Journal*, 92(12):4168–4178, Jun 2007. doi: 10.1529/biophysj.106.092650. URL http://dx.doi.org/10.1529/biophysj.106.092650.

D. Wang and R. Landick. Nuclease cleavage of the upstream half of the nontemplate strand DNA in an Escherichia coli transcription elongation complex causes upstream translocation and transcriptional arrest. *Journal Of Biological Chemistry*, 272(9):5989–5994, Feb 1997.

Anna K Woźniak, Gunnar F Schröder, Helmut Grubmüller, Claus A M Seidel, and Filipp Oesterhelt. Single-molecule FRET measures bends and kinks in DNA. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 105(47):18337–18342, Nov 2008. doi: 10.1073/pnas.0800977105. URL http://dx.doi.org/10.1073/pnas.0800977105.

K. Wüthrich. *NMR of Proteins and Nucleic Acids*. Wiley, New York, 1986.

Brian A Young, Tanja M Gruber, and Carol A Gross. Views of transcription initiation. *Cell*, 109(4):417–420, May 2002.

List of publications

Publications related to the thesis

peer reviewed journals

- A. Muschielok, J. Andrecka, A. Jawhari, F. Brückner, P. Cramer, and J. Michaelis.
  A nano-positioning system for macromolecular structural analysis.
  Nature Methods, 5(11):965–971, November 2008. ISSN 1548–7105.
  doi: 10.1038/nmeth.1259
- J. Andrecka, B. Treutlein, M.A. Izquierdo Arcusa, A. Muschielok, R. Lewis, A.C.M. Cheung, P. Cramer, and J. Michaelis.
  Nano positioning system reveals the course of upstream and nontemplate DNA within the RNA polymerase II elongation complex.
  Nucleic Acids Research, 37(17):5803–5809, September 2009. ISSN 1362-4962.
  doi: 10.1093/nar/gkp601

invited presentations

- *A nano-positioning system for macromolecular structural analysis.*
  European Workshop on Advanced Fluorescence Imaging and Dynamics 2008, München, Germany
- *A nano-positioning system for macromolecular structural analysis.*
  Single Molecule Image Analysis and Systems Biology Workshop 2009, Harwell, UK
- *The Nano-Positioning System – a FRET-based tool for macro-molecular structural analysis.*
  Biophysical Society Meeting 2010, San Francisco, CA, USA

posters

- A. Muschielok, J. Andrecka, P. Cramer, J. Michaelis,
  The Nano-Positioning System – a Bayesian approach.,
  Biophysical Society Meeting 2008, Long Beach, CA, USA
- A. Muschielok, J. Andrecka, A. Jawhari, F. Brückner, P. Cramer, and J. Michaelis
  A nano-positioning system for macromolecular structural analysis.,
  Single Molecule Biophysics Winter Workshop 2009, Aspen, CO, USA
- A. Muschielok, J. Andrecka, B. Treutlein, J. Michaelis,
  Extending the Nano-positioning system.,
  Biophysical Society Meeting 2010, San Francisco, CA, USA
- J. Michaelis, A. Muschielok, B. Treutlein, J. Andrecka,
  Extending the Nano-positioning system,
  Single Molecule Approaches to Biology (Gordon Research Conference 2010),
  Lucca (Barga), Italy

Publications not related to the thesis

- J. Michaelis, A. Muschielok, J. Andrecka, W. Kügel, and J.R. Moffitt.
  DNA based molecular motors.
  Phys Life Rev, Oct 2009. doi: 10.1016/j.plrev.2009.09.001

Curriculum Vitae

2006 - 2010	Graduate studies in the group of Prof. Jens Michaelis, LMU München - *Bayesian data analysis of single-molecule fluorescence experiments* - *Set-up of ultra-stable optical tweezers*
2004 - 2005	Diploma Thesis in the group of Dr. Andreas Volkmer, U Stuttgart *Investigation of Collagen Type I Fibrils by CARS Microscopy*
2004	Internship at WITec GmbH, Ulm *Test measurements with atomic force and Raman microscopy systems*
2004	Research student in the group of Prof. Wolfgang Zinth, LMU München *Fourier-Transformed IR spectroscopy on photoswitchable peptides*
2002 - 2003	Research student in the group of Dr. Motomu Tanaka, TU München *Production and characterization of ultra thin cellulose films on silicon wafers*
2000 - 2005	Studies of physics/biophysics, TU München
1990 - 1999	Gymnasium Miesbach (secondary school)

Acknowledgement