Simulation and Analysis of Protein-Fluorophore Systems for Comparison with Fluorescence Spectroscopy Data

Zur Erlangung des akademischen Grades eines

DOKTORS DER NATURWISSENSCHAFTEN (Dr. rer. nat.)

von der KIT-Fakultät für Physik des
Karlsruher Instituts für Technologie (KIT)
angenommene

DISSERTATION

von

Ines Reinartz, M.Sc.

Tag der mündlichen Prüfung: 03. Mai 2019
Referent: Prof. Dr. Wolfgang Wenzel
Korreferent: Prof. Dr. Gerd Ulrich Nienhaus

Simulation and Analysis of Protein-Fluorophore Systems for Comparison with Fluorescence Spectroscopy Data

Abstract

Biomolecules and in particular proteins are essential for life at the molecular level and facilitate various biological functions such as molecular transport, cell motion, or catalysis of chemical reactions. One example is the protein hemoglobin in the red blood cells, which transports oxygen in the human body. Malfunction of proteins can cause severe degenerative diseases such as Parkinson's disease, Huntington's disease, Alzheimer's disease, or variants of amyotrophic lateral sclerosis (ALS). Thus, understanding of biomolecular function is of fundamental importance for the fields of biology, pharmacology, and medical sciences and requires detailed insights into protein structures and dynamics. As proteins are too small to be directly observed with an optical microscope, indirect techniques have to be used.

One approach is the use of Förster resonance energy transfer (FRET), where fluorophores are utilized to study protein dynamics and to observe molecular processes *in vitro* and *in vivo*. Similarly, fluorophores can be employed in biosensors to measure concentrations of small biomolecules such as glucose, having many applications in medical sciences and microbiology. All of these protein-fluorophore systems are governed by the underlying physical processes such as molecular dynamics and photophysics. The fundamental mechanisms in these systems can not be observed directly and are therefore not fully understood.

A technique complementing experimental measurements are molecular simulations. They facilitate a detailed insight into the molecular systems and their microscopic function. The current methods to model protein-fluorophore systems are mostly approximations applicable for specific cases or computationally too demanding to model all relevant motions. This work introduces a new method for simulating dynamics of protein-fluorophore systems with a computationally efficient model based on coarse-grained molecular dynamics simulations. While requiring only few parameters it yields a realistic description of the system in quantitative agreement with experiments. As the presented computational method can directly be compared to experimental data, it facilitates improving planning and microscopic interpretation of experiments. It also yields information about the underlying dynamics by enabling to trace the motions of biomolecules in atomic detail. This work establishes a systematic simulation protocol to study protein-fluorophore systems *in silico* that can easily be applied to study a large range of biologically relevant applications. It shows that experiments and simulations complement each other leading to new insights into biomolecular dynamics and function.

Simulation und Analyse von Protein-Fluorophor-Systemen für den Vergleich mit Fluoreszenzspektroskopie-Daten

Zusammenfassung

Proteine sind die Grundbausteine des Lebens auf molekularer Ebene und wichtig für viele biologische Funktionen wie den Transport von Molekülen, Zellbewegungen oder die Katalyse von chemischen Reaktionen. So transportiert das Protein Hämoglobin beispielsweise den Sauerstoff im menschlichen Blut. Störungen der Proteinfunktionen können schwere degenerative Krankheiten wie zum Beispiel die Parkinson-, Huntington- oder Alzheimer-Krankheit verursachen. Das Verständnis von Proteinfunktionen, ihrer Struktur und Dynamik ist daher ein wichtiges Forschungsgebiet in Biologie, Pharmakologie und Medizin. Da Proteine aufgrund ihrer geringen Größe nicht mit Lichtmikroskopie beobachtet werden können, verwendet man stattdessen indirekte Methoden.

Eine dieser Methoden macht sich den Förster-Resonanzenergietransfer (FRET) zunutze, um damit Proteindynamik und andere molekulare Prozesse *in vitro* und *in vivo* zu untersuchen. Die Methode wird außerdem auch in Biosensoren zur Messung der Konzentrationen von kleinen Biomolekülen wie zum Beispiel Glukose eingesetzt. Die dabei verwendeten Systeme aus Proteinen und Fluorophoren unterliegen physikalischen Prozessen wie Molekulardynamik und Photophysik. Da man diese Mechanismen nicht direkt beobachten kann, ist die Funktionsweise vieler Systeme noch nicht vollständig verstanden.

Molekulare Simulationen können diese experimentellen Messungen ergänzen. Sie ermöglichen einen Einblick in molekulare Systeme und ihre Funktion auf atomarer Ebene. Die bisherigen Modellierungsmethoden für Protein-Fluorophor-Systeme sind größtenteils Näherungen, die nur für spezielle Anwendungen verwendbar sind oder zu rechenaufwändig um alle relevanten Bewegungen zu modellieren. Diese Arbeit stellt eine neue Methode für die Simulation der Dynamik in Protein-Fluorophor-Systemen vor. Sie basiert auf recheneffizienten vereinfachten Molekulardynamiksimulationen. Mit nur wenigen Parametern bietet die Methode eine realistische Beschreibung des Systems, die quantitativ mit Experimenten übereinstimmt. Sie ermöglicht den direkten Vergleich von Simulationen mit experimentellen Daten und somit eine bessere Planung und Interpretation von Experimenten. Gleichzeitig liefert sie Informationen über die zugrundeliegende Dynamik der Systeme. Diese Arbeit präsentiert ein systematisches Simulationsprotokoll für die Modellierung von Protein-Fluorophor-Systemen *in silico*, welches für die Erforschung von vielen biologisch relevanten Anwendungen verwendet werden kann. Sie zeigt wie Experimente und Simulationen einander ergänzen, um neue Einblicke in Dynamik und Funktion von Biomolekülen zu erhalten.

Contents

1
Introduction

Proteins are large macromolecules and fundamental building blocks of life on the molecular level. They are the functional components of cells and essential for various biological functions such as molecular transport, cell motion, and catalysis. Examples are e. g. the protein hemoglobin, transporting oxygen in the human body or collagen, a structural protein and one of the main components of connective tissue. Protein malfunction can cause severe degenerative diseases such as Parkinson's disease, Huntington's disease, Alzheimer's disease, or variants of amyotrophic lateral sclerosis (ALS). Thus, understanding of protein function is essential for the fundamental research in biology and biophysics, as well as for applications in pharmacology and medical sciences. The strong interrelation between protein structure, dynamics, and function is one of the main paradigms in biophysics. Studying protein structure and dynamics is therefore crucial to understand proteins and their function.

With their sizes in the nanometer range, proteins are too small to be directly observed with an optical microscope. Hence, to access information about proteins on a molecular scale, development of imaging methods capable of capturing their dynamics is essential. Techniques such as nuclear magnetic resonance (NMR) spectroscopy and X-ray crystallography are limited in their application. Where X-ray requires elaborately crystallized proteins to measure protein structure, NMR is limited by system size. One approach to access dynamic structural information about proteins is fluorescence spectroscopy. Fluorophores are utilized to study protein dynamics, to elucidate biomolecular function, and to observe molecular processes *in vitro* and *in vivo*. Similarly, specific biosensors can quantitatively measure the concentration of small molecules that interact with proteins by change in their spectroscopic properties. In both applications, only specific properties of the fluorophores can be directly measured and need to be carefully interpreted. This interpretation can be complemented by molecular simulations of the underlying dynamics of the protein-fluorophore systems. I introduce both approaches in the following.

A particularly powerful technique in fluorescence spectroscopy is Förster resonance energy transfer (FRET). Here, fluorophores, often small organic dyes or fluorescent proteins, are attached to a protein via flexible linkers. Acting as a "spectroscopic ruler" in the nanometer range [1], FRET gives information about inter- and intramolecular distances by measuring the distance dependent transfer efficiency between these fluorophores. Furthermore, FRET can be used to obtain time-resolved information about e. g. unfolded or intrinsically disordered proteins (IDPs) [2–4], protein-folding dynamics [5], folding intermediates [6, 7], or conformational transitions [8]. FRET does not directly provide distance information but rather the transfer efficiency. The efficiency depends on several parameters such as the spectroscopic properties of the fluorophores, the distance between them, their mutual orientation, and therefore also their shapes and dynamic behavior. These dependencies complicate planning and interpretation of FRET measurements, especially when the motions of the fluorophores are anisotropic, conformationally restricted, or slow. In these cases, simplifying assumptions such as neglectable relative orientations of the fluorophores are not valid.

Similar to studying proteins in FRET experiments, FRET-based biosensors utilize fluorescent proteins for measuring the concentration of small biomolecules that can bind to a protein [9–12]. The conformational change of such a sensing protein upon binding to this molecule yields an associated change in the FRET signal. FRET-based biosensors can thus be used for *in vitro* and *in vivo* measurement and visualization of small molecules such as glucose [13], with many applications in medical sciences and microbiology. Mechanisms involved in the change of FRET efficiency are reorientation of the fluorophores, changes of their rotational flexibility, or changes in the inter-fluorophore distances. These nanoscale processes are often not fully understood and the sensors are typically engineered by a cost and time extensive experimental trial and error optimization process [13, 14].

As the quality of measured FRET efficiencies is crucial for experiments, significant scientific effort is spend to reduce experimental artifacts, enhance spectroscopic fluorophore properties, and improve methods to interpret the underlying mechanisms and molecular motions. A technique complementing experimental FRET measurements are simulations of biomolecules with submolecular or atomistic resolution. Computational modeling and simulation of protein-fluorophore systems can yield additional information about the underlying dynamics of the FRET process. Thereby, they can resolve the challenges in interpretation of FRET measurements and elucidate the microscopic processes involved in FRET-based biosensors.

To date, there are several simulation approaches to improve understanding of FRET measurements as well as engineering biosensors. One approach to model small organic dyes attached to proteins is to calculate the accessible volume of these FRET dyes [15–19]. From the accessible volume, inter-dye distance distributions can be derived and used to plan experiments. This method does not account for dye dynamics, in particular it neglects the relative orientation of the dyes, as well as photon statistics. The model is thus not transferable to other fluorophores with a different chemical structure such as fluorescent proteins. Another model describes fluorescent proteins by sampling their conformational space using rigid body modeling. The sampling yields knowledge about the system's structural ensemble [20]. Still, this model does not include the influence of the system's dynamics, linker rigidity, photon statistics, or the weak dimerization tendency of some fluorescent proteins on the FRET efficiency. A third approach, analytical polymer models, can be used to analyze and interpret FRET data of dyes attached to unfolded and intrinsically disordered proteins [21]. Because this model is only valid for flexible (and thus unfolded) polymer chains, it can not be used for folded protein ensembles.

An approach which, in principle, can resolve all of the aforementioned challenges are molecular dynamics (MD) simulations. They provide insight into dynamics and function of molecular systems on an atomistic level. Aiming at an accurate model of the system's dynamics, MD simulations are computationally considerably more demanding than the other approaches. Previous work has applied MD simulations to study protein-dye systems, where biomolecular force fields have been parametrized to describe specific dye molecules [22, 23]. The simulation results have been compared to experimental observations such as fluorescence anisotropy [22] or mean FRET efficiency [23]. These MD simulations have revealed that dye and linker fluctuations [24] as well as shape and mutual orientation [25] play an important role in FRET studies. They furthermore allow to test the widely used assumption of a constant and isotropic orientation factor of $\overline{\kappa^2} = 2/3$, which has been questioned in several studies [26, 27], in particular for low linker flexibility [28]. Nevertheless, the prohibitive computational costs of MD simulations to reach sufficiently long simulation times limit their application to comparably small system sizes. MD simulations are infeasible for simulating e. g. slow or large-scale motions, structurally diverse ensembles such as unfolded proteins, or large systems, in particular systems comprising multiple fluorescent proteins.

In this work, I introduce a new method for simulating dynamics of protein-fluorophore systems. It enables simulation of proteins with small organic dye molecules as well as fluorescent proteins to gain insight into the underlying dynamics and structural ensembles. To achieve sufficiently long simulated times corresponding to a biologically relevant physical time scale, I use a coarse-grained description level of the molecular systems [29], which was originally developed to study protein folding. The model maintains full flexibility of the system and includes all heavy atoms of proteins and fluorophores. The formulation of this native structure-based model (SBM) is based on energy landscape theory and the *principle of minimal frustration* [30–33]. SBMs can sample large conformational ensembles with considerably lower computational costs than needed for regular MD simulations. The simulations yield a realistic description of the system with only few parameters in contrast to the complex force fields used in regular MD simulations.

The presented simulation method provides direct access to full atomic information of the entire system, including derived properties such as distance and orientation distributions of FRET fluorophores. It facilitates planning and interpretation of experimental FRET measurements and can be utilized to derive experimentally inaccessible properties from simulations. In contrast to the existing techniques the method presented here is particularly useful for systems with a restricted conformational space of the fluorophore or very slow fluorophore dynamics. Additionally, it allows for description of protein-dye systems both in their folded and unfolded states within a single model. As a computational model directly comparable to experimental FRET and small-angle X-ray scattering (SAXS) data, it furthermore improves our understanding of structural ensembles and microscopic function of biosensors.

The goal of this work is to establish a systematic simulation protocol to study protein-fluorophore systems computationally, which can easily be employed to study a large range of biological relevant applications.

Chapter 2 presents the theoretical background of this work. It gives an introduction to proteins, along with energy landscape theory. The latter is required to derive native structure-based models, the simulation method used throughout this work, which will be discussed in Chapter 3. Chapter 2 furthermore describes the theory of Förster resonance energy transfer (FRET) with particular focus on determination of FRET efficiencies in experiments and simulations. Moreover, the principles of fluorescence anisotropy used for an adjustment of the simulation time scale are discussed. Small-angle X-ray scattering (SAXS) measurements can be used to coarsely derive the shape of molecules. I use data from this technique as input for the biosensor simulations and investigate the interplay between FRET and SAXS in Chapter 5. Finally, an analytical polymer model to enable comparing the presented simulations to other analysis methods is introduced.

Chapter 3 describes computational methods, starting with the fundamental ideas of molecular dynamics (MD). Due to the prohibitive computational costs of MD to simulate large conformational ensembles or systems with slow dynamics, I use the computationally more efficient native structure-based models (SBMs). Moreover, a Monte Carlo method to compute photon statistics from simulations is described. It yields FRET efficiency histograms which can be used for direct comparison of experimental and simulated data.

Chapter 4 presents the method I have developed to simulate protein-fluorophore systems. I describe a systematic way to incorporate FRET fluorophores into SBM simulations and obtain structures and parameters for modeling of the whole protein-fluorophore systems.

Sections 4.1 to 4.5 explain the implementation of small organic dyes and proteins in SBMs for simulation of folded and unfolded ensembles. Additionally, I present an extension of the simulation protocol to obtain two-color and three-color FRET data.

In Section 4.6, I describe the incorporation of fluorescent proteins in SBM simulations with particular focus on a FRET-based glucose sensor. This sensor comprises two fluorescent proteins connected by linkers to a glucose sensing protein. Structures and parametrization of proteins and linkers in the model are derived. Then, I describe a procedure to merge the protein and linker structures and select starting conformations using experimental SAXS data. Finally, a simulation protocol for the sensor is given.

Chapter 5 presents the results for simulations of different dye-labeled proteins and the glucose sensor. Using test simulations, I demonstrate how they can yield new insights into the systems' dynamics by distance and orientation distributions. The influence of different dye pairs and labeling positions is analyzed and the simulation method is validated through direct quantitative comparison to experimental data. I demonstrate how simulations improve the interpretation of experimental data of the protein ClyA, including both two-color and three-color FRET measurements. Chapter 5 also compares the presented simulation method with descriptions of FRET by the accessible volume approach and analytical polymer models, being consistent with both in their respective area of applicability. Additional investigations consider the interplay of FRET and SAXS, in particular how FRET dyes affect SAXS measurements and whether both techniques measure the same quantities with respect to the radius of gyration of the studied protein. I also show that the simulation method is suitable to obtain quantitative parameters inaccessible in experiments such as diffusion parameters of dyes or rotational correlation times of restricted fluorescent proteins. Finally, I investigate differences between glucose sensor variants to study the underlying mechanisms contributing to the function of a highly sensitive sensor.

Chapter 6 discusses and summarizes the obtained results. An outlook on further studies and applications of this method is given.

2
Theoretical Background

This chapter introduces the theoretical background underlying this work. As this work studies proteins and their dynamics, at first Sec. 2.1 gives a short introduction to proteins. It discusses protein structures, essential to their function, and energy landscape theory, as one of the basic theories of protein folding. The latter is employed in the formulation of structure-based models used for the simulations in this work (see also Sec. 3.2). I study systems used for Förster resonance energy transfer (FRET) measurements, which are introduced in Sec. 2.2. Two measures to characterize the motions in the simulated systems are fluorescence anisotropy and diffusion. Both serve as tools to analyze the dynamics in the simulations and are described in Secs. 2.3 and 2.4, respectively. Sec. 2.5 describes small-angle X-ray scattering (SAXS) measurements which can be used to coarsely derive the shape of molecules. I use data from this technique as input for the biosensor simulations and investigate the interplay between FRET and SAXS in Chapter 5. Finally, to compare the developed approach to other analysis methods, Sec. 2.6 introduces the analytical polymer model which is applied to investigate unfolded and intrinsically disordered proteins.

2.1 Proteins

Proteins are large macromolecules and fundamental building blocks of life. They are essential for various biological functions as, e. g., DNA replication, molecular transport (e. g. hemoglobin which transports oxygen in the human body), cell motion, catalysis, and sensing of semiochemicals. Each protein consists of a specific sequence of amino acids, which determines its three-dimensional structure and thus function.

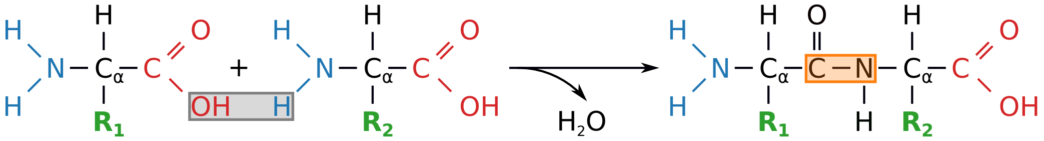
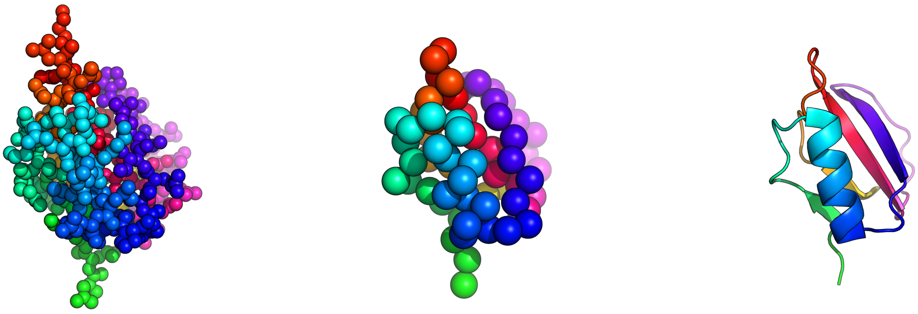


Figure 2.1: Two amino acids with C_α-atoms, carboxyl groups (red), amino groups (blue) and side chains R_1 and R_2 (green). The carboxyl group of the first amino acid reacts with the amino group of the second amino acid and forms a peptide bond (orange) with expulsion of a water molecule.



(a) All-atom representation. **(b)** C_α representation. **(c)** Cartoon representation.

Figure 2.2: The protein CI-2 (chymotrypsin inhibitor 2 [34]) in different representations. The sequences are color-coded from N-terminus (green) over blue and red to C-terminus (yellow). The all-atom representation includes all non-hydrogen atoms and the C_α representation only the C_α-atom of each residue. The cartoon representation depicts the secondary structure motifs of CI-2, i. e. an α-helix and the β-sheets.

2.1.1 Protein Structure

Proteins consist out of one or more chains of amino acids, each linked to one another by peptide bonds shown in Fig. 2.1. Each amino acid comprises a C_α-atom, a carboxyl group, an amino group, and a side chain, which is specific for each amino acid and determines its characteristics (see Fig. 2.1). In the genetic code, twenty different amino acids are encoded. The sequence of amino acids of a protein is referred to as primary structure, forming a polypeptide backbone with different side chains. The local structural motifs formed by this chain are referred to as secondary structure. Common motifs of secondary structure are the α-helix and the β-sheet, which are shown in the cartoon representation in Fig. 2.2c. The tertiary structure refers to the overall fold of a protein. If a protein contains several individual chains, their mutual arrangement is named quaternary structure.

A fundamental principle in biophysics is the paradigm of structure and function, stating that the three-dimensional native structure of a protein determines its function. Accordingly, determination of this native protein structure and understanding the process of protein folding from a disordered random coil to a functional biomolecule are of major interest in this research field.

Anfinsen's dogma postulates that protein structure is, at least for small globular proteins, exclusively dependent on the protein's sequence [35]. It also states that the native structure represents the free energy minimum and is unique, stable and kinetically accessible during folding. Irrespective of the loss of entropy, the native structure is stabilized by a shielding of hydrophobic amino acids from the surrounding solvent, additional salt bridges, hydrogen bonds, and disulfide bonds between cysteine residues, an amino acid containing sulfur. However, the conformational space for a protein chain is considerable. Levinthal's paradox states that the time needed for a protein to sequentially sample every possible conformation up to finding the minimum energy in the native state would be longer than the age of the universe [36]. Albeit, proteins are known to fold on the time scales of milliseconds to seconds [5, 37, 38]. This yields the question how proteins are able to "find" their native state so fast.

One approach to resolve this paradox is the theory of folding pathways with well-defined intermediate states between folded and unfolded conformations [39]. A different theory, named energy landscape theory [30, 33, 40], takes into account multiple folding routes and the ensemble character of conformations. It is the foundation for the computational model I use in this work.

2.1.2 Energy Landscape Theory

The free energy of a structural conformation as a function of its degrees of freedom is named energy landscape [30, 33, 40]. Every conformation corresponds to one point in this high-dimensional space. Randomly chosen amino acid sequences are likely to have competing energy contributions and therefore frustrated interactions. The frustration can lead to kinetic traps which prohibit proper folding [33]. According to the paradigm of structure and function, incorrect folding leads to malfunction. Consequently, a random amino acid sequence would be strongly disfavored by natural selection. The *principle of minimal frustration* states that evolution favors proteins with robust folding and thus minimized the frustration of interactions [30–33]. This can be described quantitatively using spin glass theory [30, 32] and leads to a "funneled" energy landscape, which is biased towards the native structure and enables fast folding (see Fig. 2.3). In this depiction, energy gains cause loss in conformational entropy as, e. g., stabilizing bonds narrow the conformational options.

A perfectly smooth unfrustrated funneled energy landscape would only include interactions stabilizing the native structure [33]. Consequently, a fully unfrustrated protein can be described only by its interactions in the native state [41]. This idea

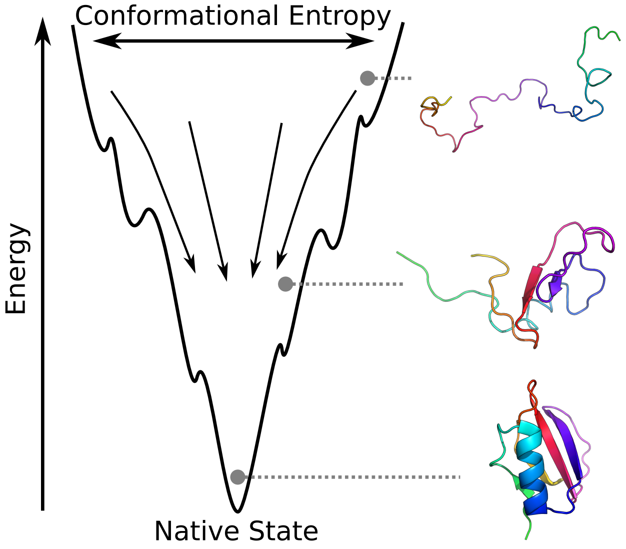


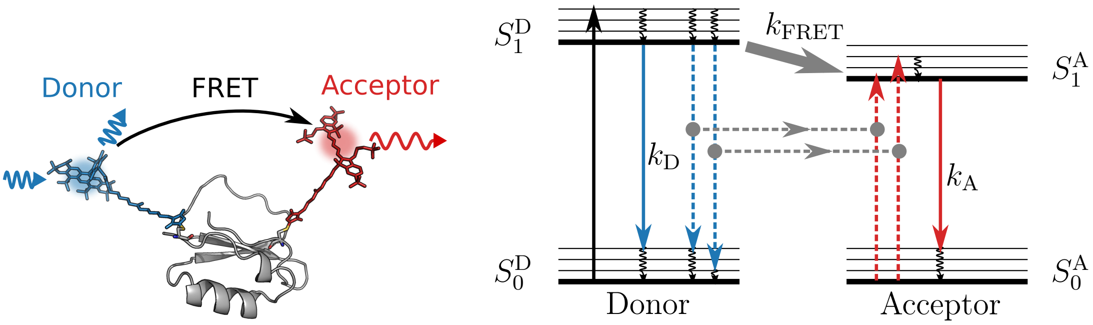
Figure 2.3: Scheme of a funneled energy landscape. The y-axis shows the energy and the width of the funnel represents the conformational entropy. Unfolded proteins have high energies and a large conformational space, whereas the native ground state occupies the energetic minimum with low conformational entropy.

is employed in the formulation of native structure-based or Gō-type models which are described in more detail in Sec. 3.2.

2.2 Förster Resonance Energy Transfer

Measuring Förster resonance energy transfer (FRET) [43] is a powerful experimental tool to get insight into dynamics and function of biomolecules. FRET refers to a non-radiative energy transfer between a donor and an acceptor fluorophore (see Fig. 2.4a). Due to its strong distance dependency in the nanometer range it is often utilized as “spectroscopic ruler” [1]. It is used in experiments to observe different protein conformations, conformational changes, or interactions between different molecules. For example, two residues of a protein are labeled with donor and acceptor fluorophores and folding transitions are observed directly through distance changes between donor and acceptor. Popular FRET applications are the observation of folding intermediates [6, 7], structure determination of biomolecules [42, 44, 45], and investigation of unfolded or intrinsically disordered proteins [2–4, 46].

Förster’s theory [43] considers the “weak coupling regime” of the fluorophores, which neglects perturbation on the energy levels due to non-coulombic interactions. The energy transfer is described as dipole-dipole coupling between the transition dipole moments of donor and acceptor. The different energy levels are shown



(a) FRET principle

(b) Jablonski diagram

Figure 2.4: (a) Principle of FRET with donor (blue) excitation and energy transfer to the acceptor (red). The blue and red wavy arrows depict the incident and emitted photons. (b) Simplified Jablonski diagram with electronic states for a donor and an acceptor fluorophore. S_0^{D} and S_0^{A} denote the ground singlet states, S_1^{D} and S_1^{A} the lowest excited singlet states of donor and acceptor, respectively. In the left, donor excitation (black arrow), de-excitation in absence of an acceptor (blue arrow) with rate k_{D}, and de-excitation via FRET (blue dashed arrows) with rate k_{FRET} are shown. In the right, acceptor excitation via FRET (red dashed arrows) and de-excitation (red arrow) with rate k_{A} are shown. Vibrational relaxations are shown as wavy black arrows. A similar depiction can be found in [42].

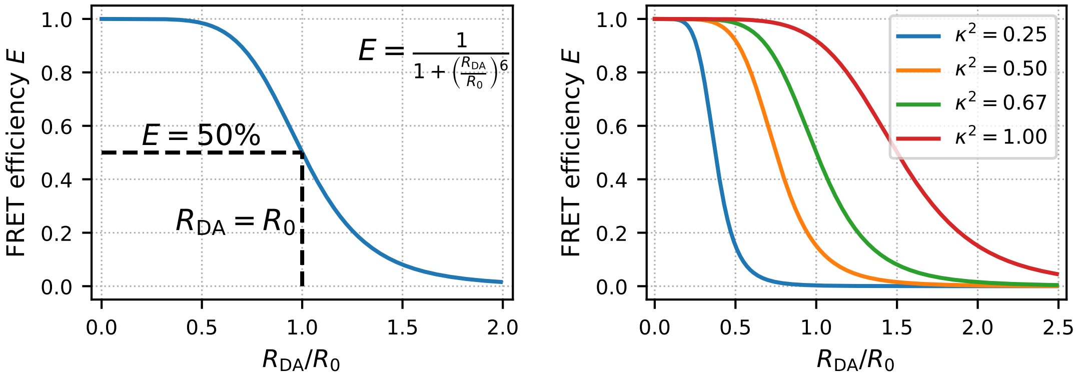
in a Jablonski diagram in Fig. 2.4b. The donor is excited by absorption of incident monochromatic light into its singlet state S_1^{D} and relaxes into the vibronic ground state. In absence of the acceptor the donor de-excites with a rate of k_{D} to its ground singlet state S_0^{D}. In presence of an acceptor it can also de-excite via FRET and transfer its energy to the acceptor with a rate of k_{FRET}. The acceptor then also de-excites with a rate of k_{A}. The given de-excitation rates for donor and acceptor include fluorescence decay via photon emission, and thermal de-excitation, where energy is dissipated as heat to the solvent. De-excitation to the ground state often occurs to a vibrationally excited substate of the ground state and then reaches thermal equilibrium [47].

At small donor acceptor distances ($< 2\,\mathrm{nm}$), the competing Dexter energy transfer, namely an electron exchange between donor and acceptor, comes into play. As the inter-fluorophore distances in the studied systems are typically larger, the Dexter energy transfer is neglected in the following.

2.2.1 FRET Efficiency

FRET experiments do not provide direct access to the distances between the fluorophores. Instead, the FRET efficiency E is measured, which is defined by [42]:

$$E = \frac{k_{\mathrm{FRET}}}{k_{\mathrm{D}} + k_{\mathrm{FRET}}}\,. \tag{2.1}$$



(a) Dependency of FRET efficiency E on donor acceptor distance R_{DA}. The Förster radius R_0 is defined as the distance corresponding to an efficiency of 50%. The highest sensitivity of FRET is in the range close to R_0.

(b) Dependency of FRET efficiency E on donor acceptor distance R_{DA} for different mutual orientations represented by the orientation factor κ^2. A value of $\kappa^2 = 2/3$ (green) corresponds to an isotropic average and is used to calculate R_0.

Figure 2.5: (a) Distance and (b) orientation dependency of FRET efficiency.

From this, the distance dependency of the efficiency can be derived [42]:

$$E = \frac{1}{1 + \left(\frac{R_{DA}}{R_0}\right)^6} , \tag{2.2}$$

where R_{DA} is the distance between donor and acceptor and R_0 is the Förster radius. In Fig. 2.5a the strong distance dependency of the FRET efficiency is displayed. The Förster radius R_0 is defined as the distance corresponding to an efficiency of 50%, i. e. de-excitation via FRET and de-excitation via other paths are equally probable. FRET efficiency reaches its highest sensitivity to the inter-fluorophore distance in proximity to R_0. The Förster radius mainly depends on the spectroscopic properties of the fluorophore pair used and the relative orientation of their transition dipole moments. It is given by [42]:

$$R_0^6 = \frac{9(\ln 10)\kappa^2 Q_D J(\lambda)}{128\pi^5 n^4 N_A} , \tag{2.3}$$

where Q_D is the fluorescence quantum yield of the donor in absence of the acceptor, $J(\lambda)$ the spectral overlap integral of donor emission and acceptor absorption spectrum, n the refractive index of the medium, and N_A the Avogadro constant. The orientation factor κ^2 describes the relative orientation of the transition dipole moments for emission of donor ($\boldsymbol{\mu}_D$) and for absorption of acceptor ($\boldsymbol{\mu}_A$) (see Fig. 2.6) and is given by [42]:

$$\kappa^2 = (\sin\theta_D \sin\theta_A \cos\phi - 2\cos\theta_D \cos\theta_A)^2 . \tag{2.4}$$

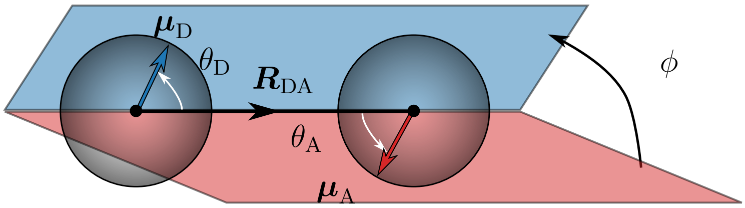

$\boldsymbol{\mu}_\mathrm{D}$
θ_D
$\boldsymbol{R}_\mathrm{DA}$
θ_A
$\boldsymbol{\mu}_\mathrm{A}$
ϕ

Figure 2.6: Parameters for the definition of the orientation factor κ^2. The centers of donor and acceptor (black dots) are connected by the vector $\boldsymbol{R}_\mathrm{DA}$ (black arrow). The transition dipole moment of donor (blue arrow) and acceptor (red arrow) are given by $\boldsymbol{\mu}_\mathrm{D}$ and $\boldsymbol{\mu}_\mathrm{A}$, respectively. The angle between $\boldsymbol{R}_\mathrm{DA}$ and $\boldsymbol{\mu}_\mathrm{D}$ ($\boldsymbol{\mu}_\mathrm{A}$) is referred to as θ_D (θ_A) and the angle between the two planes spanned by $\boldsymbol{\mu}_\mathrm{D}$ and $\boldsymbol{R}_\mathrm{DA}$ ($\boldsymbol{\mu}_\mathrm{A}$ and $\boldsymbol{R}_\mathrm{DA}$) is denoted ϕ. For the definition of centers and transition dipole moments of the fluorophores in this work, see Sec. 3.3.

Here, θ_D and θ_A are the angles between $\boldsymbol{\mu}_\mathrm{D}$ and $\boldsymbol{\mu}_\mathrm{A}$, and the vector $\boldsymbol{R}_\mathrm{DA}$ between the two centers of the fluorophores, respectively. The angle between the planes spanned by $\boldsymbol{\mu}_\mathrm{A}$, $\boldsymbol{R}_\mathrm{DA}$ and $\boldsymbol{\mu}_\mathrm{D}$, $\boldsymbol{R}_\mathrm{DA}$, respectively, is denoted as ϕ (see Fig. 2.6). The dependency of the FRET efficiency on κ^2 is shown in Fig. 2.5b.

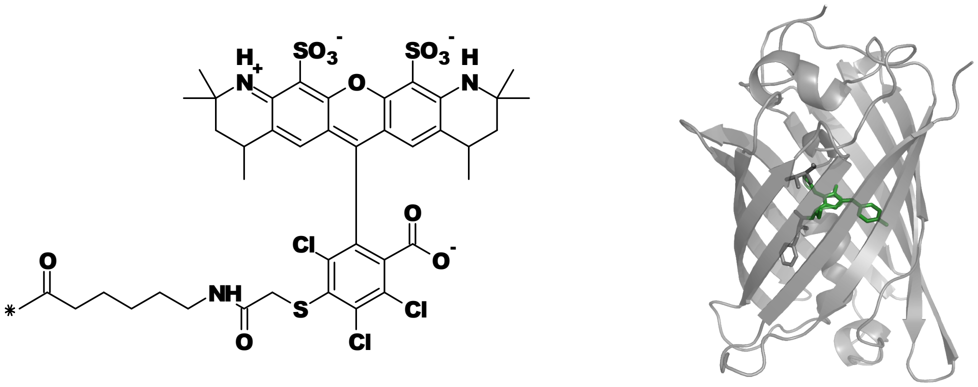
Presuming fast rotational diffusion and isotropic orientation of the fluorophores during the excited state lifetime of the donor, most studies assume a constant value of $\overline{\kappa^2} = 2/3$, which results from averaging over all possible rotations. It is referred to as the "isotropic dynamic averaging regime" [42]. This assumption has been questioned in several studies [19, 26–28]. Molecular dynamics simulations have found average $\overline{\kappa^2}$ values differing significantly from the isotropic value of $\overline{\kappa^2} = 2/3$ and also correlations between κ and R_DA despite the assumed independence [28]. The assumption of the "isotropic dynamic averaging regime" can be tested experimentally, e. g. by measuring time-dependent fluorescence anisotropy decays [23]. However, in the interpretation of FRET measurements it is difficult to account for sterically hindered or slow rotation as in studies with, for example, fluorescent proteins. The simulation method presented in this work directly provides the orientations of the fluorophores and allows to easily test the assumption for various systems.

Detailed derivations of the presented formula can be found in [43, 48].

2.2.2 FRET Fluorophores

Different kinds of fluorophore pairs are utilized for FRET measurements. Ideally, fluorophores for FRET should be bright (i. e. have high extinction coefficients and quantum yields), photostable, small, and water soluble [49]. Additionally, the donor and acceptor emission spectra should be well separated.

For these reasons small organic dye molecules as, e. g., the Alexa Fluor dyes [50] are often used in FRET studies. They are named roughly according to their ex-



(a) Alexa Fluor 546 dye [50].

(b) Green fluorescent protein (PDB: 1GFL [51]). The fluorophore inside the β-barrel (gray) is shown in green.

Figure 2.7: Different fluorophores used for FRET measurements. Examples for a small organic dye (left) and for a fluorescent protein (right) are shown.

citation maxima. As an example, the Alexa Fluor 546 dye is shown in Fig. 2.7a. Dyes are used with different linker lengths and are often bound via a maleimide to a protein residue mutated to cysteine (see also Fig. 4.1). As dyes are small and flexible, they have negligible influence on the dynamics of the molecule they are attached to and their rotational diffusion is usually fast enough to justify the isotropic averaging regime.

Despite having a lower photostability [49] and being of larger size than dyes, fluorescent proteins are also utilized for many FRET applications. They have the advantage that they can be genetically encoded and thus directly fused to the investigated protein. The primary fluorescent protein, the green fluorescent protein (GFP [52]), was discovered in the 1960s. In recent years, GFP was engineered to produce various mutants with different color spectra. It consists of a β-barrel comprising eleven β-strands with an α-helix containing the covalently bound fluorophore in the center (see Fig. 2.7b) [51]. When the protein is folded completely, induced specific cyclization reactions form the fluorophore from the tripeptide Ser65-Tyr66-Gly67. By altering the sequence of this tripeptide and the proximate amino acids in the β-barrel, color, intensity, and photostability of the fluorescent protein can be varied.

Another possible choice for FRET donors are semiconductor quantum dots, but due to their large size ($> 20\,\text{nm}$) their use is limited [49].

2.2.3 Two-Color FRET Experiments

In two-color FRET experiments, a donor and an acceptor fluorophore are attached to specific residues of biomolecules such as proteins. The FRET donor is excited by laser pulses and the resulting emitted photons are split by dichroic mirrors and detected by separate detection channels for donor and acceptor photons, respectively. In ensemble FRET measurements the observed FRET efficiency is averaged over an entire ensemble of structures and conformations. To access individual molecular states or time-resolved evolution of systems, single-molecule FRET (smFRET) is becoming more and more popular [42]. It can be performed with molecules immobilized on a surface or freely diffusing in buffer solution. In the latter case the molecules are observed by a confocal microscope during their diffusion time through the confocal volume, which is about $\Delta t \approx 1\,\mathrm{ms}$ [5]. The detected photons are collected in individual bursts, which are discarded when containing less than a specified minimal number of photons to reduce the influence of shot noise. After correcting the photon counts for crosstalk and background the FRET efficiency is calculated by [49]:

$$E = \frac{1}{1 + \gamma \frac{I_\mathrm{D}}{I_\mathrm{A}}} \,. \tag{2.5}$$

I_D and I_A are the corrected intensities of donor and acceptor, respectively. The correction factor γ accounts for the different quantum yields and detection efficiencies for donor and acceptor. From this efficiency, the inter-fluorophore distance can only be calculated approximately due to uncertainties in κ^2 and instrumental corrections [49]. Therefore, for comparison of experimental and simulated data the measured FRET efficiency is preferable compared to the derived inter-fluorophore distance.

FRET-based biosensors are one application of fluorescent proteins, where a sensing protein changes its conformation and the conformational change is translated to a change in the FRET signal. To quantify the quality of FRET sensors, experiments rather measure the FRET intensity ratio

$$R = \frac{I_\mathrm{A}}{I_\mathrm{D}} \tag{2.6}$$

instead of the FRET efficiency for different conformations of the sensor. The change between minimal and maximal ratio, ΔR, gives a characteristic parameter for the sensitivity of a sensor.

2.2.4 Three-Color FRET Experiments

An extension of regular smFRET with two fluorophores are FRET measurements using three or more fluorophores. Three-color FRET measurements [53, 54] en-

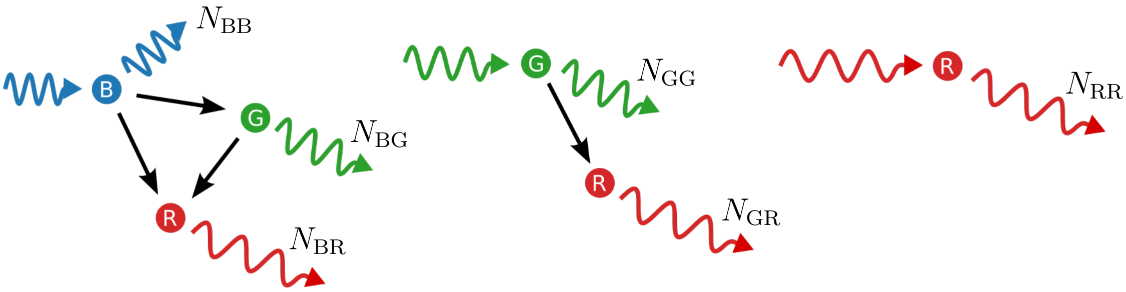


Figure 2.8: Schematic depiction of three-color FRET, represented by a blue (B), a green (G), and a red (R) fluorophore. Upon excitation the blue donor fluorophore (left) de-excites or transfers energy via FRET to one of the acceptors (black arrows). The green fluorophore then de-excites or transfers the energy to the red fluorophore. The red fluorophore then de-excites. A similar scheme for direct excitation of the green fluorophore (middle) and of the red fluorophore (right) are shown alongside. In addition, the photon counts N_{MN} with M denoting the directly excited fluorophore and N referring to the fluorophore whose photons are counted are depicted.

able observing changes in more than one distance simultaneously, e. g. to show correlated movements [55], and provide a more detailed picture of the dynamics in complex systems. Three-color FRET is widely used [56, 57], e. g. for studying intrinsically disordered proteins [58] or by employing a FRET cascade to extend the range of FRET [59]. Also some attempts on four-color FRET have been made which could be described by an analog procedure [60].

Interpretation of these measurements is more complex than two-color smFRET. Several more paths have to be considered as illustrated in Fig. 2.8. Due to many unknown rates, these experiments need additional information. Therefore, the experiments with three small dyes considered here use three successive laser pulses to directly excite the donor, the first, and the second acceptor. Here they are exemplary referred to as blue (B), green (G), and red (R) fluorophore, respectively. The resulting photons are counted after each excitation, yielding photon counts N_{MN} with M being the directly excited fluorophore and N being the fluorophore corresponding to the photons counted. Instead of a single efficiency, the quantities analyzed are the photon count rates F_{MN}:

$$F_{\mathrm{BB}} = \frac{N_{\mathrm{BB}}}{N_{\mathrm{BB}} + N_{\mathrm{BG}} + N_{\mathrm{BR}}}, \tag{2.7}$$

$$F_{\mathrm{BG}} = \frac{N_{\mathrm{BG}}}{N_{\mathrm{BB}} + N_{\mathrm{BG}} + N_{\mathrm{BR}}}, \tag{2.8}$$

$$F_{\mathrm{BR}} = \frac{N_{\mathrm{BR}}}{N_{\mathrm{BB}} + N_{\mathrm{BG}} + N_{\mathrm{BR}}}, \quad \text{and} \tag{2.9}$$

$$F_{\mathrm{GR}} = \frac{N_{\mathrm{GR}}}{N_{\mathrm{GG}} + N_{\mathrm{GR}}}. \tag{2.10}$$

In experiments, the photon counts N_{MN} are already corrected for crosstalk, detection efficiencies, quantum yields, and background. The photon count after excitation of the red fluorophore N_{RR} is used in experiments to ensure selection of only bursts from those systems where all fluorophores are present.

2.3 Fluorescence Anisotropy

When illuminating a sample of randomly oriented fluorophores with polarized light, the fluorophores with transition dipole moment parallel to the electric field of the incident light are preferentially excited. Consequentially, the excited-state population is partially oriented. Due to rotational diffusion the sample depolarizes over time with a certain time constant. This time-dependent fluorescent anisotropy $r(t)$ can be measured experimentally and is defined as [47]:

$$r(t) = \frac{I_{\parallel}(t) - I_{\perp}(t)}{I_{\parallel}(t) + 2I_{\perp}(t)} \,. \tag{2.11}$$

$I_{\parallel}(t)$ and $I_{\perp}(t)$ denote the time-dependent fluorescence intensity parallel and perpendicular to the polarization of the incident light, respectively. The fluorescence decay is characterized by the rotational correlation time τ_{rot}, giving a measure of flexibility and rotational speed of the fluorophores.

The measured fluorescence anisotropy decay can be compared to the calculation of the time-dependent anisotropy from atomic coordinates in simulations by using the normalized absorption and emission dipole vectors $\widehat{\boldsymbol{\mu}_{\mathrm{a}}}$ and $\widehat{\boldsymbol{\mu}_{\mathrm{e}}}$. Here, I assume the absorption and emission dipole moment vectors of a fluorophore to be collinear ($\widehat{\boldsymbol{\mu}_{\mathrm{a}}} = \widehat{\boldsymbol{\mu}_{\mathrm{e}}} = \widehat{\boldsymbol{\mu}}$). The fluorescence anisotropy is then calculated as [61]:

$$r(t) = r_0 \left\langle P_2 \left[\widehat{\boldsymbol{\mu}}(s) \cdot \widehat{\boldsymbol{\mu}}(s+t)\right]\right\rangle_s \,, \tag{2.12}$$

where r_0 is the fundamental anisotropy and P_2 the second-order Legendre polynomial given by $P_2(x) = \frac{1}{2}(3x^2 - 1)$. With the assumption of collinear transition dipole moments, the fundamental anisotropy is given by $r_0 = 0.4$, close to the experimentally measured value [62].

As the rotational diffusion of both fluorophores and proteins contribute, the anisotropy decay is generally described as a double-exponential function [23, 61, 63]. In the systems with small dyes in this work, the proteins are significantly larger than the dye molecules. As a result, the expected rotational correlation times of the proteins are one order of magnitude larger than the rotational correlation times of the dyes. Thus, for the experimentally measured rotational correlation times the protein motions can be neglected. At the same time this prevents measuring rotational correlation times of freely diffusing systems with fluorescent proteins.

In the analyses of the simulations of the systems with small dyes as well as the systems with fluorescent proteins, I focus on the rotational motion of the fluorophores in the inertial system of the protein. Hence, I find using only one exponential function sufficient and fit with the relation [47]

$$r(t) = r_0 \exp\left[-\frac{t}{\tau_{\text{rot}}}\right] . \tag{2.13}$$

In case of a spatially restricted rotation of the fluorophore, e. g. when using fluorescent proteins, the "wobbling-in-a-cone" model can be applied [64]. It assumes that the transition dipole moments of the fluorophore moves freely in a cone. Then the anisotropy decay can be described with

$$r(t) = r_0 \left[(1-A)\exp\left[-\frac{t}{\tau_{\text{rot}}}\right] + A\right] , \tag{2.14}$$

where A defines a measure of the rotational restriction and is related to the angle of the cone.

The fluorescence anisotropy $r(t)$ allows a direct comparison of experimental observations with simulations. It will serve as a conversion factor between the time scale in simulations in this work to the physical time scale (see Sec. 4.5).

2.4 Diffusion

Besides the rotational correlation time, the dynamic behavior of fluorophores can also be described by diffusion in the accessible volume. The diffusion constant is usually not directly accessible in experiments, but there are several methods to calculate it from simulations given velocities and positions of the respective molecule are known.

One way to calculate the diffusion constant D of a molecule in three dimensions is via the velocity autocorrelation function

$$C_{\mathbf{v}}(\tau) = \langle \mathbf{v}_i(\tau) \cdot \mathbf{v}_i(0) \rangle_i , \tag{2.15}$$

where $\mathbf{v}_i$ is the velocity and i refers to the averaging over the ensemble which is here taken as different starting times in the simulation. The diffusion constant is then given by the Green-Kubo relation [65]:

$$D = \frac{1}{3}\int_0^{\infty} C_{\mathbf{v}}(t)\, dt. \tag{2.16}$$

A second approach uses the mean square displacement of positions $\mathbf{r}_i(t)$ and the Einstein relation for three dimensions [65]:

$$\lim_{t\to\infty} \left\langle \|\mathbf{r}_i(t) - \mathbf{r}_i(0)\|^2 \right\rangle_i = 6Dt\,. \tag{2.17}$$

For small time scales the diffusion coefficient for fluorophores can be calculated with Eq. (2.17). For larger times the diffusion is confined by the accessible volume of the fluorophore due to the restriction of the linker bound to the protein. This confined diffusion is described by [66, 67]:

$$\left\langle r^2(t) \right\rangle = \left\langle r_c \right\rangle^2 \cdot \left[1 - A_1 \cdot \exp\left(-A_2 \cdot \frac{6Dt}{\left\langle r_c \right\rangle^2}\right)\right],, \tag{2.18}$$

where $\langle r_c \rangle^2$ represents the size of the accessible volume and A_1 and A_2 are fit parameters.

The diffusion constant D and the value of $\langle r_c \rangle^2$ derived from simulations can then be employed in further calculations.

2.5 Small-Angle X-Ray Scattering

Small-angle X-ray scattering (SAXS) is an experimental method to measure the sizes and shapes of molecules and is widely used for studying proteins in solution [68, 69]. In contrast to nuclear magnetic resonance (NMR), SAXS is not limited by protein size and is experimentally less extensive than structure determination via X-ray crystallography as it does not need laborious crystallization of the molecule. Moreover, it is used to study the behavior of unfolded or intrinsically disordered proteins [70].

In SAXS, samples of molecules in solution are exposed to X-rays and the scattered light is recorded by a detector. The scattering intensity I is measured as a function of the scattering angle θ and the momentum transfer $q = 4\pi \sin\theta/\lambda$, respectively, where λ is the wavelength of the incident light. SAXS measures the averaged scattering intensity over the entire ensemble and all orientations of the molecules. The scattering intensity curve of the pure solvent is then subtracted from the curve of the molecules in solution.

In SAXS theory, the spherically averaged intensity is described as a sum of elementary scatterers, e. g. atoms or amino acids. It is calculated with the Debye formula [71]:

$$I(q) = \sum_i \sum_j f_i(q) f_j(q) \frac{\sin(q r_{ij})}{q r_{ij}}\,, \tag{2.19}$$

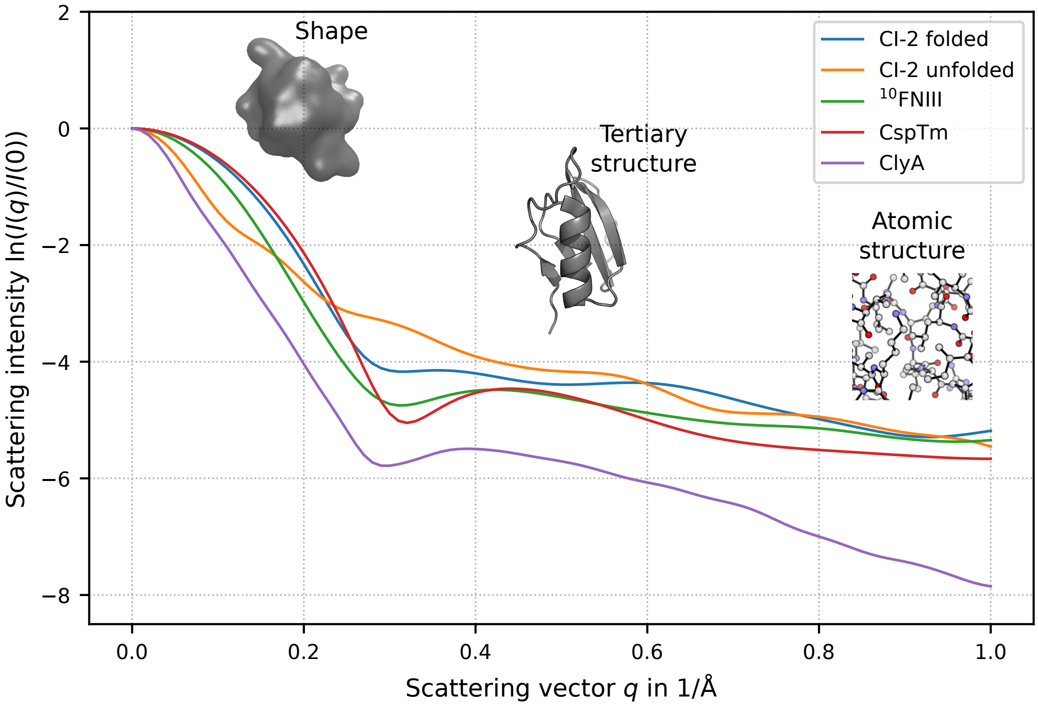


Figure 2.9: SAXS scattering curves for different proteins (see Sec. 4.3). The SAXS intensity curves are characteristic for specific sizes and shapes. The different ranges of the scattering vector q give access to different structural features. Small q values reflect the overall shape of a structure, whereas medium q values yield information on the level of tertiary structure. High q values in principle provide access to atomic structures.

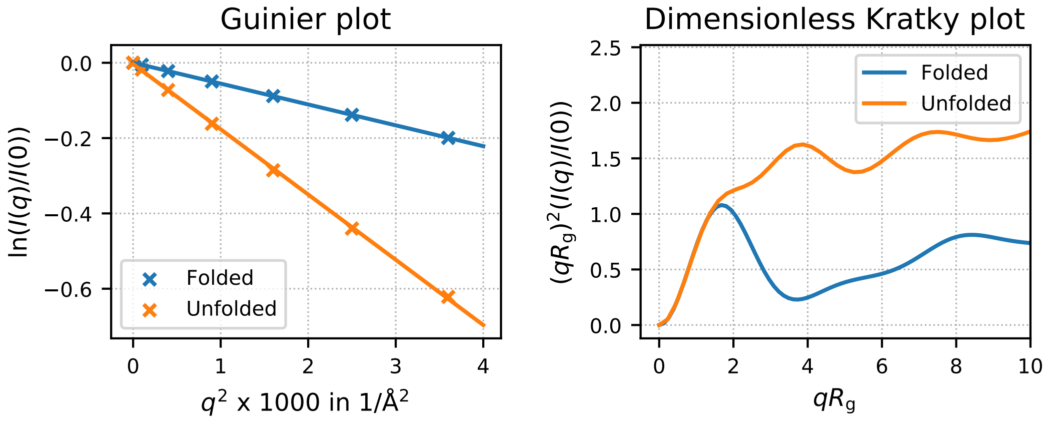


Figure 2.10: Guinier plot (left) and dimensionless Kratky plot (right) for CI-2 in folded and unfolded states. From the Guinier plot in the region of small q the radius of gyration R_g can be approximated. The dimensionless Kratky plot gives information about the configuration of the protein. For example, the rise to a plateau indicates a random coil (unfolded CI-2, orange), whereas a distinct peak indicates a globular structure (folded CI-2, blue).

where r_{ij} is the distance between two scattering particles and f_i is the form factor of the scattering particle i.

The different parts of the resulting one-dimensional SAXS intensity curve give information about different structural features, as can be seen in Fig. 2.9 for different proteins. In the range of small q values the intensity $I(q)$ can be described by the Guinier approximation [72]:

$$I(q) = I(0) \exp\left[-\frac{q^2 R_\mathrm{g}^2}{3}\right] , \tag{2.20}$$

with the radius of gyration R_g, a measure of the protein's size (see also Sec. B). With this relation, the radius of gyration can be extracted as the slope of the curve in a Guinier plot (see Fig. 2.10). The Guinier approximation is only valid in a range of $qR_\mathrm{g} < 1.3$ for globular proteins [69] and in an even smaller range for elongated structures.

The intensities at higher q values yield information about the molecule's shape and are referred to as the “power-law regime”, where the scattering can be described as:

$$I(q) \propto q^{-d_\mathrm{f}} , \tag{2.21}$$

where d_f denotes the fractal dimension. For folded macromolecules Eq. (2.21) results in Porod's law [73], where $d_\mathrm{f} = 4$. For other structures as, e. g., unfolded polymers, d_f can adopt a wide range of values. The approximation of Eq. (2.21) only holds in a part of the scattering curve as it breaks down at higher q values, where information about atomic resolution becomes significant [69].

The Kratky plot ($q^2 I(q)$ as function of q) provides an excellent tool to analyze the folding of molecules [69]. It is shown in its dimensionless form in Fig. 2.10. A rise to a plateau indicates a random coil, while a globular structure results in a distinct peak.

For comparison of theoretically calculated and experimentally measured SAXS profiles, the degree of agreement is commonly given by the value of χ^2 after fitting. Here, I use the fitting method implemented in `CRYSOL` [74], where χ^2 is defined as:

$$\chi^2 = \frac{1}{N_q} \sum_{i=1}^{N_q} \left[\frac{I_\mathrm{exp}(q_i) - c \cdot I(q_i)}{\sigma(q_i)} \right]^2 , \tag{2.22}$$

and the constant c is given by:

$$c = \left[\sum_{i=1}^{N_q} \frac{I_\mathrm{exp}(q_i) \cdot I(q_i)}{\sigma(q_i)^2} \right] \left[\sum_{i=1}^{N_q} \frac{I(q_i)^2}{\sigma(q_i)^2} \right]^{-1} . \tag{2.23}$$

N_q denotes the number of experimental points q_i, $I_\mathrm{exp}(q_i)$ the experimental scattering intensities, $\sigma(q_i)$ the experimental errors, and $I(q_i)$ the theoretical intensities. The absolute value of χ^2 is difficult to interpret, but different profiles can be compared by their χ^2 values to a reference profile to, e. g., find the theoretical profile best fitting to experimental measurements.

2.6 Polymer Model

Proteins in the unfolded states and intrinsically disordered proteins are often approximated as polymer chains. At high temperatures the model I use describes an excluded volume polymer chain [75], which follows

$$\langle r_{ij}^2 \rangle^{1/2} = C \cdot N_{ij}^{\nu} . \tag{2.24}$$

Here, apart from the constant C, r_{ij} is the spatial distance and N_{ij} the sequence distance between two chain elements i and j, respectively. The length scaling exponent ν is characteristic for different polymer models and expected to be $\nu = 3/5$ for excluded volume polymer chains [75]. It is related to the fractal dimension described in Sec. 2.5 via $\nu = 1/d_\mathrm{f}$ [76].

Polymer models are further employed to describe unfolded proteins with dyes attached and interpret the corresponding FRET experiments [77, 78]. It is often

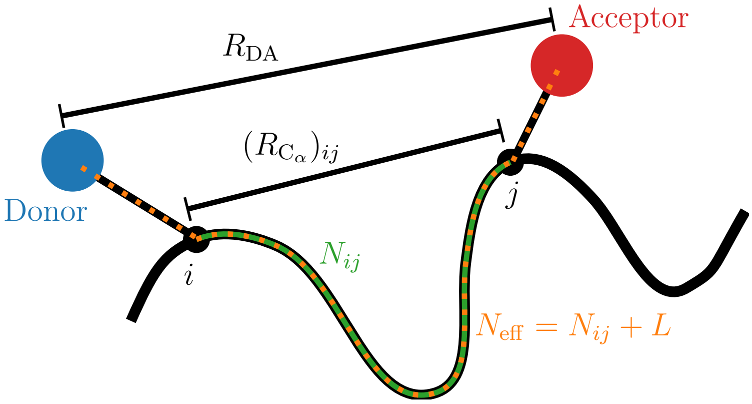

R_{DA}
Acceptor
$(R_{\mathrm{C}_\alpha})_{ij}$
Donor
j
i
N_{ij}
$N_{\mathrm{eff}} = N_{ij} + L$

Figure 2.11: Schematic depiction of an unfolded protein (black line) with donor (blue) and acceptor (red) attached. The sequence distance $N_{ij} = |i - j|$ between donor and acceptor, attached to residues i and j is shown in green. The spatial distances between donor and acceptor centers, R_{DA}, and between the respective C_α-atoms (black circles), $(R_{\mathrm{C}_\alpha})_{ij}$, are shown. The resulting effective distance N_{eff} is depicted in orange (dashed line).

assumed that FRET experiments measure the distance between the C_α-atoms of the labeled residues (in the following referred to as C_α distance), neglecting the contributions of the dyes' linkers. However, in careful analyses, the linker of donor and acceptor can be described as extension of the chain with a certain additional length L. This results in an effective sequence distance $N_{\mathrm{eff}} = N_{ij} + L$ (see Fig. 2.11). Then, a correction factor m can be calculated to relate the distance of the dyes' centers R_{DA} to the distance of the corresponding C_α-atoms $(R_{\mathrm{C}_\alpha})_{ij}$, where i, j are the residue indices:

$$m = \left(\frac{N_{ij} + L}{N_{ij}}\right)^{\nu}, \quad \text{for } N_{ij} > 0. \tag{2.25}$$

For a theoretical sequence separation of $N_{ij} = 0$, the correction factor is not defined. The respective quantities are illustrated in Fig. 2.11.

2.7 Summary

This chapter showed how proteins and their dynamics can be described by energy landscape theory. The latter is employed in the formulation of structure-based models, which in turn are used for the simulations in this work. Furthermore, it presented the experimental technique utilizing FRET as a “spectroscopic ruler” to study structures and dynamics of proteins. FRET is highly dependent on distances and mutual orientations between two fluorophores which are not directly accessible in experiments. Simulations provide a more detailed insight in biomolecular systems labeled with FRET fluorophores, as a) single molecule two- and three-color FRET with small organic dyes, and b) systems with fluorescent proteins.

In addition, I introduced SAXS as an experimental technique to get further information about the shapes of systems. The interplay of FRET and SAXS will be studied in Sec. 5.7 and SAXS data will be utilized to select suitable structures for simulation of a glucose sensor in Sec. 5.9.1.

3
Computational Methods

This chapter introduces the basics of the computational methods used in this work. The simulations I perform are based on the principles of molecular dynamics (MD), which are introduced in Sec. 3.1. MD simulations model a system's dynamics with Newton's equations of motion and a given force field. Here, I simulate proteins in solution, so a treatment of the interaction between protein and solution is necessary. This interaction is employed via Langevin dynamics, which implicitly accounts for solvent friction and random collisions with solvent molecules, as described in Sec. 3.1.1. The typical composition of an MD force field is described in more detail in Sec. 3.1.2.

For the study of large conformational transitions as, e. g., in unfolded protein ensembles, or large systems, regular MD simulations are infeasible due to prohibitive computational costs. Therefore I use the computationally more efficient native structure-based models (SBMs) presented in Sec. 3.2 throughout this work. In the SBM potential (see Sec. 3.2.1) native contacts play a crucial role, so they are discussed further in Sec. 3.2.2. Also, time and temperature units in SBMs, which due to their accelerated dynamics do not directly relate to physical units, are discussed in Sec. 3.2.3. Finally, Sec. 3.3 presents the Monte Carlo method to calculate photon statistics from simulated trajectories, necessary for comparison of the simulations to two-color and three-color FRET experiments.

3.1 Molecular Dynamics

Molecular dynamics (MD) is a simulation method to investigate systems on the atomic and molecular scale. It is based on classical mechanics, as quantum mechanical methods can not be feasibly utilized for molecules of larger size. The atomic coordinates of the simulated system are calculated in successive time steps in the order of magnitude of 1 fs. Dynamic properties can then be calculated by integration over time. The simulation is based on classical force fields represented by high-dimensional potentials $V(x_1, x_2, ..., x_N)$. For every simulation step the force is calculated by:

$$\boldsymbol{F} = -\nabla V(x_1, x_2, ..., x_N)\,. \tag{3.1}$$

The resulting system of Newton's equation of motion is solved for every atom, and the positions and velocities are updated in every time step. The coordinates of all atoms in the system are stored in trajectories, consisting of individual frames at discrete points in time, which can then be analyzed.

MD simulations are used for many applications in material science, chemical physics, and for simulation of biomolecules such as proteins. Recent applications of MD simulations are, e. g., the investigation of protein folding [79], refinement of experimental measurements [80], and the study of enzymatics of kinases involved in muscle function [81]. MD simulations have also been employed to investigate fluorescence anisotropy decay (see Sec. 2.3) [82, 83]. However, the simulations have been adjusted as the original parametrization had yielded a rotational correlation time deviating about a factor of three from experimental values [82]. This result shows that even extensively developed force fields have to be adjusted for specific applications.

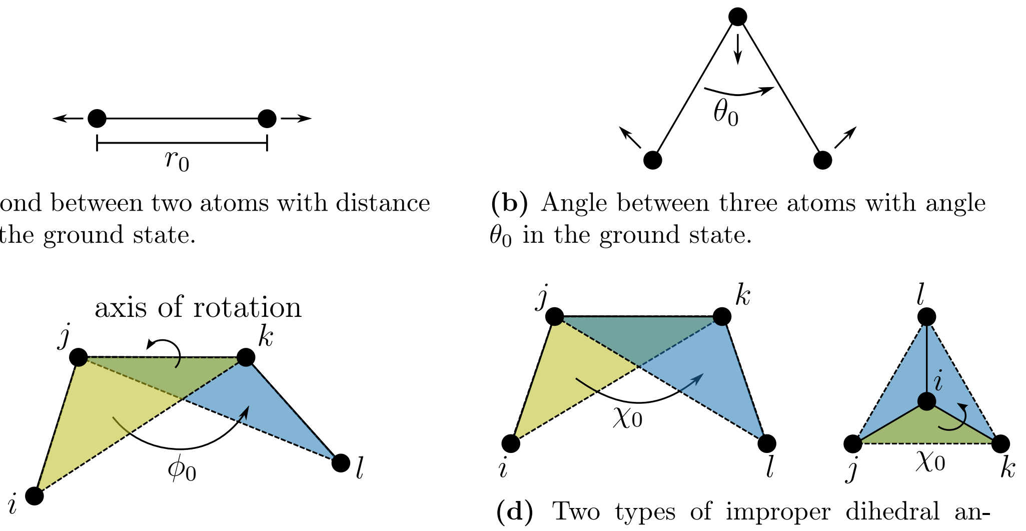
Still, MD simulations face several limitations [84]. The classical force fields do not account for quantum mechanical behavior and electronic motions are neglected. All electrons are considered to be in their ground state. Also, the force field parameters may be ambiguous. They are derived by quantum mechanical calculations and adjusted to empirical data, but there are several different force fields available which are adapted for different applications. Due to computational limitations, the coulombic interactions are cut off at a certain range and periodic boundary conditions have to be used. Both assumptions can lead to unphysical behavior.

One implementation of molecular dynamics is the `GROMACS` package [84], which is used for the simulations in this work.

3.1.1 Langevin Dynamics

Langevin dynamics [84] or stochastic dynamics refers to Newton's equation of motion with additional terms for friction and a random force. It allows for implicit modeling of solvent friction and of the random perturbations of the system by occasional collisions with solvent molecules. The Langevin equation is given by:

$$m_i \frac{\mathrm{d}^2\boldsymbol{r}_i(t)}{\mathrm{d}t^2} = \boldsymbol{F}_i(\boldsymbol{r}_i(t)) - \gamma_i m_i \frac{\mathrm{d}\boldsymbol{r}_i(t)}{\mathrm{d}t} + \boldsymbol{R}_i(t)\,, \tag{3.2}$$



(a) Bond between two atoms with distance r_0 in the ground state.

(b) Angle between three atoms with angle θ_0 in the ground state.

(c) Proper dihedral angle with angle ϕ_0 in the ground state. It is defined as the angle between the planes formed by the atoms i, j, k (yellow) and j, k, l (blue).

(d) Two types of improper dihedral angles with angle χ_0 in the ground state. The angle is defined as the angle between the planes formed by atoms i, j, k (yellow) and j, k, l (blue).

Figure 3.1: Four types of bonded interactions between atoms (black circles).

where m_i is the mass of the particle, and $\boldsymbol{r}_i(t)$ are the three-dimensional coordinates of the particle over time t. As no explicit solvent is present in this type of simulation, the friction constant γ_i implicitly models the solvent friction. The random force $\boldsymbol{R}_i(t)$ is a stationary Gaussian process, satisfying

$$\langle \boldsymbol{R}_i(0)\boldsymbol{R}_j(t)\rangle = 2m_i\gamma_i k_\mathrm{B}T\delta_{ij}\delta(t)\,. \tag{3.3}$$

Here, k_B is the Boltzmann constant, T the temperature, δ_{ij} the Kronecker delta, and $\delta(t)$ the Dirac delta. MD simulations directly provide the NVE ensemble, i. e., a constant number of particles, constant volume, and constant energy. As the calculation of quantities within the canonical ensemble (NVT) is required, the temperature has to be held constant. The temperature can be handled by explicit temperature coupling or, in Langevin dynamics, by implicit control of the temperature via the random force term in Eq. (3.3). To numerically integrate the differential equations in Eq. (3.2), `GROMACS` uses a third-order leap-frog integrator [85].

3.1.2 Force Fields

An essential part of an MD simulation is the underlying force field. A force field is a set of parameters for all bonded and non-bonded interactions present in the simulated system. The bonded interactions include all interactions between directly connected atoms, namely bonds, angles, and proper and improper dihedral angles. The different types of bonded interactions are shown in Fig. 3.1.

In the potential, the bonded interactions are represented by harmonic oscillators centered around the respective ground state. The potential terms are given by:

$$V_{\mathrm{b}} = K_{\mathrm{b}}(r - r_0)^2\,, \tag{3.4}$$

$$V_{\mathrm{a}} = K_{\mathrm{a}}(\theta - \theta_0)^2\,, \tag{3.5}$$

$$V_{\mathrm{d}} = \sum_n K_{\mathrm{d}}(1 - \cos(n\phi - \phi_0))\,, \quad \text{and} \tag{3.6}$$

$$V_{\mathrm{i}} = K_{\mathrm{i}}(\chi - \chi_0)^2\,, \tag{3.7}$$

for bonds with bond length r, angles θ, proper dihedral angles ϕ, and improper dihedral angles χ. r_0, θ_0, ϕ_0, and χ_0 are the respective values for the ground state. The periodic dihedral potential with multiplicity n allows for isomeric conformations. The corresponding force constants are K_{b}, K_{a}, K_{d}, and K_{i}. Improper dihedral angles are added to keep the involved atoms in a plane (e. g. in ring structures) and to prevent transition to unphysical isomers.

Non-bonded interactions are typically electrostatic interactions and Lennard-Jones interactions. The respective potentials are given by the following terms:

$$V_{\mathrm{LJ}} = K_{\mathrm{LJ}}\left[\left(\frac{\sigma_{ij}^0}{r_{ij}}\right)^{12} - 2\cdot\left(\frac{\sigma_{ij}^0}{r_{ij}}\right)^{6}\right] \tag{3.8}$$

$$V_{\mathrm{Coulomb}}(r_{ij}) = \frac{q_i q_j}{4\pi\epsilon_0\epsilon_r r_{ij}} \tag{3.9}$$

Here, σ_{ij}^0 is the radius for the excluded volume, and r_{ij} is the distance between two atoms i and j with charges q_i and q_j. ϵ_0 and ϵ_r are the electric constant and the dielectric constant, respectively. K_{LJ} denotes the force constant for the Lennard-Jones potential. All force constants are given by the chosen force field and can be different for different types of bonds, angles, and dihedral angles, as well as dependent on the atom types involved. Common force fields are, e. g., the AMBER (Assisted Model Building and Energy Refinement) [86, 87] or CHARMM (Chemistry at HARvard Macromolecular Mechanics) [88, 89] force fields implemented in GROMACS.

Reaching biologically relevant time scales in regular MD simulations involves huge computational costs, especially considering large systems. A simulation of a protein with explicit solvent requires a sufficiently large box of water molecules and simulation steps taking place on the femtosecond time scale. However, the interesting time scales for, e. g., protein folding are in the scale of micro- to milliseconds. On this account, regular MD simulations themselves are impractical for many applications. Therefore, there are many efforts to accelerate MD simulations by reducing the degrees of freedom of the systems in question. One possible attempt are coarse-grained models [90], which treat groups of atoms as single element or use simplified potentials.

3.2 Native Structure-Based Models

For my simulations I use the framework of native structure-based models (SBMs), also known as Gō-type models [91]. They are based on the principle of minimally frustrated energy landscapes for proteins (see also Sec. 2.1.2) [30–33]. SBMs include several simplifications in contrast to regular MD force fields. The main difference is that they only represent heavy atoms, therefore neglecting all hydrogen atoms. By using Langevin dynamics (see Sec. 3.1.1) the protein-solvent interaction is treated implicitly. Furthermore, they do not distinguish between different atom types and do not explicitly take electrostatic interactions into account. These simplifications drastically reduce the needed computational resources. SBMs are of course limited to unfrustrated systems with negligible non-native interactions [30–33, 91].

Despite these simplifications, SBM simulations show good agreement with experimental results. For example, they are able to reproduce transition state ensembles and "en-route" intermediates [92], while also yielding folding rates comparable to experimental measurements [93]. Due to their high computational efficiency they are applied to study a wide range of phenomena [94]. These are ranging from protein structure prediction [95], protein folding [93, 96], misfolding [97, 98], and conformational dynamics [99, 100] to large biomolecules as the ribosome [101] or RNA [102].

With their high computational efficiency, SBMs allow for simulations of several folding and unfolding transitions on regular desktop computers and still offer full flexibility for all parts of the system.

3.2.1 Structure-Based Potential

I use an SBM including all heavy atoms [29] as implemented in `eSBMTools` [103]. The simplified SBM potential has the following form [104]:

$$\begin{aligned} V_{\mathrm{SBM}} &= \sum_{\text{bonds}} K_{\mathrm{b}}(r-r_0)^2 + \sum_{\text{angles}} K_{\mathrm{a}}(\theta-\theta_0)^2 + \sum_{\substack{\text{improper}\\ \text{dihedrals}}} K_{\mathrm{i}}(\chi-\chi_0)^2 \\ &+ \sum_{\substack{\text{proper}\\ \text{dihedrals}}} K_{\mathrm{d}}\left[[1-\cos(\phi-\phi_0)] + \frac{1}{2}[1-\cos(3(\phi-\phi_0))]\right] \\ &+ \sum_{\text{contacts}} K_{\mathrm{c}} C_{\mathrm{G}}(r_{ij}, r_0^{ij}) \\ &+ \sum_{\substack{\text{non-native}\\ \text{contacts}}} K_{\mathrm{nc}} \left(\frac{\tilde{\sigma}}{r_{ij}}\right)^{12} . \end{aligned} \tag{3.10}$$

As the MD potential (see Eqs. (3.4)-(3.7)), the SBM potential includes harmonic potentials for bonds, angles, and improper dihedral angles. The potential for proper dihedral angles allows for isomeric conformations. The native structure is employed as the ground state with native bond lengths r_0, native angles θ_0, native improper dihedral angles χ_0, and native proper dihedral angles ϕ_0. The native structure is mainly stabilized by the contact potential $C_{\mathrm{G}}(r_{ij}, r_0^{ij})$, which introduces attractive interactions for atom pairs forming contacts in the native state. Additionally, a repulsive term is added for all possible atom pairs to account for the excluded volume. Here, r_0^{ij} and r_{ij} denote the native and the actual distance of the atom pair (i,j). $\tilde{\sigma}$ represents the excluded volume for Pauli repulsion with $\tilde{\sigma} = 0.25\,\mathrm{nm}$. The force constants are set to $K_{\mathrm{b}} = 20000\,\epsilon/\mathrm{nm}^2$, $K_{\mathrm{a}} = 40\,\epsilon/\mathrm{deg}$, $K_{\mathrm{i}} = 40\,\epsilon/\mathrm{deg}$, and $K_{\mathrm{nc}} = 0.01\,\epsilon$, where deg refers to degree and ϵ is the reduced energy unit used in these types of models [105].

The stabilizing energy is comprised of the terms for the contact potential and dihedral angles. To achieve a consistent energy scale between different systems, the total stabilizing energy E_{s} is set to the total number of atoms N_{atoms} [105]:

$$E_{\mathrm{s}} = \sum E_{\mathrm{c}} + \sum E_{\mathrm{d}} = N_{\mathrm{atoms}} \,, \tag{3.11}$$

where E_{c} is the contact energy and E_{d} is the dihedral energy. Furthermore, the relation between contact energy and dihedral energy is set to [105]:

$$R_{\mathrm{c/d}} = \frac{\sum E_{\mathrm{c}}}{\sum E_{\mathrm{d}}} = 2\,. \tag{3.12}$$

To account for multiple counting of dihedral angles, proper dihedral angles with mutual middle bond are grouped and reweighted with the number of dihedral

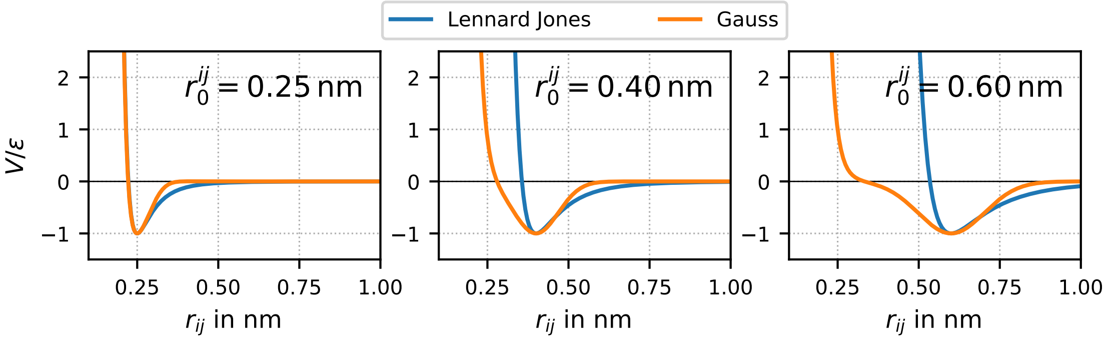


Figure 3.2: Lennard-Jones (blue) and Gaussian (orange) contact potentials for different native contact distances r_0^{ij}. The potentials are almost equal for $r_0^{ij} = 0.25\,\mathrm{nm}$, but for larger ground state distances, the repulsive part of the Lennard-Jones potential shifts to higher r_{ij} whereas the repulsive part of the Gauss potential stays the same.

angles in the group $N_{\mathrm{dihedrals}}$. The force constants for the contact potential K_{c} and for the proper dihedral angle potential K_{d} are chosen in a way that Eqs. (3.11) and (3.12) are fulfilled, namely:

$$K_{\mathrm{d}} = \frac{N_{\mathrm{atoms}}}{1 + R_{\mathrm{c/d}}} \cdot \frac{1}{N_{\mathrm{dihedrals}}} \quad \text{and} \tag{3.13}$$

$$K_{\mathrm{c}} = \frac{N_{\mathrm{atoms}}}{N_{\mathrm{contacts}}} \cdot \frac{R_{\mathrm{c/d}}}{1 + R_{\mathrm{c/d}}}\,, \tag{3.14}$$

where N_{contacts} is the number of contacts.

The contact potential for an SBM can be chosen as a Lennard-Jones like potential, including attractive and repulsive part. An alternative is the Gaussian contact potential, which is given by [104, 106]:

$$\begin{aligned} C_{\mathrm{G}}(r_{ij}, r_0^{ij}) \quad &= \left(1 + \left(\frac{\tilde{\sigma}}{r_{ij}}\right)^{12}\right) \\ &\times \left(1 - \exp\left[-\frac{(r_{ij} - r_0^{ij})^2}{2\sigma^2}\right]\right) - 1\,, \end{aligned} \tag{3.15}$$

with $\sigma^2 = (r_0^{ij})^2/(50\ln 2)$ for each native contact pair (i, j). Lennard-Jones and Gaussian contact potentials are depicted in Fig. 3.2. The Gaussian potential mimics the depth and the increasing width of the Lennard-Jones potential, but does not change in the repulsive part, which yields a fixed excluded volume independent of the contact distance. A further advantage of the Gaussian potential is the possibility to include multiple minima. This allows to model systems with multiple stable conformations and it can be used to investigate conformational transitions [8, 99, 100, 107].

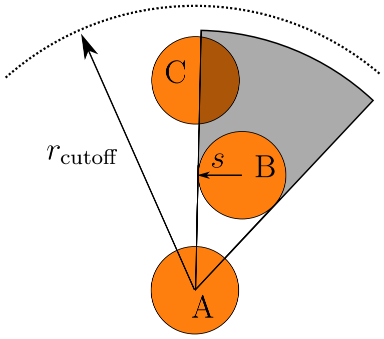


Figure 3.3: Shadow map algorithm. Atoms A, B, and C (orange circles) are considered. A cutoff radius r_{cutoff} and a screening radius s are applied to determine the atoms in contact. If an atom is seen by a third atom within the cutoff radius and not screened by another atom in between, the atom pair is counted as a contact. In this example atom A would be in contact with atom B, but not with atom C, as B is screening atom C from atom A. A similar depiction can be found in [104].

3.2.2 Determination of Contacts

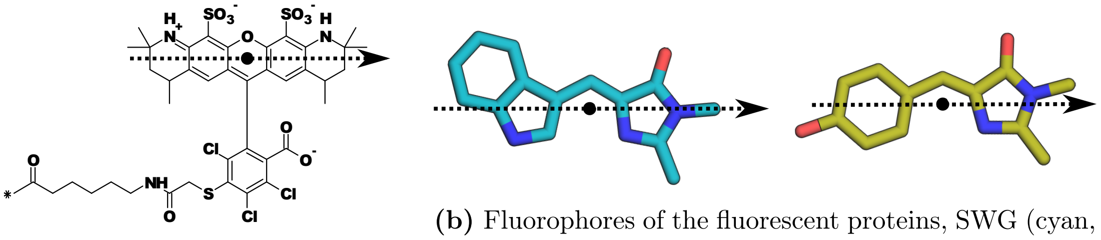
In this work I determine contacts by the shadow map algorithm [104] illustrated in Fig. 3.3. To determine the atoms in contact with an atom A, all atoms within a cutoff radius $r_{\text{AB}} < r_{\text{cutoff}}$ having a sequence distance of at least four residues are considered. Each atom not screened by another atom forms a contact with atom A (with the screening radii s_{neighbor} for directly bonded atoms and s for all other atoms). This criterion has to be satisfied in both directions. In addition to close range contacts, this algorithm accounts for e. g. salt-bridges with separations up to 0.55 nm and interactions mediated through water molecules, which can occur in distances up to 0.7 nm. Here, I use $r_{\text{cutoff}} = 0.6\,\text{nm}$, $s = 0.1\,\text{nm}$ and $s_{\text{neighbor}} = 0.05\,\text{nm}$ [104].

3.2.3 Units in Structure-Based Models

In SBMs, all atoms are handled equally with identical parameters for excluded volume and a unit mass of $m = 1.0$.

As the energies are scaled to the system size (see Eq. (3.11)), SBMs do not have an inherent temperature scale directly comparable to physical temperature. The energy scaling leads to a folding temperature which is about 1.0 in reduced units and corresponds to approximately $T = 120$ in `GROMACS` units [105]. In this work I consistently report SBM temperatures in `GROMACS` units.

Furthermore, the dynamics in SBMs is accelerated due to the smoothened energy landscape, which results in an unphysical time scale. Depending on the application, the time scale has to be adjusted. This adjustment can be done by



(a) Alexa Fluor 546 dye [50].

(b) Fluorophores of the fluorescent proteins, SWG (cyan, left) and CR2 (yellow, right).

Figure 3.4: Fluorophore structures with assumed transition dipole moments (dashed arrows) and centers (black dots). For the Alexa dyes (with Alexa Fluor 546 shown as an example) all atoms of the dye's connected ring system are used for the calculations. For the fluorophore of the cyan fluorescent protein, SWG, the imidazole ring and the pyrrole side of the indole ring are used, and for the fluorophore of the yellow fluorescent protein, CR2, the imidazole ring and the phenol ring are chosen in accordance with the calculations in [111].

comparing folding rates [108], rotational correlation times, or other experimentally accessible time constants. The times in SBM simulations are marked with an SBM subscript in the following (e. g. t_{SBM}) in contrast to physical times.

3.3 Monte Carlo Photon Generation

This section describes the generation of FRET efficiency histograms via Monte Carlo photon simulations. Monte Carlo photon simulations have been successfully applied for analysis of MD simulations of FRET dyes in several studies [15, 109, 110].

As the inter-fluorophore distance is not directly accessible in experiments, I compare the FRET efficiency histograms from experiments to the simulation results. The simulations provide the coordinates of the fluorophores over time.

For the small organic dyes I assume the active part of the dye in the energy transfer to be the ring system. Every part of the structure with at least two connected rings is denoted a "ring system". The distance between dyes is calculated as the distance between the geometric centers of these ring systems (see Fig. 3.4a for an example). The transition dipole moment is assumed to be in the plane of this ring system and chosen to be the axis belonging to its smallest principal moment of inertia (see Fig. 3.4a). Subsequently, the time-dependent inter-dye distances $R_{\mathrm{DA}}(t)$ and the orientation factors $\kappa^2(t)$ are calculated.

The characteristic photophysics of the fluorophores in the fluorescent proteins are influenced by the side chains of the surrounding amino acids. The two fluorophores used here are depicted in Fig. 3.4b. Derived from the calculations made in [111], I assume the active part of SWG in the energy transfer to be the atoms in the imidazole ring and the pyrrole side of the indole ring. For CR2 I choose

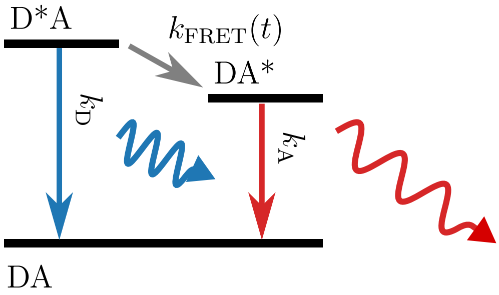


Figure 3.5: Scheme of two-color FRET with the different system states and rates of the occurring processes. The excited donor (D, blue) de-excites with rate k_{D} via internal conversion or fluorescence decay accompanied by the emission of a photon, or transfers the energy to the acceptor (A, red) via FRET with rate $k_{\mathrm{FRET}}(t)$. The acceptor de-excites with rate k_{A} (also either via internal conversion or fluorescence decay). The respective excited fluorophore is marked with an asterisk (*).

the imidazole ring and the phenol ring (see Fig. 3.4b). The centers and dipole moments are calculated from the respective part as described above.

To obtain a FRET efficiency histogram, I perform Monte Carlo photon simulations, similar to the protocol implemented in `md2fret` [110], using $R_{\mathrm{DA}}(t)$ and $\kappa^2(t)$ as input.

The processes in two-color FRET depicted in Fig. 3.5 can be described by the time-independent de-excitation rates of donor k_{D} and acceptor k_{A}, which relate to the lifetimes of donor (in absence of an acceptor) τ_{D} and acceptor τ_{A} via $k_{\mathrm{D}} = 1/\tau_{\mathrm{D}}$ and $k_{\mathrm{A}} = 1/\tau_{\mathrm{A}}$. Donor and acceptor quantum yields Q_{D} and Q_{A} then give the probability for photon emission (fluorescence decay) or internal energy conversion with probability $(1 - Q)$.

The FRET rate $k_{\mathrm{FRET}}(t)$ depends on distance and mutual orientation of the fluorophores and is therefore time-dependent. It can be calculated for every time step as [109]:

$$k_{\mathrm{FRET}}(t) = k_{\mathrm{D}} \left(\frac{R_0}{R_{\mathrm{DA}}(t)} \right)^6 \cdot \frac{\kappa^2(t)}{2/3} \,, \tag{3.16}$$

where R_0 denotes the Förster radius for assumed $\overline{\kappa^2} = 2/3$.

The total de-excitation rates then result in:

$$k_{\mathrm{D,tot}}(t) = k_{\mathrm{D}} + k_{\mathrm{FRET}}(t) \,, \quad \text{and} \tag{3.17}$$

$$k_{\mathrm{A,tot}} = k_{\mathrm{A}} \,. \tag{3.18}$$

The probabilities for the different changes of state in each time step Δt are given by:

$$p_{\mathrm{D^*A}\to\mathrm{DA}}(t) = \left(1 - \mathrm{e}^{-k_{\mathrm{D,tot}}(t)\Delta t}\right) \cdot k_{\mathrm{D}}/k_{\mathrm{D,tot}}(t)\,, \tag{3.19}$$

$$p_{\mathrm{D^*A}\to\mathrm{DA^*}}(t) = \left(1 - \mathrm{e}^{-k_{\mathrm{D,tot}}(t)\Delta t}\right) \cdot k_{\mathrm{FRET}}(t)/k_{\mathrm{D,tot}}(t)\,, \quad \text{and} \tag{3.20}$$

$$p_{\mathrm{DA^*}\to\mathrm{DA}} = \left(1 - \mathrm{e}^{-k_{\mathrm{A,tot}}\Delta t}\right)\,. \tag{3.21}$$

For the generation of each photon, a random starting point in the trajectory is chosen as excitation of the donor. Then the system is propagated in discrete time steps, where in each time step a random number is generated which determines the change of the system's state according to the probabilities p. This propagation is done up to the system's de-excitation. In case of de-excitation to the ground state, according to the quantum yield Q of the respective fluorophore it is randomly determined whether a photon is emitted. The resulting donor and acceptor photons are collected until a specified burst size is reached. The time between excitation and de-excitation can be tracked to evaluate the change in the donor's lifetime in presence of the acceptor.

To determine the burst sizes, the exponential distribution $\mathrm{e}^{-\lambda}$ is used which gives a good approximation for the burst size distributions in experiments. The exponent is set to $\lambda = 2.3$ [110] and a lower cutoff of $n_{\min}$ is applied in accord with the minimal burst size used in the experiments.

For each burst, one single efficiency is calculated by:

$$E = \frac{\frac{n_{\mathrm{A}}}{Q_{\mathrm{A}}}}{\frac{n_{\mathrm{A}}}{Q_{\mathrm{A}}} + \frac{n_{\mathrm{D}}}{Q_{\mathrm{D}}}}\,, \tag{3.22}$$

where n_{D} and n_{A} are the number of donor and acceptor photons collected after donor excitation, respectively. This efficiency value is already corrected for different quantum yields as done in experiments and therefore reflects the statistical influence of this correction. However, the algorithm does not account for crosstalk, direct acceptor excitation, and background [109]. Experimental histograms are typically already corrected for these effects.

3.3.1 Photon Simulation for Three-Color FRET

The Monte Carlo photon simulation for three-color FRET involves three dyes (here exemplary referred to as blue (B), green (G) and red (R)) and their mutual distances and orientations. For all three dyes, the centers and transition dipole moments are extracted from the trajectory and the mutual distances and orientation factors between them are calculated, respectively, as described above. For each pair (M, N) of dyes I use $R_{\mathrm{MN}}(t) := R_{\mathrm{DA,\,MN}}(t)$ and $\kappa^2_{\mathrm{MN}}(t)$ as input.

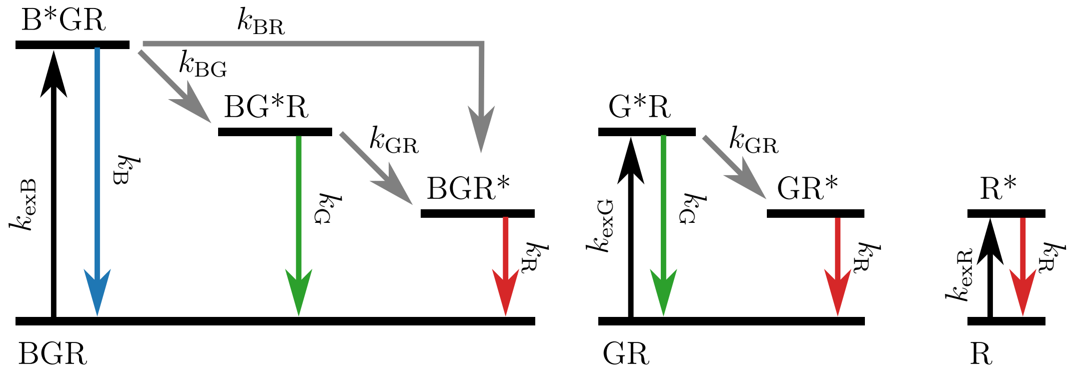


Figure 3.6: Schematic depiction of three-color FRET. The three fluorophores are represented by a blue (B), a green (G) and a red (R) fluorophore. Upon excitation (marked by an asterisk *) of the blue donor fluorophore with rate k_{exB}, it can de-excite with a rate k_{B} or transfer energy via FRET to one of the acceptors with rates k_{BG} and k_{BR}. The green fluorophore then can de-excite with a rate k_{G} or transfer the energy to the red fluorophore via FRET with rate k_{GR}. The red fluorophore then de-excites with the rate k_{R}. A similar scheme for direct excitation of the green fluorophore with rate k_{exG} and the red fluorophore with rate k_{exR} are shown alongside. A similar depiction can be found in [112].

The processes in three-color FRET along with the different rates are depicted in Fig. 3.6. For each pair the FRET rate $k_{\mathrm{MN}}(t)$ is calculated as:

$$k_{\mathrm{MN}}(t) = k_{\mathrm{M}} \left(\frac{R_{0,\mathrm{MN}}}{R_{\mathrm{MN}}(t)} \right)^6 \cdot \frac{\kappa^2_{\mathrm{MN}}(t)}{2/3}\,, \tag{3.23}$$

where $R_{0,\mathrm{MN}}$ denotes the Förster radius of the pair (M, N) with assumed $\overline{\kappa^2} = 2/3$ and k_{M} is the de-excitation rate of dye M in absence of an acceptor. The total de-excitation rates $k_{\mathrm{M,tot}}$ are then calculated by:

$$k_{\mathrm{B,tot}}(t) = k_{\mathrm{B}} + k_{\mathrm{BG}}(t) + k_{\mathrm{BR}}(t)\,, \tag{3.24}$$

$$k_{\mathrm{G,tot}}(t) = k_{\mathrm{G}} + k_{\mathrm{GR}}(t)\,, \quad \text{and} \tag{3.25}$$

$$k_{\mathrm{R,tot}} = k_{\mathrm{R}}\,. \tag{3.26}$$

The transition probabilities between the different states depicted in Fig. 3.6 for a time step with duration Δt result in [group of B. Schuler, private communication]:

$$p_{\mathrm{B^*GR \to BGR}}(t) = \left(1 - \mathrm{e}^{-k_{\mathrm{B,tot}}(t)\Delta t}\right) \cdot k_{\mathrm{B}}/k_{\mathrm{B,tot}}(t)\,, \tag{3.27}$$

$$p_{\mathrm{B^*GR \to BG^*R}}(t) = \left(1 - \mathrm{e}^{-k_{\mathrm{B,tot}}(t)\Delta t}\right) \cdot k_{\mathrm{BG}}(t)/k_{\mathrm{B,tot}}(t)\,, \tag{3.28}$$

$$p_{\mathrm{B^*GR \to BGR^*}}(t) = \left(1 - \mathrm{e}^{-k_{\mathrm{B,tot}}(t)\Delta t}\right) \cdot k_{\mathrm{BR}}(t)/k_{\mathrm{B,tot}}(t)\,, \tag{3.29}$$

$$p_{\mathrm{BG^*R \to BGR}}(t) = \left(1 - \mathrm{e}^{-k_{\mathrm{G,tot}}(t)\Delta t}\right) \cdot k_{\mathrm{G}}/k_{\mathrm{G,tot}}(t)\,, \tag{3.30}$$

$$p_{\mathrm{BG^*R \to BGR^*}}(t) = \left(1 - \mathrm{e}^{-k_{\mathrm{G,tot}}(t)\Delta t}\right) \cdot k_{\mathrm{GR}}(t)/k_{\mathrm{G,tot}}(t)\,, \quad \text{and} \tag{3.31}$$

$$p_{\mathrm{BGR^* \to BGR}} = \left(1 - \mathrm{e}^{-k_{\mathrm{R,tot}}\Delta t}\right)\,, \tag{3.32}$$

where the asterisk denotes the excited state. The probabilities are calculated for every time step.

For each cycle, a random starting point t_0 in the trajectory is chosen for excitation of the donor (B). The system is propagated in discrete time steps as done for two-color FRET up to the system's de-excitation.

In the experiment the sample is excited by alternating laser pulses of three different colors with a frequency of 20 MHz, exciting B, G, and R every 50 ns, respectively. Consequently, for the consecutive excitation of G and R, times close to the time points $t_1 = t_0 + \frac{1}{3} \cdot 50\,\mathrm{ns}$ and $t_2 = t_0 + \frac{2}{3} \cdot 50\,\mathrm{ns}$ are chosen. The system is then propagated as before. The photons are collected and new starting points generated until a specified burst size (here the number of photons collected after donor excitation) is reached. The photon counts of fluorophore N after excitation of fluorophore M, n_{MN}, are then corrected by the quantum yield of fluorophore N:

$$N_{\mathrm{MN}} = \frac{n_{\mathrm{MN}}}{Q_{\mathrm{N}}}\,. \tag{3.33}$$

For each burst, photon rates, as described in Sec. 2.2.4, are calculated. The burst sizes are determined as described above.

3.4 Summary

This chapter provided an overview over the molecular dynamics technique SBMs are based on. As simulation technique, I introduced Langevin dynamics, which models a system's dynamics with Newton's equations of motion while implicitly accounting for friction and random collisions with solvent molecules. The simplistic SBM potential employs the protein's native state as its ground state. It allows for simulation of large conformational ensembles with reasonable computational costs. SBMs do not have an explicit inherent time or temperature scale directly comparable to physical units, but physical scales can be introduced by comparison to experimental values.

Finally, Monte Carlo photon simulations that generate photon statistics which can directly be compared to experimentally measured two-color and three-color FRET data were presented.

4
Method Development

In this work I want to model the entire FRET process in simulations, including the system dynamics and photon statistics. As first requirement structures and parametrization of the systems with attached fluorophores are needed.

The first part of this chapter presents how small organic dyes are modeled. The dyes' structures, obtained by quantum-chemical geometry optimizations, and parameters are introduced in Secs. 4.1 and 4.2, respectively. Subsequently, Secs. 4.3 and 4.4 describe protein parameters for the simulation of folded and unfolded ensembles and the generation of the whole system, respectively. Sec. 4.5 presents the simulation protocol for protein-dye systems. All descriptions up to this point are based on my publication (see [113]). In addition to that, the modifications necessary in the simulation protocol for simulating three-color FRET systems are provided.

The second part of this chapter in Sec. 4.6 presents how fluorescent proteins are implemented in SBM simulations. In this work, I focus on a FRET-based glucose sensor comprising a sensing protein and two fluorescent proteins connected to the sensing protein by linkers, as described in Sec. 4.6.1. The structure and parametrization of the proteins (see Secs. 4.6.2 and 4.6.3) and of the linkers (see Secs. 4.6.4 and 4.6.5) are discussed in the subsequent subsections. Sec. 4.6.6 presents the generation and selection of a starting structure for simulations of the whole sensor. Finally, the simulation protocol for the glucose sensor is given in Sec. 4.6.7.

4.1 Dye Structure

For the simulation of the small organic dyes I want to find a minimal and robust set of parameters which sufficiently replicates and predicts experimental data. At the same time, the parameters should be systematized to be easily transferable to other dye molecules. In this work, I only consider the widely used Alexa Fluor dyes [50] and the Biotium dye CF680R (B680) [114]. Nonetheless, the method is applicable to all other small organic dyes as well.

SBM simulations require an initial structure of the system. For many dyes, only chemical structures are available. Therefore, I start by performing quantum-chemical geometry optimizations in collaboration with the group of C. Jacob to obtain three-dimensional structures of the dyes. For the Alexa Fluor dyes, the initial structures are constructed based on the chemical structures. The neutral forms of carboxyl, amide, and sulfite groups are used in all cases. These initial structures are optimized using density functional theory (DFT) as implemented in `TURBOMOLE 6.5` [115, 116]. In the DFT calculations, the BP86 exchange-correlation functional [117, 118] and the def2-TZVP basis set [119] are used. The optimized three-dimensional structures of several Alexa Fluor dyes can be found in [113].

Subsequently, I add the linker and maleimide structure to the dye structure. Maleimide groups are often used to bind the dyes to specific protein residues mutated to cysteines beforehand. In this work I use two pairs of Alexa Fluor dyes - the Alexa Fluor 488 dye with C_5-linker (AF488) and Alexa Fluor 594 dye with C_5-linker (AF594), and the Alexa Fluor 546 dye with C_5-linker (AF546) and Alexa Fluor 647 dye with C_2-linker (AF647).

For the three-color FRET measurements, the group of B. Schuler uses AF488, AF594, and Biotium dye CF680R (B680). The two Alexa Fluor dyes can be site-specifically labeled to cysteine residues [112]. To label B680 in a site specific manner as well, oxime ligation to the non-natural amino acid p-acetyl phenylalanine is used. In order to mimic this in the structure, I use a mutation to phenylalanine at the labeling site and include the missing atoms into the structure of B680. The three-dimensional structure of B680 is then determined by DFT calculations [data from P. Friederich, private communication]. In these calculations the B3-LYP functional [120] and the def2-SV(P) basis-set [121] are used.

As an example, Fig. 4.1 shows the preparation of the structure for Alexa Fluor 546.

4.2 Dye Parameters

For the initial dye parametrization in the SBM I choose the same parameters as for the proteins (see Sec. 3.2.1). To generate the dye topology, the information about bonds, angles, and dihedral angles present in the structure is needed. The bond information is already given in the chemical structure, meaning angles and dihedral angles can be determined automatically from the bond information (see

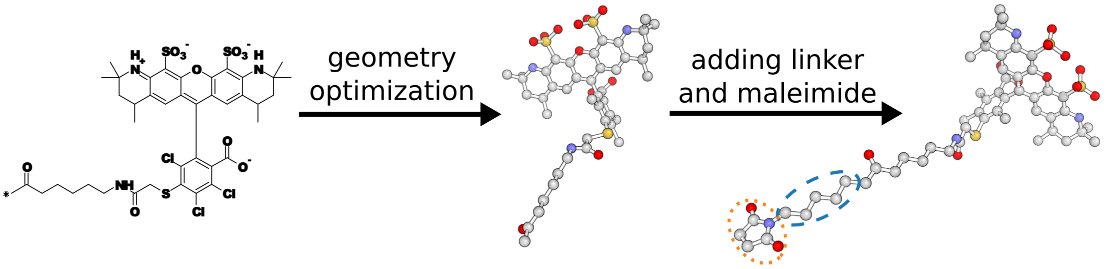


Figure 4.1: Generation of the structure for the Alexa Fluor 546 dye with C_5-linker and maleimide. A quantum-chemical geometry optimization of the two-dimensional chemical structure (left) [50] generates a three-dimensional structure of the dye (middle). Subsequently, a C_5-linker (blue dashed ellipse) and a maleimide (orange dotted ellipse) are added to the structure. The resulting structure is shown on the right.

Sec. C for details). Both dye-dye and dye-protein interactions are limited to a repulsive excluded volume term. Additional interactions can easily be employed as needed.

Dye Temperature

As mentioned in Sec. 3.2.3, SBMs do not have an inherent temperature scale. To adjust the dye parameters independently of the protein in a consistent way with a minimal number of free parameters, I use a separate temperature T_{dye} for the dyes. This is done by assigning protein and dyes to different groups in the simulation and coupling them to separate temperature baths. The separation of the temperatures uncouples the behavior of dye and protein, so I can use the same dye temperature regardless of the state of the protein investigated. In this way, dyes attached to a folded or unfolded protein, which are realized by different protein temperatures, can be treated equally. The effect of the dye temperature on the protein dynamics through possible energy transfer between the two temperature regions was tested [data not shown]. I found only a slight increase in the fluctuations of the residues the dyes are attached to, whereas the influence on the dynamics of the adjacent residues is negligible.

The dyes are highly flexible and their motions fast compared to the protein motions (see also Fig. 4.4b), therefore I use high temperatures for the dyes. Above temperatures of $T = 250$ in the SBM, numerical instabilities [122] arise in the `GROMACS` version used*. To further account for the high dye flexibility and accelerate the dye motion, I change the mass of the dye's atoms from the SBM unit atom mass of 1.0 to 0.2.

*`GROMACS v4.5.4` [84] with the extension for Gaussian contact potentials [104]

4.3 Protein Parameters

SBM simulations require a native starting structure and a topology of the system. The protein structures I use for the simulations are taken from the Protein Data Bank (PDB) [123], a collection of experimentally measured three-dimensional structures of proteins and nucleic acids.

The first protein is the chymotrypsin inhibitor 2 (CI-2, PDB: 2CI2 [34]), which I use as a test system as it is a widely studied and well understood protein. The second and third object of study are systems investigated by the groups of my collaborators, G.U. Nienhaus and B. Schuler, namely, the tenth type III module of fibronectin (^{10}FNIII, PDB: 1TTG [124]) and the cold-shock protein from the hyperthermophilic bacterium *Thermotoga maritima* (CspTm, PDB: 1G6P [125]).

Moreover, the pore-forming toxin cytolysin A (ClyA), equally studied by the group of B. Schuler in two-color as well as in three-color FRET experiments, is investigated in line with this work. This protein exists as a monomer (ClyA monomer, PDB: 1QOY [126]) and undergoes a conformational change to the protomer before assembling to the dodecameric pore complex [127] (ClyA dodecamer, PDB: 2WCD [128]). ClyA protomer and ClyA trimer conformations are taken as a single chain and three chains from the dodecamer structure, respectively. Of the whole 303 residue amino acid sequence, only residues 1 to 298 and 8 to 292 are experimentally resolved for monomer and protomer, respectively. To enable simulating the dye labeling at residues 2 and 303, I generate homology models for the structures of monomer and protomer [129]. They are used for all further studies of these two conformations. ClyA trimer and ClyA dodecamer are simulated with only residues 8 to 292. As the labeling sites are at residue 56 and 252 I do not expect any influence of the missing residues.

The topologies for all proteins are generated with the SBM implementation in `eSBMTools` [103] developed in our group (see also Sec. 3.2).

Protein Temperature

Another parameter which has to be determined in the SBM is the temperature of the protein in the folded state. As the folding temperature in SBMs is around $T = 120$ for all proteins [105], a temperature of $T = 50$ is reasonable to describe a protein in its folded state. I use this temperature for the test system CI-2. In the cases of CspTm, ^{10}FNIII, and ClyA I want to achieve a quantitative comparison against experimental measurements. To identify the SBM temperature corresponding to the experimental setup, I initially perform regular all-atom MD simulations of CspTm, ^{10}FNIII, and ClyA (in monomeric and protomeric form) using the AMBER99 force field [87] at the physiological temperatures from the

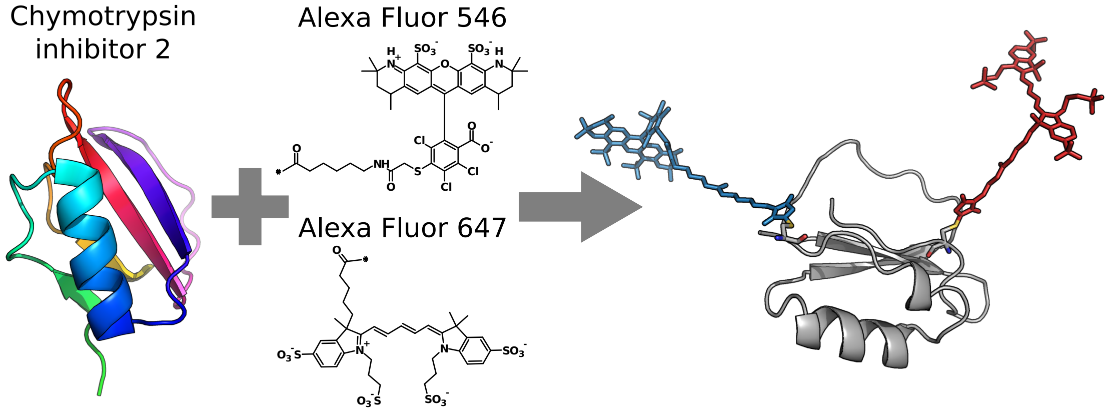


Figure 4.2: Structure of CI-2 (left), AF546 (blue) and AF647 (red) dyes [50] (middle). The dyes are attached with the respective linkers via a maleimide bound to residues 20 and 78 of CI-2, respectively. The merged structure is shown on the right.

experiments, i.e. $T = 296/295/295\,\mathrm{K}$ for ^{10}FNIII, CspTm, and ClyA, respectively. I then compare the root mean square fluctuations (see Sec. B) of the C_α-atoms in these simulations with the corresponding values in SBM simulations of different temperatures. I choose the SBM temperature resulting in the least deviation from the AMBER99 simulation, as done in [108]. An example and a detailed description can be found in Sec. D. For the resulting protein temperatures, see Tab. 4.1.

For simulations of the unfolded state, I use a temperature well above the folding temperature ensuring that the protein is unfolded throughout the simulation. In particular, I found the influence of different temperatures above the folding temperature on the results to be negligible [data not shown].

4.4 Merging of Dyes and Protein

For the simulation of the whole protein-dye system, structures of dyes and protein have to be merged. The Alexa Fluor dyes are typically attached via a maleimide group to residues of proteins mutated to cysteine. In the protein structures I use, the respective residue is mutated accordingly to a cysteine and the dye structure (already containing linker and maleimide) is attached to the sulfur atom of the cysteine. The dye structure is preferably placed orthogonally to the protein surface while respecting steric restrictions and preventing clashes between atoms. As an example, Fig. 4.2 shows CI-2 with AF546 and AF647 dyes attached to residues 20 and 78, respectively. A detailed description of the merging procedure can be found in Sec. E.

A similar procedure is conducted for B680, except that the respective residue is mutated to a phenylalanine. Then, the structure generated as described in Sec. 4.1

Table 4.1: Simulation parameters for different systems. The proteins, their temperatures for simulation of folded (T_{F}) and unfolded (T_{U}) states, and the dye pairs (donor/acceptor) are given. Labeled residues (with donor/acceptor position) and dye temperature(s) (for donor/acceptor) are shown.

Protein	T_{F}	T_{U}	Dye pair	Labeled residues	T_{dye}
			D/A	D/A	D/A
CI-2	50	170	AF546/AF647	20/78	250
CI-2	50	170	AF488/AF594	20/78	190/250
CspTm	90	150	AF488/AF594	2/68, 68/2, 11/68, 68/11, 23/68	190/250
^{10}FNIII	60	200	AF546/AF647	11/86	250
ClyA monomer	70	200	AF488/AF594	56/252, 2/303	165/250
ClyA protomer	80	200	AF488/AF594	56/252, 2/303	165/250

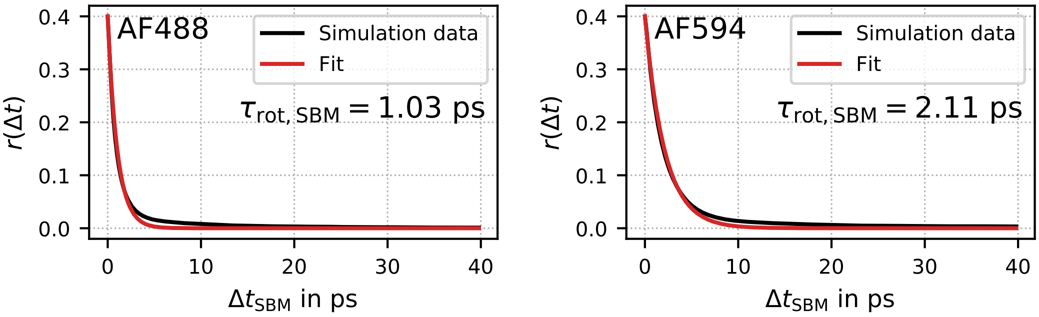


Figure 4.3: Exemplary fits of the fluorescence anisotropy $r(\Delta t)$ as a function of time Δt_{SBM} for AF488 (left) and AF594 (right) attached to CI-2. In addition, the calculated rotational correlation time $\tau_{\mathrm{rot,\ SBM}}$ is given in units of the SBM time scale.

including the additional carbon and oxygen atoms of p-acetyl phenylalanine is attached to the respective carbon atom of the phenylalanine ring.

4.5 Simulation Protocol for Protein-Dye Systems

For the simulations I use `GROMACS v4.5.4` [84] with the extension for Gaussian contact potentials [104], an SBM potential as given in Eq. (3.10), and Langevin dynamics (see Sec. 3.1.1). The protein temperatures used can be found in Tabs. 4.1 and 4.5.

Table 4.2: Time step Δt_{SBM} and total simulation times for simulation of folded state $t_{\text{tot, F, SBM}}$ and unfolded state $t_{\text{tot, U, SBM}}$ for the different systems in units of the SBM time scale.

Protein	Δt_{SBM}	$t_{\text{tot, F, SBM}}$	$t_{\text{tot, U, SBM}}$
CI-2	0.2 fs	500 ns	500 ns
CspTm	0.2 fs	500 ns	1000 ns
^{10}FNIII	0.5 fs	500 ns	500 ns
ClyA	0.5 fs	500 ns	2000 ns

To balance the dye motions against each other, I determine appropriate dye temperatures. I perform simulations with varying dye temperatures and calculate the respective rotational correlation times by using Eqs. (2.12) and (2.13). Exemplary fits are shown in Fig. 4.3 for AF488 and AF594 attached to CI-2. The rotational correlation time τ_{rot} is a measure of the dye flexibility and highly dependent on the temperature. Higher temperatures yield smaller τ_{rot}. To get a high dye flexibility, I choose a value of $T_{\text{dye}} = 250^{\dagger}$ for the faster dye. The temperature of the slower dye is adjusted in such a way that the relation of both rotational correlation times from experiments is matched (see Tab. 4.3).

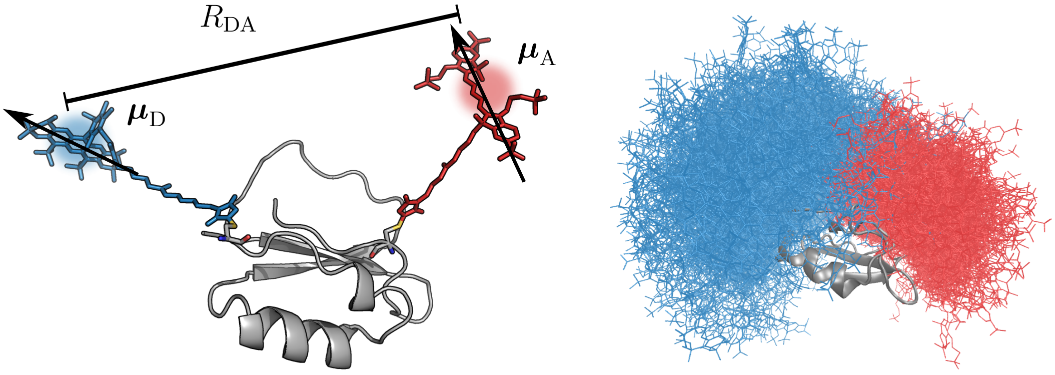
As I do not have experimental values for τ_{rot} of AF647 and of B680, I use a temperature of $T_{\text{dye}} = 250$ in these cases. Varying this temperature slightly does not alter the final results.

The final simulations are performed with the time steps and total simulation times given in Tab. 4.2. Due to its much larger conformational space, the unfolded state has to be sampled for a longer time than the folded state. The different total times needed result from the different system sizes. The temperature coupling constant is always set to $\tau_T = 0.1\,\text{ps}$.

For the subsequent Monte Carlo photon simulations, the inter-dye distance $R_{\text{DA}}(t)$ and the transition dipole moments of donor $\boldsymbol{\mu}_{\text{D}}(t)$ and acceptor $\boldsymbol{\mu}_{\text{A}}(t)$ (see Fig. 4.4a) are extracted from the simulations as described in Sec. 3.3. An example for the high dye flexibility and the variety of possible dye conformations during the simulations is shown in Fig. 4.4b.

As mentioned in Sec. 3.2.3 SBMs do not have an inherent time scale directly comparable to the physical time scale, due to their accelerated dynamics. Nonetheless, evaluation of the simulations and generation of FRET histograms comparable to experimental measurements requires a time scale. In analogy to previous work [108], I introduce a time scale based on comparison of experimental and theoretical time-dependent values. In this work, I use the rotational correlation times τ_{rot} of the dyes. I calculate the fluorescence anisotropy decay $r(t)$ in the

†Higher temperatures quickly become instable in simulations, see also Sec. 4.2.


R_{DA}
$\boldsymbol{\mu}_{\mathrm{A}}$
$\boldsymbol{\mu}_{\mathrm{D}}$

(a) Dipole moments and distance. **(b)** Dye positions over simulation.

Figure 4.4: CI-2 (gray) with AF546 (blue) and AF647 (red) dyes attached to residues 20 and 78, respectively. (a) The transition dipole moments of donor $\boldsymbol{\mu}_{\mathrm{D}}$ and acceptor $\boldsymbol{\mu}_{\mathrm{A}}$ (black arrows) and the distance between the dyes' centers R_{DA} are shown. (b) The dyes' structures for different time points during the simulation show the high dye flexibility and the variety of possible dye conformations.

simulations by using Eq. (2.12). The fit with Eq. (2.13) yields a rotational correlation time for the simulation. Comparison with the experimental value provides a conversion factor to adjust the simulation time scale.

Using this converted time scale, I gain photon statistics from Monte Carlo photon simulations as described in Sec. 3.3, which fully account for shot noise.

Simulation parameters for photon simulations of different dye pairs are given in Tabs. 4.3 and 4.4. For FRET with two dyes I generate donor and acceptor photons for the whole simulation as described in Sec. 3.3. This corresponds to around 100 µs on the physical dye time scale (for ClyA in the folded state around 280 µs and in the unfolded state around 1 ms). Photons are generated until a specified number of photons, the burst size, is collected. I use each burst to calculate a single FRET efficiency value. For each FRET efficiency histogram I generate $2 \cdot 10^4$ bursts with a minimum number of photons of $n_{\mathrm{min}} = 50$.

I perform simulations with two dyes for four different proteins (CspTm, CI-2, ^{10}FNIII, and ClyA) with two different dye pairs (AF488 and AF594, AF546 and AF647). The Förster radius R_0 depends on the refractive index of the medium, therefore it changes slightly with higher denaturant concentrations in the experiments. Hence, I choose a modified Förster radius for the evaluation of the unfolded two dye systems (see Tab. 4.4).

4.5.1 Simulation Protocol for Three-Color FRET

The simulations with three dyes are conducted accordingly. The protein and dye parameters used can be found in Tabs. 4.5 and 4.3. The total simulation times of

Table 4.3: Parameters for the different dyes used in the Monte Carlo photon simulations. Lifetimes τ, quantum yields Q, and experimentally measured rotational correlation times τ_{rot} are listed. [1]The lifetime of Biotium CF680R was not available and therefore was estimated [group of B. Schuler, private communication]. The reported rotational correlation times for Alexa Fluor 488 and Alexa Fluor 594 attached to CspTm ([2][23]) differ from measured rotational correlation times of the same dyes attached to ClyA ([3][group of B. Schuler, private communication]). The particular systems where the parameters are used are given in parentheses.

Dye	τ	Q	τ_{rot}	
Alexa Fluor 488	4.1 ns	0.92	240 ps[2]	(CI-2, CspTm)
			660 ps[3]	(ClyA)
Alexa Fluor 594	3.9 ns	0.66	460 ps[2]	(CI-2, CspTm)
			1100 ps[3]	(ClyA)
Alexa Fluor 546	4.1 ns	0.79	301 ps	(CI-2, ^{10}FNIII)
Alexa Fluor 647	1.0 ns	0.33	–	(CI-2, ^{10}FNIII)
Biotium CF680R	~3.0 ns[1]	0.34	–	(ClyA)

Table 4.4: Förster radii for proteins in water $R_{0,\mathrm{F}}$ and in 7.0 M GdmCl ([1][130]) and 4.63 M GdmCl ([2][26]) $R_{0,\mathrm{U}}$, respectively, are listed. [3]For the three-color FRET simulations of ClyA (ClyA$_{3\mathrm{C}}$) the values are taken from [112] and the same values are used for folded and unfolded state. The particular systems where the parameters are used are given in parentheses.

Donor	Acceptor	$R_{0,\mathrm{F}}$	$R_{0,\mathrm{U}}$	
AF488	AF594	5.4 nm[1]	5.1 nm	(CI-2, CspTm, ClyA)
AF546	AF647	6.6 nm[2]	6.3 nm[2]	(CI-2, ^{10}FNIII)
AF488	AF594	5.8 nm[3]		(ClyA$_{3\mathrm{C}}$)
AF488	B680	4.7 nm[3]		(ClyA$_{3\mathrm{C}}$)
AF594	B680	6.8 nm[3]		(ClyA$_{3\mathrm{C}}$)

Table 4.5: Simulation parameters for simulating ClyA with three dyes. The protein configuration, the protein temperature for simulation of folded (T_{F}) and unfolded (T_{U}) states, and the used dyes (donor/acceptor 1/acceptor 2) are given. The labeled residues (with donor/acceptor 1/acceptor 2 position) and the dye temperature(s) (for donor/acceptor 1/acceptor 2) are shown.

Protein	T_{F}	T_{U}	Dyes	Labeled residues	T_{dye}
			D/A1/A2	D/A1/A2	D/A1/A2
Monomer	70	200	AF488/AF594/B680	252/56/8	165/250/250
Protomer	80	200			

the final simulations are the same as above (see Tab. 4.2). The inter-dye distances $R_{\mathrm{DA,\ MN}}(t)$ of each dye pair (M,N) and their mutual orientations $\kappa^2_{\mathrm{MN}}(t)$ are extracted from the simulations as described in Sec. 3.3. I gain photon statistics as described in Sec. 3.3.1. The simulation parameters for the photon simulations of the different dyes are given in Tabs. 4.3 and 4.4. I generate photons for the whole simulation as described in Sec. 3.3.1, which corresponds to around 280 µs and 1 ms on the physical dye time scale for folded and unfolded states, respectively. Here, the burst size refers to the number of photons collected after donor excitation. I use each burst to calculate a single set of photon rates (see Sec. 2.2.4) and generate $2 \cdot 10^4$ bursts for each histogram with a minimum number of photons of $n_{\mathrm{min}} = 30$. The Förster radii are taken from [112] and used for all simulations (see Tab. 4.4).

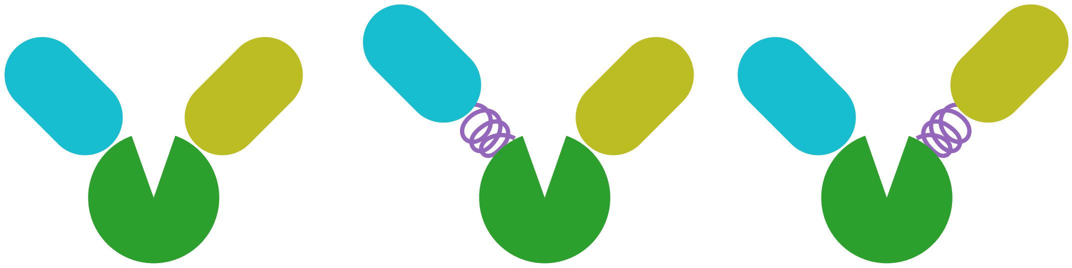
4.6 Biosensors with Fluorescent Proteins

An application where orientation of the fluorophores plays an important role is the use of fluorescent proteins instead of small dyes. Due to their size, the movement of the fluorescent proteins can be restricted and is also on a time scale orders of magnitude larger than their lifetimes (see also Sec. 5.9.4). In particular, I will discuss genetically encoded FRET-based biosensors [9, 14]. As they are typically experimentally engineered by trial and error, the mechanisms involved in their function are not always fully understood. Here, simulations can help with testing different hypotheses.

FRET-based biosensors are used for quantification of a small ligand, as they change their conformation and associated FRET signal according to binding to this ligand. This mechanism enables measurement of the concentration of a ligand *in vitro* and *in vivo* without interfering with the system, which has many applications in medical sciences and microbiology. Designing these sensors is challenging, as they are required to be as sensitive as possible, and includes an extensive optimization process [13]. A genetically encoded biosensor typically consists of a sensing protein fused to two fluorescent proteins via linkers [13].

The aim of a good sensor is to transfer small conformational changes in the sensing protein into large changes in the associated FRET signal. A good sensor makes use of reorientation or change of rotational flexibility, resulting in a changed κ^2, and/or change in distance of the fluorescent proteins, resulting in a changed R_{DA}, to maximize the difference in FRET intensity ratio ΔR (see Eq. (2.6)) upon ligand binding.

Which specific effects contribute to the change in intensity ratio is often unknown, but would help to better understand and fundamentally improve the sensor design. There have been attempts to construct biosensors and sample conformations through rigid body modeling [20] to obtain a better understanding of the



(a) Sensor 1. **(b)** Sensor 2. **(c)** Sensor 4.

Figure 4.5: Schematic depictions of sensor variants, comprising of Glc-BP (green), CFP (cyan), and YFP (yellow). Sensor 1 does not contain linkers, sensor 2 has a flexible linker at CFP, and sensor 4 has a flexible linker at YFP. The flexible linker is depicted as purple helix. A similar depiction can be found in [132]. The numbering refers to the notation in [132].

structural ensembles. The mentioned method, however, does not include influences of the system's dynamics, linker rigidity, photon statistics, or the weak dimerization tendency of some fluorescent proteins. With the approach I present in this work it is possible to include all these features.

In this work I focus on a glucose sensor [13], which is well studied and its design improved by several groups. It consists of the glucose binding protein (Glc-BP, see also Fig. 4.8) with inserted mTurquoise2 (donor, here referred to as CFP (cyan fluorescent protein), see also Fig. 4.7) and Venus (acceptor, here referred to as YFP (yellow fluorescent protein), see also Fig. 4.7). Steffen et al. have developed a toolbox of different linkers between fluorescent proteins and a sensing protein [131], which Höfig et al. have applied to this glucose sensor [132].

I consider three of these sensors, referred to as sensor 1 without linker, sensor 2 with a flexible linker at CFP, and sensor 4 with a flexible linker at YFP. All sensors behave differently in FRET experiments [132] and yield distinct FRET ratios, which rises the question why the position of the flexible linker is so important. A schematic depiction of the three sensors is shown in Fig. 4.5.

4.6.1 Sensor Structures

The sensors are constructed by genetically encoding the different proteins in an amino acid sequence. A description can be found in e. g. [131]. A schematic of the different sensor sequences is depicted in Fig. 4.6. Instead of attaching the CFP to the N-terminus of Glc-BP, it was found to be beneficial to insert it into the sequence of Glc-BP [13]. Residues 1 to 11 of Glc-BP are at the beginning of the sequence and three alanine residues are inserted as a restriction site to enable proper folding of the proteins. CFP is then attached to residue 12 of Glc-BP via either a restriction site with the sequence `GSDLV` or a flexible $(\mathrm{GGS})_4$-linker and

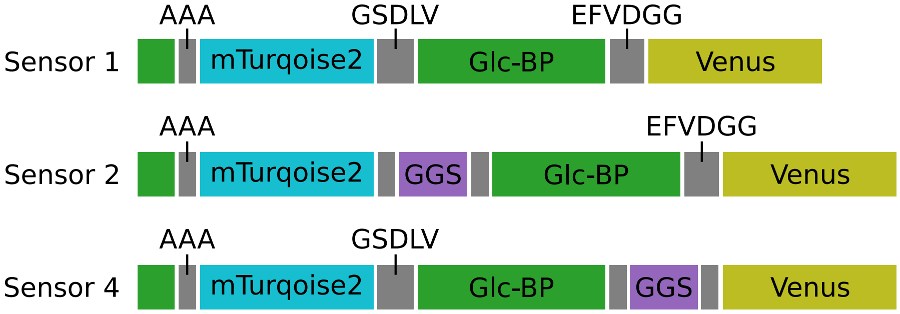


Figure 4.6: Schematic of the three sensor sequences. Glc-BP, CFP, and YFP are shown in green, cyan, and yellow, respectively. The flexible linker is colored in purple and restriction sites are shown in gray with their sequences depicted. CFP is inserted into Glc-BP, so the first 11 residues of Glc-BP are in the beginning of the sequence and the main part in the middle.

two restriction sites, each consisting of two amino acids (`GS(GGS)`$_4$`PG`). YFP is attached to the C-terminus of Glc-BP via a restriction site with sequence `EFVDGG` or a flexible $(\text{GGS})_4$-linker and two restriction sites (`EF(GGS)`$_4$`VDGG`). The complete sequences used for generation of the sensors in experiments and simulations are given in Sec. F. In the simulations, parts of the N- and C-termini are omitted as no resolved structures are available. However, they should not have considerable effects on the results.

4.6.2 Protein Structures

PDB structures are available for both fluorescent proteins. The structures of mTurquoise2 (CFP, PDB: 3ZTF [133]) and Venus (YFP, PDB: 1MYW [134]) along with the two fluorophores are shown in Fig. 4.7. The fluorophores result from the autocatalytic cyclization of residues Ser65-Trp66-Gly67 (CFP) and residues Gly65-Tyr66-Gly67 (YFP), respectively.

For Glc-BP two structures are available, i.e. the structure with glucose bound (Glc-BP$^{+\text{G}}$, PDB: 2FVY [135]) and the apo open form without glucose (Glc-BP$^{-\text{G}}$, PDB: 2FW0 [135]). In both structures, the first and the last three residues are not resolved. Whereas the first residue is omitted in the simulations, the last three residues are included in the subsequent linker structures or structures of restriction sites, respectively (see Sec. 4.6.4). According to the sequence used for the glucose sensor, residues 12 and 13 are deleted in the structures (see also Sec. F). Both structures are shown in Fig. 4.8. Their mutual root mean square deviations (RMSD, see Sec. B) is only 0.37 nm and the respective shift between the C_α-atoms of the fluorescent protein attachment positions at residues 14 and 303 (shown as cyan and yellow spheres) is very small. The C_α distance between residues 11 and 303 changes from 2.53 nm in Glc-BP$^{+\text{G}}$ to 2.62 nm in Glc-BP$^{-\text{G}}$. The C_α distance

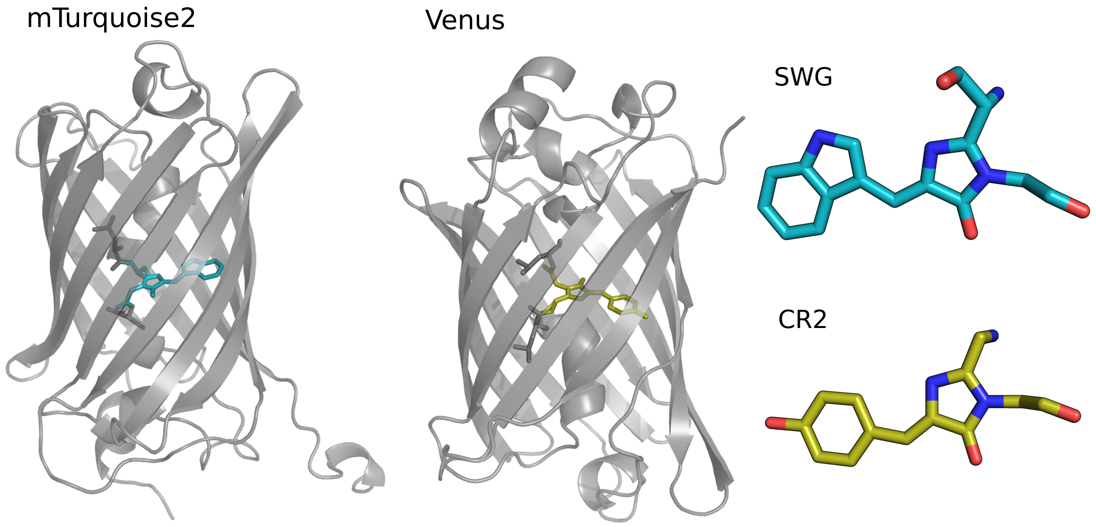


Figure 4.7: The two fluorescent proteins mTurquoise2 and Venus along with their respective fluorophores SWG (cyan) and CR2 (yellow).

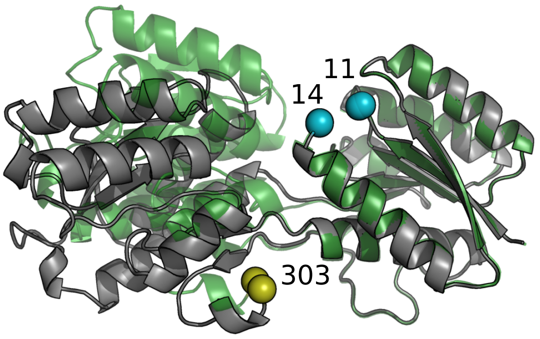


Figure 4.8: Glucose binding protein Glc-BP in the glucose free form (gray) and the glucose bound form (green), aligned to the first 100 residues. The C_α-atoms of residues 11 and 14 (the site of the CFP insertion) are depicted as cyan spheres for both structures. Their respective position is almost identical for both structures. The last structurally resolved residues of both structures (residue 303, the YFP attachment position) are depicted as yellow spheres. Due to the conformational change in Glc-BP upon glucose binding the position of this site changes slightly with respect to the CFP attachment position.

between residues 14 and 303 changes from 2.10 nm in Glc-BP^{+G} to 2.16 nm in Glc-BP^{-G}. From these observations, only a slight distance increase between the fluorescent proteins when going from Glc-BP^{+G} to Glc-BP^{-G} is expected. Similar to the negligible shift of the C_α-atoms of residues 11 and 14, the shift of the C_α-atom of residue 303 is also only 0.27 nm and thus surprisingly small to result in the large change of FRET ratios upon glucose binding observed in the FRET experiments. This hints to an unknown mechanism amplifying the change in FRET.

4.6.3 Protein Parameters

As done in Sec. 3.2, the topologies for all proteins are generated with `eSBMTools`. For the fluorophores in the fluorescent proteins I generate topologies in the same way as for the dyes (see Sec. 4.2). Then, I include the fluorophores into the topology generation in `eSBMTools` and treat them as regular amino acids to maintain proper connections between the chain elements. As for the dyes, mutual interactions between the three proteins are limited to a repulsive excluded volume term for now (see also Sec. 4.6.8).

The atom masses for both fluorescent proteins are set to $m_{\mathrm{FP}} = 0.2$ to accelerate the dynamics in the simulations, as was done for the dyes before. I determine the temperature for CFP and YFP by comparison of regular MD simulations with SBM simulations in the same way as described in Sec. 4.3 (for parameters and a detailed description see also Sec. D). The fluorophores are omitted in the regular MD simulations as there are no standardized AMBER99 parameters available for these structures, but should not influence the overall RMSF. The resulting temperatures are $T = 70$ for both CFP and YFP.

For Glc-BP, temperature comparison simulations are not reasonable as the ligand is not straightforward to parametrize and include into the AMBER99 simulations. Therefore, I use the same temperature as for the fluorescent proteins. Given that all determined temperatures for the proteins are in this range and small changes do not affect the result, this seems to be a valid assumption (see also Sec. D).

I further analyze the difference of both structures Glc-BP^{+G} and Glc-BP^{-G} in SBM simulations to test if the model is too coarse to model both structures as different states. I simulate both structures at different temperatures and calculate the RMSD to both starting structures, respectively. For small temperatures ($T = 30$) the structures are well separated for both simulations [data not shown], meaning the RMSD to the starting structure is always much lower than the RMSD to the other structure. This is still valid at $T = 70$ for Glc-BP^{+G}, as the closed conformation is stabilized by contacts (see Sec. G). For Glc-BP^{-G}, the RMSD fluctuations are larger and the structure comes closer to the structure of Glc-BP^{+G} than vice

Table 4.6: Sequences modeled for the linkers and restriction sites of the sensor variants. The last three residues of Glc-BP (green) are included in the sequences as they are not resolved in the structure. The flexible $(\mathrm{GGS})_4$-linker and the restriction sites are colored in purple and black, respectively.

Name	Sequence	Sensors
N_{rs}	`GSDLV`	Sensor 1 (CFP), Sensor 4 (CFP)
$\mathrm{N}_{\mathrm{flex}}$	$\mathtt{GS(GGS)_4PG}$	Sensor 2 (CFP)
C_{rs}	`SKKEFVDCC`	Sensor 1 (YFP), Sensor 2 (YFP)
$\mathrm{C}_{\mathrm{flex}}$	$\mathtt{SKKEF(GGS)_4VDGG}$	Sensor 4 (YFP)

versa. This is expected, as the conformational freedom compared to Glc-BP$^{+\mathrm{G}}$ is increased and probably resembles the physical behavior appropriately.

4.6.4 Linker Structures

The linker structures are a particular challenge, as there are no three-dimensional structures and little information about their behavior available. Still, the linkers are one of the crucial parts for the systems' dynamics. Furthermore, the structures of the restriction sites are not known and the three C-terminal residues in Glc-BP not resolved. To incorporate these regions into the simulations, I generate structures of the sequences shown in Tab. 4.6 in the following way.

First, I build the respective sequence in `pymol` [136] as extended amino acid chain without any secondary structure. To relax each structure, I simulate it in an all-atom AMBER99 force field with explicit water for 100 ns. To check the convergence of the simulation, I calculate the RMSD with reference to the last structure in the simulation for each system. For $\mathrm{N}_{\mathrm{flex}}$, C_{rs}, and $\mathrm{C}_{\mathrm{flex}}$, the RMSD value decreases rapidly at the beginning of the simulation and remains constant, indicating convergence to a final structure [data not shown]. In the case of N_{rs}, the convergence is not that obvious. This is expected as it is a rather short sequence. However with an overall low RMSD value it is valid to use this structure, as some flexibility is still allowed in the subsequent simulations.

As I want to attach the N- and C-termini of the linkers to the protein structures, the respective atoms need to be accessible. While this is the case for the structures of N_{rs} and C_{rs}, the relaxed structures of $\mathrm{N}_{\mathrm{flex}}$ and $\mathrm{C}_{\mathrm{flex}}$ are too compact to merge with the protein structures. They both contain the $(\mathrm{GGS})_4$-linker which is known to have a random coil like structure and being flexible [137]. To obtain a good starting point for the generation of the merged structures later, I perform AMBER99 simulations with a pulling force for both structures. I pull at N- and C-terminus with a small force to make the termini accessible. In the merged system the proteins are attached to the termini and also apply a force on the linkers, so this seems to be a valid procedure. For $\mathrm{N_{flex}}$ and $\mathrm{C_{flex}}$ I extract four different structures from different points in the simulations, respectively, and use them for the next step.

4.6.5 Linker Parameters

The linker behavior is not well known but still a crucial factor for the dynamics of the system and therefore the resulting FRET intensity ratios. Hence, to obtain a reference for the linker flexibility, I perform further simulations, starting from the last structures of the AMBER99 relaxation runs. I simulate all systems for 500 ns in an AMBER99 force field with explicit water. Additionally, I simulate the same structures (and the structures from the simulation with a pulling force if present) in an SBM with temperatures ranging from $T = 40$ to $T = 150$. Applying the temperature comparison procedure (see Sec. D) almost the entire SBM temperature range shows to have equal agreement with the AMBER99 simulation [data not shown]. As the fitting allows for a lot of freedom for small structures like the linkers, this method does not seem to give sufficient information to select a temperature for the linkers.

An important characteristic of a linker in this system is its end-to-end distance, as it plays a huge role in the distance between the two fluorophores. So I compare the means and standard deviations of the end-to-end distances from AMBER99 and SBM simulations for all structures.

$\mathrm{N_{rs}}$ and $\mathrm{C_{rs}}$ are rather rigid, as they only consist of restriction sites. I check if a choice of $T = 70$ as determined for the proteins is valid. For both there is no big change in the mean or standard deviation of the end-to-end distance as a function of T and both are close to the values from the AMBER99 simulation. Given the little information available about these structures, assuming a temperature of $T = 70$ seems to be an adequate starting point allowing for reasonable flexibility.

For $\mathrm{N_{flex}}$, attaching the completely relaxed structure to the proteins is not possible due to inwardly rotated termini. Simulations of the strongly extended structures show an immoderately high mean distance over all temperatures. In this case I choose an only slightly elongated structure, where the termini are accessible, and which has the best agreement in terms of end-to-end distance at around $T = 140$. As $\mathrm{N_{flex}}$ contains the flexible linker region $(\mathrm{GGS})_4$, I choose a temperature of $T = 140$ for the $(\mathrm{GGS})_4$-linker and the protein temperature of $T = 70$ for the restriction site residues in all further simulations.

Although the relaxed structure of $\mathrm{C_{flex}}$ can be attached, the attachment does not have a lot of conformational freedom. Thus, the result might not resemble the physical structure. On these accounts, I choose two more elongated structures, $\mathrm{C_{flexB}}$ and $\mathrm{C_{flexC}}$, besides the relaxed structure $\mathrm{C_{flexA}}$ and use all of them in further

Table 4.7: Parameters for construction of the different sensors. N- and C-terminal linkers are given (see also Tab. 4.6). The cutoff RMSD for the selection of structures after attaching the fluorescent proteins and the number of resulting structures are shown.

Sensor	N-terminal	C-terminal	$\text{RMSD}_{\text{cutoff, FP}}$	# structures
Sensor 1	N_{rs}	C_{rs}	1.2 nm	109
Sensor 2	N_{flex}	C_{rs}	1.2 nm	127
Sensor 4A	N_{rs}	C_{flexA}	1.0 nm	29
Sensor 4B	N_{rs}	C_{flexB}	1.5 nm	56
Sensor 4C	N_{rs}	C_{flexC}	1.5 nm	93

simulations. Still, the mean end-to-end distances are much higher for the elongated structures than in the reference simulation.

4.6.6 Merging of Sensing Protein and Fluorescent Proteins

Merging the structures of CFP, YFP, and Glc-BP is not as straightforward as for dyes (see Secs. 4.4 and E), as CFP has to be inserted into Glc-BP. It is not known, but assumed for now, that still all residues, including residues 1 to 11, fold into the given Glc-BP structure (see also Sec. 5.9.1).

To merge the structures, I start with the sensing protein Glc-BP. I use the structure of the glucose bound state, as the transition from the closed to the open conformation with attached fluorescent proteins is easier than vice versa. I remove residues 12 and 13 as done in experiments and attach N-terminal and C-terminal linkers with the algorithm described in Sec. E. The C-terminus of the N-terminal linker is attached to residue 14 and the N-terminus of the C-terminal linker to the last structurally resolved residue (residue 303) of Glc-BP. Here, I test all possible orientations using a set of angles α, θ, ϕ in discrete steps of $\alpha, \phi \in [0°, 20°, ..., 340°]$, $\theta \in [0°, 10°, ..., 90°]$ (see also Sec. E) and check for steric clashes. I save all sterically possible conformations.

However, pre-selections are necessary to reduce computational costs later, so I filter for similar structures which could be easily reached mutually within simulations. I start at the first found set of angles for each linker (which is a conformation preferably orthogonal to the surface of Glc-BP) and discard every set which differs less than $\Delta\alpha + \Delta\theta + \Delta\phi < 60°$ from one of the already selected sets. In a second cycle I calculate the mutual RMSD between all constructed structures with both linkers and again discard every structure, which differs less than $\text{RMSD}_{\text{cutoff}} = 0.1\,\text{nm}$ from any of the already chosen structures.

With the remaining structures I proceed by attaching CFP and YFP in the same way. The C-terminus of CFP is attached to the N-terminus of the N-terminal linker and the N-terminus of YFP to the C-terminus of the C-terminal linker. Again, I filter the resulting structures with $\Delta\alpha+\Delta\theta+\Delta\phi < 60°$ and use different values for $\mathrm{RMSD}_{\mathrm{cutoff}}$ as the structures are considerably larger now. I choose $\mathrm{RMSD}_{\mathrm{cutoff,\,FP}}$ slightly different for the different sensors, to end up with around 100 different structures each. The values together with the number of resulting structures are shown in Tab. 4.7. For sensor 4A the linker structure is rather compact as mentioned above and allows for only few conformations of the merged structure.

In a next step, I attach the three missing alanine residues at residue 11 of Glc-BP (see Fig. 4.6) using `pymol` [136]. I subsequently perform an energy minimization for 5000 steps to get rid of possible clashes caused by the insertion of the alanine residues.

The last step consists of two short simulations to connect the alanine residues with the N-terminus of CFP. To enable inclusion of a regular bond potential in the final structure, the target distance for the respective atoms is set to $d_{\mathrm{target}} = 0.14\,\mathrm{nm}$. In these simulations the temperature of the three alanine residues and the N-terminal linker is set to $T = 140$ to allow them to flexibly adjust to the rest of the structure. Also, no contacts between the different parts shown in Fig. 4.6 are included, except the contacts between the first eleven residues and the rest of Glc-BP.

As the two atoms to connect might be distant, I pull them together in two consecutive steps to avoid large forces. I introduce a soft contact potential between the N-terminus of CFP and the carbon atom of the third alanine with a minimum at the target distance, an extra broad width of $\sigma = 2\,\mathrm{nm}$, and a force constant of $K_{\mathrm{c}} = 40\,\epsilon$ and simulate for 1000 ps in the SBM. I confirm that Glc-BP stays completely folded in the simulation by checking the RMSD of Glc-BP. Then, the distance of the respective atoms is calculated over time and the structure with the distance closest to the target distance is extracted.

With this structure I proceed by introducing a weak bond potential instead of the contact potential with a minimum at the target distance and a force constant of $K_{\mathrm{b}} = 80\,\epsilon/\mathrm{nm}^2$. I perform this simulation in the same way as before for a total time of 500 ps in the SBM. For the rare case the distance does not reach $d < 0.3\,\mathrm{nm}$, the desired bond might be sterically inhibited and the structure is discarded. Otherwise, the structure with the distance closest to the target distance is extracted and used for further simulations.

This procedure results in a wide ensemble of possible structures for each sensor. To narrow the conformational space, I fit the resulting structures to experimental SAXS data [group of A. Stadler, private communication] of the respective sensor in the glucose bound state using `CRYSOL` [74]. Then I sort the structures according

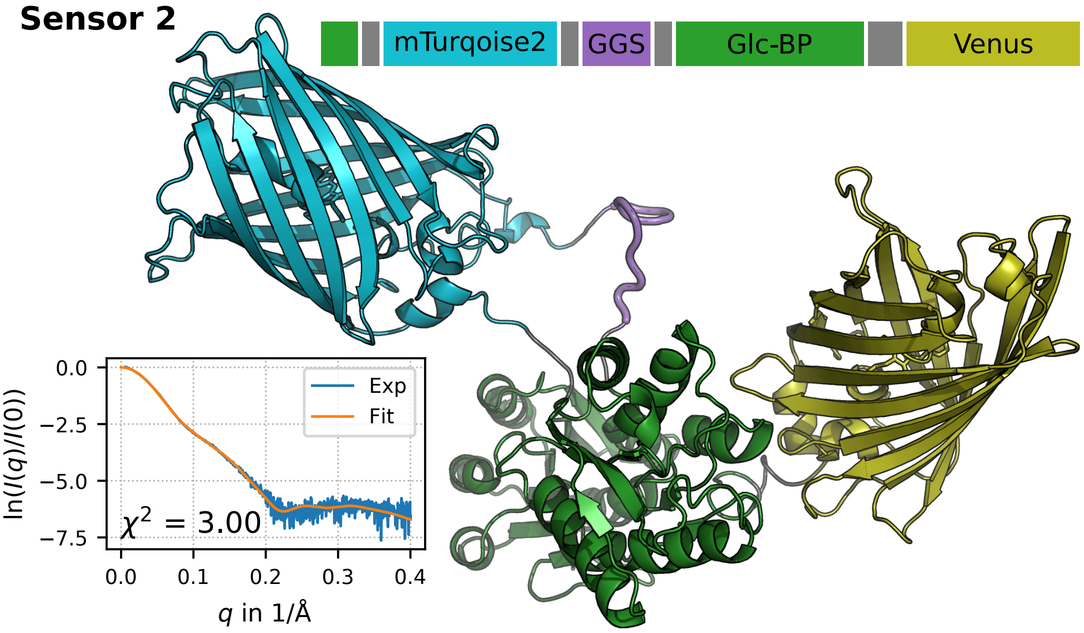


Figure 4.9: Schematic sequence and exemplary resulting structure for sensor 2. Glc-BP is shown in green, CFP in cyan, and YFP in yellow. The flexible $(GGS)_4$-linker is depicted in purple. The experimental SAXS intensity curve (blue) and a fit of the theoretical curve of the shown structure to the experimental data (orange) are shown in an inset along with the χ^2 value.

Table 4.8: Parameters of the fluorophores SWG and CR2. Lifetimes τ ([1][134], [2][138]), quantum yields Q ([3][132]), rotational correlation times τ_{rot} ([4][H. Höfig, private communication]), and Förster radius R_0 ([5][139]) are given for the fluorophore in CFP (SWG) and YFP (CR2), respectively.

Donor	Acceptor	τ	Q	τ_{rot}	R_0
SWG		$4.0\,\text{ns}^1$	0.90^3	$20.85\,\text{ns}^4$	$4.9\,\text{nm}^5$
	CR2	$3.0\,\text{ns}^2$	0.59^3	$\sim 20\,\text{ns}$	

to their χ^2 value (see Sec. 2.5) and choose the best fitting structures to perform the final simulations. As an example, the best fitting structure with $\chi^2 = 3.00$ for sensor 2 is shown in Fig. 4.9 along with the schematic sequence and the SAXS fit.

4.6.7 Simulation Protocol for Fluorescent Proteins

All simulations of the glucose sensors are performed with `GROMACS v4.5.4` [84] with the extension for Gaussian contact potentials [104], an SBM potential as given in Eq. (3.10), and Langevin dynamics (see Sec. 3.1.1). The time step and the temperature coupling constant are set to $\Delta t_{\text{SBM}} = 0.5\,\text{fs}$ and $\tau_T = 0.1\,\text{ps}$, respectively. A total time of $t_{\text{tot, SBM}} = 1000\,\text{ns}$ is simulated. The system is separated into different temperature groups, where only the temperature of the $(\text{GGS})_4$-linker is set to $T = 140$, while the rest of the system (proteins and restriction sites) is coupled to a temperature of $T = 70$.

The time scale can not be adjusted as done for the systems with small dyes (see also Sec. 4.5) because the rotational correlation times τ_{rot} of the fluorescent proteins are on a similar time scale as the rotational motion of Glc-BP. Therefore, the rotations of the fluorescent proteins are superposed by the rotational motion of Glc-BP in experiments and can not be measured in single molecule experiments in buffer solution. However, values for τ_{rot} of the free fluorescent proteins are available. The rotational correlation time of GFP is often estimated with $\tau_{\text{rot}} \sim 20\,\text{ns}$. For free CFP it was measured as $\tau_{\text{rot}} = 20.85\,\text{ns}$ [H. Höfig, private communication]. The rotational correlation time in free solution is mainly dependent on the friction with the solvent, so a comparison of SBM simulations of a freely diffusing fluorescent protein with the experimentally measured values should yield a friction which mimics the physical behavior. To find a reference time scale for these systems, I perform simulations for $t_{\text{tot, SBM}} = 50\,\text{ns}$ with free CFP and YFP including the fluorophores. Then I calculate the free rotational correlation times using Eqs. (2.12) and (2.13) and use the ratio between experimental and simulated values as conversion factor to adjust the simulation time scale. With this time scale I perform Monte Carlo photon simulations as described in Sec. 3.3 with

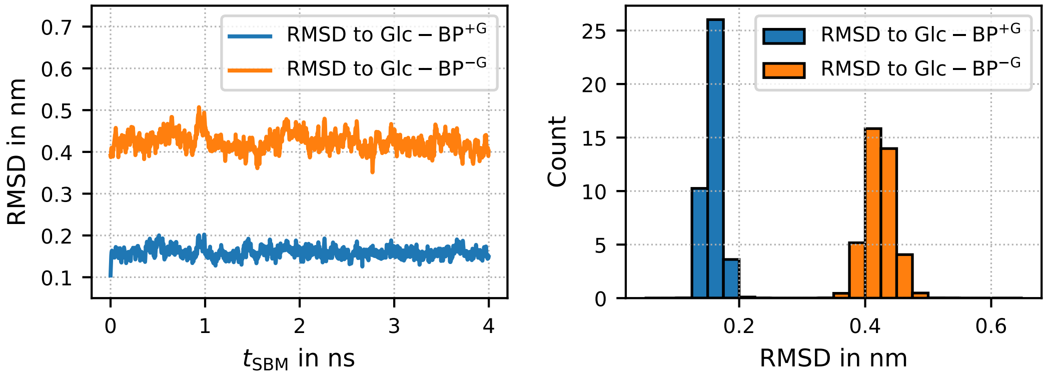


Figure 4.10: Exemplary RMSD values from simulation of sensor 2 with Glc-BP in its glucose bound state. The RMSD values of Glc-BP with reference to Glc-BP^{+G} (blue) and Glc-BP^{-G} (orange) are shown for the start of the simulation (left) and as distributions for the entire simulation (right). The conformation is stable over the entire simulation.

the parameters given in Tab. 4.8. I use the whole simulations, which correspond to around 100 µs on the physical fluorophore time scale.

Now I compare the results of Glc-BP^{+G} with respective results of Glc-BP^{-G} with the same configuration of fluorescent proteins. For the simulation of Glc-BP^{-G}, I use the same starting structure, but replace the parts of the force field (bond, angle, dihedral angle, and contact parameters) belonging to Glc-BP^{+G} with the respective values of Glc-BP^{-G}. To test if the conformation changes accordingly, I analyze the RMSD of the Glc-BP part with respect to both conformations for the simulation of Glc-BP^{+G} and Glc-BP^{-G} (see Figs. 4.10 and 4.11). The conformation of Glc-BP^{+G} stays stable during the simulation (see Fig. 4.10). For the simulation of Glc-BP^{-G}, Fig. 4.11 shows that the transition to the target state occurs in the beginning of the simulation and also stays stable, although the distributions are broader. Comparing these results to the fluctuations in the ground state discussed in Sec. 4.6.3, the procedure seems reasonable.

4.6.8 Dimerization of Fluorescent Proteins

GFP in its wild type has a tendency to form dimers which can be observed in the crystal structure [51, 133] (see Fig. 4.12). A mutation of residue 206 from the hydrophobic alanine to the charged lysine (A206K) is known to suppress this weak dimerization [140]. In attempts to optimize a CFP-YFP FRET pair, further mutations of residues 208 (from the hydrophilic serine to the hydrophobic phenylalanine, S208F) and 224 (from valine to leucine, V244L) have shown improvement in the FRET signal changes [141]. Although the authors of this work

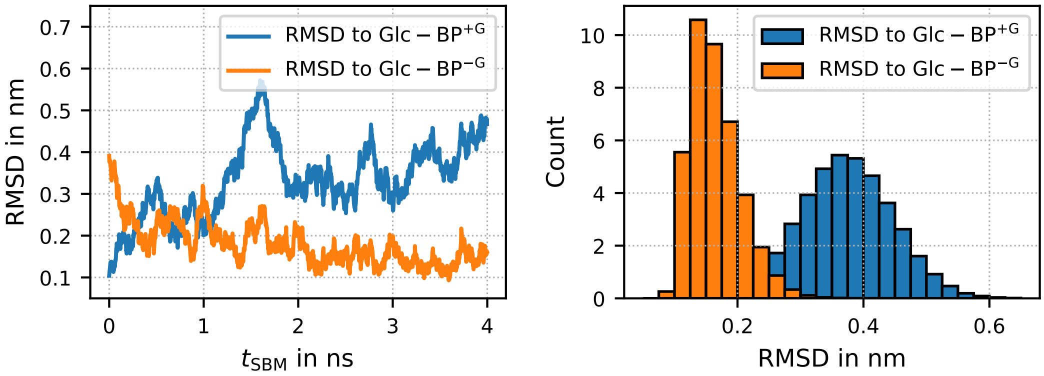


Figure 4.11: Exemplary RMSD values from simulation of sensor 2 with Glc-BP in its glucose free state. The RMSD values of Glc-BP with reference to Glc-BP^{+G} (blue) and Glc-BP^{-G} (orange) are shown for the start of the simulation (left) and as distributions for the entire simulation (right). A transition from the starting structure (Glc-BP^{+G}) to the target structure (Glc-BP^{-G}) occurs directly at the beginning of the simulation.

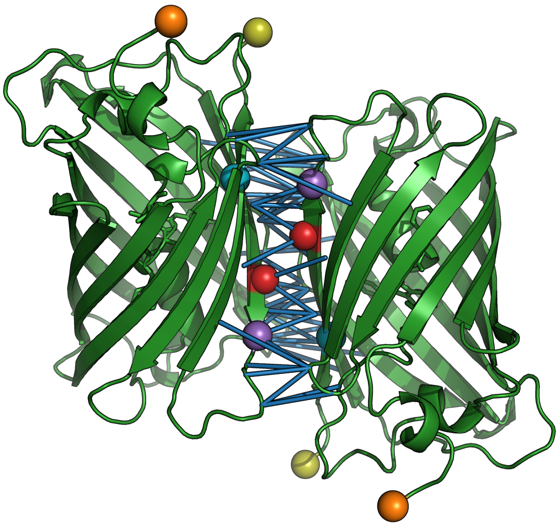


Figure 4.12: GFP in its dimer conformation as observed in the crystal structure (green, PDB: 1GFL [51]). The N-termini and C-termini are depicted in orange and yellow, respectively, to show the antiparallel configuration of the two GFP monomers. The 50 C$_\alpha$-contacts included in the simulation are illustrated by blue lines. The residues of the A206K mutation are depicted in red, the residues of the S208F and the V224L mutations are depicted in purple and cyan, respectively.

have excluded dimerization of the two fluorescent proteins as a mechanism [141], later work has found evidence for the formation of an intramolecular complex caused by the two mutations S208F and V224L [46, 142]. They have seen a signal increase due to enhanced dimerization [46] and a successive decrease in the signal when monomerizing one or both fluorescent proteins with the A206K mutation. These findings have lead to new design approaches for FRET sensors based on mutually exclusive domain interactions, where dimerization is possible in one state while prohibited in the other state [143]. Furthermore, it has been found that reversible intramolecular interactions as this heterodimerization are important for creation of FRET sensors with a large dynamic range [144, 145]. The dimerization, however, has to be critically balanced as overdimerization as well as overmonomerization can restrict the FRET range [144].

Now the question arises, what role dimerization plays in the glucose sensor studied here. Of the fluorescent proteins used in this work, CFP contains the A206K mutation which should prevent dimerization, whereas YFP lacks this mutation. It is not known whether the CFP-YFP pair can form a heterodimer in the studied glucose sensor. I want to investigate the effects of a possible temporary dimerization on the results, so I implement dimer contacts into the simulations in the following way.

I determine all atom-atom contacts in the dimer structure of GFP between the two monomers according to the description in Sec. 3.2.2. The contacts are almost symmetrical between the two structures as their interface consists of the same residues in both structures. To generalize these contacts, I translate them into residue-residue contacts and choose the mutual ones determined in both directions. I implement these contacts between the C_α-atoms of the respective residues with a distance in the ground state corresponding to the mean distance of both C_α-atom pairs in the GFP dimer structure. For a depiction of the included 50 contacts see Fig. 4.12. Depth and width of the contact potential then can be varied to analyze the effects on the simulated structures. It should be noted that some of the contacts in the dimer crystal structure are caused by the crystallization and are probably not present in solution. This is neglected here, but could be considered in future work.

4.7 Summary

This chapter described the implementation of the whole FRET process in simulations. I discussed a systematic way to obtain structures and parametrization with a minimal amount of parameters for small organic dyes and proteins. The simulation temperatures turn out to be essential for adjusting the respective dynamics to be comparable to experimental data. The chapter showed how to merge dyes or fluorescent proteins with the proteins under study and generate starting structures of the whole systems for the simulations. For the glucose sensor, I used additional experimental SAXS data to select the most suitable structures. In addition, I described an approach to enable dimerization of the two fluorescent proteins in simulations to study its effects.

Finally, I presented detailed simulation protocols for two-color and three-color FRET with small dyes as well as with fluorescent proteins. Parametrization procedures and simulation protocols are in principle applicable to arbitrary systems involving dyes or fluorescent proteins.

5 Simulation Results

This chapter presents the results for the simulations of different dye-labeled proteins and the glucose sensor. I start with test simulations of CI-2 in Sec. 5.1 to demonstrate how the simulation method can provide new insights into the system by yielding distance and orientation distributions. Furthermore, in Sec. 5.2 the effects of (in-)sufficient sampling are discussed.

In Sec. 5.3, CspTm is studied as an example of an experimentally investigated system with different labeling positions. Subsequently, experimental results are compared to simulated data in Sec. 5.4, showing the high agreement with experiments and the validity of the presented approach. All these results are published in my publication [113].

An application to the system of ClyA, including two- and three-color FRET, as well as new observations regarding differences observed in experiments between distinct conformations of this system are presented in Sec. 5.5. The presented method enables to study the underlying dynamics of this system and helps in the interpretation of the experimental measurements of labeled ClyA.

Sec. 5.6 compares the presented approach to simple models of data analysis, the accessible volume approach for folded proteins and a polymer model for unfolded proteins. The developed model is consistent with both, leading to a single model suitable for investigating both folded and unfolded proteins.

As the interplay of FRET and SAXS measurements is a highly discussed topic, I show the influence of FRET dyes on SAXS intensity profiles and compare different methods of calculating the radius of gyration in Sec. 5.7.

Sec. 5.8 shows for different systems how simulations can be employed to obtain quantitative parameters, using the example of the diffusion constant.

Finally, I present the simulation results for studying different variants of a glucose sensor in Sec. 5.9. I want to obtain insight in the underlying mechanisms responsible for certain configurations showing higher FRET efficiency ratios than others. This insight facilitates better understanding and improving the function of this sensor in the future.

5.1 Test Simulations with CI-2

As an example test system, I simulate the protein CI-2 with two different dye pairs, AF488 and AF594, and AF546 and AF647, attached to residues 20 and 78, respectively (published in [113]). The results are shown in Fig. 5.1.

Figs. 5.1a,d show the distributions of inter-dye distances R_{DA} and C_α distances for folded and unfolded CI-2. As expected, the unfolded protein has much broader distance distributions than the folded protein. Also the C_α distance distributions for the folded states clearly differ from the R_{DA} distributions. This result shows that the distribution of R_{DA} for the folded states is dominated by dye dynamics. The inter-dye distance distributions for both dye pairs are similar, but the dye pairs differ in Förster radius. As the ideal Förster radius to obtain best-separated states should be between distances of folded and unfolded states, the distributions and mean distances imply that the pair AF488/AF594 is more suitable for studying CI-2 experimentally.

The distributions of the orientation factors κ^2 for folded and unfolded conformations are shown in Figs. 5.1b,e. They are all similar and the mean values for unfolded proteins $\overline{\kappa^2}_{\text{U}}$ are in good agreement with the approximation of freely rotating dyes with no steric restrictions ($\overline{\kappa^2} = 2/3$). For the mean values in the folded state $\overline{\kappa^2}_{\text{F}}$, however, I observe slight deviations of 2-5%. These deviations probably are due to steric restrictions imposed by the protein.

Figs. 5.1c,f show the resulting FRET efficiency histograms. They are fitted with a Gaussian curve and the mean efficiency values for folded states $\langle E\rangle_{\text{F}}$ and unfolded states $\langle E\rangle_{\text{U}}$ are given. Using log-normal distributions for the fits or the median instead of the mean efficiency has only negligible effects on the quantitative results. The presented method further allows to easily test for different hypothetical Förster radii.

5.2 Effects of the Sampling Length

To ensure sufficient convergence of the simulations, I carefully investigate the effect of the sampled simulation length [113]. Here, I exemplarily look at the FRET efficiency and the distance distributions for different lengths of the simulation of CI-2 with the dye pair AF488/AF594. The total simulation length is denoted by t_{tot}. Fig. 5.2a shows the mean inter-dye distances $\overline{R}_{\text{DA}}$ and the peak positions of

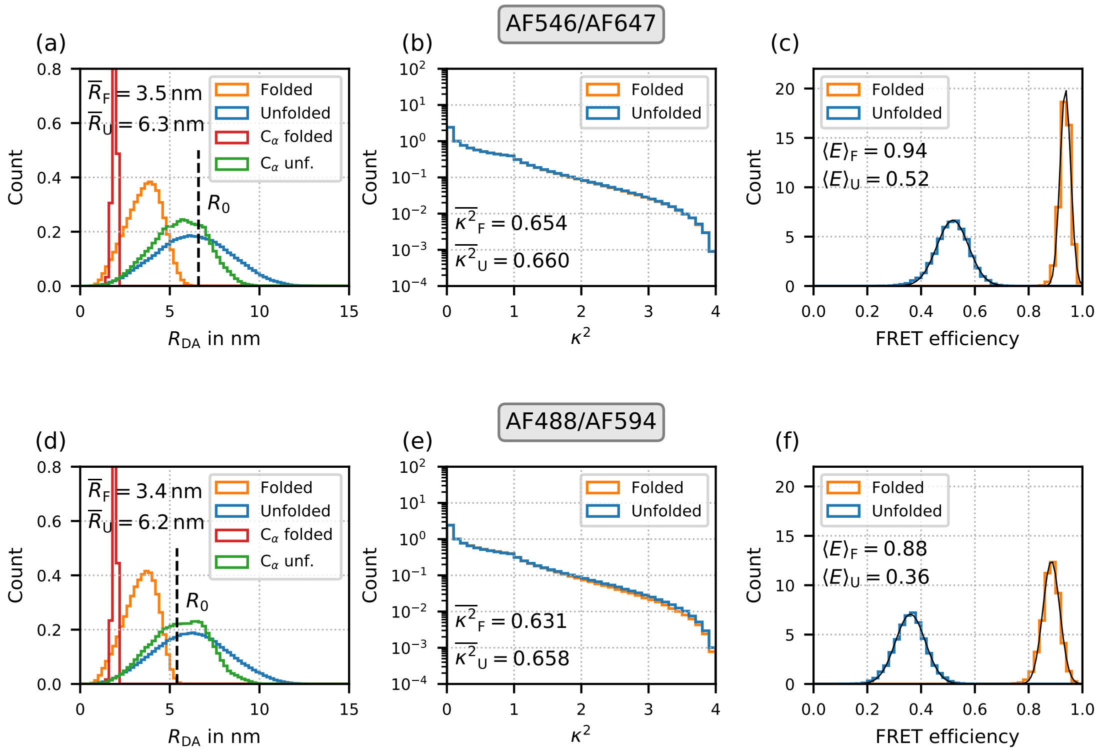
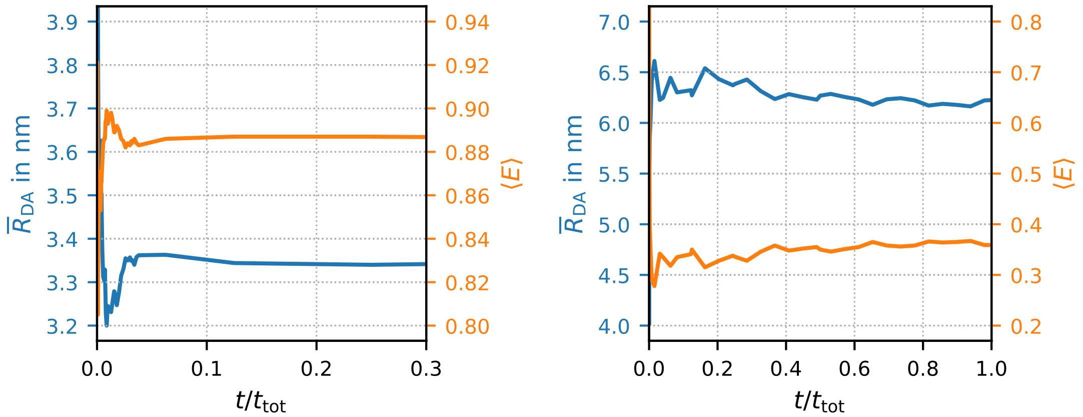


Figure 5.1: Simulation results for CI-2 labeled with AF546 and AF647 (top) and AF488 and AF594 (bottom) [113]. Donor and acceptor dye are attached to residues 20 and 78, respectively. The results for folded and unfolded states are depicted in orange and blue, respectively. (a) and (d) show the distributions of the inter-dye distances along with distributions of the respective C_α distances for folded (red) and unfolded states (green). The Förster radii R_0 are shown as dashed black lines. The mean distances for folded state ($\overline{R}_F$) and unfolded state ($\overline{R}_U$) are given. (b) and (e) show the distributions of the orientation factor κ^2 and the mean values for folded ($\overline{\kappa^2}_F$) and unfolded states ($\overline{\kappa^2}_U$). (c) and (f) show the resulting FRET efficiency histograms with Gaussian fits (black lines) and the peak positions of the fits for folded ($\langle E\rangle_F$) and unfolded ($\langle E\rangle_U$) states.



(a) Folded states.

(b) Unfolded states.

Figure 5.2: Mean inter-dye distances $\overline{R}_{\mathrm{DA}}$ (blue) and peak positions of the Gaussian fit for the FRET efficiencies $\langle E\rangle$ (orange). The values are given for the parts of the simulation used in the calculation. The time t is given with respect to the total simulated time t_{tot} for simulations of the folded protein (left) and simulations of the unfolded protein (right).

the Gaussian fits for the FRET efficiency $\langle E\rangle$ over time for CI-2 in the folded conformations. The values for both quantities already stabilize in the first 10% of the simulation. For the simulation of unfolded CI-2, where chain dynamics plays a crucial role, the respective values stabilize much later after about half of the considered simulation (shown in Fig. 5.2b).

As the mean values do not contain all information about the underlying distributions, the distance distributions for the folded state are given in Fig. 5.3 for different lengths of the simulation, ranging from $t_{\mathrm{tot}}/128$ to the whole simulation t_{tot}. Even though the mean distance stabilizes early, the distribution still varies slightly between $t_{\mathrm{tot}}/32$ and $t_{\mathrm{tot}}/8$. The variation in the FRET distributions for the folded state is negligible [data not shown]. Clearly, the folded state is rather uncomplicated regarding sampling issues. Nevertheless, mean values should be handled with care as the distributions might still vary.

In Figs. 5.4 and 5.5, distance distributions and efficiency histograms are shown for the unfolded state, respectively. Here, the distributions still vary a lot. CI-2 is a small system and seems to be sufficiently sampled. For larger systems, especially in the unfolded state, the simulation length has to be adjusted.

5.3 CspTm with Different Dye Positions

The next system I investigate to test the influence of different labeling positions is CspTm [113]. I perform simulations of CspTm with AF488 and AF594 at dif-

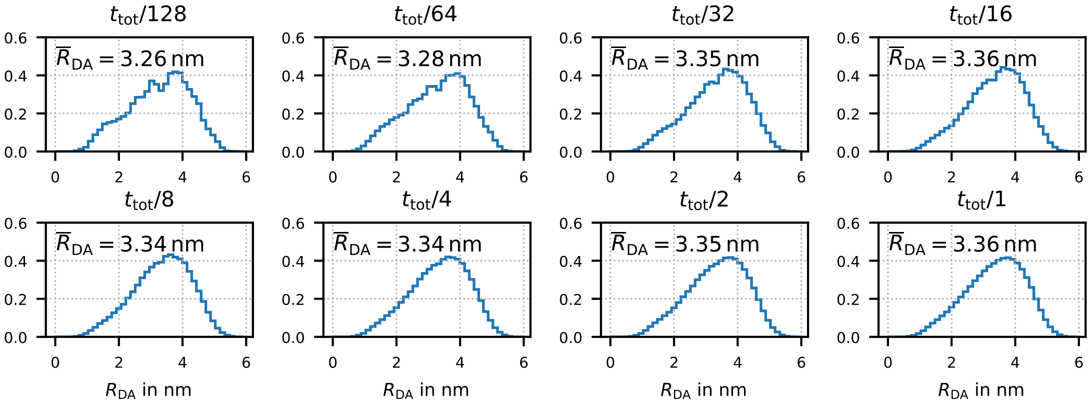


Figure 5.3: Inter-dye distance distributions of folded states for different fractions of the total length of the simulation t_{tot}. Additionally, the respective mean inter-dye distances are given.

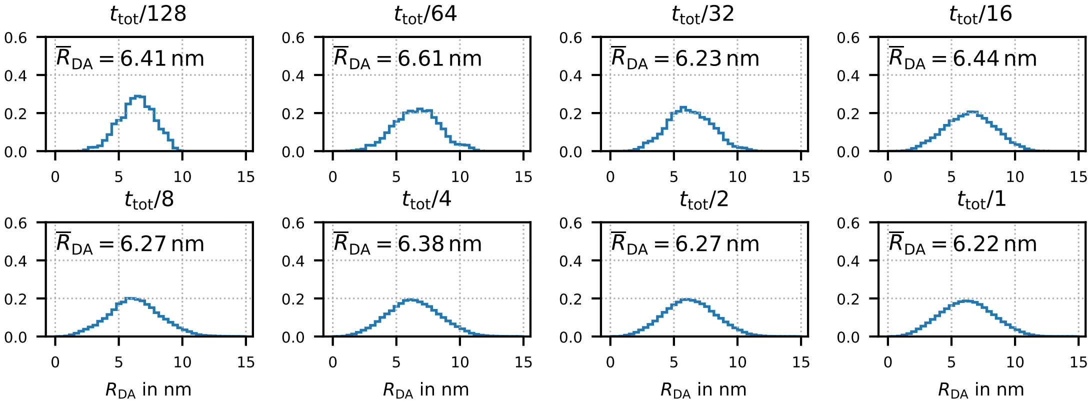


Figure 5.4: Inter-dye distance distributions of unfolded states with respective mean values $\overline{R}_{DA}$ for different fractions of the total length of the simulation t_{tot}.

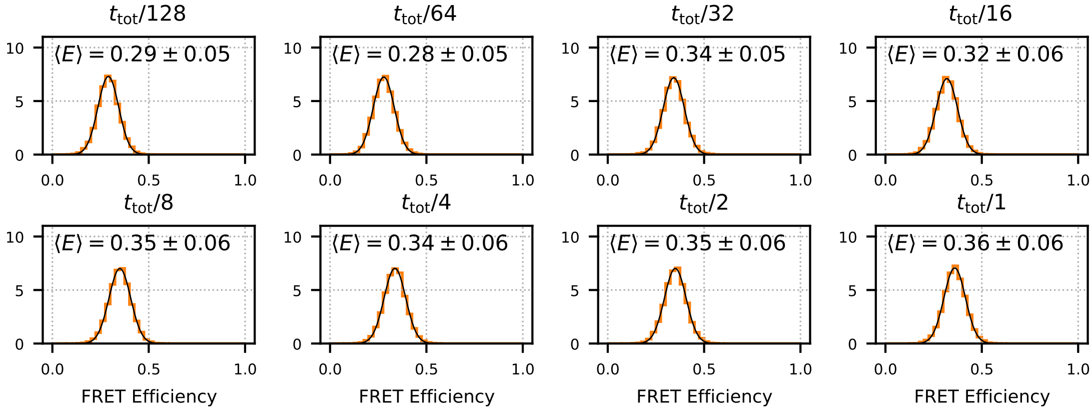


Figure 5.5: FRET efficiency distributions of unfolded states with respective peak positions $\langle E \rangle$ and standard deviations of the Gaussian fits (black lines) for different lengths of the simulation with total length t_{tot}.

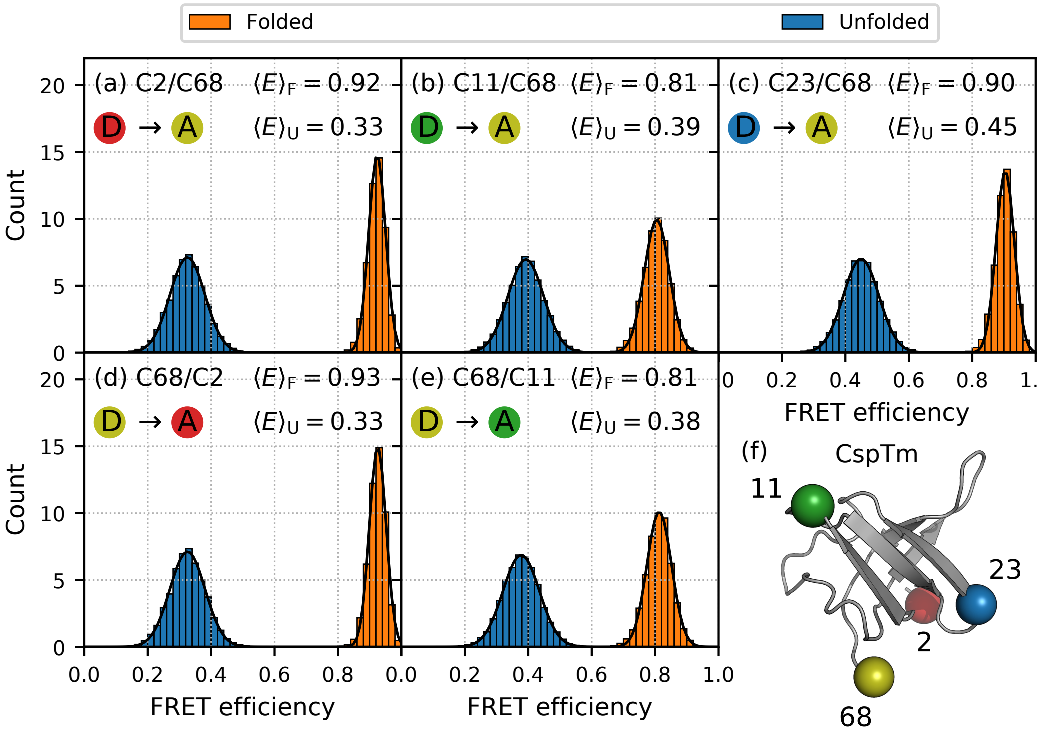


Figure 5.6: FRET efficiency distributions for CspTm labeled with AF488 (donor) and AF594 (acceptor) at different labeling sites [113]. Donor (D) and acceptor (A) positions are color coded in (a), (b), (c), (d), and (e). The colors refer to different residues, namely residue 2 (red), 11 (green), 23 (blue), and 68 (yellow). The positions are shown in (f). Residue numbering within the protein is given in Sec. H. The efficiency distributions are shown with the peak values of the Gaussian fits (black lines) for folded states (orange, $\langle E \rangle_{\text{F}}$) and unfolded states (blue, $\langle E \rangle_{\text{U}}$).

ferent labeling sites, which are also used in experiments [146, 147]. The positions are the residue pairs C2/C68, C68/C2, C11/C68, C68/C11 and C23/C68 for donor/acceptor (for the residue numbering scheme, see Sec. H), illustrated in Fig. 5.6f. Figs. 5.6a-e show the resulting FRET efficiency histograms. Unsurprisingly, the peaks for the unfolded states shift to higher efficiencies with shorter sequence separation of the labeled residues, while the dye permutations with respect to the attachment point have no effect.

The results presented in Secs. 5.1 and 5.3 prove the strength of the presented simulation method. It establishes a way of directly relating FRET efficiency distributions to distance distributions. Further, it allows to vary and test different parameters as dye pair, Förster radius, linker length, or labeling sites, which facilitates improvement in planning, interpretation, and validation of experimental results.

5.4 Comparison of Simulation and Experiment

For a validation of the model, I compare the resulting FRET efficiency histograms directly against experimental data for four different systems [113]. Fig. 5.7a shows results for ^{10}FNIII with the dye pair AF546 and AF647. Figs. 5.7b-d show results for CspTm with AF488 and AF594 at different labeling positions. The experimental data was measured for the protein in 0.0 M GdmCl (folded) and in 4.63 M and 7.0 M GdmCl (unfolded) for ^{10}FNIII [113] and CspTm [147], respectively. The experimental efficiency histograms are already corrected for background, different quantum yields of donor and acceptor, different detection efficiencies, crosstalk, and direct acceptor excitation. Only different quantum yields are reflected in my simulation protocol and also corrected for in the simulated efficiency histograms.

Simulated and experimental data are in high agreement except for the system CspTm C11/C68 shown in Fig. 5.7c. The transfer efficiency of the folded state in the simulation of CspTm C11/C68 is shifted to lower values compared to the experimental measurements. In this system the dyes are attached on opposite sides of the protein, which explains a lower FRET efficiency compared to the simulations of the other two labeling schemes. This deviation could be explained by residual attractive interactions between dyes and protein surface, which are not reflected by the simulations. This aspect can be tested by detailed time-resolved fluorescence anisotropy measurements and comparison of the different systems.

In Tab. 5.1 means and standard deviations of the Gaussian fits of the efficiency distributions from Fig. 5.7 are summarized. In the simulations, the widths of the efficiency distributions are dominated by shot noise caused by the limited number of photons collected for each burst.

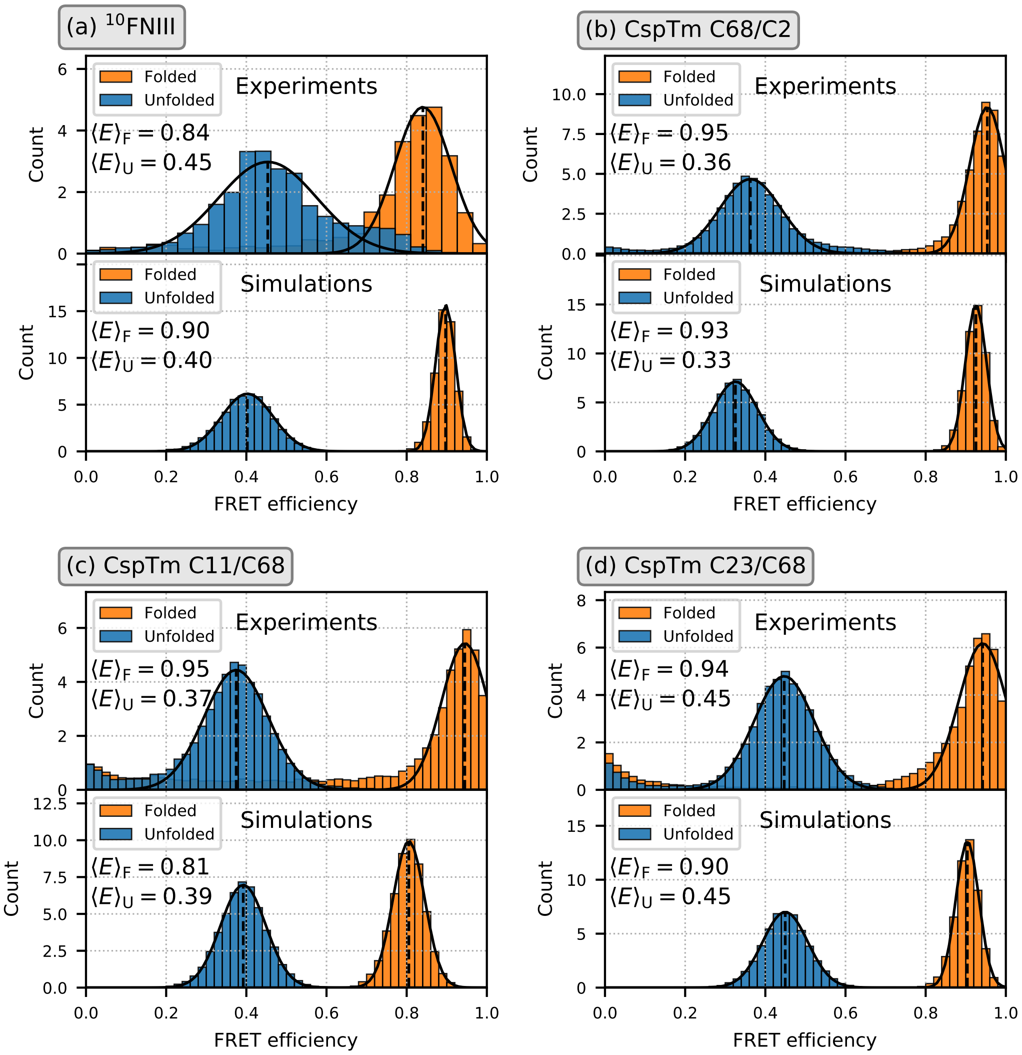


Figure 5.7: Comparison of FRET efficiency histograms from simulations and experiments [113]. (a) Results for ^{10}FNIII with AF546 and AF647 attached to residues 11 and 86. (b), (c), (d) Results for CspTm with AF488 and AF594 attached to (b) residues 68 and 2, (c) residues 11 and 68, (d) residues 23 and 68, respectively [147]. Distributions are given for folded (orange) and unfolded states (blue) and fitted with Gaussian curves (black lines). The peak positions (dashed lines) for folded ($\langle E\rangle_F$) and unfolded ($\langle E\rangle_U$) states are shown for simulations and experiments. The respective standard deviations can be found in Tab. 5.1. Experimental efficiency values below 0.0 and above 1.0 are not shown.

Table 5.1: Results of the Gaussian fits for the FRET efficiency distributions from simulations (sim) and experiments (exp) [113]. Peak positions ($\langle E\rangle_\mathrm{F}$, $\langle E\rangle_\mathrm{U}$) and respective standard deviations (σ_F, σ_U) are given for folded and unfolded states. (a) ^{10}FNIII with AF546 and AF647 attached to residues 11 and 86. (b), (c), (d) CspTm with AF488 and AF594 attached to (b) residues 68 and 2, (c) residues 11 and 68, (d) residues 23 and 68, respectively.

System		Folded		Unfolded	
		$\langle E\rangle_\mathrm{F}$	σ_F	$\langle E\rangle_\mathrm{U}$	σ_U
(a) ^{10}FNIII	exp	0.84	0.07	0.45	0.12
	sim	0.90	0.03	0.41	0.06
(b) CspTm C68/C2	exp	0.95	0.04	0.36	0.08
	sim	0.93	0.03	0.34	0.06
(c) CspTm C11/C68	exp	0.95	0.06	0.37	0.08
	sim	0.81	0.04	0.40	0.06
(d) CspTm C23/C68	exp	0.94	0.06	0.45	0.07
	sim	0.90	0.03	0.46	0.06

Deviations between experiments and simulations can result from the aspect that experiments are additionally influenced by the dye photophysics, e. g. donor and acceptor quenching or blinking. Furthermore, they may deviate due to the necessity to correct for background, crosstalk, and detection efficiencies. The lack of site-specific labeling or other chemical heterogeneity and further experimental artifacts can lead to additional broadening of the FRET efficiency distribution [42, 148]. In experiments, incomplete labeling or photobleaching can lead to a donor-only peak near zero FRET efficiency which is also not present in the simulations.

For CspTm the experimental distribution widths are only slightly larger than the ones from the simulation (see Figs. 5.7b-d). This indicates that the width in the experimental histograms is already close to the shot noise limit.

To conclude, this section demonstrates the validity of the presented simulation approach as it is in excellent agreement with experimental data.

5.5 Interpretation of FRET Measurements of ClyA

The next system I consider is ClyA, where a single chain can adopt a monomer or a protomer conformation with twelve protomers forming a pore. With my simula-

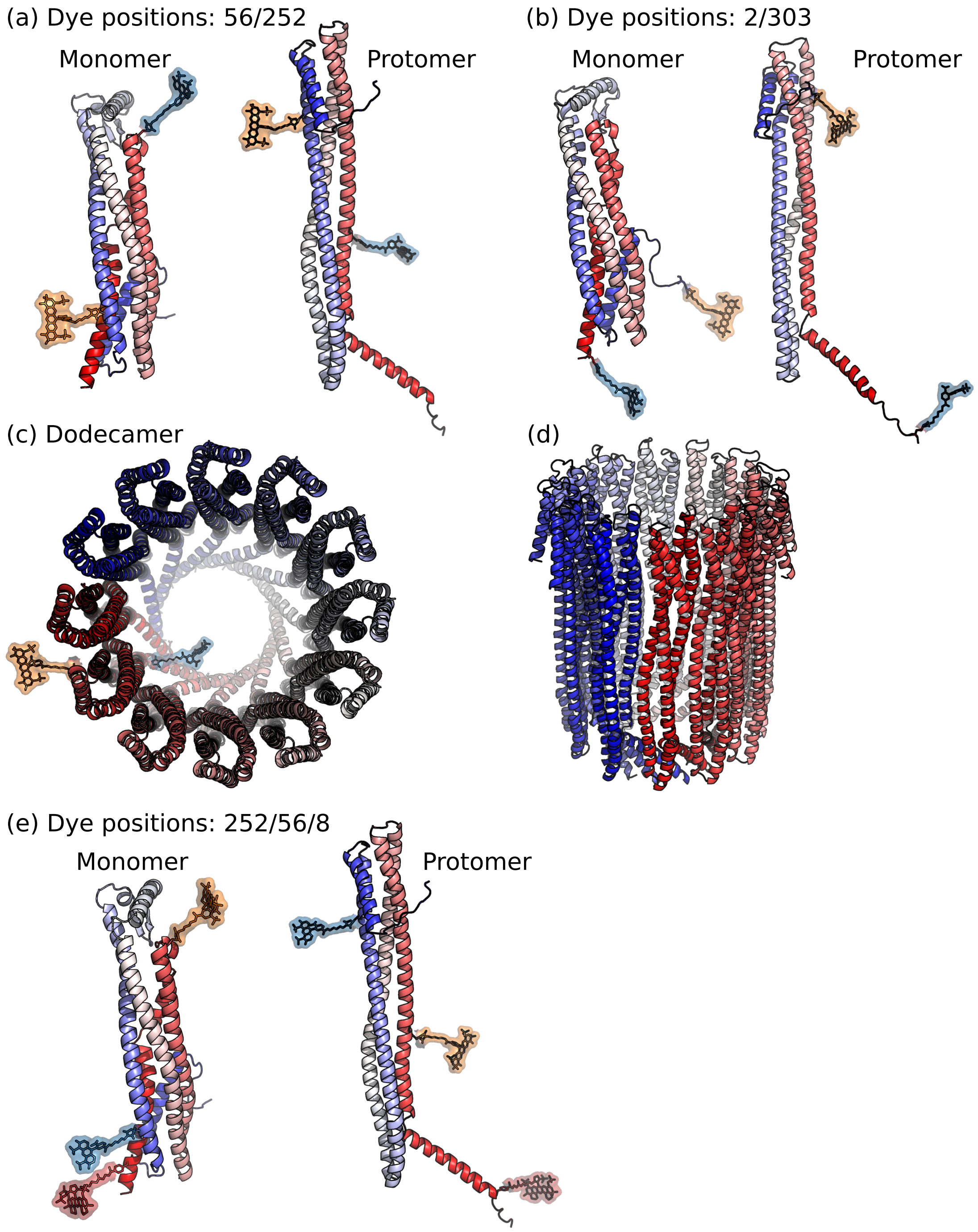


Figure 5.8: Structures of different ClyA conformations with different labeling sites. (a), (b) Monomer and protomer conformations are shown with AF488 (blue) and AF594 (orange) attached to residues 56 and 252, and 2 and 303, respectively. (c) Top view of the dodecamer structure with AF488 and AF594 attached to residues 56 and 252, respectively. (d) Side view of the dodecamer structure. (e) Monomer and protomer structures with the three dyes AF488 (blue), AF594 (orange), and B680 (red) attached to residue 252, 56, and 8, respectively.

tion protocol I can investigate differences in dye behavior for all of these conformations. In experiments, ClyA is studied with the dye pair AF488/AF594 at different positions (residues 56/252 and residues 2/303) for monomer and protomer conformations as shown in Figs. 5.8a,b. In Figs. 5.8c,d the dodecamer conformation with dyes attached to residues 56 and 252 is shown as seen from top and from side without dyes, respectively. In analogy to the experiments [112], additional three-color FRET simulations are conducted with the three dyes AF488, AF594, and B680 at positions 252, 56, and 8, respectively, depicted in Fig. 5.8e.

5.5.1 ClyA in Two-Color FRET Experiments

The FRET efficiency histograms for simulations of ClyA with AF488 and AF594 attached to residues 56 and 252, respectively, are shown in Fig. 5.9. The simulations are performed for monomer and protomer (see also Fig. 5.8a), trimer, dodecamer (see also Figs. 5.8c,d), and unfolded conformations of ClyA.

All histograms from simulations are in high agreement with the experimental measurements. The peak positions for monomer, protomer, and dodecamer in the simulations deviate from the experimental data. However, these efficiencies are in the most sensitive range of FRET. That means small deviations in distance can cause large deviations in efficiencies (see also below). The order of the considered systems in terms of peak efficiencies is identical for simulations and experiments.

Trimer and dodecamer, where for both the labeled protomer is clamped between two other protomers, unsurprisingly result in similar efficiencies. However, they are clearly shifted with respect to the efficiencies of the protomer conformation. This shift is probably caused by steric restrictions due to the adjacent protomer structures in trimer and dodecamer. The steric dye distribution over a simulation for dodecamer and protomer is shown in Fig. 5.10. In the protomer conformation the dyes can obviously adopt many more configurations and are able to move closer to each other than in the dodecamer conformation.

In Fig. 5.11, corresponding C_α and inter-dye distance distributions are shown for the different ClyA conformations. The C_α distance distributions of trimer and dodecamer are almost identical, whereas the distribution of the protomer is slightly different. This difference is likely to be caused by the flexibility of the protomer, as the positions of the C_α-atoms in trimer and dodecamer are more stable. The inter-dye distances of protomer, trimer, and dodecamer are still similar and only slightly shifted. As these distance distributions are close to the Förster radius of $R_0 = 5.4\,\mathrm{nm}$, this slight shift in distances causes a clearly visible shift in the efficiencies depicted in Fig. 5.9.

Tab. 5.2 shows the mean C_α and inter-dye distances of all simulated conformations of ClyA along with the respective standard deviations. For the protomer, the width of the C_α distance distribution is already slightly higher than for trimer

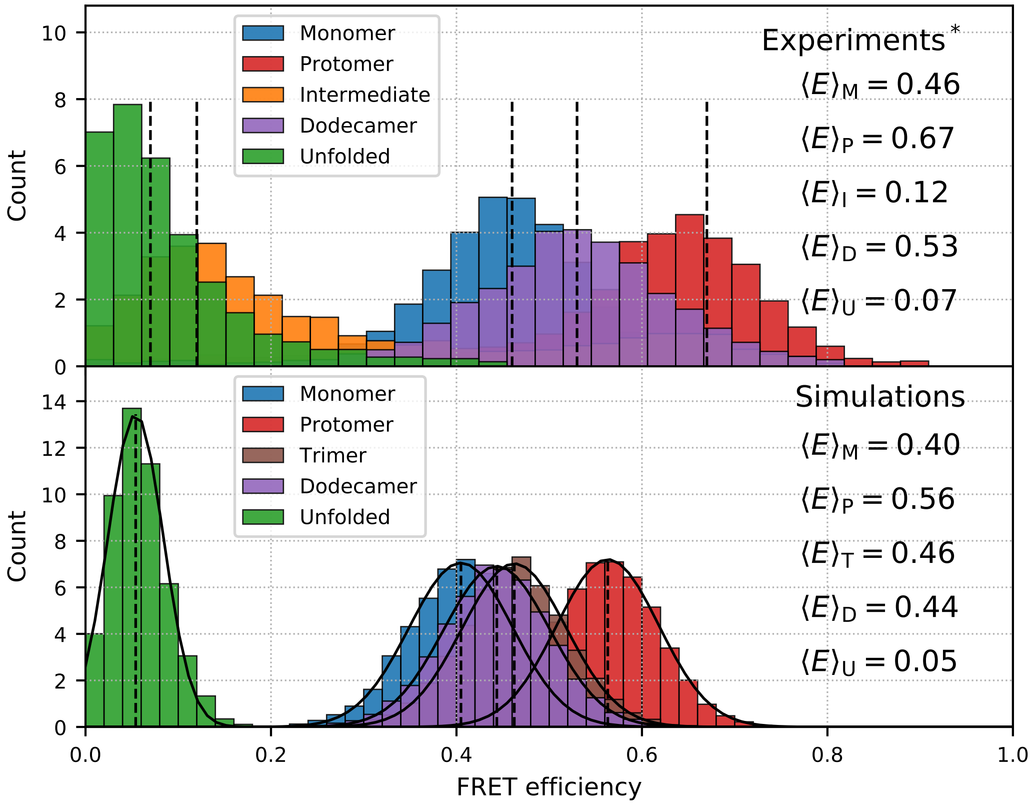
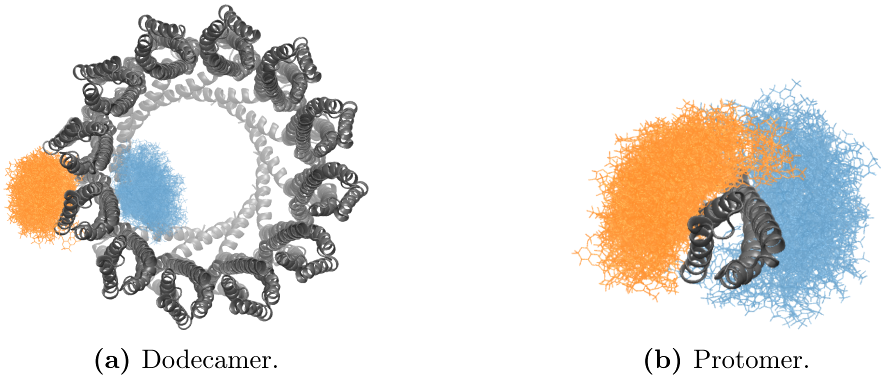


Figure 5.9: FRET efficiency histograms for ClyA with dyes AF488 and AF594 attached to residues 56 and 252, respectively. Shown are the histograms from experimental measurements *[127] (top) and from simulations (bottom) along with the mean values of the Gaussian fits $\langle E \rangle$. Simulations are conducted for monomer (blue), protomer (red), trimer (brown), dodecamer (purple), and unfolded conformation (green). For the experiments, a histogram for the intermediate conformation occurring in the transition from monomer to protomer conformation [127] is shown additionally (orange).



(a) Dodecamer. **(b)** Protomer.

Figure 5.10: Dodecamer and protomer conformations of ClyA (gray). The dyes AF488 (blue) and AF594 (orange) are shown for multiple time points spread over the simulation.

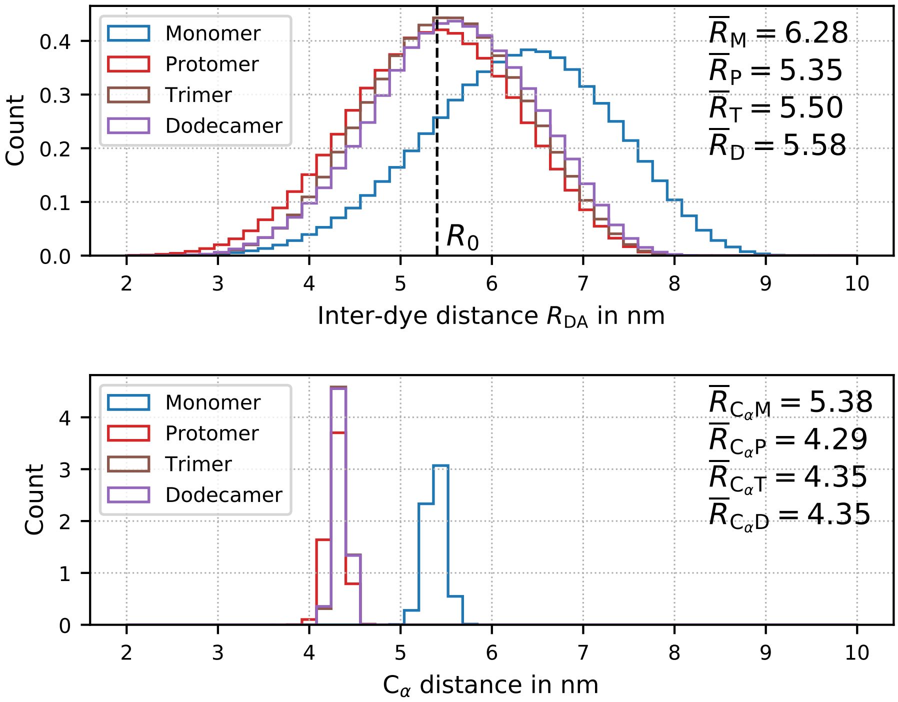


Figure 5.11: Inter-dye distances R_{DA} and C_α distances for different conformations of ClyA. The distributions are shown along with the mean values for ClyA in monomer (M, blue), protomer (P, red), trimer (T, brown), and dodecamer (D, purple) conformations. The Förster radius R_0 of the used dye pair is indicated by a dashed line.

Table 5.2: C_α distances and inter-dye distances for different conformations of ClyA. The respective mean C_α distances ($\overline{R}_{\mathrm{C}_\alpha}$) and standard deviations ($\sigma_{\mathrm{C}_\alpha}$), as well as mean inter-dye distances ($\overline{R}_{\mathrm{DA}}$) and standard deviations (σ_{DA}) for monomer, protomer, trimer, and dodecamer in different conformations with different labeling sites for donor and acceptor (D/A), are given in nm.

System	D/A	C_α distance		Inter-dye distance	
		$\overline{R}_{\mathrm{C}_\alpha}$	$\sigma_{\mathrm{C}_\alpha}$	$\overline{R}_{\mathrm{DA}}$	σ_{DA}
Monomer folded	56/252	5.38	0.11	6.28	1.02
Protomer folded	56/252	4.29	0.10	5.35	0.91
Trimer	56/252	4.35	0.07	5.50	0.85
Dodecamer	56/252	4.35	0.07	5.58	0.87
Monomer folded	2/303	3.98	0.58	5.09	1.24
Protomer folded	2/303	9.97	1.80	10.28	2.34
Monomer unfolded	56/252	11.93	3.99	12.15	4.17
Protomer unfolded	56/252	10.57	3.41	10.85	3.62
Monomer unfolded	2/303	13.10	5.20	13.40	5.35
Protomer unfolded	2/303	14.24	5.01	14.49	5.17

Table 5.3: Parameters for different conformations of ClyA with the same labeling scheme (AF488 and AF594 at residues 56/252). The mean $\overline{\kappa^2}$ value along with rotational correlation times of donor $\tau_{\text{rot, D}}$ and acceptor $\tau_{\text{rot, A}}$ are given. The donor lifetime in absence of an acceptor is set to $\tau_{\text{D}} = 4.1\,\text{ns}$. The donor lifetime in presence of the acceptor τ_{DA} is calculated from fits to simulated data and given in the table. Furthermore, efficiencies calculated from donor lifetimes $E(\tau_{\text{DA}})$ and efficiency histograms $\langle E \rangle$ as well as anisotropy values for donor in absence r_{D} and presence of an acceptor r_{DA} are listed. The value of $\tau_{\text{rot, D}}$ for the monomer (highlighted) is taken from experimental measurements and used for the adjustment of the time scale required for calculation of all remaining values.

Conformation	$\overline{\kappa^2}$	$\tau_{\text{rot, D}}$ in ns	$\tau_{\text{rot, A}}$ in ns	τ_{DA} in ns	$E(\tau_{\text{DA}})$	$\langle E \rangle$	r_{D}	r_{DA}
Monomer	0.665	**0.666**	1.16	2.81	0.32	0.33	0.06	0.08
Protomer	0.662	0.68	1.28	2.14	0.48	0.50	0.06	0.10
Trimer	0.636	0.69	1.25	2.28	0.44	0.46	0.06	0.09
Dodecamer	0.629	0.66	1.17	2.36	0.43	0.44	0.06	0.09

and dodecamer, supporting the earlier statements. In comparison with the dodecamer, the protomer mean inter-dye distance is lower, while the corresponding fluctuations are higher.

5.5.2 Dye Flexibility in Different ClyA Conformations

In the experiments, the shift from protomer to dodecamer is assumed to mainly be caused by acceptor quenching [127], as the efficiency calculated via donor lifetimes yields a better, however not full, agreement between protomer and dodecamer measurements. Tab. 5.3 summarizes additional derived parameters for monomer, protomer, trimer, and dodecamer conformations of ClyA. Once more Tab. 5.3 shows the steric restriction in the dodecamer by a changed $\overline{\kappa^2}$ and additionally differences in the rotational dynamics of the acceptor ($\tau_{\text{rot, A}}$). Although being similar, peak efficiencies determined by donor lifetime $E(\tau_{\text{DA}})$ (see Eq. (I.1)) are systematically smaller than those from efficiency histograms $\langle E \rangle$. The anisotropy r calculated by the Perrin equation (see Eq. (I.2)) in absence and presence of an acceptor is considerably lower than the experimentally measured values with a range of $r = 0.16...0.21$ [127], which hints to additional effects as e. g. sticking of the dyes to the surface. These effects are not accounted for in the present simulation, but could be included by additional attractive interactions.

The results show that the remaining difference in transfer efficiency between protomer and dodecamer could also be caused by the respective steric restrictions of the different conformations.

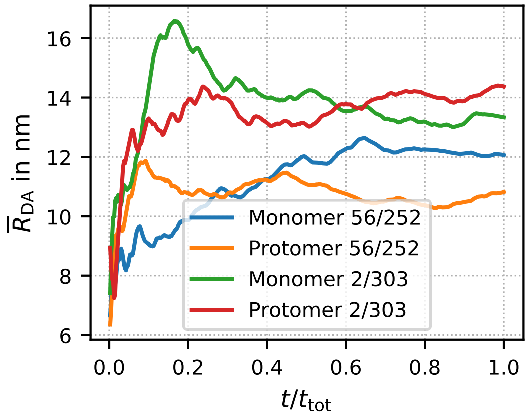


Figure 5.12: Mean inter-dye distances $\overline{R}_{\mathrm{DA}}$ over different parts of the simulations of unfolded ClyA monomer and protomer. The x-axis gives the time t as fraction of the total simulated time t_{tot}. The results are shown for monomer (blue) and protomer (orange) with dyes attached to residues 56 and 252, respectively, and monomer (green) and protomer (red) with dyes attached to residues 2 and 303, respectively.

In Tab. 5.2, furthermore the results for monomer and protomer with dyes attached to residues 2 and 303 are given. Missing parts of the structures were complemented using homology modeling and are rather flexible. This flexibility manifests in the large width of the C_α distances and the even larger width of the inter-dye distances, which has to be considered when using these labeling positions in experiments.

For all simulations of folded systems, the mean inter-dye distances are clearly larger and the distributions broader than the distributions of the C_α distances, which is in accordance with my expectations. Furthermore, the fluctuations are varying, depending on the labeling positions and the surrounding environment, e. g. when directly comparing protomer and dodecamer.

The results presented for ClyA indicate that the introduced method can give valuable new insights into the underlying dynamics of labeled proteins, helping to understand experimental measurements.

5.5.3 Sampling of Unfolded ClyA Conformations

Simulations of unfolded monomer and protomer conformations only differ in the starting structures and in the native contacts included in the potential, which should play a negligible role in the unfolded states. If these simulations sample the conformational space exhaustively, the mean C_α and inter-dye distances are expected to converge to similar values. However, they evidently differ in Tab. 5.2, which is why I test the convergence of the simulation. In Fig. 5.12, the mean distances over parts of the simulation of unfolded monomer and protomer with two different labeling schemes are shown, respectively. In the beginning, protomer and monomer apparently differ strongly and start to converge later on, but still do not fully converge by the end of the simulation. This could be caused either by residual contacts differing in the two structures or insufficient sampling. For a system this large, the conformational space of the unfolded states might not be sampled sufficiently, despite the much longer simulation time as for CI-2 (see also Sec. 5.2).

5.5.4 ClyA in Three-Color FRET Simulations

Finally, the results for three-color FRET for ClyA with dyes AF488, AF594, and B680 attached to residues 252, 56, and 8, respectively (see Fig. 5.8e), are shown in Fig. 5.13. Simulations of monomer, protomer, and unfolded conformations yield clearly different photon rate contributions.

For three-color FRET, different labeling positions, the effect of labeling isomers, and hypotheses regarding the underlying protein dynamics can now be tested. This may improve planning and interpretation of experiments studying these even more complex systems.

5.6 Comparison to Simple Models for Data Analysis

To test the presented simulation method against established methods, I compare the simulation results with different models for data analysis [113]. Fig. 5.14 shows the inter-dye distance distributions for the folded states as well as its relation to the C_α distances in the unfolded states for three different protein-dye systems.

For the folded states, means and standard deviations of the distance distributions from simulations are shown (black crosses and error bars). In addition, the respective values of an accessible volume calculation [18, 19] are shown in blue. The accessible volume approach is based on dye parameters, such as three-dimensional dye extension and linker length, and calculates all sterically possible dye positions within this linkage length at a given labeling position. All dye positions are considered equally probable and the mean distance of the dyes $\overline{R}_{\mathrm{DA}}$ is computed from this distribution. Figs. 5.14a,c,e show that both methods are in good agreement. The simulations have a lower standard deviation, originating from the dynamics in the simulation. In contrast to the accessible volume method assuming all states to be equally probable, the dynamic simulations entropically disfavor dye states close to the protein surface.

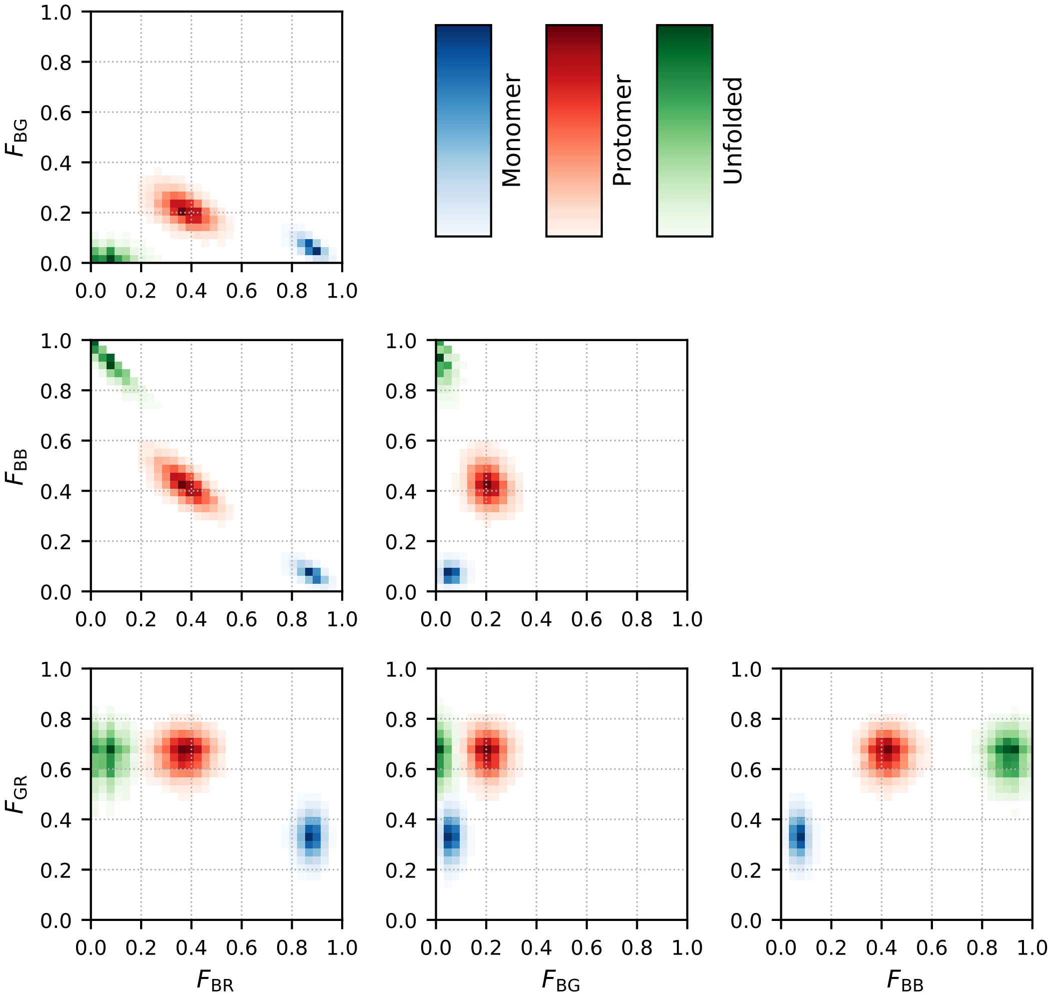


Figure 5.13: Results for three-color FRET simulations of ClyA. In the simulations, the dyes AF488, AF594, and B680 are attached to residues 252, 56, and 8, respectively (see also Fig. 5.8e). The two-dimensional histograms of the photon rates (see Sec. 2.2.4) are shown for monomer (blue), protomer (red), and unfolded conformations (green). Each histogram is scaled to its maximal value. For better visibility of overlapping distributions, values below a small threshold are discarded.

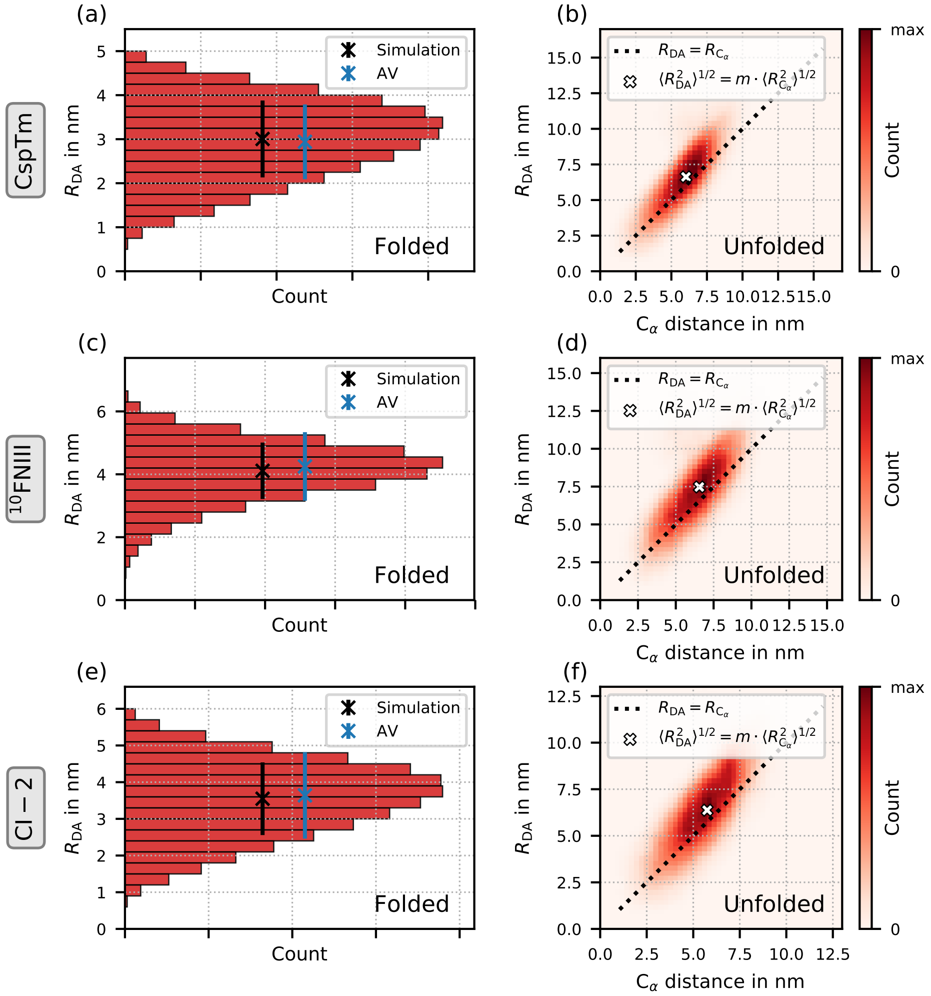


Figure 5.14: Distributions of inter-dye distances R_{DA} and C$_\alpha$ distances of the respective residues [113]. Results are shown for (a), (b) CspTm with AF488 and AF594 at residues 2 and 68, (c), (d) ^{10}FNIII with AF546 and AF647 at residues 11 and 86, and (e), (f) CI-2 with AF546 and AF647 at residues 20 and 78. (a), (c), and (e) show distance distributions (red), mean distances (crosses), and standard deviations (error bars) from simulations (black) and accessible volume calculations (blue) for the folded states. (b), (d), and (f) show two-dimensional histograms of the inter-dye distances and corresponding C$_\alpha$ distances for the unfolded states. The dotted lines show the expected dependency for equality of both values. Expectations corrected by an effective segment length of the chain accounting for the linkers are depicted as white crosses. Each histogram count is scaled individually according to its maximal value.

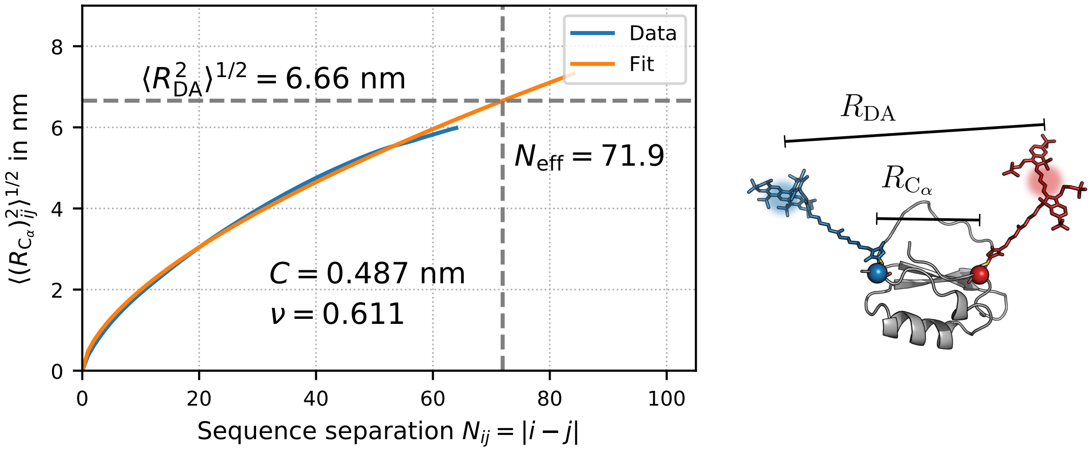


Figure 5.15: Exemplary calculation of effective segment length N_{eff} for CI-2 with AF546 and AF647 at residues 20 and 78, respectively, (left) and depiction of the system (right). The averaged squared C_α distance between residues i and j $\langle (R_{\mathrm{C}_\alpha})^2_{ij}\rangle^{1/2}$ is shown as a function of sequence separation $N_{ij} = |i-j|$. The data from the simulations is shown in blue and fitted with Eq. (2.24) (orange). Fit parameters for the constant C and the scaling exponent ν are given. The effective values for the protein-dye system ($\langle R^2_{\mathrm{DA}}\rangle^{1/2}$ and N_{eff}) are indicated by gray dashed lines. On the right, CI-2 is shown in gray along with AF546 (blue) and AF647 (red) and the respective C_α-atoms (blue and red spheres).

In the unfolded state, the accessible volume approach faces the challenge of properly treating the whole unfolded ensemble, as it consists of diverse protein conformations with different possible dye distributions. In contrast, my model can directly simulate the unfolded ensemble for the whole system with dyes.

To properly compare the inter-dye distances with the C_α distances, I use the model of unfolded proteins as polymer chains (see Sec. 2.6). At high simulation temperatures in the unfolded state, where the contact potential plays a negligible role, the SBM describes an excluded volume polymer chain [75]. To show this, I consider the dependency of the mean C_α distance $\langle (R_{\mathrm{C}_\alpha})^2_{ij}\rangle^{1/2}$ between residues i and j on the sequence separation $N_{ij} = |i-j|$. This dependency is expected to behave like

$$\langle (R_{\mathrm{C}_\alpha})^2_{ij}\rangle^{1/2} = C \cdot N_{ij}^{\nu}\,, \tag{5.1}$$

where $\langle (R_{\mathrm{C}_\alpha})^2_{ij}\rangle^{1/2}$ is averaged over the simulation (see also Eq. (2.24)). Fig. 5.15 shows this dependency exemplarily for CI-2 with AF546 and AF647.

The resulting scaling exponent ν from a fit corresponds well with the expected value for an excluded volume chain of $\nu = 3/5$ [75]. The resulting values of ν for the different protein-dye systems are given in Tab. 5.4. From this fit, I determine an effective segment length $N_{\text{eff}} = N_{ij} + L$, accounting for the dye pair

Table 5.4: Sequence separation N_{ij} and fitted scaling exponent ν for different protein-dye systems [113]. The correction factors m for the three systems shown in Fig. 5.14 are given.

Protein	Dye pair	N_{ij}	ν	m
CspTm C2/C68	AF488/AF594	66	0.613	1.10
CspTm C68/C2	AF488/AF594	66	0.602	
CspTm C11/C68	AF488/AF594	57	0.580	
CspTm C68/C11	AF488/AF594	57	0.597	
CspTm C23/C68	AF488/AF594	45	0.597	
^{10}FNIII	AF546/AF647	75	0.598	1.11
CI-2	AF488/AF594	58	0.595	
CI-2	AF546/AF647	58	0.611	1.14

with an additional length L. From the dyes' mean squared separation $\langle R_{\mathrm{DA}}^2\rangle^{1/2}$, I assign a length N_{eff} from the respective fit. It turns out that the length for the AF488/AF594 dye pair is about $L = 11.3 \pm 1.3$ residues, which is in the same range as found experimentally in [78]. For the AF546/AF647 dye pair, L is 14.15 ± 0.65 residues.

The effective length can be used to calculate a correction factor m for the relation between $\langle R_{\mathrm{DA}}^2\rangle^{1/2}$ and the distance of the respective C$_\alpha$-atoms $\langle R_{\mathrm{C}_\alpha}^2\rangle^{1/2}$ (see Eq. (2.25)). The resulting correction factors for the systems shown in Fig. 5.14 are given in Tab. 5.4.

The histograms relating the inter-dye distance to the corresponding C$_\alpha$ distance for the unfolded states are shown in Figs. 5.14b,d,f. Clearly, the values are not identical (equality is indicated by the dotted lines). Taking into consideration the relation between R_{DA} and the C$_\alpha$ distance for the corrected chain length, I expect a different relation as indicated by the white crosses. It provides a good approximation of the dependency observed in the simulations.

To summarize, I achieve to capture both folded and unfolded states within the same simulation method. This enables to simulate e. g. complex large-scale conformational transitions between multiple states in the future.

5.7 Combination of FRET and SAXS

Currently, there is an ongoing discussion on how to interpret SAXS measurements in comparison to FRET measurements, especially regarding intrinsically disordered and unfolded proteins. Questions are whether the FRET dyes influence the SAXS profile and how to assess derived values for the radius of gyration obtained by the two different methods.

Several studies report that FRET measures compaction of intrinsically disordered proteins in water in comparison to high denaturant concentrations, whereas SAXS experiments have not observed this compaction [149, 150]. These diverging results have been explained by decoupling of size and shape fluctuations, making FRET and SAXS complementary methods [151]. Fuertes et al. assume that proteins undergo a sequence-specific decoupling of end-to-end distance R_e measured by FRET and radius of gyration R_g measured by SAXS. As proteins are not homopolymers, they may have different R_g-R_e relationships [152]. Other studies claim that the main reason for the found discrepancies lies in the analysis methods [153]. Borgia et al. have performed combined FRET and SAXS measurements of unfolded and intrinsically disordered proteins. They state that while SAXS measurements are principally model free, interpretation of FRET measurements always depends on the underlying model relating R_g and R_e [154]. Possible models are e. g. a Gaussian chain or an excluded volume chain. As one possible approach to overcome the differences, they propose reweighting of structural ensembles. Furthermore, regular MD simulations with explicit solvent of one intrinsically disordered protein comprising 79 residues conducted by Zheng et al. [155] have resolved possible discrepancies between measurement methods for this specific protein.

Still, the questions reside, if FRET dyes have an effect on SAXS measurements and how the different calculation methods for R_g in general affect the results. With the framework introduce here, I can tackle these questions. Simulations can be used to analyze the influence of FRET dyes on SAXS profiles in the shown examples and also provide values for the radius of gyration calculated from SAXS data and other methods.

5.7.1 Influence of FRET Dyes on SAXS Measurements

First, I consider the direct influence of FRET dyes on SAXS measurements and derived values. As SAXS experiments measure and average over structural ensembles, I calculate the mean intensity curves of the folded and unfolded ensembles by taking structures distributed over the entire simulation in time intervals of $\delta t_\mathrm{SBM} = 100\,\mathrm{ps}$. For each structure, I determine a SAXS intensity profile with `CRYSOL` [74]. Subsequently, I calculate average and standard deviation of the resulting intensities to get an impression of how broad the distributions resulting from the different conformations are. This is done for the simulations of the systems both with and without dyes.

The resulting SAXS intensity curves for ^{10}FNIII in Fig. 5.16 show that the curves with and without dyes differ considerably for the folded conformations. This is to be expected as the dyes change the size and shape of the system. In the unfolded conformations, there still is a visible difference, as the size of the chain

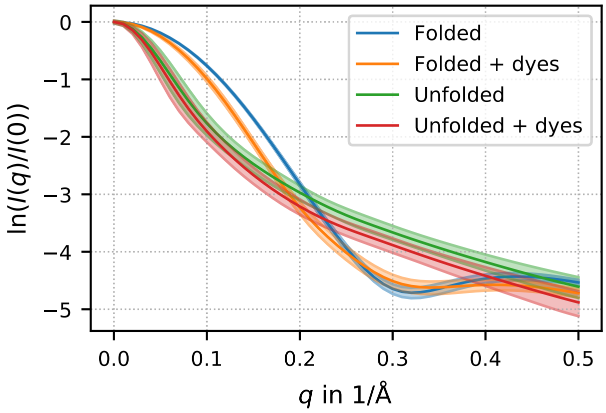


Figure 5.16: SAXS intensities without and with dyes for ^{10}FNIII. The mean SAXS intensities (solid lines) are shown along with the distributions' standard deviations (shaded area) as functions of the scattering vector q. Curves are depicted for ^{10}FNIII in the folded states without (blue) and with dyes (orange) and ^{10}FNIII in the unfolded states without (green) and with dyes (red).

is different with and without dyes. However, it is rather small and the shapes of the curves are alike. Consequently, the influence of the dyes in the unfolded states is small, but present.

In addition to this rather small system consisting of 94 residues, Fig. 5.17 shows the respective curves for the larger system of ClyA monomer consisting of 303 residues. There, the visible differences are considerably smaller, according to my expectations.

The derived plots for ^{10}FNIII are depicted in Fig. 5.18. The Guinier plot (see also Sec. 2.5) shows that the Guinier approximation (Eq. (2.20)) holds in a certain region and $\ln(I(q)/I(0))$ as a function of q^2 can be fitted linearly. The slope of the curve, representing the radius of gyration, certainly is different for folded and unfolded states and there is also a visible difference between the calculation with and without dyes, respectively. The Kratky plot in Fig. 5.18 (see also Sec. 2.5), illustrates the specific features of a distinct peak for the folded states and a plateau for the unfolded states. The broadening of the peak width for the folded states including dyes points to a less compact structure.

In Fig. 5.19, the Porod plot of the SAXS intensity (see also Sec. 2.5) is shown for ^{10}FNIII. The so-called "power-law regime" refers to the region where the intensity can be described with an exponential (see Eq. (2.21)). The derived scaling exponent ν is similar for both simulations with and without dyes and agrees well with the parameter determined in Sec. 5.6.

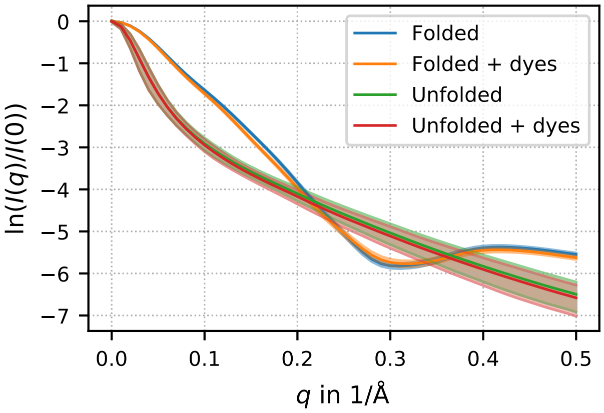


Figure 5.17: SAXS intensities without and with dyes at positions 56 and 252 for the ClyA monomer. The mean SAXS intensities (solid lines) with the distributions' standard deviations (shaded area) as functions of the scattering vector q are depicted. Curves are shown for the folded ClyA monomer without (blue) and with dyes (orange) and the unfolded ClyA monomer without (green) and with dyes (red).

5.7.2 Determination of the Radius of Gyration

To investigate the different methods to determine the radius of gyration R_g, I first calculate its value from the molecular model of the protein without dyes using GROMACS, here denoted as $R_\mathrm{g,\ gmx}$ (see Eq. (B.3)) as a "true" reference value. The second value considered is the value from the molecular model with dyes, $R^{+\mathrm{dyes}}_\mathrm{g,\ gmx}$.

Analysis of the Guinier region in the SAXS measurements yields two additional values $R_\mathrm{g,\ saxs}$ and $R^{+\mathrm{dyes}}_\mathrm{g,\ saxs}$ (as described in Sec. 2.5). These can contain errors due to the narrow Guinier region, where only few data points remain for the fit, especially for large systems as the unfolded states of ClyA monomer and protomer.

Furthermore, I calculate R_g for the unfolded proteins as done in experimental work [154]. I assume the proteins to behave like excluded volume chains (for validation, see Sec. 5.6), but only consider atoms from residues between the dye positions, neglecting the remaining residues in all calculations to mimic dyes attached to the termini*. From this data, I extract the end-to-end C_α distances R_e and calculate the apparent R_g, which is given by:

$$R^\mathrm{app}_{\mathrm{g,\ C}_\alpha} = \sqrt{\langle R_\mathrm{e}^2 \rangle} / \sqrt{6.26} \tag{5.2}$$

*The influence of the remaining chain has to be neglected here, but could be tested with additional simulations.

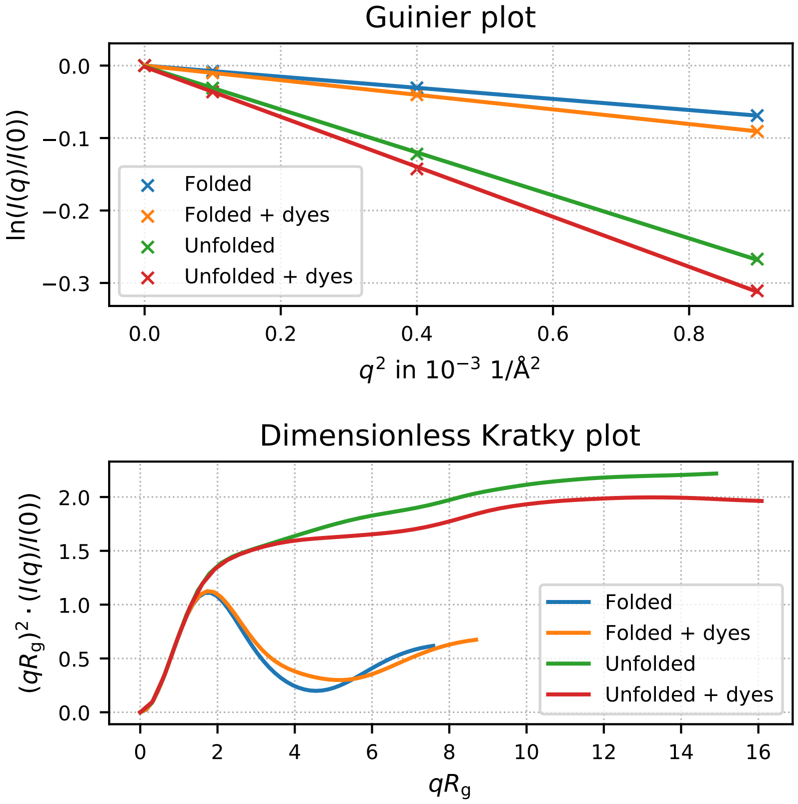


Figure 5.18: Guinier plot and dimensionless Kratky plot for ^{10}FNIII. The Guinier plot (top) shows the behavior of the SAXS profile in the Guinier region (crosses) along with linear fits (solid lines). The dimensionless Kratky plot (bottom) gives information about the protein conformations. Both plots are shown for ^{10}FNIII in folded states (blue, orange) and unfolded states (green, red) without and with dyes, respectively.

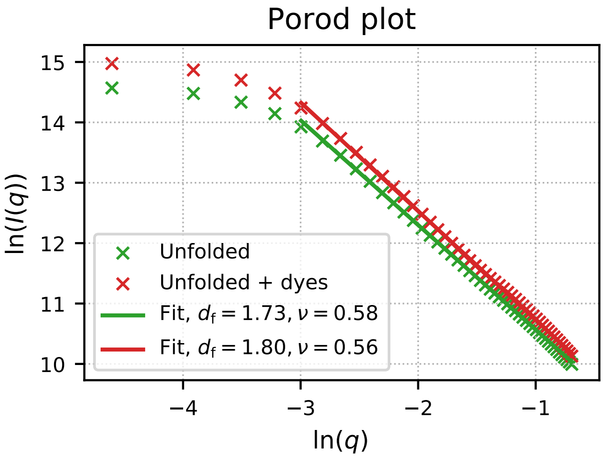


Figure 5.19: Porod plot of SAXS intensity curve for ^{10}FNIII. The “power-law regime” is the region of the linear fit (solid lines) shown here along with the data from simulation (crosses) for ^{10}FNIII without (green) and with dyes (red) for the unfolded conformations. From the fits, the fractal dimensions d_f and the corresponding scaling exponents ν are calculated.

for excluded volume chains [154]. As a second approach [154], I use the inter-dye distance R_DA and rescale it to an end-to-end distance via the factor

$$f = \left(\frac{N_\text{end-to-end}}{N_\text{end-to-end} + L}\right)^\nu \tag{5.3}$$

with scaling exponent ν and additional sequence length L for the dye pairs from Sec. 5.6. With this factor, I obtain

$$R^\mathrm{app}_{\mathrm{g},\, R_\mathrm{DA}} = \sqrt{\langle R^2_\mathrm{DA}\rangle} \cdot f/\sqrt{6.26}\,. \tag{5.4}$$

All calculated values reflect an average R_g over the simulated ensemble. I compare all of these values of R_g for the different simulated systems in Figs. 5.20, 5.21, and 5.22.

Fig. 5.20 shows the results for the folded proteins in relation to $R_\mathrm{g,\,gmx}$, which is given in the bottom. As expected, $R_\mathrm{g,\,gmx}$ is only dependent on the protein. Including the dyes, $R^\mathrm{+dyes}_\mathrm{g,\,gmx}$ is apparently higher, and the effect is more pronounced in the smaller systems. The only case where $R^\mathrm{+dyes}_\mathrm{g,\,gmx}$ is actually lower than $R_\mathrm{g,\,gmx}$ is the ClyA protomer with dyes at positions 56/252, as the dyes probably shift the center of mass in favor of a smaller R_g (see also Fig. 5.8a). The dye position seems to have a small effect (as can be seen for CspTm, ClyA monomer and protomer), whereas the choice of dyes has a larger effect (see CI-2).

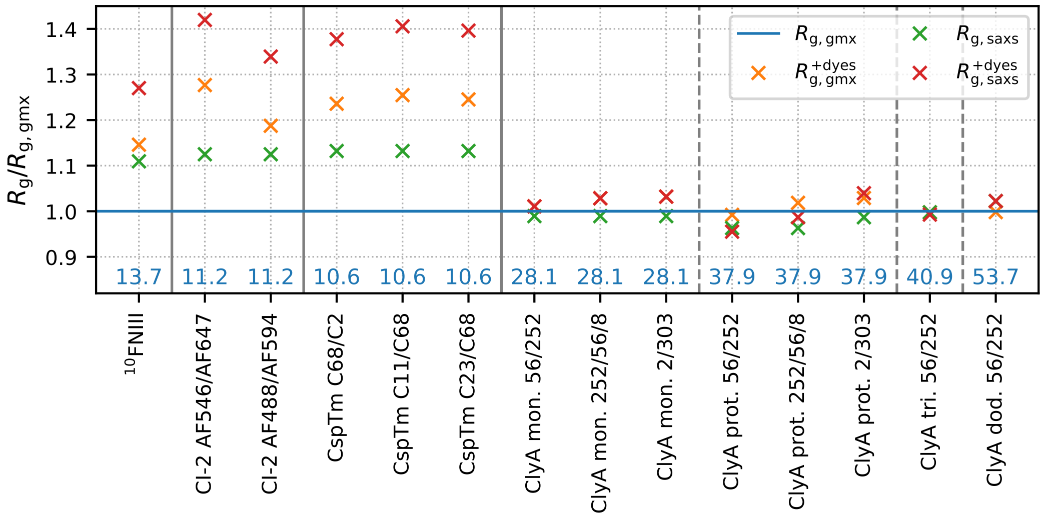


Figure 5.20: R_g values for different folded systems with respect to $R_{g,\text{ gmx}}$ (blue line), which is given at the bottom in Å. The systems studied are ^{10}FNIII, CI-2 with two different dye pairs, CspTm with AF488/AF594 at three different labeling positions, and ClyA monomer, protomer, trimer, and dodecamer with AF488/AF594/(B680) at different labeling sites. Shown are the values for R_g calculated from the atomic structure with dyes ($R^{+\text{dyes}}_{\text{g, gmx}}$, orange) and the values for R_g derived from SAXS curves without ($R_{\text{g, saxs}}$, green) and with dyes ($R^{+\text{dyes}}_{\text{g, saxs}}$, red).

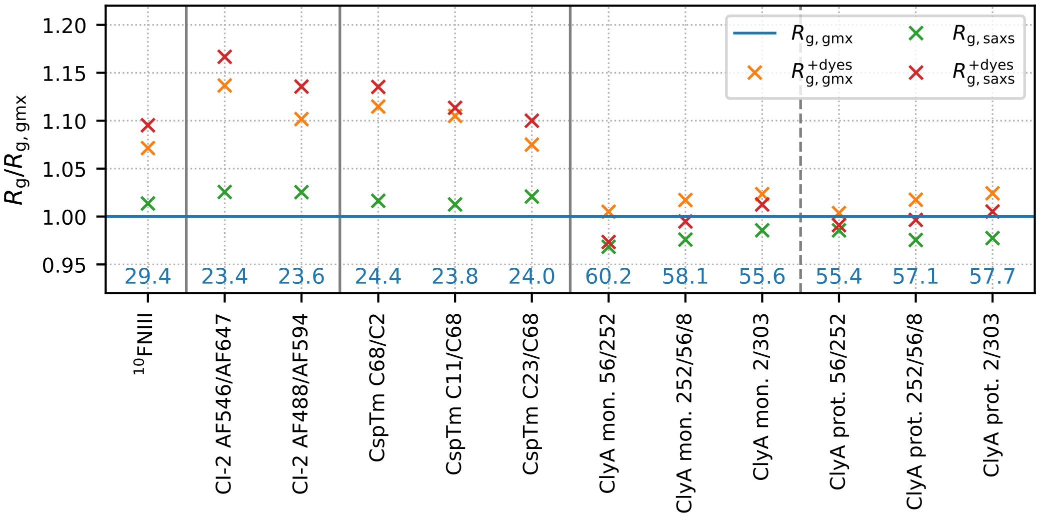


Figure 5.21: R_g values for different systems in the unfolded states with respect to $R_{g,\,gmx}$ (blue line), which is given at the bottom in Å. The systems studied are ^{10}FNIII, CI-2 with two different dye pairs, CspTm with AF488/AF594 at three different labeling positions, and ClyA monomer and protomer with AF488/AF594/(B680) at different labeling sites. Shown are the values for R_g calculated from the atomic structure with dyes ($R^{+dyes}_{g,\,gmx}$, orange) and the values for R_g derived from SAXS curves without ($R_{g,\,saxs}$, green) and with dyes ($R^{+dyes}_{g,\,saxs}$, red).

The analysis of the SAXS curves appears to overestimate R_g for small systems, while slightly underestimating R_g for larger systems. The overestimation could be caused by the calculation method of CRYSOL, which includes a hydration shell, and the neglect of hydrogen atoms in the molecular model. As the ClyA conformations are rather elongated than globular, the deviation in these systems could arise from the narrow Guinier region. $R^{+dyes}_{g,\,saxs}$ shows the same shift to higher values as for $R^{+dyes}_{g,\,gmx}$.

The results for the unfolded proteins are shown in Fig. 5.21. Here, $R_{g,\,gmx}$ is not independent of the dye pair and position (see CI-2, CspTm), indicating that there is a however small influence of the dyes on the chain dynamics. The same effect occurs for ClyA monomer and protomer, although this could also arise from insufficient sampling (see also Sec. 5.5). As for the folded proteins, $R^{+dyes}_{g,\,gmx}$ is always shifted to higher values. Here, the effect is correlated with the sequence distance of the dyes. The larger their separation, the larger the shift, as the dyes have more influence on the occupied volume at the termini than in the middle of the sequence (see CspTm, ClyA monomer and protomer). For the small systems of CI-2, CspTm, and ^{10}FNIII, $R_{g,\,saxs}$ and $R^{+dyes}_{g,\,saxs}$ show the same effect for the unfolded states as for the folded proteins. For ClyA, the values of $R^{+dyes}_{g,\,gmx}$ are

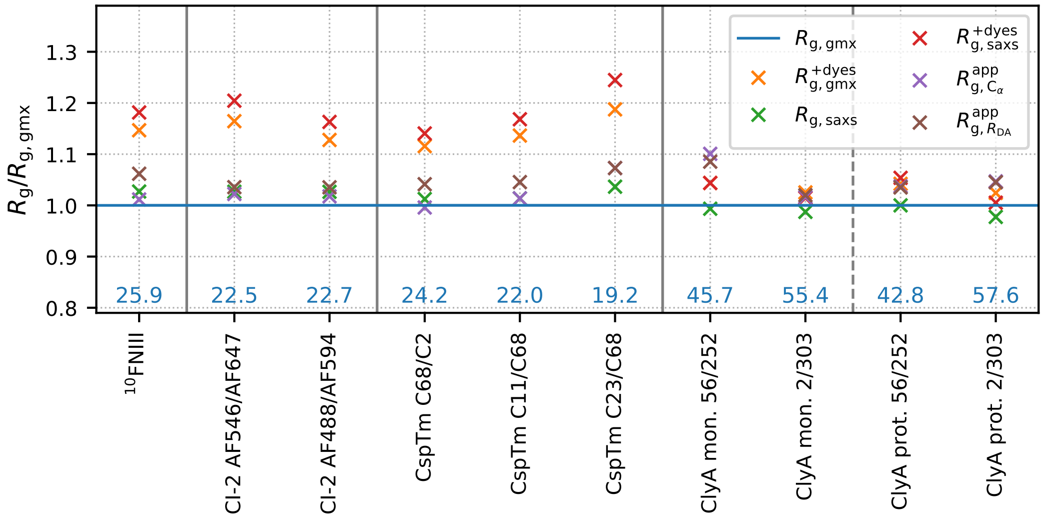


Figure 5.22: R_g values for different truncated systems in the unfolded states with respect to $R_{g,\,gmx}$ (blue line), which is given at the bottom in Å. The systems studied are ^{10}FNIII, CI-2 with two different dye pairs, CspTm with AF488/AF594 at three different labeling positions, and ClyA monomer and protomer with AF488/AF594 at different labeling sites. Shown are the values for R_g calculated from the atomic structure with dyes ($R^{+dyes}_{g,\,gmx}$, orange), the values for R_g derived from SAXS curves without ($R_{g,\,saxs}$, green) and with dyes ($R^{+dyes}_{g,\,saxs}$, red), and the apparent values of R_g calculated from C$_\alpha$ end-to-end distance ($R^{app}_{g,\,C_\alpha}$, purple) and inter-dye distance ($R^{app}_{g,\,R_{DA}}$, brown).

higher than the values of $R^{+dyes}_{g,\,saxs}$, which could originate from the smaller Guinier region, causing a larger error in the calculation. Furthermore, the simulation of unfolded ClyA might not have converged (see also Sec. 5.5), as also the scaling exponent (see Sec. 5.6) differs strongly between monomer and protomer with the different dye positions [data not shown].

Finally, I investigate the R_g values obtained from end-to-end distances shown in Fig. 5.22. Here, $R^{+dyes}_{g,\,gmx}$ and $R^{+dyes}_{g,\,saxs}$ are similar, but both shifted to higher values with respect to $R_{g,\,gmx}$ as before. $R_{g,\,saxs}$, $R^{app}_{g,\,C_\alpha}$, and $R^{app}_{g,\,R_{DA}}$ are all close to $R_{g,\,gmx}$. However, $R^{app}_{g,\,R_{DA}}$ is always higher than $R_{g,\,saxs}$, suggesting that there is a small systematic difference in the quantity that FRET and SAXS experiments measure.

A study of intrinsically disordered proteins in different denaturant concentrations as done in experimental studies [154] and regular MD simulations with explicit solvent [155] is currently not possible with the presented simulation method. So far, the model only resembles the unfolded protein in its excluded volume polymer state. Different denaturant concentrations could be included in future work by e. g. introducing a varying overall attractive potential.

To conclude, FRET dyes have a considerable effect on measured SAXS curves, especially in small systems. The measured R_g differs considering the system with and without dyes and also depends on the choice of dyes. The dyes might further have a slight influence on the chain dynamics. For the unfolded simulations with larger dye separation in the chain, the influence on the measured R_g becomes more pronounced. Finally, R_g values measured by FRET and SAXS deviate systematically, which hints to another possible mechanism leading to diverging results for the two experimental methods. However, when including the known corrections, all values are in good agreement.

5.8 Diffusion Parameters

The diffusional behavior of the dyes attached to a protein as quantified by e. g. the diffusion constant D is not known a priori. D is correlated to the rotational correlation time, but also dependent on size and shape of the occupied volume. For some calculations [156], approximated diffusion constants have been used. Here, I show how the performed simulations can be used to obtain quantitative values for the diffusion constant D and the confining volume $\langle r_c \rangle$, which can be employed in further calculations.

As described in Sec. 2.4, the diffusion constant can be calculated from velocities $\mathbf{v}_i$ via the velocity autocorrelation function (here denoted as $D_\mathbf{v}$, see Eqs. (2.15) and (2.16)) and from positions $\mathbf{x}_i$ via the the mean square displacement (here denoted as $D_\mathbf{x}$, see Eq. (2.17)). Having adjusted the time scale to the system specific rotational correlation time, I apply both methods on dye diffusion in each system considered.

For each calculation, I use a simulation of in total $t_\text{tot, SBM} = 250\,\text{ps}$ and a time step for the evaluation of $\Delta t_\text{SBM} = 0.2\,\text{fs}$ ($\Delta t_\text{SBM} = 0.5\,\text{fs}$ for ClyA/ $\Delta t_\text{SBM} = 5\,\text{fs}$ for determination of $D_\mathbf{x}$ for ClyA dodecamer). The integration for $D_\mathbf{v}$ is done in the interval $\tau_\text{SBM} \in [0, 1]\,\text{ps}$.

For the calculation of the parameter $\langle r_c \rangle$ characterizing the confining volume, I use the full simulations of $t_\text{tot, SBM} = 500\,\text{ns}$ with $\Delta t_\text{SBM} = 0.2\,\text{ps}$ ($\Delta t_\text{SBM} = 0.5\,\text{ps}$ for ClyA) and fit the trajectory to the protein to remove the overall protein motion. Then I fit with Eq. (2.18), where A_1 is set to 1.0.

As Fig. 5.23 shows, the mean square displacement can indeed be fitted with a linear function for small times $\tau_\text{SBM} \in [0, 2]\,\text{ps}$ and a diffusion constant can be calculated. For larger τ_SBM, the confinement of the movement emerges. The ambiguity in Eq. (2.18) does not allow to calculate a diffusion constant here, but $\langle r_c \rangle$ can be determined.

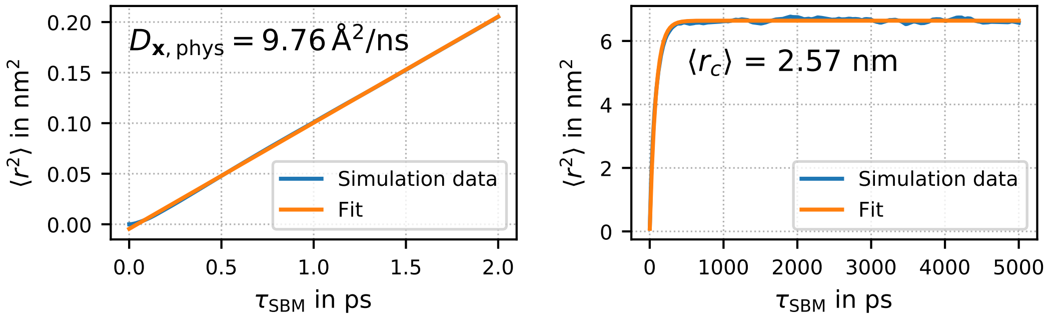


Figure 5.23: Mean square displacement as a function of time delay τ_SBM for AF546 attached to CI-2 in different regimes. For small τ_SBM (left), the mean square displacement (blue) is linear and the diffusion constant $D_{\mathbf{x},\mathrm{phys}}$ can be calculated from a fit (orange). The diffusion constant is given in converted physical units. For larger τ_SBM (right), the curve reaches a plateau due to the restricted dye motion. The value corresponding to the occupied volume of the dye $\langle r_c \rangle$ is calculated from a fit with Eq. (2.18).

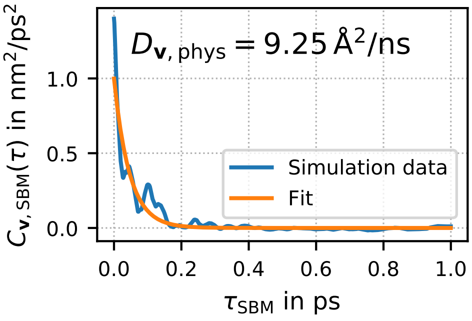


Figure 5.24: Velocity autocorrelation function $C_\mathbf{v}(\tau)$ as a function of τ_SBM for AF546 attached to CI-2. The data (blue) can be fitted with an exponential function (orange). The integral over $C_\mathbf{v}(\tau)$ (see Eq. (2.15)) yields a value for the diffusion constant via the Green-Kubo relation (see Eq. (2.16)). The resulting diffusion constant is given in converted physical units.

Table 5.5: Diffusion parameters calculated from simulations. The diffusion constants are given for the different dyes, attached to the different proteins, calculated from the velocity autocorrelation function ($D_{\mathbf{v}}$) and from the mean square displacement ($D_{\mathbf{x}}$), respectively. With $\langle r_c \rangle$, a measure for the occupied volume of the dyes is listed, comparable to the extend of the dye l_{dye}, the distance between the assumed center of the dye and the carbon atom of the maleimide which is attached to the protein (see Secs. 3.3 and E). All values are given in physical units.

Dye	Protein	$D_{\mathbf{v}}$ in Å^2/ns	$D_{\mathbf{x}}$ in Å^2/ns	$\langle r_c \rangle$ in nm	l_{dye} in nm
AF488	CI-2	8.20	8.23	2.05	1.69
	CspTm	7.25	8.04	1.80	
	ClyA prot.	2.20	2.68	2.05	
	ClyA dod.	3.17	2.87	1.50	
AF546	CI-2	9.25	9.76	2.57	2.69
	^{10}FNIII	11.15	11.51	2.26	
AF594	CI-2	7.98	8.38	1.64	1.66
	CspTm	7.96	8.28	2.04	
	ClyA prot.	3.51	3.38	1.77	
	ClyA dod.	1.90	3.19	1.43	
AF647	CI-2	7.91	8.40	1.47	1.70
	^{10}FNIII	9.51	9.76	1.63	

For calculation of $D_{\mathbf{v}}$, the main contributions to the integral come from small times as evident in Fig. 5.24. However, it has to be noted that the time range used for the calculation affects the results and is not unambiguous.

The values determined for the different dyes in several systems are given in Tab. 5.5. The two methods of determining D yield results in good agreement. Deviations likely arise from the time step ($D_{\mathbf{v}}$ might yield better estimations with an even smaller time step), the choice of time frame used in the integration for $D_{\mathbf{v}}$, and the fact that, due to the confinement, the mean square displacement can not be calculated up to infinity as required. On these accounts, the range used for the linear fit is an additional ambiguous parameter. Still, the values for AF488 and AF594 agree well with the assumed diffusion constants of previous work [156].

As the rotational correlation times already differ between CspTm and ClyA, it is to be expected that the diffusion parameters differ as well. Despite having similar rotational correlation times in different systems, the values for AF546 or AF647 differ with respect to the system. This result shows the influence of the surrounding protein on behavior and occupied volume of the dye.

For comparison, also values of l_{dye}, the distance between the assumed dye center and the carbon atom of the maleimide attached to the protein (see also Secs. 3.3 and E), are given. These values serve as a reference for the dyes' linker length. $\langle r_c \rangle$ clearly correlates with l_{dye}, but diverges for the different systems.

Comparing values for ClyA protomer and ClyA dodecamer, it is not surprising that the volume occupied by the dyes is larger for the protomer (see also Sec. 5.5). Due to the smaller volume and the repulsion by adjacent protein chains, also the diffusion of AF488 is faster in the dodecamer.

This section showed the ability of the presented simulation method to determine approximations for experimentally inaccessible parameters, which can be used for further calculations and analyses.

5.9 Modeling of a Glucose Sensor

The different variants of a the glucose sensor introduced in Sec. 4.6 have shown different behavior in FRET experiments [132]. They yield different changes in FRET ratio (see Eq. (2.6)) upon glucose binding and have therefore considerably different sensitivities. With the simulation-based approach presented in this work I want to identify the effects contributing to this distinct behavior, which facilitates better understanding and improving the function of this sensor in the future.

The glucose sensor variants I study here (see Sec. 4.6) have been produced by J. Otten of the group of M. Pohl and investigated by several experimental groups. H. Höfig from the group of J. Fitter has performed FRET measurements, leading to the result that sensor 2 yields a significantly larger FRET ratio than sensor 1 and sensor 4 [132]. The underlying mechanism is not fully understood, as the only difference between sensor 2 and sensor 4 is the position of the flexible linker. To get a better insight into the systems, M. Sarter from the group of A. Stadler has conducted SAXS measurements of the different variants [unpublished data]. In the following, I use the experimental results measured by both groups.

In the next subsections, I start with an overview of the simulations, followed by analysis of the resulting distance, orientation, and efficiency distributions. Furthermore, I consider the convergence of the simulations and the flexibility of the fluorescent proteins. Finally, I present the how including dimerized conformations affects the results.

5.9.1 Starting Structures and Simulations

I perform SBM simulations of the glucose sensor variants with several starting structures (see Sec. 4.6.6). The different sensor models are denoted with sensor 2-1, sensor 2-2, etc. according to sensor variant (1, 2 or 4) and rank of their

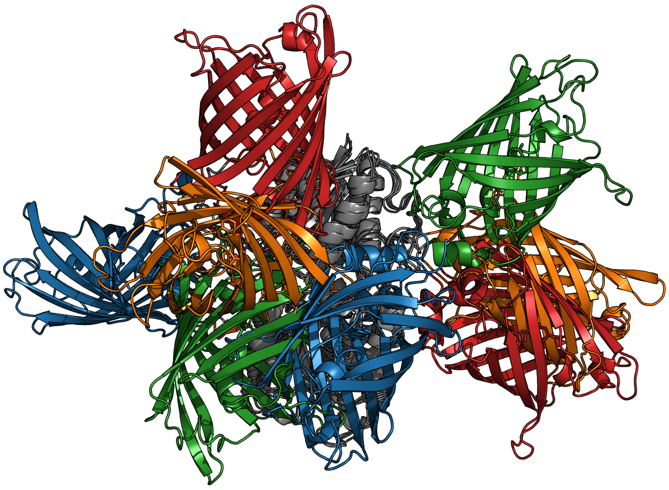


Figure 5.25: Starting structures for the simulation of sensor 2 with Glc-BP in the glucose bound state. All structures are aligned to Glc-BP (gray). The different pairs of fluorescent proteins are depicted in blue (sensor 2-1), orange (sensor 2-2), green (sensor 2-3), and red (sensor 2-4). The fluorescent protein on the left is CFP, the fluorescent protein on the right YFP, respectively.

starting structure in the fitting to the experimental SAXS data (see Sec. 4.6.6). The best fitting sensor model is denoted sensor X-1, the second best sensor X-2, etc. Using the example of sensor 2, the four starting structures with the lowest χ^2 values with respect to the experimental SAXS data [M. Sarter, private communication] are depicted in Fig. 5.25. The span of conformations in agreement with experimental data is large. Starting structures for simulation of sensor 1 and sensor 4 are given in Sec. J.

Still, considering only these single structures and assuming they are stiff, the inter-fluorophore distance and orientation of each structure would yield small FRET efficiencies in contrast to the experimental measurements [132]. In regard of this aspect and the diversity of starting structures, it is suggested that a dynamic structural ensemble is required instead of a single structure to explain the behavior of the glucose sensor.

To obtain a larger structural ensemble and investigate the dynamics of the sensor variants, I perform simulations with different starting structures for each sensor with Glc-BP in the glucose bound state and in the glucose free state, respectively. The simulations cover $t_{\text{tot, SBM}} = 1000\,\text{ns}$ (see Sec. 4.6.7), which represents approximately 100 µs on the physical time scale. Comparison of the respective RMSD values with respect to Glc-BP^{+G} and Glc-BP$^{\text{-G}}$ shows that all structures stay stable in their respective Glc-BP state [data not shown].

For glucose bound sensor 2-1 I exemplarily present a detailed picture of the results. I generate SAXS intensity profiles of structures distributed evenly over the entire simulation and average the intensities (see also Sec. 5.7.1). Fig. 5.26

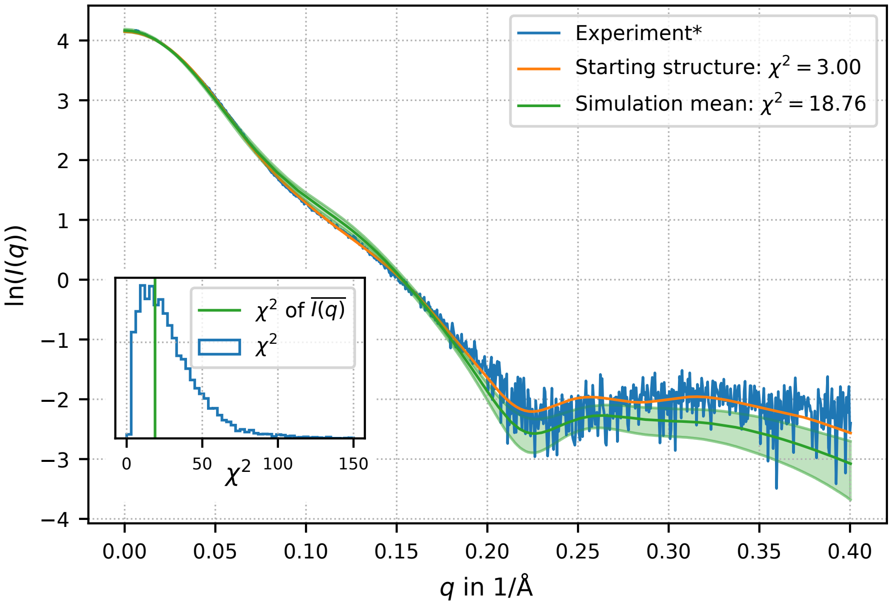


Figure 5.26: SAXS intensity curves for sensor 2-1 with Glc-BP in the glucose bound state. The experimental SAXS data (*[M. Sarter, private communication]) is shown in blue, the fit of the starting structure in orange, and the mean intensity in green, respectively. The width of the intensity distribution in the simulation is depicted as shaded area. Furthermore, the respective χ^2 values for starting structure and mean intensity are given. The inset shows the χ^2 distribution of individual structures over the simulation.

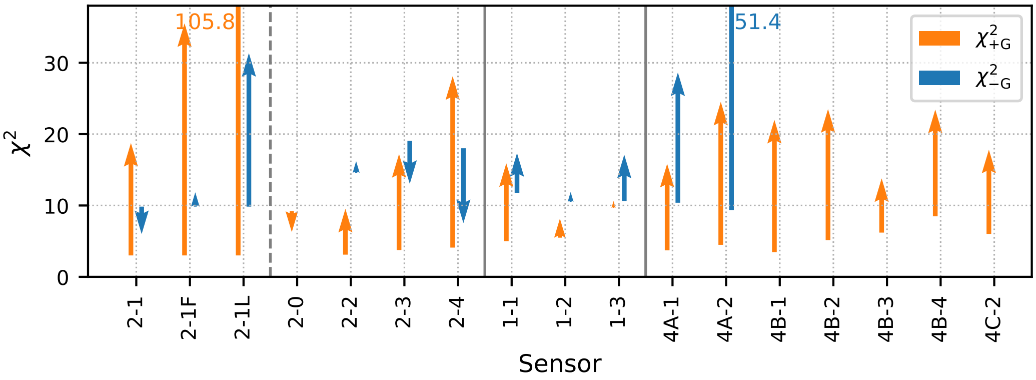


Figure 5.27: Change in χ^2 from starting structure to mean intensity of the simulation with respect to the experimental data, respectively. The arrows start at the χ^2 value of the starting structure and end at the χ^2 value of the mean intensity. For the different sensors, the respective values are depicted for Glc-BP^{+G} (orange) and Glc-BP^{-G} (blue) states. The endpoints of the two arrows exceeding the figure are denoted alongside. For sensors 2-0, 4B-1, 4B-2, 4B-3, 4B-4, and 4C-2, no simulations of the glucose free state are performed.

depicts the SAXS intensity curves for the starting structure of sensor 2-1 along with the mean intensity curve over the simulation. The inset in Fig. 5.26 displays large χ^2 fluctuations over the simulation, which average out when calculating the mean intensity. This shows the suitability of a structural ensemble instead of a single structure to explain the experimental data. The χ^2 value of the mean intensity curve in comparison to that of the starting structure, however, shows a divergence of the simulation from the SAXS data.

Fig. 5.27 depicts the change in χ^2 from starting structure to mean intensity curve of the simulation for all sensor models. For almost all sensors the simulations diverge from the experimental data, regardless of the Glc-BP state.

To test different hypotheses in addition to the sensor variants described in Sec. 4.6, I simulate three variants of sensor 2. They are denoted as sensor 2-1F, sensor 2-1L, and sensor 2-0. The two former sensors both use the same starting structure as sensor 2-1, but differ in contacts and flexibility. In sensor 2-1F, beside the flexible (GGS)$_4$-linker, all restriction sites are kept flexible. As the restriction sites are added artificially to the structure and their structures are not known, this could be a valid model. However, the high χ^2 values in the simulation of sensor 2-1F indicate that its agreement with experimental data is worse than for the other models of sensor 2. The original parametrization chosen for the restriction sites seems to be preferable.

To test if Glc-BP refolds completely including residues 1 to 11 with CFP inserted (see also Sec. 4.6.6), I simulate sensor 2-1L with no contacts between residues 1 to

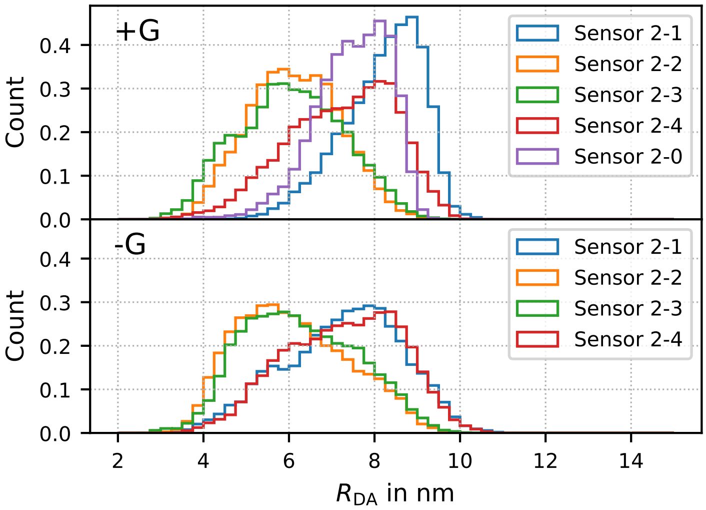


Figure 5.28: Inter-fluorophore distance distributions from simulations of different starting structures for sensor 2 with glucose (+G, top) and without glucose (-G, bottom).

11 and the remaining part of Glc-BP. In the simulation, the N-terminal residues can loosen itself from the rest of the structure accordingly. Sensor 2-1L clearly shows larger divergence from the SAXS data than sensor 2-1, making this scenario unlikely.

Sensor 2-0 referred to in Fig. 5.27 and the following figures denotes a structure constructed by taking the first possible rotations in the merging process of sensor 2 (see Sec. 4.6.6), regardless of the agreement with the SAXS data. The parameters are taken as for the other sensor 2 variants. It also yields low χ^2 values which even decrease for the simulation (see Fig. 5.27).

5.9.2 Distances, Orientations, and FRET Efficiencies

The simulations provide direct access to the dynamics and structural ensembles of the different sensor models. To determine their differences, I compare distances and orientation factors between the distinct models with and without glucose. Fig. 5.28 exemplarily shows the distance distributions for sensor 2. The distributions of the different starting structures vary considerably. Although expected to be larger in the Glc-BP$^{\text{-G}}$ state (see Sec. 4.6.2), the mean distances decrease and the distributions become broader in comparison to the Glc-BP$^{+\text{G}}$ state. A possible reason for the discrepancy of the different models is that the conformational space is not sufficiently sampled (see also Sec. 5.9.3). It should further be noted that

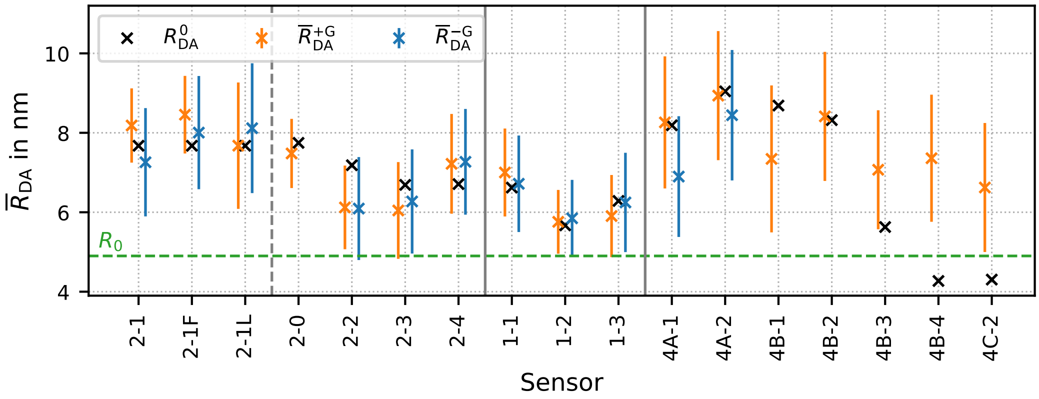


Figure 5.29: Distance distributions for the different sensors. The mean distances $\overline{R}_{\mathrm{DA}}$ (crosses) and the standard deviations σ_{DA} of the distribution (error bars) are depicted for simulations with glucose (+G, orange) and without glucose (-G, blue). The values of the respective starting structures R^0_{DA} are shown as black crosses. The Förster radius of the fluorophore pair is illustrated by a green dashed line. For sensors 2-0, 4B-1, 4B-2, 4B-3, 4B-4, and 4C-2, no simulations without glucose are performed.

as it represents the energetic minimum the influence of the starting structure in SBMs is not negligible.

In Figs. 5.29, 5.30, and 5.31, distance distributions, mean $\overline{\kappa^2}$ values, and resulting peak FRET efficiencies are depicted, respectively.

As evident from Fig. 5.29, the distance distributions of the simulations without glucose are always broader than that of the simulations with glucose. The fluorescent proteins are more flexible in this glucose free state (see also Sec. 5.9.4). For all sensors, the distributions vary noticeably. Even in the models with small inter-fluorophore distance in the starting structure, the distances tend to increase during the simulation so that most distances are well above the Förster radius of the fluorophore pair $R_0 = 4.9\,\mathrm{nm}$. This includes most of the starting structures, which are in good agreement with the SAXS data, and also sensor 2-0 with a low χ^2 in the simulation. This is surprising, considering the experimentally measured FRET efficiencies ranging from about 0.1 to 0.8. The distance distributions alone seem to contradict the measured FRET data. Furthermore, the distance distributions show large differences for the different sensors, depending on the starting structures. This may hint to a not fully converged ensemble and could be approached in future work by either extending the simulations or by e. g. averaging over the different simulations for each sensor variant.

The mean $\overline{\kappa^2}$ values in Fig. 5.30 are diverse and clearly deviate from the isotropic value of $\overline{\kappa^2} = 2/3$. This deviation could be caused by the steric restrictions of

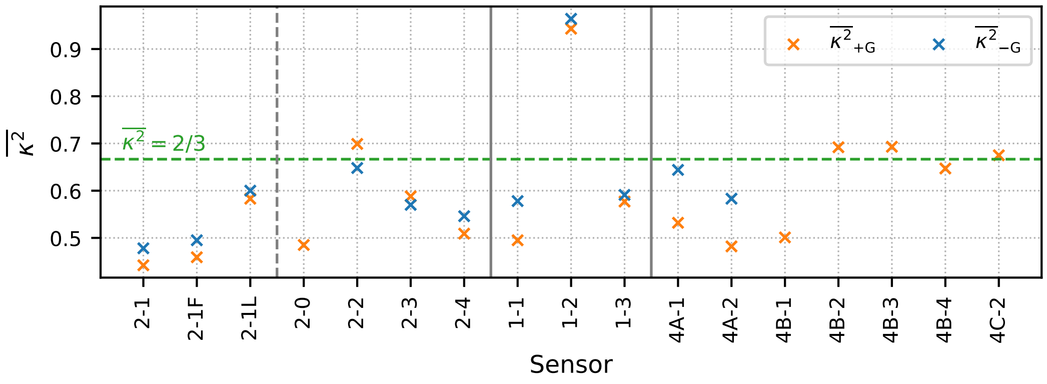


Figure 5.30: Mean $\overline{\kappa^2}$ values for the different sensors. The mean values are given for simulations with glucose (+G, orange) and without glucose (-G, blue). As a reference, the green dashed line illustrates the value $\overline{\kappa^2} = 2/3$ valid in the isotropic averaging regime. For sensors 2-0, 4B-1, 4B-2, 4B-3, 4B-4, and 4C-2, no simulations without glucose are performed.

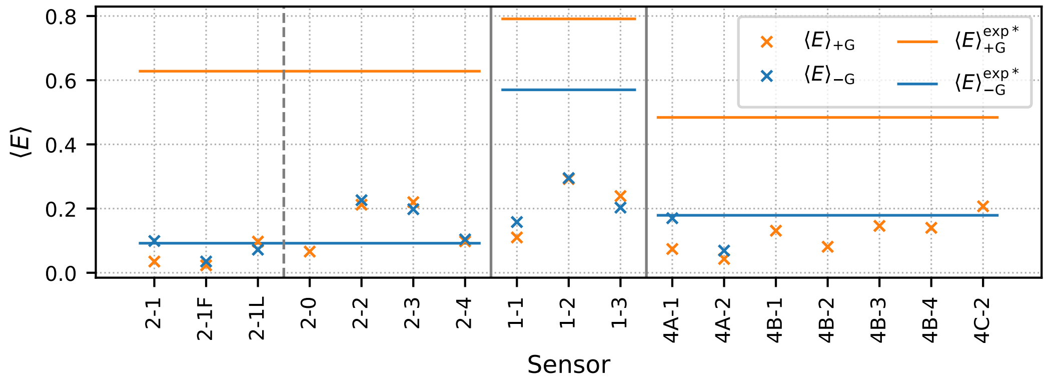


Figure 5.31: FRET efficiencies for the different sensors. The peak efficiencies $\langle E \rangle$ are given for simulations with glucose (+G, orange crosses) and without glucose (-G, blue crosses). The experimentally measured peak efficiencies are given as lines for sensor 1, sensor 2, and sensor 4, respectively (*[132]). For sensors 2-0, 4B-1, 4B-2, 4B-3, 4B-4, and 4C-2, no simulations without glucose are performed.

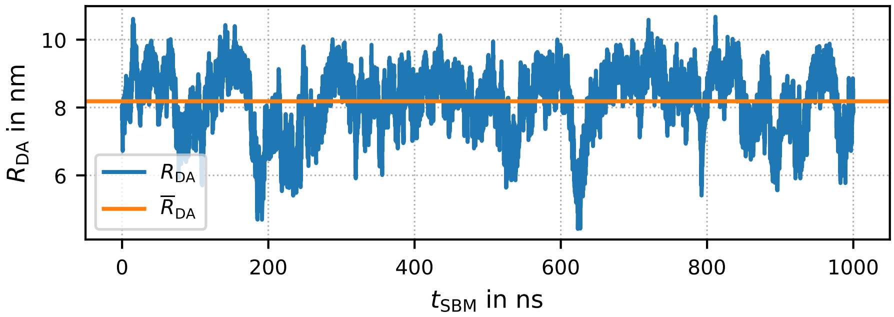


Figure 5.32: Inter-fluorophore distances as a function of time for the entire simulation of sensor 2-1 with glucose (blue). The mean value is depicted in orange. The fluctuations are large and occur on a long time scale compared to the simulation time.

the fluorescent proteins imposed by the linkers, their slow motions or insufficient sampling (see also Sec. 5.9.3). There is however no apparent systematic difference between the states with and without glucose. Still, the different orientations have an impact on the results and should not be neglected.

Finally, the peak efficiencies resulting from Monte Carlo photon simulations of the systems are overall too low compared to the experimental measurements (see Fig. 5.31) [132]. All results indicate additional effects that are not accounted for in the simulations so far, such as additional attractive interactions between the fluorescent proteins or between the fluorescent proteins and Glc-BP, respectively.

5.9.3 Convergence of Simulations

I investigate the extend of interconversion between distinct starting structures in the simulations and test if the systems converge within the simulated time.

Comparing the structures against each other, they seem to be able to mutually reach each other. Sensor 1 is rather stiff, but all three simulations are adopting structures close (with an RMSD below 1 nm) to the sensor 1-2 starting structure. The structures of sensors 2-2, 2-3, and 2-4 are mutually reachable, however, sensor 2-1 is further apart. This deviation is reflected in the distributions in Fig. 5.28. When looking at the rather similar starting structures of sensors 2-2, 2-3, and 2-4 in contrast to sensor 2-1, this is to be expected. All structures of sensor 4 are mutually linked but rather far apart and get only rarely close to one another. This probably arises from the long and flexible linker, which yields a large space of possible conformations for YFP.

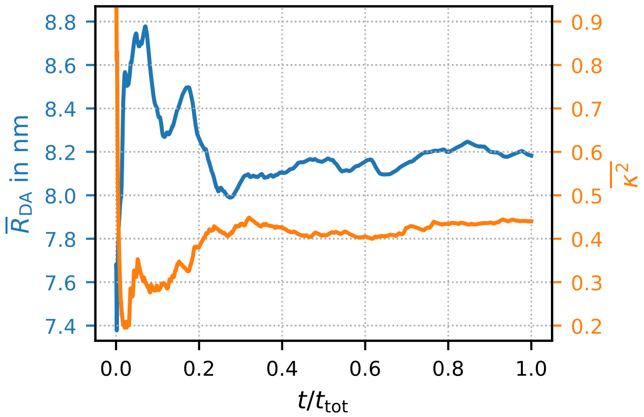


Figure 5.33: Mean inter-fluorophore distances $\overline{R}_{\mathrm{DA}}$ (blue) and mean orientation factors $\overline{\kappa^2}$ (orange) for sensor 2-1. The values are given over the parts of the simulation used for the calculation. The time t is given with respect to the total simulated time t_{tot}.

Considering a single simulation, Fig. 5.32 shows the time-dependent inter-fluorophore distance for sensor 2-1 exemplarily. The distance fluctuations are large and occur on a time scale rather large in comparison to the total simulation time.

The mean inter-fluorophore distances and orientation factors over time for sensor 2-1 are shown exemplarily in Fig. 5.33. Both values still vary up to the end of the simulation and are probably not fully converged. Concerning the times needed for the systems to change distances significantly, it should be considered to extend the simulations of the systems.

5.9.4 Flexibility of Fluorescent Proteins

To compare the flexibility of both fluorophores in the different sensors, I calculate the fluorescence anisotropy and the rotational correlation times of the fluorophores bound to Glc-BP. By fitting the simulated trajectory to Glc-BP, I can observe the fluorophore motion independently of the overall rotation of the system. Two exemplary fits are shown in Fig. 5.34. Due to the restriction of the motions, the "wobbling-in-a-cone" model (see Eq. (2.14)) is more appropriate here than a simple exponential fit. Considering the parameter A representing the spatial restriction, it turns out that CR2 is rather free in every sensor with values of $A = 0.01 - 0.05$ for sensor 2, $A = 0.02 - 0.06$ for sensor 1, and $A = 0.00 - 0.03$ for sensor 4. SWG is hindered to different extents in the distinct sensors, revealing one of the most significant differences between the three sensor variants. The values for SWG are in the ranges of $A = 0.07 - 0.17$ for sensor 2, $A = 0.21 - 0.42$ for sensor 1, and $A = 0.18 - 0.56$ for sensor 4. Furthermore, SWG always gains flexibility, i.e. A decreases, in the structures without glucose, whereas the change for CR2

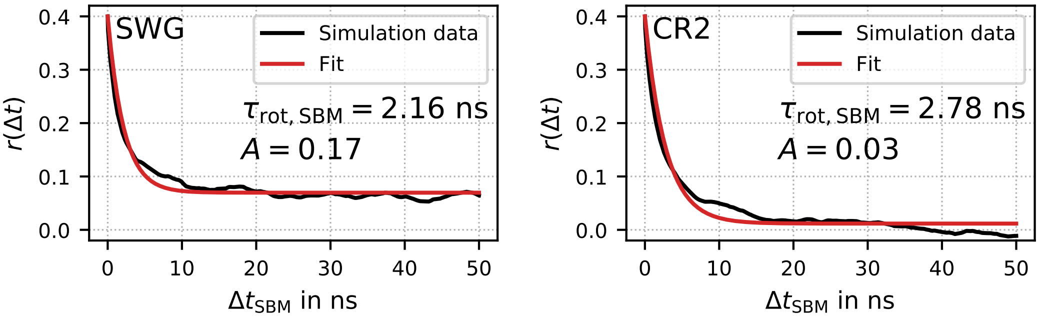


Figure 5.34: Exemplary fits of the fluorescence anisotropy $r(\Delta t)$ as a function of time Δt_{SBM} for the two fluorophores SWG and CR2 in the simulation of sensor 2-1. The calculated rotational correlation times $\tau_{\text{rot, SBM}}$ and the fit parameters A are given. The respective rotational correlation times on a physical time scale are $\tau_{\text{rot, SWG}} = 223\,\text{ns}$ and $\tau_{\text{rot, CR2}} = 287\,\text{ns}$.

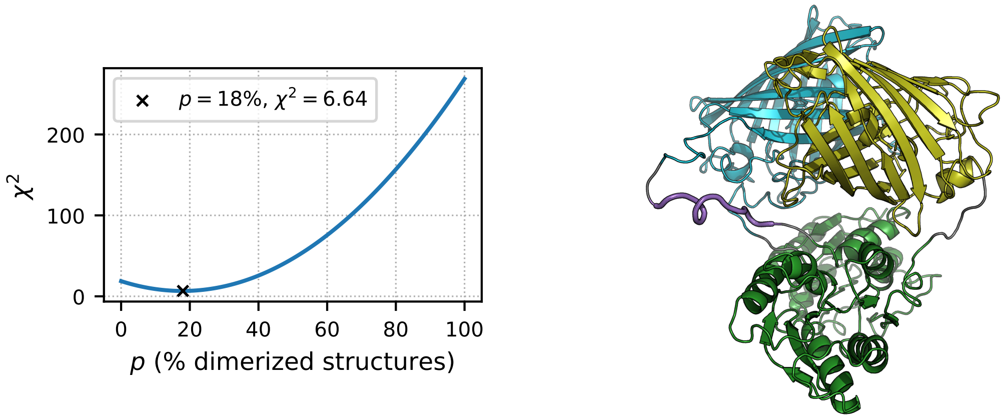
is negligible. It should be noted that SWG is mainly hindered in its rotation about the cylinder axis of the β-barrel, resulting from its two attachment points to Glc-BP.

This type of simulation is suited to determine approximations of rotational correlation times of the fluorescent proteins in the sensor, which are not directly accessible in the conducted experiments. The derived values for rotational correlation times of SWG and CR2 in sensor 2-1 are $\tau_{\text{rot, SWG}} = 223\,\text{ns}$ and $\tau_{\text{rot, CR2}} = 287\,\text{ns}$, respectively. Compared to the much shorter lifetimes of both fluorophores (see Tab. 4.8), the reorientation and dynamics of the fluorophores can not be neglected here as is done in the systems with small dyes.

5.9.5 Including Dimerized Conformations

The results so far yield good agreement with experimental SAXS data, but could not explain the experimentally observed high FRET efficiencies. However, the original fluorescent protein GFP is known to have a tendency to form dimers (see Sec. 4.6.8). I test if the missing mechanism might be due to CFP and YFP temporarily forming a heterodimer. This temporary dimerization could be simulated explicitly in future work. Here, I test how combination of dimerized structures with “free” simulations affects the results.

I simulate different sensor structures with strong dimer contacts (see Sec. 4.6.8) to obtain an ensemble of structures comprising a heterodimer between CFP and YFP. The dimer formation shows to be sterically possible for each sensor model. From these simulations, I take an ensemble of dimer structures of the respective sensor. For the dimer simulations by themselves, comparison with the experi-



(a) χ^2 values of the combination of dimer and free simulations of sensor 2-1. The χ^2 values after fitting I_{mixed} for different fractions of dimerized structures p to the experimental SAXS data, respectively, are shown. The minimum is at $p = 18\%$ (cross).

(b) Exemplary dimer structure of sensor 2-1. Glc-BP, CFP, and YFP are depicted in green, cyan, and yellow, respectively. The $(\text{GGS})_4$-linker is shown in purple.

Figure 5.35: Combination of free and dimer simulations.

mental data yields χ^2 values of around 130 to 270. However, combination of the dimerized ensemble with the free simulations in different fractions results in the behavior shown exemplarily in Fig. 5.35a for sensor 2-1. A combination of the dimer intensity curve $I_{\text{dimer}}(q)$ with the intensity curve of the free simulation $I_{\text{free}}(q)$ is calculated as:

$$I_{\text{mixed}}(q) = p \cdot I_{\text{dimer}}(q) + (1 - p)I_{\text{free}}(q) \tag{5.5}$$

with the fraction of dimerized structures p. Fig. 5.35a shows a clear minimum, meaning that combination of 18% dimer simulation and 82% free simulation improves the congruence with the experimental data significantly. An example of the dimerized structure of sensor 2-1 is shown in Fig. 5.35b.

Interestingly, Fig. 5.36 shows that dimerization improves the fitting to experimental data for all sensors. Temporary dimerization of the fluorescent proteins could compensate the large distances and thus explain the FRET data with high efficiencies. In Fig. 5.36, also differences between the sensors become visible. In contrast to sensor 1 and sensor 4, sensor 2 reveals to have a huge difference in the fraction of dimer structures that can improve the fitting to the SAXS data in the glucose bound state versus the glucose free state. In the glucose bound state, it seems to form a dimer in fractions of 11 to 21%, whereas almost no dimerization occurs in the glucose free state (0-7%). For the sensor 2 models without glucose, also the FRET efficiencies are already in good agreement with the experimental data (see Fig. 5.31).

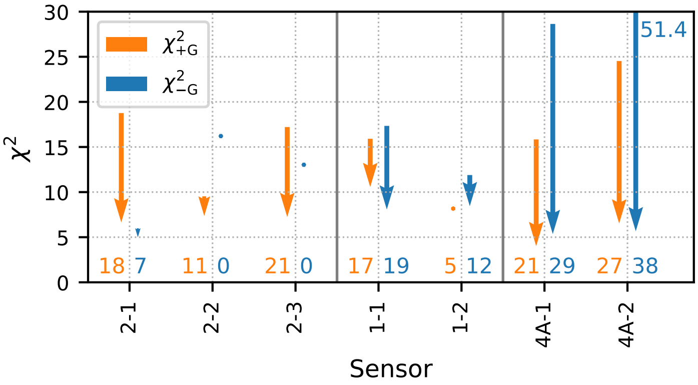


Figure 5.36: Change of χ^2 values upon adding dimerized structures to the free simulation. The arrows start at χ^2 of the intensity of the free simulation and end at χ^2 of the mixed intensity with respect to the experimental data, respectively. The numbers at the bottom denote the fraction of dimer structures p that yield the minimal χ^2 in percent. Orange refers to the simulations with glucose (+G) and blue to those without glucose (-G). Beside the arrow not completely depicted in the figure its start point is denoted. The dots indicate no or negligible change in χ^2.

The dimerization in the glucose bound state in contrast to almost no dimerization in the glucose free state can possibly explain the large FRET ratio for sensor 2 in comparison to sensor 1 and sensor 4. Furthermore, it is in accordance with the experimental observation that sensor 2 shows a compaction upon adding glucose while the other sensors show only slight differences [M. Sarter, private communication].

Finally, Fig. 5.37 shows the SAXS intensity curves and the Kratky plot (see also Sec. 2.5) for sensor 2-1 with glucose for the free simulation, the dimer simulation, and the mixed intensity curve. The latter is in high agreement with the experimental SAXS data. Also the radius of gyration determined from the mixed curve fits better with the experimental measurement than the value of the free simulation. It should be noted here that for most sensors the R_g values do not agree as well with the experimentally derived value as for sensor 2-1 [data not shown]. A systematic difference in the calculation methods for experiments and simulations can not be excluded and is subject to further investigation.

A possibly important difference between sensor 2 and the other sensors could lie in its by comparison only moderately flexible linkers. They allow for a lot of conformations, while still often residing in the distance range necessary for dimerization. Sensor 1, however, is rather stiff and thus can not explore this wide range. Sensor 4 with one very rigid and one very flexible linker (see also Sec. 5.9.4) in principle shows an even wider conformational range than sensor 2. The results suggest that it adopts the dimerized state more frequently than sensor 2. This

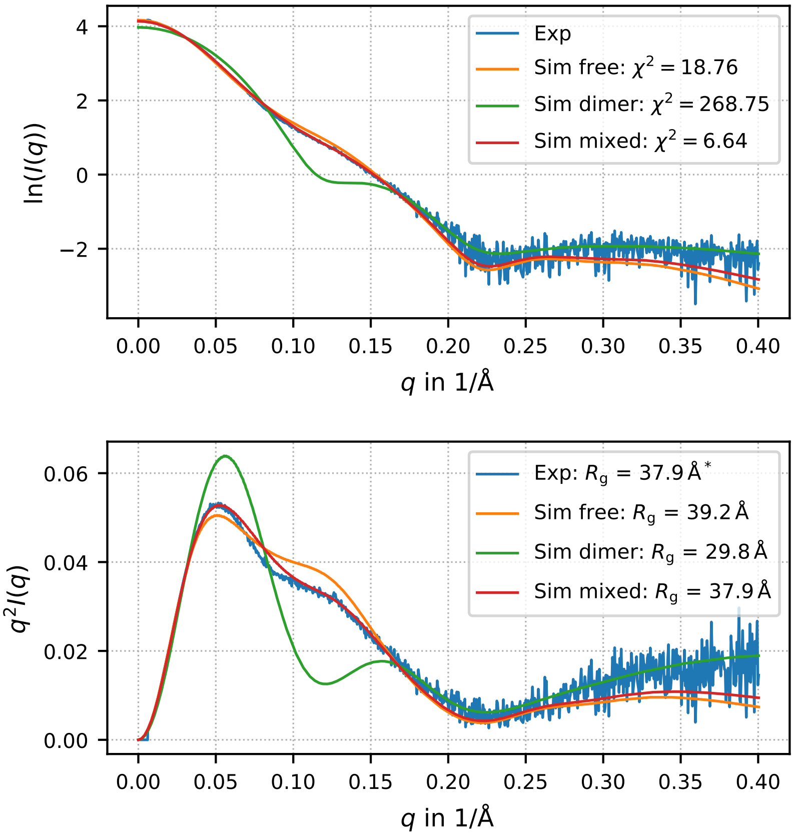


Figure 5.37: SAXS intensity curves for sensor 2-1. The curves for free (orange) and dimer simulation (green) are shown along with the mixed intensity curve (red) and the experimental data (blue). The SAXS intensities (top) and a Kratky plot (bottom) are depicted, together with the respective radii of gyration derived from the SAXS curves. The R_g value for the experimental data (*) has been calculated by M. Sarter.

behavior of sensor 4 can not directly be explained by the results so far. As only sensor 4A is simulated with and without glucose, the bias due to the few merging possibilities (see also Sec. 4.6.6) might affect the results. However, further studies are necessary to investigate this behavior.

Besides the possible formation of a heterodimer, other mechanisms can not be excluded. A different compacted structure, e. g. sticking of a fluorescent protein to Glc-BP, or weak attractive interactions between the fluorescent proteins and Glc-BP might also explain the data.

To summarize, the presented method achieves a considerable improvement in modeling the glucose sensor. It yields different structural ensembles which can already partly explain the experimental data. The overall flexibility turns out to be higher for all sensors in the glucose free state in comparison to the glucose bound state. Also, the results indicate that the flexibility of CFP marks a parameter significantly distinguishing the three sensor variants. Considering temporary dimerization of the two fluorescent proteins improves the agreement with experimental measurements. However, the different starting structures yield significantly different distance distributions. Thus, the conformational space still is considerably large and too heterogeneous to answer all questions with the simulations so far.

In future work, the simulations could be further evaluated by e. g. applying ensemble optimization methods [157] to the obtained ensemble. Another approach would be using the structures from simulations with minimal χ^2 values as new starting structures for additional simulations. The dimerization parameters can be adjusted for simulation of a system which only temporarily dimerizes and dissociates again. With simulations including both, dimerized and free structures, FRET efficiencies for the system can be calculated. It then can be tested in what respect the different sensors behave differently and how this affects the FRET efficiencies. FRET efficiency histograms can be considered further and used as additional information to choose the model which best explains the experimental results. A slightly higher flexibility in the restriction sites could overcome possible biases by the choice of their starting structures.

5.10 Summary

In this chapter, I showed results for simulations of different dye-labeled proteins and a glucose sensor. I demonstrated that the presented simulation method gives new insights into the systems' dynamics by yielding e. g. distance and orientation distributions between FRET fluorophores, which can directly be related to FRET efficiency distributions. This facilitates testing of different experimental settings, improving planning, interpretation, and validation of experimental measurements.

For example, different dye pairs, Förster radii, linker lengths, labeling sites, or the effects of three-color FRET can be tested.

The direct comparison of experimental with simulated data yielded high agreement with experiments and validated the presented approach. Investigation of different conformations of ClyA provided new insights regarding the differences observed in experiments and the underlying dynamics. The method achieves to capture both folded and unfolded states in the same simulation method, and still is in agreement with the accessible volume approach for folded proteins and the polymer model for unfolded proteins. I also elucidated the non-negligible effect of FRET dyes on SAXS measurements and determination of R_{g}, especially in small systems.

I furthermore employed the simulations to determine approximations for experimentally inaccessible quantities such as diffusion constants or rotational correlation times for fluorescent proteins, which can be used for further calculations and analyses.

Finally, the simulation method can be applied to study FRET-based biosensors comprising fluorescent proteins and obtain information about their dynamics in atomic detail. I achieved a considerable improvement in modeling the glucose sensor, yielding structural ensembles and information about the flexibility of the fluorescent proteins in these systems. By including the effects of a temporary dimerization of the two fluorescent proteins, I found possible explanations for the experimental data.

6 Conclusion

6.1 Summary and Discussion

In this work, I introduced a new method for simulating protein-fluorophore systems with a reduced set of parameters compared to regular MD force fields. A systematic way to generate structures and parameters for modeling small dyes attached to proteins was described. Furthermore, I established a detailed simulation protocol to obtain FRET efficiency histograms for two-color FRET and photon count rates for three-color FRET directly comparable to experimental measurements.

The treatment of fluorescent proteins attached to proteins via different linkers was discussed with the example of a FRET-based glucose sensor. I included additional SAXS data to determine a structural ensemble for the glucose sensor and established a protocol which enables the study of the formation of a heterodimer of the fluorescent proteins *in silico*.

The introduced parametrization procedures and simulation protocols are in principle applicable to arbitrary systems that involve dyes or fluorescent proteins.

The simulations provided new insights into the systems' dynamics by yielding relative distance and orientation distributions of FRET dyes. This was shown by testing different experimental settings such as different dye pairs, labeling sites, or the effects of three-color FRET. The C_α distance turned out to be an inappropriate measure for the inter-dye distance, as the protein restricts the dyes' motions. Permutation of the dyes with respect to labeling positions showed no altered results for CspTm, but it can be of importance in other systems. For example, in simulations of ClyA, the dyes were attached in different regions (inside and outside the pore), which resulted in different dynamics. Testing of different simulation lengths proved the suitability of the simulation method for sampling even conformationally diverse ensembles such as unfolded proteins. Simulations of three-color FRET facilitate the interpretation of various experiments, e. g., test different hypotheses of protein dynamics, labeling positions, or the effect of labeling isomers.

The presented model can improve planning experimental measurements as it allows to test different parameters such as dye pair, Förster radius, linker length, and labeling sites *in silico*. This simplifies selection of settings to e. g. best distinguish conformational states of interest.

As shown by direct comparison and quantitative agreement of FRET efficiency histograms from simulations with experimental data, the approach is able to achieve a realistic description of the systems. This directly complements experimental FRET efficiency distributions with atomically resolved structural ensembles from simulations. Hence, the simulation method provides novel and detailed insights into biomolecular processes. For example, simulations of ClyA revealed that steric limitations can explain the experimentally observed differences between protomer and dodecamer conformations.

Moreover, I probed the estimations of dye distance distributions based on simple models for data analysis. The presented method is in good agreement with both the accessible volume approach for folded proteins and the polymer model for unfolded proteins. In contrast to the accessible volume approach, it further includes orientational dynamics and is also applicable to unfolded structural ensembles or fluorescent proteins. Compared to the theory of unfolded proteins as polymer chains, it can be used to calculate effective lengths for the dye pairs to correct the relation between C_α and inter-dye distances.

The investigation of the interplay between FRET and SAXS measurements revealed an influence of the FRET dyes on SAXS intensity profiles, especially for small systems. The derived value of the radius of gyration R_g depends on the choice of dyes and labeling sites. R_g as calculated from simulations equivalent to the different experimental methods FRET and SAXS deviates systematically, which gave another possible reason for the diverging results seen in the two experiments. After inclusion of the known corrections, all derived R_g values are in good agreement.

Beyond that, the simulations can produce approximations for experimentally inaccessible quantities such as diffusion constants or rotational correlation times for conformationally restricted fluorescent proteins.

By studying a FRET-based glucose sensor comprising two fluorescent proteins, I demonstrated the suitability to use simulations to complement experimental measurements. I achieved a considerable improvement in modeling the glucose sensor, yielding structural ensembles and identifying possible effects contributing to the resulting experimental FRET data.

Furthermore, the approximation of $\overline{\kappa^2} = 2/3$ could be tested for the investigated systems. Considering the typical systems with small dyes, the assumption turned out to be mostly appropriate, whereas the orientation factor deviates for fluorescent proteins.

Due to its computational efficiency, the model is able to sufficiently sample even complex scenarios involving large structural ensembles such as intrinsically disordered or unfolded proteins, conformational transitions, folding intermediates, or large systems such as the presented biosensors with modest computational resources. In particular, it is most useful in the absence of non-native interactions as implemented in the framework of structure-based models. As the example of the glucose sensor has shown, it nevertheless is easily extendable to these interactions.

6.2 Outlook

The introduced simulation method enables various follow-up applications. To test different hypotheses for the underlying protein dynamics, two-, three-, or four-color FRET systems can be simulated and compared to experimental data to find the most probable scenarios. Properties determined by these simulations such as diffusion constants, residual anisotropy, or effective lengths for dye pairs in the simulation of unfolded proteins can be used for further calculations. Especially systems with restricted or slow dye motions and short or inflexible linkers can be tested with respect to their effect on κ^2 and FRET efficiencies. Knowing these previously inaccessible quantities enables the utilization of dyes with such restrictions for quantitative measurements.

To study different scenarios, additional interactions can be included, for example temporary sticking of the dyes to the protein surface. With complete atomically resolved trajectories including orientations and distances, this approach can be used to investigate the effects of fluorophore quenching as done in [156]. Through the inclusion of different van der Waals interactions between dyes and surface and comparison to residual anisotropies from experiments, the currently hardly understood dye-protein interactions can be investigated.

The low computational costs of the simulation method in comparison to regular MD simulations also make new scenarios accessible for simulation. Intrinsically disordered proteins (IDPs) have become of great interest in recent years, as they are a large and functionally important class of proteins. They do not adopt a single structure, but rather a structural ensemble, where structural dynamics plays an important role for their function. Modeling IDPs is a big challenge for the conventional methods [158]. The presented method now also provides a tool to model FRET studies of IDPs. It allows to adjust for studying them in different denaturant concentrations, e. g. by introduction of an overall attractive potential.

Furthermore, FRET studies of protein folding, large systems, or large structural ensembles as e. g. unfolded proteins can be simulated. Even complex large-scale conformational transitions between multiple states [159] can now be modeled with FRET fluorophores to study their dynamics.

As both SBM simulations and FRET measurements are further used to study ribonucleic acids (RNA), the model could be extended to describe these systems easily.

Finally, the glucose sensor can be further investigated by testing different interactions between fluorescent proteins and sensing protein or dimerization interactions at different rates. Comparison of the resulting FRET data to experiments can identify valid scenarios. The same simulation method can also be applied to other FRET-based biosensors to study and subsequently improve their function.

A Abbreviations

AMBER	Assisted Model Building with Energy Refinement
CI-2	Chymotrypsin Inhibitor 2
CHARMM	Chemistry at HARvard Macromolecular Mechanics
CFP	Cyan Fluorescent Protein (here: mTurquoise2)
ClyA	Cytolysin A
CspTm	Cold-shock protein from the hyperthermophilic bacterium *Thermotoga maritima*
DFT	Density Functional Theory
^{10}FNIII	Tenth type III domain of FibroNectin
FRET	Förster Resonance Energy Transfer
Glc-BP	Glucose-Binding Protein
GROMACS	GROningen MAchine for Chemical Simulation
IDP	Intrinsically Disordered Protein
MD	Molecular Dynamics
MSD	Mean Square Deviation
PDB	Protein Data Bank
RMSD	Root Mean Square Deviation
RMSF	Root Mean Square Fluctuation
RNA	RiboNucleic Acid
SAXS	Small Angle X-ray Scattering
SBM	Structure-Based Model
smFRET	single molecule Förster Resonance Energy Transfer
YFP	Yellow Fluorescent Protein (here: Venus)

B
Analysis of Simulations

One important quantity to analyze simulated trajectories is the root mean square deviation (RMSD) between two structures. After a least square fitting to the reference structure, it is defined by:

$$\mathrm{RMSD}(t) = \sqrt{\frac{1}{N_{\mathrm{atoms}}} \sum_{i=1}^{N_{\mathrm{atoms}}} \left| \boldsymbol{r}_i(t) - \boldsymbol{r}_{i,0} \right|^2}\,, \tag{B.1}$$

where $\boldsymbol{r}_i(t)$ is the position of atom i at time t and $\boldsymbol{r}_{i,0}$ is the position of atom i in the reference structure.

A measure for flexibility of a structure over sequence is given by the root mean square fluctuations (RMSF) which are calculated by:

$$\mathrm{RMSF}(i) = \sqrt{\frac{1}{t_{\mathrm{tot}}} \sum_{t_j=1}^{t_{\mathrm{tot}}} \left| \boldsymbol{r}_i(t_j) - \overline{\boldsymbol{r}}_i \right|^2}\,. \tag{B.2}$$

After a least mean square fit of the structure to a reference structure, the RMSF values give the time averaged spatial fluctuations of atoms or residues i around their time averaged position $\overline{\boldsymbol{r}}_i$.

A measure for the extent of a molecule is the radius of gyration R_{g}. `GROMACS` determines it as [84]:

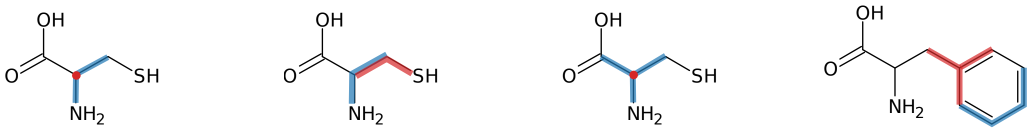
$$R_{\mathrm{g}} = \sqrt{\frac{\sum_i r_i^2 m_i}{\sum_i m_i}}\,, \tag{B.3}$$

where r_i is the distance of atom i to the center of mass of the molecule and m_i is the atom's mass, which is selected from a `GROMACS` parameter file according to the respective atom name.

C
Generation of Fluorophore Topology

For generation of the SBM force field parameters I need a topology for each fluorophore, i.e., the included atoms for all bonds, angles, proper and improper dihedral angles. This is done in a systematic way as follows (see also [113], SI).

I extract the bond information from the chemical structure. Then I assign an angle for every two bonds sharing a common atom (shown in Fig. C.1a). A proper dihedral angle is assigned to every two angles sharing a bond and not the middle atom (see Fig. C.1b). If two or more atoms of this dihedral angle are part of a ring (see Fig. C.1d), I change the dihedral angle into an improper dihedral angle to stabilize the rings. Similar to the implementation of amino acids in `eSBMTools` (see Sec. 3.2.1) I group and count all proper dihedral angles with mutual middle bond. In the parameter generation the respective force constants are divided by the number of group members to avoid overcounting.



(a) Angle.

(b) Proper dihedral angle.

(c) Improper dihedral angle.

(d) Dihedral angles in ring.

Figure C.1: Determination of angles and dihedral angles from bond information. (a) An angle is added for every two bonds (blue) sharing an atom (red). (b) A proper dihedral angle is added for every two angles (red and blue) sharing a bond. (c) An improper dihedral angle is added for three atoms bound to the same middle atom (red). (d) The proper dihedral angles involved in a ring (blue) are converted to improper dihedral angles. Improper dihedral angles with three or more atoms involved in a ring (red) are removed.

Also, I include an improper dihedral angle for each three atoms that are bound to the same forth atom (see Fig. C.1c). As the rings are already stabilized by improper dihedral angles, I remove the improper dihedral angles with three or more atoms in a ring (see Fig. C.1d).

The ordering scheme for the atoms of the improper dihedral angles follows the conventions for the CHARMM force field [160]. In the case of a single atom bound to four other atoms, all four improper dihedral angles are assigned to maintain symmetry, and the respective force constants are divided by four.

D Temperature Comparison

The temperature comparison is performed adhering to the description in [108] by comparing a regular MD simulation with SBM simulations. To determine a reasonable temperature for the SBM simulations I perform initial all-atom simulations of CspTm, ^{10}FNIII, ClyA monomer and protomer, and the fluorescent proteins CFP and YFP in the AMBER99 force field [87] with explicit water. The simulations are performed at the physiological temperatures used in the experiments. The time steps Δt and total simulation times t_{tot} for the simulations are shown in Tab. D.1. The first part of each simulation is discarded to prevent from equilibration artifacts (see the values for the time span used t_{used} in Tab. D.1).

The SBM simulations are performed over a wide temperature range, which along with the used time step Δt_{SBM} and the total simulation time $t_{\text{tot, SBM}}$ can be found in Tab. D.1.

Table D.1: Parameters used for the temperature comparison simulations. For each protein, physiological temperature T, time step Δt, total simulation time t_{tot}, and simulation time used for the temperature comparison t_{used} are given for the AMBER99 simulation. Furthermore, temperature range T_{SBM}, time step Δt_{SBM}, and total simulation time $t_{\text{tot, SBM}}$ for the simulations in the structure-based model are listed.

Protein	T	Δt	t_{tot}	t_{used}	T_{SBM}	Δt_{SBM}	$t_{\text{tot, SBM}}$
^{10}FNIII	296 K	1.5 fs	300 ns	20-300 ns	10-225	1.5 fs	30 ns
CspTm	295 K	2.0 fs	500 ns	100-500 ns	10-120	2.0 fs	200 ns
ClyA	295 K	2.0 fs	500 ns	100-500 ns	40-150	0.5 fs	2.5 ns
CFP	295 K	2.0 fs	500 ns	100-500 ns	30-100	0.5 fs	50 ns
YFP	295 K	2.0 fs	500 ns	100-500 ns	30-100	0.5 fs	50 ns

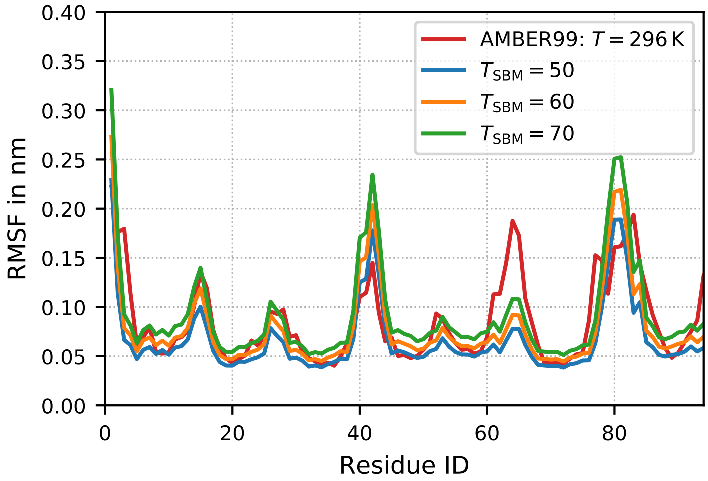


Figure D.1: RMSF values over residue index for ^{10}FNIII. The RMSF values are shown for the reference AMBER99 simulation (red) and SBM simulations for three different temperatures.

As an example, I show the results for ^{10}FNIII. In Fig. D.1 the root mean square fluctuations (RMSF) (see Sec. B) of the C$_\alpha$-atoms in the AMBER99 simulation and in the SBM simulations with three different temperatures are shown. The SBM RMSF curves all reflect the same overall behavior as the RMSF values of the AMBER99 simulation. The different SBM curves do not differ in shape, higher temperatures result in overall higher RMSF values.

I want to find the SBM temperature which gives the best approximation of the behavior in the AMBER99 force field, so I calculate the mean square deviation (MSD) of the RMSF curves with respect to the corresponding curve of the AMBER99 simulation. The results for ^{10}FNIII are shown in Fig. D.2 for a range of SBM temperatures. As expected, a large rise of the MSD values can be observed around the expected folding temperature of about $T = 120$. Furthermore, a clear minimum can be seen at a temperature of $T = 60$, which I choose as the temperature for ^{10}FNIII in the subsequent simulations. However, the variation of MSD around this temperature is small, so small changes in temperature are expected to make a negligible difference in the results of the simulations. This also justifies the temperature choice for CI-2 (see Sec. 4.3).

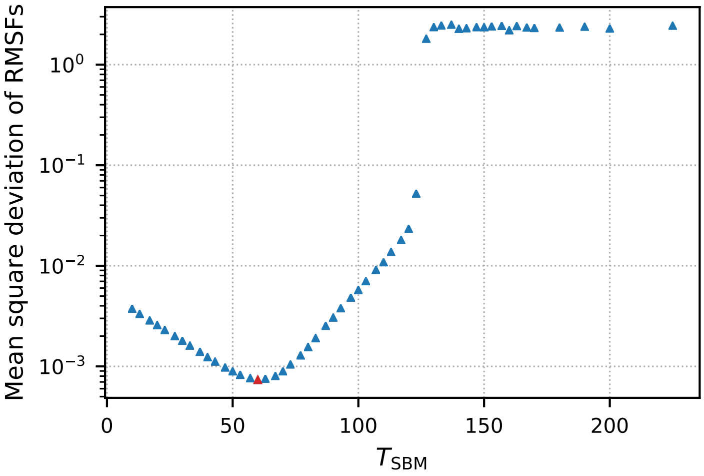


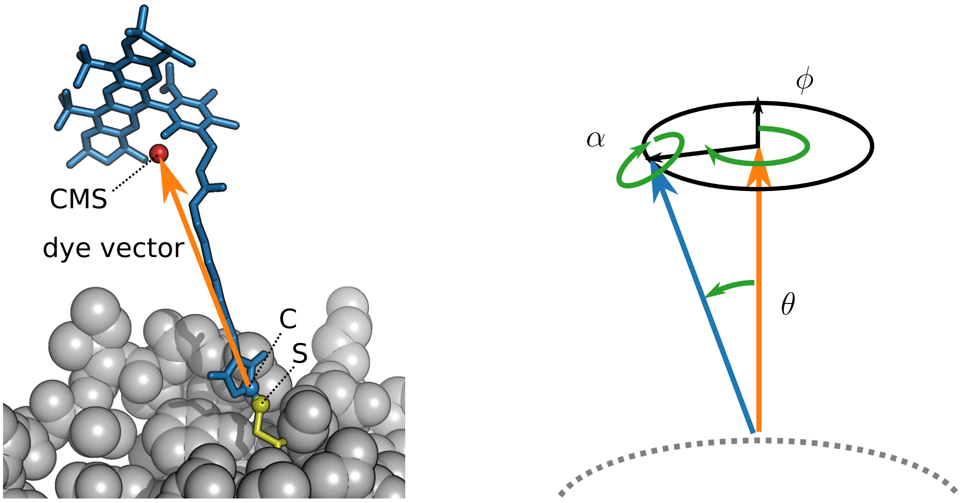
Figure D.2: Mean square deviations of the RMSF curves of ^{10}FNIII for different SBM temperatures in reference to the AMBER99 simulation. The lowest value occurs at $T = 60$ (red).

E
Merging of Structures

To merge two structures, e. g. the structure of a dye or a linker to a protein or a fluorescent protein to the respective linker, I use the following algorithm (see also [113], SI). As example, the case of attaching the structure of AF546 to CI-2 is shown in Fig. E.1a.

In the first step I determine the connection point C (blue sphere in Fig. E.1a) as the position for the dye atom that is bound to the sulfur atom S of the cysteine residue (yellow sphere in Fig. E.1a). It is chosen to be ideally pointing orthogonally away from the protein surface in a distance of 0.14 nm to the sulfur atom while avoiding steric clashes with other atoms.

The vector connecting S and C serves as a starting orientation for the "dye vector" (orange arrow in Fig. E.1a). The dye vector connects the dye's center of mass (CMS, red sphere in Fig. E.1a) with C. Then I adjust the orientation of the dye vector by gradually rotating the dye in a cone around the starting vector while checking for clashes. The angles used for this are depicted in Fig. E.1b, where $\theta \in [0, 90°]$, and $\phi, \alpha \in [0, 360°[$.



(a) CI-2 (gray spheres) and AF546 (blue). The connection point C is shown as blue sphere, the sulfur atom S of the cysteine residue the dye is attached to is depicted in yellow. The dyes' center of mass (CMS, red sphere) and the dye vector (orange arrow) are shown.

(b) Definition of the different angles for rotation. The starting dye vector (orange), the resulting vector (blue) and the angles θ (between starting and resulting vector), ϕ and the angle α as the rotation around the dye axis itself are shown.

Figure E.1: Example of merging two structures and definition of rotation angles.

F
Sequences of Glucose Sensor Variants

The sequences of the glucose sensor variants used in the SAXS experiments. The underlined parts are omitted in the simulations but should not have a considerable effect on the results. The amino acids are colored according to the protein they belong to in green (Glc-BP), cyan (CFP), and yellow (YFP). The flexible $(GGS)_4$-linker is colored in purple and the restriction sites and His-tags are shown in black.

The mutation of residue 206 of CFP (alanine) into a lysine done in the experiments is neglected here.

The sequences used for the experimental FRET measurements [132] deviate slightly, as sensor 2 and sensor 4 have an N-terminal His-tag instead of a C-terminal His-tag. As no considerable effect is expected this was neglected for the simulations. The sequences used for the FRET measurements can be found in [132].

Sensor 1

MRGSHHHHHHGMASMTGGQQMGRDLYDDDDKEPGRADTRIGVTIYKAAAMVSKGEELFTGVVPI
LVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLSWGVQCFARYPDHM
KQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLE
YNYFSDNVYITADKQKNGIKANFKIRHNIEDGGVQLADHYQQNTPIGDGPVLLPDNHYLSTQSK
LSKDPNEKRDHMVLLEFVTAAGITLGMDELYGSDLVDNFMSVVRKAIEQDAKAAPDVQLLMNDS
QNDQSKQNDQIDVLLAKGVKALAINLVDPAAAGTVIEKARGQNVPVVFFNKEPSRKALDSYDKA
YYVGTDSKESGIIQGDLIAKHWAANQGWDLNKDGQIQFVLLKGEPGHPDAEARTTYVIKELNDK
GIKTEQLQLDTAMWDTAQAKDKMDAWLSGPNANKIEVVIANNDAMAMGAVEALKAHNKSSIPVF
GVDALPEALALVKSGALAGTVLNDANNQAKATFDLAKNLADGKGAADGTNWKIDNKVVRVPYVG
VDKDNLAEFSKKEFVDGGMVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLK
LICTTGKLPVPWPTLVTTLGYGLQCFARYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTR

AEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYITADKQKNGIKANFKIRHNIED
GGVQLADHYQQNTPIGDGPVLLPDNHYLSYQSALSKDPNEKRDHMVLLEFVTAAGITLGMDELY
K

Sensor 2

MADTRIGVTIYKAAAMVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFIC
TTGKLPVPWPTLVTTLSWGVQCFARYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEV
KFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYFSDNVYITADKQKNGIKANFKIRHNIEDGGV
QLADHYQQNTPIGDGPVLLPDNHYLSTQSKLSKDPNEKRDHMVLLEFVTAAGITLGMDELYGSG
GSGGSGGSGGSPGDNFMSVVRKAIEQDAKAAPDVQLLMNDSQNDQSKQNDQIDVLLAKGVKALA
INLVDPAAAGTVIEKARGQNVPVVFFNKEPSRKALDSYDKAYYVGTDSKESGIIQGDLIAKHWA
ANQGWDLNKDGQIQFVLLKGEPGHPDAEARTTYVIKELNDKGIKTEQLQLDTAMWDTAQAKDKM
DAWLSGPNANKIEVVIANNDAMAMGAVEALKAHNKSSIPVFGVDALPEALALVKSGALAGTVLN
DANNQAKATFDLAKNLADGKGAADGTNWKIDNKVVRVPYVGVDKDNLAEFSKKEFVDGGMVSKG
EELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKLICTTGKLPVPWPTLVTTLGYGL
QCFARYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKE
DGNILGHKLEYNYNSHNVYITADKQKNGIKANFKIRHNIEDGGVQLADHYQQNTPIGDGPVLLP
DNHYLSYQSALSKDPNEKRDHMVLLEFVTAAGITLGMDELYKHHHHHH

Sensor 4

MADTRIGVTIYKAAAMVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFIC
TTGKLPVPWPTLVTTLSWGVQCFARYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEV
KFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYFSDNVYITADKQKNGIKANFKIRHNIEDGGV
QLADHYQQNTPIGDGPVLLPDNHYLSTQSKLSKDPNEKRDHMVLLEFVTAAGITLGMDELYGSD
LVDNFMSVVRKAIEQDAKAAPDVQLLMNDSQNDQSKQNDQIDVLLAKGVKALAINLVDPAAAGT
VIEKARGQNVPVVFFNKEPSRKALDSYDKAYYVGTDSKESGIIQGDLIAKHWAANQGWDLNKDG
QIQFVLLKGEPGHPDAEARTTYVIKELNDKGIKTEQLQLDTAMWDTAQAKDKMDAWLSGPNANK
IEVVIANNDAMAMGAVEALKAHNKSSIPVFGVDALPEALALVKSGALAGTVLNDANNQAKATFD
LAKNLADGKGAADGTNWKIDNKVVRVPYVGVDKDNLAEFSKKEFGGSGGSGGSGGSVDGGMVSK
GEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKLICTTGKLPVPWPTLVTTLGYG
LQCFARYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFK
EDGNILGHKLEYNYNSHNVYITADKQKNGIKANFKIRHNIEDGGVQLADHYQQNTPIGDGPVLL
PDNHYLSYQSALSKDPNEKRDHMVLLEFVTAAGITLGMDELYKHHHHHH

G
Simulation of Glc-BP with and without Glucose

I perform SBM simulations of both Glc-BP$^{+\mathrm{G}}$ and Glc-BP$^{\text{-G}}$ at $T = 70$ for $t_{\mathrm{SBM}} = 50\,\mathrm{ns}$ to check whether the two structures are well distinguishable in the simulations, even without explicit simulation of the glucose molecule. Fig. G.1 shows that for the simulation of Glc-BP$^{+\mathrm{G}}$, the RMSD to the starting structure is overall lower than the RMSD to the glucose free structure. In the simulation of Glc-BP$^{\text{-G}}$ the RMSD values come closer, which is mostly explained by the higher flexibility of the glucose free structure. The two parts of Glc-BP are further apart so there are less contacts to stabilize their respective position. It is not surprising that the structure also occupies conformations close to the structure of Glc-BP$^{+\mathrm{G}}$. Still, the overall RMSD shows that the structures are distinguishable in simulations.

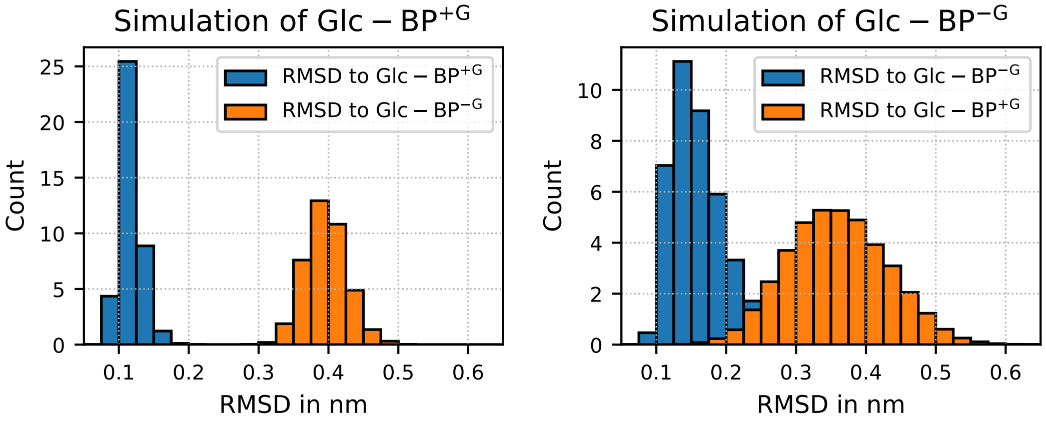


Figure G.1: SBM simulations of Glc-BP^{+G} and Glc-BP^{-G} at T = 70. The distributions of RMSD values during the simulation with respect to the starting structure (blue) and the respective other structure (orange) are shown.

H

Residue Numbering Scheme for CspTm

Table H.1: Variants of CspTm for different labeling schemes. The amino acid sequences of the CspTm variants used in the simulations and the experiments are shown, along with the numbering scheme of the residues. The cysteine residue inserted for the C2/C68 variant is shown in red and the residues mutated to cysteine in the respective variants are shown in green.

C2/C68	M	C	R	G	K	V	K	F	F	D	S	K	K	G	Y	G	F
C11/C68	M	-	R	G	K	V	K	F	F	D	C	K	K	G	Y	G	F
C23/C68	M	-	R	G	K	V	K	F	F	D	S	K	K	G	Y	G	F
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
	I	T	K	D	E	G	G	D	V	F	V	H	F	S	A	I	E
	I	T	K	D	E	G	G	D	V	F	V	H	F	S	A	I	E
	I	T	K	D	E	C	G	D	V	F	V	H	F	S	A	I	E
	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
	M	E	G	F	K	T	L	K	E	G	Q	V	V	E	F	E	I
	M	E	G	F	K	T	L	K	E	G	Q	V	V	E	F	E	I
	M	E	G	F	K	T	L	K	E	G	Q	V	V	E	F	E	I
	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51
	Q	E	G	K	K	G	G	Q	A	A	H	V	K	V	V	E	C
	Q	E	G	K	K	G	G	Q	A	A	H	V	K	V	V	E	C
	Q	E	G	K	K	G	G	Q	A	A	H	V	K	V	V	E	C
	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68

I

Additional Formulas for FRET Efficiency and Anisotropy

As an alternative to the commonly used FRET efficiency histogram described in Sec. 2.2.1, the FRET efficiency can be calculated via the donor lifetime in absence of the acceptor (τ_{D}) and in presence of the acceptor (τ_{DA}), by:

$$E = 1 - \frac{\tau_{\mathrm{DA}}}{\tau_{\mathrm{D}}} \,. \tag{I.1}$$

The fluorescence r in experiments denotes the integrated fluorescence anisotropy decay. In simulations, assuming a single exponential decay of fluorescence, the anisotropy can be calculated via the Perrin equation [47]:

$$r = \frac{r_0}{1 + \tau/\tau_{\mathrm{rot}}} \,, \tag{I.2}$$

where r_0 is the fundamental anisotropy, τ is the fluorophore lifetime, and τ_{rot} the rotational correlation time of the fluorophore.

J
Starting Structures for Sensor 1 and Sensor 4

The structures resulting for sensor 1 and sensor 4 with the lowest χ^2 values in comparison to the respective experimental SAXS data are depicted in Figs. J.1 and J.2, respectively. Apparently, a variety of structures fits the experimental SAXS curves.

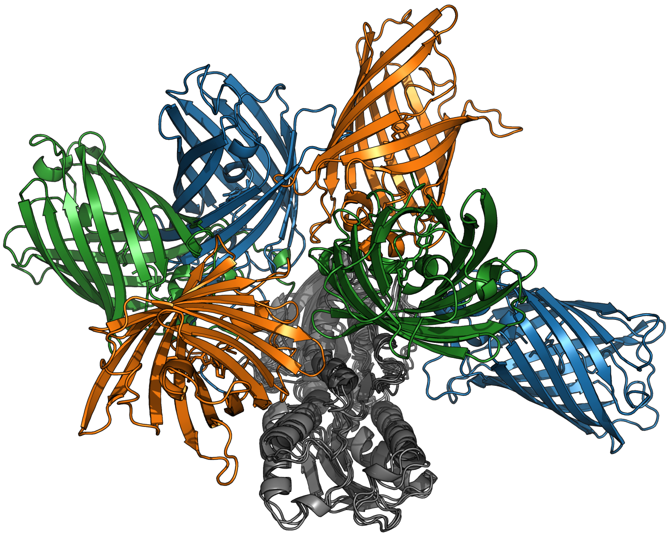


Figure J.1: Starting structures for sensor 1. Glc-BP is depicted in gray, the different pairs of fluorescent proteins are depicted in blue (sensor 1-1), orange (sensor 1-2), and green (sensor 1-3). The fluorescent protein on the left is CFP, the fluorescent protein on the right YFP, respectively.

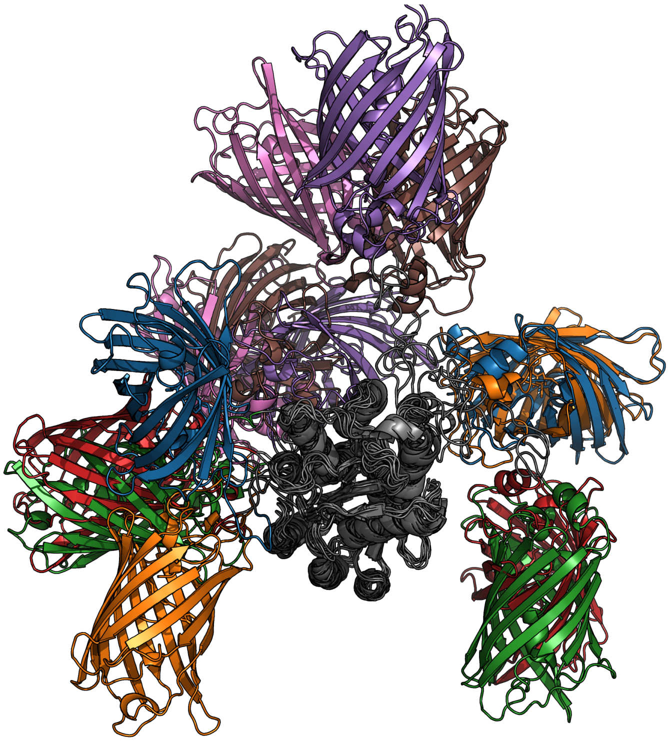


Figure J.2: Starting structures for sensor 4. Glc-BP is depicted in gray, the different pairs of fluorescent proteins are depicted in blue (sensor 4A-1), orange (sensor 4A-2), green (sensor 4B-1), red (sensor 4B-2), purple (sensor 4B-3), brown (sensor 4B-4), and pink (sensor 4C-2).

Bibliography

[1] L. Stryer. *Fluorescence Energy Transfer as a Spectroscopic Ruler.* Annual Review of Biochemistry 1978, 47(1), 819–846.

[2] S. Müller-Späth, A. Soranno, V. Hirschfeld, H. Hofmann, S. Ruegger, L. Reymond, D. Nettels, and B. Schuler. *Charge interactions can dominate the dimensions of intrinsically disordered proteins.* Proceedings of the National Academy of Sciences 2010, 107, 14609–14614.

[3] D. Nettels, S. Müller-Späth, F. Kuster, H. Hofmann, D. Haenni, S. Ruegger, L. Reymond, A. Hoffmann, J. Kubelka, B. Heinz, K. Gast, R. B. Best, and B. Schuler. *Single-molecule spectroscopy of the temperature-induced collapse of unfolded proteins.* Proceedings of the National Academy of Sciences 2009, 106(49), 20740–20745.

[4] E. Sherman and G. Haran. *Coil-globule transition in the denatured state of a small protein.* Proceedings of the National Academy of Sciences 2006, 103(31), 11539–11543.

[5] B. Schuler and H. Hofmann. *Single-molecule spectroscopy of protein folding dynamics—expanding scope and timescales.* Current Opinion in Structural Biology 2013, 23(1), 36–47.

[6] R. Rieger, A. Kobitski, H. Sielaff, and G. U. Nienhaus. *Evidence of a Folding Intermediate in RNase H from Single-Molecule FRET Experiments.* ChemPhysChem 2011, 12(3), 627–633.

[7] R. Rieger and G. U. Nienhaus. *A combined single-molecule FRET and tryptophan fluorescence study of RNase H folding under acidic conditions.* Chemical Physics 2012, 396(1), 3–9.

[8] Y. Gambin, A. Schug, E. A. Lemke, J. J. Lavinder, A. C. M. Ferreon, T. J. Magliery, J. N. Onuchic, and A. A. Deniz. *Direct single-molecule observation of a protein living in two opposed native structures.* Proceedings of the National Academy of Sciences 2009, 106(25), 10153–10158.

[9] R. E. Campbell. *Fluorescent-Protein-Based Biosensors: Modulation of Energy Transfer as a Design Principle.* Analytical Chemistry 2009, 81(15), 5972–5979.

[10] H. J. Carlson and R. E. Campbell. *Genetically encoded FRET-based biosensors for multiparameter fluorescence imaging.* Current Opinion in Biotechnology 2009, 20(1), 19–27.

[11] M. Mohsin, A. Ahmad, and M. Iqbal. *FRET-based genetically-encoded sensors for quantitative monitoring of metabolites.* Biotechnology Letters 2015, 37(10), 1919–1928.

[12] L. Sanford and A. Palmer. *Chapter One - Recent Advances in Development of Genetically Encoded Fluorescent Sensors. Enzymes as Sensors.* Vol. 589. Methods in Enzymology. Academic Press, 2017, 1 –49.

[13] K. Deuschle, S. Okumoto, M. Fehr, L. L. Looger, L. Kozhukh, and W. B. Frommer. *Construction and optimization of a family of genetically encoded metabolite sensors by semirational protein engineering.* Protein science : a publication of the Protein Society 2005, 14(9), 2304–14.

[14] L. Lindenburg and M. Merkx. *Engineering Genetically Encoded FRET Sensors.* Sensors 2014, 14(7), 11691–11713.

[15] Z. Wang and D. E. Makarov. *Nanosecond Dynamics of Single Polypeptide Molecules Revealed by Photoemission Statistics of Fluorescence Resonance Energy Transfer: A Theoretical Study.* The Journal of Physical Chemistry B 2003, 107(23), 5617–5622.

[16] A. Muschielok, J. Andrecka, A. Jawhari, F. Brückner, P. Cramer, and J. Michaelis. *A nano-positioning system for macromolecular structural analysis.* Nature Methods 2008, 5(11), 965–971.

[17] A. Muschielok and J. Michaelis. *Application of the Nano-Positioning System to the Analysis of Fluorescence Resonance Energy Transfer Networks.* The Journal of Physical Chemistry B 2011, 115(41), 11927–11937.

[18] S. Sindbert, S. Kalinin, H. Nguyen, A. Kienzler, L. Clima, W. Bannwarth, B. Appel, S. Muller, and C. A. M. Seidel. *Accurate Distance Determination of Nucleic Acids via Forster Resonance Energy Transfer: Implications of Dye Linker Length and Rigidity.* Journal of the American Chemical Society 2011, 133(8), 2463–2480.

[19] S. Kalinin, T. Peulen, S. Sindbert, P. J. Rothwell, S. Berger, T. Restle, R. S. Goody, H. Gohlke, and C. A. M. Seidel. *A toolkit and benchmark study for FRET-restrained high-precision structural modeling.* Nature Methods 2012, 9(12), 1218–1225.

[20] E. Pham, J. Chiang, I. Li, W. Shum, and K. Truong. *A Computational Tool for Designing FRET Protein Biosensors by Rigid-Body Sampling of Their Conformational Space.* Structure 2007, 15(5), 515–523.

[21] B. Schuler, A. Soranno, H. Hofmann, and D. Nettels. *Single-Molecule FRET Spectroscopy and the Polymer Physics of Unfolded and Intrinsically Disordered Proteins.* Annual Review of Biophysics 2016, 45(1), 207–231.

[22] G. F. Schröder, U. Alexiev, and H. Grubmüller. *Simulation of Fluorescence Anisotropy Experiments: Probing Protein Dynamics.* Biophysical Journal 2005, 89, 3757–3770.

[23] R. B. Best, H. Hofmann, D. Nettels, and B. Schuler. *Quantitative Interpretation of FRET Experiments via Molecular Simulation: Force Field and Validation.* Biophysical Journal 2015, 108(11), 2721–2731.

[24] R. B. Best, K. A. Merchant, I. V. Gopich, B. Schuler, A. Bax, and W. A. Eaton. *Effect of flexibility and cis residues in single-molecule FRET studies of polyproline.* Proceedings of the National Academy of Sciences 2007, 104(48), 18964–18969.

[25] M. J. Shoura, R. U. Ranatunga, S. A. Harris, S. O. Nielsen, and S. D. Levene. *Contribution of Fluorophore Dynamics and Solvation to Resonant Energy Transfer in Protein-DNA Complexes: A Molecular-Dynamics Study.* Biophysical Journal 2014, 107(3), 700–710.

[26] E. V. Kuzmenkina, C. D. Heyes, and G. U. Nienhaus. *Single-molecule Forster resonance energy transfer study of protein dynamics under denaturing conditions.* Proceedings of the National Academy of Sciences 2005, 102(43), 15471–15476.

[27] M. M. Reif and C. Oostenbrink. *Molecular dynamics simulation of configurational ensembles compatible with experimental FRET efficiency data through a restraint on instantaneous FRET efficiencies.* Journal of Computational Chemistry 2014, 35(32), 2319–2332.

[28] D. B. VanBeek, M. C. Zwier, J. M. Shorb, and B. P. Krueger. *Fretting about FRET: Correlation between κ and R.* Biophysical Journal 2007, 92(12), 4168–4178.

[29] P. C. Whitford, J. K. Noel, S. Gosavi, A. Schug, K. Y. Sanbonmatsu, and J. N. Onuchic. *An all-atom structure-based potential for proteins: Bridging minimal models with all-atom empirical forcefields.* Proteins: Structure, Function, and Bioinformatics 2009, 75(2), 430–441.

[30] J. D. Bryngelson and P. G. Wolynes. *Spin glasses and the statistical mechanics of protein folding.* Proceedings of the National Academy of Sciences 1987, 84(21), 7524–7528.

[31] P. E. Leopold, M. Montal, and J. N. Onuchic. *Protein folding funnels: a kinetic approach to the sequence-structure relationship.* Proceedings of the National Academy of Sciences 1992, 89(18), 8721–8725.

[32] H. Frauenfelder, S. G. Sligar, and P. G. Wolynes. *The energy landscapes and motions of proteins.* Science (New York, N.Y.) 1991, 254(5038), 1598–603.

[33] J. N. Onuchic and P. G. Wolynes. *Theory of protein folding.* Current Opinion in Structural Biology 2004, 14(1), 70–75.

[34] C. A. McPhalen, I. Svendsen, I. Jonassen, and M. N. G. James. *Crystal and molecular structure of chymotrypsin inhibitor 2 from barley seeds in complex with subtilisin Novo.* Proceedings of the National Academy of Sciences 1985, 82(21), 7242–7246.

[35] C. B. Anfinsen. *Principles that Govern the Folding of Protein Chains.* Science 1973, 181(4096), 223–230.

[36] C. Levinthal. *How to Fold Graciously. Mossbauer Spectroscopy in Biological Systems: Proceedings of a meeting held at Allerton House, Monticello, Illinois.* University of Illinois Press, 1969, 22–24.

[37] L. Qiu, S. A. Pabit, A. E. Roitberg, and S. J. Hagen. *Smaller and Faster: The 20-Residue Trp-Cage Protein Folds in 4 μs.* Journal of the American Chemical Society 2002, 124(44), 12952–12953.

[38] J. Kubelka, J. Hofrichter, and W. A. Eaton. *The protein folding 'speed limit'.* Current Opinion in Structural Biology 2004, 14(1), 76–88.

[39] R. A. Goldbeck, Y. G. Thomas, E. Chen, R. M. Esquerra, and D. S. Kliger. *Multiple pathways on a protein-folding energy landscape: Kinetic evidence.* Proceedings of the National Academy of Sciences 1999, 96, 2782–2787.

[40] K. A. Dill and H. S. Chan. *From Levinthal to pathways to funnels.* Nature Structural & Molecular Biology 1997, 4(1), 10–19.

[41] A. Schug and J. N. Onuchic. *From protein folding to protein function and biomolecular binding by energy landscape theory.* Current Opinion in Pharmacology 2010, 10(6), 709–714.

[42] E. Sisamakis, A. Valeri, S. Kalinin, P. J. Rothwell, and C. A. Seidel. *Chapter 18 - Accurate Single-Molecule FRET Studies Using Multiparameter Fluorescence Detection. Single Molecule Tools, Part B:Super-Resolution, Particle Tracking, Multiparameter, and Force Based Methods.* Vol. 475. Methods in Enzymology. Academic Press, 2010, 455 –514.

[43] T. Förster. *Zwischenmolekulare Energiewanderung und Fluoreszenz.* Annalen der Physik 1948, 437, 55–75.

[44] T.-O. Peulen and C. A. M. Seidel. *Struktur und Dynamik von Biomolekülen mit high precision-FRET.* BIOspektrum 2011, 17(7), 765–767.

[45] G. Schröder and H. Grubmüller. *FRETsg: Biomolecular structure model building from multiple FRET experiments.* Computer Physics Communications 2004, 158(3), 150–157.

[46] T. Ohashi, S. D. Galiacy, G. Briscoe, and H. P. Erickson. *An experimental study of GFP-based FRET, with application to intrinsically unstructured proteins.* Protein Science 2007, 16(7), 1429–1438.

[47] J. R. Lakowicz. *Principles of Fluorescence Spectroscopy.* Springer US, 2006.

[48] D. L. Andrews and D. S. Bradshaw. *Virtual photons, dipole fields and energy transfer: a quantum electrodynamical approach.* European Journal of Physics 2004, 25(6), 845–858.

[49] R. Roy, S. Hohng, and T. Ha. *A practical guide to single-molecule FRET.* Nature Methods 2008, 5(6), 507–516.

[50] URL: http://www.thermofisher.com.

[51] F Yang, L. G. Moss, and G. N. Phillips. *The molecular structure of green fluorescent protein.* Nature biotechnology 1996, 14(10), 1246–51.

[52] R. Y. Tsien. *The green fluorescent protein.* Annual Review of Biochemistry 1998, 67(1), 509–544.

[53] H. M. Watrob, C.-P. Pan, and M. D. Barkley. *Two-Step FRET as a Structural Tool.* Journal of the American Chemical Society 2003, 125(24), 7336–7343.

[54] S. Lee, J. Lee, and S. Hohng. *Single-Molecule Three-Color FRET with Both Negligible Spectral Overlap and Long Observation Time.* PLoS ONE 2010, 5(8), e12270.

[55] S. Hohng, C. Joo, and T. Ha. *Single-Molecule Three-Color FRET.* Biophysical Journal 2004, 87(2), 1328–1337.

[56] S. Lee and S. Hohng. *An Optical Trap Combined with Three-Color FRET.* Journal of the American Chemical Society 2013, 135(49), 18260–18263.

[57] J. Jung, K. Y. Han, H. R. Koh, J. Lee, Y. M. Choi, C. Kim, and S. K. Kim. *Effect of Single-Base Mutation on Activity and Folding of 10-23 Deoxyribozyme Studied by Three-Color Single-Molecule ALEX FRET.* The Journal of Physical Chemistry B 2012, 116(9), 3007–3012.

[58] S. Milles, C. Koehler, Y. Gambin, A. A. Deniz, and E. A. Lemke. *Intramolecular three-colour single pair FRET of intrinsically disordered proteins with increased dynamic range.* Molecular BioSystems 2012, 8(10), 2531.

[59] S. Voss, L. Zhao, X. Chen, F. Gerhard, and Y.-W. Wu. *Generation of an intramolecular three-color fluorescence resonance energy transfer probe by site-specific protein labeling.* Journal of Peptide Science 2014, 20(2), 115–120.

[60] J. Lee, S. Lee, K. Ragunathan, C. Joo, T. Ha, and S. Hohng. *Single-Molecule Four-Color FRET.* Angewandte Chemie International Edition 2010, 49(51), 9922–9925.

[61] G. Lipari and A. Szabo. *Effect of librational motion on fluorescence depolarization and nuclear magnetic resonance relaxation in macromolecules and membranes.* Biophysical Journal 1980, 30(3), 489–506.

[62] F. Hillger, D. Hänni, D. Nettels, S. Geister, M. Grandin, M. Textor, and B. Schuler. *Probing Protein-Chaperone Interactions with Single-Molecule Fluorescence Spectroscopy.* Angewandte Chemie International Edition 2008, 47(33), 6184–6188.

[63] D. Nettels, A. Hoffmann, and B. Schuler. *Unfolded Protein and Peptide Dynamics Investigated with Single-Molecule FRET and Correlation Spectroscopy from Picoseconds to Seconds.* The Journal of Physical Chemistry B 2008, 112(19), 6137–6146.

[64] K. Kinosita, S. Kawato, and A. Ikegami. *A theory of fluorescence polarization decay in membranes.* Biophysical Journal 1977, 20(3), 289–305.

[65] M. P. Allen and D. J. Tildesley. *Computer Simulation of Liquids.* New York, NY, USA: Clarendon Press, 1989.

[66] A. Kusumi, Y. Sako, and M. Yamamoto. *Confined lateral diffusion of membrane receptors as studied by single particle tracking (nanovid microscopy). Effects of calcium-induced differentiation in cultured epithelial cells.* Biophysical Journal 1993, 65(5), 2021–2040.

[67] M. J. Saxton and K. Jacobson. *SINGLE-PARTICLE TRACKING: Applications to Membrane Dynamics.* Annual Review of Biophysics and Biomolecular Structure 1997, 26(1), 373–399.

[68] D. I. Svergun and M. H. J. Koch. *Small-angle scattering studies of biological macromolecules in solution.* Reports on Progress in Physics 2003, 66(10), 1735–1782.

[69] C. D. Putnam, M. Hammel, G. L. Hura, and J. A. Tainer. *X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution.* Quarterly reviews of biophysics 2007, 40(3), 191–285.

[70] A. G. Kikhney and D. I. Svergun. *A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins.* FEBS Letters 2015, 589(19), 2570–2577.

[71] P. Debye. *Zerstreuung von Röntgenstrahlen.* Annalen der Physik 1915, 351(6), 809–823.

[72] A Guiner, G. Fournet, and C Walker. *Small angle scattering of X-rays.* Jahn Willey-Champan, New-York 1955.

[73] G. Porod. *Die Röntgenkleinwinkelstreuung von dichtgepackten kolloiden Systemen.* Kolloid-Zeitschrift 1951, 124(2), 83–114.

[74] D. Svergun, C. Barberato, and M. H. Koch. *CRYSOL - A program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates.* Journal of Applied Crystallography 1995, 28(6), 768–773.

[75] P. J. Flory. *The Configuration of Real Polymer Chains.* The Journal of Chemical Physics 1949, 17, 303–310.

[76] D. Johansen, J. Trewhella, and D. P. Goldenberg. *Fractal dimension of an intrinsically disordered protein: Small-angle X-ray scattering and computational study of the bacteriophage λ N protein.* Protein Science 2011, 20(12), 1955–1970.

[77] H. Hofmann, A. Soranno, A. Borgia, K. Gast, D. Nettels, and B. Schuler. *Polymer scaling laws of unfolded and intrinsically disordered proteins quantified with single-molecule spectroscopy.* Proceedings of the National Academy of Sciences 2012, 109(40), 16155–16160.

[78] M. Aznauryan, L. Delgado, A. Soranno, D. Nettels, J.-R. Huang, A. M. Labhardt, S. Grzesiek, and B. Schuler. *Comprehensive structural and dynamical view of an unfolded protein from the combination of single-molecule FRET, NMR, and SAXS.* Proceedings of the National Academy of Sciences 2016, 113(37), E5389–E5398.

[79] D. E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R. O. Dror, M. P. Eastwood, J. A. Bank, J. M. Jumper, J. K. Salmon, Y. Shan, and W. Wriggers. *Atomic-Level Characterization of the Structural Dynamics of Proteins.* Science 2010, 330(6002), 341–346.

[80] O. F. Lange, N.-A. Lakomek, C. Fares, G. F. Schröder, K. F. A. Walter, S. Becker, J. Meiler, H. Grubmüller, C. Griesinger, and B. L. de Groot. *Recognition Dynamics Up to Microseconds Revealed from an RDC-Derived Ubiquitin Ensemble in Solution.* Science 2008, 320(5882), 1471–1475.

[81] E. M. Puchner, A. Alexandrovich, A. L. Kho, U. Hensen, L. V. Schafer, B. Brandmeier, F. Grater, H. Grubmüller, H. E. Gaub, and M. Gautel. *Mechanoenzymatics of titin kinase.* Proceedings of the National Academy of Sciences 2008, 105(36), 13385–13390.

[82] G. F. Schröder, U. Alexiev, and H. Grubmüller. *Simulation of Fluorescence Anisotropy Experiments: Probing Protein Dynamics.* Biophysical Journal 2005, 89(6), 3757–3770.

[83] B. Corry and D. Jayatilaka. *Simulation of Structure, Orientation, and Energy Transfer between AlexaFluor Molecules Attached to MscL.* Biophysical Journal 2008, 95(6), 2711–2721.

[84] D. van der Spoel, E. Lindahl, B. Hess, A. R. van Buuren, E. Apol, P. J. Meulenhoff, D. P. Tieleman, A. L. T. M. Sijbers, K. A. Feenstra, R. van Drunen, and H. J. C. Berendsen. *Gromacs User Manual version 4.5.6.* URL: http://www.gromacs.org, 2010.

[85] W. F. Van Gunsteren and H. J. C. Berendsen. *A Leap-frog Algorithm for Stochastic Dynamics.* Molecular Simulation 1988, 1, 173–185.

[86] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. *A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules.* Journal of the American Chemical Society 1995, 117(19), 5179–5197.

[87] J. Wang, P. Cieplak, and P. A. Kollman. *How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules?* Journal of Computational Chemistry 2000, 21(12), 1049–1074.

[88] A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, and M. Karplus. *All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins.* The Journal of Physical Chemistry B 1998, 102(18), 3586–3616.

[89] A. D. Mackerell, M. Feig, and C. L. Brooks. *Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations.* Journal of Computational Chemistry 2004, 25(11), 1400–15.

[90] G. A. Voth. *Coarse-Graining of Condensed Phase and Biomolecular Systems.* Boca Raton, FL: CRC Press/Taylor and Francis Group, 2009.

[91] J. D. Bryngelson, J. N. Onuchic, N. D. Socci, and P. G. Wolynes. *Funnels, pathways, and the energy landscape of protein folding: A synthesis.* Proteins: Structure, Function, and Genetics 1995, 21(3), 167–195.

[92] C. Clementi, H. Nymeyer, and J. N. Onuchic. *Topological and energetic factors: what determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? an investigation for small globular proteins.* Journal of Molecular Biology 2000, 298(5), 937–953.

[93] L. L. Chavez, J. N. Onuchic, and C. Clementi. *Quantifying the roughness on the free energy landscape: entropic bottlenecks and protein folding rates.* Journal of the American Chemical Society 2004, 126(27), 8426–32.

[94] C. Sinner, B. Lutz, S. John, I. Reinartz, A. Verma, and A. Schug. *Simulating Biomolecular Folding and Function by Native-Structure-Based/Go-Type Models.* Israel Journal of Chemistry 2014, 54(8-9), 1165–1175.

[95] A. Schug, M. Weigt, J. N. Onuchic, T. Hwa, and H. Szurmant. *High-resolution protein complexes from integrating genomic information with molecular simulation.* Proceedings of the National Academy of Sciences 2009, 106(52), 22124–22129.

[96] C. Sinner, B. Lutz, A. Verma, and A. Schug. *Revealing the global map of protein folding space by large-scale simulations.* The Journal of Chemical Physics 2015, 143(24), 243154.

[97] M. B. Borgia, A. Borgia, R. B. Best, A. Steward, D. Nettels, B. Wunderlich, B. Schuler, and J. Clarke. *Single-molecule fluorescence reveals sequence-specific misfolding in multidomain proteins.* Nature 2011, 474(7353), 662–665.

[98] P. Tian and R. B. Best. *Structural Determinants of Misfolding in Multidomain Proteins.* PLOS Computational Biology 2016, 12, 1–28.

[99] P. C. Whitford, O. Miyashita, Y. Levy, and J. N. Onuchic. *Conformational Transitions of Adenylate Kinase: Switching by Cracking.* Journal of Molecular Biology 2007, 366(5), 1661–1671.

[100] A. Schug, P. C. Whitford, Y. Levy, and J. N. Onuchic. *Mutations as trapdoors to two competing native conformations of the Rop-dimer.* Proceedings of the National Academy of Sciences 2007, 104(45), 17674–17679.

[101] P. C. Whitford and K. Y. Sanbonmatsu. *Simulating movement of tRNA through the ribosome during hybrid-state formation.* The Journal of Chemical Physics 2013, 139(12), 121919.

[102] P. C. Whitford, A. Schug, J. Saunders, S. P. Hennelly, J. N. Onuchic, and K. Y. Sanbonmatsu. *Nonlocal Helix Formation Is Key to Understanding S-Adenosylmethionine-1 Riboswitch Function.* Biophysical Journal 2009, 96(2), L7–L9.

[103] B. Lutz, C. Sinner, G. Heuermann, A. Verma, and A. Schug. *eSBMTools 1.0: enhanced native structure-based modeling tools.* Bioinformatics 2013, 29(21), 2795–2796.

[104] J. K. Noel, A. Schug, A. Verma, W. Wenzel, A. E. Garcia, and J. N. Onuchic. *Mirror Images as Naturally Competing Conformations in Protein Folding.* The Journal of Physical Chemistry B 2012, 116(23), 6880–6888.

[105] J. K. Noel, P. C. Whitford, K. Y. Sanbonmatsu, and J. N. Onuchic. *SMOG@ctbp: simplified deployment of structure-based models in GROMACS.* Nucleic Acids Research 2010, 38(Web Server issue), W657–W661.

[106] H. Lammert, A. Schug, and J. N. Onuchic. *Robustness and generalization of structure-based models for protein folding and function.* Proteins: Structure, Function, and Bioinformatics 2009, 77(4), 881–891.

[107] J. K. Noel, P. C. Whitford, and J. N. Onuchic. *The Shadow Map: A General Contact Definition for Capturing the Dynamics of Biomolecular Folding and Function.* The Journal of Physical Chemistry B 2012, 116(29), 8692–8702.

[108] B. Lutz, M. Faber, A. Verma, S. Klumpp, and A. Schug. *Computational Analysis of Co-Transcriptional Riboswitch Folding.* Biophysical Journal 2014, 106(2), 284a.

[109] M. Hoefling, N. Lima, D. Haenni, C. A. M. Seidel, B. Schuler, and H. Grubmüller. *Structural Heterogeneity and Quantitative FRET Efficiency Distributions of Polyprolines through a Hybrid Atomistic Simulation and Monte Carlo Approach.* PLoS ONE 2011, 6(5), e19791.

[110] M. Hoefling and H. Grubmüller. *In silico FRET from simulated dye dynamics.* Computer Physics Communications 2013, 184(3), 841–852.

[111] T. Ansbacher, H. K. Srivastava, T. Stein, R. Baer, M. Merkx, and A. Shurki. *Calculation of transition dipole moment in fluorescent proteins—towards efficient energy transfer.* Physical Chemistry Chemical Physics 2012, 14(12), 4109.

[112] S. Benke. *The Cytolytic Pore Toxin ClyA Studied With Single-Molecule Spectroscopy.* PhD thesis. Universität Zürich, 2015.

[113] I. Reinartz, C. Sinner, D. Nettels, B. Stucki-Buchli, F. Stockmar, P. T. Panek, C. R. Jacob, G. U. Nienhaus, B. Schuler, and A. Schug. *Simulation of FRET dyes allows quantitative comparison against experimental data.* The Journal of Chemical Physics 2018, 148(12), 123321.

[114] URL: https://biotium.com.

[115] R. Ahlrichs et al. *Turbomole.* URL: http://www.turbomole.com.

[116] R. Ahlrichs, M. Bär, M. Häser, H. Horn, and C. Kölmel. *Electronic structure calculations on workstation computers: The program system turbomole.* Chemical Physics Letters 1989, 162, 165–169.

[117] A. D. Becke. *Density-functional exchange-energy approximation with correct asymptotic behavior.* Phys. Rev. A 1988, 38, 3098–3100.

[118] J. P. Perdew. *Density-functional approximation for the correlation energy of the inhomogeneous electron gas.* Phys. Rev. B 12 1986, 33, 8822–8824.

[119] F. Weigend and R. Ahlrichs. *Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy.* Phys. Chem. Chem. Phys. 2005, 7, 3297–3305.

[120] A. D. Becke. *A new mixing of Hartree–Fock and local densityfunctional theories.* The Journal of Chemical Physics 1993, 98(2), 1372–1377.

[121] A. Schäfer, H. Horn, and R. Ahlrichs. *Fully optimized contracted Gaussian basis sets for atoms Li to Kr.* The Journal of Chemical Physics 1992, 97(4), 2571–2577.

[122] M. Abraham, D. van der Spoel, E. Lindahl, B. Hess, and the GROMACS development team. *GROMACS User Manual version 5.0, p. 78–79,84*. URL: http://www.gromacs.org, 2014.

[123] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne. *The protein data bank*. Nucleic acids research 2000, 28(1), 235–242.

[124] A. L. Main, T. S. Harvey, M. Baron, J. Boyd, and I. D. Campbell. *The three-dimensional structure of the tenth type III module of fibronectin: An insight into RGD-mediated interactions*. Cell 1992, 71(4), 671–678.

[125] W. Kremer, B. Schuler, S. Harrieder, M. Geyer, W. Gronwald, C. Welker, R. Jaenicke, and H. R. Kalbitzer. *Solution NMR structure of the cold-shock protein from the hyperthermophilic bacterium Thermotoga maritima*. European Journal of Biochemistry 2001, 268(9), 2527–2539.

[126] A. J. Wallace, T. J. Stillman, A. Atkins, S. J. Jamieson, P. A. Bullough, J. Green, and P. J. Artymiuk. *E. coli hemolysin E (HlyE, ClyA, SheA): X-ray crystal structure of the toxin and observation of membrane pores by electron microscopy*. Cell 2000, 100(2), 265–76.

[127] S. Benke, D. Roderer, B. Wunderlich, D. Nettels, R. Glockshuber, and B. Schuler. *The assembly dynamics of the cytolytic pore toxin ClyA*. Nature Communications 2015, 6(1), 6198.

[128] M. Mueller, U. Grauschopf, T. Maier, R. Glockshuber, and N. Ban. *The structure of a cytolytic α-helical toxin pore reveals its assembly mechanism*. Nature May 2009, 459, 726–730.

[129] K. Arnold, L. Bordoli, J. Kopp, and T. Schwede. *The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling*. Bioinformatics 2006, 22(2), 195–201.

[130] B. Schuler, E. A. Lipman, and W. A. Eaton. *Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy*. Nature 2002, 419(6908), 743–7.

[131] V. Steffen, J. Otten, S. Engelmann, A. Radek, M. Limberg, B. Koenig, S. Noack, W. Wiechert, and M. Pohl. *A Toolbox of Genetically Encoded FRET-Based Biosensors for Rapid l-Lysine Analysis.* Sensors 2016, 16(10), 1604.

[132] H. Höfig, J. Otten, V. Steffen, M. Pohl, A. J. Boersma, and J. Fitter. *Genetically Encoded Förster Resonance Energy Transfer-Based Biosensors Studied on the Single-Molecule Level.* ACS Sensors 2018, 3(8), 1462–1470.

[133] A. Rekas, J.-R. Alattia, T. Nagai, A. Miyawaki, and M. Ikura. *Crystal Structure of Venus, a Yellow Fluorescent Protein with Improved Maturation and Reduced Environmental Sensitivity.* Journal of Biological Chemistry 2002, 277(52), 50573–50578.

[134] J. Goedhart, D. von Stetten, M. Noirclerc-Savoye, M. Lelimousin, L. Joosen, M. A. Hink, L. van Weeren, T. W. Gadella, and A. Royant. *Structure-guided evolution of cyan fluorescent proteins towards a quantum yield of 93%.* Nature Communications 2012, 3(1), 751.

[135] M. J. Borrok, L. L. Kiessling, and K. T. Forest. *Conformational changes of glucose/galactose-binding protein illuminated by open, unliganded, and ultra-high-resolution ligand-bound structures.* Protein Science 2007, 16(6), 1032–1041.

[136] Schrödinger, LLC. *The PyMOL Molecular Graphics System, Version 1.8.* 2015.

[137] E. M. W. M. van Dongen, T. H. Evers, L. M. Dekkers, E. W. Meijer, L. W. J. Klomp, and M. Merkx. *Variation of Linker Length in Ratiometric Fluorescent Sensor Proteins Allows Rational Tuning of Zn(II) Affinity in the Picomolar to Femtomolar Range.* Journal of the American Chemical Society 2007, 129(12), 3494–3495.

[138] G.-J. Kremers, J. Goedhart, E. B. van Munster, and T. W. J. Gadella. *Cyan and Yellow Super Fluorescent Proteins with Improved Brightness, Protein Folding, and FRET Förster Radius.* Biochemistry 2006, 45(21), 6570–6580.

[139] G. H. Patterson, D. W. Piston, and B. Barisas. *Förster Distances between Green Fluorescent Protein Pairs.* Analytical Biochemistry 2000, 284(2), 438–440.

[140] D. A. Zacharias, J. D. Violin, A. C. Newton, and R. Y. Tsien. *Partitioning of lipid-modified monomeric GFPs into membrane microdomains of live cells.* Science 2002, 296(5569), 913–916.

[141] A. W. Nguyen and P. S. Daugherty. *Evolutionary optimization of fluorescent proteins for intracellular FRET.* Nature Biotechnology 2005, 23(3), 355–360.

[142] J. L. Vinkenborg, T. H. Evers, S. W. A. Reulen, E. W. Meijer, and M. Merkx. *Enhanced Sensitivity of FRET-Based Protease Sensors by Redesign of the GFP Dimerization Interface.* ChemBioChem 2007, 8(10), 1119–1121.

[143] M. Merkx, M. V. Golynskiy, L. H. Lindenburg, and J. L. Vinkenborg. *Rational design of FRET sensor proteins based on mutually exclusive domain interactions.* Biochemical Society Transactions 2013, 41(5), 1201–1205.

[144] I. Kotera, T. Iwasaki, H. Imamura, H. Noji, and T. Nagai. *Reversible dimerization of Aequorea victoria fluorescent proteins increases the dynamic range of FRET-based indicators.* ACS chemical biology 2010, 5(2), 215–22.

[145] C. A. Jost, G. Reither, C. Hoffmann, and C. Schultz. *Contribution of Fluorophores to Protein Kinase C FRET Probe Performance.* ChemBioChem 2008, 9(9), 1379–1384.

[146] A. Hoffmann, A. Kane, D. Nettels, D. E. Hertzog, P. Baumgartel, J. Lengefeld, G. Reichardt, D. A. Horsley, R. Seckler, O. Bakajin, and B. Schuler. *Mapping protein collapse with single-molecule fluorescence and kinetic synchrotron radiation circular dichroism spectroscopy.* Proceedings of the National Academy of Sciences 2007, 104(1), 105–110.

[147] A. Soranno, B. Buchli, D. Nettels, R. R. Cheng, S. Müller-Späth, S. H. Pfeil, A. Hoffmann, E. A. Lipman, D. E. Makarov, and B. Schuler. *Quantifying internal friction in unfolded and intrinsically disordered proteins with single-molecule spectroscopy*. Proceedings of the National Academy of Sciences 2012, 109(44), 17800–17806.

[148] B. Schuler. *Application of Single Molecule Förster Resonance Energy Transfer to Protein Folding. Protein Folding Protocols*. Totowa, NJ: Humana Press, 2006, 115–138.

[149] J. A. Riback, M. A. Bowman, A. M. Zmyslowski, C. R. Knoverek, J. M. Jumper, J. R. Hinshaw, E. B. Kaye, K. F. Freed, P. L. Clark, and T. R. Sosnick. *Innovative scattering analysis shows that hydrophobic disordered proteins are expanded in water*. Science 2017, 358(6360), 238–241.

[150] J. A. Riback, M. A. Bowman, A. Zmyslowski, C. R. Knoverek, J. Jumper, E. B. Kaye, K. F. Freed, P. L. Clark, and T. R. Sosnick. *Response to Comment on "Innovative scattering analysis shows that hydrophobic disordered proteins are expanded in water"*. Science 2018, 361(6405), eaar7949.

[151] G. Fuertes, N. Banterle, K. M. Ruff, A. Chowdhury, D. Mercadante, C. Koehler, M. Kachala, G. Estrada Girona, S. Milles, A. Mishra, P. R. Onck, F. Gräter, S. Esteban-Martín, R. V. Pappu, D. I. Svergun, and E. A. Lemke. *Decoupling of size and shape fluctuations in heteropolymeric sequences reconciles discrepancies in SAXS vs. FRET measurements*. Proceedings of the National Academy of Sciences 2017, 114(31), E6342–E6351.

[152] G. Fuertes, N. Banterle, K. M. Ruff, A. Chowdhury, R. V. Pappu, D. I. Svergun, and E. A. Lemke. *Comment on "Innovative scattering analysis shows that hydrophobic disordered proteins are expanded in water"*. Science 2018, 361(6405), eaau8230.

[153] R. B. Best, W. Zheng, A. Borgia, K. Buholzer, M. B. Borgia, H. Hofmann, A. Soranno, D. Nettels, K. Gast, A. Grishaev, et al. *Comment on "Innovative scattering analysis shows that hydropho-*

bic disordered proteins are expanded in water". Science 2018, 361(6405), eaar7101.

[154] A. Borgia, W. Zheng, K. Buholzer, M. B. Borgia, A. Schüler, H. Hofmann, A. Soranno, D. Nettels, K. Gast, A. Grishaev, R. B. Best, and B. Schuler. *Consistent View of Polypeptide Chain Expansion in Chemical Denaturants from Multiple Experimental Methods.* Journal of the American Chemical Society 2016, 138(36), 11714–11726.

[155] W. Zheng, A. Borgia, K. Buholzer, A. Grishaev, B. Schuler, and R. B. Best. *Probing the Action of Chemical Denaturant on an Intrinsically Disordered Protein by Simulation and Experiment.* Journal of the American Chemical Society 2016, 138(36), 11702–11713.

[156] T.-o. Peulen, O. Opanasyuk, and C. A. M. Seidel. *Combining Graphical and Analytical Methods with Molecular Simulations To Analyze Time-Resolved FRET Measurements of Labeled Macromolecules Accurately.* The Journal of Physical Chemistry B 2017, 121(35), 8211–8241.

[157] G. Tria, H. D. T. Mertens, M. Kachala, and D. I. Svergun. *Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering.* IUCrJ 2015, 2(2), 207–217.

[158] A. Borgia, M. B. Borgia, K. Bugge, V. M. Kissling, P. O. Heidarsson, C. B. Fernandes, A. Sottini, A. Soranno, K. J. Buholzer, D. Nettels, B. B. Kragelund, R. B. Best, and B. Schuler. *Extreme disorder in an ultrahigh-affinity protein complex.* Nature 2018, 555(7694), 61–66.

[159] J. Jackson, K. Nguyen, and P. Whitford. *Exploring the Balance between Folding and Functional Dynamics in Proteins and RNA.* International Journal of Molecular Sciences 2015, 16(12), 6868–6889.

[160] E. Małolepsza, B. Strodel, M. Khalili, S. Trygubenko, S. N. Fejer, and D. J. Wales. *Symmetrization of the AMBER and CHARMM force fields.* Journal of Computational Chemistry 2010, 31, 1402–1409.

L
List of Publications

Articles

In preparation

- Oskar Taubert, **Ines Reinartz**, Henning Meyerhenke, and Alexander Schug
  diSTruct v1.0: A Solution to the Distance Geometry Problem for Biological Macromolecules

Submitted

- Jakob Rosenbauer, Benjamin Mattes, **Ines Reinartz**, Kyle Wedgwood, Simone Schindler, Claude Sinner, Steffen Scholpp, and Alexander Schug
  Modeling of Wnt-mediated Tissue Patterning in Vertebrate Embryogenesis

- Mathias Bockwoldt, Dorothée Houry, Marc Niere, Toni I. Gossmann, **Ines Reinartz**, Alexander Schug, Mathias Ziegler, and Ines Heiland
  Identification of Evolutionary and Kinetic Drivers of NAD-dependent Signalling

Published

- Marie Weiel, **Ines Reinartz**, and Alexander Schug
  Rapid Interpretation of Small-angle X-ray Scattering Data
  PLOS Computational Biology, 2019

- Arun A. Gupta, **Ines Reinartz**, Gogulan Karunanithy, Alessandro Spilotros, Venkateswara Rao Jonna, Anders Hofer, Dmitri I. Svergun, Andrew J. Baldwin, Alexander Schug, and Magnus Wolf-Watz
  Formation of a Secretion-Competent Protein Complex by a Dynamic Wraparound Binding Mechanism
  Journal of Molecular Biology, 2018

- **Ines Reinartz**, Claude Sinner, Daniel Nettels, Brigitte Stucki-Buchli, Florian Stockmar, Pawel T. Panek, Christoph R. Jacob, Gerd Ulrich Nienhaus, Benjamin Schuler, and Alexander Schug
  Simulation of FRET Dyes Allows Quantitative Comparison Against Experimental Data
  The Journal of Chemical Physics, 2018

- Anna-Lena Winkler, Joachim von Wulffen, Lisa Rödling, Annamarija Raic, **Ines Reinartz**, Alexander Schug, Robert Gralla, Udo Geckle, Alexander Welle, and Cornelia Lee-Thedieck
  Significance of Nanopatterned and Clustered DLL1 for Hematopoietic Stem Cell Proliferation
  Advanced Functional Materials, 2017

- Claude Sinner, Benjamin Lutz, Shalini John, **Ines Reinartz**, Abhinav Verma, and Alexander Schug
  Simulating Biomolecular Folding and Function by Native-Structure-Based/Go-Type Models
  Israel Journal of Chemistry, 2014

Conference contributions

Oral presentations

- 03/2018 “DPG Spring Meeting” in Berlin, Germany
  Simulation of FRET Dyes Allows Direct Comparison against Experimental Data

- 02/2017 “61st Annual Meeting of the Biophysical Society” in New Orleans, Louisiana, USA
  Simulation of FRET Dyes Allows Direct Comparison against Experimental Data

- 06/2015 “2015 Soft Matter Summer School: Polymers in Biology” in Seoul, Korea
  Simulation of FRET Dyes with Native Structure-based Models

Poster presentations

- 07/2016 “Gordon Research Conference: Single Molecule Approaches to Biology” in Hong Kong, Special Administrative Region of the People’s Republic of China
  Simulation of FRET Dyes with Native Structure-based Models

- 02/2016 “60th Annual Meeting of the Biophysical Society” in Los Angeles, California, USA
  3D Simulations of Morphogen Transport in an Early Fish Embryo

- 01/2016 “Gordon Research Conference: Protein Folding Dynamics” in Galveston, Texas, USA
  Simulation of FRET Dyes with Native Structure-based Models

- 06/2015 “2015 Soft Matter Summer School: Polymers in Biology” in Seoul, Korea
  Simulation of FRET Dyes with Native Structure-based Models

- 04/2015 Workshop “Computer Simulation and Theory of Macromolecules” in Huenfeld, Germany
  Simulation of FRET Dyes with Native Structure-based Models

- 04/2014 Workshop “Computer Simulation and Theory of Macromolecules” in Huenfeld, Germany
  Integration of FRET fluorophores into native structure based models

- 02/2014 “58th Annual Meeting of the Biophysical Society” in San Francisco, California, USA
  Integration of FRET fluorophores into native structure based models
- 02/2014 “552nd WE-Heraeus Seminar: Physics of Biomolecular Folding and Assembly: Theory meets Experiment” in Bad Honnef, Germany
  Integration of FRET fluorophores into native structure based models

A
Acknowledgments

Ich danke meinen beiden Referenten Herrn Prof. Wolfgang Wenzel und Herrn Prof. Gerd Ulrich Nienhaus für die Betreuung meiner Arbeit und die wertvollen Rückmeldungen.

Dr. Alexander Schug danke ich für die vielen Möglichkeiten die ich in seiner Arbeitsgruppe hatte, das hilfreiche Feedback und die Ermutigungen.

Meinen Kollegen, darunter besonders Claude Sinner, Oskar Taubert und Marie Weiel-Potyagaylo, danke ich für die schöne Zeit im Büro und die angenehme Arbeitsatmosphäre. Ein Dankeschön gilt auch den restlichen derzeitigen und ehemaligen Mitgliedern der AG Wenzel für die vielen Kuchen und das gelegentliche Feierabendbier/-kaffee/-tee.

Außerdem danke ich den Doktoranden und Postdocs des SCC für die schöne Zeit an diesem Institut.

Weiterhin vielen Dank an die Kaffeegruppe und die Meditationsgruppe des SCC, die mir dabei geholfen haben wach zu werden und meine Nerven zu behalten.

Danke an die Mitarbeiter des SCC, insbesondere diejenigen die sich um ein reibungsloses Funktionieren der Cluster gekümmert haben.

Weiterhin möchte ich herzlich der Gruppe von Prof. Ben Schuler danken, die mir bei meinem Aufenthalt ermöglicht haben viel über ihre experimentelle Arbeit zu lernen, darunter besonders Daniel Nettels und Stephan Benke. Sebastian König und Erica Fiorini danke ich, ohne die ich wohl auch nie diese schöne Zeit in Zürich verbracht hätte.

Für die Zusammenarbeit bei der Bestimmung der Farbstoffstrukturen möchte ich der Gruppe von Prof. Christoph Jacob danken, sowie Pascal Friederich.

Für die Zusammenarbeit an dem Projekt des Glukosesensors möchte ich allen voran Henning Höfig, sowie Prof. Jörg Fitter, Mona Sarter, Dr. Andreas Stadler, Julia Otten und Prof. Martina Pohl herzlich danken.

Vielen Dank auch an alle die so viel Zeit in das Korrekturlesen dieser Arbeit investiert haben, insbesondere Marie Weiel-Potyagaylo, Markus Götz, Oskar Taubert, Pascal Friederich und Claude Sinner.

Außerdem danke ich meinen Freunden für ihre Unterstützung in dieser Zeit und der 7G-Runde für wunderschöne Samstage. Ich danke auch meinem Freund Nico für seine Geduld, Ruhe und die Ermutigungen vor allem in der Endphase dieser Arbeit.

Ein großer Dank geht an meine Eltern, die mich ermutigt und unterstützt haben und immer ein offenes Ohr hatten. Außerdem Danke an meine Familie, meine Großeltern und insbesondere meinen Großvater, der für mich mit seinem Eifer und seiner Begeisterung ein Vorbild für das Schreiben dieser Arbeit war.