
High-Resolution Single-Molecule Spectroscopy to Probe Conformational Dynamics of Proteins

Christian Gebhardt



München 2022

High-Resolution Single-Molecule Spectroscopy to Probe Conformational Dynamics of Proteins

Christian Gebhardt

Dissertation
an der Fakultät für Physik
der Ludwig-Maximilians-Universität
München

vorgelegt von
Christian Gebhardt
aus Heidenheim an der Brenz

München, den 1.3.2022

Erstgutachter: Prof. Dr. Jan Lipfert
Zweitgutachter: Prof. Dr. Thorben Cordes
Tag der mündlichen Prüfung: 5.4.2022

“Remember that all models are wrong; the practical question is how wrong do they have to be to not be useful.” George E.P. Box

Zusammenfassung

Fluoreszenzspektroskopie und -mikroskopie sind unverzichtbare Methoden in zahlreichen Forschungsbereichen und der modernen Medizin, einschließlich Molekularbiologie, Biophysik und Biochemie. Die Anwendungen reichen von der Mikroskopie und Spektroskopie mit Einzelmoleküldetektion bis zur Sensorik und der Entwicklung neuartiger fluoreszenter Proben.

Diese Arbeit trägt zur grundlegenden Untersuchung und spektroskopischen Charakterisierung etablierter sowie neu verfügbarer fluoreszierender Proben für die Einzelmolekülmikroskopie bei. Darüber hinaus ermöglicht die Entwicklung von quantitativen Einzelmolekül-Fluoreszenzmessungen ein systematisches Versuchsdesign und standardisierte Messroutinen für quantitative Messungen der Proteinkonformation und -dynamik. In dieser Arbeit konzentriere ich mich auf das Design und die Analyse von lösungsbasierten Einzelmolekülexperimenten des Förster-Resonanz-Energie-Transfers (FRET) mittels konfokaler Mikroskopie, um die Dynamik von mehreren Bindeproteinen als exemplarische biologische Systeme zu untersuchen. Als Grundlage für qualitativ hochwertige Fluoreszenzmessungen wird die weit verbreitete Klasse der Cyanin-Fluorophore mit optischer Spektroskopie, Massenspektrometrie und NMR-Spektroskopie untersucht. Ich habe einen umfassenden Vergleich zwischen Alexa-Fluor- und AF-Farbstoffen mittels kontinuierlicher Absorptions- und Emissionsspektroskopie, Bestimmung der Quantenausbeute, der Fluoreszenzlebensdauer und Anisotropie-Spektroskopie von freien und an Proteine gebundenen Farbstoffen durchgeführt, um die Auswirkungen der Position von negativ geladenen sulfonierten Gruppen auf das gesamte photophysikalische Verhalten und deren Auswirkungen auf Einzelmolekül-FRET-Messungen (smFRET) zu verstehen.

Um photophysikalische Einflüsse von biologischen Vorgängen zu trennen, wird die gepulste Elektronen-Elektronen-Doppelresonanzspektroskopie als komplementäre Technik zu FRET für die Untersuchung der Proteinkonformation durch die Anbringung von - in diesem Fall - Spin-Markern eingesetzt. Ein Vergleich der derzeitigen Limitierungen und Vorzüge beider Methoden wird für mehrere Proteinsysteme durchgeführt, die Konformationsänderungen auf unterschiedlichen Zeitskalen aufweisen. Ich habe gemessene Abstände mit den Abstandsvorhersagen aus grobkörnigen Strukturmodellen verglichen und Bereiche für weitere Verbesserungen in Bezug auf die Auswahl der Markierungsstellen und Fluorophore identifiziert.

Um die Zuverlässigkeit von mit FRET bestimmten Abständen und Abstandsunsicherheiten bei Proteinen weiter zu erhöhen, habe ich eine Blindstudie mit 19 Laboren durchgeführt, die die Fähigkeit von smFRET bestätigt, die Konformationsdynamik auf verschiedenen Zeitskalen von Submillisekunden bis zu Sekunden aufzudecken und die Grenzen der Abstandsgenauigkeit bei stochastischer Markierung von Proteinen zu quantifizieren. Darüber hinaus gebe ich einen detaillierten Überblick über die etablierten

Routinen bei smFRET-Messungen für Abstands- und Dynamikmessungen und bestimme die Grenzen der von FRET abgeleiteten Abstände auf eine Abstandsgenauigkeit von ≤ 2 Å und eine Genauigkeit ≤ 5 Å.

Ich habe die Auswahl der Fluorophore und der Markierungsstellen auf dem zu untersuchenden Protein als limitierenden Faktor für die Qualität und Aussagekraft der Ergebnisse in den experimentellen Studien identifiziert. Aus diesem Grund habe ich einen theoretischen Ansatz für das Studiendesign unter Verwendung der grundlegenden Wahrscheinlichkeitstheorie entwickelt. Dazu habe ich die Ergebnisse von mehr als 100 Veröffentlichungen ausgewertet, um die wichtigsten Parameter für erfolgreiche Einzelmolekül-Fluoreszenzexperimente zu ermitteln und ein Punktesystem für die Auswahl einzuführen. Ich habe die Analyse als eigenständiges Softwarepaket und als frei zugänglichen Webserver implementiert, um allen Forschern auf diesem Gebiet Zugang zu dem Analyseverfahren zu ermöglichen. Die Ergebnisse werden durch eine experimentelle Validierung an einigen beispielhaften Proteinsystemen untermauert.

Abstract

Fluorescence spectroscopy and microscopy are indispensable tools in numerous research fields and modern medicine, including molecular biology, biophysics, and biochemistry. Applications stretch from microscopy and spectroscopy with single molecule detection to sensor technology and the development of novel fluorescent probes.

This work contribute to a fundamental investigation and spectroscopic characterization of established as well as newly available fluorescent probes for single-molecule microscopy. Furthermore, the development of quantitative single-molecule fluorescence measurements enables a systematic assay design and standardized measurement routines for quantitative measurements of protein conformation and dynamics. In this thesis, I focus on the design and analysis of solution-based single-molecule experiments of Förster resonance energy transfer (FRET) via confocal microscopy to study the dynamics of multiple binding proteins as exemplary biological systems.

As the basis for high-quality fluorescence measurements, the widely used class of cyanine fluorophores is investigated with optical spectroscopy, mass spectrometry, and NMR spectroscopy. I conducted a comprehensive comparison between Alexa Fluor and AF dyes by continuous-wave absorption and emission spectroscopy, determination of quantum yield, fluorescence lifetime and anisotropy spectroscopy of free and protein-attached dyes to understand the impact of the location of negatively charged sulfonated groups on the overall photophysical behavior and its implications for single-molecule FRET (smFRET) measurements.

In order to unravel photophysical effects from biological findings, pulsed electron-electron double resonance spectroscopy is used as a complementary technique to FRET for the investigation of protein conformation by the attachment of - in this case - spin labels. A comparison of the current limitations and advantages of both methods is conducted for multiple protein systems, which undergo conformational changes on different timescales. I compared measured distances to distance predictions from coarse-grained structural models and identified areas for further improvements in terms of label site and fluorophore selection.

To further increase the reliability of FRET-derived distances and distance uncertainties in proteins, I conducted a blind study comprising 19 labs, which confirms the ability of smFRET to discover conformational dynamics on different timescales from sub-milliseconds to seconds and quantify the limits of distance accuracies for stochastic labeling of proteins. Further, I provide a detailed overview of the established routines in smFRET measurements for distance and dynamics measurements and determine the limitations of FRET-derived distances to a distance precision of ≤ 2 Å and an accuracy ≤ 5 Å.

I identified the selection of fluorophores and labeling sites on the protein under investigation as limiting factor for the quality and significance of the findings in the experimental studies. Therefore, I established a theoretical framework for the assay design using basic probability theory. To do so, I exploited the results of more than 100 publications to identify the most relevant parameters for successful single-molecule fluorescence experiments and to introduce a scoring system for the selection. I implemented the analysis as stand-alone software package and open-access webserver to provide access to the analysis framework to all researchers in the field. The results are backed up with an experimental validation on a few exemplary protein systems.

Contents

1	Introduction	1
2	Fundamentals of fluorescence - concepts and applications	5
2.1	Fluorescence	5
2.1.1	General concepts	5
2.1.2	Fluorescence lifetime	7
2.1.3	Fluorescent quantum yield	7
2.1.4	Fluorescence anisotropy	8
2.1.5	Organic fluorophores	10
2.2	Förster resonance energy transfer	12
2.2.1	Basic theory	12
2.2.2	Lifetime-based FRET efficiency determination	14
2.2.3	Intensity-based FRET efficiency determination	14
2.3	Confocal microscopy	15
2.3.1	Basic principles	15
2.3.2	Single-molecule burst detection and photon statistics	17
2.3.3	Fluorescence correlation spectroscopy	19
3	Measurement and evaluation techniques	21
3.1	Protein labeling	21
3.1.1	Labeling protocol	21
3.1.2	Size exclusion chromatography	22
3.1.3	labeling efficiency calculation	23
3.2	Fluorescence quantum efficiency determination	24
3.2.1	Spectrum normalization	25
3.3	Lifetime and anisotropy measurements	27
3.3.1	Setup and measurement	27
3.3.2	IRF measurements	29
3.3.3	Lifetime reconvolution fitting	30
3.3.4	Fundamental anisotropy	33
3.3.5	Anisotropy reconvolution fitting	33
3.3.6	Protein relaxation	34
3.4	ALEX and PIE measurements	36
3.4.1	Setup and measurement	36
3.4.2	Binding kinetics	39
3.4.3	Accurate FRET determination	40
3.4.4	Dynamics evaluations	43

4	Molecular and spectroscopic characterization of green and red cyanine fluorophores from the Alexa Fluor and AF series	47
4.1	Introduction	48
4.2	Material and Methods	50
4.3	Results	56
4.4	Discussion	70
4.5	Acknowledgements and Author Contributions	71
4.6	Supplementary Information	73
5	Cross-validation of distance measurements in proteins by PELDOR/DEER and single-molecule FRET	91
5.1	Introduction	91
5.2	Results	94
5.3	Discussion	111
5.4	Acknowledgements and Author Contributions	117
5.5	Materials and Methods	117
5.6	Supporting Information	120
6	Distance and dynamics measurements with FRET	137
6.1	Introduction	137
6.1.1	Introduction of the model protein systems	139
6.2	Results	143
6.2.1	Precision and reproducibility of smFRET values on proteins	143
6.2.2	Discovery and quantification of conformational dynamics in proteins via smFRET	145
6.3	Discussion	150
6.4	Acknowledgements and Author Contributions	153
6.5	Methods	154
6.5.1	Protein labeling	154
6.5.2	Sample handling	155
6.5.3	Data analysis	156
6.6	Supplementary Figures and Tables	157
6.7	Supplementary Notes	175
6.7.1	Primer design and sequences	175
6.7.2	Dynamics of U2AF65	176
6.7.3	Evaluation procedure for U2AF65	177
6.7.4	FRET with different fluorophores	178
6.7.5	Error simulation	178
6.7.6	Förster radius determination	180
7	Labelizer: systematic selection of labeling sites for fluorescent dyes on proteins	183
7.1	Introduction	183
7.2	Results	186

7.3	Discussion and conclusion	198
7.4	Acknowledgements and Author Contributions	199
7.5	Materials and Methods	200
7.5.1	Database generation and evaluation	200
7.5.2	Protein production and labeling	203
7.5.3	smFRET spectroscopy	203
7.6	Supplementary Notes	203
7.6.1	Database parameter evaluation	203
7.6.2	Förster resonance energy transfer	206
7.7	Supplementary Information	208
8	Conclusion	219
	Bibliography	225
	Publications and conference contributions	257

1 | Introduction

The absorption and emission of light has already been observed and studied already for centuries. The term fluorescence dates back to George Gabriel Stokes (1819-1903) in his work *On the Change of Refrangibility of Light* [1], where Stokes describes the absorption and emission of light. Although classical Maxwell's equations constitute the fundamental description of light propagation and light-matter interaction, we only obtain a full picture using the quantum-mechanical framework developed in the early 20th century, which provides the basis to explain the absorption and emission of light in discrete energy packages associated with electronic state transitions.

Advances in the understanding of energy levels and band structures led to the discovery of a variety of light emitters of different classes, e. g. organic fluorophores, (modified) fluorescent proteins, and quantum dots, where the spectroscopic properties such as emission wavelength, spectral width, and quantum yields can be - to some extent - tailored for the respective requirements [2]. This evolution paved the way to multiple applications on fluorescence spectroscopy, sensing, imaging and microscopy in the biosciences [2] or the development of fluorescence dye laser [3, 4].

Fluorescence microscopy and spectroscopy. The development of advanced microscopy techniques and the ability of single-molecule detection [5, 6] in combination with tailored fluorophores (e.g. blinking, photo-activation) [6–8] have made the techniques an indispensable tools in numerous research fields ranging from molecular biology to biophysics and biochemistry. The development of new microscopy concepts was honored with the Nobel prize in chemistry 2014 awarded to Eric Betzig, Stefan Hell, and William E. Moerner “for the development of super-resolved fluorescence microscopy” such as stimulated emission depletion microscopy (STED), photo-activated localization microscopy (PALM), or stochastic optical reconstruction microscopy (STORM) [9]. Beside the pure intensity based measurements, further spectroscopic methods based on the readout of wavelength or lifetime provide additional or complementary insights to these imaging techniques.

Single-molecule Förster resonance energy transfer. The general near-field dipole-dipole coupling also applies to fluorescent emitters, which results in a non-radiative energy transfer from one dipole to the other dipole. The interaction of standard fluorophores via dipole-dipole coupling was fully described by Theodor Förster in 1948 [10]. Later, this so-called Förster resonance energy transfer (FRET) was exploited to measure distances between two adjacent fluorophores in the range of a few nanometers (2-10 nm), and therefore, is referred to as ‘molecular ruler’ [11]. Especially in bioscience, the

techniques of microscopy and fluorescence spectroscopy on a single-molecular level play a crucial role in monitoring biological processes, observing interactions of biomolecules and detecting changes of conformational states [12]. After the first realization of single-molecule FRET measurements [13], several techniques were developed to reduce the noise floor and remove the signal of undesired background, such as alternating-laser excitation microscopy (ALEX) [14, 15] or pulsed-interleaved excitation microscopy (PIE) [16] and to gain further information, such as lifetime or anisotropy, with multi-fluorescent detection (MFD) [17, 18].

Biomolecules in single-molecule measurements. Structural information on proteins from the continuously increasing protein database [19, 20] and the possibility of site-specific labeling of proteins with organic fluorophores [21, 22] paved the way for smFRET measurements as a standard tool to obtain structural, conformational, or kinetic information on the biological system under investigation *in vitro* or *in vivo* [12]. During the period of two decades, developments in single-molecule fluorescence spectroscopy enable the study of dynamics of macromolecular structures and biological processes on a molecular level under biologically relevant conditions [12], which was enabled by advancements in three different fields: First, technical improvement were made in the field of pulsed laser sources, optical components such as objectives, single-photon detectors and time-synchronization and real-time-recording electronics. A second point is the advancement in chemistry and biochemistry that provided fluorescence dyes with improved and tailored properties as well as novel methods of attaching them to the molecule under investigation in an highly specific and efficient manner. Lastly, improvements in analysis methods and the development of new algorithms reveal deeper insights into the investigated fundamental biological processes and results in a deeper understanding of their nature.

This work aims to contribute further to the last two aspects by the fundamental investigation and spectroscopic characterization of established as well as newly available fluorescent probes and the development of quantitative single-molecule fluorescence measurements allowing a systematic and standardized assays design and standardized measurements for quantitative findings. The work focuses on the design and analysis of solution-based single-molecule measurements of Förster resonance energy transfer (FRET) on confocal microscopes. One challenge of the fluorescent techniques in combination with advanced statistical methods is to extract the maximum amount of information from noisy data of single molecules to be able to answer relevant biological questions on the level of individual molecules.

Although the methods are routinely used to monitor the conformation and interactions of biomolecules through spectroscopic readout, environmental effects and intrinsic measurement limitations restrict the informative value. Several fluorescence parameters, such as the spectrum, the quantum efficiency, lifetime and polarization, are highly sensitive to the environment of the fluorophore [2, 23]. A better understanding of the protein-fluorophore interaction and a careful selection of labeling sites can minimize these effects, and thus, allow quantitate measurements on protein conformations and conformational

transitions. Therefore, it is crucial to investigate and benchmark smFRET measurements under varying conditions of the fluorophore environment in different protein systems to identify and improve the limits of the techniques. A comparison to complementary methods is helpful to disentangle the complex interplay of biological effects and method-intrinsic artifacts.

Outline of this thesis In chapter 2, I will concentrate on the fundamentals of fluorescence and some basics on the commonly used organic fluorophores. Further, a theoretical description of their near field dipole-dipole interaction will be introduced. The principles of fluorescent confocal microscopy and its usage for single-molecule spectroscopy measurements will be discussed next. In the end, environmental effects on the fluorophore properties shall be addressed in the context of fluorophore-protein-interaction.

Chapter 3 deals with technical considerations on the sample preparation and high resolution measurements of fluorescent quantum yield, lifetime, and anisotropy. I will further give an in-depth introduction into the details on smFRET measurements on continuous-wave and time-resolved confocal microscopes. Furthermore, the standard evaluation and correction routines will be discussed.

In chapter 4, I will demonstrate an in-depth characterization of common fluorophores from the cyanine-class for the use in smFRET-measurements.

Chapter 5 presents a comparison of the two complimentary methods - smFRET and electron paramagnetic resonance (EPR) - for distance measurements in protein samples and their ability to resolve dynamics.

A thorough study on the accuracy and precision of distance determination with smFRET is presented in chapter 6, where a multi-lab benchmark study within 19 different research groups investigates the potentials and limits of smFRET.

In chapter 7, a global analysis of suitable labeling sites was performed based on a meta-study of more than 100 publications in the field of single-molecule spectroscopy to develop a theoretical framework for the prediction of suitable labeling sites. The findings were experimentally validated and resulted in an open-access analysis software for the planning of single-molecule fluorescence and FRET experiments building upon the results of two decades of smFRET measurements.

The last chapter summarises the work and gives an outlook.

2 | Fundamentals of fluorescence - concepts and applications

In this thesis, organic fluorophores are used as nanoscale probes to determine inter-residue distances and conformational dynamics in bio-molecules. The biophysical assays used rely on a proper physical understanding of fluorophores and their properties. In this chapter, I will provide an overview of the basic concepts of fluorescence, single-molecule microscopy, and spectroscopy applications. Fluorescence occurs in atoms, molecules, and in the solid state, yet this chapter focuses on the fundamentals of fluorescence in organic fluorophores with large π -orbitals. Furthermore, I will describe the theoretical background of Förster resonance energy transfer (FRET) between two fluorophores, which represent the basis for many assays used in my thesis. Although the basics of FRET is well understood and has a sound theoretical standing, the energy transfer is always affected by other phenomena with rate-influencing effects such as quenching or steric interference, which will also be discussed. Further, I shall introduce the basic concepts of confocal microscopy as the main measurement technique in this thesis. Other groundbreaking advances, especially in the field of single-molecule fluorescence and super-resolution microscopy shall not be discussed here, but can be found in ‘Far-Field Optical Nanoscopy’ (2015, Tinnefeld, Eggeling, and Hell) [6].

2.1 Fluorescence

Fluorescence is a type of quantum mechanical process that describes the radiative de-excitation of an electronically excited state and is also called photoluminescence. It occurs during the relaxation of an atom or molecule from an excited (higher energy) state to its (lower energy) ground state accompanied by the emission of a photon. Depending on the duration of the emission process (lifetime of the excited state), one distinguishes between fluorescence ($\sim 10^{-9}$ - 10^{-6} s) and phosphorescence ($\sim 10^{-3}$ - 1000 s) [2]. The following section introduces the main concepts based on the textbooks from Sauer, Hofkens, and Enderlein ([24], chapter 1), Lakowicz ([2], chapter 1/3/10/11), and Sakurai ([25], chapter 5), where more details can be found.

2.1.1 General concepts

Fluorescence is the spontaneous deexcitation of an excited electronic state under the emission of a photon, which occurs in the transition between atomic energy levels or molecular orbitals. Fluorophores with visible absorption and emission comprise typically conjugated π -orbitals (e.g. a chain of double bonds such as polymethine, or conju-

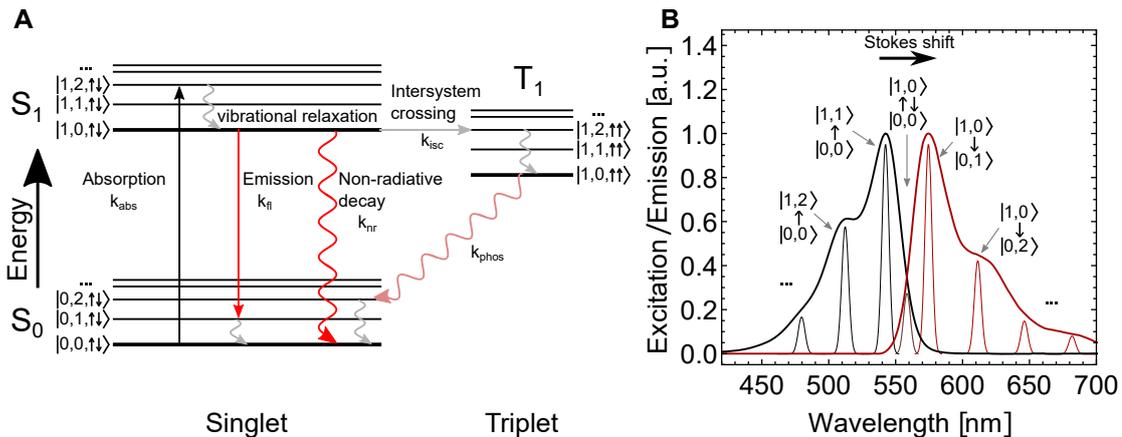


Figure 2.1.1: Fluorescence principle. **A** Jablonski diagram of a fluorophore with energy levels in the ground state S_0 , the excited state S_1 , and the triplet state T_1 . The excitation and relaxation pathways are shown as curly arrows. **B** Absorbance and emission spectrum of Cy3 in aqueous solution. The underlying discrete state transitions are broadened in standard experiments and only become visible at low temperature [24].

gated aromatic rings), which exhibit a molecular energy gap on the order of 1.5 to 3 eV (visible range) between the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO). The electronic wavefunctions of the electron pair occupying the HOMO in this delocalized system are well described with three quantum numbers, which define the energy levels in the system: a principle quantum number (ground state, first excited state), a vibrational quantum number, and the spin quantum number (singlet, triplet). Considering the energy levels of a molecule as illustrated in the Jablonski diagram in Figure 2.1.1A, we observe absorbance and emission of light at several wavelengths (or energies) with allowed transitions from an initial state a to a final state b . Due to the very fast vibrational relaxation on the timescale of 10^{-15} s, the excitation occurs mainly from the vibrational ground state ($v=0$) in the HOMO-level S_0 to excited vibrational states in the LUMO-level S_1 and vice-versa for the emission of light from S_1 to S_0 (Kasha's rule) [25]. This effect results in an energy shift between absorbance and emission (Stokes shift) as seen in Figure 2.1.1B.

In general, any transition between two quantum mechanical states a and b in a molecule is described by the transition dipole moment

$$\boldsymbol{\mu}_{ab} = \langle \psi_b | e\mathbf{r} | \psi_a \rangle, \quad (2.1.1)$$

where ψ_i are the wave-functions of state i , e is the elementary charge, and \mathbf{r} the position operator [25]. Fluorophores in the visible spectrum consist of tens of atoms giving rise to manifold vibrational modes and their overtones and therefore multiple transitions between the HOMO and LUMO energy level are possible. At room temperature, these discrete states are broadened and the spectrum is completely blurred due to the interaction and collisions with solvent molecules (see Figure 2.1.1B) [24]. The relative contribution

of the transitions between different vibrational states depends on the overlap of their wavefunctions, assuming a separation of the electronic and vibrational wavefunctions according to the Born–Oppenheimer approximation [25]. The typical shape of the spectrum with an unlikely transition from $v=0$ for S_0 to $v=0$ in S_1 follows from the so-called Franck-Condon principle [24].

2.1.2 Fluorescence lifetime

The fluorescence lifetime τ_{fl} is defined as the average time of a fluorophore remaining in the excited state before deexcitation and reads as

$$\tau_{fl} = \langle I(t) \rangle_t = \frac{\int_0^\infty t I(t) dt}{\int_0^\infty I(t) dt} \quad (2.1.2)$$

with the fluorescence intensity decay $I(t)$. For the simple mono-exponential case with $I(t) = I_0 e^{-t/\tau}$, we get the straightforward relation $\tau_{fl} = \tau$. For multi-exponential decays of the form

$$I(t) = \sum_i \alpha_i e^{-t/\tau_i}, \quad (2.1.3)$$

with multiple amplitudes α_i and corresponding lifetimes τ_i , equation 2.1.2 becomes the amplitude-weighted average

$$\tau_{fl} = \frac{\sum_i \alpha_i \tau_i^2}{\sum_i \alpha_i \tau_i}. \quad (2.1.4)$$

The amplitude-averaged lifetime is an alternative figure of merit reading as

$$\tau = \frac{\sum_i \alpha_i \tau_i}{\sum_i \alpha_i}, \quad (2.1.5)$$

which corresponds to a mono-exponential lifetime that results in the same time-integrated intensity as a multi-exponential decay.

2.1.3 Fluorescent quantum yield

The fluorescent quantum yield is the ratio of emitted and absorbed photons of an emitter and is given by

$$\Phi = \frac{k_{rad}}{k_{rad} + \sum_i k_{nr,i}} \quad (2.1.6)$$

Here, k_{rad} denotes the radiative fluorescent decay rate and $k_{nr,i}$ denotes any non-radiative decay paths (internal conversion, inter-system-crossing, collisional quenching, etc.). The denominator is the total rate of all pathways depopulating the excited state S_1 and relates to the fluorescent lifetime as

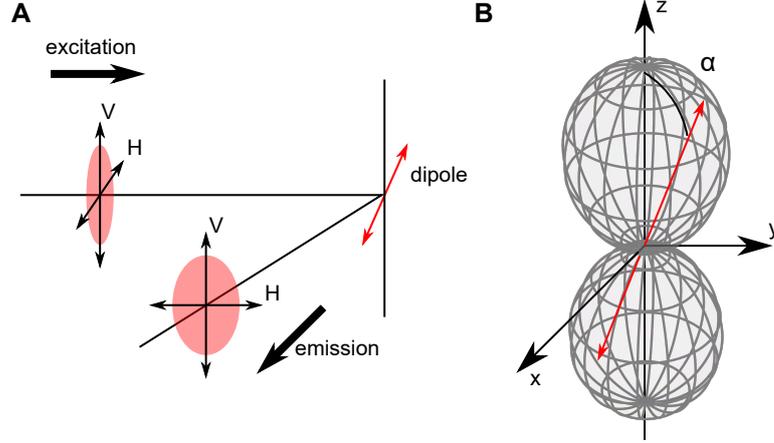


Figure 2.1.2: Fluorescence anisotropy. **A** Scheme of anisotropy of a dipole (red). **B** The surface (grey) represents the angle dependent photon selection of a dipole (red) depending on its the spatial angle with respect to the excitation polarization.

$$\tau_{fl} = (k_{rad} + \sum_i k_{nr,i})^{-1} \quad (2.1.7)$$

for ideal emitters with mono-exponential decays. For the case of vanishing non-radiative decay channels ($k_{nr,i} \rightarrow 0$), the quantum yield becomes unity and the lifetime is given by the so-called natural-lifetime

$$\tau_n = 1/k_{rad}. \quad (2.1.8)$$

This is an intrinsic upper bound of the lifetime, which can be calculated based on basic quantum mechanical principles according to Fermi's golden rule as a direct consequence of Heisenberg's uncertainty principle [25].

2.1.4 Fluorescence anisotropy

The anisotropy is defined as

$$r(t) = \frac{I_{\parallel}(t) - I_{\perp}(t)}{I_{\parallel}(t) + 2I_{\perp}(t)} \quad (2.1.9)$$

based on vertically $I_{\parallel}(t)$ and horizontally $I_{\perp}(t)$ polarized emission intensity from a vertically excited sample (see Figure 2.1.2A). It is directly related to the polarization of the emission light $P(t) = (I_{\parallel}(t) - I_{\perp}(t))/(I_{\parallel}(t) + I_{\perp}(t))$ as $r(t) = 2P(t)/(3 - P(t))$. The photoselection for the emission of a photon in a particular polarization axis shows an angular dependence proportional to $\cos^2(\alpha)$, where α is the angle between the emission dipole moment and the polarization of the emitted light (see Figure 2.1.2B). For a single emitter, the expressions for the parallel and perpendicular intensities may be

expressed as a function of the dipole orientation α with respect to the parallel axis as $I_{\parallel}(t) = \cos^2(\alpha(t))$ and $I_{\perp}(t) = \frac{1}{2}\sin^2(\alpha(t))$. One finds that

$$r(t) = \frac{3 \langle \cos^2(\alpha(t)) \rangle - 1}{2}, \quad (2.1.10)$$

where all excited molecular orientations are averaged. For the static isotropic case (fixed, isotropically orientated dipoles), the average is given by the spherical integral

$$\langle \cos^2(\alpha) \rangle = \frac{\int_0^{\pi/2} \cos^2(\alpha) \cos^2(\alpha) \sin(\alpha) d\alpha}{\int_0^{\pi/2} \cos^2(\alpha) \sin(\alpha) d\alpha} = \frac{3}{5} \quad (2.1.11)$$

and equation 2.1.10 results in a maximal anisotropy of $r_0 = 2/5$. It needs to be mentioned that we intrinsically assumed parallel orientation of the absorption and emission dipole moment in the previous consideration. However, the transition dipole moments of transitions between different vibrational states (see figure 2.1.1B) might be differently oriented, such that there is an average angular displacement β of the dipoles. Considering this effect in equation 2.1.10 the fundamental anisotropy reads as

$$r_0 = \frac{2}{5} \left(\frac{3 \langle \cos^2(\beta) \rangle - 1}{2} \right). \quad (2.1.12)$$

The steady-state anisotropy is the intensity-weighted average anisotropy and is given by

$$r = \frac{\int_0^{\infty} I(t) r(t) dt}{\int_0^{\infty} I(t) dt}. \quad (2.1.13)$$

This is the anisotropy, which is observed for measurements with continuous excitation.

Spherical rotational relaxation For spherical objects the depolarization, and thus the time-resolved anisotropy decay can be derived from first principles on diffusion using the Perrin equation [26, 27]

$$r(t) = r_0 e^{-6Dt} = r_0 e^{-t/\theta}, \quad (2.1.14)$$

where D is the rotational diffusion coefficient and $\theta = (6D)^{-1}$ is the rotational correlation time

$$\theta = \frac{\eta V}{RT} \quad (2.1.15)$$

with the viscosity η , the sphere volume V , the gas constant R , and the temperature T .

For the simple mono-exponential depolarization $r(t) = r_0 e^{-t/\theta}$ and mono-exponential intensity $I(t) = I_0 e^{-t/\tau}$, equation 2.1.13 reduces to

$$r = \frac{r_0}{1 + \tau/\theta}. \quad (2.1.16)$$

Non-spherical rotational relaxation Most fluorescent emitters are far from spherical and show more complex rotational diffusion. Thus, the anisotropy relaxation exhibits multi-exponential behavior. For prolate or oblate ellipsoids (two axes are equivalent), the anisotropy decay reads

$$r(t) = r_1 e^{-t/\theta_1} + r_2 e^{-t/\theta_2} + r_3 e^{-t/\theta_3}, \quad (2.1.17)$$

with $r_0 = r_1 + r_2 + r_3$ and the characteristic rotational correlation times $\theta_1 = (D_{\parallel} + 5D_{\perp})^{-1}$, $\theta_3 = (4D_{\parallel} + 2D_{\perp})^{-1}$, and $\theta_2 = (6D_{\perp})^{-1}$ as the inverse of linear combinations of the diffusion constant along the main axis D_{\parallel} and the orthogonal axis D_{\perp} [28]. The values of the amplitudes r_1 , r_2 , and r_3 depend on the actual location and orientation of the absorbance and emission dipole moment with respect to the ellipsoid (reference [2], chapter 12).

2.1.5 Organic fluorophores

Synthetic organic fluorophores are the common choice for single-molecule applications due to their high absorbance cross section ($\epsilon > 10^4 M^{-1} cm^{-1}$) and large fluorescence quantum yields ($\Phi > 0.1$) leading to a bright fluorescence signal. This thesis focuses mainly on fluorophores from the classes of rhodamines (Rhodamine 6G, Alexa Fluor 488, Alexa Fluor 546, TMR) and cyanines (Cy3, Alexa Fluor 555, Cy5, Alexa Fluor 647) as illustrated in Figure 2.1.3A/B. Oftentimes, the core structures are modified by attaching charged side groups (e.g. SO_3^-) in order to improve water solubility (see chapter 4). Further common fluorophore classes used in single-molecule studies are oxazines, carbopyronines, carborhodamines, perylenes, or BODIPYs [2, 29].

Besides excited state deactivation via radiative (fluorescence) and non-radiative pathways (internal conversion), typical organic fluorophores undergo spin-forbidden transitions called intersystem-crossing to a long-lived triplet state with lifetimes on the order of micro- to milliseconds (see Figure 2.1.1A). This long-lived state is chemically reactive due to the two unpaired electrons and the excess energy, which makes it susceptible to electron transfer reactions (redox reactions) yielding radicals [31, 32]. As a consequence, fluorophores can photobleach due to chemical modifications that often originates from triplet-state formation in the first step [33]. Removal of oxygen (which causes photooxidation or generation of singlet oxygen [33]) and the addition of triplet-state quenchers that enhance the photon count rates and suppress photobleaching in applications where high signal or long measurement times are needed (e.g. immobilized molecules) [31, 34, 35]. However, this thesis focuses on solution-based measurements with short observation times and does not request photostabilizer additives.

The right choice of fluorophores for the different application depends on several aspects as described in later chapters of this thesis (chapter 4-7). Beside (ideal) spectroscopic properties, environmental effects due to the surrounding solvent or the attached biochemical target play a crucial role for the decision. Ideally, fluorophores show environment and solvent independent spectroscopic properties for the application selected. The following effects have to be considered:

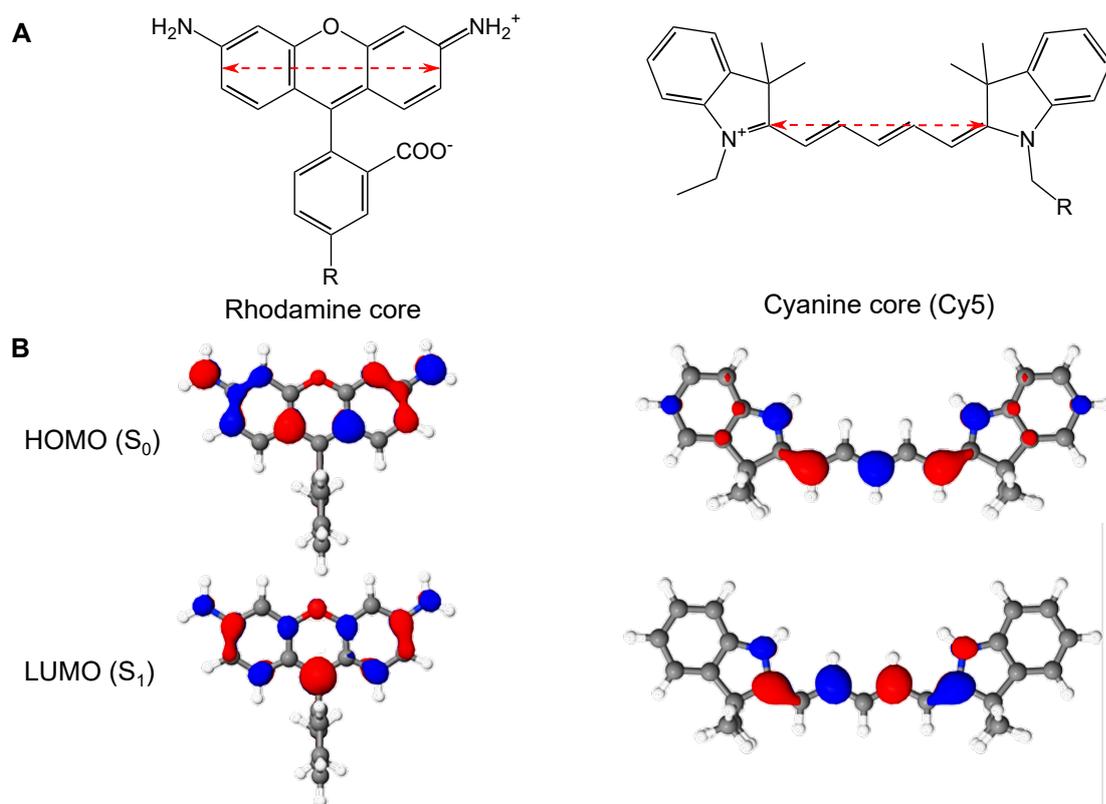


Figure 2.1.3: Structure and conjugated π -orbital system of common organic fluorophores. **A** Chemical structure of a rhodamine and a cyanine (Cy5) core with illustrated transition dipole (red arrow). **B** Illustration of the wave functions in the electronic ground state S_0 and the first excited state S_1 for the fluorophores from **A** (color indicating the sign of the wave function) [30].

- Uncharged (hydrophobic) fluorophores can dimerize and form aggregates in a polar environment, which alters the absorption and emission spectra [24].
- Fluorophores from the cyanine class undergo light-induced isomerization between a fluorescent trans-state and a dark cis-state [36]. The equilibrium between the two states is influenced by the viscosity or steric hindrance of the surrounding. For labeled proteins, this effect is known as protein-induced fluorescence enhancement (PIFE), since the quantum yield increases if the fluorophore interacts with the protein [37, 38].
- Solvents with different polarities and/or viscosities may lead to spectral shifts and a change in quantum yield due to an interaction with the fluorophore states [39, 40].
- Aromatic molecules, such as some amino acids (e.g. tryptophan), or metal ions quench the fluorophores by collisional quenching [40–43].

2.2 Förster resonance energy transfer

This thesis details the use of biophysical assays based on fluorophore-fluorophore interaction for nanometer distance measurements in bio-macromolecules. The previous section focused on the properties of an isolated fluorescent emitter and illustrated the general concepts of the underlying fluorescence emission. Considering a pair of fluorophores, we need to take the interaction into account, which can be separated into three distance regimes. For distances smaller than 2 nm (which is approximately the size of common fluorophores), collisional quenching [2], Dexter energy transfer [44], and photoinduced electron transfer (PET) [45] dominate the interaction of the two fluorophores. Further, fluorophore dimerization may change the energy levels under the formation of aggregates [24]. At larger distances ($R > 10$ nm), far field interaction occurs in form of classical light emission and absorption, which scales with R^2 . In between these two distance regimes (2-10 nm), the non-radiative dipole-dipole interaction of the two fluorescent emitters dominates the interaction. This is known as Förster resonance energy transfer (FRET).

2.2.1 Basic theory

The basic principle behind the energy transfer rate k_{FRET} from the donor fluorophore to an acceptor fluorophore is described via the absolute square of the transition matrix

$$k_{\text{FRET}} \sim \left| \langle \psi_{D_a} \psi_{A_b'} | V | \psi_{D_b} \psi_{A_b'} \rangle \right|^2, \quad (2.2.1)$$

where V is the perturbation, which takes the donor from the excited state ψ_{D_b} and the acceptor in its ground state $\psi_{A_b'}$ to a quenched donor ψ_{D_a} and an acceptor in the excited state $\psi_{A_b'}$ according to Fermi's golden rule [25, 46]. The perturbation term V scales with $1/R^3$ for the dipole-dipole interaction, which leads to the typical distance dependency of $k_{\text{FRET}} \sim 1/R^6$ for the energy transfer rate.

There are several implicit and explicit assumptions in theoretical description of FRET. The theory derived by Theodor Förster [10] relies on the explicit assumptions that

- the interaction of the two fluorophores is dominated by dipole-dipole interactions, which is valid for fluorophore distances well beyond the extent of the involved molecules and well below the far-field regime dominated by the photon emission.
- the fluorophore has a continuum of states, which can be well described by a density function.

Based on these assumptions, one can connect the state transition rate k_{FRET} directly to spectroscopic and geometric quantities.

$$k_{\text{FRET}} = (k_{\text{rad}} + k_{\text{nr}}) \left(\frac{R_0}{R} \right)^6, \quad (2.2.2)$$

where the so-called Förster-radius R_0 reads

$$R_0 := \sqrt[6]{\frac{9 \ln(10)}{128 \pi^5 N_A} \frac{\kappa^2}{n^4} Q_D \frac{\int_0^\infty F_D(\lambda) \varepsilon_A(\lambda) \lambda^4 d\lambda}{\int_0^\infty F_D(\lambda) d\lambda}}. \quad (2.2.3)$$

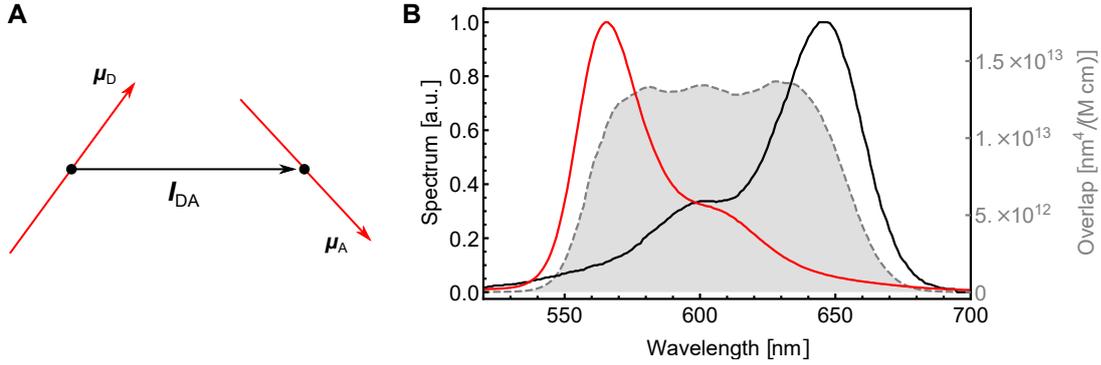


Figure 2.2.1: FRET theory. **A** For two dipoles μ_D and μ_A separated by a distance $R = |\mathbf{I}_{DA}|$, the transfer rate is averaged over all possible dipole orientation. **B** The emission spectrum of Cy3 (red) and the absorbance spectrum of Cy5 (black) result in a spectral overlap $\tilde{F}_D(\lambda)\varepsilon_A(\lambda)\lambda^4$ (gray).

Here, N_A is the Avogadro number, Q_D the donor quantum yield, and n the effective refractive index between donor and acceptor. κ^2 is an orientation factor averaging the dipole-dipole interaction over all potential orientations of the donor transition dipole moment $\mu_{D_{ba}}$ and acceptor transition dipole moment $\mu_{A_{ab}}$ as

$$\kappa^2 = \left\langle \left(\mu_{D_{ba}} \cdot \mu_{A_{ab}} - 3(\mathbf{I}_{DA}\mu_{D_{ba}})(\mathbf{I}_{DA}\mu_{A_{ab}}) \right)^2 \right\rangle_{\theta_D, \phi_D, \theta_A, \phi_A}, \quad (2.2.4)$$

where \mathbf{I}_{DA} is the normalized vector from donor to acceptor position (see Figure 2.2.1A). For a complete isotropic sample, the orientation value reads as

$$\kappa^2 = \int_0^\pi d\theta_D \int_0^{2\pi} d\phi_D \int_0^\pi d\theta_A \int_0^{2\pi} d\phi_A \left(\mu_{D_{ba}} \cdot \mu_{A_{ab}} - 3(\mathbf{I}_{DA}\mu_{D_{ba}})(\mathbf{I}_{DA}\mu_{A_{ab}}) \right)^2 = \frac{2}{3}, \quad (2.2.5)$$

which is normally used for the calculation of R_0 .

The term

$$J = \int_0^\infty \tilde{F}_D(\lambda)\varepsilon_A(\lambda)\lambda^4 d\lambda \quad (2.2.6)$$

is referred to as overlap integral with the normalized emission spectrum of the donor $\tilde{F}_D(\lambda) = F_D(\lambda)/\left(\int_0^\infty F_D(\lambda)d\lambda\right)$ and the extinction coefficient of the acceptor ε_A (see Figure 2.2.1B).

A rigorous derivation of FRET theory can be found in *Fluorescence resonance energy transfer* (Clegg, 1996) [46].

It has to be noted that the presented relations rely on several implicit assumption which should be considered in every experiment:

- A single transition rate only holds for a homogeneous sample without heterogeneous broadening due to environmental effects. This is often the case for stochasti-

cally labeled proteins, where donor and acceptor fluorophore are randomly labeled at the two labeling positions.

- The averaging of dipole orientations assumes fast rotation ($k_{\text{rot}} \ll k_{\text{rad}}$) and uniform and isotropic dipole orientation distribution.
- A simple effective refractive index describes the medium surrounding donor and acceptor homogeneously.

The FRET-efficiency E is defined as the ratio of the deexcitation rate of the donor via the energy transfer and the total deexcitation rate (equivalent to a yield of energy transfer) and reads as (equation 2.2.2):

$$E := \frac{k_{\text{FRET}}}{k_{\text{FRET}} + k_{\text{rad}} + k_{\text{nr}}} = \frac{1}{1 + \left(\frac{R}{R_0}\right)^6}. \quad (2.2.7)$$

Considering potential interaction and sticking with non-uniform and/or non-isotropic dipole orientations, it may be useful to exclude the sample-specific orientation factor κ^2 from the Förster-radius constant and rewrite the FRET-efficiency as

$$E = \frac{1}{1 + \frac{2}{3\kappa^2} \left(\frac{R}{R'_0}\right)^6}, \quad (2.2.8)$$

where κ^2 is in the range $0 < \kappa^2 < 4$ and R'_0 the Förster radius constant under isotropic conditions. Normally, this value cannot be measured directly, but it can be confined by anisotropy measurements, or alternatively, estimated from geometric models[47] or from molecular dynamics (MD) simulations [30, 48].

2.2.2 Lifetime-based FRET efficiency determination

From the definition in equation 2.2.7, we can directly link the FRET efficiency E to the donor lifetime $\tau = (k_{\text{FRET}} + k_{\text{rad}} + k_{\text{nr}})^{-1}$ under energy transfer conditions and the fluorescent lifetime $\tau_{fl} = (k_{\text{rad}} + k_{\text{nr}})^{-1}$ in the absence of an acceptor molecule (see equation 2.1.7) as

$$E = 1 - \frac{\tau}{\tau_{fl}}. \quad (2.2.9)$$

Therefore, the lifetime of the donor is a direct read-out of the FRET-efficiency E , and thus, directly related to the fluorophore distance R .

2.2.3 Intensity-based FRET efficiency determination

Alternatively, splitting the fluorescence emission signal (number of photons) from donor F_{DD} and F_{DA} after excitation of the donor allows to approximate the FRET efficiency as

$$E^* \approx \frac{F_{DA}}{F_{DD} + F_{DA}}. \quad (2.2.10)$$

This equation is strictly true only for fluorophores with (i) equal quantum yield, (ii) without background signal, where (iii) the signal from donor and acceptor can be separated completely, and (iv) no contribution of a directly excited acceptor is present. Since these assumptions fail in practical experiments, the value is called raw FRET efficiency and correction steps for the recorded signals are needed to derive the ‘true’ (accurate) FRET efficiency E as described in chapter 5 and 6 in more detail.

2.3 Confocal microscopy

A confocal arrangement is commonly used in fluorescence microscopy because of its small observation volume due to an overlaid excitation and detection focus, resulting in a high signal-to-noise ratio and good suppression of Raman scattering from solvent molecules [49]. The basic principle was invented and patented by Marvin Minsky in 1957 [50], where a pinhole blocks out-of-focus background light in the detection path.

2.3.1 Basic principles

The focus of a laser beam is well approximated by a Gaussian beam under the paraxial approximation with an Gaussian intensity profile

$$I(x, y, z) = I_0 \left(\frac{w_0}{w(z)} \right)^2 e^{-2\frac{x^2+y^2}{w(z)^2}} \quad (2.3.1)$$

with the maximum intensity I_0 , the beam waist w_0 , and the z -dependent $1/e^2$ beam radius $w(z)$ with

$$w(z) = w_0 \sqrt{1 + \left(\frac{z}{z_R} \right)^2}, \quad (2.3.2)$$

where $z_R = \frac{\pi w_0^2 n}{\lambda}$ is the so-called Rayleigh range (n is the refractive index and λ the laser wavelength) [51]. For collimated laser beams, which are focused by a lens or objective, the beam waist can be approximated by $w_0 \approx \frac{2\lambda}{\pi \text{NA}}$. However, this relation becomes less accurate for high numerical apertures $\text{NA} > 1$ (where the paraxial approximation is no longer fully valid), but still with reasonable good accuracy [51]. In the limit of the high intensity region of the beam (focus), the intensity profile can be simplified by the approximation

$$I(x, y, z) = I_0 e^{-2\left(\frac{x^2+y^2}{w_0^2} + \frac{z^2}{w_z^2}\right)}, \quad (2.3.3)$$

with the $1/e^2$ focal length w_z in z -dimension. For a standard excitation wavelength of 532 nm and a high numerical aperture objective ($\text{NA} = 1.2 - 1.4$), the typical dimensions

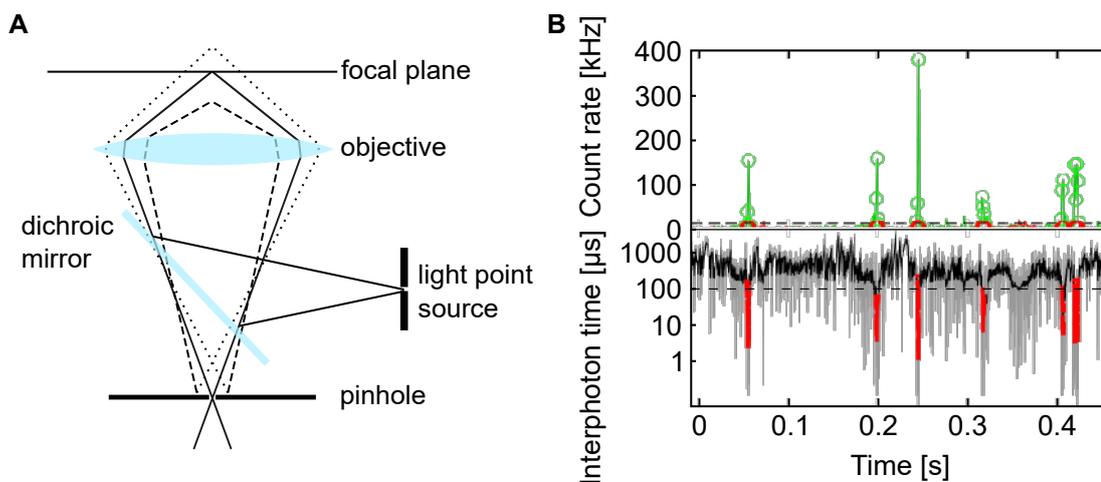


Figure 2.3.1: Single-molecule confocal microscopy. **A** Illustration of the confocal microscope principle. The light from a point source is focused with the objective. The detection light is collected and focused onto the detection pinhole, where out-of-plane light (dotted/dashed) is blocked and only light from the focal plane can pass. **B** Photon counts time trace of a picomolar fluorophore concentration (top). The average background noise is on the level of $\sim 1 - 5$ kHz and the fluorescent signal of ~ 100 kHz. The interphoton-time (grey) varies over several orders of magnitudes (bottom). An averaged or smoothed interphoton-time allows to threshold between background (black) and signal (red).

of the excitation volume yields $w_0 \approx 200 - 300$ nm and $w_z \approx 1 - 2$ μm . In fluorescence confocal microscopy, the detection light is filtered by a pinhole, which suppresses out-of-focus scattering and background light by blocking non-collimated light behind the infinity-corrected objectives (see Figure 2.3.1A) [51]. Typically, the optimal pinhole diameter is around 1 AU (airy-unit, distance of the first minima on the airy disc), which is a good compromise between maximum detection yield and axial/lateral resolution [52]. In single-molecule spectroscopic applications, an underfilling of the objective lens (laser beam diameter is smaller than the objective diameter) may be used to reduce the effective numerical aperture, which increases the size of the point-spread function (PSF), and thus the diffusion time of molecules inside the excitation volume, despite the use of a high NA objective lens.

Modern confocal microscopes use infinity-corrected objectives, where all light originating from the sample is collimated. This implementation requires a collimated light source in front of the objective and an additional lens to focus the detection light onto the pinhole [53]. This setup simplifies the alignment and the insertion of filters, polarizers, and other optical elements, which is possible without beam distortion (see section 3.4 for more details).

2.3.2 Single-molecule burst detection and photon statistics

The use of photostable and bright organic fluorophores (see section 2.1.5) on modern microscopy setups with single-photon detection allows the detection of single molecules. From equation 2.3.1, we can derive an effective confocal volume $V_{\text{eff}} = (\pi/2)^{3/2} w_0^2 w_z$, which is on the order of ~ 1 fL for typical high-NA microscope objectives and wavelengths in the visible range. Considering a realistic confocal volume of 0.5 fL and a fluorescent particle concentration of 50 pM in solution, the average number of particles in the volume is $\langle p(N) \rangle \sim 0.015$. As for typical counting experiments, the probability of N fluorescent particles in the focus follows a Poissonian distribution with $\lambda = \langle p(N) \rangle$:

$$p(N) = \frac{\lambda^N}{N!} e^{-\lambda} \stackrel{N \text{ large}}{\approx} \frac{1}{\sqrt{2\pi\lambda}} e^{-\frac{(N-\lambda)^2}{2\lambda}}. \quad (2.3.4)$$

Therefore, the probability to find one particle is $P(N = 1) = 1.4\%$ and for more than one particle at a time the probability is $P(N > 1) = 1 - P(N \leq 1) \approx 0.011\%$. Thus, single-molecule detection is >100 times more frequent than multiple-molecule detection. It should be mentioned that for this calculation example, 98.5% of the time, no fluorescent particle is in the focus, which leads to the characteristic burst-signal of solution-based single-molecule microscopy (see Figure 2.3.1B).

Burst detection algorithms In the typical application of single-molecule Förster resonance energy transfer (smFRET), the intensity traces are dissected into parts containing background or actual signal from fluorescent molecules. These photon packages from an individual molecule (so-called burst marked with circles in Figure 2.3.1B) can be extracted from the time trace with multiple approaches: A simple approach is to bin the data in time windows with a window size on the order of the diffusion time of the fluorescent molecule and a consecutive thresholding. An analysis of the interphoton time smoothed with a Lee filter avoids that photons of one molecule are chopped into multiple bins [54, 55]. Alternatively, sliding window algorithms are often used to identify the starting and ending time of a burst and integrate all photons in between. Here, two alternatives are present: The all photon burst search (APBS) algorithm, where all photons in different detection channels are summed up to one photon stream [17, 56, 57]. For multiple detection channels with two excitation sources (see section 3.4), a separate analysis of the detection signal is performed with a dual channel burst search (DCBS) to detect photo-bleaching, photo-blinking events and random coincident of multiple molecules in the focal spot [58]. For the sliding window analysis, the typical time window is 500 μs , a minimal required photon number per window is 5-15 and an overall minimum required number of photons per burst event N_{min} is in the range of 30-200 photons.

Photon statistics The photon statistics in solution-based single molecule measurements consists of two contributions of Poissonian background and the individual signals from fluorescent molecules. An analysis of inter-photon delay times allows to extract

the background signal from a log-linear fit to the event number of inter-photon delay times [59, 60]. A more simplistic approach estimates the background by binning the data in fixed bins (e.g. 1 ms) and averaging over all bins with low counting numbers (e.g. <10), which is a first order approximation to estimate $\lambda_{\text{bkgr}} = \langle N_{\text{bkgr}} \rangle \approx \langle N_{\text{signal}} \rangle_{N < 10}$. For FRET measurements with multiple detection channels, this is done for all channels individually.

The fluorescent signal distributed into the two detection channels in FRET experiments follows binomial statistics as

$$p(F_{DA} | \langle E^* \rangle, N) = \binom{N}{F_{DA}} \langle E^* \rangle^{F_{DA}} (1 - \langle E^* \rangle)^{N - F_{DA}} \quad (2.3.5)$$

for a given number of photons N with $N = F_{DD} + F_{DA}$.

The standard width σ_{E^*} of the FRET efficiency distribution for a homogeneous, single distance fluorophore pair with theoretical raw FRET efficiency $\langle E^* \rangle$ is given as $\sigma_{E^*}^2 = \left\langle \left(\frac{F_{DA}}{F_{DD} + F_{DA}} - \langle E^* \rangle \right)^2 \right\rangle_{F_{DA}, F_{DD}}$, which depends on the actual counting probability distribution $p(N)$ to detect N photons. In practice, the detected photon events (bursts) are typically filtered by a lower threshold $N \geq N_{\text{min}}$, which defines an upper bound for the standard width σ_E as

$$\sigma_{E^*}^2 \leq \frac{E^*(1 - E^*)}{N_{\text{min}}}, \quad (2.3.6)$$

where the equal case occurs if $N = N_{\text{min}}$ for all bursts [61].

A full description of the photon statistics based on an assumed distribution $p(N)$ (estimated from the measurement) allows to simulate the FRET data with photon distribution analysis (PDA). Here, the Poissonian background statistics, the binomial distribution of donor (F_{DD}) and acceptor channel photons (F_{DA}) and the photon count statistics $p(N)$ are considered to fully reproduce the complete FRET efficiency histogram [62]. It can be expanded by dynamic terms simulating molecular dynamics to describe the interchange between different (static) FRET states [63, 64].

A simplified version to detect molecular dynamics (exchange between two FRET efficiency states $E_1^* \neq E_2^*$) based on photon statistics is the burst variance analysis (BVA), which analyses multiple subset of photons for each burst [65]. Dividing the N burst photons of a single burst in $\lfloor N/n \rfloor$ fixed sets (s_i) of n photons (often $n = 5$ is used), similar to equation 2.3.6, we expect an average standard deviation of

$$\langle \sigma_{E^*} \rangle_{s_i} = \sqrt{\frac{E^*(1 - E^*)}{n}}, \quad (2.3.7)$$

for a given raw FRET efficiency E^* . Thus, the burst variance analysis can be used for multiple bursts at a given raw FRET efficiency to detect dynamics if a statistically significant deviation

$$\left\langle \langle \sigma_{E^*} \rangle_{s_i} \right\rangle_{\text{bursts}} > \sqrt{\frac{E^*(1-E^*)}{n}} \quad (2.3.8)$$

is observed (see section 3.4.4 for more details). Since the inequality is only significant, if the dynamics occur on a time scale of the observation time ($\sim 1 \text{ ms}$) and slower than the average detection time for multiple packages of n consecutive photons, the sensitivity of this analysis method is limited to the time range between $\sim 200 \mu\text{s} - 5 \text{ ms}$ [66].

2.3.3 Fluorescence correlation spectroscopy

Based on the intensity profile following from equation 2.3.3 and the Poissonian nature of photon emission, the fluorescence signal can be fully described for diffusing molecules. The analysis of fluctuations of fluorescence intensity with fluorescence correlation spectroscopy uses the auto-correlation of the signal to obtain insights into dynamic processes. The second-order auto-correlation function is defined by

$$G(\tau) = \frac{\langle \delta I(t) \delta I(t + \tau) \rangle_{\tau}}{\langle I(t) \rangle_{\tau}^2}, \quad (2.3.9)$$

where $I(t)$ is the fluorescent intensity at time t and $\delta I(t) = I(t) - \langle I(t) \rangle$ the time-dependent fluctuation.

For the confocal intensity profile, the second-order correlation function for a diffusing molecules with diffusion constant D reads as

$$G(\tau) = \frac{1}{\langle N \rangle} \frac{1}{(1 + 4D\tau/w_0) \sqrt{1 + 4D\tau/w_z}}, \quad (2.3.10)$$

where $\langle N \rangle$ is the average number of molecules in the effective confocal-volume $V_{\text{eff}} = (\pi/2)^{3/2} w_0^2 w_z$ as described above. A full derivation on the auto-correlation can be found in reference [24] (chapter 5). The above derivation assumes photon emitters with constant photon emission rates over time, which only depend on the position in the focus. For common fluorophores, the photon emission rates vary due to rate-interfering processes as briefly discussed in the beginning of this chapter. Considering all effects, the full auto-correlation function is described as the product of all contributions

$$G(\tau) = G_{\text{diffusion}}(\tau) \cdot G_{\text{photophysics}}(\tau) \cdot G_{\text{dynamics}}(\tau) \cdot G_{\text{antibunching}}(\tau), \quad (2.3.11)$$

where the different factors contribute on distinct timescales ranging from picoseconds (antibunching) to hundreds of milliseconds or seconds (triplet relaxation, protein dynamics). Due to the fact, that the signal of different molecules is fully uncorrelated and does not contribute to the auto-correlation, the overall auto-correlation for multiple species is fully separable and reads

$$G(\tau) = \Upsilon_1 G_1(\tau) + \Upsilon_2 G_2(\tau) \quad (2.3.12)$$

with the fractional intensity $\mathcal{Y}_i = \varepsilon_i \langle N_i \rangle / (\varepsilon_1 \langle N_1 \rangle + \varepsilon_2 \langle N_2 \rangle)$, where ε_i are the species intensities [67].

3 | Measurement and evaluation techniques

While the previous chapter introduced fluorescence principles and gave theoretical concepts to single-molecule measurements, this chapter is supposed to shed light on practical aspects of the measurements. This starts with a brief introduction to protein labeling and followed by a section about spectroscopic measurements of quantum efficiency, lifetime and anisotropy. Then, two different smFRET-measurements will be introduced including standard evaluation and correction procedures.

3.1 Protein labeling

A common strategy for the fluorescent labeling of proteins is the attachment of organic dyes to selected amino-acids via reactive groups of the fluorophore [22]. Binding to the abundant lysine residues via NHS-ester functionalized fluorophores as well as His-tag or N-terminal labeling gives no choice regarding the labeling position. Thus a widely used strategy is to label inserted cysteine residues site-specifically via a maleimide group, which enables the attachment of an organic fluorophore by a thiol-maleimide reaction [68, 69] (see Figure 3.1.1A) at, in principle, any position with high labeling efficiency and specificity [22, 70]. Alternatively, incorporation of unnatural amino-acids (UAAs) has become an alternative when native cysteine hinder site-specific incorporation [71–76]. However, here we focus solely on cysteine-labeling in this thesis. While the careful selection of suited labeling locations is a central part of this thesis (see chapter 7), the generation of the protein samples is not addressed. More details on the cloning, protein expression, and purification can be found in references [77–80].

3.1.1 Labeling protocol

In general, proteins were stored in a buffer at -20 °C containing 50% glycerol to avoid crystallization and 1 mM DTT (dithiothreitol) to avoid cysteine oxidation and disulfide bond formation (protein dimer formation).

In short, His-tagged proteins were immobilized on a Ni-Sepharose 6 Fast Flow resin (GE Healthcare) and the DTT-containing buffer was washed off. The resin was incubated 4-12 h at 4°C in the labeling buffer at pH 7.4 - 8.0 with 25 nmol of each fluorophore. At pH 7.4, the reaction rate of maleimide with thiols is much faster than the reaction rate of maleimide with amines. However, at higher pH, free primary amines react competitively with thiols at the maleimide C=C bond [81]. The column was washed sub-sequentially

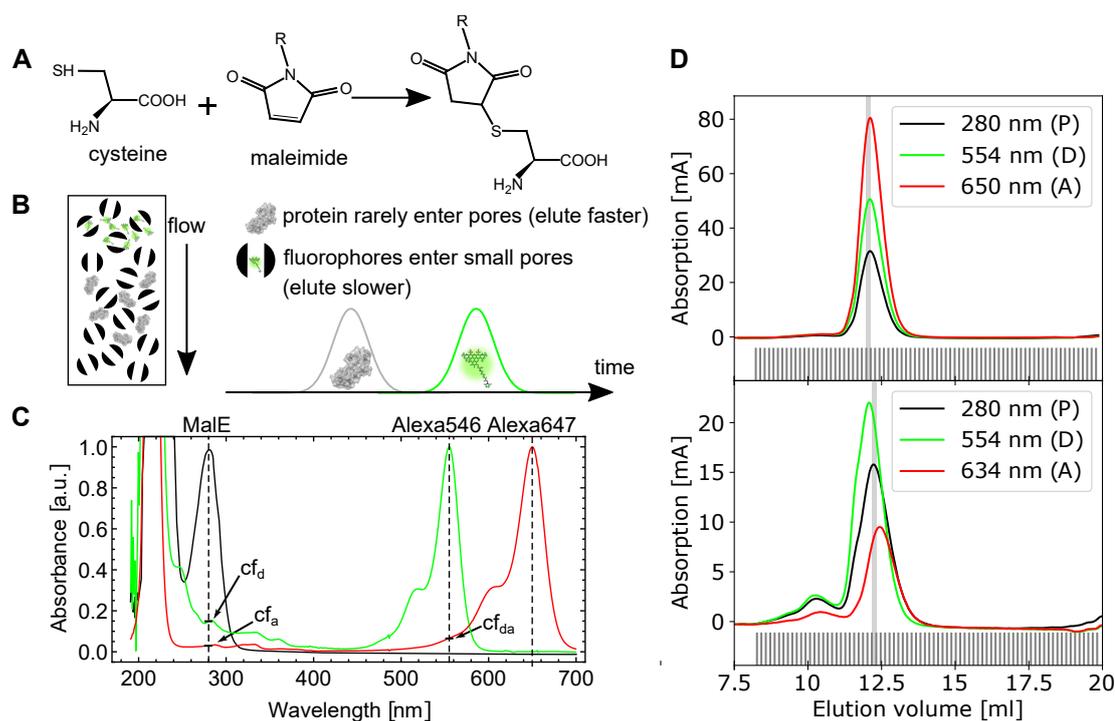


Figure 3.1.1: Protein labeling principle. **A** Thiol-maleimide reaction links the cysteine and the fluorophore. **B** Size exclusion chromatography with porous material separates molecules by size due to different flow velocities. **C** Absorbance spectra of a protein (MalE, black), Alexa Fluor 546 (green), and Alexa Fluor 647. The vertical lines illustrate the measurement points in the size exclusion with marked correction factors. **D** Exemplary size exclusion chromatograms of MalE mutants 83-205 labeled with Alexa Fluor 546/Alexa Fluor 647 (top) and mutant 29-352 labeled with Alexa Fluor 546/Star 635P (bottom).

with labeling buffer with 5% and 50% glycerol to remove unbound fluorophores. Bound proteins were eluted with 500 μ l of elution buffer containing 500 mM imidazole.

3.1.2 Size exclusion chromatography

The labeled and eluted protein is further purified by size-exclusion chromatography, where species of different size (protein aggregates, proteins, free fluorophores, DTT) can be separated through size-dependent diffusion velocities in a porous material (see Figure 3.1.1B). The column (Superdex 75 Increase 10/300 GL, GE Healthcare) was chosen to be selective in the range of protein masses with 3 - 70 kDa. The elution of the protein, fluorophores and contaminants is monitored by absorbance measurements at the maximum wavelengths of protein (280 nm) and the labeled fluorophores (see Figure 3.1.1C, dashed line). Figure 3.1.1D shows two exemplary size-exclusion chromatograms of two different maltose-binding protein (MalE) mutants from *E. coli* [82] labeled with Alexa Fluor 546 / Alexa Fluor 647 (MalE 83-205, top) and Alexa Fluor 546 / Star 635P (MalE 29-352, bottom). The upper profile shows a very symmetric and clean

protein peak without aggregation, where the size exclusion mainly serves to remove free fluorophore and imidazol (both elute at larger volumes). However, the lower profile shows a clear aggregation peak (around 10.5 ml) and an asymmetric absorption profiles at the different wavelengths. A careful selection of one fraction (gray area) allows to optimize the labeling ratio and eliminate protein aggregates. Distinct retardation of fluorophore-labeled proteins occurs for hydrophobic dyes (Atto647N) but also for Star635P (phosphorylated rhodamine).

3.1.3 labeling efficiency calculation

The labeling ratio (number of fluorophores per protein) can be calculated from the absorbance values, which are composed of the different protein and fluorophore contributions as

$$\underbrace{\begin{pmatrix} A_{280} \\ A_d \\ A_a \end{pmatrix}}_{\vec{A}} = l \underbrace{\begin{pmatrix} \varepsilon_p & cf_d \varepsilon_d & cf_a \varepsilon_a \\ 0 & \varepsilon_d & cf_{da} \varepsilon_a \\ 0 & \underbrace{cf_{ad} \varepsilon_d}_{\approx 0} & \varepsilon_a \end{pmatrix}}_{\Sigma} \underbrace{\begin{pmatrix} c_p \\ c_d \\ c_a \end{pmatrix}}_{\vec{c}}. \quad (3.1.1)$$

In this formula, c_p , c_d , and c_a are the protein, donor fluorophore, and acceptor fluorophore concentrations and ε_p , ε_d , and ε_a their extinction coefficients, respectively, and l is the pathlength. The correction factors cf_i of the fluorophores account for contributions due to spectral overlap at the measured wavelengths and can be determined from the absorbance spectra (see Figure 3.1.1C). The correction factors for the donor in the acceptor channel cf_{ad} can be neglected for most fluorophore pairs, if the spectra are well separated.

The extinction coefficient for the proteins is either taken from literature or estimated as

$$\varepsilon_p = N_W \cdot 5500 \frac{1}{Mcm} + N_Y \cdot 1490 \frac{1}{Mcm} + N_C \cdot 125 \frac{1}{Mcm}, \quad (3.1.2)$$

where N_W , N_Y , and N_C are the number of tryptophan, tyrosin, and cysteine residues in the protein, respectively [83]. This approach assumes the extinction coefficients from the absorbing amino acids in the protein to be similar to the value for isolated amino acids.

From equation 3.1.1, we obtain the concentrations as $\vec{c} = 1/l \Sigma^{-1} \vec{A}$ and calculate the donor and acceptor labeling efficiencies c_d/c_p and c_a/c_p , respectively. For the example in 3.1.1D (gray fraction), we obtain 1.11/0.91 (top) and 1.12/0.46 (bottom) donor/acceptor per protein for the two labeling sites. Thus the combined labeling efficiency is 101% and 78%, respectively.

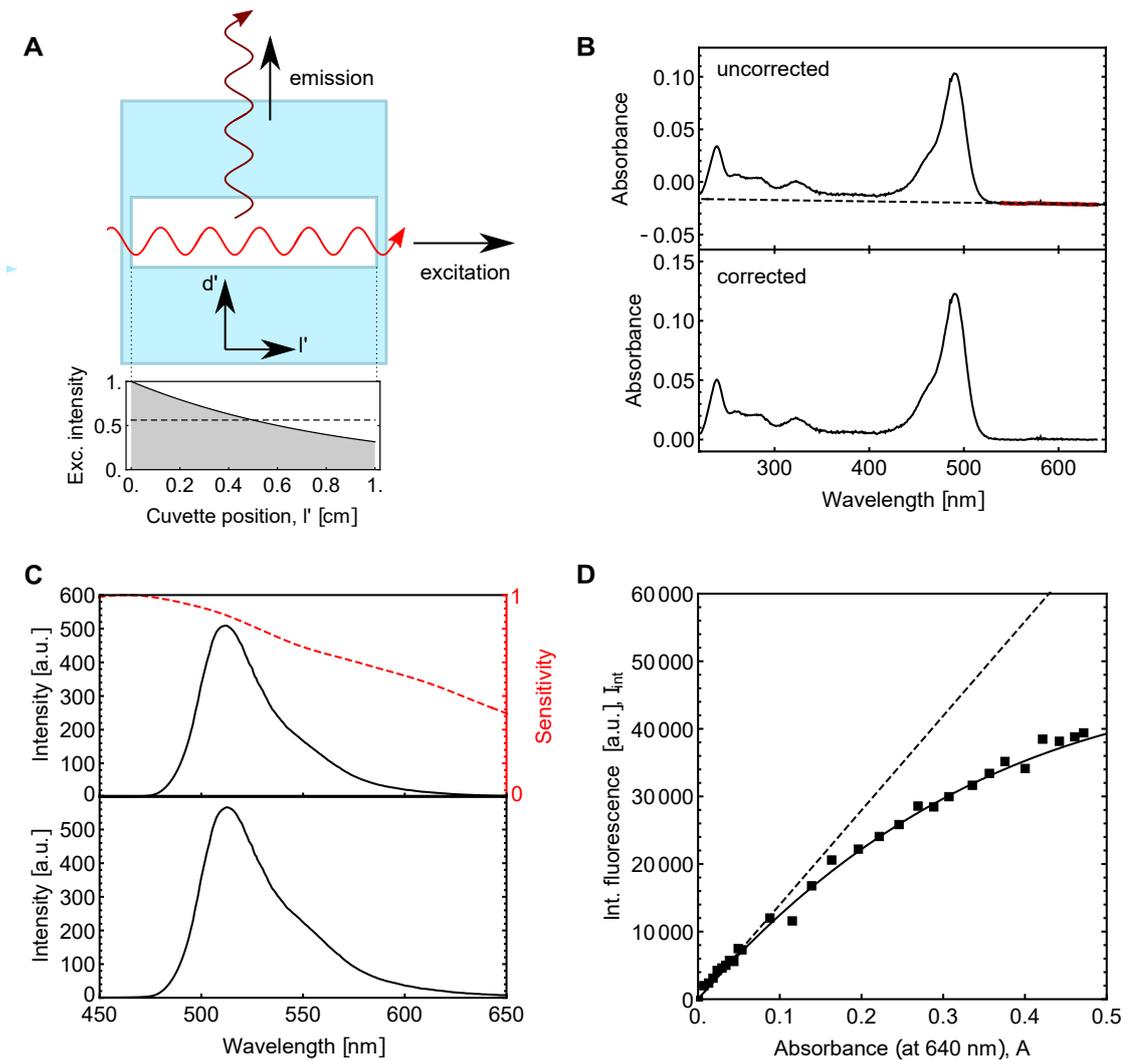


Figure 3.2.1: Quantum efficiency measurement. **A** The cross section of a cuvette (blue) illustrates the excitation (light red) and emission path (dark red) of the light. The intensity profile of the light decays exponentially across the light path due to absorption (bottom graph). **B** Uncorrected (top) and background-corrected absorbance spectrum of GFP. **C** Uncorrected emission spectrum (top, black) and sensor sensitivity (top, red) and sensor corrected emission spectrum of GFP (bottom). **D** The integrated emission from a titration curve is plotted versus the absorbance at 640 nm for Alexa Fluor 647 (data points). The data is fitted according to equation 3.2.3 (solid, black) with the initial slope (dashed, black).

3.2 Fluorescence quantum efficiency determination

According to the definition of the quantum efficiency (equation 2.1.6), the quantum efficiency is determined by the rates of the competing deexcitation pathways of the excited

state. These rates are difficult to be measured directly. Alternatively, the quantum efficiency can be expressed as $\Phi = \frac{\text{\#emitted photons}}{\text{\#absorbed photons}}$, which is used to measure the quantum efficiency of standard samples with calibrated integrated-sphere setups according to *Mello's method* [84]. In practice, the quantum efficiency can be determined in reference to a quantum efficiency standard by standard absorbance and emission measurements [85, 86].

3.2.1 Spectrum normalization

For the quantum efficiency determination, absorbance measurements are performed in buffer in a 10 mm (L) x 2 mm (H) cuvette on a continuous-wave UV/VIS spectrometer (LAMBDA 465, Perkin Elmer). Fluorescence emission is recorded on the same sample with a fluorescence spectrometer (LS 55, Perkin Elmer) with excitation/emission slit width of 5 nm and gain values set to 775 V for the photomultiplier tube (PMT R928, Hamamatsu).

In these cuvette measurements, the emitted light is detected perpendicular to the excitation light as shown in Figure 3.2.1A in contrast to the absorption measurements in one direction. According to the Lambert-Beer law, the absorbance A scales linearly with the sample concentration

$$A = \varepsilon \cdot l \cdot c, \quad (3.2.1)$$

where ε is the molar extinction coefficient, l the path length of the light, and c the concentration. While the absorption can be measured very precisely, an offset correction is needed for samples with significant background (in comparison to the standard reference). Therefore, a linear offset is estimated by a linear fit (see Figure 3.2.1B, dashed line) to the long wavelength range (red data points), where no contribution from the fluorophore or biological sample is expected. The extracted absorbance wavelength (at the excitation of the emission measurements) is averaged over the slit width (5 nm).

In contrast, the emission intensity highly depends on the sample concentrations due to various absorption effects of the excitation and the emission light as well as the wavelength due to wavelength-dependent detection sensitivity of the PMT. Therefore, the spectra are corrected by the theoretical detection sensitivity from manufacturer specifications (see Figure 3.2.1C, dashed line). From the normalized spectra in Figure 3.2.1B/C, the integrated intensity $I_{int} = \int_0^\infty I(\lambda) d\lambda$ shows a non-linear dependency of the absorbance (and concentration) and it could even decrease upon increase of the fluorophore concentration. This effect is due to absorption effects in emission measurements at higher concentrations, when the light is mainly collected from a small part of the cuvette, where the light enters (see Figure 3.2.1A).

Overall, the measured intensity is expressed as

$$I = I_o \int_0^d \int_0^l \underbrace{10^{-\varepsilon_x l' c}}_{\text{excitaton light}} \cdot \sigma \cdot \Phi \cdot \underbrace{(1 - 10^{-\varepsilon_{em} d' c})}_{\text{emission light}} \cdot D_{\text{eff}}(l', d') dd' dl' \quad (3.2.2)$$

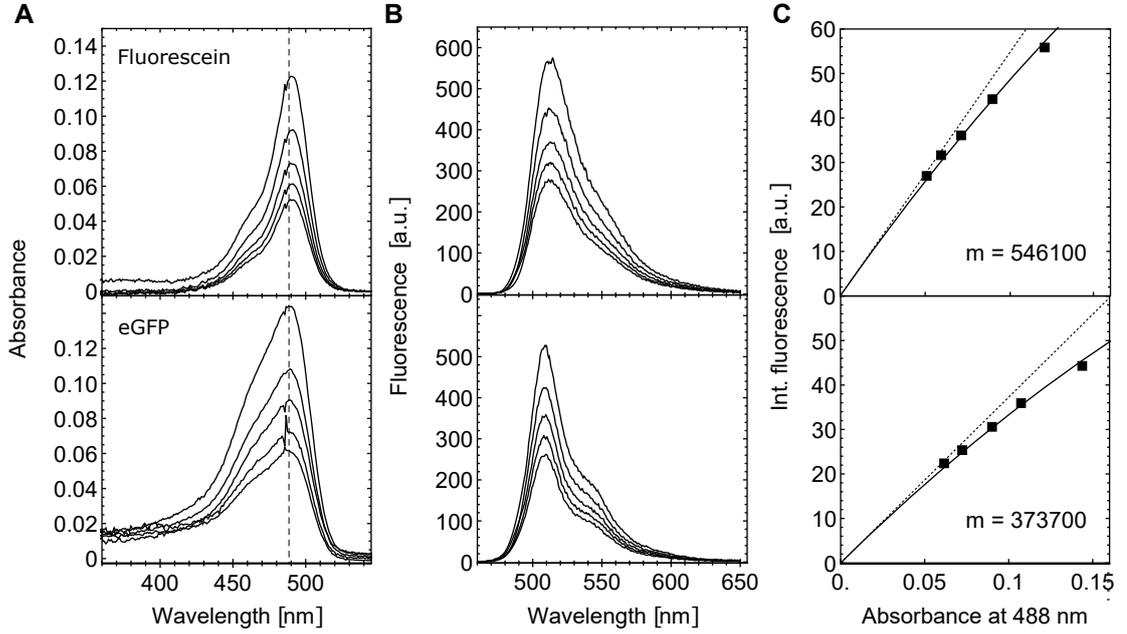


Figure 3.2.2: Quantum efficiency measurement of eGFP in reference to fluorescein. **A** Absorbance spectra of fluorescein (top) and eGFP (bottom) at 5 different concentrations. The dashed line indicates the extracted absorbance values from the excitation wavelength at 488 nm. **B** Emission spectra of fluorescein (top) and eGFP (bottom) of the sample in (A) excited at 488 nm. **C** The integrated emission from (B) is plotted against the absorbance extracted from A and fitted to the function $I_{int} = mA \cdot 10^{-A/2}$ (solid line). The dashed line indicates the slope $I_{int} = mA$ without absorption effects (dashed line).

where I_0 is an initial excitation intensity, ε_{ex} and ε_{em} the extinction coefficients at the excitation and the emission wavelength, c the fluorophore concentration, σ the absorbance cross-section, Φ the quantum efficiency, d and l the width and the length of the cuvette and D_{eff} a location-dependent detection efficiency (additional spectral components are neglected in this consideration). At low concentrations ($c \lesssim 1 \mu\text{M}$) the effect of re-absorption of the emission light can be omitted, since $\varepsilon_{em} \ll \varepsilon_{ex}$ (and $d < l$ for the used cuvettes). Assuming that the detection efficiency D_{eff} is rather constant over the cuvette cross-section, equation 3.2.2 simplifies to $I = \tilde{I}_0 \cdot c \cdot \int_0^l 10^{-\varepsilon_{ex}l'} c dl' \approx \tilde{I}_0 \cdot c \cdot 10^{-\varepsilon_{ex}l/2c}$, where \tilde{I}_0 summarizes all concentration independent factors. Therefore, the integrated intensity I_{int} can be expressed as a function of the absorbance A as

$$I_{int} = m \cdot A \cdot 10^{-A/2} \quad (3.2.3)$$

for small concentrations, where reabsorption effects are estimated to be linear. A calibration measurement of the reference sample Alexa Fluor 647 in Figure 3.2.1D shows this non-linear relation. A fit to the function $I_{int} = m \cdot A \cdot 10^{-rA}$ reveals $r = 0.48 \pm 0.03$

with good fit quality, which justifies the simplified relation in equation 3.2.3 in the absorbance range $0 < A < 0.5$, despite the overall complex relation between concentration and intensity. An example of fluorescence quantum efficiency determination is shown in Figure 3.2.2 for eGFP in comparison to fluorescein¹. Here, the quantum efficiency was determined for eGFP in PBS in comparison to the quantum efficiency standard fluorescein in 0.1 M NaOH with a literature value $\Phi_{ref} = 0.95$ [84, 88]. We obtain absorption and integrated emission from five different concentrations and extracted the quantum efficiency of $\Phi_s = m_s/m_{ref} \cdot \Phi_{ref} = 0.63 \pm 0.02$ according to equation 3.2.3. The reported values and standard deviations result from three independent experiments. The choice of fluorescein as reference is due to its large spectral overlap with eGFP, which minimizes deviations due to errors in the detector sensitivity correction. Alternative reference dyes are quinine (emission at ~ 450 nm), rhodamine 6G (~ 550 nm) and rhodamine 101 (~ 600 nm) [84]. For red fluorophores, we use the less well established fluorophore Alexa Fluor 647 (with maleimide group), due to the lack of available standards.

For high precision quantum efficiency measurements, a deeper understanding of the instruments, the measurement principles and the sample is important to avoid or minimize various errors due to polarization effects, concentration instability and sample sticking, spectral errors, etc. as described in more detail in [89].

3.3 Lifetime and anisotropy measurements

The lifetime and time-dependent anisotropy of fluorophores are theoretically well described (see section 3.3.1) and are nowadays typically measured with high resolution via time-correlated single photon counting (TCSPC) systems [90].

3.3.1 Setup and measurement

Bulk lifetime and polarization decay measurements were performed here using a home-built setup (Figure 3.3.1A) as also described in reference [2] (Chapter 11). This setup was established experimentally as part of this thesis. In the measurement procedure, 400 μ l of sample was measured in a 10 mm (L) x 1.5 mm (H) cuvette at a concentration of 50-150 nM. The sample concentration was always tuned to obtain a ~ 50 kHz photon count rate to ensure high signal-to-noise ratio and avoiding to work in the non-linear range of the detector. On this setup, the samples are excited by a pulsed laser (LDH-P-FA-530B for green fluorophores/LDH-D-C-640 for red fluorophores with PDL 828 "Sepia II" controller, Picoquant). The excitation polarization is set with a $\lambda/2$ -waveplate (ACWP-450-650-10-2-R12 AR/AR, Laser Components) and a linear polarizer (glass polarizer #54-926, Edmund Optics). The emission light is polarization filtered (wire grid polarizer #34-315, Edmund Optics). The emission light is collected with a lens (AC254-100-A, Thorlabs) and scattering light or Raman contributions are blocked with filters (green: 532 LP Edge Basic & 596/83 BrightLine HC, AHF; red: 635 LP Edge Basic & 685/80

¹The presented results are an extract of fluorescence quantum efficiency measurements in "Characterization of fluorescent proteins with intramolecular photostabilization" [87]

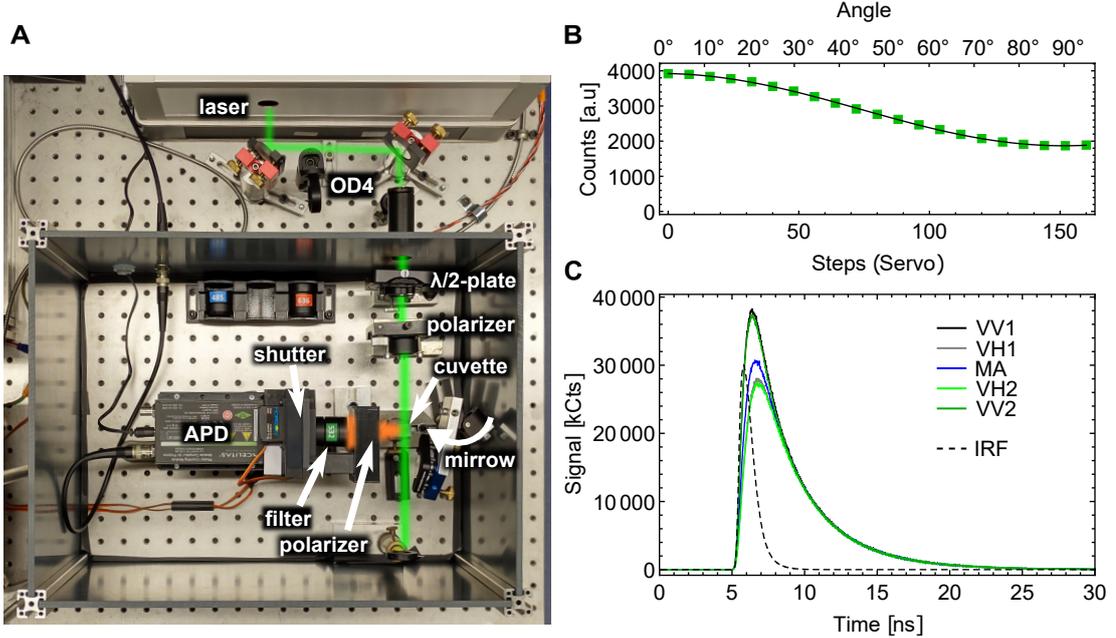


Figure 3.3.1: Lifetime and anisotropy measurements. **A** Picture of the home-build setup. The laser beam is guided from the top through the cuvette. The emission light is collected by an APD. For the IRF measurements, a flip mirror replaces the cuvette to directly guide the attenuated (OD=4) laser beam into the detector. **B** Calibration curve of the emission light, where the linear polarizer is rotated by a servo-motor. The measurement points (green) are fitted with a cosine-function to determine the 0° , 90° , and 54.7° positions. **C** Exemplary anisotropy measurement with one magic angle (MA) and two vertical (VV) / horizontal measurements (VH). The IRF is measured separately with the flip mirror.

ET Bandpass, AHF). The signal is recorded with an avalanche-photo-diode (SPCM-AQRH-34, Excelitas) and a TCSPC module (HydraHarp400, Picoquant). Polarization optics are mounted in home-built, 3D-printed rotation mounts and the APD is protected from light with a 3D-printed shutter unit. The angle settings of 0° , 90° , and 54.7° (magic angle) are calibrated with a cosine-fit to a measurement of the full rotation range (Figure 3.3.1B). In a typical experiment, the excitation power is set to $10 \mu\text{W}$ at a repetition rate of 20 MHz. For anisotropy and lifetime measurements, data sets are recorded for each polarization setting for 5 min in the order vertical (VV1), horizontal (VH1), magic angle (MA), horizontal (VH2), and vertical polarization (VV2) under vertical excitation. The anisotropy is calculated based on the sum of two vertical and horizontal measurements to compensate for small drifts in laser power or slow changes in fluorophore concentration due to sticking (see Figure 3.3.1C). With $VV(t) = VV1(t) + VV2(t)$ and $VH(t) = VH1(t) + VH2(t)$, we obtain the anisotropy decay as

$$r(t) = \frac{VV(t) - G \cdot VH(t)}{VV(t) + 2G \cdot VH(t)}, \quad (3.3.1)$$

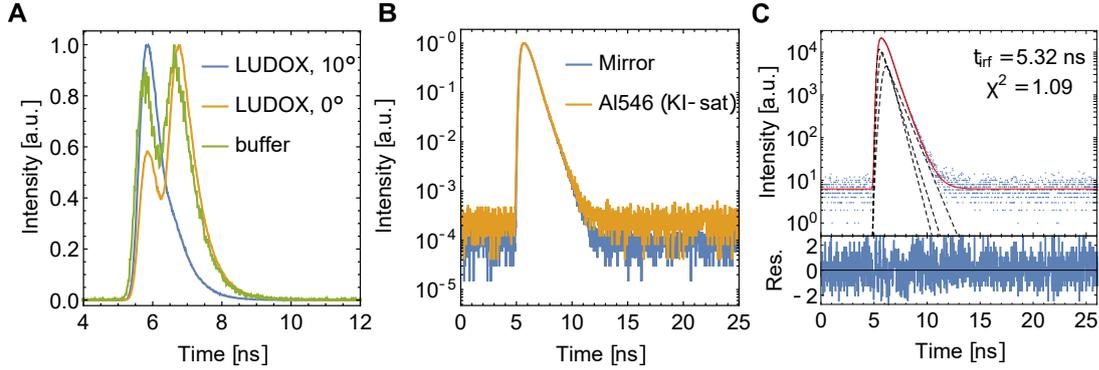


Figure 3.3.2: Instrument response function (IRF) measurement and fitting.

A Measured IRF profiles with LUDOX dispersion and two tilting angles of the linear polarizer (with respect to the light path, see Figure 3.3.1A) and scattering signal of the buffer with significantly lower intensity and smaller signal-to-noise ratio. **B** Measured IRF profiles with the flip-mirror and highly quenched Alexa Fluor 546 fluorophores in saturated KI solution. **C** Exemplary IRF profile fitted with three Gaussians reconvoluted with an exponential decay according to equation 3.3.4.

where G is the correction factor obtained by measuring with horizontal excitation $G = HV/HH$ (HV and HH is the total signal in the vertical or horizontal channel, respectively) [2].

3.3.2 IRF measurements

As described in section 2.1.2, the signal of excited fluorophores follows a single or multi-exponential deexcitation according to $I(t) = \sum_i \alpha_i e^{-t/\tau_i}$. In experiments, this theoretical decay is overlaid with the time profile of the excitation source, which deviates from an instantaneous excitation in practice. Further, delays in the detection and recording of the signal lead to a broadening of the measured intensity profile. This instrument-specific response is overlaid with the actual decay according to

$$I_{\text{measured}}(t) = \int_{-\infty}^t \text{IRF}(t') I(t - t') dt'. \quad (3.3.2)$$

Equation 3.3.2 is a general concept of a signal response based on an impulse, which can hardly be resolved solely based on the measured signal I_{measured} [91]. However, we can approximately determine the IRF with a response that is close to a delta-function ($I(t) \rightarrow \delta(t)$), which gives us

$$I_{\text{measured}}(t) \approx \int_{-\infty}^t \text{IRF}(t') \delta(t - t') dt' = \text{IRF}(t). \quad (3.3.3)$$

A common approach in fluorescent lifetime measurements is the use of scattering elements such as LUDOX (dispersion of colloidal silica) [92, 93]. However, the setup

shows high sensitivity to reflections of polarization optics, which changes the shape of the IRF significantly (see Figure 3.3.2A). Therefore, a flip-mirror as a sample replacement (and a neutral density filter with OD = 4) was used (see Figure 3.3.1A) to guide the laser directly into the detection path. This method improved the robustness for the measurement of the instrument response function. Alternatively, the response of a highly quenched fluorophore approximates the delta function, $I(t) = \sum_i \alpha_i e^{-t/\tau_i} \xrightarrow{\tau_i \rightarrow 0} \delta(t)$. For the green laser (532 nm), we found that Alexa Fluor 546 in saturated KI-solution shows the desired behaviour (Figure 3.3.2B). However, there was no adequate fluorophore candidate for the red laser (640 nm) available in this work.

Heating and cooling effects of the laser (operating at different powers) and, especially, the APD (operating at different count rates) were found to shift the instrument response and thus broaden the decay curves. Therefore, measurements of the IRF and different measurement conditions in the experiment were performed with fixed laser settings and comparable count-rates without turning the APD off and on between the measurements.

In order to reduce the noise of the reconvolution fit function for lifetime and anisotropy fits, I approximated the IRF by the sum of multiple IRF-components IRF_i consisting of a Gaussian ($e^{-(t-t_{irf})^2/(2\sigma^2)}$) reconvoluted with an exponential decay ($e^{-\lambda_i t}$), which reads

$$\text{IRF}_i(t) = A_i e^{-\lambda_i \left(t - (t_{irf} + (1 - \delta_{i1}) \Delta t_i) - \frac{\sigma_i^2 \lambda_i}{2} \right)} \text{erf} \left(-\frac{t - (t_{irf} + (1 - \delta_{i1}) \Delta t_i) - \frac{\sigma_i^2 \lambda_i}{2}}{\sqrt{2\sigma_i}} \right) \quad (3.3.4)$$

$$\text{IRF}(t) = \sum_{i=1}^{N_{irf}} \text{IRF}_i(t) + bk g_{\text{IRF}},$$

where the term $(1 - \delta_{i1}) \Delta t_i$ accounts for small shifts between the different summands and $bk g_{\text{IRF}}$ for the background.

Especially for the fit of anisotropy decays, I found that a small mismatch of IRF and measurement signal is well compensated and the fits yield more robust results (see section 3.3.5).

3.3.3 Lifetime reconvolution fitting

The typical signal in fluorescent lifetime measurements with polarized excitation light results in a superimposition of a polarization decay (see section 2.1.4) and a lifetime-dependent decay due to deexcitation. One can obtain a polarization-independent signal by a calculated intensity $I(t) = I_{VV}(t) + 2 \cdot G \cdot I_{HV}(t)$, which directly follows from the geometric consideration of isotropically oriented emitters excited by polarized light. Alternatively, measurements under magic angle conditions with an emission angle of $\Theta_{MA} = \cos^{-1}(\sqrt{1/3}) \approx 54.7^\circ$ result in the same polarization independent signal [2]. This signal can be described as a convolution of $\text{IRF}_0(t) = \text{IRF}(t) - bk g_{\text{IRF}}$ (background-free IRF) from equation 3.3.4 with a mono-exponential (N=1) or multi-exponential (N>1) fit according to:

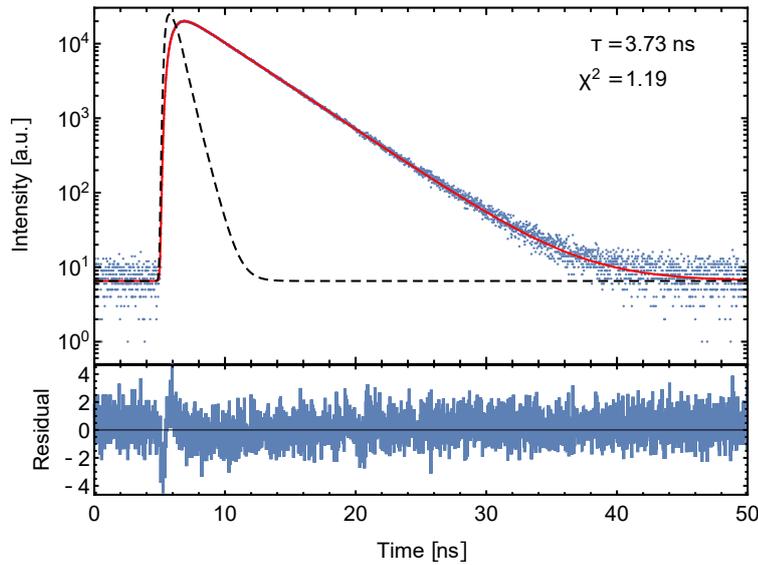


Figure 3.3.3: Lifetime data analysis. Fluorescent lifetime signal of Alexa Fluor 546 (blue, dots). The fitted IRF (black, dashed) is reconvoluted with an exponential decay and fitted to the signal (red). The residual of the fit (bottom) shows statistically limited de-excitation profiles with a minor mismatch at the signal raise at around 5 ns.

$$I(t) = \sum_{d=1}^N I_d \cdot \left(\text{IRF}_0 * e^{-\lambda_d t} \right) (t) + bkg. \quad (3.3.5)$$

A typical approach is a fully numeric reconvolution fit, where the measured IRF enters directly into equation 3.3.5 and the signal is reconstructed via reconvolution fitting [91]. The advantage of the analytic formula from equation 3.3.4 is that the signal of the lifetime decay can be fully described by an analytic equation. Figure 3.3.3 shows an exemplary mono-exponential fit ($N=1$) to the lifetime decay of Alexa Fluor 546 in aqueous solution.

The residuals are calculated based on the statistical uncertainty of photon counting, which follows a Poissonian distribution (eq. 2.3.4) with variance $\sigma_i^2 = n_i$, where n_i are the counts in each bin of the decay curve. The plotted residual is computed as $(n_i - f_i)^2 / \max(n_i, 1)$. Here, f_i are the fitted values of equation 3.3.5. Based on this, one obtains a reduced Chi-square of

$$\chi^2 = \frac{1}{N - p} \sum_{i=1}^N \frac{(n_i - f_i)^2}{\max(n_i, 1)}, \quad (3.3.6)$$

where N is the number of data points, p the number of fit parameters.

The lifetime decay of a fluorophore often deviates from the ideal case of a mono-exponential decay of the fluorescence due to a heterogeneous sample which different lifetimes due to a varying environment of the fluorophore. Section 2.1.5 describes typical

effects of quenching or fluorescence enhancement that changes the lifetime of a fluorophore. An example of multi-exponential fluorescence lifetime measurements is shown in Figure 3.3.4 for the DNA-binding fluorophore SYBR Gold at varying fluorophore and DNA concentrations. SYBR Gold is an unsymmetrical cyanine, which has a low quantum efficiency and short lifetime when it is free in solution due to very fast cis-trans relaxation [94–97]. If the fluorophore intercalates with DNA, this non-radiative relaxation is suppressed².

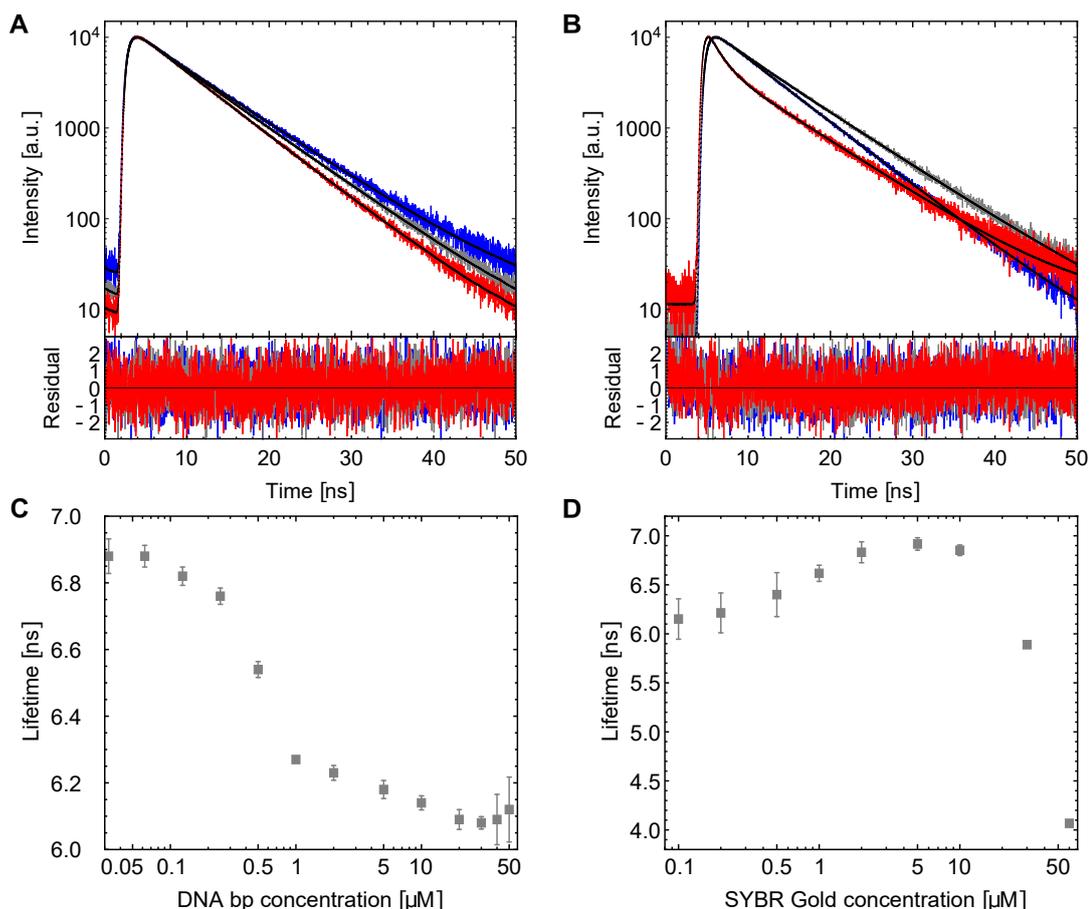


Figure 3.3.4: Multi-exponential lifetime fits. **A** Fluorescence lifetime measurements for 0.0325 (blue), 0.5 (grey) and 30 (red) μM · basepair (bp) DNA at a SYBR-Gold concentration of 2.5 μM fitted with a mono-exponential decay. **B** Fluorescence lifetime measurements for 0.2 μM (blue), 10.4 μM (gray) and 124 μM (red) SYBR Gold in the presence of a constant DNA concentration of 2 μM · bp DNA fitted with a bi-exponential decay. **C** Fluorescence lifetimes as a function of DNA concentration determined from single-exponential fits as in A. **D** Amplitude-weighted fluorescence lifetime as a function of SYBR Gold concentration from measurements at 2 μM · bp DNA determined from two exponential fits.

²The presented results are an extract of lifetime measurements in "Molecular structure, DNA binding mode, photophysical properties and recommendations for use of SYBR Gold" [98].

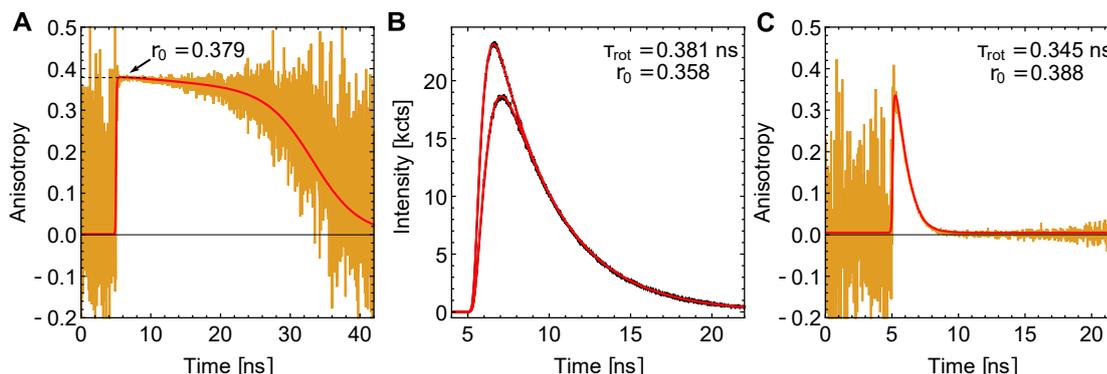


Figure 3.3.5: Time-resolved anisotropy. **A** Time-resolved anisotropy decay of Alexa Fluor 546 in 98% glycerol. The rotational relaxation is almost completely hindered and the anisotropy value approaches asymptotically the fundamental anisotropy at short times (~ 5 ns). **B** Intensity curves of vertical and horizontal polarizations fitted with the functions in equation 3.3.7. **C** Time-resolved anisotropy calculated from B based on equation 3.3.1.

Depending on the packing density, the lifetime slightly changes, since the fluorophore is assumed to be sterically stabilized, the more fluorophores are intercalated in the DNA (see Figure 3.3.4A/C). On the other hand, the fluorophores can quench each other via collisional quenching at high fluorophore densities. The described effects are very sensitive to the actual binding and the local environment of a fluorophore. Therefore, the decay cannot be described by a single exponential anymore (see Figure 3.3.4C/D).

3.3.4 Fundamental anisotropy

The fundamental anisotropy varies for different fluorophores mostly in the range from 0.3-0.4 due to different angles between absorption and emission dipole moment (see section 2.1.4). It can directly be measured in solvents with high viscosity such as propylene glycol or glycerol (preferably at low temperatures), where molecular rotation is hindered. In this thesis, high glycerol concentrations are used to extract the fundamental anisotropy from the time-resolved anisotropy decays (see Figure 3.3.5A). The decrease in anisotropy is mainly due to the decrease in signal and a higher contribution of the background, as the fit with a rotational correlation time >100 ns also indicates (see next section for details). For Alexa Fluor 546, a fundamental anisotropy of $r_0 = 0.38$ was obtained. This corresponds to an angle β between excitation and emission dipole moment of $\beta = 10.5^\circ$.

3.3.5 Anisotropy reconvolution fitting

The vertical and horizontal intensity decay of a fluorophore including a mono-exponential depolarization according to equation 2.1.14 reads

$$\begin{aligned}
 I_{VV}(t) &= \sum_{d=1}^N I_d \cdot \left(IRF_0 * \left(e^{-\lambda_d t} (1 + r(t)) \right) \right) (t) + bkg \\
 I_{VH}(t) &= \sum_{d=1}^N I_d \cdot \left(IRF_0 * \left(e^{-\lambda_d t} (1 - 2r(t)) \right) \right) (t) + bkg
 \end{aligned}
 \tag{3.3.7}$$

with $r(t) = r_0 e^{-t/\theta}$ for the generic multi-exponential fluorescence decay [2]. A global fit of vertical and horizontal decay with this function yields the lifetime components λ_d , the fundamental anisotropy r_0 and the rotational correlation time θ .

Alternatively, the calculated time-resolved anisotropy (according to equation 3.3.1) can be fitted with an analytic model for the anisotropy decay (see Figure 3.3.5C), where the lifetime is fixed based on an initial fit of the magic-angle condition (see section 3.3.3). In principle, both methods should be equivalent. However, the fitting of the anisotropy decay provides better results for the analysis in this thesis, since it tends to be more robust against smaller mismatches of the IRF and the decay curves (see Figure 3.3.3) and for multi-exponential rotation times close to the fluorescence lifetime (see next section).

3.3.6 Protein relaxation

The previous section shows that a fluorophore in solution possesses a rather simple mono-exponential anisotropy decay. For fluorophores attached to a protein, the anisotropy becomes more complex due to multiple rotational modes (see Figure 3.3.6A). The smaller fluorophore exhibits a fast rotation θ_{fl} of around ~ 1 ns within the accessible volume, which is slower but still on the same order as a free fluorophore (see previous section). This rotation is superimposed with the rotation of the protein, which can be approximated by an ellipsoid. Although the rotation of an ellipsoidal protein is expected to be multi-exponential according to equation 2.1.17, this can hardly be resolved in practice. More details on the complex rotation relaxation for ellipsoidal structures can be found in references [99, 100].

A valid simplification is the splitting of the anisotropy decay into one fast rotation of the fluorophore θ_{fl} and one slow rotation of the protein θ_p according to

$$r(t) = r_0 e^{-t/\theta_p} \left(\alpha e^{-t/\theta_{fl}} + (1 - \alpha) \right),
 \tag{3.3.8}$$

where α determines how free the fluorophore is to move. Here, $\alpha = 1$ means that the fluorophore is completely free to move and $\alpha = 0$ means that the fluorophore completely sticks to the protein. Due to the different rotation times around the principle axes θ_x , θ_y , and θ_z , the overall protein relaxation in the measurement depends on the actual labeling position (see Figure 3.3.6D). At the tips of the protein (z-coordinate large), the slow rotations θ_x and θ_y dominate the protein relaxation (see Figure 3.3.6A/B) in contrast to the central residues (z-coordinate approximately 0), while the fast rotation θ_z reduces the overall protein relaxation time. Using the theory on rotational correlation time, one can estimate the expected values according to equation 2.1.15 considering the ellipsoidal form and a hydration shell around the protein:

$$\theta_x = 15.9 \text{ ns } (h = 0) \mid 19.9 \text{ ns } (h = 0.2 \text{ ml/g})$$

$$\theta_y = 14.0 \text{ ns } (h = 0) \mid 17.8 \text{ ns } (h = 0.2 \text{ ml/g})$$

$$\theta_z = 10.2 \text{ ns } (h = 0) \mid 12.9 \text{ ns } (h = 0.2 \text{ ml/g})$$

The hydration parameter h accounts for water molecules at the protein surface, which interact electrostatically with polar and charged amino acids. The determined values for the rotational correlation time with $h = 0.2 \text{ ml/g}$ are comparable to the measured values of 15-25 ns, although, the estimation neglects the additional friction of the attached fluorophore (see Figure 3.3.6D).

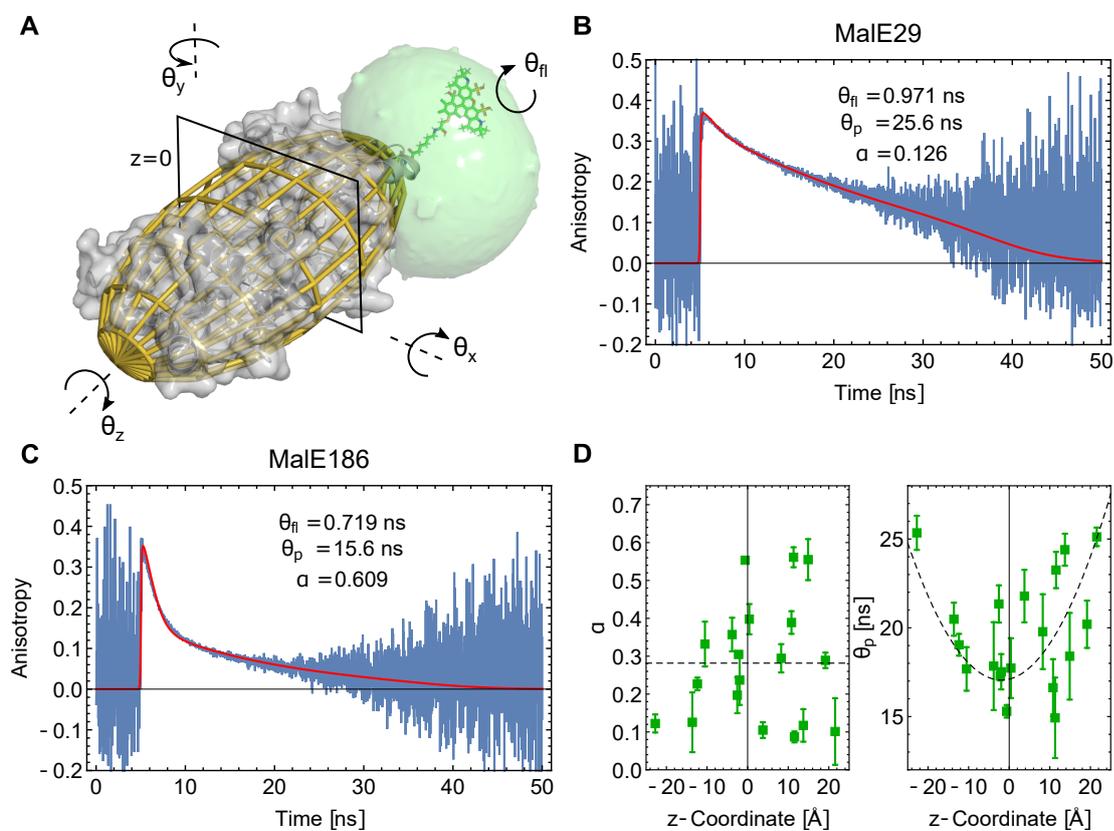


Figure 3.3.6: Protein anisotropy decay. **A** Protein structure of MalE (pdb: 1omp, gray) fitted with an ellipsoid (yellow) and the fluorophore attached at residue 29 (green structure). The fluorophore can move and rotate within the accessible volume (pale green). **B** Anisotropy decay of Alexa Fluor 546 labeled at residue 29. **C** Anisotropy decay of Alexa Fluor 546 labeled at residue 186. **D** Fraction of free fluorophore rotation α plotted against the projection of the $C\alpha$ -atom to the z -axis (left) does not show a clear correlation. The protein relaxation time θ_p plotted against the projection of the $C\alpha$ -atom to the z -axis (right) shows clear tendency of larger relaxation times at the outer positions indicated by a parabola fit (black, dashed).

3.4 ALEX and PIE measurements

3.4.1 Setup and measurement

In this work, single-molecule FRET experiments were carried out on two types of solution based confocal microscopy systems that were set up as part of the work. One is an alternating-laser excitation microscopy (ALEX) setup as sketched in Figure 3.4.1, where the two excitation lasers are alternately switched on and off in cw-mode to excite donor and acceptor molecules separately [14, 15].

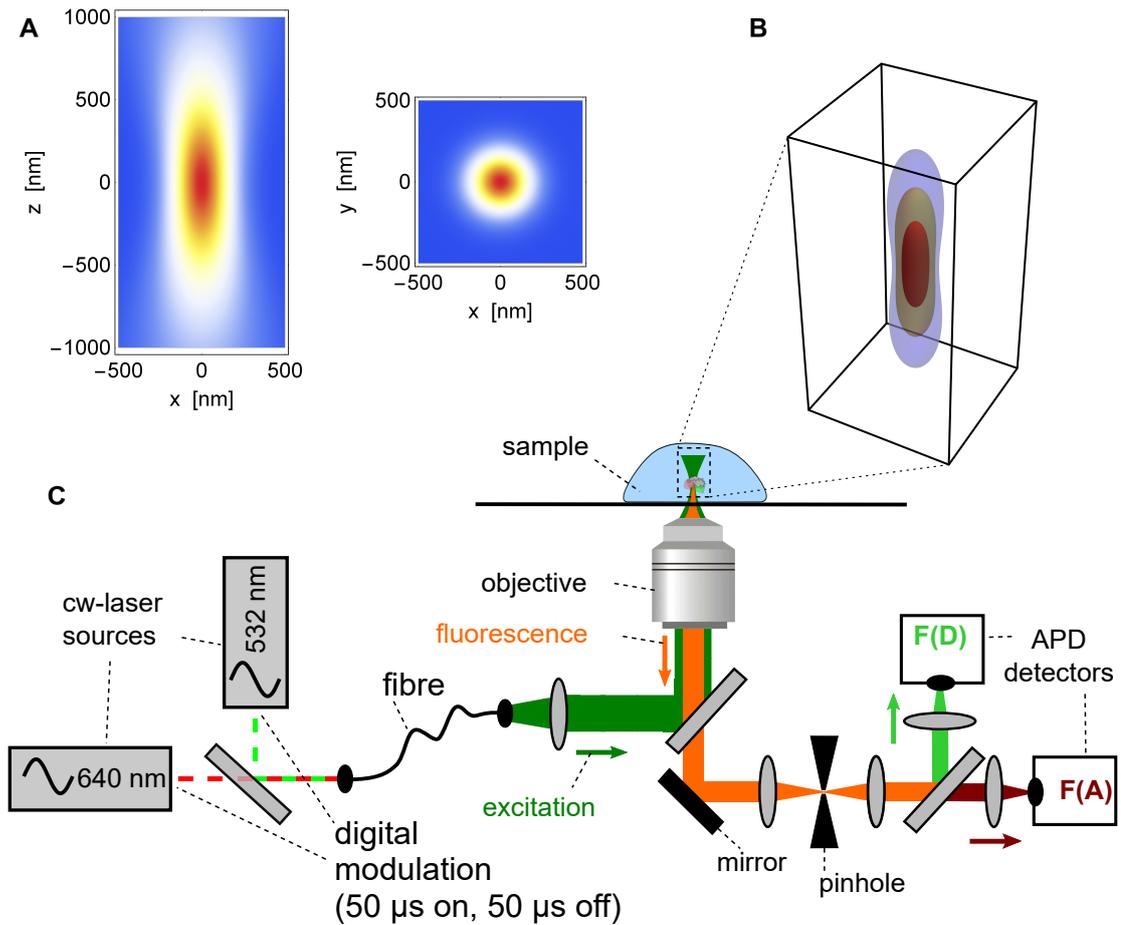


Figure 3.4.1: ALEX microscope principle. **A** Intensity profile of a focused Gaussian beam in xz - (left) and xy -direction (right). **B** 3D focal volume with contour lines at $1/2$, $1/4$, and $1/e^2$ of the peak intensity. **C** Scheme of an ALEX setup with two alternating laser sources, which are guided into the objective and focused in the solution above the coverslip. The emitted light is collected with the same objective and separated from the excitation light with a dichroic mirror. The emission light is spatially filtered with a pinhole and spectrally split before it is focused onto an APD chip.

The second setup is a pulsed-interleaved excitation (PIE) microscope, which imple-

ments the alternation of the two excitation lasers with pulsed lasers on a nanosecond timescale (typical alternation and pulse intervals ~ 25 -100 ns). A synchronized detection with a TCSPC module allows to determine the lifetime of each individual burst [16].

ALEX setup Single-molecule FRET experiments with ALEX were performed on a home-built confocal microscope as described previously. On this setup, the fluorescent donor molecules are excited by a diode laser OBIS 532-100-LS (Coherent, USA) at 532 nm operated at 60 μ W at the sample in alternation mode. The fluorescent acceptor molecules are excited by a diode laser OBIS 640-100-LX (Coherent, USA) at 640 nm operated at 25 μ W at the sample in alternation mode (100 μ s alternation period). The lasers are coupled into a polarization maintaining single-mode fiber P3-488PM-FC-2 (Thorlabs, USA). The laser light is guided into the epi-illuminated confocal microscope (Olympus IX71, Hamburg, Germany) by a dual-edge beamsplitter ZT532/640rpc (Chroma/AHF) focused by a water immersion objective (UPlanSApo 60x/1.2w, Olympus Hamburg, Germany). The emitted fluorescence is collected through the objective and spatially filtered using a pinhole with 50 μ m diameter and spectrally split into donor and acceptor channel by a single-edge dichroic mirror H643 LPXR (AHF). Fluorescence emission is filtered (donor: BrightLine HC 582/75 (Semrock/AHF), acceptor: Longpass 647 LP Edge Basic (Semrock/AHF) and focused on avalanche photodiodes (SPCM-AQRH-64, Excelitas). The detector outputs were recorded by a NI-Card PCI-6602 (National Instruments, USA). Data analysis was performed using home written software package as described in references [101, 102].

Single-molecule events were identified using an APBS algorithm with a threshold of 15, a time window of 500 μ s and a minimum total photon number of typically 150 (number is varying for different experiments) or a DCBS algorithm with a thresholds of 15/15 for both excitation time windows (green/red) separately (other parameters as in APBS).

PIE setup PIE-FRET experiments were carried out on a home-built confocal microscope. On this setup, the samples are excited by two pulsed lasers (LDH-P-FA-530B at 32 μ W, LDH-D-C-640 at 20 μ W) with the PDL 828 “Sepia II” controller, Picoquant, GER) at a repetition rate of 20 MHz. The laser-pulses are altered on the nanosecond timescale by a multichannel picosecond diode laser driver (PDL 828 “Sepia II”, PicoQuant GmbH) with an oscillator module (SOM 828, PicoQuant GmbH). The lasers were coupled into a single mode fiber (P3-488PM-FC, Thorlabs GmbH) to obtain a Gaussian beam profile. The laser light is guided into a confocal microscope (Olympus IX73, Hamburg, Germany) by a dual-edge beamsplitter ZT532/640rpc (Chroma/AHF, Germany) and focused to a diffraction-limited excitation spot by an oil immersion objective (UPLSAPO 60XO, Olympus Hamburg, Germany). The emitted light is collected through the same objective, spatially filtered using a pinhole with 50 μ m diameter and spectrally split into green and red channel by a single-edge dichroic mirror H643 LPXR (AHF). The emission is filtered (donor: BrightLine HC 582/75 (Semrock/AHF), acceptor: Longpass 647 LP Edge Basic (Semrock/AHF)) and the signal is recorded with avalanche-photodiodes (SPCM-AQRH-34, Excelitas) and a TCSPC module (HydraHarp400, Picoquant). The setup

was controlled by a commercial software package (SymPhoTime64, Picoquant GmbH).

Data analysis is performed using the PAM software package as described in reference [103].

Single-molecule events are identified using an APBS algorithm with a threshold of 5, a time window of 500 μs and a minimum total photon number of 100 or a DCBS algorithm with a thresholds of 5/5 for both excitation time windows (green/red) separately (other parameters as in APBS).

Data extraction and stoichiometry The samples were measured in ALEX and PIE mode at approximately 25-50 pM protein concentrations on a confocal microscope as sketched in Figure 3.4.1. The photon counts extracted from the result of an all photon burst search (APBS) and dual channel burst search (DCBS) algorithm are separated into the excitation (donor excitation or acceptor excitation) time windows to calculate three intensity numbers for every burst:

$$\begin{aligned} \text{Donor excitation, donor emission:} & I_{DD} \\ \text{Donor excitation, acceptor emission (FRET signal):} & I_{DA} \\ \text{Acceptor excitation, acceptor emission:} & I_{AA} \end{aligned}$$

The fourth option of donor emission after acceptor excitation I_{AD} is neglected, since it only contains background noise. While FRET solely relies on the donor excitation, the acceptor excitation is a readout for the stoichiometry (relative brightness) of donor and acceptor fluorophores. Therefore, we define the stoichiometry as

$$S := \frac{I_{DD} + I_{DA}}{I_{DD} + I_{DA} + I_{AA}}. \quad (3.4.1)$$

Molecules with donor only or acceptor only exhibit S-values of $S > 0.95$ and $S < 0.2$, respectively. Therefore, this number can be used for filtering purposes.

Data evaluation Since the total photon numbers of a burst are in the range of 50 to 500, which are split into the different channels, the FRET efficiency and the stoichiometry distributions show highly non-monotonic behavior due to the combinatorics (see Nir et al. [58]). The data is typically binned into 61 to 101 bins in the range between 0 and 1 for the FRET efficiency and stoichiometry to obtain a more monotonic distribution, which can be approximated by a 1D-Gaussian. The description of the two-dimensional ES-histogram considers a potential correlation of the efficiency and stoichiometry (e.g. due to slight confocal misalignment) and reads as

$$Gauss(E, S) = A \cdot e^{-\frac{1}{2(1-\rho^2)} \left(\frac{(E-E_0)^2}{\sigma_E^2} + \frac{(S-S_0)^2}{\sigma_S^2} - \frac{2(E-E_0)(S-S_0)}{\sigma_E \sigma_S} \right)}, \quad (3.4.2)$$

where the parameter $\rho = corr(E, S)$ is the correlation of E and S .

3.4.2 Binding kinetics

Single-molecule FRET experiments allow to monitor conformational changes [12, 104, 105] and ligand binding [102, 106, 107] when the fluorescent labels change distance within this process (see subsequent chapter). Figure 3.4.2 shows such a titration experiment monitoring the ligand binding of VcSiap, where the distance decreases upon ligand binding and the FRET efficiency increases³.

The dissociation constant K_D can be determined from the concentration of bound c_c and unbound proteins c_o by fitting the share of bound proteins $r_c = c_c/(c_o + c_c)$ according to

$$r_c = \frac{\left(1 + \frac{K_D}{P_0} + \frac{L}{P_0}\right) - \sqrt{\left(1 + \frac{K_D}{P_0} + \frac{L}{P_0}\right)^2 - 4\frac{L}{P_0}}}{2}, \quad (3.4.3)$$

where $P_0 = c_o + c_c$ is the total protein concentration, L the ligand concentration [109]. For many measurements at low concentrations, e.g. for single-molecule measurements at picomolar concentrations with $P_0 \ll K_D$, equation 3.4.3 can be simplified to

$$r_c = L/(K_D + L). \quad (3.4.4)$$

In smFRET experiments, the ratio of both states can be well derived from a fit of a summed Gaussian to the counts in the ES-histogram $N(E, S)$ as

$$N(E, S) = \text{Gauss}_o(E, S) + \text{Gauss}_c(E, S) \quad (3.4.5)$$

by fixing the parameters $E_o, S_o, \sigma_E, \sigma_S$ to the apo (opened, no ligand) and holo state (closed, 200 μM Neu5Ax) values (derived from a single Gaussian fit) and only keep the amplitudes as free fit parameters to get $r_c = A_c \sigma_{E,c} \sigma_{S,c} / (A_o \sigma_{E,o} \sigma_{S,o} + A_c \sigma_{E,c} \sigma_{S,c})$. Figure 3.4.2C shows a plot of the ratio of closed molecules against the ligand concentration and a fit to the simplified binding curve from equation 3.4.4.

³The presented results are an extract of ALEX measurements in "Triggering Closure of a Sialic Acid TRAP Transporter Substrate Binding Protein through Binding of Natural or Artificial Substrates" [108].

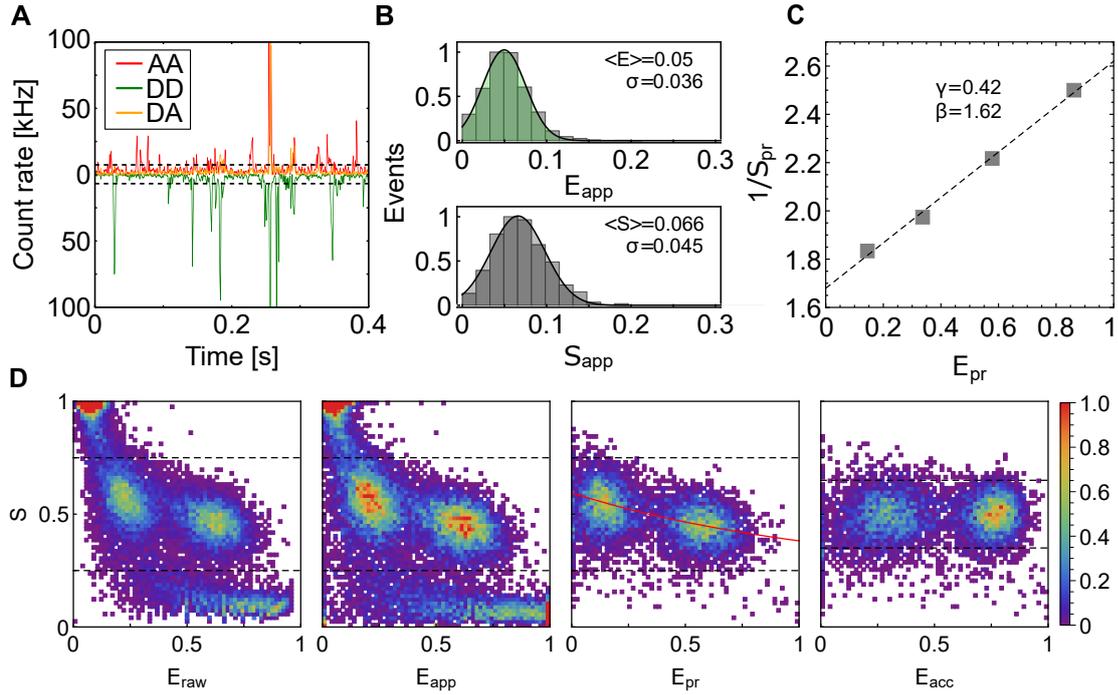


Figure 3.4.3: Intensity-based FRET efficiency corrections. **A** Background is determined based on the average inter-burst photon rates. All three channels are averaged over the times where the photon rate is below 10 kHz (black, dashed). **B** Exemplary E-histograms and S-histograms of donor only (top, $S > 0.95$) and acceptor only (bottom, $E > 0.8$ & $S < 0.25$) populations are fitted with a Gaussian to determine $\langle E_{app}^{(DO)} \rangle$ and $\langle S_{app}^{(AO)} \rangle$, respectively. **C** Linear fit to the 2D-Gaussian mean positions of the inverse stoichiometry value ($1/S$) to the FRET efficiency E (after cross-talk corrections) of DNA with fluorophore distance of 8, 13, 18, and 23 basepairs. The resulting γ and β values are determined according to equation 3.4.10. **D** Exemplary ES-histograms for 13 and 23 basepair distances (joined data sets) at every correction step (from left to right): without correction (APBS), after background correction (APBS), after crosstalk correction (DCBS), and after all corrections (DCBS).

Background correction. Background is calculated by averaging over the signal of all three photon traces, where the total photon count is less than 10 cts/ms or from a background measurement of the buffer (see Figure 3.4.3A). Alternatively, it can be derived from pure buffer measurements or from a fit to the interphoton delay times [59]. The three background rates $I_{DD}^{(BG)}$, $I_{DA}^{(BG)}$, and $I_{DA}^{(BG)}$ are subtracted from all three burst photon numbers ${}^i I_{DD}$, ${}^i I_{DA}$, and ${}^i I_{AA}$ of every burst obtained from the DCBS burst search algorithm:

$$\begin{aligned}
 {}^{ii}I_{DD} &= {}^iI_{DD} - \tau I_{DD}^{(BG)} \\
 {}^{ii}I_{DA} &= {}^iI_{DA} - \tau I_{DA}^{(BG)} , \\
 {}^{ii}I_{AA} &= {}^iI_{AA} - \tau I_{AA}^{(BG)}
 \end{aligned} \tag{3.4.6}$$

where τ is the burst duration. We obtain a background-corrected apparent FRET efficiency E_{app} and apparent stoichiometry S_{app} as:

$$\begin{aligned}
 E_{app} &= \frac{{}^{ii}I_{DA}}{{}^{ii}I_{DA} + {}^{ii}I_{DD}} \\
 S_{app} &= \frac{{}^{ii}I_{DA} + {}^{ii}I_{DD}}{{}^{ii}I_{DA} + {}^{ii}I_{DD} + {}^{ii}I_{AA}}
 \end{aligned} \tag{3.4.7}$$

Leakage and direct excitation. Leakage and direct excitation are calculated by fitting a 1-dimensional Gaussian to the data points in E with $S > 0.95$ and in S with $E > 0.8$ and $S < 0.25$ (APBS, after background correction) to determine the apparent FRET efficiencies $E_{app}^{(DO)}$ and apparent stoichiometries $S_{app}^{(AO)}$ of the donor only (DO) and acceptor only (AO) populations, respectively (see Figure 3.4.3B). With the correction factors $\alpha = \langle E_{app}^{(DO)} \rangle / (1 - \langle E_{app}^{(DO)} \rangle)$ and $\delta = \langle S_{app}^{(DO)} \rangle / (1 - \langle S_{app}^{(DO)} \rangle)$, the photon numbers coming from acceptor emission upon donor excitation are corrected for leakage and direct excitation according to

$${}^{iii}I_{DA} = {}^{ii}I_{DA} - \alpha {}^{ii}I_{DD} - \delta {}^{ii}I_{AA}. \tag{3.4.8}$$

From these photon numbers, proximity FRET efficiency and stoichiometry is obtained by

$$\begin{aligned}
 E_{pr} &= \frac{{}^{iii}I_{DA}}{{}^{iii}I_{DA} + {}^{ii}I_{DD}} \\
 S_{pr} &= \frac{{}^{iii}I_{DA} + {}^{ii}I_{DD}}{{}^{iii}I_{DA} + {}^{ii}I_{DD} + {}^{ii}I_{AA}}
 \end{aligned} \tag{3.4.9}$$

Global gamma correction. A line $ax + b$ was fitted to the reciprocal proximity stoichiometry $1/S_{pr}$ in relation to the proximity FRET efficiency E_{pr} for all measurements (see Figure 3.4.3C). Gamma and beta correction factors were determined from the linear fit as

$$\begin{aligned}
 \gamma &= \frac{a-1}{a+b-1} \\
 \beta &= a + b - 1
 \end{aligned} \tag{3.4.10}$$

From the correction factors, the fully corrected FRET efficiency E and stoichiometry S were calculated as

$$E = \frac{E_{pr}}{\gamma + E_{pr} - \gamma E_{pr}}$$

$$S = \frac{1 + \left(\frac{1}{E_{pr}} - 1\right)}{1 + \gamma \left(\frac{1}{E_{pr}} - 1\right) + \frac{1}{\beta E_{pr}} \left(\frac{1}{S_{pr}} - 1\right)} \quad (3.4.11)$$

The effect of every correction step and change in FRET efficiencies is illustrated for a merged dataset of DNA labeled with Alexa Fluor 546 and Alexa Fluor 647 at 13 and 23 basepair separation in Figure 3.4.3D.

3.4.4 Dynamics evaluations

For the detection of molecular dynamics with FRET, two methods are applied in this work. While the BVA method can be applied on every single-molecule data from ALEX or PIE in diffusion-based as well as immobilized samples, the lifetime-based E-tau analysis is only applied on data from pulsed excitation at the PIE setup.

Burst-variance analysis. As described in chapter 2, an analysis of basic photon statistics gives insights on dynamics or heterogeneity of the sample and artifacts of the measurements. For samples from the ALEX and PIE setup, the BVA analysis was applied with a binning of $n=5$, which defines the static line for a purely shot-noise limited standard deviation $\sigma = \sqrt{E^*(1 - E^*)/5}$.

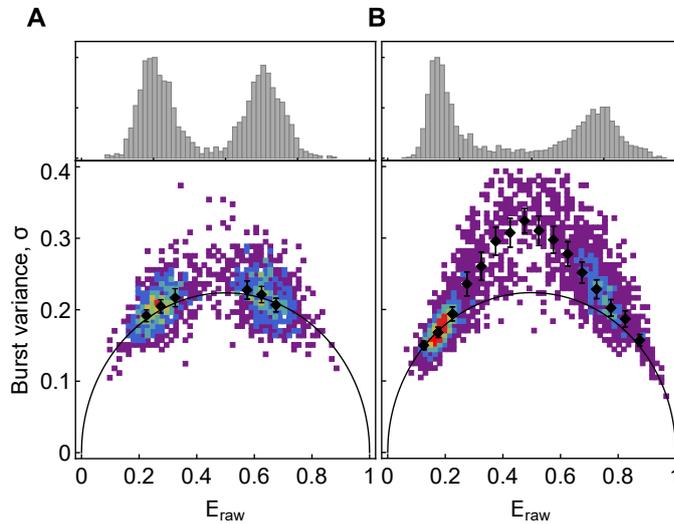


Figure 3.4.4: Burst variance analysis. **A** Mixture of DNA labeled with Alexa Fluor 546 and Alexa Fluor 647 separated by 13 and 23 basepairs. **B** DNA hairpin measured in PBS buffer with 100 mM NaCl switching between a low and high FRET state as described in reference [59, 112].

Figure 3.4.4A shows a mixture of labeled DNA double-strands with different fluorophore separations. Since the DNA is considered to be static on length scales below the persistence length, the BVA analysis reveals two separated distributions centered on the static line. Figure 3.4.4B shows a DNA hairpin, which interchanges between two states with close and far fluorophore distances. We see that most molecules stay in one of the two states between the diffusion time forming two populations on the static line. However, a minor share changes from one state to the other forming the dynamic bridge, which clearly deviates from the static line.

Efficiency-lifetime analysis. The lifetime information in PIE experiments allows an alternative of dynamics, which is not available for the cw-excitation on ALEX setups. Assuming a sample which interchanges between two FRET states E_1 and E_2 , one can derive an average FRET efficiency $\langle E \rangle$ from basic theory (see section 2.2), which reads as

$$\langle E \rangle = t_1 E_1 + t_2 E_2, \quad (3.4.12)$$

where $t_1 = T_1/T$ and $t_2 = T_2/T = 1 - t_1$ are the relative times spend in state one (T_1) and two (T_2) during the measurement time $T = T_1 + T_2$.

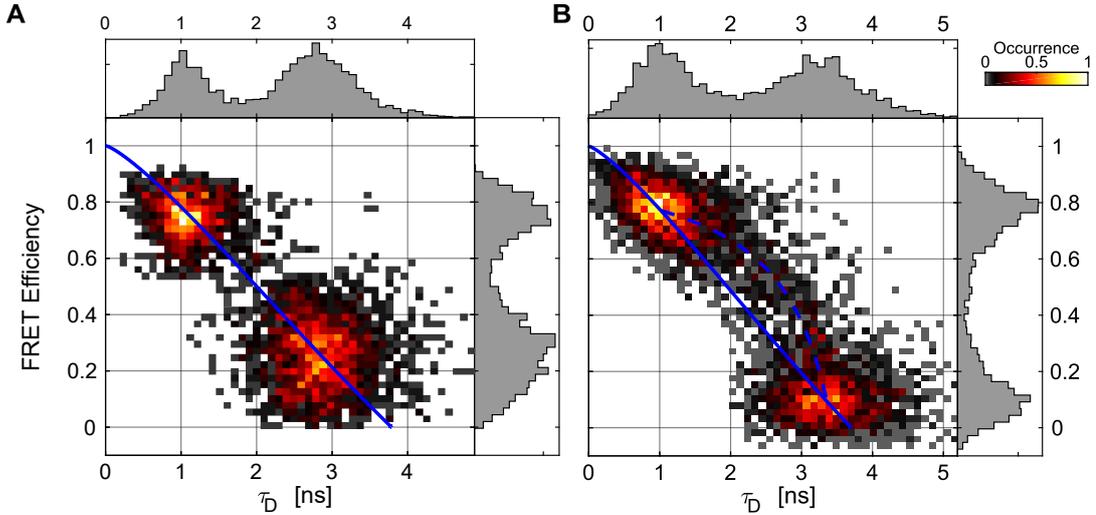


Figure 3.4.5: E-tau dynamics analysis. **A** E- τ -plot of a mixture of DNA labeled with Alexa Fluor 546 and Alexa Fluor 647 separated by 13 and 23 basepairs. **B** E- τ -plot of DNA hairpin measured in PBS buffer with 100 mM NaCl as in Figure 3.4.4.

While the interchange between two distances (and FRET efficiency states) during the detection time leads to a linear mixing of the mean FRET efficiency $\langle E \rangle$, the corresponding average lifetime $\tau_{D(A)}$ shows a non-linear behavior with respect to t_1 due to a FRET efficiency dependent photon count rate. From the average FRET efficiency

and lifetime, one obtains the relation as the so-called dynamic-FRET line, which is given by

$$E_{dyn} = 1 - \frac{\tau_1 \tau_2}{\tau_{fl}(\tau_1 + \tau_2 - \tau_{D(A)})} \quad (3.4.13)$$

as derived in reference [64]. The free donor lifetime τ_{fl} is obtained from a lifetime fit to the donor-only population. Figure 3.4.5 shows the E- τ -plot of a fully static DNA sample (left, similar to Figure 3.4.4A), which shows two populations centered on the static FRET line (solid line) according to the linear relation from equation 2.2.9 (corrected for linker dynamics at very high FRET efficiencies [64]). The DNA hairpin sample (similar sample as in Figure 3.4.4B) shows a bridging between the two states following the non-linear dynamic FRET line from equation 3.4.13 (dashed line). In principle, the dynamics can be observed on significantly faster timescales compared to BVA down to the order of ~ 1 μ s. However, a robust deviation from the static line can only be observed for well-separated FRET states and a proper correction of the intensity-based FRET efficiency (see chapter 6 for more details).

4 | Molecular and spectroscopic characterization of green and red cyanine fluorophores from the Alexa Fluor and AF series

The use of fluorescence techniques has an enormous impact on various research fields including imaging, biochemical assays, DNA-sequencing and medical technologies. This has been facilitated by the development of numerous commercial dyes with optimized photophysical and chemical properties. Often, however, information about the chemical structures of dyes and the attached linkers used for bioconjugation remain a well-kept secret. This can lead to problems for research applications where knowledge of the dye structure is necessary to predict or understand (unwanted) dye-target interactions, or to establish structural models of the dye-target complex. Using a combination of optical spectroscopy, mass spectrometry, NMR spectroscopy and molecular dynamics simulations, we here investigate the molecular structures and spectroscopic properties of dyes from the Alexa Fluor (Alexa Fluor 555 and 647) and AF series (AF555, AF647, AFD647). Based on available data and published structures of the AF and Cy dyes, we propose a structure for Alexa Fluor 555 and refine that of AF555. We also resolve conflicting reports on the linker composition of Alexa Fluor 647 maleimide. We also conducted a comprehensive comparison between Alexa Fluor and AF dyes by continuous-wave absorption and emission spectroscopy, quantum yield determination, fluorescence lifetime and anisotropy spectroscopy of free and protein-attached dyes. All these data support the idea that Alexa Fluor and AF dyes have a cyanine core and are a derivative of Cy3 and Cy5. In addition, we compared Alexa Fluor 555 and Alexa Fluor 647 to their structural homologs AF555 and AF(D)647 in single-molecule FRET applications. Both pairs showed excellent performance in solution-based smFRET experiments using alternating laser excitation. Minor differences in apparent dye-protein interactions were investigated by molecular dynamics (MD) simulations. Our findings clearly demonstrate that the AF-fluorophores are an attractive alternative to Alexa- and Cy-dyes in smFRET studies or other fluorescence applications.¹

¹This chapter was reproduced from Gebhardt *et al.* [113] with the permission of Wiley Online Library. Copyright © 2021 Wiley Online Library. For details of the individual contribution see the “Author Contributions” statement in section 4.5.

4.1 Introduction

The exploitation of fluorescence techniques has impacted various research fields and specific applications such as optical imaging, biochemical assays, DNA-sequencing, and medical technologies. The molecular contrast agents, i.e., the light absorbing and emitting molecules used, and their properties govern the success of these applications (for instance in PCR-based amplification of disease-related genomes [114]) and the information depth of state-of-the-art techniques in specialized research fields such as single-molecule [24] and super-resolution microscopy [33, 115–120]. Whereas fluorescent proteins are more frequently used in live-cell applications, in most other settings, where high photostability and tailored functional properties are required [24, 121], synthetic organic fluorophores dominate.

The common molecular scaffolds of modern synthetic organic fluorophores are fluoresceins, rhodamines, carbon- and silicon-pyrone, rylenes, bodipys, and cyanines [121]. They all feature intense absorption and emission in the visible spectrum [121]. Years of structural optimization has resulted in commercially available compounds with favorable photophysical properties and reactive linkers for flexible bioconjugation. The general structural design of such commercial fluorophores aims at high absorbance cross sections, high fluorescence quantum yields, and low rates for internal conversion and intersystem crossing e.g., achieved by the exclusion of heavy atoms to reduce the latter (see e.g., ref. [24] and references cited therein). In addition, self-healing [4, 7, 34, 122–124], self-blinking [125], and photoactivatable dyes [8], fluorescent sensors for ions [126] and pH etc. have become (commercially) available. Small-molecule additives [29, 31, 32, 127–129] are frequently used as intermolecular reaction partners for dyes to either improve their performance by reduction of photodamage by triplet-states [33], oxygen [33, 130] and other reactive fluorophore species, or to achieve photoswitching [120, 131].

The increasing availability of dyes from commercial sources has been a boon to research and medicine. Companies, however, have often not been forthcoming with information on the chemical structures of the dyes (and their linkers), which has been an obstacle for some applications. Prominent examples are the nucleic acids stains of the SYBR family (SYBR Green, SYBR Gold) [94, 98], the Alexa Fluor dye series (Alexa Fluor 555) [132], and the ATTO dye family (ATTO643). A few structures from these suppliers have recently been made available (SYBR Green [95], ATTO647N [133], ATTO655 [134]). There are many applications where knowledge of the chemical structure of a dye is not compulsory, e.g., when using Alexa Fluor 555 in imaging [135–143] and spectroscopic studies [144–147]. However, Alexa Fluor 555 and Alexa Fluor 647 are dyes that are frequently and successfully used for single-molecule FRET in combination with other fluorophores [111, 148–151], or as a donor-acceptor pair [107, 152–156]. They have become a popular choice because of their favorable performance in many assays, and this is largely due to their high water solubility and the absence of strong (unwanted) interactions between dye and target after bioconjugation.

For Alexa Fluor 647, the chemical structure is known, but there are conflicting reports in the literature on the linker length connecting the two sulfonated SO₃⁻ groups (both

3-carbon [111, 157] or 4-carbon atoms [158–161] have been reported), as well as the structure of the maleimide-linker connecting the chromophore to a biological target [159, 162]. For Alexa Fluor 555, on the other hand, there is no verified information on its fluorophore class or molecular structure. Fluorescent lifetimes were found to be similar to Cy3 [7] suggesting that Alexa Fluor 555 might have a cyanine core [157]. Also chemical structures were proposed, but never verified experimentally [76, 161, 163, 164]. The lack of unequivocal structural data for all these fluorophores limits their proper use for FRET-restrained structural modelling, and in situations where an understanding of dye-target interactions are important [165], such as in MD simulations [156, 161, 166].

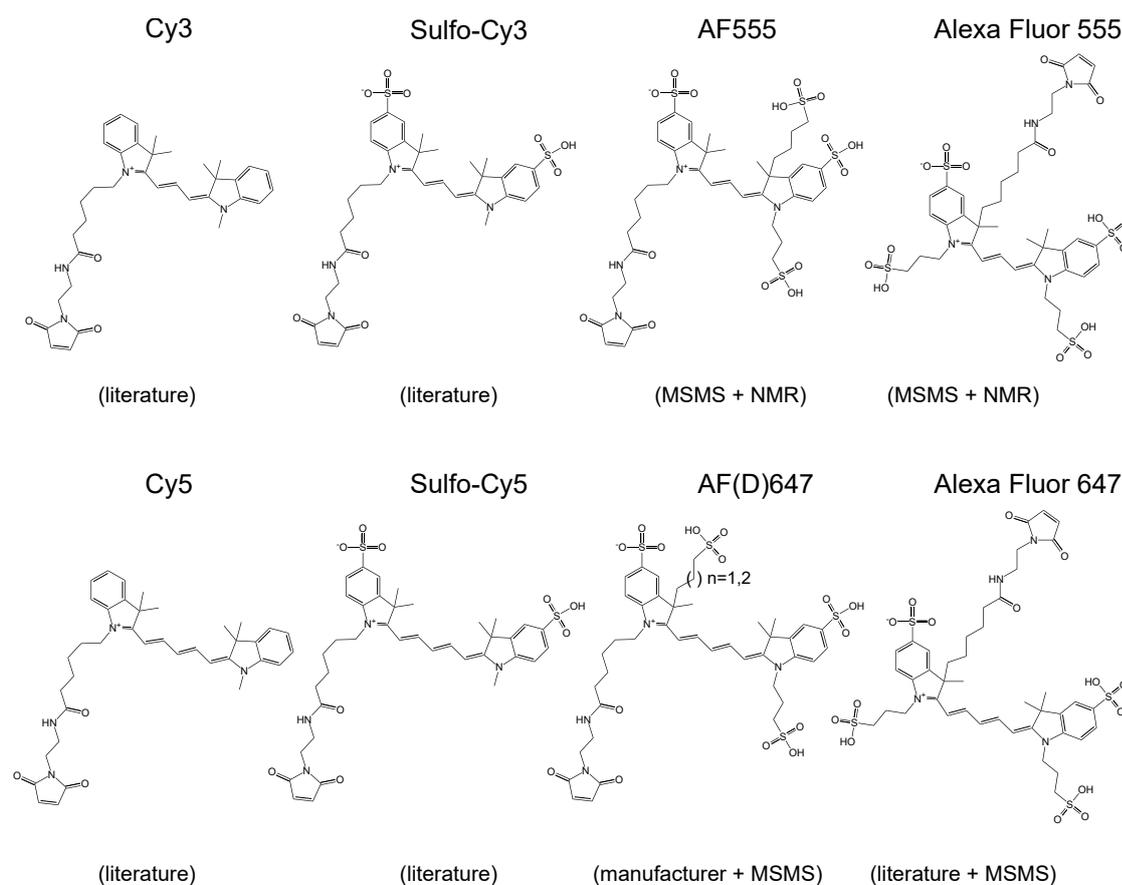


Figure 4.1.1: Confirmed chemical structures of cyanine fluorophores from the Cy-, Alexa Fluor and the AF-series. For Cy3, Cy5, Sulfo-Cy3, Sulfo-Cy5 all structures were obtained from the literature [158, 159, 162]. The structures of AF555 and Alexa Fluor 555 were determined by NMR and mass spectrometry in this manuscript. For AF647, AFD647 and Alexa Fluor 647, we confirmed the published structures (AF(D)647 in [167], Alexa Fluor 647 structure in [157]) by mass spectrometry. Note that the AF-fluorophore homologues of Cy5 are available in two versions called AFD647 ($n=1$) and AF647 ($n=2$).

We have here studied the molecular and spectroscopic properties of Alexa Fluor 555

and Alexa Fluor 647 in relation to other cyanine fluorophores with known molecular structures (Cy3, sulfo-Cy3, AF555, Cy5, sulfo-Cy5, AF(D)647, Figure 4.1.1). Using a combination of visible spectroscopy, mass spectrometry and NMR spectroscopy, we show that Alexa Fluor 555 has a cyanine-based fluorophore core identical to Cy3. Based on the available data, we propose the structure of Alexa Fluor 555 closely related to Cy3 (Figure 4.1.1). We further identified inconsistencies between NMR data and the proposed structure of AF555, which presumably had an incorrect assignment of the locations of alkylsulfonate sidechains. We thus present a refined structure of AF555 that is compatible with our NMR data (Figure 4.1.1). We finally clarified the precise molecular structure of commercial Alexa Fluor 647 maleimide to settle contradicting reports on its linker composition. Using a combination of spectroscopic techniques including continuous-wave absorption and emission spectroscopy, quantum yield determination, fluorescence lifetime and anisotropy spectroscopy of free and protein-attached dyes, we finally compared Alexa Fluor 555 and Alexa Fluor 647 to their structural homologs Cy3/AF555 and Cy5/AF(D)647 in single-molecule FRET applications. Based on the high similarity of the molecular and spectroscopic parameters presented in this manuscript, we explored and characterized the performance of donor-acceptor pairs AF555-AFD647 for smFRET in direct comparison to Alexa Fluor 555-Alexa Fluor 647. Both dye pairs showed good performance in solution-based smFRET experiments using alternating laser excitation. Subtle differences in protein-dye interactions in spectroscopic experiments were further investigated with molecular dynamics simulations. In summary our study suggests that the AF-fluorophores are an attractive alternative to Alexa- and Cy-dyes for smFRET studies but also other fluorescence applications.

4.2 Material and Methods

Sample preparation and labelling of proteins. MalE single and double cysteine variants were obtained and fluorophore-labelled as described previously [102, 107]. The cysteine positions for fluorophore attachment were chosen based on the open and closed x-ray crystal structures of MalE (1OMP, 1ANF, respectively). The double cysteine variants were (i) stochastically labelled with the maleimide derivative of the dyes Alexa Fluor 555 and Alexa Fluor 647 (ThermoFischer Scientific, A20346 & A20347), and AF555, AFD647 & AF647 (Jena Bioscience, APC-007, APC-009 and APC-009) for FRET measurements. (ii) Corresponding single cysteine variants were labelled with one fluorophore as indicated. His-tagged proteins were incubated in buffer containing 1 mM DTT to keep all cysteine residues in a reduced state. Subsequently proteins were immobilized on a Ni Sepharose 6 Fast Flow resin (GE Healthcare). The resin was incubated 2-4 h at 4°C with 25 nmol of each fluorophore dissolved in labelling buffer 1 (50 mM Tris-HCl pH 7.4-8.0, 50 mM KCl, 5% glycerol) and subsequently washed sequentially with 1 CV labelling buffer 1 and 2 (50 mM Tris-HCl pH 7.4-8.0, 50 mM KCl, 50% glycerol) to remove unbound fluorophores. Bound proteins were eluted with 500 μ l of elution buffer (50 mM Tris-HCl pH 8, 50 mM KCl, 5% glycerol, 500 mM imidazole) The labelled protein was further purified by size-exclusion chromatography (ÄKTA pure, Superdex

75 Increase 10/300 GL, GE Healthcare) to remove remaining fluorophores and aggregates. For all proteins, the labelling efficiency was higher than 80% for each labelling site (Supplementary Figure 4.6.1).

Sample handling for quantum-yield, time-resolved anisotropy, and single-molecule FRET measurements. The labelled MalE proteins were stored in 50 mM Tris-HCl pH 7.4, 50 mM KCl with 1 mg/ml bovine serum albumin (BSA) at 4 °C for less than 7 days. The samples were stored at protein concentrations between 100-500 nM and diluted for the measurements indicated as described below.

Single-molecule FRET measurements and data analysis. ALEX experiments were carried out by diluting the labelled proteins to concentrations of ≈ 50 pM in 50 mM Tris-HCl pH 7.4, 50 mM KCl supplemented with the ligand maltose as described in the text and figures. Before each experiment, the coverslip was passivated for 5 minutes with a 1 mg/ml BSA solution in PBS buffer. The measurements were performed without photostabilizer, which showed little effects on the resulting data quality (Supplementary Figure 4.6.2), which is in contrast to the pair Cy3B/ATTO647N used previously for amino-acid binding proteins [102, 106, 107] and ribosome recycling factor ABCE1 [168] where the addition of TX/MEA had a significant positive impact.

Data acquisition and correction procedures were performed for confocal measurements similar to the procedure as described by Hellenkamp et al. [111]. Solution based sm-FRET experiments were performed on a homebuilt confocal ALEX microscope as described in [14]. All samples were studied using a 100 μ l drop of buffer (50 mM Tris-HCl pH 7.4-8.0, 50 mM KCl) on a coverslip. The fluorescent donor molecules were excited by a diode laser at 532 nm (OBIS 532-100-LS, Coherent, USA) operated at 60 μ W at the sample in alternation mode (50 μ s alternating excitation and a 100 μ s alternation period). The fluorescent acceptor molecules were excited by a diode laser at 640 nm (OBIS 640-100-LX, Coherent, USA) operated at 25 μ W at the sample. The lasers were combined and coupled into a polarization maintaining single-mode fiber (P3-488PM-FC-2, Thorlabs, USA). The laser light was guided into an epi-illuminated confocal microscope (Olympus IX71, Hamburg, Germany) by a dual-edge beamsplitter ZT532/640rpc (Chroma/AHF, Germany) and focused to a diffraction-limited excitation spot by a water immersion objective (UPlanSApo 60x/1.2w, Olympus Hamburg, Germany). The emitted fluorescence was collected through the same objective, spatially filtered using a pinhole with 50 μ m diameter and spectrally split into donor and acceptor channel by a single-edge dichroic mirror H643 LPXR (AHF). Fluorescence emission was filtered (donor: BrightLine HC 582/75 (Semrock/AHF), acceptor: Longpass 647 LP Edge Basic (Semrock/AHF)) and focused onto avalanche photodiodes (SPCM-AQRH-64, Excelitas). The detector outputs were recorded by a NI-Card (PCI-6602, National Instruments, USA).

Data analysis was performed using a home written software package as described in [102]. Single-molecule events were identified using an all-photon-burst-search algorithm with a threshold of 15, a time window of 500 μ s and a minimum total photon number of 150 [56]. E-histograms of double-labelled FRET species with Alexa555 and Alexa647

were extracted by selecting $0.25 < S < 0.75$. E-histograms of the open state without ligand (apo) and closed state with saturation of the ligand (holo) were fitted with a Gaussian distribution $A e^{-(E-\mu)^2/(2\sigma^2)}$.

Visible absorption and fluorescence spectroscopy. Absorbance measurements were performed in buffer (50 mM Tris-HCl pH 7.4, 50 mM KCl) on a continuous-wave UV/VIS spectrometer (LAMBDA 465, Perkin Elmer). Absorbance spectra were recorded at a maximum absorbance of ~ 0.4 and base-line corrected to remove background.

Fluorescence emission was recorded in buffer (50 mM Tris-HCl pH 7.4, 50 mM KCl) on a fluorescence spectrometer (LS 55, Perkin Elmer) with excitation/emission slit width of 5 nm and gain values set to 775 V (PMT R928, Hamamatsu). The spectra were corrected for wavelength-dependent detection efficiencies. For data representation and Förster radius calculation, the mean of three repeats of absorbance and emission spectra was calculated and normalized.

Quantum yield measurements. For quantum yield measurements, three dilution series at 5 different concentrations were recorded in absorbance and emission. The absorbance value at the excitation wavelength was averaged over the interval 510 ± 2.5 nm for green and 610 ± 2.5 nm for red fluorophores. The integrated fluorescence was calculated according to $I_F = \int_0^\infty I(\lambda) d\lambda$. The respective absorbance values $A_{\lambda_{ex}}$ at 510 nm (green fluorophore) and 610 nm (red fluorophore) were fitted to the function $I_F(A_{\lambda_{ex}}) = m A_{\lambda_{ex}} \cdot 10^{-A_{\lambda_{ex}}/2}$, where the factor $10^{-A_{\lambda_{ex}}/2}$ accounts for the absorption of the excitation light of the emission spectra measurements. The fit returns the initial slopes m (m_{fluo} for the fluorophore under investigation and m_{ref} for a reference fluorophore with known quantum yield). The fluorescence quantum yield of the fluorophores is calculated from the slopes m_{fluo} and m_{ref} as

$$\Phi_{fluo} = \frac{m_{fluo}}{m_{ref}} \Phi_{ref} \quad (4.2.1)$$

where we used the literature values $\Phi_{ref} = 91\%$ for Rhodamine 6G [89] (green fluorophores) and $\Phi_{ref} = 33\%$ for Alexa Fluor 647 [169] (red fluorophores) as reference. The reported values and standard deviations result from three independent experiments.

Förster radius calculation. The Förster radius R_0 was calculated according to

$$R_0 = \sqrt[6]{\frac{9 \ln(10)}{128 \pi^5 N_A} \frac{\kappa^2}{n^4} Q_D \frac{\int_0^\infty F_D(\lambda) \varepsilon_A(\lambda) \lambda^4 d\lambda}{\int_0^\infty F_D(\lambda) d\lambda}}, \quad (4.2.2)$$

with the following values set to theoretical values:

The other parameters were derived from absorption/emission spectra and quantum yield measurements as described above.

Orientation factor κ^2 :	2/3
Averaged refractive index n:	1.33 in buffer / 1.4 for protein (ref. [170])
Extinction coefficient at maximum:	270 000 1/(M cm) (ref. [169, 171])

Time-correlated single-photon counting for lifetime and anisotropy determination. Bulk lifetime and polarization decay measurements were performed using on a homebuilt setup (Supplementary Figure 4.6.3A) as also described in ref. [2] (Chapter 11): 400 μ l of sample was measured in a 1.5x10 mm cuvette at a concentration of around 100 nM. The samples were excited by a pulsed laser (LDH-P-FA-530B for green fluorophores/LDH-D-C-640 for red fluorophores with PDL 828 “Sepia II” controller, Picoquant, GER). Excitation polarization was set with a lambda-half-waveplate (ACWP-450-650-10-2-R12 AR/AR, Laser Components) and a linear polarizer (glass polarizer #54-926, Edmund Optics). Emission light was polarization filtered (wire grid polarizer #34-315, Edmund Optics). The emission light was collected with a lens (AC254-100-A, Thorlabs) and scattering light or Raman contributions were blocked with filters (green: 532 LP Edge Basic & 596/83 BrightLine HC, AHF; red: 635 LP Edge Basic & 685/80 ET Bandpass, AHF). The signal was recorded with an avalanche-photo-diode (SPCM-AQRH-34, Excelitas) and a TCSPC module (HydraHarp400, Picoquant). Polarization optics were mounted in homebuilt, 3D-printed rotation mounts and the APD is protected from light with a 3D-printed shutter unit. An additional neutral density filter with OD = 4 in combination with a flip-mirror was used to guide the laser directly into the detection path for the measurement of the instrument response function.

In a typical experiment, the excitation power was set to 10 μ W at a repetition rate of 20 MHz. The sample concentration was always tuned to obtain a \sim 50 kHz photon count rate. For anisotropy and lifetime measurements, data sets were recorded for each polarization setting for 5 min in the order vertical (VV1), horizontal (VH1), magic angle (MA), horizontal (VH2), and vertical polarization (VV2) under vertical excitation. The anisotropy was calculated based on the sum of two vertical and horizontal measurements to compensate for small drifts in laser power or slow changes in fluorophore concentration due to sticking. With $VV(t) = VV1(t) + VV2(t)$ and $VH(t) = VH1(t) + VH2(t)$, we obtained the anisotropy decay as $r(t) = (VV(t) - GVH(t))/(VV(t) + 2GVH(t))$, where G is the correction factor obtained by measuring with horizontal excitation $G = HV/HH$ (HV and HH is the total signal in the vertical or horizontal channel, respectively) [2]. The IRF was approximated as a sum of (up to) 3 Gaussians convoluted with a fast exponential decay, which fitted and reproduced the IRF in our setup:

$$IRF(t) = \sum_{i=1}^3 A_i e^{-\lambda_i(t - (t_{irf} + (1 - \delta_{i1})\Delta t_i) - \sigma_i^2 \lambda_i/2)} \text{erf} \left(\frac{-\lambda_i(t - (t_{irf} + (1 - \delta_{i1})\Delta t_i) - \sigma_i^2 \lambda_i/2)}{\sqrt{2}\sigma_i} \right). \quad (4.2.3)$$

The times of the Gaussian-exponential convolutes for $i > 1$ are defined as relative time shifts Δt_i with respect to t_{irf} in order to enable that the instrument response function can be shifted with one single time parameter t_{irf} . Please note that the choice for this function was due to the possibility to analytically convolute the IRF with exponential

decays. Alternatively, other functions could be used to describe the IRF, e.g., a sum of gamma distribution, with the same benefit. Alternatively, well-established numerical re-convolution fits could have been performed [2, 91]. For our system, however, the fits were more robust with respect to small IRF mismatches with the described analytical approach. The parameters were derived from a fit of $(IRF(t) + bkg)$ to the measured instrument response function (Supplementary Figure 4.6.3B). The lifetime decays were fitted as convolution of the background-free IRF with a single ($N = 1$) or double exponential decay ($N = 2$), where the fitted IRF parameter were fixed, except of t_{irf} to compensate for small shifts due to heating/cooling effects (Supplementary Figure 4.6.3C/E):

$$I_{MA}(t) = \sum_{d=1}^N I_d(IRF * e^{-\lambda_d t})(t) + bkg \quad (4.2.4)$$

The polarization intensities read as

$$\begin{aligned} I_{VV}(t) &= \sum_{d=1}^N I_d(IRF * e^{-\lambda_d t})(t) + 2r_0 \sum_{d=1}^N I_d(IRF * e^{-(\lambda_d + \lambda_{rot})t})(t) + bkg \\ I_{VH}(t) &= \sum_{d=1}^N I_d(IRF * e^{-\lambda_d t})(t) - r_0 \sum_{d=1}^N I_d(IRF * e^{-(\lambda_d + \lambda_{rot})t})(t) + bkg \end{aligned} \quad (4.2.5)$$

(see also ref. [93, 172]), where the parameters I_d , λ_d , t_{irf} , and bkg are fixed. The calculated anisotropy was fitted with the model $r(t) = (I_{VV}(t) - G I_{VH}(t))/(I_{VV}(t) + 2G I_{VH}(t))$, where the inverse rotational correlation time λ_{rot} and the intrinsic anisotropy r_0 are the only free parameter (Supplementary Figure S4.6.3D/F). All fits were performed as least square fits with weighted residuals according to Poissonian photon statistics.

Mass spectrometry. For mass spectrometric analysis fluorophore standards were run on an ultra-high-performance liquid chromatographic (UHPLC) system including a diode array detector (DAD; Dionex Ultimate 3000 UHPLC, Thermo Fisher Scientific, Waltham, USA) coupled to a timsTOF MS (Bruker Daltonik, Bremen, Germany). Five microliter of each fluorophore sample was injected and separated using a C8 reversed phase column (Ultra C8, 3 μ m, 2.1 x 100 mm, Restek GmbH, Bad Homburg, Germany) with 300 μ l flow per minute at 60°C. Solvents were water (A) and a mixture (70/30 v/v) of acetonitrile and isopropanol (B), both containing 1% ammonium acetate and 0.1% acetic acid. The gradient started with 1 min at 55% B followed by a slow ramp to 99% B and a fast ramp within 14 min. This was kept constant for 7 min and returned to 55% B with additional 4 min of re-equilibration.

Using the DAD the absorption spectra of 190-800 nm were recorded. In parallel mass spectra were acquired by otofControl 4.0 in negative MSMS mode from 100-1300 m/z mass range. The most important parameters are set as followed: capillary voltage 4000 V, nebulizer pressure 1.8 bar, nitrogen dry gas 8 l min⁻¹ at 200°C, collision energy 70

eV, Collision RD 800 Vpp (volt peak to peak). The evaluation was performed by Data Analysis 4.5 and MetaboScape 4.0. All software tools were provided by Bruker (Bruker Daltonik, Bremen, Germany).

NMR spectroscopy. The maleimide derivatives of the dyes Alexa Fluor 555 (ThermoFischer Scientific, A20346, 5*1 mg) and AF555 (Jena Bioscience, APC-007, 2*5 mg) were dissolved in 0.6 mL DMSO-d₆ (99.96% ²H, Eurisotope, St-Aubin, France). NMR spectra were acquired at 298 K on different NMR spectrometers (Avance III / III HD, Bruker Rheinstetten; ¹H frequenc 500 / 600 / 800 / 950MHz, mostly equipped with helium-cooled HCN- inverse triple-resonance cryoprobes) using the Topspin3.5pl7 software package (Bruker, Rheinstetten). In addition to 1D ¹H and ¹³C spectra, various 2D spectra were acquired to achieve the complete signal assignment and structure elucidation of both compounds (2D DQF-COSY, 2D TOCSY, 2D long-range COSY, 2D NOESY, 2D ROESY, 1H,13C-DEPT, ¹H,¹³C -HSQC, ¹H,¹³C -HMBC, ¹H,¹⁵N -HSQC, ¹H,¹⁵N -HMBC). DOSY diffusion spectra were used to identify impurities in the samples. ¹³C chemical shifts were also simulated with the help of the nmrd.org website [173] and compared to the experimentally assigned values in Table 4.6.1.

The exact positions of the alkylsulfonate sidechains and the maleimide linker in Alexa Fluor 555 and AF555 were established from ¹H,¹³C-HMBC long-range correlations between their CH₂ groups and the signals of the cyanine core. The connectivities were further confirmed by observation of the expected NOE correlations between sidechain and cyanine protons. While the assignment allowed us to propose one unique structure for Alexa Fluor 555, we found that for AF555 both the HMBC and NOESY correlations were in disagreement with the previously published structure and suggested an alternative sidechain arrangement as shown below.

Molecular dynamics simulations. Molecular dynamics (MD) simulations of mutants A186C and S352C of the E. coli maltose binding protein (PDB ID 1OMP [174]), each labelled with either AF555, or Alexa Fluor at the mutated site, were performed using the GROMACS MD simulation engine [175]. The initial structures of the fluorophore-labelled proteins were built in PyMOL [176]. The fluorophore was initially oriented away from the protein. For the protein, the amber99sb [177] force-field description was used. Fluorophore parameters were obtained as follows. The fluorophore was cut off including the linking cysteine residue and the cysteine termini were capped with an N-methyl amide group at the C-terminus and an acetyl group at the N-terminus. The AM1 method [178] in the AMBER antechamber package [179] was used to optimize the geometries and determine partial charges for the two dye structures, as well as to determine atom types based on the gaff force field [180]. The resulting partial charges are very similar in equivalent functional groups in the two dyes (Supplementary Figure 4.6.4). The partial charges assigned to the sulphonated groups SO₃⁻ were found to be similar to other existing dye parameterizations [181] (Supplementary Figure 4.6.4/4.6.5, Supplementary Table 4.6.2). If available, covalent interaction terms were taken from the AMBER-DYES force field [159]. Missing terms were taken from the gaff-based antechamber parameterization.

The fluorophore-labelled proteins were solvated in cubic computational water boxes of edge length 9.5-10.2 nm. The TIP3P water model [182] was used. A neutralizing amount of sodium counterions was added to the solvent. The systems were energy-minimized using the steepest descent algorithm. Position restraints with a force constant of $1000 \text{ kJ mol nm}^{-2}$ were put on all solute heavy atoms and the system was simulated for 100 ps at constant volume and a temperature of 100 K. Throughout, the Berendsen thermostat [183] with a coupling time of 0.1 ps was used for temperature control. In a second and third equilibration step, the system was simulated with reduced (force constant $500 \text{ kJ mol nm}^{-2}$) and vanishing position restraints, respectively, at temperatures of 200 and 300 K, respectively, for 100 ps. In a final equilibration step of 100 ps length, pressure control was introduced via the Berendsen barostat96 using a target pressure of 1 bar, a coupling time of 1.0 ps and an isothermal compressibility of $4.5 \cdot 10^{-5} \text{ bar}^{-1}$. For all MD simulations, a time step of 0.002 ps was used, bond lengths were kept constant with the LINCS algorithm [184], van der Waals interactions were described with the Lennard-Jones potential [185] and a cutoff of 1.4 nm and electrostatic interactions were described with the reaction-field method [186], a cutoff of 1.4 nm and a dielectric constant of 80. Coordinates were written to file every 6 ps.

For each of the four equilibrated systems, four long production runs at 300 K and 1 bar, differing in the set of initial velocities assigned from the Maxwell-Boltzmann distribution, of 200 ns length were performed. From these simulations, the minimum distances between the SO₃⁻ sulfur atoms and any protein heavy atom were determined. Distinct fluorophore-dependent behavior concerning the terminal SO₃⁻ in the indole ring attached to the protein linker was detected which is why a set of 19-21 configurations were sampled from the compiled 800 ns simulations per fluorophore-protein system such that these configurations reflect the 800 ns-simulation data in terms of the probability distribution of minimum distances between the terminal SO₃⁻ in the indole ring attached to the protein linker and any protein heavy atom (Supplementary Figure 4.6.6). These structures were used as initial structures in multiple short simulations (20 ns) to calculate the rotational anisotropy decay,

$$r(t) = \frac{2}{5} \langle P_2(\boldsymbol{\mu}(s) \cdot \boldsymbol{\mu}(s+t)) \rangle, \quad (4.2.6)$$

where $P_2(x) = (3x^2 - 1)/2$ is the second Legendre polynomial and $\boldsymbol{\mu}(t)$ is the transition dipole moment vector at time t and the averaging denoted by angular brackets is done over time origins [48, 187].

4.3 Results

Spectroscopic characterization of Alexa and AF dyes. We started our investigation of Alexa Fluor 555 and Alexa Fluor 647 properties by a comparison of absorbance and fluorescence spectra and the determination of spectroscopic parameters such as fluorescence lifetime and anisotropy against well-characterized green dyes such as Cy3, sulfo-Cy3, AF555 (cyanines), and Alexa546 (rhodamine). For comparison of Alexa Fluor 647,

we selected Cy5, sulfo-Cy5, AF647 (cyanines) and ATTO647N (carbopyronine); data see Figure 2.

By inspection of normalized spectra of the green-absorbing dyes in both absorption and emission (Figure 4.3.1A), we see a clear bathochromic shift when SO_3^- groups are attached to the Cy3-core structure ($\text{Cy3} \mapsto \text{sulfo-Cy3} \mapsto \text{AF555}$). All dyes show three vibronic peaks, e.g., for Cy3 at 540 nm, 510 nm and 475 nm, which are also seen for sulfo-Cy3, AF555 and Alexa Fluor 555, yet at higher wavelengths. The spectra of Alexa Fluor 555 and AF555 are almost indistinguishable. These spectral characteristics of the cyanine dyes can be distinguished from e.g., rhodamine dyes such as Alexa Fluor 546 which shows absorption and emission in a similar spectral window, but with different ratios of the vibronic levels [188]. Additional indication for a cyanine fluorophore-core in Alexa Fluor 555 is provided by fluorescence lifetimes experiments and relative quantum yields in comparison to AF555 (Table 4.3.1). Both the lifetime decays of Alexa Fluor 555 and AF555 and the relative quantum yields are highly similar. Any observed discrepancy was likely due to different background levels during our experiments. A reconvolution fitting procedure revealed similar lifetimes of 0.35 ± 0.05 ns and 0.33 ± 0.04 ns for Alexa Fluor 555 and AF555, respectively, in agreement with literature values for free Alexa Fluor 555 of 0.3 ns [132]. Time-resolved anisotropy decays also revealed comparable anisotropy decays of Alexa Fluor 555 and AF555 with steady-state anisotropies of 0.20 ± 0.01 for both Alexa Fluor 555 and AF555. Rotational decay times and errors based on fit uncertainties were found for Alexa Fluor 555 of 0.40 ± 0.04 ns and 0.45 ± 0.04 ns for AF555. All this is in agreement with previously determined steady-state anisotropy values of ~ 0.19 for Alexa Fluor 555 [8]. Similar systematic trends can be observed for Alexa Fluor 647, AF(D)647 in comparison to Cy5 and sulfo-Cy5 related to spectral shifts and variation of oscillator strength of vibronic transitions (Figure 4.3.1B). Also, the lifetime analysis of Alexa Fluor 647 and AFD647 and AF647 showed similar decays. A reconvolution fitting procedure revealed lifetimes of 1.12 ± 0.04 ns, 1.10 ± 0.04 ns, and 1.08 ± 0.05 ns for Alexa Fluor 647, AFD647 and AF647, respectively, all in agreement with literature values reported for Alexa Fluor 647 of 1.0 ns [132]. Time-resolved anisotropy decays revealed comparable anisotropy decays of Alexa Fluor 647 and AFD647, and AF647 with steady-state anisotropies of 0.13 ± 0.01 for all three fluorophores, in agreement with published values of 0.16 for Alexa Fluor 647 [8]. The rotational decay time was determined to be 0.58 ± 0.06 ns, 0.54 ± 0.04 ns, and 0.52 ± 0.07 ns for Alexa Fluor 647, AFD647, and AF647, respectively. The differences in rotational correlation times were not significant for the green fluorophores (Alexa Fluor 555, AF555) and red fluorophores (Alexa Fluor 647, AFD647, and AF647). We observed, however, a clear difference between the green fluorophores ($\tau_{rot} \approx 0.4 - 0.45$ ns) and the red fluorophores ($\tau_{rot} \approx 0.55$ ns), which is in good agreement with reported values for Cy3 of 0.33 ns [189] and 0.38 ns [23] and for Cy5 of 0.54 ns [189]. The difference can be explained by the larger size of the red fluorophores and the corresponding hydrodynamic radii (Stokes radii) of 0.75 and 0.82 nm, respectively (according to Stokes–Einstein–Debye equation under the assumption of a sphere with $\tau_{rot} = \eta V / k_B T$, where η is the viscosity, V the sphere volume, and $k_B T$ the thermal energy). Quantum yields of Alexa Fluor 647 and AFD647 were also found to

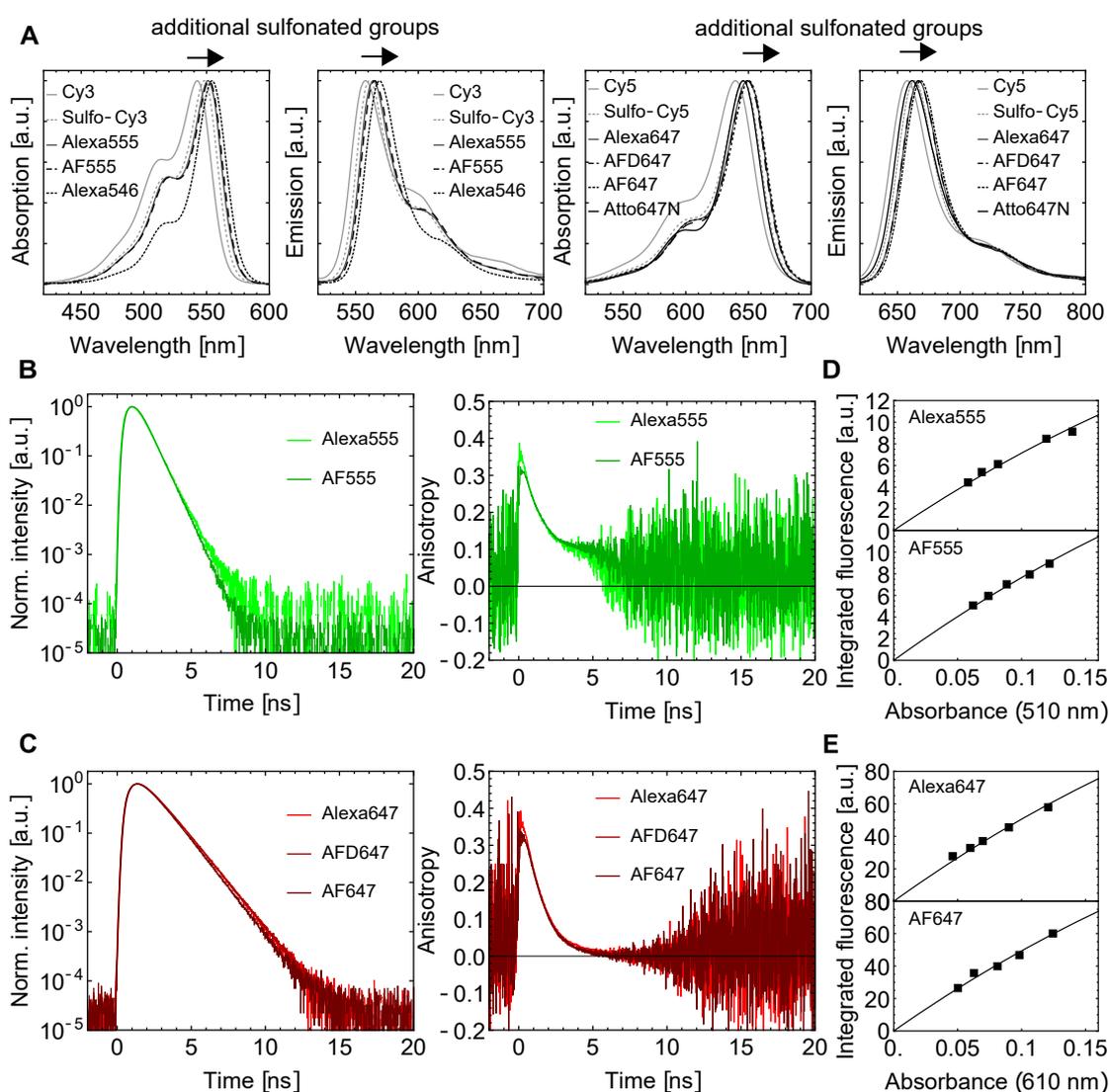


Figure 4.3.1: Spectroscopic characterization of bulk solutions of free green and red cyanine fluorophores. **A** Absorbance and emission spectra of Alexa Fluor 555 and AF555 in comparison to Cy3, Sulfo-Cy3 and Alexa Fluor 546 (left) show red-shifted spectra for increased number of SO₃⁻-groups and difference in spectral shape compared to the rhodamine-derivative Alexa Fluor 546. Absorbance and emission spectra of Alexa Fluor 647, AFD647 and AF647 in comparison to Cy5, Sulfo-Cy5 and Atto647N (left) show red-shifted spectra for increased number of SO₃⁻-groups with small difference in spectral shape compared to Atto647N. **B** Lifetime (left) and time-resolved anisotropy measurements (right) of free Fluor Alexa 555 (lighter green) and AF555 (darker green) at 100 nM concentration. **C** Lifetime (left) and time-resolved anisotropy measurements (right) of free Fluor Alexa 647 (lighter red) and AFD647 (darker red) at 100 nM concentration. **D** Integrated intensity versus absorbance at 510 nm at five different concentrations (squares) for Alexa Fluor 555 (top) and AF555 (bottom) with absorbance-corrected curve fit (solid line, see methods). **E** Integrated intensity versus absorbance at 610 nm at five different concentrations (squares) for Alexa Fluor 647 (top) and AF647 (bottom) with absorbance-corrected curve fit as in D.

Label	Extinction coefficient [1/(M*cm)]	Quantum yield	Lifetime [ns]	Chemical formula	Molecular weight [u]
Alexa Fluor 555 – C2 Maleimide	155.000 [169],(a)	0.1 [169] 0.09 (m)	0.3 [169] 0.35 (m)	C ₄₀ H ₅₀ N ₄ O ₁₅ S ₄ (m)	955.10 (m)
AF555 Maleimide	158.000 [190, 191]	0.09 (m)	0.34 (m)	C ₄₁ H ₅₂ N ₄ O ₁₅ S ₄ [190, 191] (m)	969.12 [190, 191] (m)
Alexa Fluor 647 – C2 Maleimide	270.000 [169],(b)	0.33 [169]	1.0 [169] 1.12 (m)	C ₄₂ H ₅₂ N ₄ O ₁₅ S ₄ (m)	981.14 (m)
AFD647 Maleimide	270.000 (p)	0.33 (m*)	1.1 (m)	C ₄₂ H ₅₂ N ₄ O ₁₅ S ₄ (m)	981.14 (m)
AF647 Maleimide	270.000 [167, 171]	-	1.08 (m)	C ₄₃ H ₅₄ N ₄ O ₁₅ S ₄ [167, 171] (m)	995.16 [167, 171] (m)

- (a) other reported values of 150.000 [192] and 158.000 [193] can be found on manufacturer page
(b) other reported values of 239.000 [192] and 265.000 [194] can be found on manufacturer page
(m) determined in this manuscript
(m*) determined in this manuscript in reference to Alexa Fluor 647
(p) Jena Bioscience, personal communication, Mai 6, 2020

Table 4.3.1: Photophysical and chemical parameters of Alexa Fluor 555/647 and AF555/AF(D)647.

be similar. Overall, our spectroscopic observations support the idea that all Alexa Fluor and AF-fluorophores studied here contain a cyanine fluorophore-core (Figure 4.1.1/4.3.1, Table 4.3.1).

Molecular characterization of Alexa and AF dyes. To verify the molecular structure of Alexa Fluor 555 and Alexa Fluor 647, we first performed mass spectrometry experiments. For these we used the AF dyes as calibration standards, since the structures of both AF555 and AF647 were available on the supplier webpages. With this approach we determined the molecular mass of the maleimide-derivatives of the fluorophores (Table 4.3.1) and identified characteristic molecular fragments in the MSMS spectrum based on the available structures of AF555 and AF647 (see methods). Alexa Fluor 555 maleimide showed a total mass of 955.10 u (C₄₀H₅₀N₄O₁₅S₄), which is smaller than AF555 (969.12 u; C₄₁H₅₂N₄O₁₅S₄) by the mass of exactly one methylene-fragment (~14 u); Figure 4.3.2. For Alexa Fluor 647, we found a mass of 979.21 u (C₄₂H₅₂N₄O₁₅S₄) and 993.23 u (C₄₃H₅₄N₄O₁₅S₄) for AF647; both findings are consistent with the published structure of AF647 and the idea that Alexa Fluor 647 contains two sulfonated propyl-groups and the standard maleimide 5-carbon linker which is also used for various cyanine fluorophores including Cy3, Cy5 and its sulfonated versions (see also NMR); Figure 4.3.4. To verify the proposed structural differences, we studied the fragmentation patterns of all fluorophores (Figure 4.3.2/4.3.3). Alexa Fluor 555, Alexa Fluor 647 and AFD647 showed a richer fragmentation pattern in comparison to AF555 and AF647. Surprisingly, the mass spectrum for AFD647 was indistinguishable from Alexa Fluor 647, which we also verified by repeats of the experiments with different batches of the dyes. This high similarity raised doubts about structural differences between both compounds, since much

smaller structural variations (between AF647 and AFD647) had a bigger impact on the observed spectra (Figure 4.3.3).

Most dye maleimides showed loss of small fragments with ~80 u (sulfonate group) and/or ~123 u (sulfonated propyl group); Figure 3 and 4. The presence of the standard pentenyl maleimide linker used for various cyanine fluorophores was verified as loss of ~237 u (pentenyl maleimide linker) for Alexa Fluor 647 and AFD647 (Figure 4) or seen as part of a larger ~566 u fragment for Alexa Fluor 555 (Figure 4.3.2). Importantly, for both AF555 and AF647, we could verify that these show an additional ~137 u mass loss related to a sulfonated butyl group, which is not present in the other dyes (Figure 4.3.2 and 4.3.3). Based on the mass spectrometry data, we were able to restrict the pool of potential structures for Alexa Fluor 555 to two isomers, where the maleimide linker and the sulfonate-groups are placed at opposing sites of the fluorophore core (see below). For Alexa Fluor 647, we could verify the structure shown in Figure 4.3.3.

For univocal determination of the Alexa Fluor 555 structure, we next performed NMR. 2D correlation NMR spectra allowed to reconstruct the molecular structure of both Alexa Fluor 555 and AF555 via a complete assignment of all ^1H , ^{13}C and ^{15}N NMR signals. The positions of the alkylsulfonate sidechains and the maleimide linker were then established from ^1H , ^{13}C -HMBC long-range correlations between CH₂ groups and signals of the cyanine core. The connectivity was confirmed by observation of NOE correlations between sidechain and cyanine protons (Figure 4.3.4A).

For Alexa Fluor 555, ^1H - ^{13}C long-range correlations to C2 observed in a 2D HMBC NMR spectrum indicate that the two side chains C10ff and C31ff are linked to the same indole moiety (a, correlations indicated by curved lines). However, the clearly distinct chemical shifts of the H10 signals (4.25/4.29 ppm) vs. the H31 signals (2.10 / 2.35 ppm) prove that C10 must be bound to the nitrogen (similar to C21), and C31 to the aliphatic carbon atom, which rules out the discarded structure. The structure is further confirmed by typical through-space correlations observed in NOESY spectra (green lines). Figure 4.3.4 and additional data (Supplementary Table 4.6.1, Supplementary Figure 4.6.7/4.6.8) thus confirms the Alexa Fluor 555 structure (Figure 4.3.4A, confirmed) and discards an isomeric version (Figure 4.3.4A, discarded). Since we benchmarked the Alexa Fluor 555 data against AF555, we also came to note inconsistencies in the NMR data set of AF555 and the structure that was available from the supplier (Figure 4.3.5). For AF555 both the HMBC and NOESY correlations were inconsistent with the structure available from the supplier (Figure 4.3.5A, discarded), but our data suggested an alternative sidechain different arrangement as shown in Figure 4.3.5A (refined).

The NMR data clearly show that both alkylsulfonate sidechains are linked to the same half of the cyanine core: curved lines indicate ^1H , ^{13}C long-range correlations observed in HMBC spectra (Figure 4.3.5). However, in both possible assignment schemes (broken or solid set), impossible correlations are visible that contradict the previously published structure of AF555 (red lines; Figure 4.3.5A, discarded). In the proposed refined structure, all experimentally observed HMBC correlations correlate with the sidechain arrangement. In addition, through-space correlations from NOESY spectra (green lines, thickness correlates with intensity) are also in agreement with this sidechain

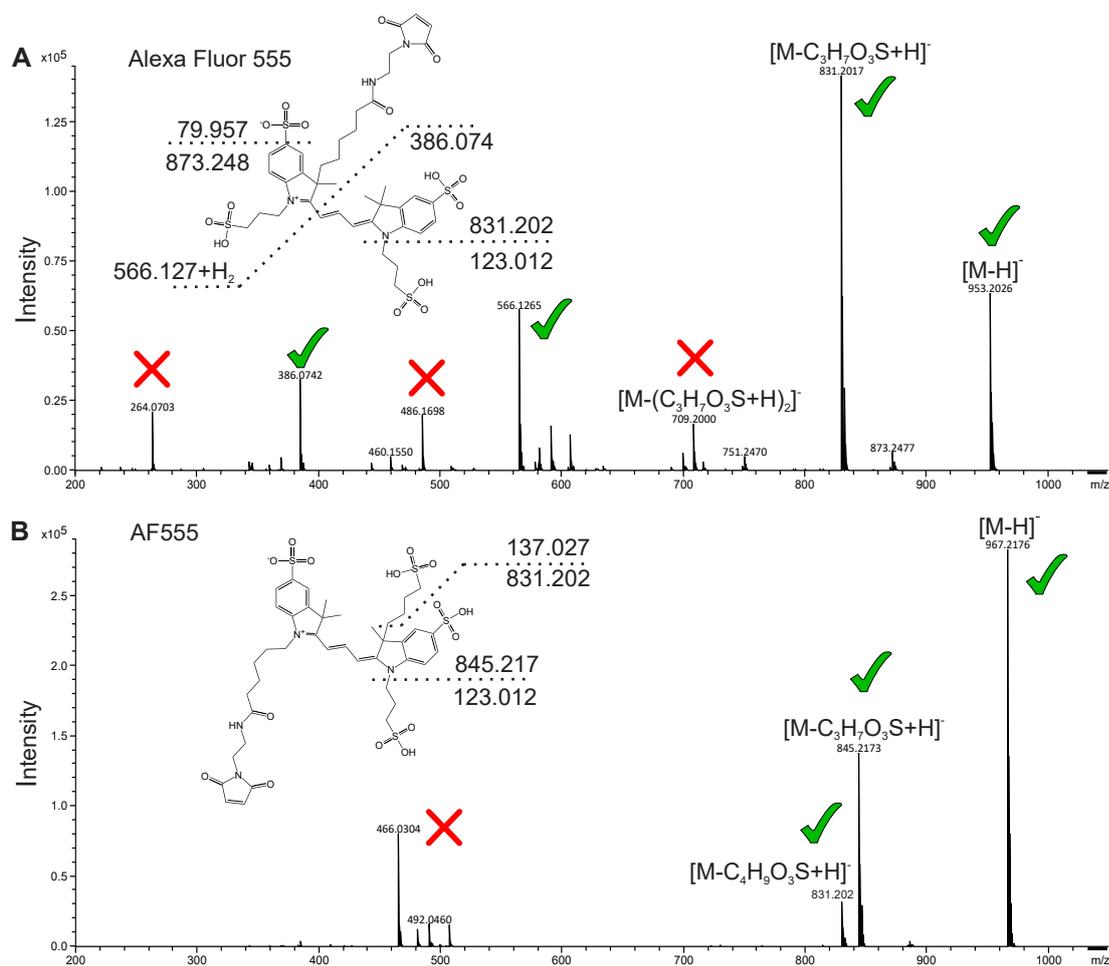


Figure 4.3.2: Mass spectrometry-based structure elucidation. Fragmentation mass spectra of the different fluorophores **A** Alexa Fluor 555 and **B** AF555. Mass range was set to 200-1050 m/z. Mass accuracy was 0.23 ± 0.09 ppm. Each fragmentation spectrum includes the predicted or confirmed structure. Dashed lines represent fragmentation events with resulting fragment masses or empirical formulas. M, molecular ion. Red crosses indicate fragment masses that could not be assigned, green ticks for mass fragments that are compatible with the proposed molecular structures.

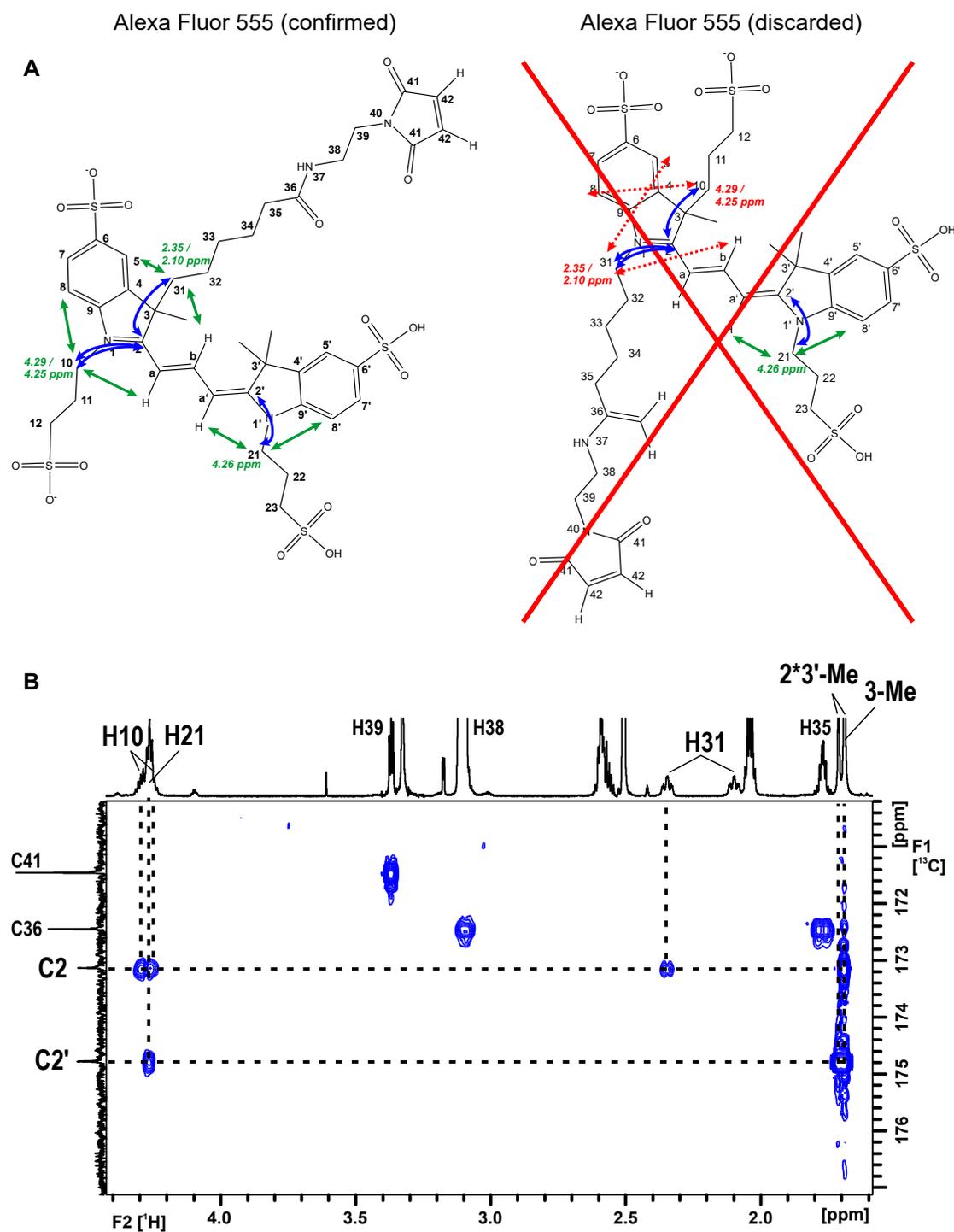


Figure 4.3.4: Structure determination of Alexa Fluor 555 via NMR. **A** Based on mass spectrometry we were able to propose two possible Alexa Fluor 555 structures, from which the left structure was confirmed by NMR spectroscopy. **B** Region from the ^1H , ^{13}C -HMBC spectrum showing long range correlation via two or three bonds, proving that the H10 and H31 protons are close to C2, and H21 to C2' (indicated by blue arrows in the structures in panel A. From the characteristic ^1H chemical shifts, it also becomes clear that both alkyl sulfonate sidechains are connected to a nitrogen atom (H10 & H21 at 4.2-4.3 ppm), unlike the maleimide sidechain (H31 at 2.1-2.4 ppm), which rules out the alternative structure for Alexa Fluor 555. In addition, ^1H , ^1H through space correlations from NOESY spectra are in complete agreement with the confirmed structure (green arrows), but are widely incompatible with the alternative structure (red arrows).

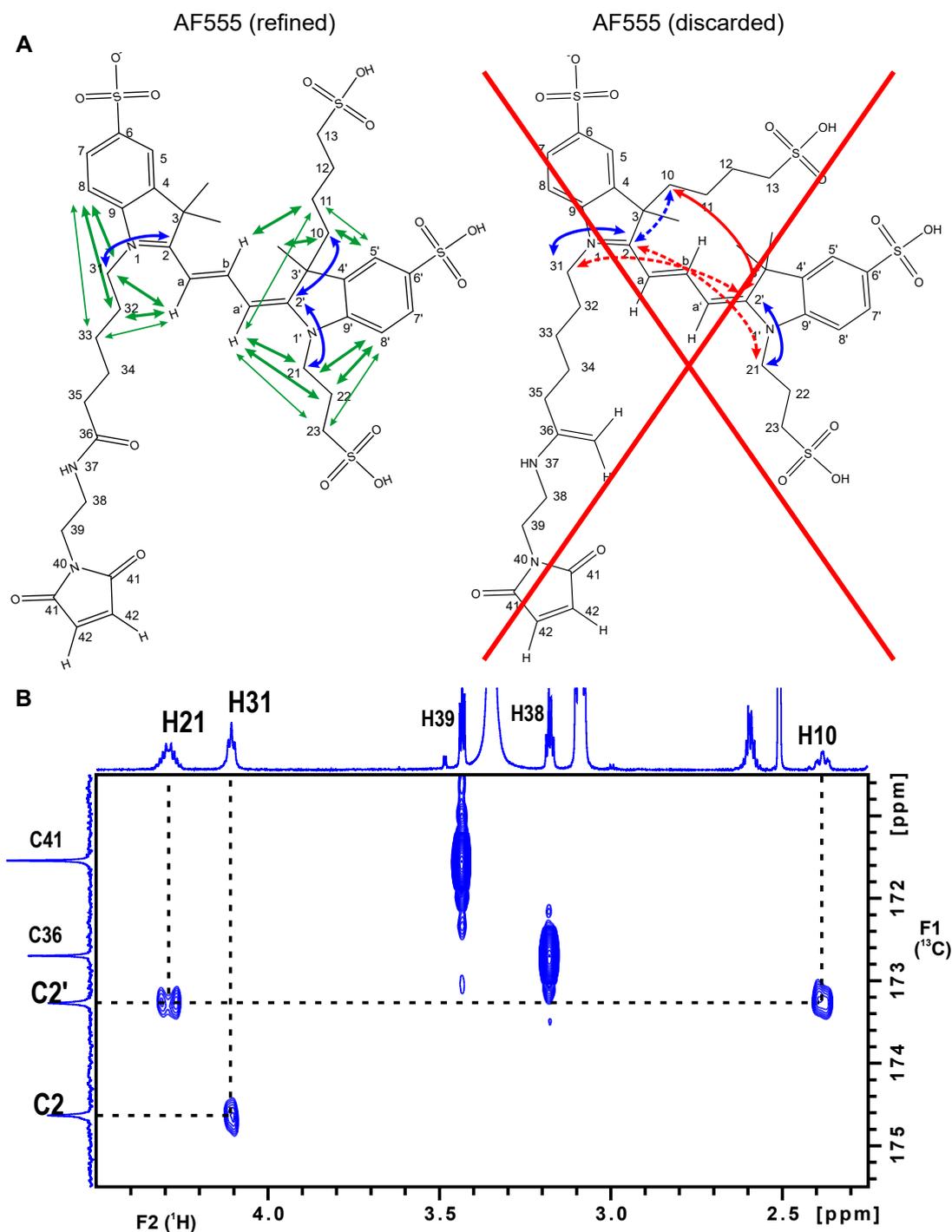


Figure 4.3.5: Structural analysis of AF555 by NMR. **A** Confirmed and refined structure of AF555 (left) and previously published structure (right, discarded). **B** Region from the ^1H , ^{13}C -HMBC spectrum containing long-range correlations via two or three bonds. The cross-peaks clearly show that both alkyl-sulfonate sidechains (with their terminal CH_2 groups H10 and H21) are linked to the same half of the cyanine core ($\text{C}2'$), while the maleimide sidechain is connected to the other half (correlation between H31 and $^{64}\text{C}2$). These correlations are depicted by the blue (possible) and red (impossible) arrows in the structures in A. In addition, experimentally observed through-space correlations from NOESY spectra further confirm this side-chain arrangement (green lines, thickness correlates with intensity).

pattern (Figure 4.3.5A/B). Full NMR data sets for AF555 are available in the supporting information (Supplementary Table 4.6.1, Supplementary Figure 4.6.7/4.6.8).

Alexa and AF dyes for protein labelling. Next, we compared the performance of fluorophores from the Alexa Fluor and AF series for different applications in protein biophysics. The major goal was to use the different dyes for smFRET assays. We selected the periplasmic maltose binding protein (MalE) as a model system (Figure 4.3.6A). MalE is part of the ATP binding cassette transporter MalFGK2 of *E. coli* [195–197]. For our studies, we created both single- and double-cysteine variants of MalE (Figure 4.3.6A) that allow fluorophore labelling via maleimide chemistry at strategic positions. These protein variants were (stochastically) labeled with fluorophores AF555, Alexa Fluor 555, AF(D)647 and Alexa Fluor 647. The selected residues allowed us to create three distance pairs in MalE to monitor ligand-induced structural changes of MalE by maltose. Two of the variants visualized conformational motion (36-352, 87-186) and show inverse effects for addition of maltose, i.e., increase or decrease of inter-residue distances (Figure 4.3.6b/c). We further had one MalE variant that served as a negative control, where no ligand-induced conformational change was expected (85-352); Figure 4.3.6c. The functionality of all variants was verified by microscale thermophoresis experiments which showed the expected change of tryptophane fluorescence upon ligand addition. This allowed us to calculate the ligand affinity for maltose, which we found to be in the low micromolar range (Figure 4.6.9) as found for wildtype MalE [107, 196]. To define the dynamic range of FRET-assays using AF-dyes, we determined the Förster radii (R_0) for different dye combinations based on our data. We calculated R_0 to be 49 ± 1 Å for Alexa Fluor 555-Alexa Fluor 647 and 50 ± 1 Å for AF555-AFD647 in buffer with a refractive index of $n = 1.33$. For labeled proteins [170], the refractive index is often assumed to be an average of $n = 1.40$, decreasing the Förster radii R_0 to 47 ± 1 Å for the Alexa- and 48 ± 1 Å for the AF-pair. Both determined values are in good agreement with reported values of R_0 for Alexa Fluor 555 and Alexa Fluor 647 on RNA (47 - 48 Å) [198] and values provided by the supplier (51 Å) [169]. The distances of our selected variants cover a substantial part of the dynamic range of smFRET for a Förster radius of ~ 5 nm. The cyanine nature of the dyes can, however, impose changes in the donor-quantum yield (see data in Figure 4.3.6). This implies that R_0 can change up to 10% provided the quantum yield does not change more than two-fold.

Alexa and AF dyes for smFRET studies of proteins. We benchmarked the performance of the fluorophore pair AF555-AFD647 in smFRET experiments of diffusing molecules against Alexa Fluor 555-Alexa Fluor 647 for the three different MalE variants (Figure 4.3.6). Labelling of MalE was conducted using established procedures and resulted in similar labelling efficiencies for each site ($>80\%$, see Supplementary Figure 4.6.1) with at least 30% donor-acceptor containing proteins (Supplementary Figure 4.6.10). Notably, a strong interaction and sticking of AF555 when attached to residue 352 led to a skewed profile on the size exclusion chromatogram and retarded the protein on the column (Figure 4.6.1a). In solution-based ALEX-measurements of

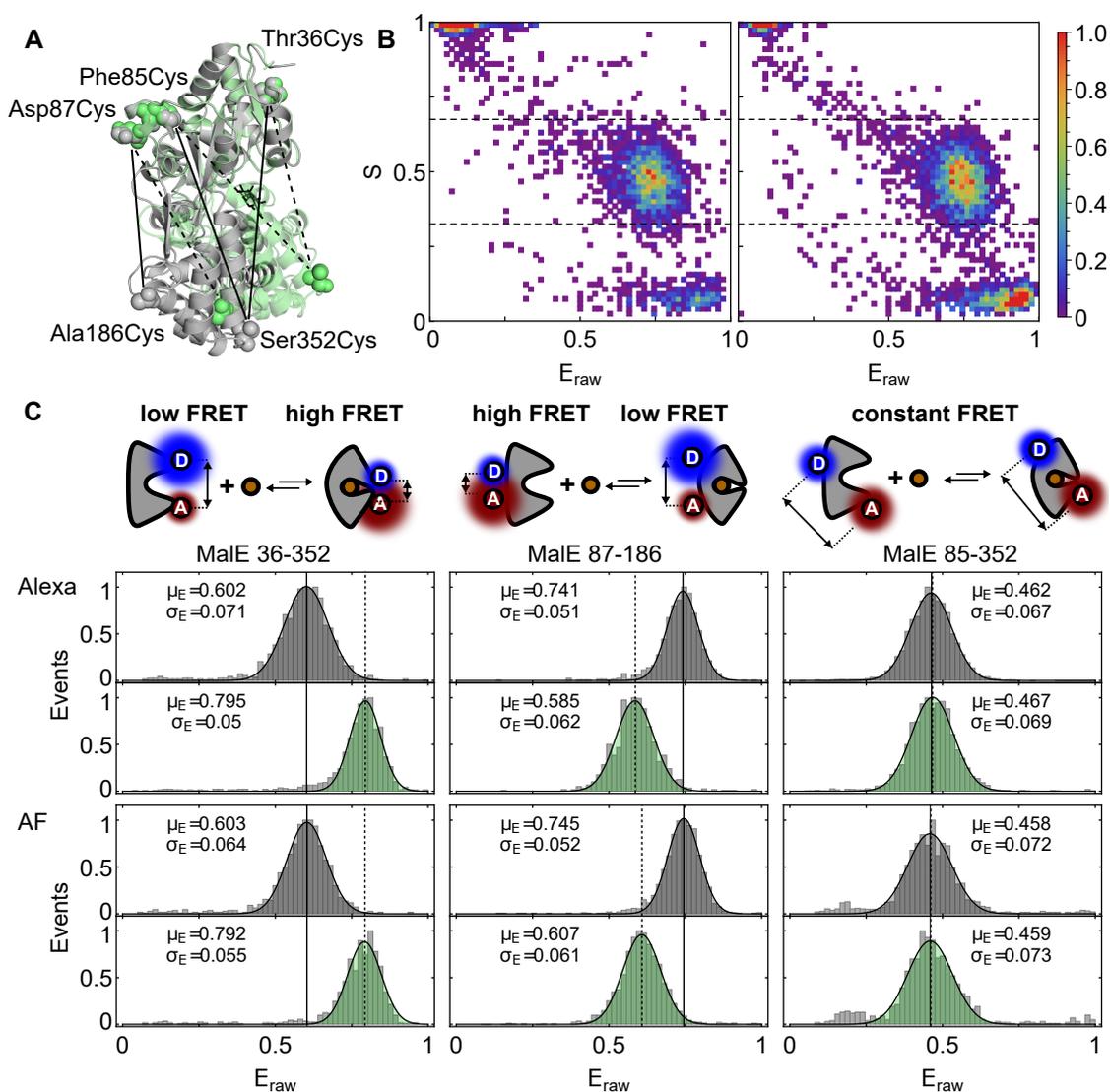


Figure 4.3.6: smFRET measurement comparison of Alexa Fluor 555 – Alexa Fluor 647 and AF555 – AFD647. **A** Overlaid crystal structures of MalE in apo state (gray, PDB 1omp) and holo state (green, PDB 1anf). Residues 4-103 are aligned and labeled residues are marked with spheres (PyMol [176]). The FRET pairs 36-352, 87-186, and 85-352 are indicated with lines for apo (solid) and holo state (dashed). **B** Representative FRET efficiency vs. stoichiometry plots (ES-plots) for MalE variant 87-186 labeled with Alexa Fluor 555/Alexa Fluor 647 (left) and AF555/AFD647 (right) to show data quality and ratio of double labeled donor-acceptor pairs. **C** FRET efficiency histograms (uncorrected, raw FRET values) for the three FRET variants labeled with the Alexa Fluor and AF pair are fitted with a Gaussian fit function with mean μ_E and standard deviation σ_E of the fit. The distributions show similar FRET efficiency values for all variants in apo state (top) and holo state with 1 mM maltose (bottom). Please note that A shows the apo and holo conformation of the protein with insufficient detail to see that MalE85-352 has no change of inter-dye distance upon maltose binding.

all three variants, we obtained very good data quality and similar photon count rates for both dye pairs (see also Supplementary Figure 4.6.11). The comparison of the 2D-histograms showed a clear FRET-related population with coincident detection of donor- and acceptor-signal. We also found little bleaching/blinking effects and shot-noise limited broadening of the FRET populations for both dye combinations (Supplementary Figure 4.6.10/4.6.12). The high data quality can be seen by an inspection and comparison of the 2D-ES histograms, where a substantial donor-acceptor population is observed, which is well separated from both donor- and acceptor-only species (Figure 4.3.6B). The latter suggest that significant blinking- or bleaching effects are absent, since no bridging between the three populations [199] (donor-only: $S > 0.7$, donor-acceptor: $0.7 > S > 0.3$, acceptor-only: $S < 0.3$) are observed; Figure 4.3.6B. All three double-cysteine variants also show the expected trends for the addition of ligand: low-to-high FRET (36-352), high-to-low FRET (87-186) and constant FRET (85-352); see Figure 4.3.6c. Furthermore, the mean uncorrected apparent FRET values were nearly identical for both dye pairs, i.e., their absolute E-value varied only by about $\sim 1\%$. The width of the distributions, which is characterized by the σ -values of the Gaussian fits, varied only in a moderate yet non-systematic way in between both pairs. Also, no sub-ms dynamics due to dye-photophysics were seen in burst-variance analysis (Figure 4.6.12). We noted, however, a slightly elevated bridge component for the dye combination AF555/AFD647 when the residue 352 was present in the protein. This suggests that residue 352 allows for stronger sticking, which is further supported by the skewed size-exclusion chromatography profile of AF555 at this position (see Figure 4.6.1) and other data shown below. In summary, the similarity of the dyes Alexa Fluor 555 / AF555 and Alexa Fluor 647 / AF(D)647 regarding their spectroscopic properties was faithfully reflected in the FRET efficiency distributions. This was expected since both dye pairs should have similar correction factors (direct excitation, leakage, quantum yield ratios) and Förster radii (see Figure 4.3.6). All this establishes the AF-pair as a credible alternative for smFRET investigations.

Lifetime and anisotropy decay of Alexa and AF dyes on proteins. Notably, the smFRET experiments with the AF dyes showed a broadened FRET-population for the variants 36-352 and 85-352 (Figure 4.6.10). To investigate this further, we characterized the environment of dyes and their protein-interactions at positions 186 and 352. We knew from previous steady-state anisotropy experiments that distinct residues in proteins can behave very differently in terms of interactions with dyes, which was indeed observed for MalE [107]. In our experiments, we observed generally faster anisotropy decays and thus less dye-protein interactions at position 186 and stronger interactions at position 352 (slower anisotropy decay) for all tested dyes (Figure 4.3.7 and Supplementary Figure 4.6.13). No detectable difference was seen for the comparison of Alexa Fluor 555 and AF555 at position 186 (Figure 4.3.7A), a result that was similar for Alexa Fluor 647 and AFD647. At residue 352, however, we identified a much slower anisotropy decay, which is indicative of strong protein-fluorophore interactions, in comparison to residue 186. Yet there were no apparent differences between Alexa Fluor 647 and AFD647 (Figure

4.3.7B). To our surprise and despite the structural similarity of Alexa Fluor 555 and AF555, we observed significant differences in protein-fluorophore interactions between both dyes at residue 352. This is interesting since both the confirmed dye-structures of Alexa Fluor 555 and AF555 differ mostly in the orientation of their protein-dye linker and the symmetric placement of the sulfonate residues.

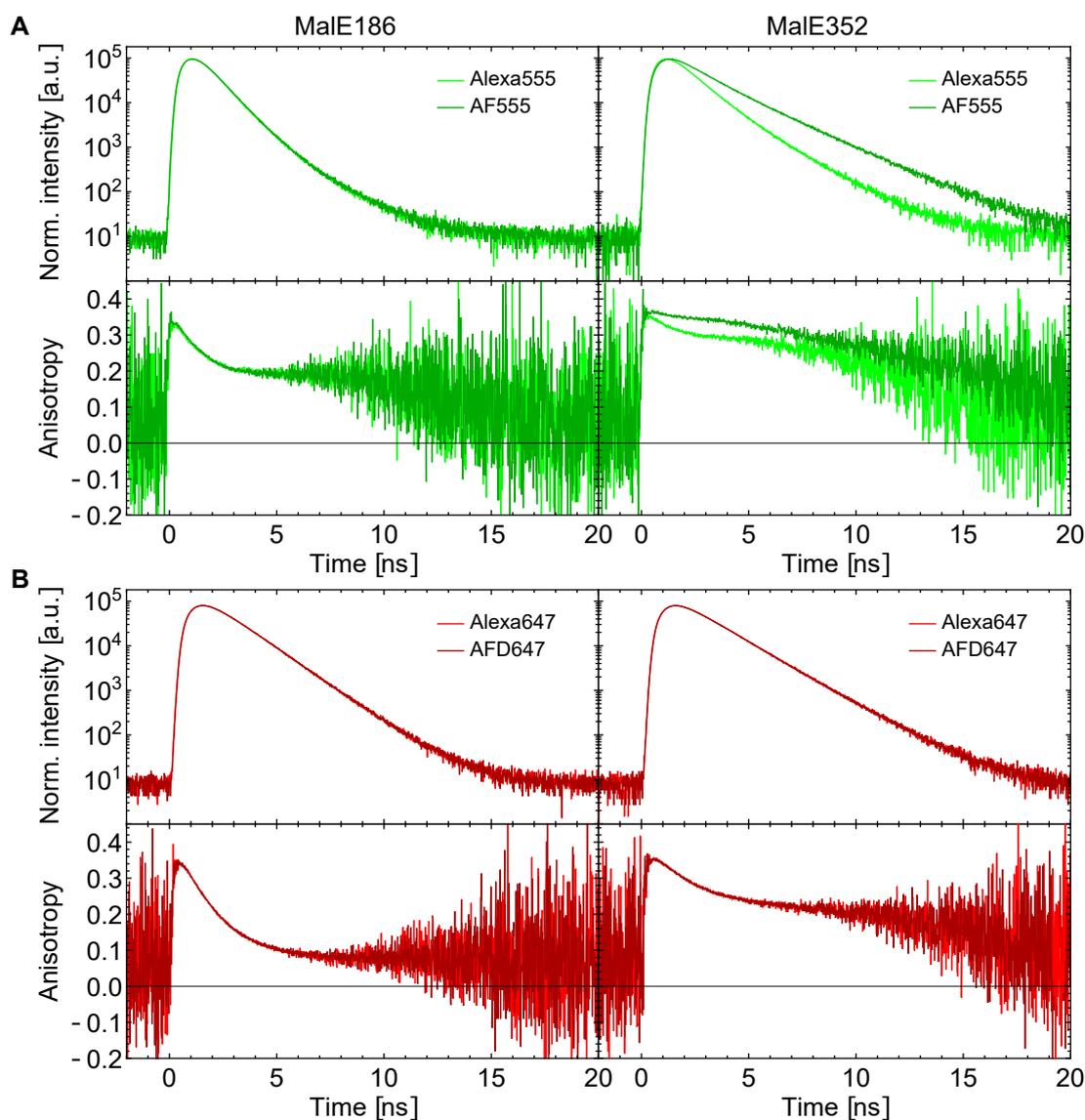


Figure 4.3.7: Characterization of anisotropy and lifetime decays of cyanine dyes on MalE. **A** Lifetime (top) and time-resolved anisotropy measurements (bottom) of Alexa Fluor 555 (lighter green) and AF555 (darker green) labelled at residues 186 (left) and 352 (right) in the ligand-free state of MalE. **B** Lifetime (top) and time-resolved anisotropy measurements (bottom) of Alexa Fluor 647 (lighter red) and AFD647 (darker red) labelled at residues 186 (left) and 352 (right) in the ligand-free state of MalE.

Additional differences appear in the fluorescence lifetime analysis, where AF555 sticking reduces non-radiative de-excitation of the dye molecule and thus AF555 displays a longer fluorescence lifetime as compared to Alexa Fluor 555 at position 352. This effect is related to protein-induced fluorescence enhancement (PIFE) [38, 200] and supports the idea that restricted motion is responsible for longer excited state lifetimes and increased brightness in the green cyanine fluorophores. The structural similarity of Alexa Fluor 555 and AF555 revealed in this paper, also provides a clear explanation for the fact that both can be used as PIFE fluorophores [37, 201–205]. This also why Alexa Fluor 555-labelled dsDNA shows PIFE upon binding to T7 DNA Polymerase gp5/trx, which we could not explain previously due to lacking knowledge of the Alexa Fluor 555 structure. Again this emphasizes the relevance of verifying fluorophore structures and labelling locations in relation to possible interactions of dyes with their environment. To further explain the observed differences between Alexa Fluor 555 and AF555 in their interactions with MalE at specific positions, we performed molecular dynamics simulations (Figure 4.3.8). As described in the methods section, the rotational anisotropy decay $r(t)$ was calculated for multiple short simulations of the maltose binding protein labelled with the confirmed structural variants of AF555 and Alexa Fluor 555 (Figure 4.3.8).

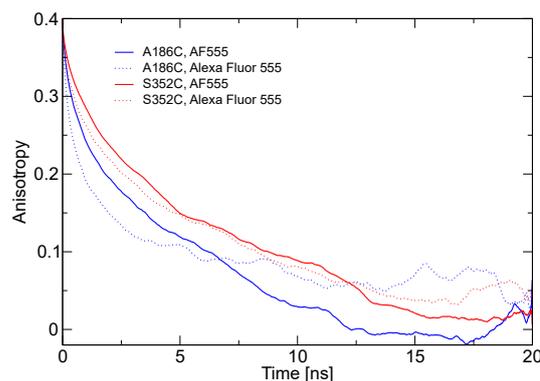


Figure 4.3.8: Average rotational anisotropy decay calculated from MD simulations of fluorophore-labelled MalE. Mutants A186C and S352C were combined with fluorophores AF555 and Alexa Fluor 555, and multiple (~20) simulations of 20 ns length were performed per protein-dye system. The curves depict the average of the anisotropy decay $r(t)$ over these simulations. Time origins for the averaging per simulation are separated by 6 ps.

The rotational anisotropy decay at 186C is faster than at 352C, reflecting, as observed in experiment, reduced dye-sticking at the former site (Figure 4.3.8). The simulations suggest distinct interaction sites, which are extremely sensitive to specific structural features of the fluorophore. Relevant protein-dye interactions that possibly contribute to the reduced motion at the 352C site and selected distances between fluorophore- and

protein-atoms are illustrated in Supplementary Figure 4.6.14 and 4.6.15 for Alexa Fluor 555. For the 352C site, the simulation also shows an (albeit limited) qualitative match to the experimental data for Alexa Fluor 555, where a faster anisotropy decay is found in contrast to AF555 (slow anisotropy decay; Figure 4.3.8).

4.4 Discussion

Using a combined investigation of the spectroscopic and molecular properties of Alexa Fluor 555, AF555 and Alexa Fluor 647, we were able to confirm that all indeed have a cyanine fluorophore core (Figure 4.1.1-4.3.4). NMR analysis also reveals the structure of Alexa Fluor 555 and we were able to propose a refined structure for AF555. Similar studies on Alexa Fluor 647 allowed us to accurately determine its molecular structure and with that settle conflicting reports on its linker structure and the linkers of the sulphonated groups. Our spectroscopic analysis and tests of the dyes in smFRET experiments on proteins (Figure 4.3.6/4.3.7) showed good performance of all dyes in the experiments and a high degree of similarity between the Alexa Fluor and AF dyes as was expected based on the structural similarity. Having determined the fluorophore structures, we were able to derive force-field parameters for MD simulations (Supplementary Table 4.6.2, Supplementary Figure 4.6.4/4.6.5) and parameters for in silico-prediction of accessible volumes of the dyes when used as a FRET label. These parameters are important for predictions of observed mean inter-fluorophore distances and FRET-efficiencies for a combination of smFRET experiments with structural modelling and simulations [206]. Using the structure AF(D)647 and the corrected one of AF555 and the ones of Alexa Fluor 555 and 647 (Figure 4.1.1), we derived all relevant parameters for AV calculations following the method by Kalinin et al. [206] (Table 4.4.1). For these simulations a parametrization for linker and fluorophore core – modelled as an ellipsoid – is provided in Table 4.4.1 according to the proposed procedure (Supplementary Figure 4.6.16) [206].

We finally note that the high structural resemblance of the dyes might render it

Label	Linker length [Å]	W [Å]	R1 [Å]	R2 [Å]	R3 [Å]
Alexa Fluor 555 – C2 Maleimide (d)	21	4.5	8.8	4.2	1.5
AF555 Maleimide (d)	20.5	4.5	8.8	4.4	1.5
Alexa Fluor 647 – C2 Maleimide [207]	21	4.5	11	4.7	1.5
AF(D)647 Maleimide (d)	20.5	4.5	11	4.9	1.5

(d) derived from approximated ellipsoid to fluorophore core and measured atom distances in fluorophore structure (see Supplementary Figure 4.6.16).

Table 4.4.1: Geometric Parameters for in silico predictions of FRET labels using the FRET-restrained positioning system.

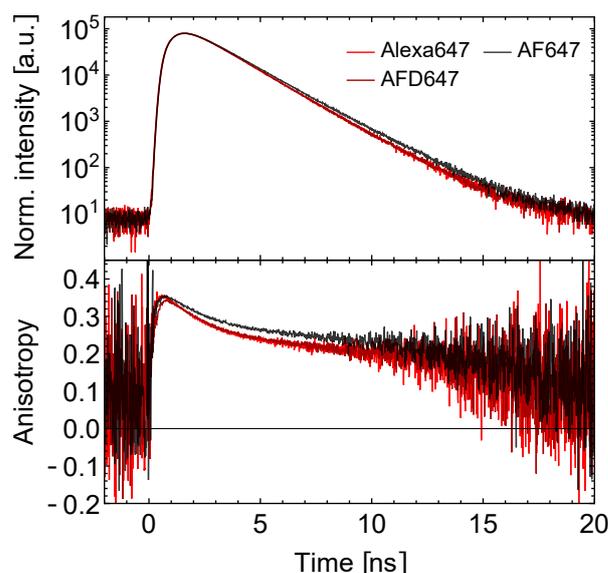


Figure 4.4.1: Characterization of anisotropy and lifetime decays of Alexa Fluor 647 and AF(D)647 on MaleE (position 352) in the ligand-free state of MaleE. The change in one methylene-group from AFD647 to AF647 shows a significant increased anisotropy and lifetime.

reasonable to use either of the dyes without considering the small differences. Yet as shown above, slight structural variations of the dyes can impact dye-protein interactions greatly, e.g., as seen for Alexa Fluor 555 and AF555 with differing lifetimes/anisotropy decays (Figure 4.3.7/4.3.8). Such effects, which we also observed for a comparison of Alexa Fluor 647 and AF647 (but not with AFD647) can largely alter various parameters in a biophysical assay (Figure 4.4.1). Here, we observed significant differences in lifetime and anisotropy decay for a mere addition of a methylene-bridge of the sulfonated group (sulfonated butyl-group instead of propyl-groups). While changes in the donor lifetime can alter the Förster radius, strong dye-protein interactions can produce a large number of additional artifacts ranging from long rotational correlation times of the respective dye to an impact of the dye on the biochemical properties of the protein.

Overall, we conclude that the gathered structural knowledge on Alexa Fluor 555 and 647 will finally enable their applications wherever precise chemical information is required. Furthermore, we conclude that AF555 and AF(D)647 are suitable replacements of the Alexa Fluor dyes in applications for which similar spectroscopic and molecular parameters are required.

4.5 Acknowledgements and Author Contributions

Acknowledgements. This work was financed by an ERC Starting Grant (ERC-StG 638536 - SM-IMPORT to T.C.), Deutsche Forschungsgemeinschaft within GRK2062

(project C03 to T.C.) and SFB863 (project A13 to T.C., project A10 and A13-111166240 to M.Z.), LMUexcellent (start-up funding to T.C.), the Center for integrated protein science Munich (CiPSM) and the Center for Nanoscience (CeNS). C.G. acknowledges a PhD fellowship from the Studienstiftung des deutschen Volkes. Compute resources for this project were partially provided by the Regionales Rechenzentrum Erlangen (RRZE). NMR measurements were performed on spectrometers of the “Bayerisches NMR-Zentrum (BNMRZ)”, jointly supported by TU München and the Helmholtz Zentrum für Gesundheit und Umwelt. We thank D.A. Griffith for carefully reading and commenting on the manuscript.

Author contributions. C.G. and T.C. conceived and designed the study. T.C. supervised the study. C.G. performed research and analyzed data. M.L. performed mass spectrometry. G.G. performed and analyzed NMR measurements. M.M.R. and M.Z. designed the simulation setup, M.M.R. performed and analyzed simulations. C.G. and T.C. wrote the manuscript. All authors discussed and interpreted the results and approved the final version of the manuscript.

4.6 Supplementary Information

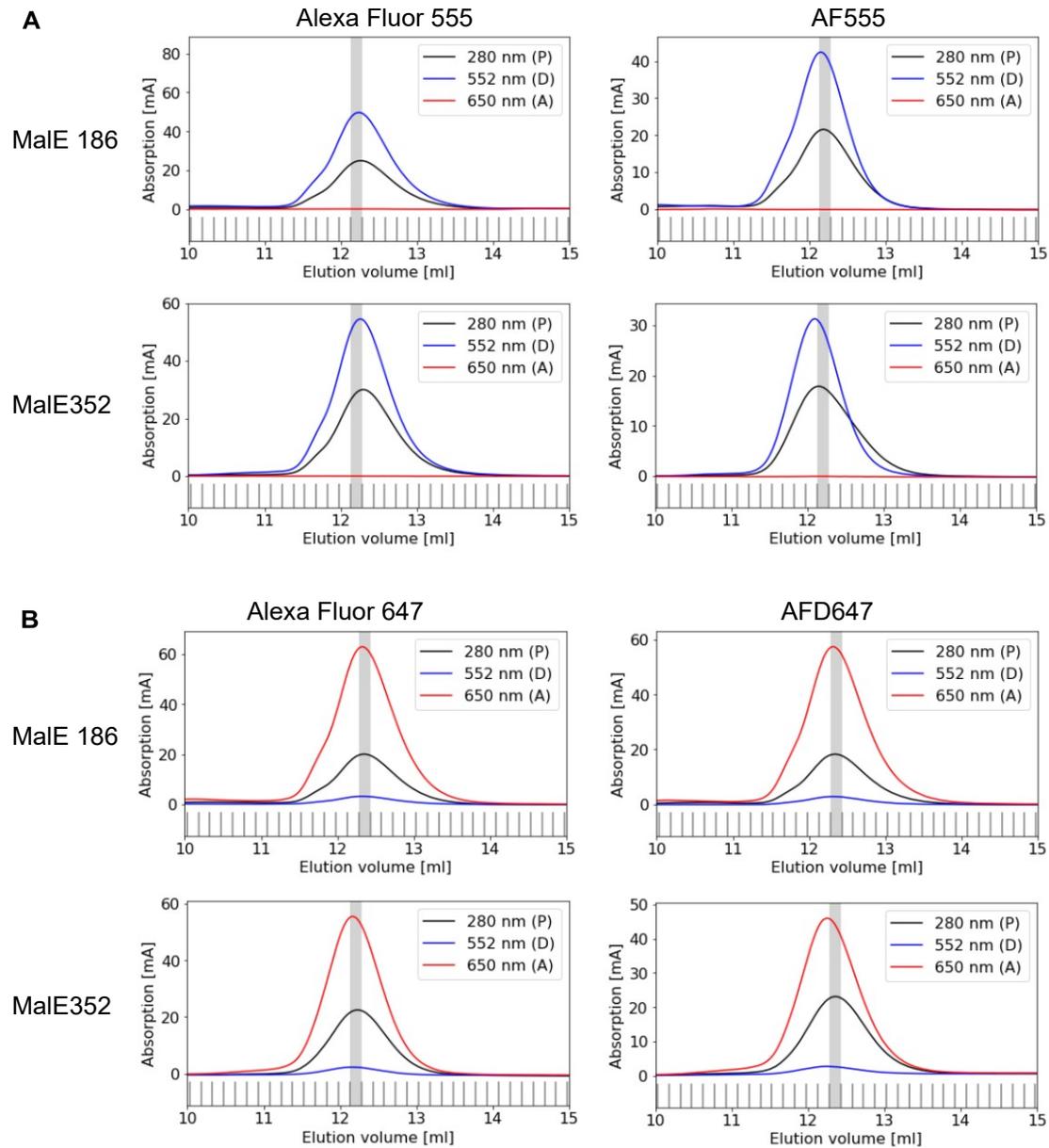


Figure 4.6.1: Investigation of protein quality and labeling efficiency. **A** Size exclusion chromatography (SEC) profiles of Alexa Fluor 555 and AF555 labeled at MaIE mutants 186 and 352. **B** SEC profiles of Alexa Fluor 647 and AFD647 labeled at the same mutants as in A.

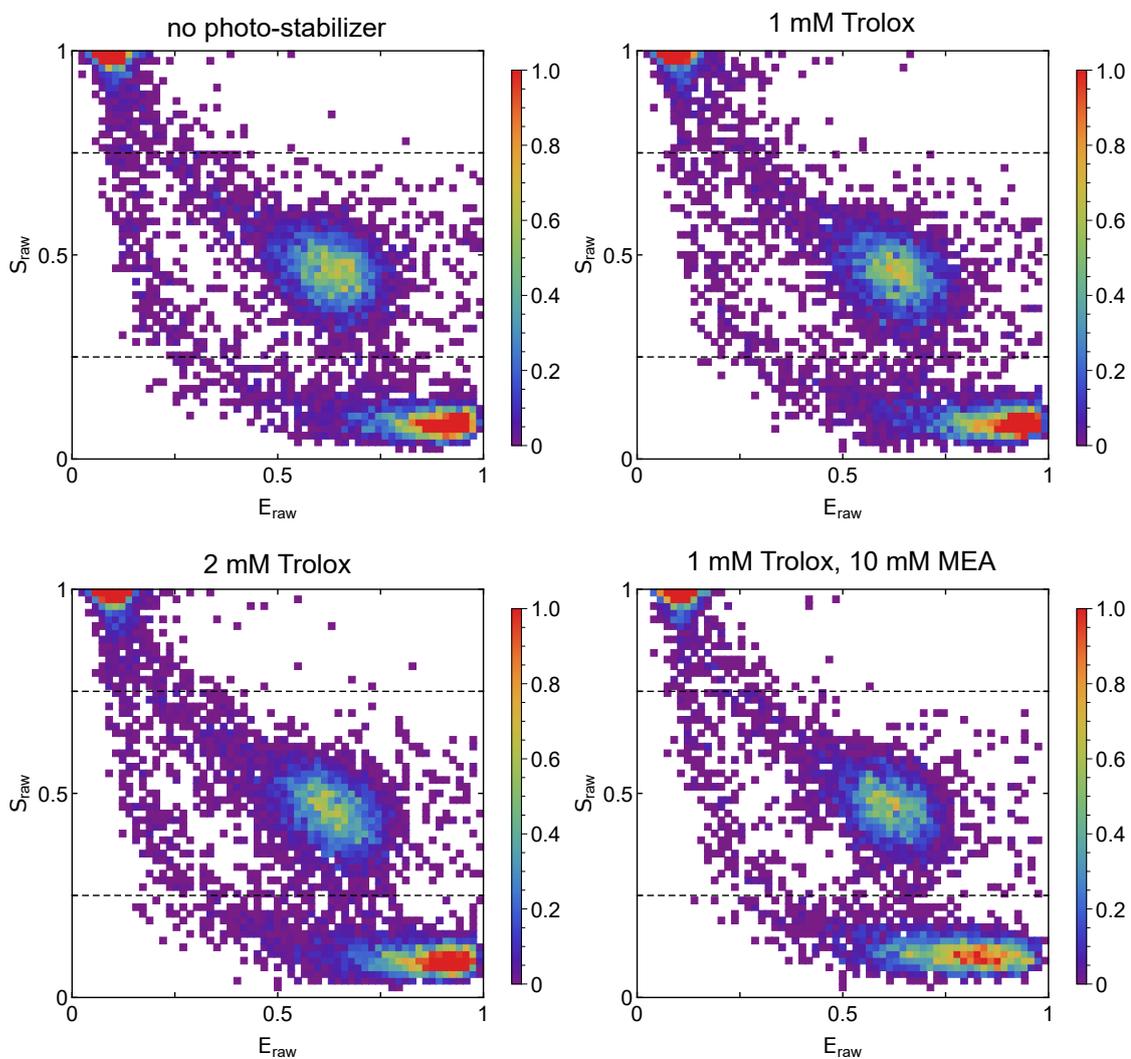


Figure 4.6.2: Photo-stabilizer comparison in FRET measurements with Alexa Fluor 555 – Alexa Fluor 647. ES-histogram of MalE mutant 36-352 in unbound state with photo-stabilizing agents do not show measurable differences in data quality and signal intensity.

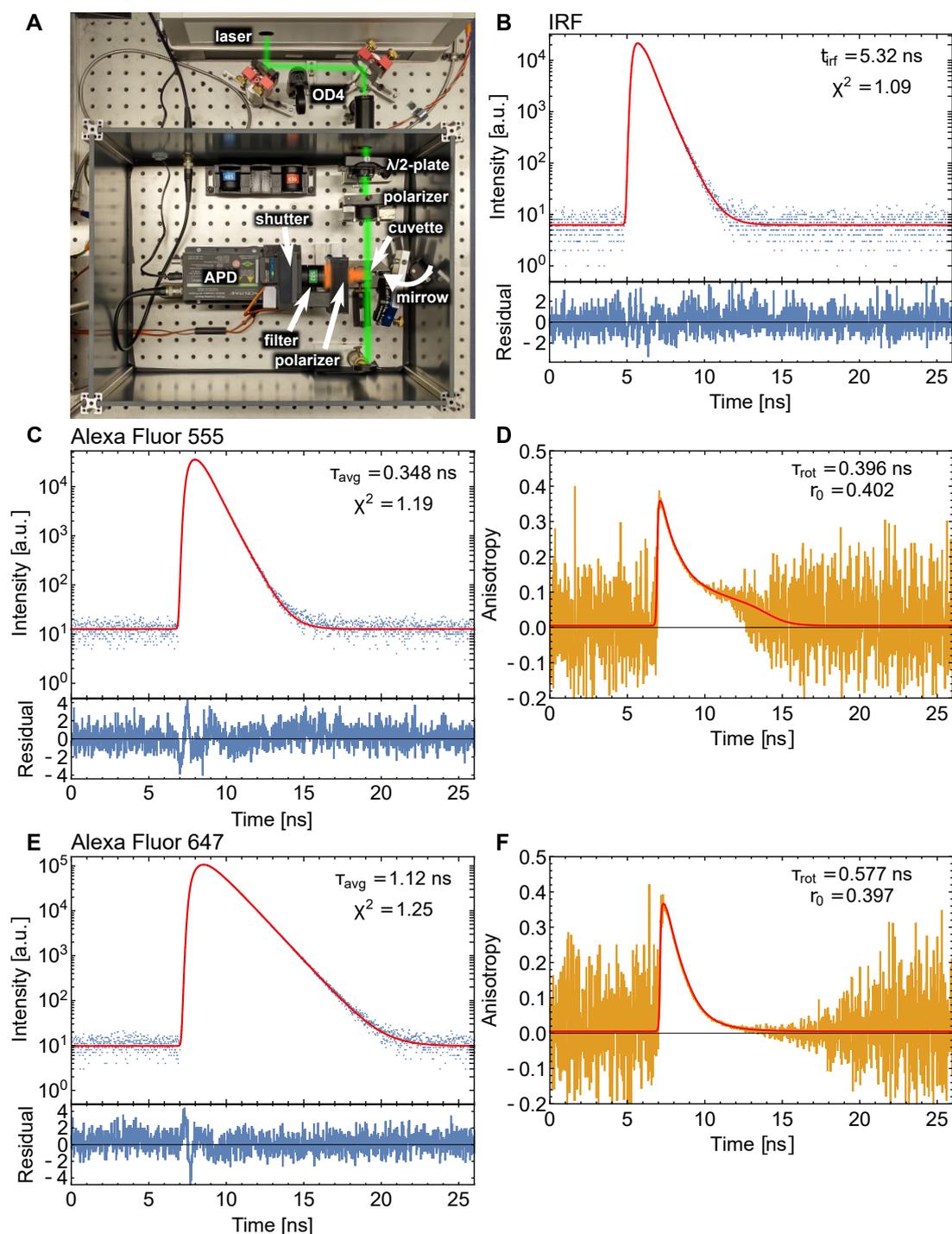


Figure 4.6.3: Lifetime and time-resolved anisotropy measurements. **A** Setup for lifetime and anisotropy measurements (see section 3.3 for details). **B** Intensity profile of green laser (blue) with fitted IRF function (red) as sum of 3 Gaussians convoluted with exponential decays (top) and fit residuals (bottom). **C** Intensity profile of free Alexa Fluor 555 (blue) fitted with IRF function from **B** convoluted with a biexponential decay (red) (top) and fit residuals (bottom). The stated lifetime is the amplitude weighted mean of the two lifetime components. **D** Calculated time-resolved anisotropy of Alexa Fluor 555 (orange) fitted with reconstructed anisotropy based on intensity fit from **C** and measured 75 detection correction factor G (red). The stated rotational correlation time τ_{rot} and intrinsic anisotropy r_0 are the only free fit parameters. **E** Intensity profile of free Alexa Fluor 647 (blue) fitted with IRF function for red laser as in **B** convoluted with a biexponential decay (red) similar to **C**. **F** Calculated time-resolved anisotropy of Alexa Fluor 647 (orange) fitted with reconstructed anisotropy (red) similar to **D**.

4 Molecular and spectroscopic characterization of green and red cyanine fluorophores

Alexa Fluor 555				AF555			
Pos.	¹³ C	¹ H	¹⁵ N	Pos.	¹³ C	¹ H	¹⁵ N
1			163.4	1			162.0
1'			162.6	1'			163.9
2	173.12			2	174.63		
2'	174.77			2'	173.28		
3	53.87			3	49.44		
3'	49.45			3'	53.89		
3-Me	27.85	1.692		3-Me	27.92 27.86	1.702 1.708	
3'-Me	27.85 27.79	1.712 1.689		3'-Me	27.90	1.691	
4	138.76			4	140.64		
4'	140.65			4'	138.78		
5	120.07	7.759		5	120.27	7.801	
5'	120.24	7.795		5'	120.17	7.769	
6	146.21			6	146.30		
6'	146.33			6'	146.24		
7	126.8	7.664		7	126.74	7.681	
7'	126.62	7.654		7'	126.82	7.667	
8	111.07	7.498		8	111.18	7.404	
8'	111.28	7.509		8'	111.06	7.492	
9	143.24			9	142.33		
9'	142.39			9'	143.15		
a	103.88	6.619		a	103.57	6.490	
a'	103.62	6.581		a'	104.09	6.678	
b	149.73	8,345		b	149.71	8.350	
10	43.55	4.297		10	41.60	2.385	
10	43.55	4.258		10	41.60	2.074	
11	24.11	2.042		11	24.01	0.861	
				11	24.01	0.550	
12	48.51	2.567		12	25.59	1.433	
				13	51.54	2.224	
21	43.63	4.265		21	43.52	4.291	
22	24.03	2.042		22	23.99	2.031	
23	48.44	2.596		23	48.35	2.595	
31	41.53	2.348 2.103		31	44.45	4.108	
32	24.10	0.826 0.472		32	27.22	1.725	
33	28.82	1.08 1.04		33	26.24	1.358	
34	24.99	1.261		34	25.21	1.524	
35	35.31	1.773		35	35.52	1.999	
36	172.44			36	172.69		
37		7.834	112.5	37		7.954	112.7
38	37.13	3.099		38	37.19	3.179	
39	37.65	3.367		39	37.76	3.434	
40			148.4	40			148.4
41	171.45			41	171.53		
42	134.95	6.955		42	134.99	6.996	
*	46.24	3.105		*	46.24	3.090	
*	9.11	1.160		*	9.10	1.160	

76 **Table 4.6.1:** Full assignment of the ¹H, ¹³C and ¹⁵N NMR signals of Alexa Fluor 555 and AF555. The numbering scheme is described in the figures of the structures below. The asterisk * denotes signals presumably belonging to trimethylamine counter ions.

SO ₃ ⁻ group	q(S) [e]	q(o) [e]
AF555, refined structure		
1,6	1.46	-0.74
2,5C	1.42	-0.75
2,5N	1.45	-0.76
2,6	1.46	-0.74
Alexa Fluor 555, confirmed structure		
1,5N	1.43	-0.75
1,6	1.48	-0.76
2,5N	1.43	-0.75
2,6	1.46	-0.74
Alexa Fluor 555, discarded structure		
1,5C	1.43	-0.76
1,6	1.46	-0.75
2,5N	1.44	-0.76
2,6	1.46	-0.74
Alexa 488, Best et al. [181]		
	1.38	-0.71
Alexa 594, Best et al. [181]		
	1.46	-0.77
Alexa 350, Graen et al. [159]		
	1.09	-0.68

Table 4.6.2: Partial charges of the atoms of the SO₃⁻ groups in the present fluorophore force-field description compared to the force-field descriptions of Best et al. [181] and Graen et al. [159]. The SO₃⁻ groups are labelled i, j depending on their position, where i = 1 denotes the indole ring attached to the protein linker, i = 2 denotes the second indole ring, j = 5N denotes a five-membered ring with the nitrogen as attachment site, j=5C denotes a five-membered ring with a carbon as attachment site and j = 6 a six-membered ring. For simplicity, the partial charges are here rounded to 2 digits behind the comma. The actual unrounded charges are given in Supplementary Figure 4.6.4.

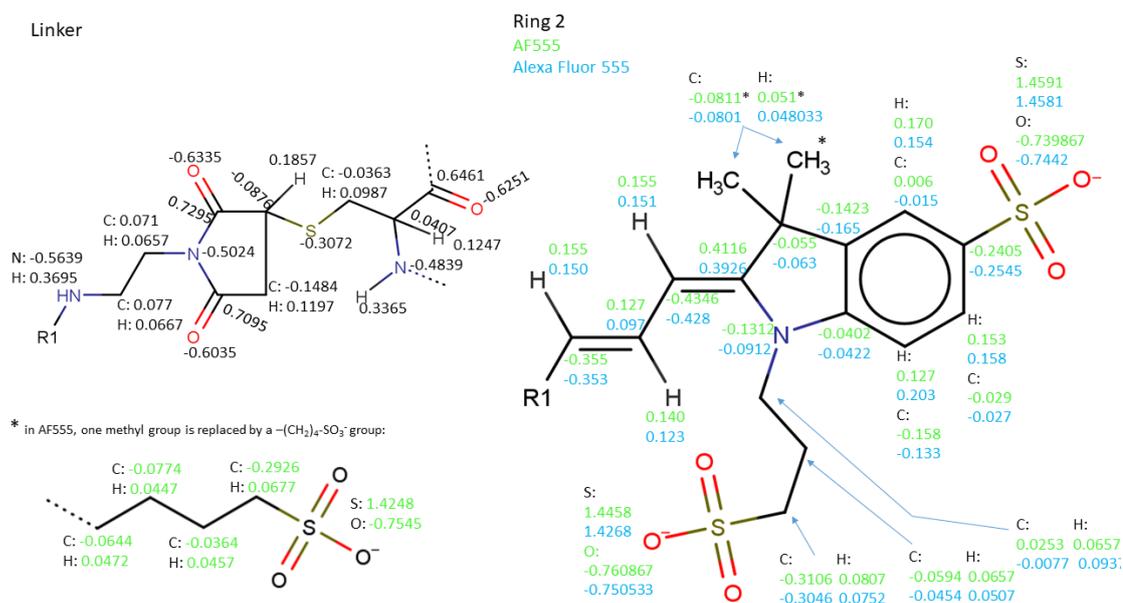
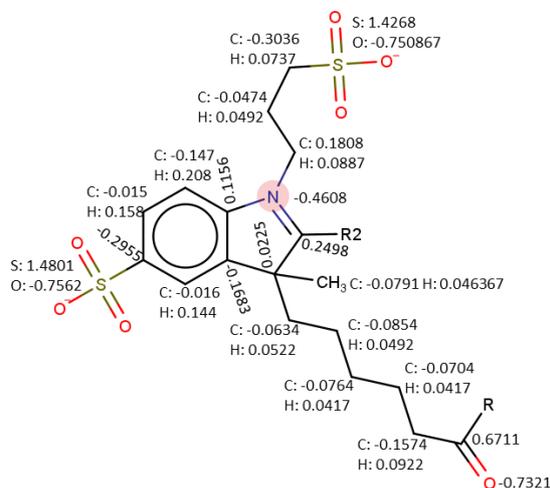


Figure 4.6.4: Partial fluorophore charges. Partial charges determined for the shared structural regions of AF555 and Alexa Fluor 555 with the antechamber package [179] as described in the main article. Abbreviations “R”, “R1” and “R2” refer to the linker, indole ring 1 and indole ring 2 with the bridging $(\text{CH})_3$ groups, respectively.

Ring 1, Alexa Fluor 555



Ring 1, AF555

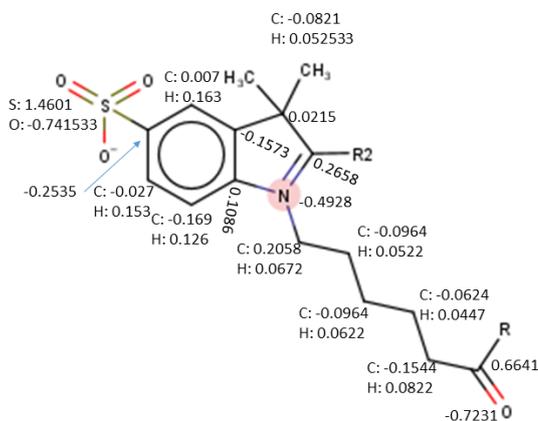


Figure 4.6.5: Partial fluorophore charges. Partial charges determined for the fluorophores Alexa Fluor 555 and AF555 in the distinct structural regions with the antechamber package [179] as described in the main article. Abbreviations “R”, “R1” and “R2” refer to the linker, indole ring 1 and indole ring 2 with the bridging (CH)₃ groups, respectively.

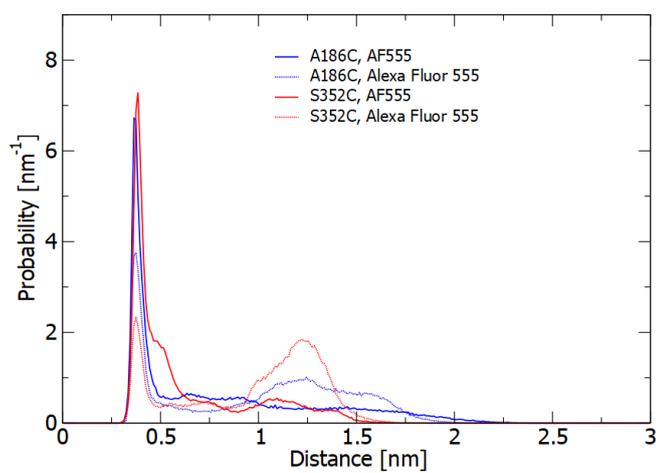


Figure 4.6.6: Distance distribution from MD simulation. Probability distribution of the minimum distance between the sulfur atom of the terminal SO_3^- group in the indole ring attached to the protein linker and any protein heavy atom, for the different dye-protein combinations, sampled during four 200 ns simulations per dye-protein combination. Large distances >1.5 nm are sampled more frequently at the A186C site than at the S352C site.

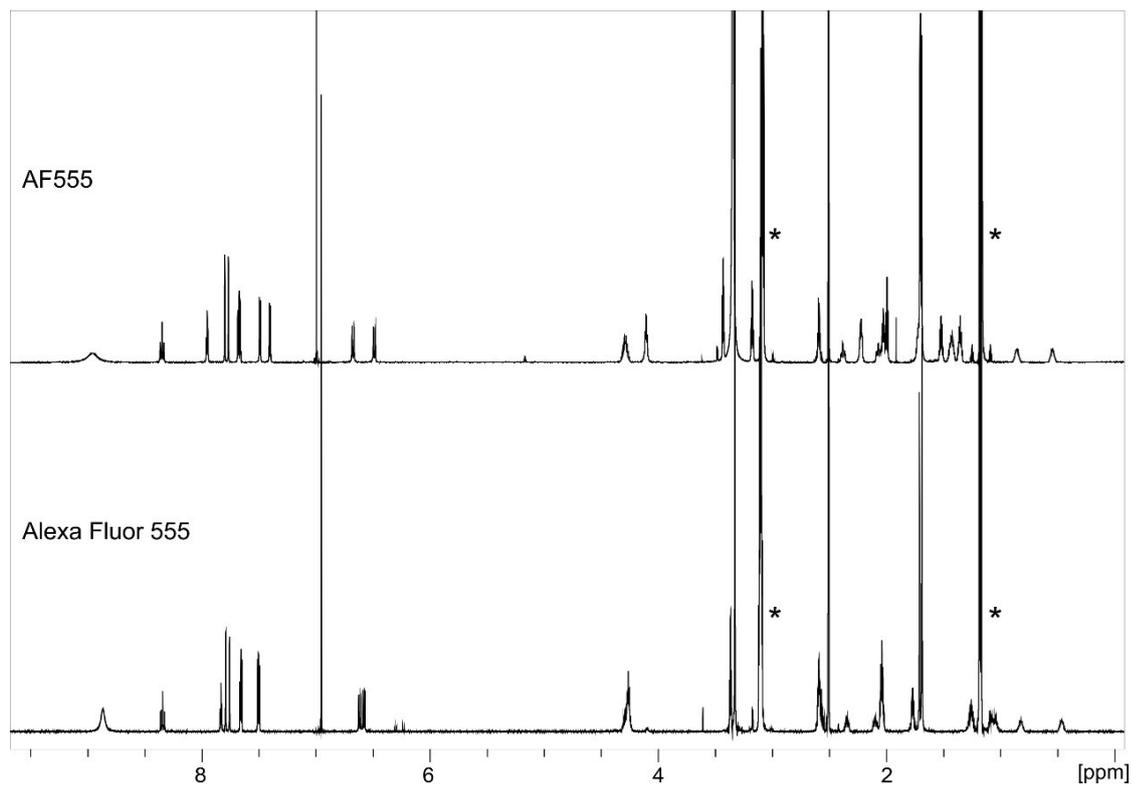


Figure 4.6.7: 1D ^1H NMR spectra of AF555 and Alexa Fluor 555. Both spectra were acquired at 298 K and 800 MHz ^1H resonance frequency (the asterisks * mark signals stemming from the trimethylamine counter ion).

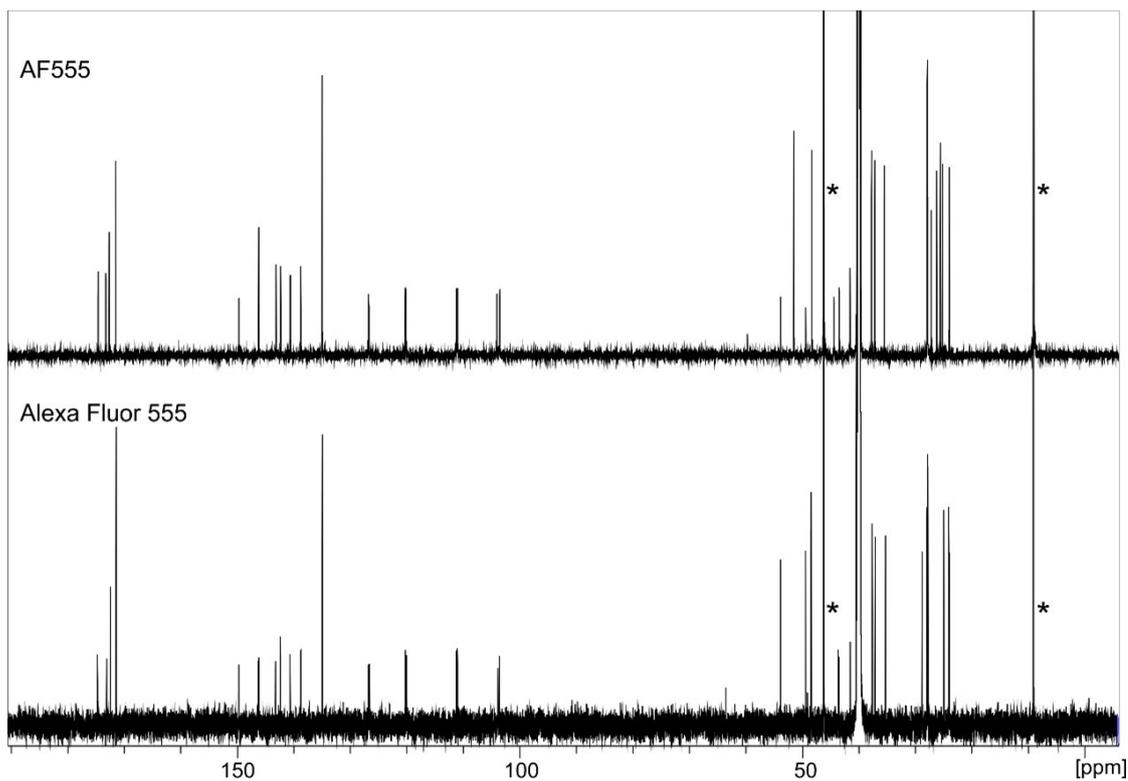


Figure 4.6.8: 1D ^{13}C NMR spectra of AF555 and Alexa Fluor 555. Spectra were acquired at 298 K and 800 MHz (AF555) resp. 950 MHz ^1H resonance frequency (Alexa Fluor 555) with ^1H decoupling (the asterisks * mark signals stemming from the trimethylamine counter ion).

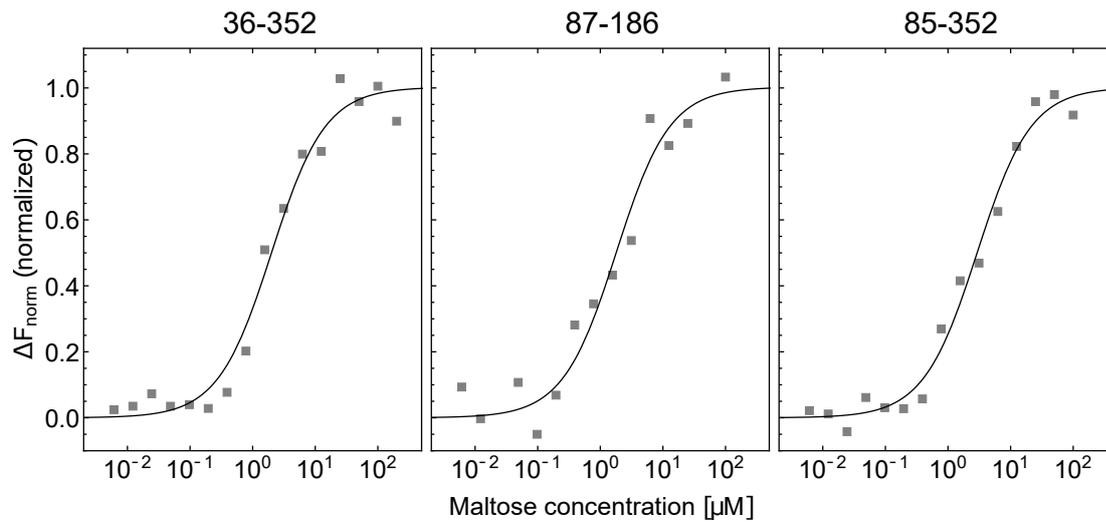


Figure 4.6.9: Binding affinity measurements using microscale thermophoresis.

Binding affinities were measured with microscale thermophoresis (Monolith NT.LabelFree, Nanotemper), where the ratio of fluorescence before and after heating $\Delta F_{norm} = F_{cold}/F_{hot}$ was recorded at different maltose concentrations [208]. Data points show ΔF_{norm} normalized to minimal and maximal fluorescence intensities for unlabelled mutants 36-352 (left), 87-186 (middle), and 85-352 (right). The curves were fitted with a standard model for receptor-ligand kinetics $\Delta F_{norm} = (K_d + c_P + c_{malt} - \sqrt{((K_d + c_P + c_{malt})^2 - 4c_P c_{malt})}) / (2c_P)$, where K_d is the dissociation constant, c_P the protein concentration set to 0.25 μM in the experiment, and c_{malt} the maltose concentration. The fits to the binding model (solid line) gave K_d -values of 2.2 ± 0.4 μM (left), 1.6 ± 0.3 μM (middle), and 2.8 ± 0.3 μM (right), respectively.

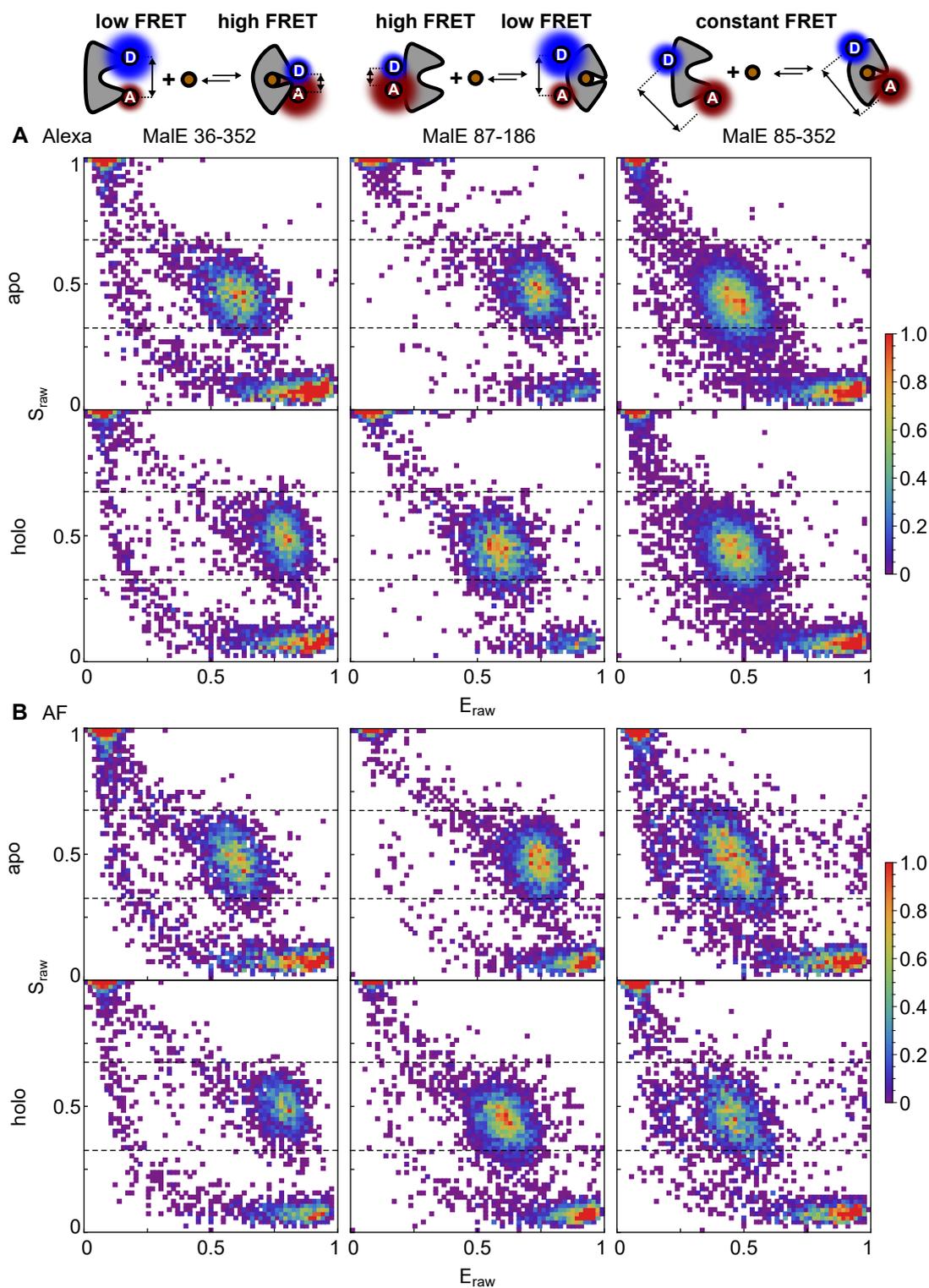


Figure 4.6.10: ES-histograms of smFRET experiments. **A** ES-histograms from all-photon burst search (threshold 150 photons) of all FRET mutants in their apo and holo states for the Alexa-fluorophore-pair and **B** the AF-fluorophore-pair. Data are binned into 61 bins.

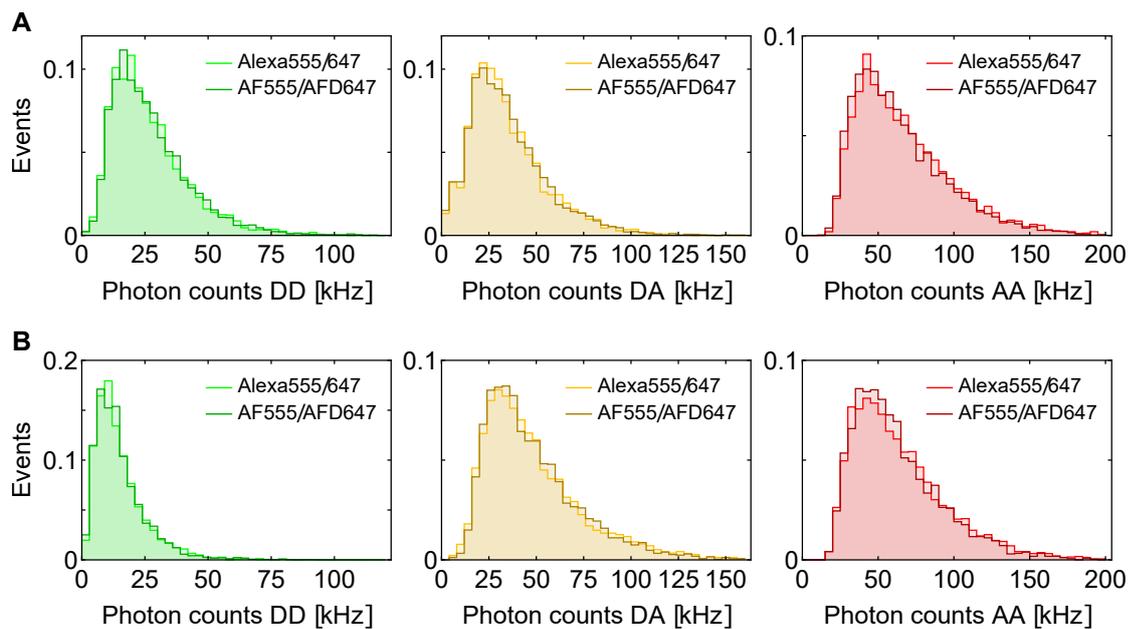


Figure 4.6.11: Investigation of data quality of smFRET measurements between Alexa Fluor 555/647 and AF555/AFD647. A Photon count histograms of selected FRET population ($0.325 < S < 0.675$) for MalE mutant 36-352 in apo state. The different detection channels show very similar photon rates for donor excitation – donor emission (DD, left), donor excitation – acceptor emission (DA, middle), and acceptor excitation – acceptor emission (AA, right). **B** Evaluation as in A for MalE 36-352 mutant in maltose bound state (1 mM maltose in buffer).

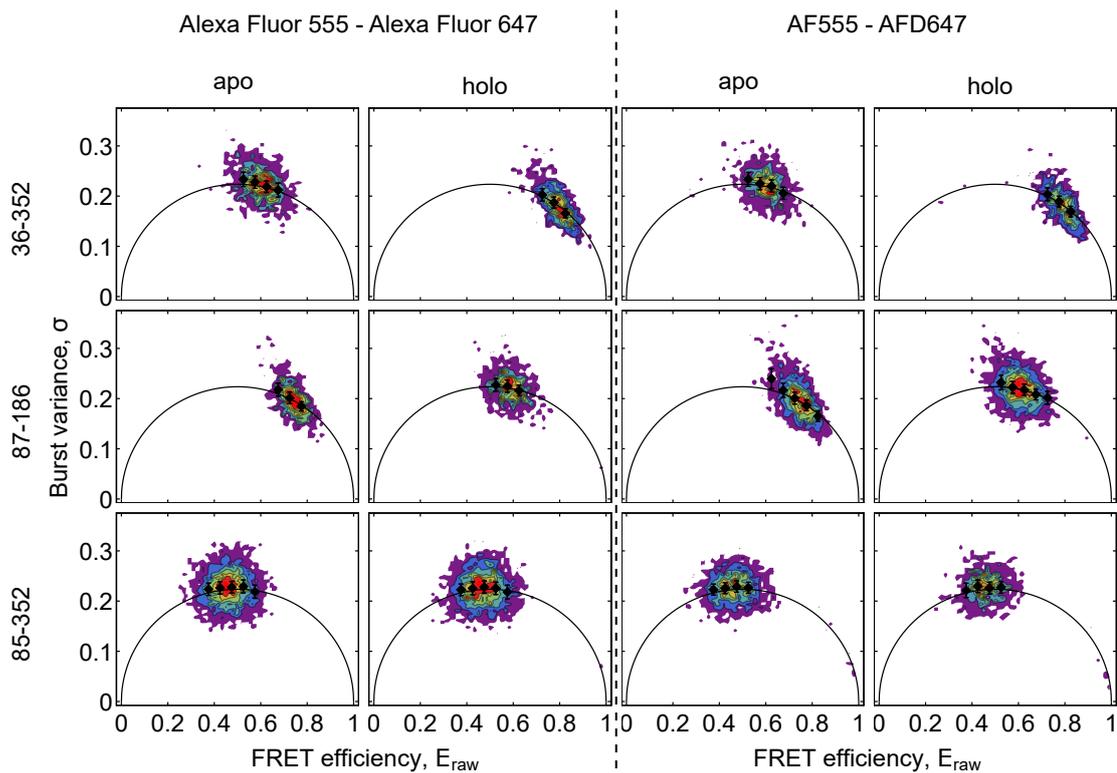


Figure 4.6.12: Burst variance analysis. Burst variance analysis of all FRET examples in their apo and holo states for the Alexa-fluorophore-pair (left) and the AF-fluorophore-pair (right). Data are binned into bins of 0.05 and mean and standard error of mean are shown (black).

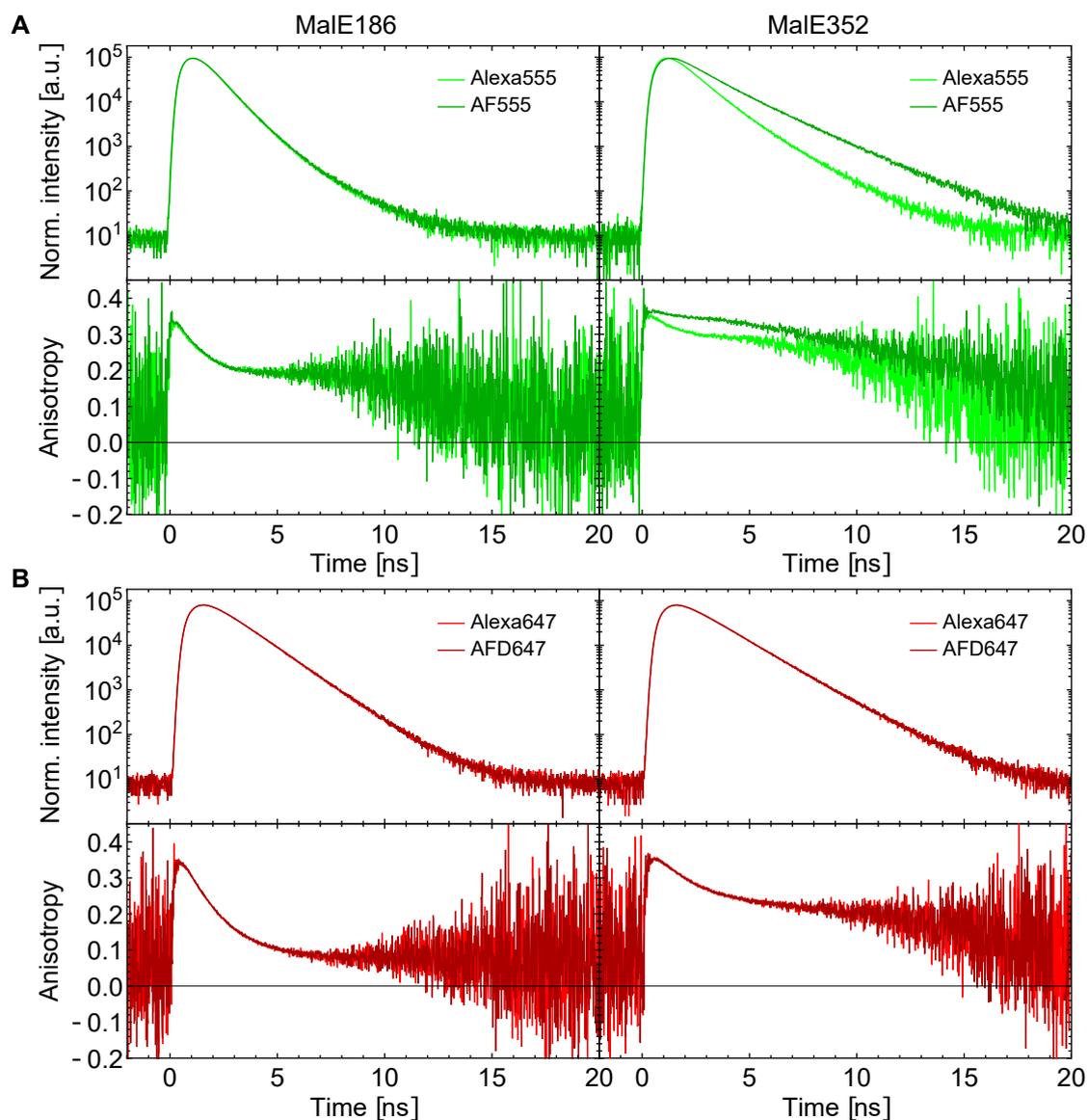


Figure 4.6.13: Characterization of anisotropy and lifetime decays of cyanine dyes on MalE. **A** Lifetime (top) and time-resolved anisotropy measurements (bottom) of Alexa Fluor 555 (lighter green) and AF555 (darker green) labelled at residues 186 (left) and 352 (right) in bound state of MalE with 1 mM maltose in buffer. **B** Lifetime (top) and time-resolved anisotropy measurements (bottom) of Alexa Fluor 647 (lighter red) and AFD647 (darker red) labelled at residues 186 (left) and 352 (right) in unbound state.

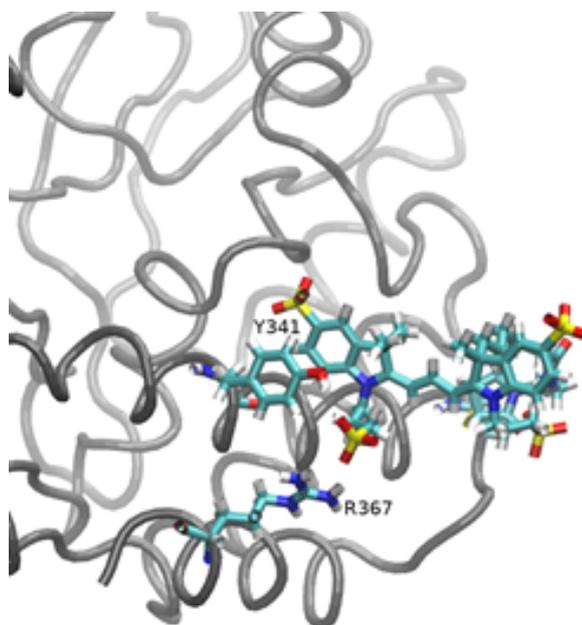


Figure 4.6.14: Fluorophore Alexa Fluor 555 (confirmed structure) attached to mutant S352C of the maltose binding protein. Residues Y341 and R367 as well as the fluorophore are highlighted in stick representation. The depicted picture is a representative snapshot sampled during a 200 ns simulation throughout which frequent hydrogen bond formation (occurrence > 20%) between the oxygen atoms of the SO_3^- group attached to the five-membered ring of the second fluorophore indole ring and the OH group of Y341 or the NH_1 and NH_2 groups of R367 occurs. The sampled distances between the involved acceptor oxygen atoms and the donor atoms of the amino acids are shown in Fig. 4.6.15. In comparison, none of the performed simulations of fluorophores attached to the A186C site presents a hydrogen bond with an occurrence > 20%, involving the oxygen atoms of a SO_3^- group. Moreover, it was found that at the A186C site, Alexa Fluor 555 has significantly less hydrogen bonding interactions with the protein than AF555 (data not shown).

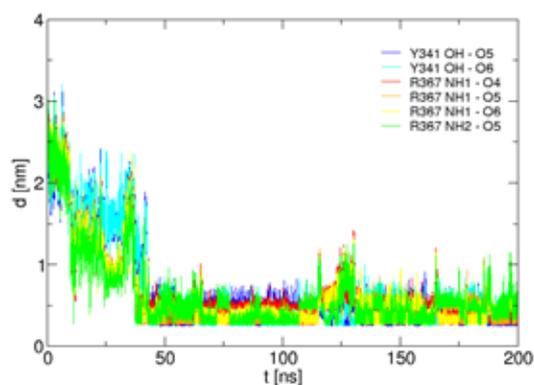


Figure 4.6.15: Distances between selected atoms in MD simulation. Distances between oxygen atoms (O4, O5 or O6) of the SO_3^- group attached to the five-membered ring of the second fluorophore indole ring and the oxygen atom of the OH group of Y341 or the nitrogen atoms of the NH_1 and NH_2 groups of R367 sampled during a 200 ns simulation of fluorophore Alexa Fluor 555 attached to mutant S352C of the maltose binding protein.

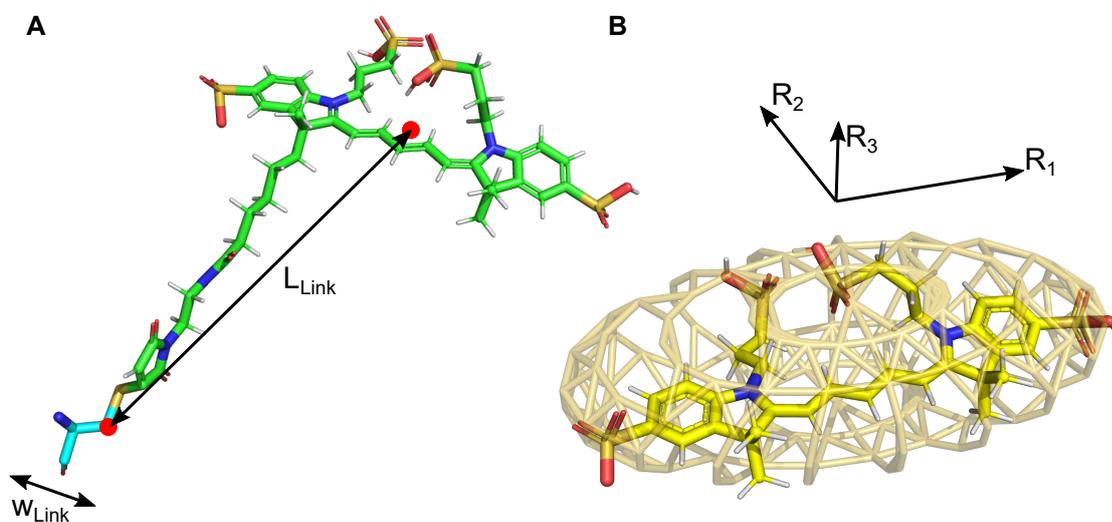


Figure 4.6.16: Fluorophore parametrization for FPS with Alexa Fluor 647. **A** The linker width is set to be 4.5 Å as suggested for comparable linkers⁵. The linker length is measured between the C-Beta atom and the center of the fluorophore core. **B** The fluorophore core is approximated by an ellipsoid with three radii R_1 , R_2 , and R_3 .

5 | Cross-validation of distance measurements in proteins by PELDOR/DEER and single-molecule FRET

Pulsed electron-electron double resonance spectroscopy (PELDOR/DEER) and single-molecule Förster resonance energy transfer spectroscopy (smFRET) are often used to determine conformational changes, structural heterogeneity and also interprobe distances in biological macromolecules. With that they provide qualitative information that facilitates mechanistic understanding of biochemical processes and quantitative data for structural modeling. To provide a comprehensive comparison of the accuracy of PELDOR/DEER and smFRET, we prepared a library of double cysteine variants of four proteins, all of which undergo large-scale conformational changes upon ligand binding. Inter-probe distances in the presence and absence of ligands were determined with either method, using established standard experimental protocols and data analysis routines. We compared our results to distance predictions from structural models. Despite an overall satisfying and similar distance accuracy, some inconsistencies were identified, which we could attribute to the use of cryoprotectants for PELDOR/DEER and label-protein interactions for smFRET. This large-scale cross-validation of PELDOR/DEER and smFRET highlights the strengths, weaknesses, and synergies of these two complementary tools in integrative structural biology. ¹

5.1 Introduction

Since the determination of the first macromolecular structures in the 1950s, our knowledge about their structure and function has dramatically increased. The protein database (PDB, www.rcsb.org) contained more than 178,000 structures at the time of writing. In many cases, multiple PDB entries represent the same macromolecule but in different conformations. The latter illustrates the dynamic nature of proteins, i.e., the presence of large- or small-scale structural fluctuations that are in many cases crucial to their biological function [210–213]. Until now, most macromolecular structures were determined by either X-ray crystallography (~ 90 %), nuclear magnetic resonance (NMR, ~ 10 %), or electron microscopy (EM, ~ 2 %) (<https://www.rcsb.org>). The recent release

¹This chapter was reproduced from Peter/Gebhardt *et al.* [209]. For details of the individual contribution see the “Author Contributions” statement in section 5.4.

of the AlphaFoldDB database provided many further albeit often not yet experimentally verified structures [214]. Undoubtedly, cryo-EM, X-ray crystallography, and alphafold2 predictions can deliver detailed insights into the molecular scaffolds of proteins. Nevertheless, they have the disadvantage that such structures are not determined in liquid solution, but in a crystal lattice, frozen on an EM grid, or even in silico. While the underlying macromolecular dynamics can be inferred by determining multiple structures and combining them into a molecular “movie” [214], this requires additional (biochemical) support to verify the selected order of structural states. Traditionally, the study of such dynamic processes is the strength of NMR. But, NMR is limited to the study of relatively small proteins (typically < 70 kDa; larger homo-oligomers are an exception), which renders the analysis of many proteins unfeasible. Due to these limitations, other “integrative” methods have become increasingly popular in the last decade [215–218]. The idea behind the concept of integrative structural biology is to combine models from either of the three classical approaches with data from e.g., hydrogen-exchange mass spectrometry HDX-MS [219], cross-linking mass spectrometry [220], Förster resonance energy transfer (FRET) [10, 13, 14, 221], small angle X-ray scattering (SAXS) [222] or electron paramagnetic resonance (EPR) in the form of pulsed electron-electron double resonance spectroscopy PELDOR (also known as double electron-electron resonance spectroscopy, DEER) [223, 224].

These orthogonal techniques allow scientists to study conformational dynamics, to visualize conformational heterogeneity, to derive distance constraints between selected residues, and to determine entire contact interfaces even for heterogeneous samples in a near-physiological environment [225]. Such information is often time-consuming or even impossible to obtain with the classical structural biology techniques alone. The hybrid models produced by such integrative approaches can be deposited in the PDB-Dev database [219].

In this study we tested whether two popular integrative methods, single-molecule FRET (smFRET) and PELDOR/DEER spectroscopy, deliver conclusive and consistent results when they are applied to the same proteins. Both techniques are suitable to determine inter-probe distances at the nanometer scale and can also detect conformational changes of macromolecules in their (frozen) solution state. But, the two methods have only rarely been applied to the same macromolecular systems using a large number of identical labelling sites [226–233]. Hence, a systematic comparison of the two methods has been lacking. Here, we provide such a comparative study using the following model proteins: (i) HiSiaP, the periplasmic substrate binding protein (SBP) from the sialic acid TRAP transporter of *Haemophilus influenzae* [232, 233], (ii) MalE, also known as MBP (maltose binding protein) from *Escherichia coli*, which plays an important role in the uptake of maltose and maltodextrins by the maltose transporter, MalEFGK2 [234, 235], (iii) SBD2, the second of two substrate binding domains that are constituents of the glutamine ABC transporter GlnPQ from *Lactococcus lactis* [7, 102, 236, 237] and (iv) YopO from *Yersinia enterocolitica*, a type-III-secretion system effector protein that is injected into macrophages of the host organism where it becomes activated by forming a tight complex with actin [238, 239]. The first three proteins belong to the class of

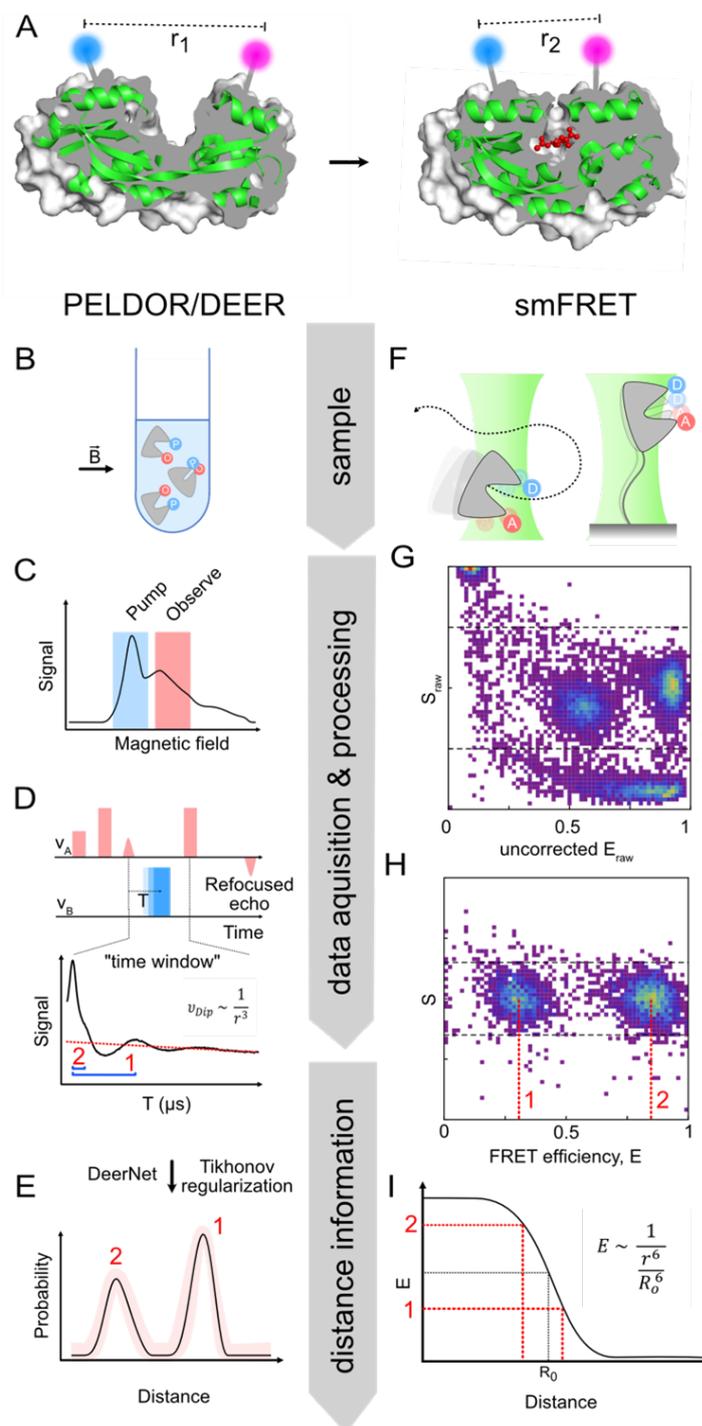


Figure 5.1.1: Following conformational changes of proteins via PELDOR/DEER or smFRET. **A** Two conformations (apo, left and holo, right) of HiSiaP (green cartoon), a substrate binding protein that binds sialic acid (red balls and sticks). A cutaway of the protein surface (grey) is shown to visualize the conformational change of the substrate binding cleft. The position of two labels is indicated by the blue and magenta spheres. **B-E** Workflow of a PELDOR experiment. **F-I** Workflow of smFRET experiments. The individual steps (B-I) are described in detail in the main text.

substrate binding proteins (SBPs) from distinct structural SBP categories [240]. It is well established that upon binding of substrate, SBPs undergo a large conformational shift ($> 10 \text{ \AA}$ for selected residues) from an open unliganded conformation (apo) to a closed conformation (holo) (Figure 5.1.1A) [240–242]. The virulence factor YopO has a kinase- and a guanine nucleotide dissociation inhibitor (GDI) activity, which both interfere with cytoskeletal dynamics of the host. The crystal structure of the GDI-domain alone and that of the YopO/actin complex have been determined [243, 244] and the large conformational changes between apo- and actin bound forms have been studied by PELDOR/DEER in combination with SAXS [245]. The protein is an example of a macromolecule that can be switched between a presumably dynamic apo state and a rigid ligand-bound state.

All our model proteins are “well-behaved” and have been used previously for either PELDOR/DEER or smFRET experiments. The proteins are therefore well suited for our purpose, i.e., to apply standard procedures for distance determination and then to objectively compare the results. Since PELDOR/DEER and smFRET are often independently used to validate structural models, a cross validation is important to objectively judge their accuracy, to gauge the severity of their distinct limitations and to choose the most suitable method for distance measurements for a biological system of interest.

5.2 Results

A brief comparison of PELDOR/DEER and smFRET. Although the methods PELDOR/DEER and smFRET are used for similar applications (Figure 5.1.1A), they both have distinct advantages and disadvantages (See Table 5.2.1 for a side-by-side comparison). The general workflow of each method is illustrated in Figure 5.1.1 and both methods are briefly described below. We refer the reader to reviews and textbooks (e.g. [207, 246–249]) for a comprehensive description of each methods’ theoretical background.

PELDOR/DEER (Figure 5.1.1A-E) is a pulsed EPR (electron paramagnetic resonance) experiment that is used to determine the distribution of distances between two or more paramagnetic centers, such as spin labels, which are attached to a macromolecule (Figure 5.1.1B) [246]. At the outset of the experiment, an EPR spectrum of the sample (usually a frozen solution) is recorded to determine suitable microwave frequencies for the pump- and observer pulses (Figure 5.1.1C). By using two different microwave frequencies, it is possible to selectively excite sub-ensembles of the spin centers in the sample, which then serve as either pump spins or observer spins. This explains why only one type of label is needed. During the actual PELDOR/DEER experiment, the electron spin echo signal (“refocused echo”) is produced by the observer pulse sequence (Figure 5.1.1D). If the pump- and observer spins are coupled via a dipole-dipole interaction, the pump pulse sequence (blue square, Figure 5.1.1D) causes characteristic oscillations of the refocused echo. A plot of its intensity for different times T is called PELDOR/DEER “time trace” (Figure 5.1.1D, bottom). The oscillation frequencies of the time trace encode the magnitudes of the dipolar coupling constant that is inversely proportional to

Structural biology question	PELDOR/DEER	smFRET
Which range of distances can be measured?	Normally 15-80-Å, but up to 160-Å with fully deuterated proteins	Normally 30-80-Å; 100-150-Å are possible with multiple acceptor dyes
How many types of labels are typically needed?	1	2*
Can multimeric (homomeric) proteins be studied?	Yes. Multiple distances can be measured in one experiment, using just one type of label and standard equipment.	Yes, but technically demanding.
Amount of sample needed?	For Q-band (standard frequency) measurements: 80-µl of ~ 10-30-µM labelled protein. Measurements with sub-µM spin concentrations have been performed [250].	100 – 400-µl of 15-100-pM labelled protein
Physical state of the sample?	Usually frozen solution (50 K). Measurements in aqueous solution are possible but require specific labels (e. g. trityl) and very large or immobilized macromolecules (e.g. [251]). Depending on the label, other pulse sequences than PELDOR/DEER might be required.	Liquid solution at room temperature or cell culture conditions (e.g., 37°C)
In vivo measurements?	Not a standard experiment, especially not under physiological conditions. But, measurements in manually injected frog oocytes or using paramagnetic unnatural amino acids in <i>E. coli</i> have been done [252, 253].	Yes
Time resolution	Freeze quenched samples can be measured. The time-resolution depends on the freeze quench equipment. Microsecond times have been reached with such equipment [254].	Down to micro- and nanoseconds
Time frame for measurements	Normally several hours per measurement for Q-band	Diffusing molecules: 30-60 minutes Immobilized molecules: minutes to hours

*This number does not consider homoFRET approaches where identical fluorophores are used in combination with fluorescence depolarization experiments [255].

Table 5.2.1: Common practical questions concerning the applicability of PELDOR/DEER and FRET to structural biology questions.

the third power of the inter-spin distance (blue brackets in Figure 5.1.1D, bottom). Once recorded, the time-traces can be converted into distance distributions (Figure 5.1.1E) by applying Tikhonov regularization [256, 257] or by using artificial neural networks such as DeerNet that were trained on large sets of simulated data [258]. While PELDOR/DEER is by far the most used experiment for pulsed EPR distance measurements, other pulse sequences have been developed (DQC, SIFTER, RIDME [259]). Their applicability to a particular system of interest depends on the spin centers that are used in the experiment (see below). Distance determination via smFRET (Figure 5.1.1A) is performed in liquid solution on single diffusing or immobilized molecules (Figure 5.1.1F) and relies on the dipole-dipole coupling between two spectrally distinct fluorophores to determine the efficiency of non-radiative energy transfer from the electronically excited donor to the acceptor (Figure 5.1.1G). The energy transfer efficiency depends on the presence of isoenergetic transitions in both molecules (donor emission and acceptor absorption), their relative orientation and the distance between the fluorophores and must be corrected for setup-dependent parameters when a ratiometric, i.e., intensity-based approach is used (Figure 5.1.1H). Accurate FRET efficiencies E can then be converted into distances r using the Förster equation (Figure 5.1.1I), which requires knowledge of the Förster radius R_0 (inter-probe distance with $E = 0.5$). An overview of correction parameters for conversion of setup-dependent to accurate FRET efficiencies and distances is provided in Supplementary Table 5.6.1. Because most proteins are diamagnetic and devoid of any suitable fluorophores, both PELDOR/DEER and smFRET experiments usually require the attachment of spin- or fluorescence labels (Figure 5.1.1). This can be accomplished by the site-specific introduction of cysteines. The sulfhydryl groups of cysteines can be reacted with functionalized labels containing maleimides or thiosulfate esters (see Figure 5.2.1 for some typical examples). If the introduction of cysteines is not possible, alternative labelling approaches such as labelled nanobodies [260] or unnatural amino acids can be used [252, 261–263]. The latter can either be fluorescent or paramagnetic themselves or bear functional groups that can be labelled, for instance by click-chemistry [252, 261–263]. Although the types of labels used for PELDOR/DEER and FRET are quite different, the requirements for suitable labelling positions in proteins are essentially the same: the residue should be solvent-accessible and its labelling should not impair folding or functional properties. For PELDOR/DEER spectroscopy the distance between the labels ought to be in the range of 1.5 to 8.0 nm (longer distances of up to 16 nm are usually only accessible with fully deuterated samples) [246, 264, 265]. The ideal distance for FRET experiments is around the Förster radius of the selected donor-acceptor pair. This provides a typical dynamic range between 3–8 nm (Figure 5.1.1), but in principle, also longer distances up to 10–15 nm are accessible [266]. Usually, labelling positions are chosen such that the distance change between conformations is as large as possible. In practice, the pool of suitable labelling sites is often surprisingly small. Fortunately, software programs exist to assist in the identification of optimal labelling positions in the case of an available structure or model of the target protein [24, 113, 198, 206, 267].

Our aim for the comparison was to choose commonly used labels and well-established

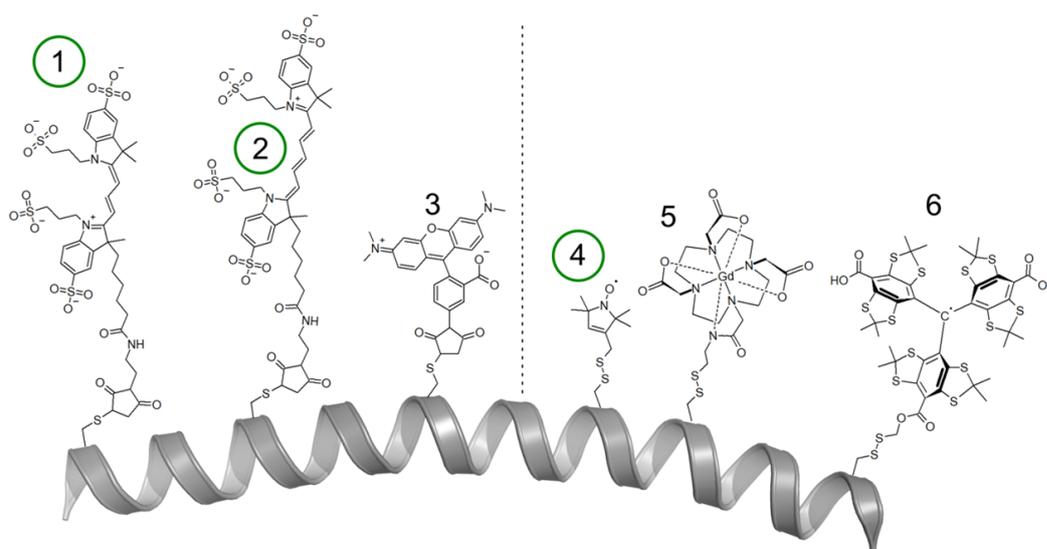


Figure 5.2.1: Chemical structures of labels commonly used for smFRET (left) and PELDOR/DEER (right). Green circles identify the commonly used probes for the two methods. (1) Maleimide-thiol adducts of Alexa Fluor 555 [113] (2) Alexa Fluor 647 [113] were used in the present study. (3) TMR (tetramethylrhodamin-5-maleimide) [24] and (4) the MTSSL- [268], (5) DOTA-Gd- [269] and (6) trityl-spin labels [270, 271] attached to a cysteine residue via a disulfide bridge. The polypeptide chain is represented as a grey cartoon. Note that many other labels with differing coupling chemistries ranging from click-chemistry to unnatural amino acids have been developed, as well as labels that are specific for nucleic acids.

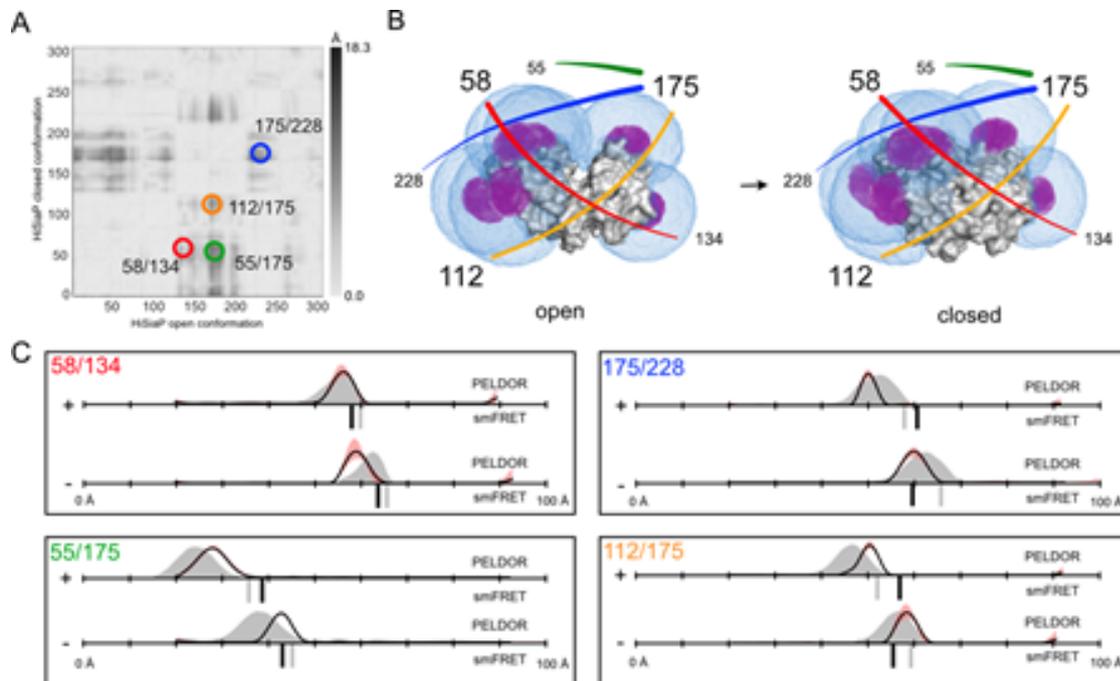
experimental conditions for either method. The PELDOR experiments were thus performed at 50 K with cryo-protected samples using a commercial pulsed Q-band spectrometer. The samples were labelled with MTSSL (S-(1-oxyl-2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl methanesulfonothioate) [268]. The distance distributions were determined using Thikonov regularization [257] and DeerNet [258, 272] and the distance predictions were calculated with mtsslWizard [273]. The smFRET experiments of diffusing protein molecules were conducted in buffer at room temperature using standard procedures suitable for microsecond alternating laser excitation (ALEX) as described before [113]. The experiments were performed with Alexa Fluor 555 (donor) and Alexa Fluor 647 (acceptor) using a homebuilt confocal microscopy setup with 2-colour excitation/detection [113].

Comparison 1: Sialic acid binding protein HiSiaP. Figure 5.2.2A shows a difference distance map of HiSiaP, based on the open- and closed crystal structures. The map represents all distance changes between the C atoms of the two states. We picked pairs of sites with pronounced distance changes (dark areas) of up to 1.8 nm for labelling. Figure 5.2.2B shows the open (PDB-ID: 2CEY) and closed structures (PDB-ID: 3B50) of the protein with the predicted accessible volumes of the spin- (magenta) and FRET-

(blue) labels at the selected labelling sites (residues 55, 58, 134, 175, 228). For PELDOR/DEER, all double variants (58/134, 55/175, 175/228 and 112/175) were labelled with MTSSL (SI Figure 5.6.1). In each case, two PELDOR/DEER measurements were performed, one in the presence and one in the absence of sialic acid (Neu5Ac). Similar to the previously published PELDOR/DEER data for the *Vibrio cholerae* homolog VcSiaP [108, 274], which shares 49 % amino acid sequence identity and 69 % sequence similarity with HiSiaP, the PELDOR/DEER-time traces obtained for HiSiaP were of excellent quality with clearly visible oscillations and high signal to noise ratios (SNR, SI Figure 5.6.2). The distance distributions had a single, well-defined peak (Figure 5.2.2C, black curves) with very small uncertainties (red shades) and a clear shift towards shorter distances in the presence of substrate. The corresponding *in silico* predictions, based on the crystal structures (Figure 5.2.2C, grey areas), were in good agreement with the experimental PELDOR/DEER results, considering the known error of $\pm 2\text{-}4$ Å for such predictions. This error margin is mainly due to difficulties in correctly predicting the rotameric state of the spin label, as discussed below and in reference [273]. In summary, for each double variant, distance changes were measured that were similar in magnitude to those calculated from the crystallographic models, and in agreement with those from the previously published VcSiaP data [274].

We next assessed substrate-induced conformational changes in the HiSiaP double variants by smFRET spectroscopy using Alexa Fluor 555 and Alexa Fluor 647 as donor and acceptor dye, respectively (Figure 5.2.1). This popular FRET pair was chosen for its high photostability, signal intensity and proven compatibility with various protein samples [102, 107, 277]. Labelling quality and sample purity were assessed by size exclusion chromatography (SEC). All samples showed high labelling efficiencies ($> 90\%$) and donor-to-acceptor labelling ratios up to $\sim 1:1$ (SI Figure 5.6.3). The experiments were conducted with freely diffusing molecules at ~ 50 pM concentration to derive mean FRET-efficiency values for the apo- and holo states of HiSiaP. All FRET measurements gave high quality ES-histograms with clearly defined populations (SI Figure 5.6.4). However, some of the donor-acceptor populations appeared to be broader than expected, indicating that they were either composed of molecules with conformational flexibility, or there were unwanted photophysical effects arising from the choice of fluorophores and labelling positions (see discussion below).

Figure 5.2.2C summarizes the FRET distance measurements in direct comparison with the PELDOR/DEER distance distributions. For smFRET (black bars) only variants 58/134 and 55/175 gave the expected trend for shorter inter-probe distances in the presence of ligand. The two other variants (175/228, 112/175) failed to reproduce the trends from PELDOR/DEER and structural predictions (grey bars). Instead, smFRET data suggested that the apo protein adopted a conformation that was “more closed” than the substrate-bound conformation. Since variants 58/134 and 55/175 agreed with both the PELDOR/DEER results and the models based on the crystal structures, we thought it unlikely that a completely unexpected structural feature of the protein was responsible for the observed discrepancies. Considering the known ± 5 Å experimental accuracy of FRET [111], one might argue that the two states were simply not discernable for the



“offending” double variants, (contradicting the simulation results in Figure 5.2.2C). To examine whether HiSiaP was undergoing FRET dynamics in the apo state, we carried out a burst-variance analysis (BVA) [65] of three HiSiaP variants (SI Figure 5.6.5). In BVA the shot-noise-limited standard deviation for a given mean FRET efficiency (STD of FRET, SI Figure 5.6.5) is compared against the experimental standard deviation within the burst related to statistical noise (SI Figure 5.6.5, black solid half circle). The results showed that the protein exists in stable FRET states and does not switch rapidly between distinct states (on the millisecond timescale). As an example, a DNA hairpin structure undergoing such changes is shown in SI Figure 5.6.5. These results however do not rule out transitions on a timescale below ~ 500 μ s and the characterization of such dynamics would require pulsed interleaved excitation (PIE) or multi-parameter fluorescence detection (MFD) analysis [18]. A possible explanation for the discrepancy between crystal structures and smFRET distances was that the fluorescence labels were partly immobilized by an interaction with a surface feature of the protein. For instance, the sulfonic acid groups of the fluorophores could interact with positively charged patches on the protein surface. Because these effects are highly location dependent, the stochastic labelling combination of donor-acceptor pair and acceptor-donor pair might result in a heterogeneous mix of two different types of labelled proteins, where the fluorophores “feel” a different environment depending on either cysteine. Interestingly, we could observe such a broadened population [150] caused by these two labelling combinations very clearly for a distinct dye-combination (Alexa Fluor 546 – Star 635P) for variant 112/175 (SI Figure 5.6.6). Notably, we found this behavior of broadened apo populations for many HiSiaP variants suggesting fluorophore interactions with the protein surface (SI Figure 5.6.4 and SI Table 5.6.2).

To explore the possibility of such unwanted dye-protein interactions, we investigated fluorescence anisotropy and lifetime decays of labelled HiSiaP for two different amino acid positions with a variety of fluorophores. FRET remains a reliable distance ruler for the scenario that at least one of the two fluorophores undergoes free rotation, which is characterized by fast decay of initial anisotropy values and low residual anisotropies. Because the smFRET results for the variant 58/134 were in good agreement with the simulations, we selected the single variant at position 58 as a positive control, where we expected low dye-protein interactions. And because position 175 was involved in both double variants that showed unexpected mean FRET values and broad FRET distributions, particularly in the apo state, the single variant at this position was chosen as a likely negative example.

The anisotropy decays observed for Alexa Fluor 555 revealed interactions of the dye with the protein at both positions (175 and 58) in the apo/olo states with residual anisotropies at long delay times just below 0.3 (Figure 5.2.3A). Yet, Alexa Fluor 555 was fully immobile in the apo conformation for residue 58 (Figure 5.2.3A, apo). Here, we could not identify a short decay component of the time-resolved anisotropy signal on the timescale of fluorophore rotation < 1 ns. For Alexa Fluor 647, both positions showed smaller residual anisotropies (Figure 5.2.3B), but also revealed a distinction between apo- and olo state for position 175. In addition, slower anisotropy decays were

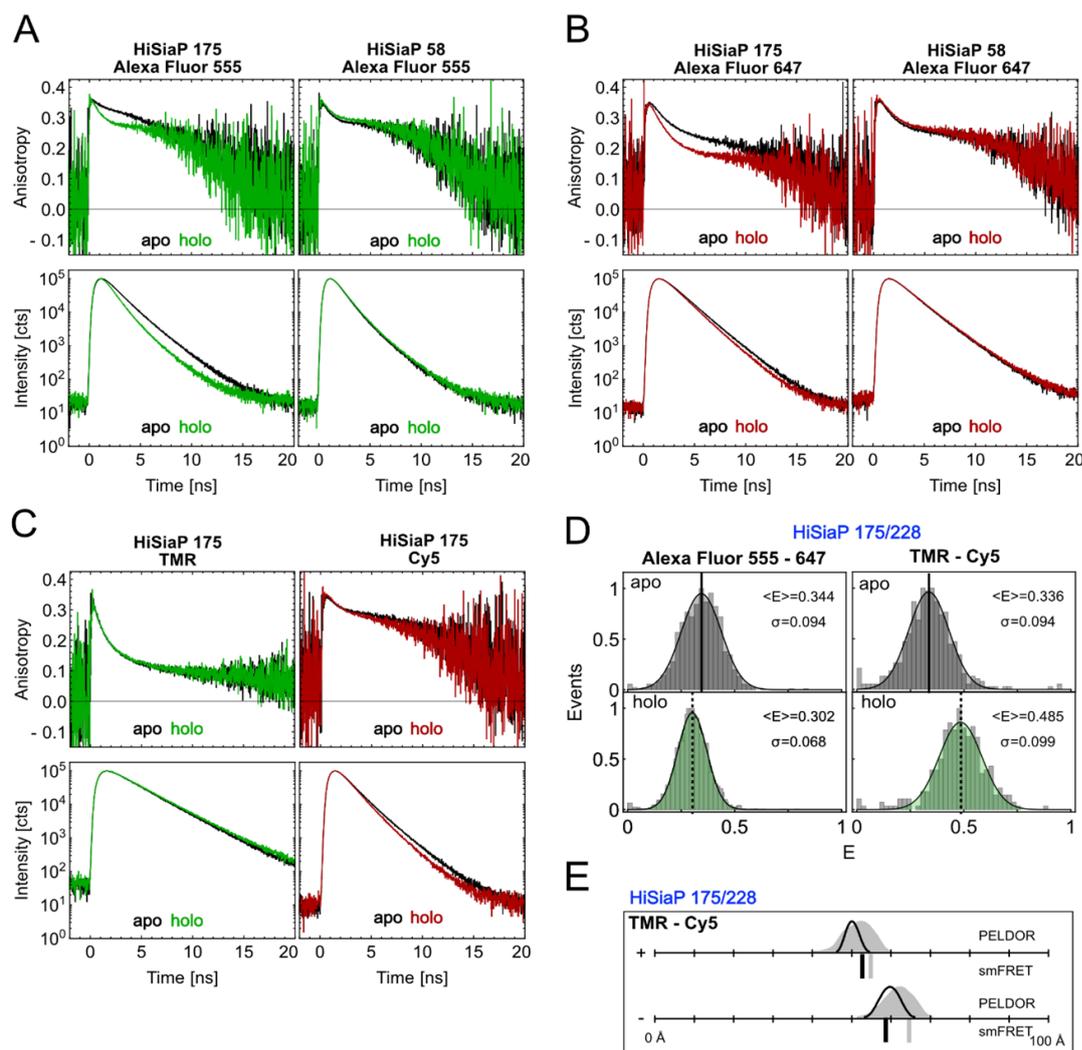


Figure 5.2.3: Time resolved fluorescence anisotropy and lifetime measurements on HiSiaP. **A** Anisotropy decay curves of Alexa Fluor 555 (top row) at residue 175 (left) and 58 (right) and lifetime decay curves (bottom row) under magic angle conditions for apo (black) and holo state (green). **B** Same measurements as in A with Alexa Fluor 647 in apo (black) and holo state (red). **C** Anisotropy decay curves of TMR (top left) and Cy5 (top right) at residue 175 and lifetime decay curves of TMR (bottom left) and Cy5 (bottom right) under magic angle conditions for apo (black) and holo state (colored). **D** FRET efficiency distributions (center & bottom) of HiSiaP variant 175/228 for Alexa Fluor 555 – Alexa Fluor 647 (left) and TMR – Cy5 (right) in apo (grey) and holo state (green). **E** Converted distances from the mean FRET efficiencies are shown as black bars in comparison to simulation (grey bar) and PELDOR/DEER results from Figure 5.2.2.

accompanied by an increase in the fluorescence lifetime of the fluorophores, a finding that was more pronounced for Alexa Fluor 555 [201] than for Alexa Fluor 647 (compare apo/holo decays at position 175, Figure 5.2.3).

The described photophysical effects with fluorophores at position 175 can explain two observations in our measurements: Firstly, the larger lifetime increases of Alexa Fluor 555 compared to Alexa Fluor 647 would lead to a broadening of the FRET distribution, because having either the donor or the acceptor at position 175 would result in two distinct FRET states. Secondly, the changes in lifetime (quantum yield), orientation and fluorophore disposition experimentally change the Förster radius and thus impact the proper conversion of FRET-efficiency to distance. To avoid these problems, we altered the FRET fluorophore pair to TMR and Cy5 and repeated the full set of experiments. In contrast to Alexa Fluor 555 and Alexa Fluor 647, these fluorophores are not negatively charged and the linker of TMR is significantly shorter as compared to Alexa Fluor 555 (Figure 5.2.1). The TMR/Cy5 label pair was not our first choice, because it is inferior to Alexa Fluor 555/647 in terms of signal intensity and photostability. Also, for many proteins, charged labels are known to be less prone to stick to the protein surface than hydrophobic labels. In anisotropy and lifetime measurements on the 175 variant in apo and holo state (Figure 5.2.3C), however, TMR showed almost ideal behavior with high rotational freedom and a fluorescence lifetime that was unaffected by the conformational state of HiSiaP (Figure 5.2.3C). Cy5 revealed a slightly longer fluorescence lifetime in the apo state. In smFRET experiments using variant 175/228 with TMR/Cy5, this reduced fluorophore-protein interaction resulted in qualitative consistency between FRET-derived and simulated distances (Figure 5.2.3D/E).

Taken together, this meant that with the TMR/Cy5 fluorophore pair, the expected substrate-induced domain closure was observed (Figure 5.2.3D/E), and the absolute distance measurements were in much better agreement with the crystal structures. Thus, it appears that the unexpected FRET results with protein variants labelled at position 175 and the discrepancy between the experimentally determined and simulated distance for the apo protein, were indeed due to interactions of both donor and acceptor fluorophores with the protein surface at this position. It should be noted that also for spin labels, intricate protein/label interactions are known to occur and have been shown to explain initially puzzling results [278].

Comparison 2: Maltose binding protein MalE. MalE has previously been studied by smFRET using Alexa Fluor 555 and Alexa Fluor 647 to elucidate the transport mechanism of maltose ABC importer MalFGK2-E (SI Figure 5.6.7/5.6.8) [107]. In the present work, we used four MalE cysteine double variants (87/127, 36/352, 29/352 and 134/186) with distinct ligand induced distance changes; see difference distance matrix in Figure 5.2.4A. Variants 36/352, 29/352 were designed to show a decrease of distance upon maltose addition. For the 87/127 variant, the labels were located on the opposite surface of the protein to the substrate binding site near the hinge region, and therefore the expectation from the crystal structures was that the distance would increase for the holo state compared to the apo state. We also included a negative control (134/186),

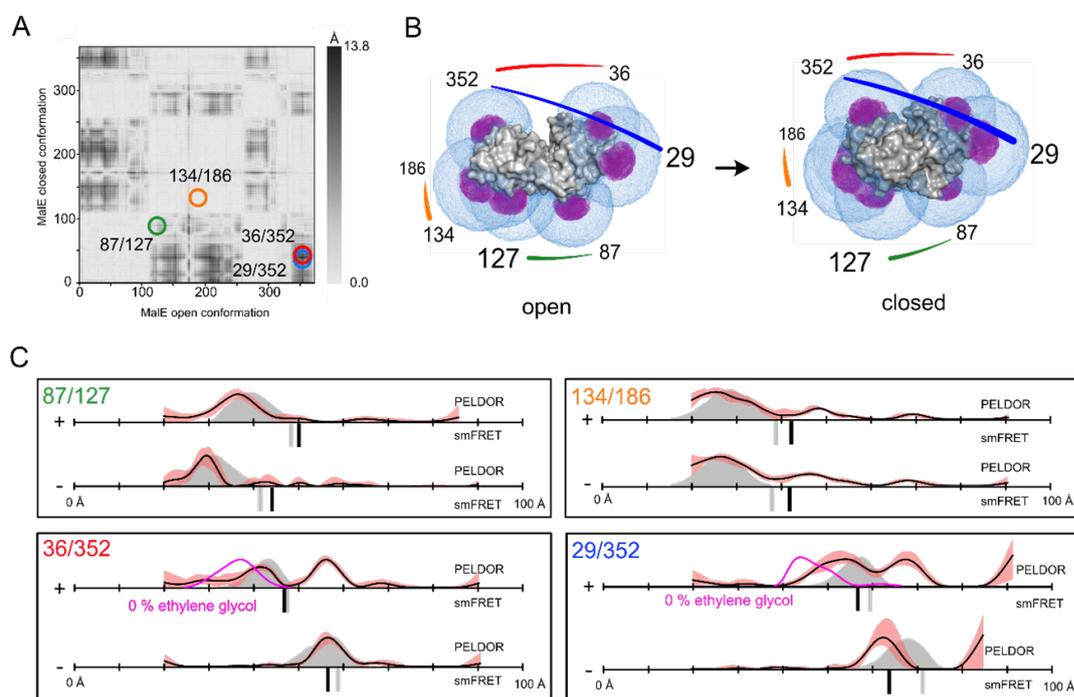


Figure 5.2.4: Distance measurements on MalE. **A** Difference distance map of MalE in the open (PDB-ID: 1OMP [174]) and closed (PDB-ID: 1ANF [82]) conformation [276]. Protein regions with high conformational changes are indicated as dark spots. The double variants for distance measurements are marked with circles. **B** Surface presentation of MalE (grey) in the open (left) and closed (right) conformation. The accessible volume of the spin label on seven different labelling positions, calculated with mtsslWizard, is represented by magenta meshes. The accessible volume of FRET label maleimide-Alexa Fluor 647, calculated with FPS [206] is shown as blue meshes. **C** Distance measurements with four different double variants of MalE without (-) and with (+) substrate. The PELDOR/DEER results are shown above (grey curves for simulation, black curves for experiment) and the FRET distance below the axis (grey bars for simulation, black bars for experiment). PELDOR/DEER results without cryoprotectant are shown as magenta curves. The red shade around the PELDOR/DEER data is the error margin calculated using the validation tool of DeerAnalysis [257]. The underlying principle of the validation tool is explained in the discussion section below.

in which the two labels were in the same domain of the protein (Figure 5.2.4C) with the expectation that no substrate-induced distance change occurs (Figure 5.2.4C). All variants showed very good agreement between experimental and simulated values in smFRET experiments. Ligand binding was confirmed for all variants using microscale thermophoresis [113].

PELDOR/DEER distance measurements were performed on the same set of variants using MTSSL (Figure 5.2.4 and SI Figure 5.6.1). The phase memory times of the MalE samples were significantly shorter than for the HiSiaP samples (SI Figure 5.6.16). Nev-

ertheless, it was possible to measure time traces with sufficient length to resolve the expected distances, albeit at a lower SNR compared to the HiSiaP samples (see Discussion and SI Figure 5.6.9). Three of the four MalE variants yielded good-quality PELDOR time traces. The 87/127 variant had a relatively low modulation depth and SNR but still provided data of sufficient quality (SI Figure 5.6.9). For all variants, except 29/352, the measured distances closely matched the predictions obtained for the crystal structure of the apo form of the protein (Figure 5.2.4C). Notably, this variant also had the worst match between the simulation and experiment for the smFRET experiments. The addition of 1 mM maltose to all four variants had different effect of the corresponding PELDOR distance distributions. As expected, the ligand did not significantly change the position of the distance peak for 134/186 variant (our negative control). The distance distribution obtained for the variant 87/127 suggested that this variant adopts a closed state. However, the distance distributions obtained for the variant 36/352 corresponded to a mixture of the holo- and apo states. A similar result was obtained for variant 29/352 with the only difference that the distance assigned to the apo state was ~ 5 Å longer than the one obtained from the experiment without maltose. Since the SNR of the PELDOR/DEER time trace for the 87/127 variant was low and the distance change between open and closed conformation was relatively small, we cannot exclude the possibility that this variant also existed in an open-closed mixture after substrate addition (SI Figure 5.6.9).

Because binding constants are temperature dependent, and the PELDOR/DEER samples were frozen, we checked, whether complete closure of MalE was achievable at a higher substrate concentration of 10 mM maltose. Within error, these experiments yielded the same mixtures of the holo- and apo states as seen with 1 mM maltose (SI Figure 5.6.9). Since the lack of complete closure did not result from a sub-saturating maltose concentration in the frozen samples, and it was not observed in the smFRET data, we reasoned that perhaps the cryoprotectant used for PELDOR/DEER experiments might be the culprit. Figure 5.2.4C shows the PELDOR/DEER results for variants 36/352 and 29/352 in the presence of 1 mM maltose, and in the presence (black lines) or absence (magenta lines) of 50 % ethylene-glycol cryoprotectant. Further measurements with 25 % ethylene-glycol and glycerol are given in SI Figure 5.6.10. Interestingly, when the ethylene glycol concentration was lowered to 25 %, the closed state of MalE became more dominant and in the absence of cryoprotectant, the closed state was the only state that was detected. Note that for the measurements without cryoprotectant, the length of the PELDOR/DEER time traces had to be shortened to achieve a sufficient SNR. Inevitably, this made the corresponding distance distributions less reliable for longer distances (see Discussion). However, SI Figure 5.6.10 shows that also the shape of the time traces themselves were clearly different, clearly revealing the influence of the cryoprotectant on our measurements.

In summary, for MalE, both methods were able to detect the substrate-induced closure of the protein. A reasonable consistency between the two methods and also in relation to structure-based predictions was found. The cryoprotectant used in PELDOR/DEER measurements was identified as the source of initial discrepancies.

Comparison 3: Glutamate/Glutamine binding protein SBD2. Previously to this work, we studied the SBD2 domain of the GlnPQ amino acid transporter by smFRET spectroscopy to elucidate its binding mechanism and its involvement in amino-acid transport [102]. Here, we conducted smFRET experiments to determine accurate FRET efficiencies for variants 319/392 and 369/451 using Alexa Fluor 555 and Alexa Fluor 647 (SI Figure 5.6.11/5.6.12). The experimental FRET distances were in good agreement with the in silico predictions (Figure 5.2.5C), where both apo and holo proteins showed a single population. In contrast, in the PELDOR/DEER experiments, the apo form of both SBD2 variants displayed at least two prominent distance peaks (Figure 5.2.5C and SI Figure 5.6.13).

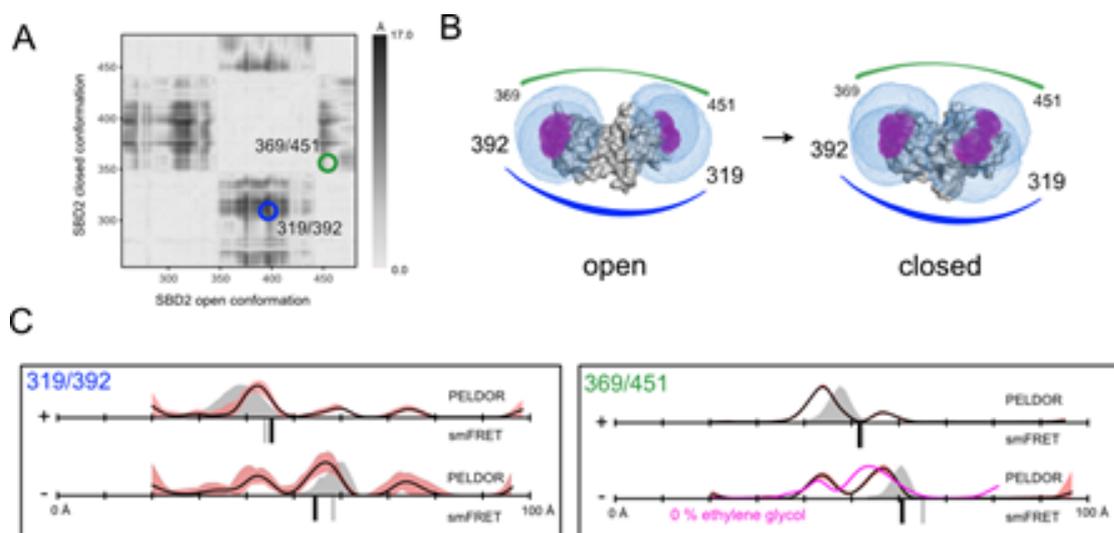


Figure 5.2.5: Distance measurements on SBD2. **A** Difference distance map of SBD2 in the open (PDB-ID: 4KR5 [236]) and closed (PDB-ID: 4KQP [236]) conformation. Protein regions with high conformational changes are indicated as dark spots. The double variants for distance measurements are marked with circles. **B** Surface presentation of SBD2 (grey) in the open (left) and closed (right) conformation. The accessible volume of the spin label at four different labelling positions, calculated with mtsslWizard, are represented by magenta meshes and the accessible volume of FRET label maleimide-Alexa Fluor 647, calculated with FPS [206] is shown as blue meshes. **C** Distance measurements with two different double variants of SBD2 without (-) and with (+ = 1 mM) substrate. The PELDOR/DEER results are shown above (grey curves for simulation, black curves for experiment) and the FRET distance below the axis (grey bars for simulation, black bars for experiment). PELDOR/DEER results without cryoprotectant are shown as magenta curves. The red shade around the PELDOR/DEER data is the error margin calculated using the validation tool of DeerAnalysis [257]. The underlying principle of the validation tool is explained in the discussion section below.

The in silico predictions (grey curves) indicated that these distance peaks can be assigned to the open- and closed conformations. After addition of the substrate glutamine to the protein sample, the relative ratio between the two PELDOR/DEER distances

shifted towards shorter distances. Thus, the closed conformation became dominant, but at least ~10 % of the protein remained in the open conformation (Figure 5.2.5C). In this case and within error, the removal of the cryoprotectant had no significant effect on the PELDOR/DEER results (the time traces of the variant 369/451 with cryo-protectant are shown in SI Figure 5.6.10).

To investigate whether the presence of the closed conformation of the apo protein seen for both variants was due to co-purified glutamine, we performed liquid chromatography mass spectrometry (LC-MS) experiments. Evaluation of the supernatant from purified variants showed that they did not contain detectable glutamine traces (μM concentrations would be needed to explain our observations, SI Figure 5.6.14).

In summary, for the SBD2 protein, both methods were able to discern the open and closed state. But, the reason for the differences between the PELDOR/DEER and smFRET experiments remained elusive.

Comparison 4: Type-III-secretion system effector protein YopO with a dynamic apo state. PELDOR/DEER experiments are almost always conducted on bulk frozen samples and it is then not possible to study real-time molecular motions. However, dynamic processes can be studied indirectly by interpreting the width of distance distributions or by using freeze-quench experiments [254, 279]. Recently, a combination of PELDOR/DEER and SAXS was used to investigate the conformational flexibility of the YopO protein from *Yersinia enterocolitica* (Figure 5.2.6A) [245]. The study revealed that the YopO/actin complex is a rigid entity with rather sharp PELDOR/DEER distance distributions, while the apo form of YopO adopts multiple conformations leading to a broader distance distribution (Figure 5.2.6A/B/D). This is reflected in the PELDOR/DEER time trace for apo YopO where a sharp initial decay and a more damped oscillation is observed. These distinct features are due to the broad distribution of shorter inter-probe distances in the sample. By generating molecular models of the protein that simultaneously explained the PELDOR/DEER data and SEC-SAXS curves it was possible to obtain a coarse-grained insight into possible conformations of the apo protein [245]. This approach could however not answer the question whether the generated structures are stable individual conformations of the apo protein or rather conformational states that each individual protein samples in a short period of time, i.e., on the sub-millisecond scale.

Such questions can be addressed by smFRET experiments. We hence labelled the YopO double variant 113/497 with the Alexa Fluor 555 / 647 dye combination and conducted smFRET measurements of YopO in the presence and absence of actin (Figure 5.2.6C/D). Indeed, a clear shift from a high (with actin) to a low (without actin) FRET efficiency was observed. Interestingly, also the width of both FRET distributions was wider than expected (Figure 5.2.6C), i.e., for YopO apo by 2.3-fold and 1.4-fold for holo (Table 5.6.2). We next employed BVA [65] (Figure 5.2.6E/F) to investigate, whether in contrast to HiSiaP (see above and SI Figure 5.6.5), sub-millisecond dynamics in apo YopO were responsible for this broadening. In BVA the shot-noise-limited standard deviation for a given mean FRET efficiency (STD of FRET, Figure 5.2.6E/F) is compared

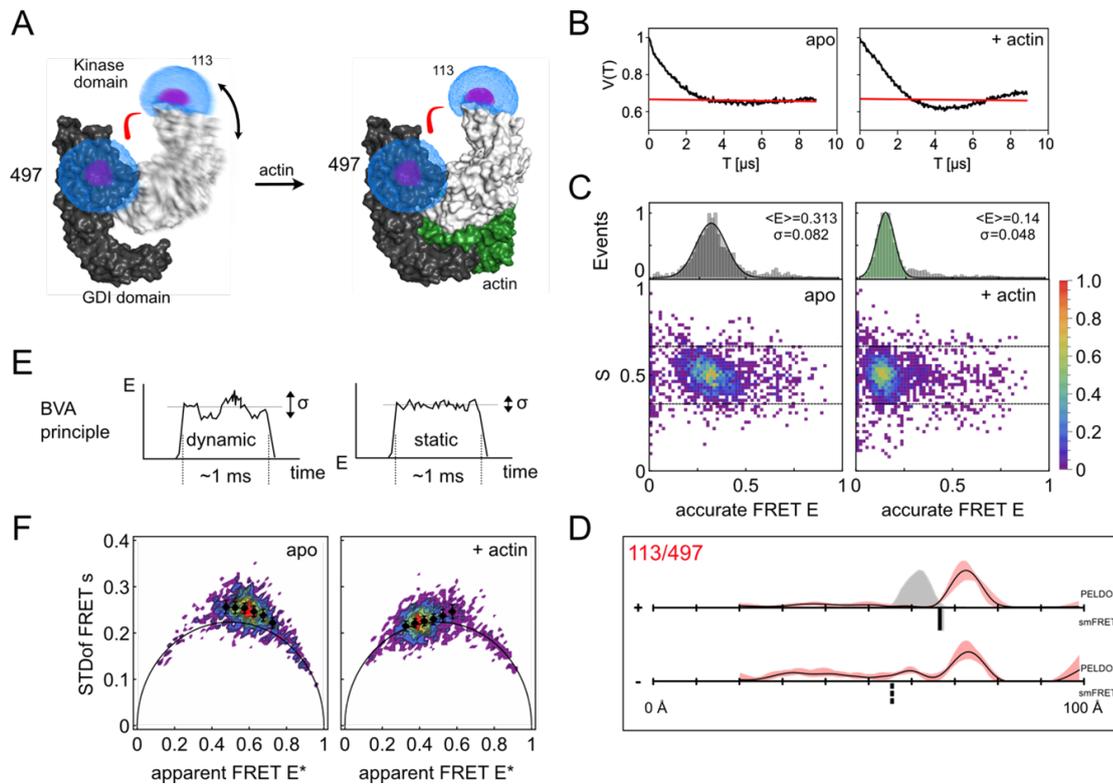


Figure 5.2.6: Distance measurements on dynamic system YopO. **A** Model of the stabilisation of YopO (kinase domain in grey, GDI domain in black) by binding of actin (green) (PDB-ID: 4CI6). The accessible volume of the spin label at two different labelling positions, calculated with mtsslWizard, are represented by magenta meshes and the accessible volume of FRET label maleimide-Alexa Fluor 647, calculated with FPS [206] is shown as blue meshes. **B** PELDOR time traces of spin labelled variant YopO89-729 113R1/497R1 without and with actin from our previously published study [245] (reproduced with permission). The red line indicates the background correction from DeerAnalysis 50. **C** smFRET data for Alexa Fluor 555 – Alexa Fluor 647 YopO89-729 113/497 in the presence and absence of actin. **D** Previously published distance measurements with the YopO double variant without (-) and with actin [245]. The PELDOR/DEER results are shown above (grey curves for simulation, black curves for experiment) and the FRET distance below the axis (grey bars for simulation, black bars for experiment). The red shade around the DEER data is the error margin calculated using the validation tool of DeerAnalysis [257] (reproduced with permission). The underlying principle of the validation tool is explained in the discussion section below. **E** Simplified schematic explaining the principle of the burst variance analysis: Dynamic systems show an increased variance in FRET efficiency during the measurement period (left) compared to the purely shot-noise limited variance of a static sample (right); please note that exchange of conformational states on timescales much faster than 100 μ s can also give rise to an apparent static behaviour of the burst in BVA. **F** Burst variance analysis of the YopO measurement in C reveals a dynamic FRET state in the absence of actin (left) and a stabilized state in the presence of actin (right).

against the experimental standard deviation within the burst (Figure 5.2.6F). In accordance with PELDOR/DEER experiments (Figure 5.2.6B/D [245]), smFRET suggests a dynamic apo state with sub-ms conformational dynamics and a static actin-bound conformation of YopO. This can be seen in the BVA plots where only deviations of the STD of FRET from the static line are seen for apo YopO (Figure 5.2.6F) similar to a dynamic DNA hairpin structure and in contrast the static HiSiaP (SI Figure 5.6.5). These data underline the importance to identify FRET dynamics in bursts since only static distributions allow a meaningful interpretation of the mean FRET efficiency as a single distance in the macromolecular complex. With this knowledge we suggest that the average distances determined for the apo- and holo states of YopO by smFRET are as expected from the molecular models and show the same trends as the PELDOR/DEER data. In conclusion we find a long distance in the holo state, which matches the simulated distance from the crystal structure, and an “apparent” shorter distance caused by multiple averaged conformations of YopO in the apo state (Figure 5.2.6D). Interestingly, the PELDOR/DEER data [245] did not match the prediction from the crystal structure (Figure 5.2.6D). This effect had been confirmed by additional PELDOR/DEER measurements and SAXS and could be explained by a slight re-orientation of the kinase domain [245]. It is possible that the longer linker of the smFRET labels masks this effect.

Estimating the influence of linker length on the accuracy of predicted distance distributions. While both PELDOR/DEER and smFRET can accurately measure inter-probe distances, one is ultimately interested in inter-residues distance between the labelled amino acids, which are different from the inter-probe distances. The common solution to interpret the distance data is therefore to build an *in silico* model of the labelled protein by placing the structure of the label (or a geometric model thereof) onto the molecular surface of the biomacromolecule and to calculate its accessible volume [206, 224, 273, 280]. Some programs refine this accessible volume by considering preferred rotameric states of the label [224, 280] or by selecting such conformations that are close to the molecular surface [206, 273]. Unfortunately, it is difficult to build a model that accurately reflects the rotameric state of the labels, their interactions with the protein surface or solvent and the molecular motion of the protein backbone. Hence, a major part of the apparent distance inaccuracy (i.e., the mismatch between predicted and experimental distance) of PELDOR/DEER and smFRET is actually caused by the *in silico* models [206, 224, 273, 280]. Nevertheless, many examples in the literature have shown that PELDOR/DEER distance distributions can be predicted rather well (± 2 - 4 Å) by determining all possible distances between the modelled spin centers in such ensembles [273, 281, 282]. For smFRET, it was shown that the average distance can be predicted with an accuracy of ± 5 Å [111]. Calculating distance distributions is more difficult for smFRET due to the indirect way of distance determination via FRET efficiency [206, 283, 284], unless lifetime-based approaches are used [18]. The resulting averaging effects between simulated distance distributions and FRET-averaged measured distances are discussed in detail in [283].

No matter which method is used, the prediction accuracy in the end depends on the correctness of the modelled label. Considering the topic of this work, we asked ourselves, how the prediction uncertainty is influenced by geometrical factors, i.e., the distinct linker length of typical spin- and fluorescence labels (Figure 5.2.1) and protein-label interactions (Figure 5.2.7). For the following simulations, we used a slightly modified version of *mtsslWizard* [273]. The program was used to attach *in silico* models of two different labels (MTSSL and Alexa Fluor 647) to the open-state HiSiaP protein (PDB-ID: 2CEY) and to generate ensembles of rotamers at positions 55, 58, 112, 134, 175 and 225. To simulate the effect of protein-label interactions, we replaced a certain percentage of randomly selected rotamers in each ensemble by one randomly chosen rotamer. In our simulation, these “sticky” rotamers represented an immobilized label.

We arbitrarily defined a weak interaction to lead to 10 % of the rotamers in a given ensemble to occupy the same position, while the remaining 90 % of rotamers were randomly distributed in the accessible volume. The simulation was run for immobilized/mobile ratios of 10/90, 50/50 and 100/0 (Figure 5.2.7). We then determined the average distances between each pair of the “sticky” ensembles (this simulates a set of distance measurements), as well as the average distances between the corresponding pairs of the original ensembles (i.e. without “sticky” rotamers, which corresponds to the prediction of a particular distance by the accessible volume approach) and calculated the prediction errors. The whole procedure was repeated 1000 times to achieve a statistical distribution of the “interaction site” within the ensembles. The results are summarized in Figure 5.2.7B: For a weakly immobilized label (Figure 5.2.7B, magenta), the prediction error was low, for both MTSSL (width of error histogram: $\sigma = 0.4 \text{ \AA}$) and for Alexa Fluor 647 ($\sigma = 0.6 \text{ \AA}$) (Figure 5.2.1). This changed markedly when the label interacted more strongly with the protein surface. If the label occupied a fixed position half of the time, the prediction error increased considerably (Figure 5.2.7B, beige). Whereas the error was still relatively small ($\sigma = 1.8 \text{ \AA}$) for MTSSL, larger errors of up to $\pm 10 \text{ \AA}$ ($\sigma = 3.2 \text{ \AA}$) quite frequently occurred for Alexa Fluor 647. If the interaction was even stronger, i.e., for a completely immobilized label (100 %, blue), it was quite likely to observe a very high prediction error for MTSSL ($\sigma = 3.8 \text{ \AA}$) but especially for Alexa Fluor 647 ($\sigma = 6.6 \text{ \AA}$). Note that our approach neglects movements of the protein backbone, which will likely further increase the observed errors. An example for such an immobilized label is the crystal structure of the matrix metalloproteinases MMP-12 in complex with a fluorophore labelled Cy5.5 inhibitor (Figure 5.2.7B) [285], which has a chemical structure related to Alexa Fluor 647 (Figure 5.2.1). In the complex, a large part of the label is bound to the surface of the protein with its sulfonic acid groups interacting with the positive charges of lysine and arginine residues. Figure 5.2.7B shows the size of the accessible volume for the label (blue) for comparison. Similar observations have been made for the MTSSL label, for instance in the case of the Spa15 chaperone [278].

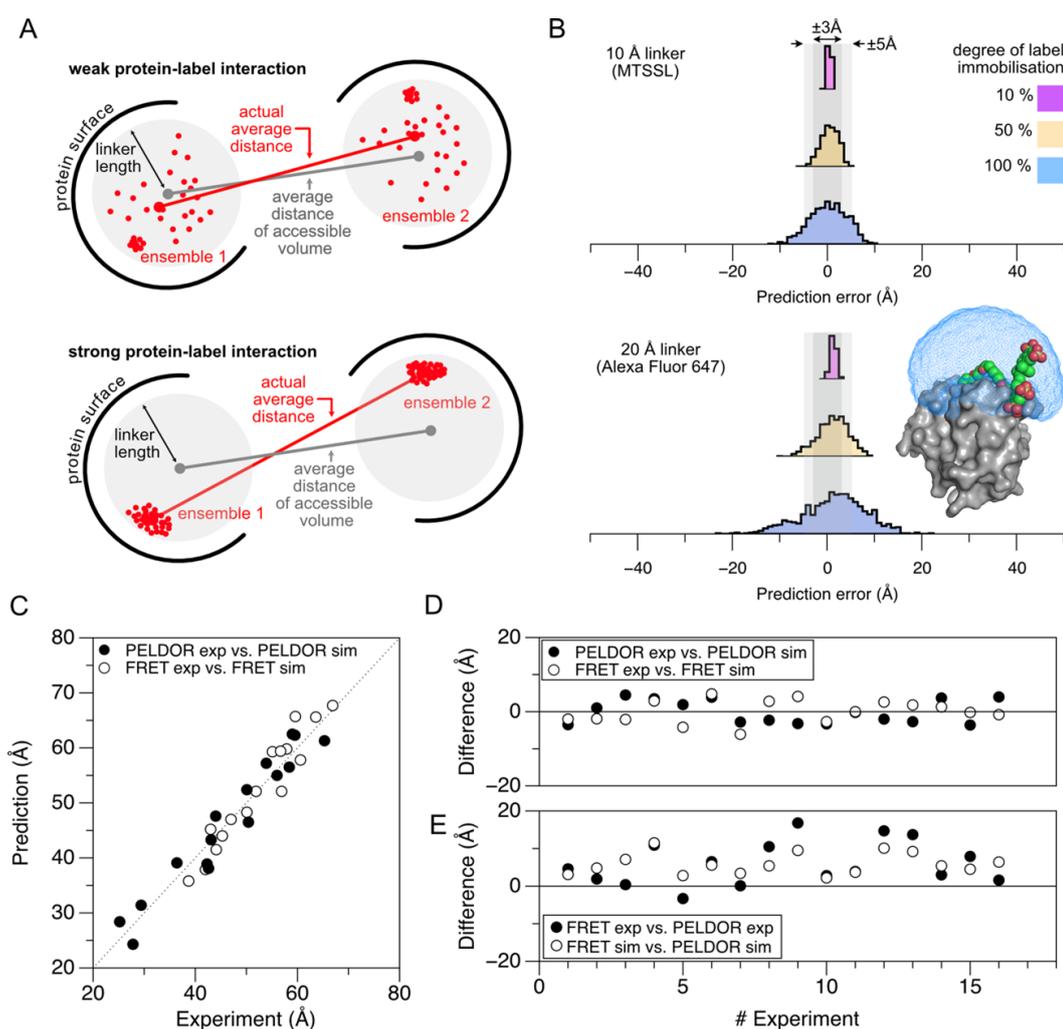


Figure 5.2.7: Comparison of PELDOR/DEER and smFRET measurements and the influence of linker length on the correlation between experimental and predicted distances. **A** Multiple ensembles of spin- and fluorescence labels were simulated with mtsslWizard using the open form of HiSiaP (PDB-ID: 2CEY [232]) and the labelling sites 55, 58, 112, 134, 175 and 225. In the schematic, the radius of the sphere represents the length of the linker that connects the fluorophore or spin center to the C-alpha atom of the labelled residue. Interactions with the protein surface (grey arcs) are indicated and lead to a clustering of labels at that position. Depending on the degree of interaction between protein and label, the accessible volume approach becomes less accurate. **B** Histograms of 1000 experiments described in A with a 10 Å linker (upper plot) and 20 Å linker (lower plot) and varying degree of protein label interaction. The percentage indicates how many percent of the 1000 dummy atoms are localized at the interaction site. As example for a long (20 Å) and immobilized linker, the protein structure of MMP-12 (matrix metalloproteinases, PDB-ID 5L79 [285]) in conjugation with a Cy5.5 fluorophore (K241, colored spheres) was selected. The surface of the protein is shown in grey and the accessible volume of the fluorophore, calculated with FPS [206] is shown as a blue mesh. **C** Predicted vs experimental smFRET or PELDOR/DEER distances of datasets that were measured with both methods in this study. **D** As C, but the differences are plotted against the experiment number in SI Table 5.6.3. **E** Comparison of the raw distances determined by PELDOR/DEER or smFRET and the simulation for the same experiment.

5.3 Discussion

We performed PELDOR/DEER and smFRET experiments on three substrate binding proteins (HiSiaP, MalE, SBD2) and the YopO protein to conduct a comprehensive cross-validation of the two methods. One of our goals was to determine the distance accuracy by comparison of simulated mean inter-probe distances vs. experimental ones and to see if both techniques correctly monitor ligand-induced structural changes. For this purpose, we used the same labelling sites for both methods and measured the inter-probe distances in the presence or absence of the respective ligands. Both methods showed a good consistency with each other and towards structural models. For a quantitative comparison of both methods, we selected 15 datasets that were measured in this study and where a monomodal distance distribution was observed. We computed the difference between the average experimental and simulated distances for both PELDOR/DEER and smFRET and plotted the data in Figure 5.2.7C/D. Within the dataset, no systematic offset was observed for PELDOR/DEER or smFRET. The PELDOR/DEER and smFRET measurements on the same double variant differed by about 5 Å with an overall spread of ± 10 Å (Figure 5.2.7E).

To an extent, the above-described differences between the PELDOR/DEER and smFRET results will be related to the different nature of the labels and in particular to their differing linker lengths (compare Figure 5.2.1). Figure 5.2.7E shows that for most of our measurements, the difference between the two methods is larger than can be explained by the different linker lengths for a freely rotating label (otherwise, the black and white data points in Figure 5.2.7E should coincide). The simulations in Figure 5.2.7A/B reveal that already moderate protein-label interactions can lead to distance measurements that are seemingly inexplicable. This is vividly reflected in our first example (Figure 5.2.2/5.2.3). Luckily, there are experimental approaches to detect strong protein label interactions in smFRET (Figure 5.2.3) as well as for PELDOR/DEER. For the latter, the shape of room temperature cw spectra and abnormally shaped Pake patterns provide hints towards strongly immobilized spin labels [286, 287]. Depending on the label, its degree of immobilization and especially at high magnetic fields, it can be necessary to consider orientation selection effects by collecting multiple PELDOR/DEER time traces at different frequency offsets [287–289].

Based on the above, three seemingly obvious solutions emerge to improve the distance accuracies of both techniques:

(i) The use of probes with shorter linkers. It is challenging to shorten the length of e.g., the MTSSL label any further (Figure 5.1.1), and short inflexible linkers such as the two-armed Rx spin label [290] are more likely to disturb or artificially stiffen the protein structure. Alternative EPR labels such as Gd-, Cu- or trityl labels with different linker types and lengths are under active research [291–293]. For smFRET labels, shorter linkers are not only a challenge of chemical synthesis, but also for the implementation of the method itself, because the free rotation of the dye is the basic requirement for distance simulations based on the accessible volume and to exclude orientational effects on FRET efficiency [206]. This requirement would be hard to meet with shorter and

thus also more rigid linkers.

(ii) The prevention of label-protein interactions. A systematic reduction of label-protein interactions is challenging and its feasibility depends on the type of macromolecule. For nucleic acids, the molecular surface will be predominately negatively charged. Still, many parameters such as the counter-ion concentration and the fold of the nucleic acid (especially for RNA) can strongly impact label-nucleic acid interactions. For proteins, it is much more difficult to predict how the label will interact with the macromolecular surface and unexpected results are more likely. Nevertheless, we showed for HiSiaP that switching fluorophores to alter charge and linker length allowed us to circumvent this problem and to detect the expected conformational change and distances (Figure 5.2.3).

(iii) Improving in silico labelling approaches. So far, even time-consuming molecular dynamics approaches have not been shown to be much more accurate than the accessible volume approach in large scale benchmarks [273, 281, 282]. It should be noted that new and promising approaches to tackle this problem are constantly developed [294, 295]. Still, no matter how sophisticated the prediction algorithm, it will be difficult to obtain absolute certainty.

PELDOR/DEER-specific aspects. Some PELDOR/DEER Other, more specific considerations possibilities to improve PELDOR/DEER experiments emerge from our results. The use of cryo-protectants had a significant impact on our distance measurements on MalE. These substances are routinely, which are added to prevent protein aggregation during freezing and thereby to improve the phase memory time of the samples. had a significant impact on our distance measurements on MalE. At the high concentrations that are typically used (10-50% v/v glycerol or ethylene-glycol), the small molecules might interact with the protein and induce a different conformation of the spin label or the protein itself [296, 297]. Note that the proteins in our study (such as MalE) might well be particularly prone to such problems, because they have deep surface crevices that can easily bind such small molecules. We did not see any a significant influence of the cryoprotectant for the SBD2 example and the same was true for previous measurements on VcSiaP (a close homolog of HiSiaP from example 1) (SI Figure 5.6.15). Nevertheless, especially when unexpected results are found, a control measurement without cryo-protectant or a different cryo-protectant from the large arsenal of such substances should be performed [298]. Efforts to develop experimental procedures that allow to reduce the amount of cryo-protectant or to completely avoid their addition, for example by rapid freeze quenching are a promising route to circumvent this problem [296, 299]. Ultimately, it would be desirable to perform the PELDOR/DEER measurements at room temperature. Conformational equilibria are temperature dependent and are thus affected by the freezing procedure, which, despite all efforts, is still slow compared to the time scale of most molecular motions. The quantitative differences between PELDOR/DEER and smFRET that were observed in the case of SBD2 might be caused by the different experimental temperatures. Unfortunately, room temperature PELDOR/DEER measurements on proteins in solution are very challenging

and it is therefore no simple task to test this. Newly developed labels, such as the trityl spin labels are an important step on the road towards room temperature PELDOR/DEER experiments on proteins in solution [270, 271]. Another factor that might contribute to the observed deviations between PELDOR/DEER and smFRET, is the sample concentration. Whereas smFRET experiments are conducted in extremely dilute solutions, standard Q-band PELDOR/DEER experiments are performed at micromolar to nanomolar concentrations [300].

It is a strong point of PELDOR/DEER that distance distributions rather than average distances are obtained. However, it is important to remember that the shape of these distributions depends on a number of parameters such as the quality and especially the length of the underlying PELDOR/DEER time trace [246, 301]. Unfortunately, its maximum length cannot be arbitrarily chosen, since the refocused echo signal quickly decreases with an increasing time window (Figure 5.1.1D). Hence, a tradeoff between signal strength and length of the time-window must be made for each sample, where it is usually more important to have a longer time window than high SNR. The conversion of time traces into distance distributions has traditionally been solved by a two-step analysis of first fitting and removing the intermolecular background (Figure 5.1.1D, dashed red line) and then applying Tikhonov regularization to extract the distance information [256, 257]. The procedure introduces a regularization parameter α that describes a compromise between the smoothness of the distance distribution and how well it reproduces the experimental time-trace [256]. Because the true shape of the distribution is unknown, this procedure inevitably introduces a degree of uncertainty. This two-step procedure works well for high-quality data, where more than a complete oscillation period of the signal was recorded (e.g. SI Figure 5.6.2). In practice, this is not always the case and the separation of the intermolecular background becomes a major source of uncertainty. The evaluation feature of the DeerAnalysis software can be used to visualize the impact of this problem on the shape of the distance distribution. The software systematically varies parameters, most importantly the starting time of the background fit to obtain a mean value, a standard deviation, as well as upper- and lower limits for each point of the distance distribution [302]. In our comparisons above, the red shade around the distance distribution was produced with this feature. Recently, new data processing algorithms have been developed that for example calculate the distance distribution in a more robust one-step analysis and also consider the noise level in the raw data to estimate the uncertainty of the distance distribution [303, 304]. Yet another approach is “DeerNet” [258], an artificial neural network that was trained on a large database of simulated data. Importantly, this latter method is independent of user-adjustable parameters. As a comparison, we processed the data measured in this study with DeerNet and reassuringly found very similar results. In a recent study, aliquots of the same PELDOR/DEER samples were analyzed by seven leading PELDOR/DEER laboratories [302]. While the resulting distance distributions were overall quite similar, the variation between the individual labs was interestingly not fully covered by the error margins calculated with the different processing algorithms alone. Hence, a comparative analysis was suggested, where the uncertainties calculated from neural network analysis

and regularization approaches are combined [302].

smFRET-specific aspects. A major obstacle for determination of inter-probe distances via smFRET is the conversion of the experimentally derived setup-dependent apparent FRET efficiency E^* values into (i) accurate FRET E values (see Figure 3.4.3) and (ii) inter-probe distances. We provide a full overview of all parameters required for distance determination via smFRET in SI Table 1 for different protein variants. This correction is not required for lifetime-based approaches, which allow to obtain distance distributions from the lifetime decays directly [18]. However, a quite common procedure is the ratiometric determination of FRET efficiency with subsequent conversion into accurate FRET E and later into distances [110, 111, 221]. We will here discuss the most problematic steps of this approach. The correction for systematic errors introduced by the apparatus (background, spectral crosstalk and differences in donor/acceptor detection and fluorescence quantum efficiency; Figure 3.4.3) is directly carried over to the corrected FRET efficiency E values (step i). Later, in step (ii) Förster-radius determination is required for conversion of accurate E values into distances.

The introduced systematic errors in step (i) are largely dominated by the γ correction factor describing differences in acceptor-to-donor detection- and fluorescence-quantum efficiency. Obtaining a reliable γ factors is thus particularly important, also because cyanine-based donor and acceptor dyes, which are used in the manuscript, show large changes of fluorescence quantum yield depending on their specific environment [163, 200–202]. The correction step, however, requires multiple experiments with the same pair of labels (if possible, in the same environment), but with distinct FRET efficiencies, e.g., obtained via a conformational change. Such local values are, however, often not accessible and fluorophores often experience changes of their fluorescence quantum yield impacting . In our case ligand-induced conformational changes provided access to two FRET efficiency states with (often) identical fluorophore properties (Figure 5.2.2-5.2.5). This assumption is, however, not always valid as was seen in smFRET experiments of HiSiaP for distinct conformational states (Figure 3/4). In cases where only a single FRET efficiency state is available for each label position, a global γ factor has to be used, e.g., derived from the same combination of labels at different label-positions on the target, using the crude assumption that the fluorophore environment remains identical.

Another key problem related to step (ii) of the analysis procedure is the determination of the Förster radius R_0 . Since the same fluorophore may show different fluorescence quantum yields, spectra etc. subject to the its biochemical environment, R_0 will vary in the range of +/- 0.3 nm (impacting the determined distances by 0.3-0.5 nm). Consequently, calculation of R_0 can improve the distance determination significantly [111], yet it requires information on the donor fluorescence spectrum, donor quantum yield, acceptor absorption spectrum, the refractive index of the donor-acceptor intervening medium and their relative orientation. Problematically, many approaches currently in use for their determination overlook heterogeneity, particularly when stochastic labelling approaches are used. Particularly challenging is also the determination of the refractive index and the dye orientation factor κ^2 . The refractive index is often simply estimated

to be an average value of $n = 1.4$ considering the values for water ($n = 1.33$) and that of proteins ($n = 1.5$) [170]. κ^2 on the other hand is often idealized by the idea of non-interacting fluorophores that undergo free rotation during the donor fluorescence lifetime to a value of $\sim 2/3$. The distance-error resulting from an incorrect κ^2 can be estimated, and studies showed the effects κ^2 values can have on the FRET-derived distance estimation [305]. As a rule of thumb fluorescence anisotropy of the individual dyes, e.g., measured via steady-state anisotropy should be less than 0.2. This approximation, however, does not hold for large organic dyes or dyes with short lifetimes, since rotational correlation times (determined indirectly from fluorescence anisotropy decays) will be similar to the fluorescence lifetime.

This is the case for Alexa Fluor 555 as donor dye and prohibits its proper use for lifetime-based experiments. A general concern might thus be that Alexa Fluor 555 has varying R_0 -values and with that a poor distance accuracy. Based on this the question might arise why Alexa Fluor 555 was chosen as a label for this study. Our goal was to provide a comparison of smFRET and EPR using the most commonly used labels. By doing so we will allow other users to adapt the analysis routines, e.g., for obtaining accurate FRET values and distances. Notably, the most commonly used dye for smFRET studies is Alexa Fluor 555. Furthermore, our study shows a very good agreement between smFRET and EPR in both a qualitative (trends of distance changes) and quantitative sense (distance accuracy in comparison to AV calculations for smFRET or rotamer libraries for EPR). In the instances where discrepancies were observed between the techniques (or simulations), we managed to identify the reason for it. To validate the idea that Alexa Fluor 555 is as useful as other dyes for quantitative smFRET, we compared smFRET-derived distances of Alexa Fluor 555 (a cyanine) to Alexa Fluor 532 and 546 (both rhodamine dyes). Using two sets of cysteine pairs in both MalE and SBD2 we derived simulated and calculated distances as shown in SI Figure 5.6.17 and SI Table 5.6.4. The data and a plot of R-values reveals that the deviations (theoretical vs. experimental interprobe distance) are indeed very similar for all dye pairs and sometimes Alexa Fluor 555 is even more "spot-on" in comparison to Alexa Fluor 532 or 546 dyes. Importantly, we can see that certain labelling positions show larger deviations due to sticking of the fluorophores, which also seems independent on the choice of dye.

In summary, the distance variations seen in smFRET are fairly consistent with an error assessment described by Hellenkamp et al. Since all distances we studied here are in the range of $0.8 \cdot R_0$ to $1.3 \cdot R_0$, we can neglect background and spectral cross-talk (α , and δ errors play a minor role, i.e., $\Delta R < 1 \text{ \AA}$) as suggested Hellenkamp et al. [111]. Therefore, the major contribution is based on wrong R_0 determination or incorrect γ -values (as discussed above). Based on this we can estimate $\Delta R(\gamma) \approx 1 \text{ \AA}$ and $\Delta R(R_0) \approx 3\text{-}4.5 \text{ \AA}$ (depending on the distance) based on a relative error in γ and R_0 of 10% and 7%, respectively, which is in full agreement with the accuracy found in our study (Figure 5.2.7).

The impact of structural dynamics. Ironically, structural dynamics are usually the reason why PELDOR/DEER and smFRET are applied in the first place, but at the same time, these processes heavily impair the possibility to interpret and use inter-probe distances (or distance distributions) in a straightforward manner. In essence, this is because (time-)averaged distances cannot be interpreted easily in light of static structures. Thus, the use of inter-probe distances in structural modelling requires a verification of the degree of structural dynamics. Since smFRET experiments are performed with individual molecules in liquid solution at room temperature, it allows to characterize fast structural dynamics over a range of different timescales ranging from nano- to milliseconds [12, 225]. In contrast, PELDOR/DEER experiments are almost always conducted on frozen bulk samples and it is not possible to directly study dynamics in real-time. Yet, PELDOR/DEER data can be readily converted to distance distributions and the shape of the latter contains information about conformational heterogeneity of the sample (see above for important caveats on interpreting the shape of distance distributions). The YopO example demonstrates this difference between the bulk-point of view of PELDOR/DEER and the single-molecule point of view of smFRET. From the broad distance distributions of the original PELDOR/DEER data, it was correctly inferred that the YopO protein is flexible and adopts different conformational states in its apo state [245]. However, it was not possible to differentiate between static or dynamic nature of the conformational heterogeneity in the sample. With smFRET investigations it became clear that apo YopO is indeed dynamic and shows sub-millisecond conformational motion that depend on the presence or absence of actin (Figure 5.2.6). Recently, we published another study of YopO using multi-parameter photon-by-photon hidden Markov modelling [306], where the underlying apo distribution of YopO could be resolved. We showed that it consists of two major FRET peaks that average on a sub 100 μ s timescale to give one apparent “static” populations. The ligand stabilized one of the two conformers (low FRET), yet both conformational states and FRET efficiencies could in principle be interpreted by structural means.

Conclusion. Both PELDOR/DEER and smFRET are valuable tools for the emerging toolkit of integrative structural biology. Overall, we found a reasonable agreement of the determined distances with an average ± 5 Å spread between the two methods. However, our experiments also revealed discrepancies that might have led to wrong interpretations if only one of the methods had been used. We could show that these differences were partly due to the distinct labels and label-protein interactions. Thus, reliable methods to predict or prevent label-protein interactions are urgently needed. Also, the much longer linkers used for smFRET can be problematic, while for PELDOR/DEER, the use of cryogenic temperatures and cryoprotectants was shown to influence conformational changes in an unwanted fashion. A positive outcome of our study is the observation that a combination of PELDOR/DEER and smFRET provides highly complementary and synergistic insights into the conformational states of macromolecules. Hence, the development of spectrometers and microscopes, as well as standardized data processing approaches [12, 225], which can also be used by non-experts, would be very beneficial to

structural biology [307].

5.4 Acknowledgements and Author Contributions

Acknowledgments. This project was financed by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) in projects no. HA 6805/4-1 and HA 6805/5-1 (to GH), GRK2062 project C03 (to TC), SFB863 project A13 (to TC), the Center for integrated protein science Munich CiPSM (start-up funding to TC), an ERC starting grant ERC-StG 638536 SM-IMPORT (to TC), and the Center of Nanoscience Munich (project funding to TC). MFP acknowledges a PhD fellowship from the Konrad Adenauer Stiftung. CG acknowledges a PhD fellowship from the Studienstiftung des deutschen Volkes. MFP, JG and GH thank Prof. Olav Schiemann, University of Bonn, for support and access to EPR spectrometers and Dr. D. Abdullin for very valuable discussions and helpful comments on the manuscript. TC thanks Nicola Gericke for help with purification of substrate binding domain 2. GH and MFP thank Dr. Frank Eggert and Prof. Stefanie Kath-Schorr for help with the MS analysis. We thank E. Lerner for the gift of HP3 and D. Griffith for proofreading and commenting on the manuscript.

Author contributions. GH and TC conceived and supervised this study. MFP and JG performed PELDOR/DEER experiments. CG performed single-molecule FRET experiments. MFP and RM provided protein samples. MFP, CG, TC and GH have designed the experiments and analyzed the data. MFP, CG, TC and GH wrote the manuscript. All authors contributed to the discussion of the results and the final version of the manuscript.

5.5 Materials and Methods

Selection of labelling sites. Dependent on the particular method (PELDOR/DEER or smFRET) for which the protein variants were produced, we used different software to calculate suitable labelling positions. For spin label positions we used mtsslSuite (www.mtsslsuite.isb.ukbonn.de) and calculated a difference distance map between the open and closed conformation (as shown in results) [276]. With this map we identified regions with large conformational changes and selected amino acids inside these regions, which are located on the surface of the protein to obtain a good accessibility. For smFRET studies, residues were rated based on different parameters such as solvent exposure or conservation to obtain a labelling feasibility estimate. Residues were selected that showed large distance changes between apo and holo (or no distance change as negative control).

Protein expression and purification. The TRAP SBPs HiSiaP and VcSiaP were expressed and purified as described before [274]. To prevent co-purification of the substrate, the *E. coli* cells were cultured in M9 minimal medium. For purification, the protein was loaded onto a benchtop Ni-affinity chromatography, followed by an ion exchange

chromatography and a size exclusion chromatography. In all steps the buffers were supplemented with 1 mM Tris(2-carboxyethyl)phosphine (TCEP) to avoid dimerization of the cysteine variants and the purity was checked after each step with SDS-PAGE. The purified protein solution was concentrated to 20 mg/mL and stored at -80 °C until labelling. SBD2 and MalE were expressed and purified as described before [107]. YopO was produced as described previously [245].

Liquid chromatography mass spectrometry (LC-MS). The LC-MS analysis was performed on an HTC esquire (Bruker Daltonic) in combination with an Agilent 1100 Series HPLC system (Agilent Technologies). Analysis gradient: 5→100% MeCN (solvent B)/0.1% formic acid (solvent A) in 20 min at a flow rate of 0.4 mL min⁻¹ using a Zorbax Narrow Bore (2.1x50 mm, 5 µm) C18 column (Agilent Technologies).

Protein labelling.

Spin labelling: In the first step, the reducing agents in the protein solution were removed with a PD10 desalting column (GE Healthcare), using a buffer based on 50 mM Tris, 50 mM NaCl, pH 8 without TCEP or DTT. Immediately after elution from the column, the protein eluate was treated with 5 times excess per cysteine of the nitroxide spin label MTSSL (Toronto Research Chemicals, Canada), dissolved in DMSO. The labelling was carried out for one hour at room temperature under gentle shaking. Afterwards, the protein was concentrated and another PD10 desalting column was used to remove free spin label. The protein eluate was again concentrated to ~ 20 mg/mL. The labelling was verified and quantified with continuous-wavelength EPR spectroscopy (cw-EPR) [286] using an EMXnano X-band EPR spectrometer from Bruker (Billerica, MA). The spin labelled proteins were diluted to a concentration of 25 µM with standard buffer and a total volume of 10 µL sample was prepared into a glass capillary, sealed with superglue. The magnetic field of the cw-EPR spectrometer at room temperature were set to a center field of 3448 G and the microwave frequency to 9.631694 GHz. The microwave power was set to 2.5 mW, the power attenuation to 16 dB and the receiver gain to 68 dB. The cw-EPR spectra were recorded with a sweep width of 150 G, a sweep time of 10.03 s with 20.48 ms time constant and 1 G modulation amplitude. For every sample 350 cw-EPR spectra were averaged to obtain a good SNR. The concentration of the spin label and the labelling efficiency were determined with the Bruker software Xenon by double integration of the cw-EPR spectrum.

Fluorophore labelling: Proteins were labelled as described previously [102, 107]. The cysteines were stochastically labelled with the maleimide derivative of the dyes TMR, Alexa Fluor 555, Alexa Fluor 647 and Cy5 (ThermoFischer Scientific). His-tagged proteins were incubated in 1 mM DTT to keep all cysteine residues in a reduced state and subsequently immobilized on a Ni Sepharose 6 Fast Flow resin (GE Healthcare). The resin was incubated 2-4 h at 4°C with 25 nmol of each fluorophore dissolved in 1 ml of labelling buffer 1 (50 mM Tris-HCl pH 7.4, 50 mM KCl, 5% glycerol) and subsequently washed sequentially with 3 ml labelling buffer 1 and buffer 2 (50 mM Tris-HCl pH 7.4, 150 mM KCl, 50 % glycerol) to remove unbound fluorophores. Bound proteins were eluted

with 500 ml of elution buffer (50 mM Tris-HCl pH 8, 50 mM KCl, 5% glycerol, 500 mM imidazole) The labelled protein was further purified by size-exclusion chromatography (ÄKTA pure, Superdex 75 Increase 10/300 GL, GE Healthcare) to eliminate remaining fluorophores and remove soluble aggregates. For all proteins, labelling efficiencies were higher than 70% and donor-acceptor pairing at least 20%.

PELDOR/DEER spectroscopy.

If not indicated in the results, the standard EPR samples were prepared and the measurements were set up as described in the following. The proteins and additives were mixed and diluted to a concentration of 15 μM in a volume of 40 μL with PELDOR/DEER buffer (100 mM TES pH 7.5, 100 mM NaCl in D₂O). The substrate concentrations were 1 mM N-acetyl neuraminic acid for HiSiaP, 1 mM maltose for Male and 100 μM glutamine for SBD2. The solutions were supplied with 40 μL deuterated ethylene glycol, transferred into a 3 mm quartz Q-band EPR tube and immediately flash-frozen and stored in liquid nitrogen.

The PELDOR/DEER experiments were measured on an ELEXSYS E580 pulsed spectrometer from Bruker in combination with an ER 5106QT-2 Q-band resonator. The temperature was set to 50 K with a continuous flow helium cryostat (CF935, Oxford Instruments) and a temperature control system (ITC 502, Oxford Instruments). The PELDOR/DEER time traces were recorded with the pulse sequence $\pi/2(\nu_A)-\tau_1-\pi(\nu_A)-(\tau_1+t)-\pi(\nu_B)-(\tau_2-t)-\pi(\nu_A)-\tau_2$ -echo. The frequency ν_A of the detection pulses were set 80 MHz lower than the frequency of the pump pulse ν_B , which was set to the resonator frequency and the maximum of the nitroxide spectrum. Typically, the shot repetition time was 1000 μs and the lengths of τ_1 and τ_2 were 12 and 24 ns, respectively. The contribution of deuterium ESEEM to the PELDOR/DEER time trace was suppressed by addition of 8 observed time traces with variable τ_1 time ($\Delta = 16$ ns). The PELDOR/DEER background was fitted by a monoexponential decay. The distance distributions were calculated and validated by means of DeerAnalysis 2018 [257].

smFRET spectroscopy.

Solution based smFRET experiments were performed on a homebuilt confocal ALEX microscope as described in section 3.4. All sample solutions were measured with 100 μl drop on a coverslip with concentration of around 50 pM in buffer (50 mM Tris-HCl pH 7.4, 50 mM KCl).

Data analysis was performed using home written software package as described in detail in section 3.4. E-histogram of double-labelled FRET species with Alexa Fluor 555 and Alexa Fluor 647 was extracted by selecting $0.25 < S < 0.75$. E-histograms of open state without ligand (apo) and closed state with saturation of the ligand (holo) were fitted with a Gaussian distribution $Ae^{-(E-\mu)^2/(2\sigma^2)}$. The burst variance analysis (BVA) [65] was performed on the same data with a photon binning of 5 photons for selected bursts with stoichiometry $0.25 < S < 0.75$.

Distance conversion was calculated according to equation 2.2.7. The Förster radius R_0 is calculated according to equation 2.2.3 to be $R_0 = 50$ Å for Alexa Fluor 555 – Alexa Fluor 647 and $R_0 = 52$ Å for TMR – Cy5 (see SI Table 5.6.5), where the parameters for the donor quantum yield Q_D , the donor emission spectrum F_D , and the acceptor

absorbance spectrum ε_A were derived from absorption and emission spectra of singly labelled donor and acceptor mutants [113].

In silico distance simulations.

For both methods, we used available programs and combined the in silico distance simulations with the experimental distances in the result parts. For PELDOR/DEER simulations, we used mtsslWizard (www.mtssluite.isb.ukbonn.de) where an ensemble of rotamers is calculated for each labelling position by rotation of the bonds from the spin label (see results) [273]. After this, the average distance and the distance distribution between two of these ensembles were determined.

For smFRET we used the FRET-restrained positioning and screening method established by the Seidel lab [206]. This method allows the determination of a FRET-efficiency-averaged model distance between the two dyes using the crystal structure information. For distance simulations we employed a simple dye model, in which three parameters were used to determine the accessible volume the dye can sample: (i) linker-length (linker), linker-width (W), and the fluorophore volume, which can be derived from an ellipsoid using R1, R2 and R3. With this information, the average distance between two of these spheres was calculated. The dye parameter for the different fluorophores are shown in Table 5.6.3. An average distance was calculated with the FPS software by exchanging donor and acceptor positions and vary the linker length (± 1 Å) as well as linker width and radii (± 0.5 Å).

Fluorophore lifetime and time-resolved anisotropy measurements

Lifetime and anisotropy decay measurements were performed as described in section 3.3. 400 μ l of sample is measured in a 1.5x10 mm cuvette at a concentration of around 100 nM. The excitation power was 10 μ W and the concentration was finetuned to have ~50 kHz count rate under magic angle conditions. All anisotropy and lifetime measurements were recorded for 5 min in the order vertical (VV1), horizontal (VH1), magic angle (MA), horizontal (VH2), and vertical (VV2) under vertical excitation.

5.6 Supporting Information

System	BG _{DD} [kHz]	BG _{DA} [kHz]	BG _{AA} [kHz]	α	δ	γ	β
HiSiaP	1.33	0.83	0.77	0.084	0.065	2.17	0.82
MalE	1.51	0.99	1.79	0.064	0.059	1.56	1.00
SBD2	1.04	0.76	1.17	0.079	0.072	1.97	0.77
YopO	1.67	0.85	0.93	0.087	0.068	2.26	0.64

Table 5.6.1: Correction parameter of FRET measurements with Alexa Fluor 555 – Alexa Fluor 647. Overview of all correction factors for smFRET measurements on HiSiaP, MalE, and YopO.

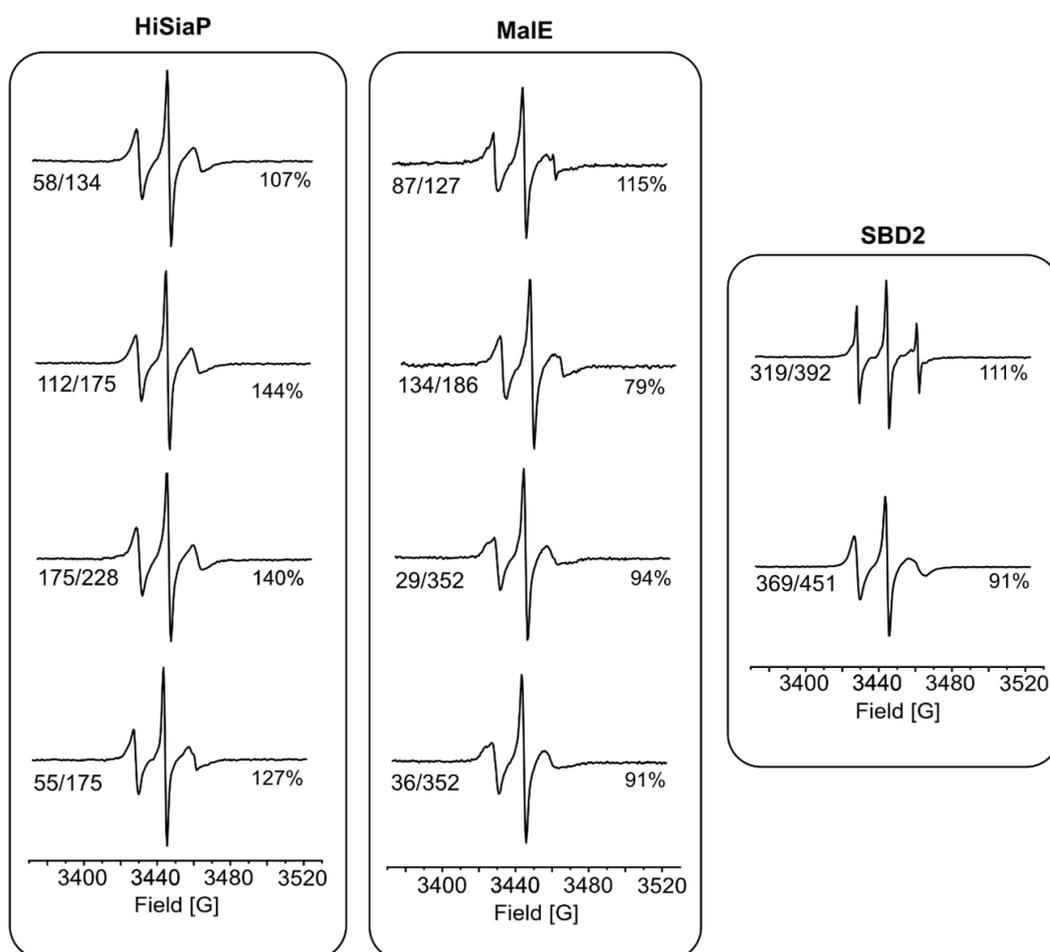


Figure 5.6.1: cw-EPR spectra of spin labelled double variants. The labelling of each variant with MTSSL was verified with room temperature cw-EPR spectroscopy (X-band). The labelling efficiencies were determined with the spectrometer software and is given next to the spectra.

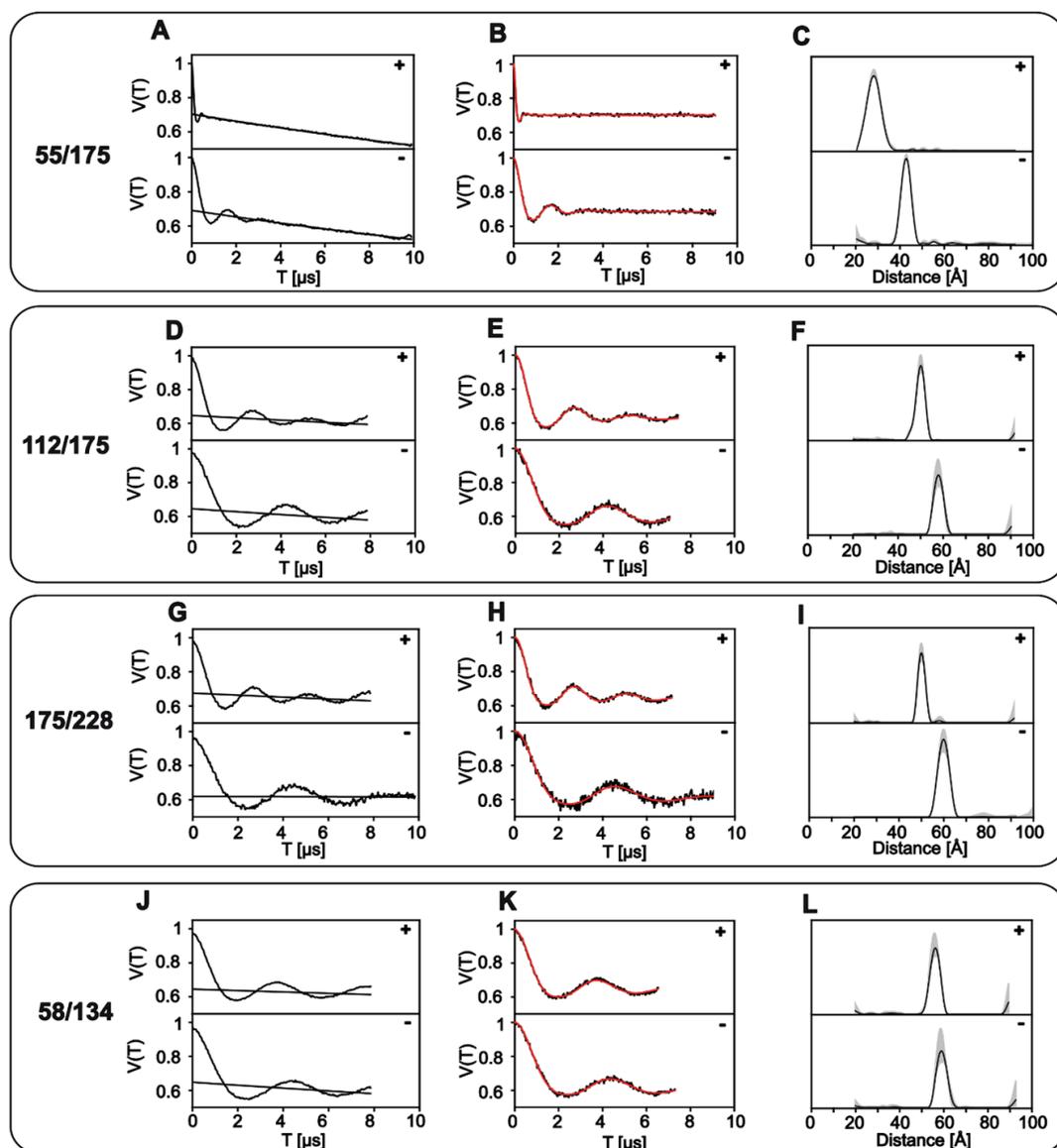


Figure 5.6.2: PELDOR/DEER data of HiSiaP variants. A/D/G/J Raw PELDOR/DEER time traces for apo (-) and holo (+) measurements of each double variant. The background, which was used for correction of the signal, is indicated as a black line. B, E, H, K) Background-corrected PELDOR/DEER time traces (black) and fits of the signal (red). C/F/I/L Distance distributions from PELDOR/DEER time traces (black) with validation of the distribution (grey).

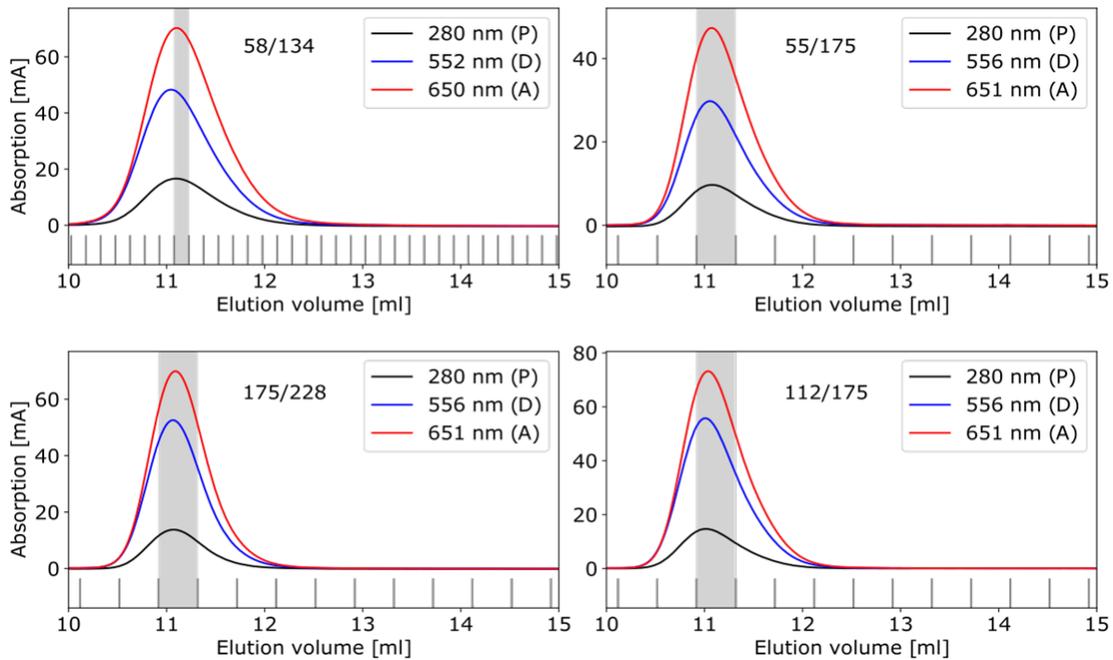


Figure 5.6.3: Size exclusion chromatography of HiSiaP variants labelled with Alexa Fluor 555 – Alexa Fluor 647. Absorption profile of the size exclusion chromatography (ÅKTA, Superdex 75 Increase 10/300 GL, GE Healthcare) for all tested HiSiaP variants 58/134, 175/228, 55/175 and 112/175 to monitor protein concentration (280 nm) and Alexa Fluor 555 (552 nm) / Alexa Fluor 647 (650 nm). The grey area indicates the fraction used in the smFRET experiments, where labelling efficiencies of >90% was achieved for all samples.

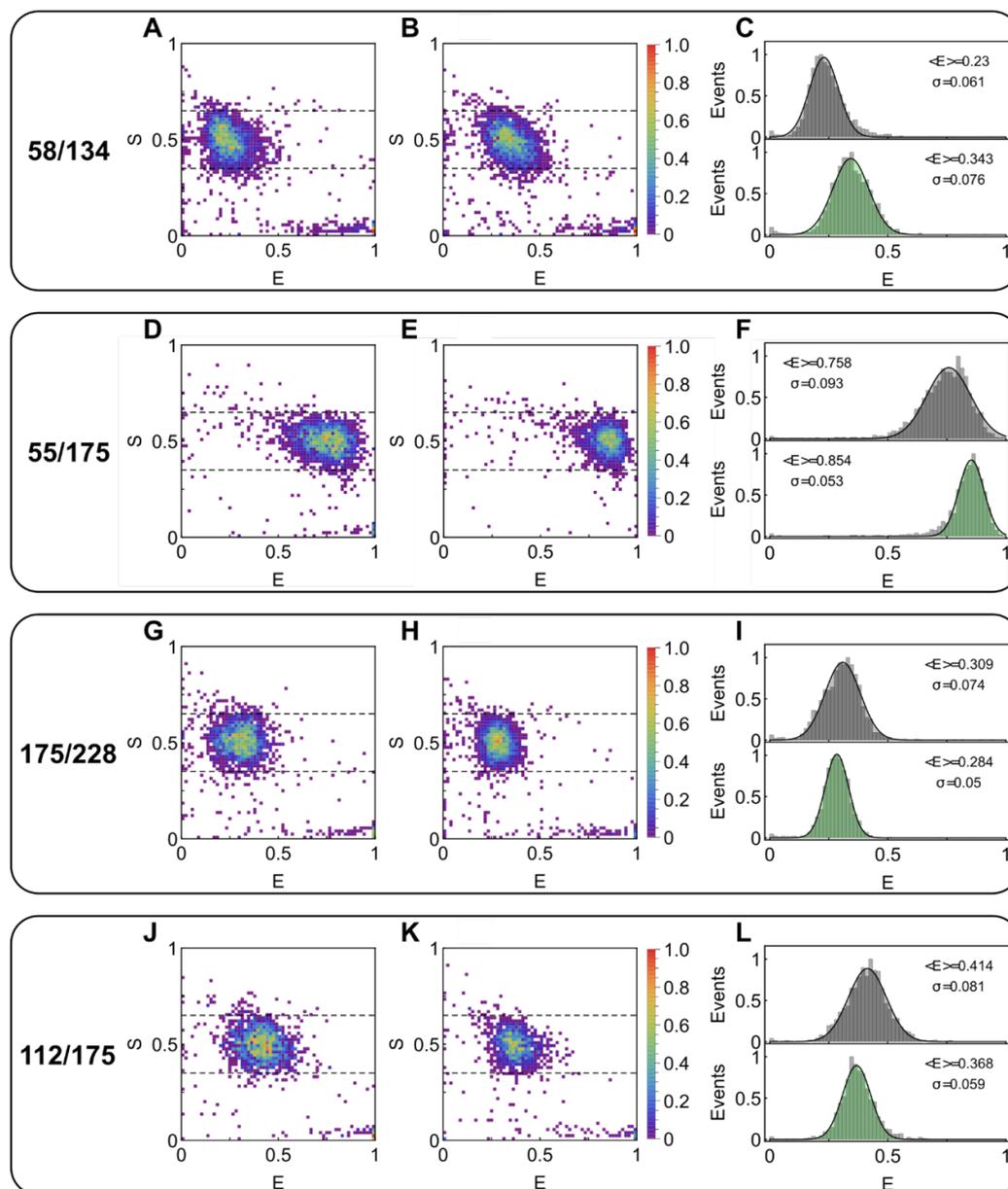


Figure 5.6.4: smFRET data of HiSiaP with Alexa Fluor 555 – Alexa Fluor 647. **A/C/E/G** ES-2D-Histograms of HiSiaP variants 58/134, 175/228, 30/175, 55/175, and 112/175 in apo state (-) and holo state (+). **B/D/F/H** 1D-E-Histograms extracted from the ES-data for apo (grey) and holo (green) are fitted with a 1D-Gaussian distribution. Mean $\langle E \rangle$ and standard deviation σ are labelled. The numbers of considered bursts N in apo/holo states are **C** 2435/2316, **F** 2129/1777, **I** 1932/2082, and **L** 1529/1367, respectively.

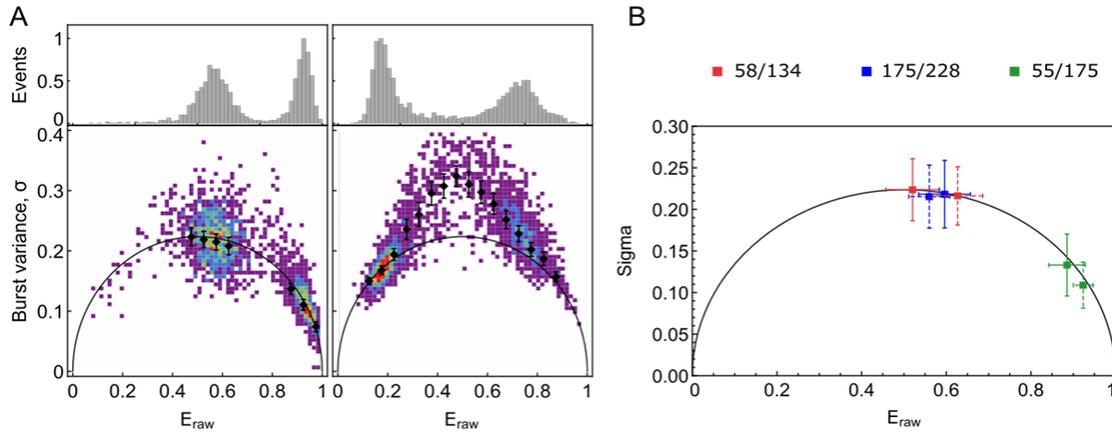


Figure 5.6.5: Burst-variance analysis of HiSiaP ALEX data. **A** Burst variance analysis example of 55/175 and 175/228 in their holo states (left, joined data set, $N=2718/1933$) and dynamic control experiments with a fluctuating DNA-hairpin (right, 2824) from reference [112]. Data are binned into bins of 0.05 and mean and standard error of mean are shown (black). **B** Population mean and standard deviation of all bursts of one measurement of burst variance analysis for three HiSiaP variants from Figure 5.2.2 in apo (solid) and holo state (dashed).

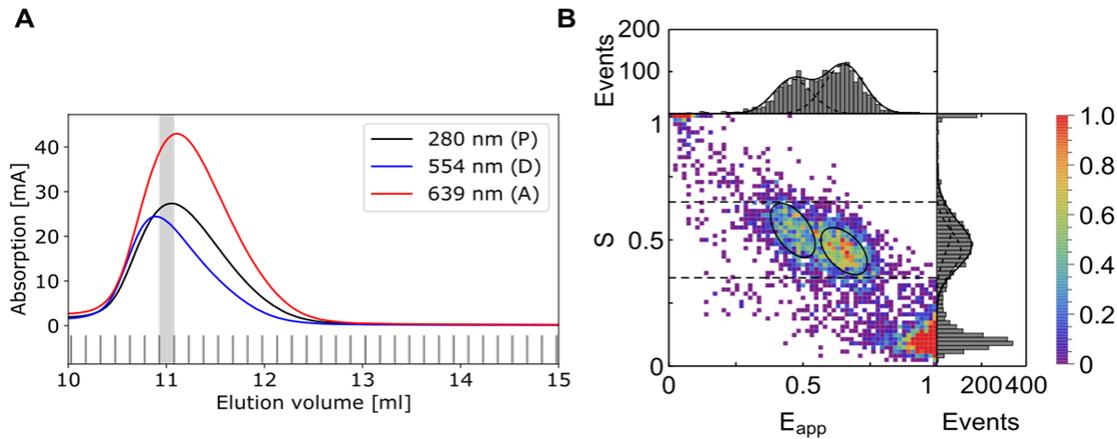


Figure 5.6.6: Extreme example for environmental effect on FRET efficiency. HiSiaP variant 112/175 labelled with Alexa Fluor546 – Star635P. **A** Absorption profile of the size exclusion chromatography (ÄKTA, Superdex 75 Increase 10/300 GL, GE Healthcare) to monitor protein concentration (280 nm) and Alexa Fluor546 (554 nm) / Star635P (639 nm). The grey area indicates the fraction used in the smFRET experiments, where labelling efficiencies of $>90\%$. **B** ES-2D-Histograms of variant from A) in apo state. 1D-E-Histograms and 1D-S-Histogram are shown on top and on the right, respectively. The ES-data are fitted with a 2D-Gaussian distribution where the 1D-integrals are shown in the 1D-histograms (black lines).

Sample	Width of Eraw data	Theoretical width	Ratio
HiSiaP 58/134 apo	0.075	0.04	1.88
HiSiaP 58/134 holo	0.073	0.039	1.87
HiSiaP 55/175 apo	0.047	0.023	2.04
HiSiaP 55/175 holo	0.024	0.02	1.20
HiSiaP 112/175 apo	0.067	0.037	1.81
HiSiaP 112/175 holo	0.054	0.039	1.38
HiSiaP 175/228 apo	0.07	0.039	1.79
HiSiaP 175/228 holo	0.05	0.039	1.28
MalE 87/127 apo	0.047	0.034	1.38
MalE 87/127 holo	0.059	0.041	1.44
MalE 36/352 apo	0.075	0.048	1.56
MalE 36/352_holo	0.054	0.038	1.42
MalE 29/352 apo	0.072	0.048	1.50
MalE 29/352 holo	0.07	0.048	1.46
MalE 134/186 apo	0.033	0.025	1.32
MalE 134/186_holo	0.032	0.025	1.28
SBD2 319/392 apo	0.075	0.046	1.63
SBD2 319/392_holo	0.07	0.043	1.63
SBD2 369/451 apo	0.066	0.044	1.50
SBD2 369/451 holo	0.049	0.034	1.44
YopO 113/497 apo	0.106	0.046	2.30
YopO_113/497_holo	0.067	0.046	1.46

Table 5.6.2: Width and mean positions of smFRET experiments in comparison to the expected width based on pure statistical noise (“shot-noise”). The width (standard deviation) of the FRET histogram is obtained from a Gaussian fit to the uncorrected FRET data. The theoretical width is calculated based on the photon count histograms assuming pure statistical noise.

Label	Linker length [Å]	W [Å]	R1 [Å]	R2 [Å]	R3 [Å]
Alexa Fluor 555 – C2 Maleimide [113]	21	4.5	8.8	4.2	1.5
Alexa Fluor 647 – C2 Maleimide [113]	21	4.5	11	4.7	1.5
TMR [206]	12	4.5	6	4.2	1.5
Cy5 [206]	21	4.5	11	3	1.5

Table 5.6.3: Geometric Parameters for in silico predictions of FRET labels.

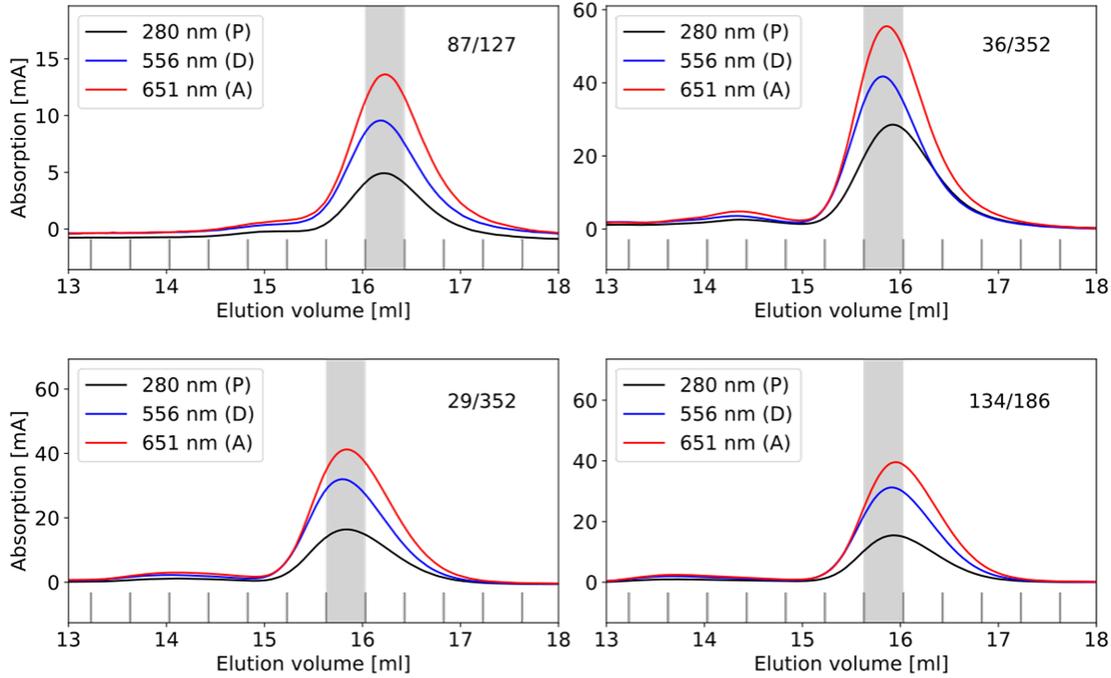


Figure 5.6.7: Size exclusion chromatography of MaleE variants labelled with Alexa Fluor 555 – Alexa Fluor 647. Absorption profile of the size exclusion chromatography (ÄKTA, Superdex 75 Increase 10/300 GL, GE Healthcare) for all tested MaleE variants 87/127, 36/352, 29/352 and 134/186 to monitor protein concentration (280 nm) and Alexa Fluor 555 (552 nm) / Alexa Fluor 647 (650 nm). The grey area indicates the fraction used in the smFRET experiments, where labelling efficiencies of >90% was achieved for all samples.

	Simulated distance	Measured distance	Simulated distance	Measured distance
Mutant	Alexa Fluor 555 – Alexa Fluor 647		Alexa Fluor 532 – Alexa Fluor 647*	
MaleE 29-352, apo	70.1	63.1	72.0	66.1
MaleE 29-352, holo	57.7	56.2	59.1	59.7
MaleE 87-186, apo	47.5	48.2	50.2	52.1
MaleE 87-186, holo	54.2	54.6	57.2	58.2
Mutant	Alexa Fluor 555 – Alexa Fluor 647		Alexa Fluor 546 – Alexa Fluor 647**	
SBD2 369-451, apo	65.1	60.5	65.4	60.8
SBD2 369-451, holo	52.1	52.5	51.2	49.3
SBD2 319-392, apo	58.3	54.7	57.6	56
SBD2 319-392, apo	44.0	46.1	44.2	45.8

*Förster radius: $R_0=61 \text{ \AA}$, ** Förster radius: $R_0=66 \text{ \AA}$

Table 5.6.4: Fluorophore comparison. Overview of simulated and measured distances with varying donor fluorophores for selected mutants.

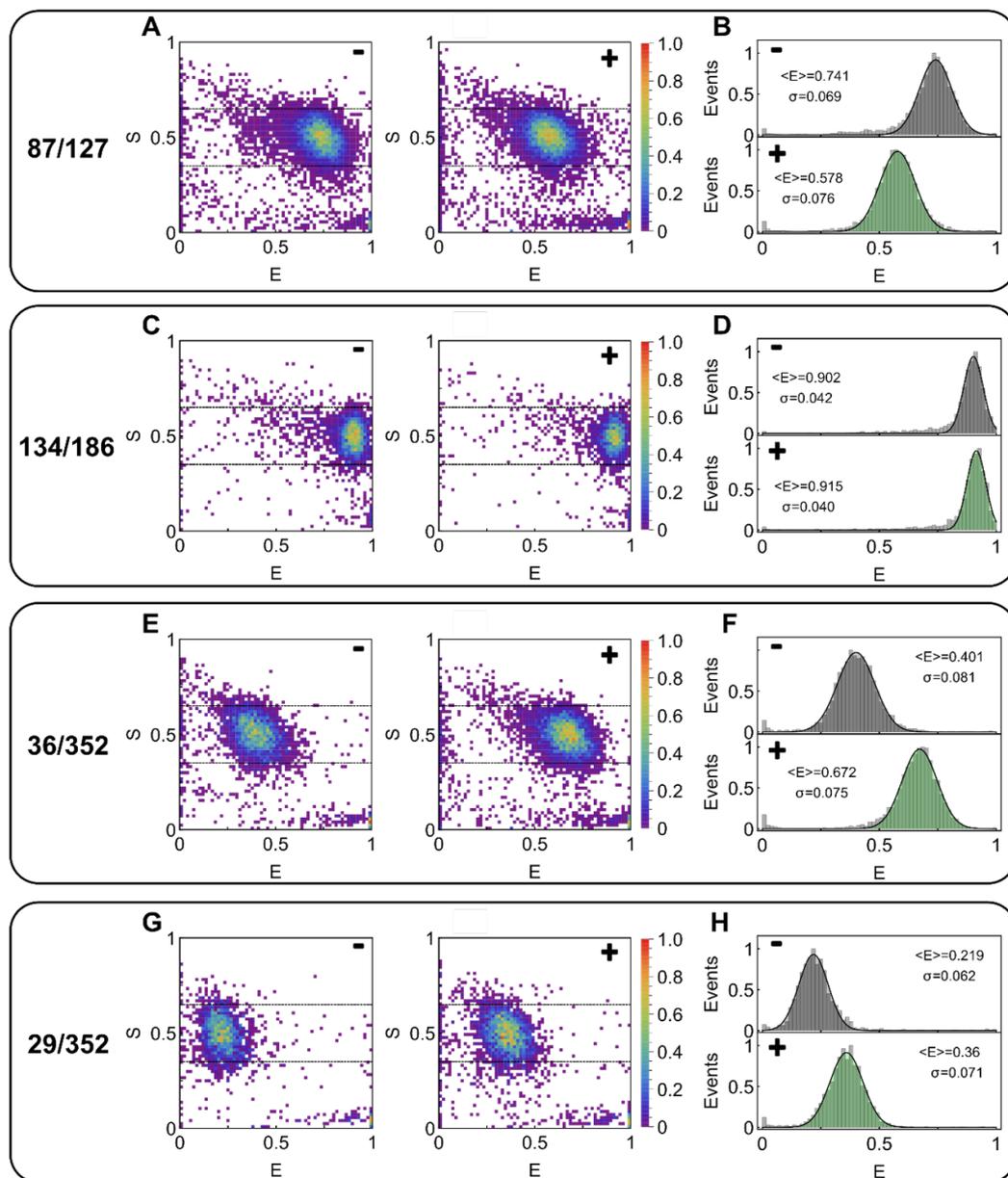


Figure 5.6.8: smFRET data of MalE with Alexa Fluor 555 – Alexa Fluor 647. **A/C/E/G** ES-2D-Histograms of MalE variants 87/186, 134/186, 36/352, and 29/352 in apo state (-) and holo state (+). **B/ D/ F/H** 1D-E-Histograms extracted from the ES-Data for apo (grey) and holo (green) are fitted with a 1D-Gaussian distribution. Mean $\langle E \rangle$ and standard deviation σ are labelled. The numbers of considered bursts N in apo/olo states are **B** 6168/6181, **D** 6902/5672, **F** 2593/4293, and **H** 1392/2249, respectively.

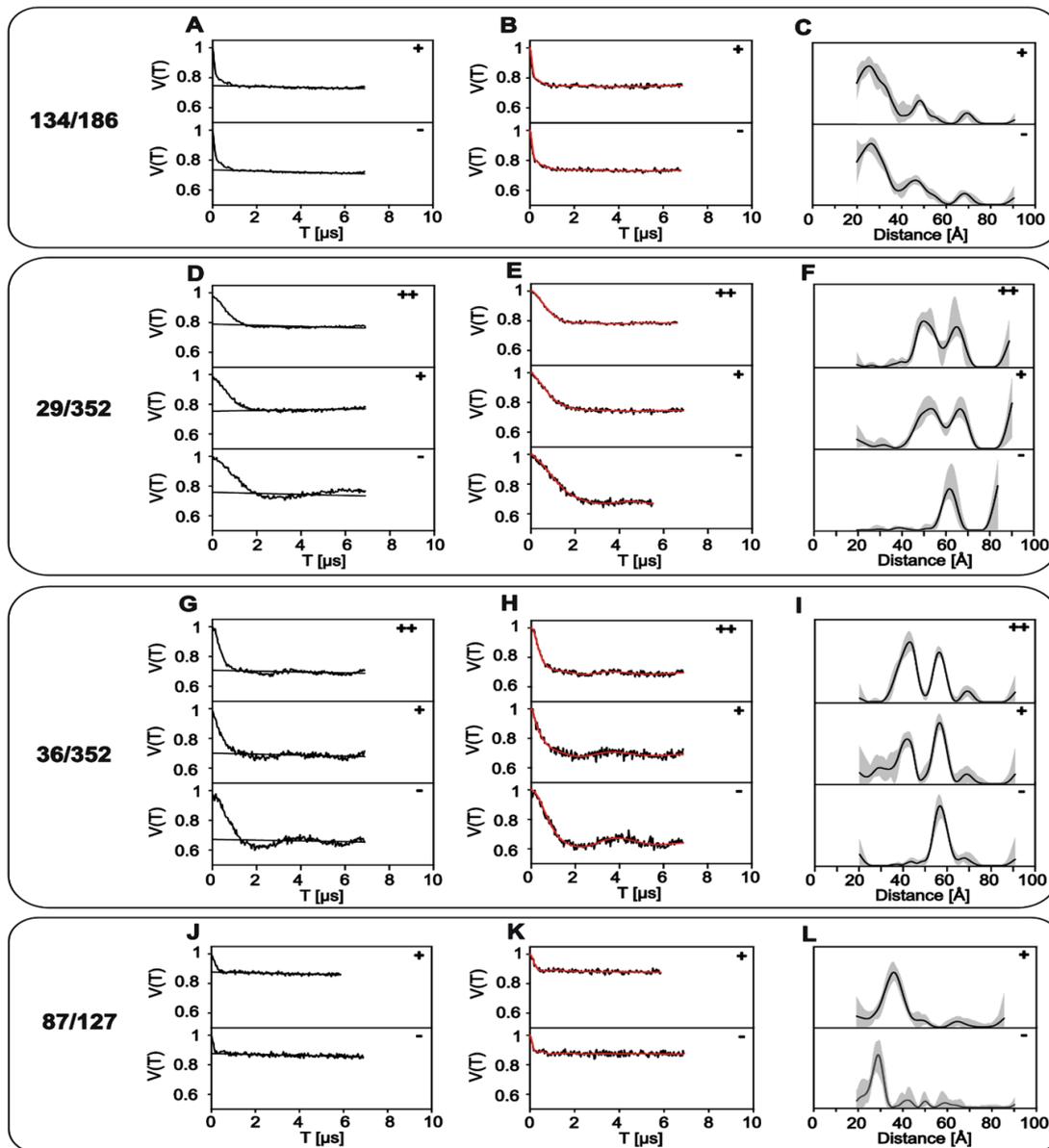


Figure 5.6.9: PELDOR/DEER data of MaleE variants from DEERanalysis. **A/D/G/J** Raw PELDOR/DEER time traces for apo (-) and holo (+, 1 mM; ++, 10 mM maltose) measurements of each double variant. The background, which was used for correction of the signal, is indicated as black line. **B/E/H/K** Background-corrected PELDOR/DEER time traces (black) and fits of the signal (red). **C/F/I/L** Distance distributions from PELDOR/DEER time traces (black) with validation of the distribution (grey).

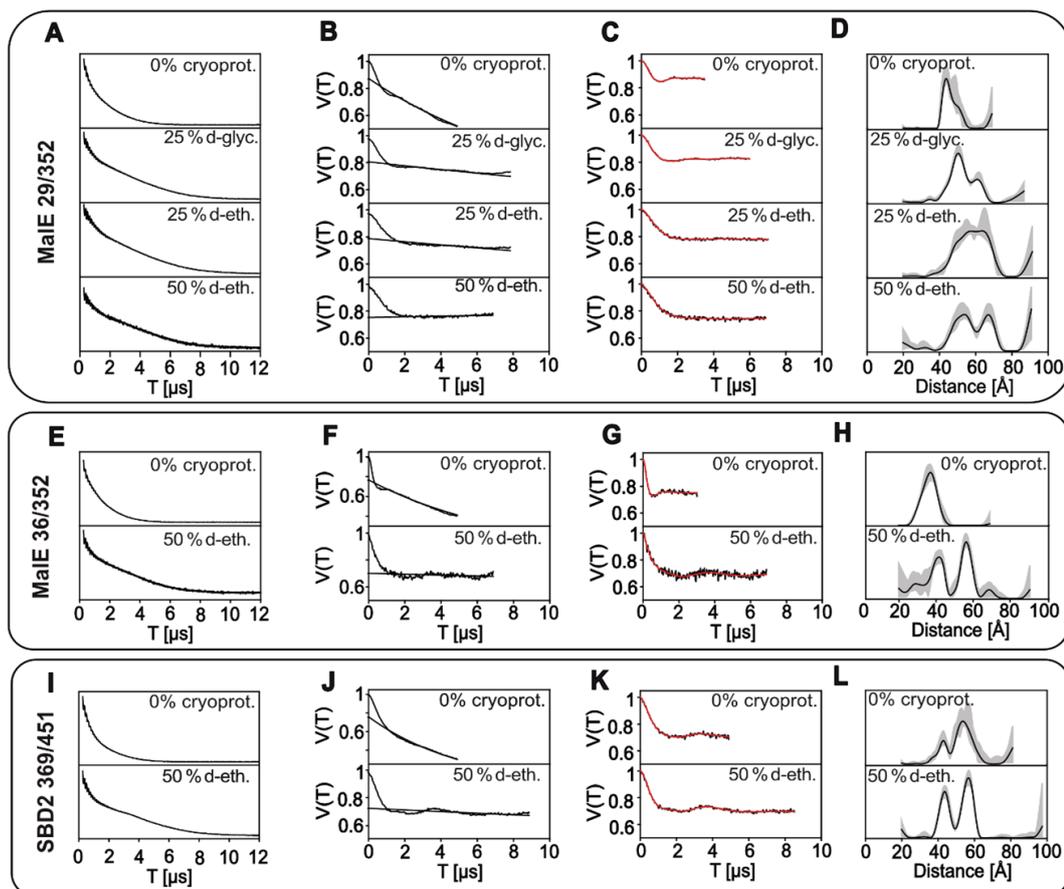


Figure 5.6.10: Data of PELDOR/DEER experiments with variations of cryoprotectant. **A/E/I** 2PESEEM (2-pulse electron spin echo envelope modulation) spectra for each variant with variations of cryoprotectant and the original 50% d-ethylene glycol measurements. **B/F/J** Raw PELDOR/DEER time traces for each measurement. The background, which was used for correction of the signal, is indicated as black line. **C/G/K** Background-corrected PELDOR/DEER time traces (black) and fits of the signal (red). **D/H/L** Distance distributions from PELDOR/DEER time traces (black) with validation of the distribution (grey).

Pair	Alexa Fluor 555 – Alexa Fluor 647	TMR – Cy5
Orientation factor κ^2	2/3	2/3
Average refractive index	1.4	1.4
Extinction coefficient	265,000 1/(M cm)	250,000 1/(M cm)
Quantum yield donor	0.14	0.11
Overlap integral	$8.12 \times 10^{15} \text{ nm}^4 / (\text{M cm})$	$1.13 \times 10^{16} \text{ nm}^4 / (\text{M cm})$
R_0	51 Å	52 Å

Table 5.6.5: Förster radius calculation for Alexa Fluor 555 – Alexa Fluor 647 and TMR-Cy5. Overview of all used parameters (measured/literature).

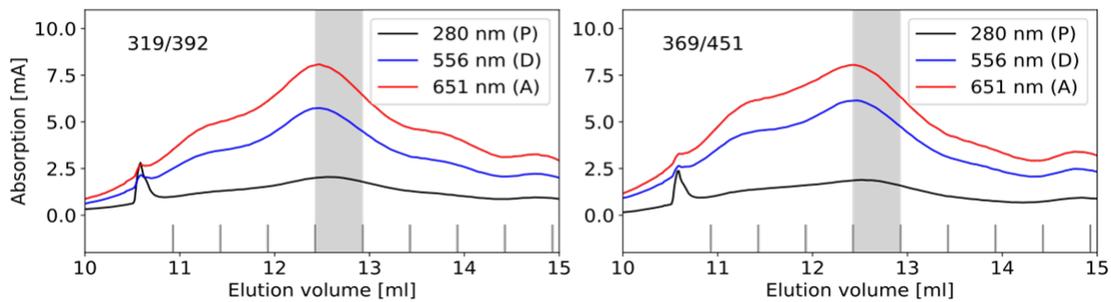


Figure 5.6.11: Size exclusion chromatography of SBD2 variants labelled with Alexa Fluor 555 – Alexa Fluor 647. Absorption profile of the size exclusion chromatography (ÄKTA, Superdex 75 Increase 10/300 GL, GE Healthcare) for all tested SBD2 variants 319/392 and 369/451 to monitor protein concentration (280 nm) and Alexa Fluor 555 (552 nm) / Alexa Fluor 647 (650 nm). The grey area indicates the fraction used in the smFRET experiments, where labelling efficiencies of $>90\%$ was achieved for all samples.

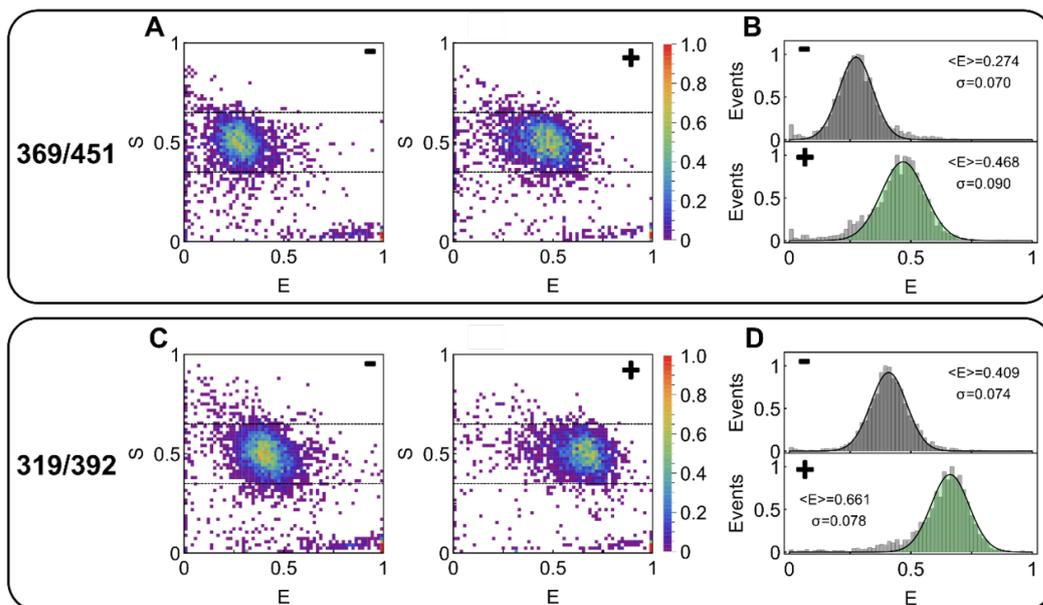


Figure 5.6.12: smFRET data of SBD2 with Alexa Fluor 555 – Alexa Fluor 647. **A/C** ES-2D-Histograms of SBD2 variants 369/451 and 319/392 in apo state (-) and holo state (+). **B/D** 1D-E-Histograms extracted from the ES-Data for apo (grey) and holo (green) are fitted with a 1D-Gaussian distribution. Mean $\langle E \rangle$ and standard deviation σ are labelled. The numbers of considered bursts N in apo/holo states are **B** 2050/2189 and **D** 3377/2060, respectively.

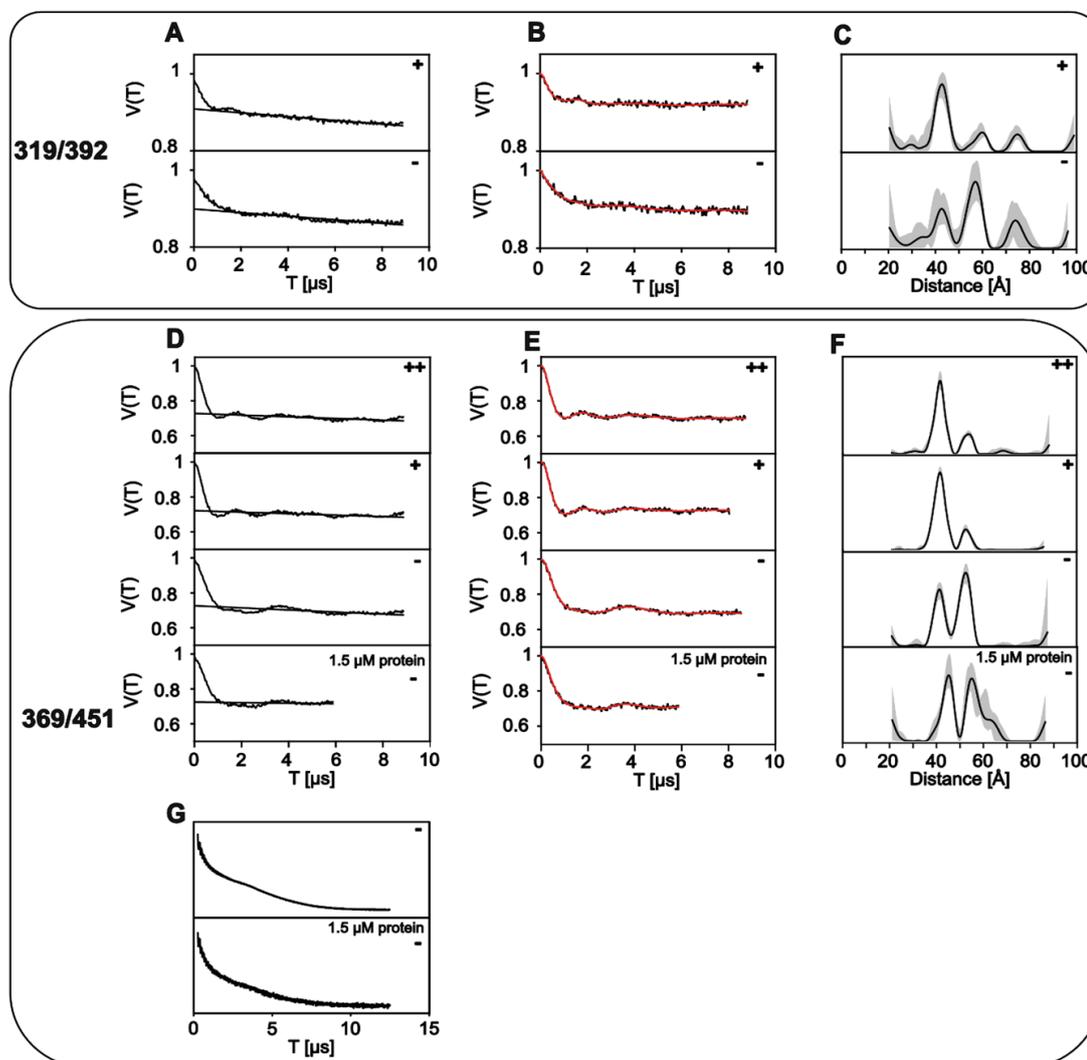


Figure 5.6.13: PELDOR/DEER data of SBD2 variants from DEERanalysis. **A/D** Raw PELDOR/DEER time traces for apo (-) and holo (+ and ++) measurements of each double variant. The background, which was used for correction of the signal, is indicated as black line. **B/E** Background-corrected PELDOR/DEER time traces (black) and fits of the signal (red). **C/F** Distance distributions from PELDOR/DEER time traces (black) with validation of the distribution (grey). **G** 2PESEEM spectra of apo measurements for 369/451 for 15 μM and 1.5 μM protein concentrations.

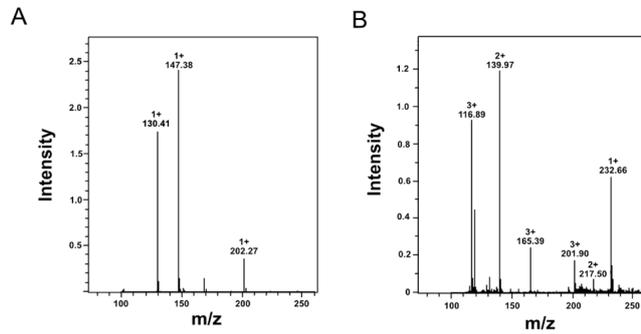
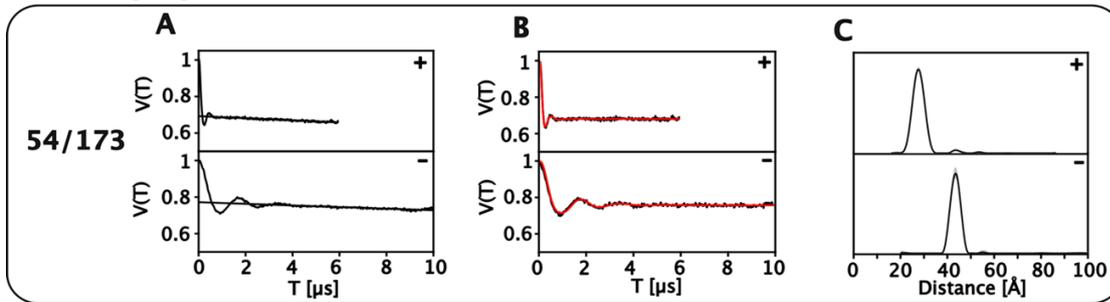


Figure 5.6.14: LC-MS for detection of glutamine. **A** Mass spectrum of glutamine in standard protein buffer as positive control ($MW_{\text{glutamine}}$: 146.15 g/mol). **B** Mass spectrum of supernatant after precipitation and centrifugation of SBD2 protein sample.

with cryo-protectant



without cryo-protectant

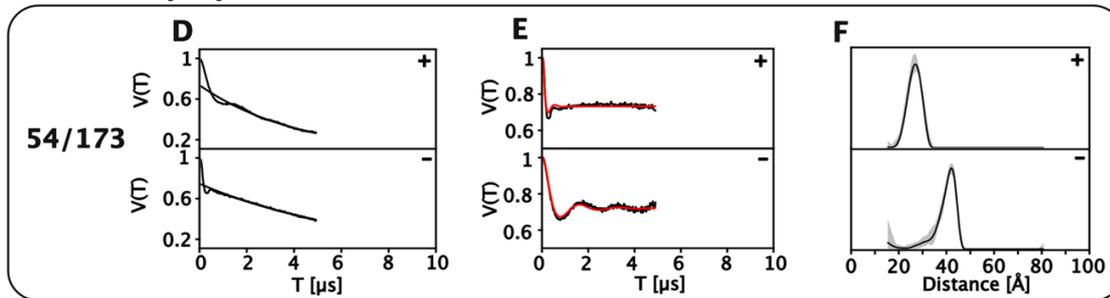


Figure 5.6.15: Comparison of PELDOR/DEER measurements on SiaP from *V. cholerae* with and without cryo-protectant. **A/D** Raw PELDOR/DEER time traces for apo (-) and holo (+, 1 mM Neu5Ac) measurements of each double variant. The background, which was used for correction of the signal, is indicated as black line. **B/E** Background-corrected PELDOR/DEER time traces (black) and fits of the signal (red). **C/F** Distance distributions from PELDOR/DEER time traces (black) with validation of the distribution (grey).

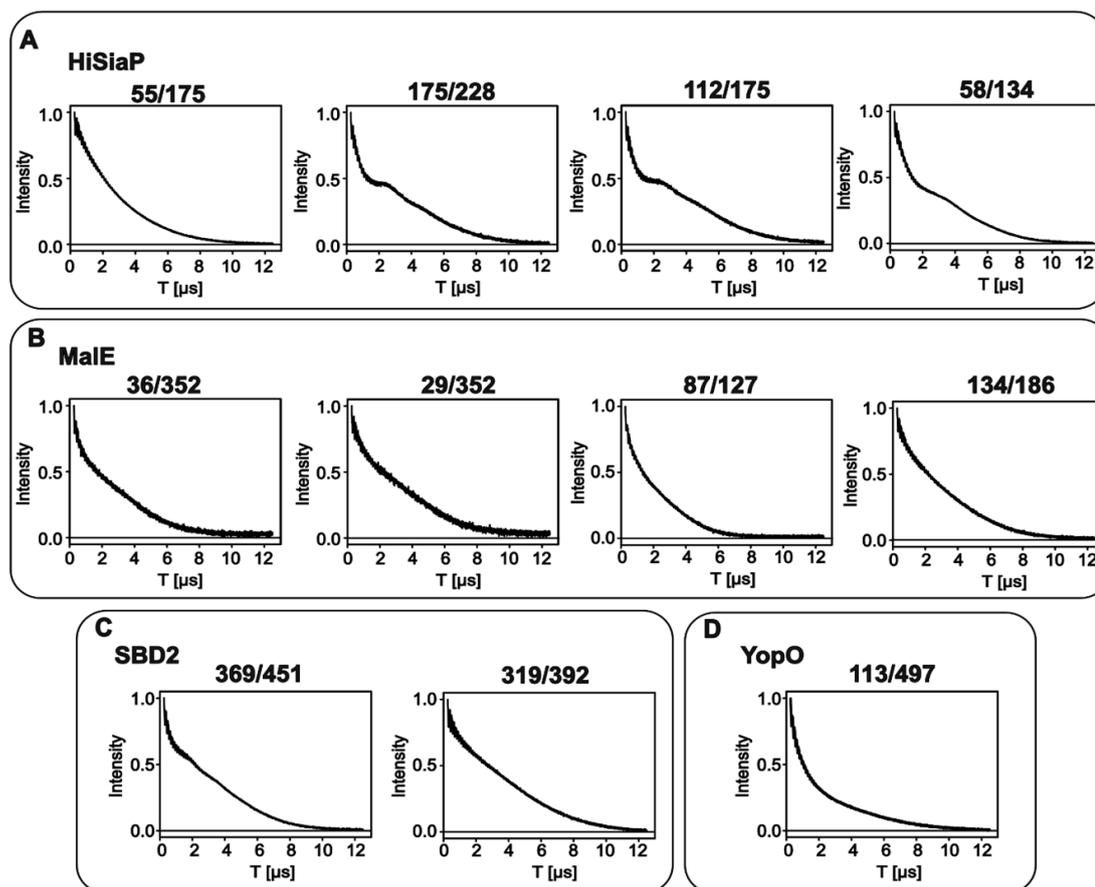


Figure 5.6.16: 2-pulse-ESEEM spectra for all double spin labeled proteins. **A** 2-pulse-ESEEM spectra for spin labeled HiSiap double variants. **B-D** Same as **A** but for MalE, SBD2 and YopO spin labeled variants.

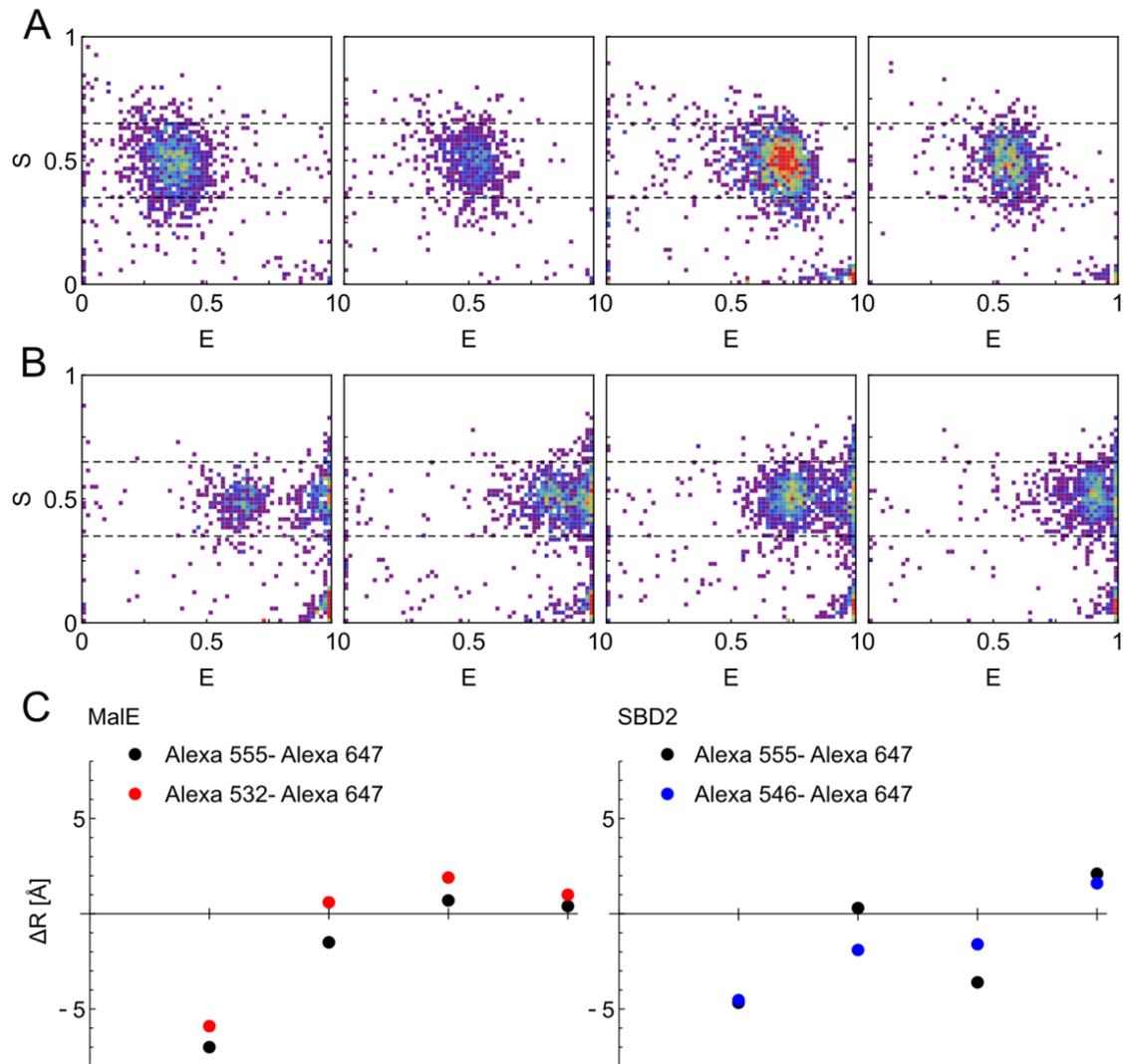


Figure 5.6.17: smFRET data of MalE with Alexa Fluor 532 – Alexa Fluor 647 and SBD with Alexa Fluor 546 – Alexa Fluor 647. A ES-2D-Histograms of MalE variant 29/352 in apo/olo state and variant 87/186 in apo/olo state (left to right). The numbers of considered bursts N are 1532/863/1286/733 (left to right). **B** ES-2D-Histograms of SBD2 variant 369/451 in apo/olo state and variant 319/392 in apo/olo state (left to right). The numbers of considered bursts N are 559/919/1509/1108 (left to right). **C** Deviation of calculated distance and simulated distance for MalE (left) and SBD2 (right) for the four measurement conditions show only small variations (<2 Å) for different fluorophores (see Table 5.6.4).

6 | Distance and dynamics measurements with FRET

Single-molecule Förster resonance energy transfer (smFRET) has evolved into a mature toolkit for the study of biomolecular structure and dynamics in physiologically relevant contexts *in vitro* and *in vivo*. Recently, a worldwide blind benchmark study provided confidence in the accuracy and precision of FRET-derived distances for static model structures such as oligonucleotides. It remains, however, unclear how reliable smFRET can discover and quantify structural dynamics in biomacromolecules – especially in proteins – while simultaneously determining distances. These aspects still hamper smFRET to become fully accepted in the developing field of integrative structural biology, e.g., via contribution of distance constraints or providing information on (dynamic) structural heterogeneity. Here, we take important steps to solve these problems and benchmark the ability of smFRET to discover and quantify structural dynamics and to accurately determine distances in proteins in a dynamic setting. We identified two model proteins with distinct conformational dynamics to address the following problems: (i) discovery and quantification of conformational dynamics in proteins, (ii) precision and accuracy of inter-dye distances in proteins with slow conformational motion and (iii) aspects related to stochastic labelling, long-term storage and stability. In a blind study comprising 18 labs, we successfully confirmed the ability of smFRET to discover and quantify conformational dynamics on different timescales from seconds to milliseconds. Furthermore, we also quantified FRET-derived distances and distance uncertainties for stochastic labelling of proteins that undergo slow structural changes on the sub-second time scale. In this challenging setting, we demonstrate reproducible FRET efficiency values with uncertainties of less than ± 0.06 corresponding to a distance precision of ≤ 2 Å and an accuracy ≤ 5 Å. This supports the established experimental and data analysis routines for determining FRET efficiencies. Our study provides vital advance for smFRET to become an accepted and standardized method in the prospering toolkit of integrative structural biology.¹

6.1 Introduction

Single-molecule Förster resonance energy transfer (smFRET) has become a powerful and accepted tool to complement classical structural biology techniques with information on heterogeneity and conformational dynamics [198, 206, 308, 309]. It allows the quantita-

¹This chapter is part of the publication *Gebhardt/Agam/Popara et al.* (in preparation). For details of the individual contribution see the “Author Contributions” statement in section 6.4.

tive assessment of interconversion dynamics and heterogeneity of conformational ensembles, which are not accessible by crystallography or cryo-EM, and can also be used to resolve parts of structures “*de novo*” [216, 284, 308, 310]. Over the past two decades, smFRET experiments have opened the avenue for structural studies of bio-macromolecules that go beyond detecting interactions and collecting static snapshots, and have provided unique insights into structure-function relationships of fundamental biological processes, such as DNA replication, repair and maintenance, transcription, protein translation and membrane transport [12, 170, 221]. The success of smFRET measurements is based on a simple experimental approach, where a single donor- and acceptor pair monitor a distance (or a distance change) in the biomacromolecule or complex of interest. The design of the assay is such that the energy transfer efficiency can be directly related to a relevant reaction coordinate of the system, e.g., intramolecular conformational changes or interactions between biomolecules. All this can be done with single-molecule sensitivity using the dynamic distance range of FRET of 3 to 12 nm on timescales ranging from nanoseconds to seconds [12, 170, 221].

Hellenkamp et al. recently presented a worldwide quantitative smFRET study (involving 20 labs) of static oligonucleotide ruler structures that demonstrated the high precision and accuracy of FRET-derived distances [111]. In this initiative, a standardized measurement and data evaluation routine was established to determine FRET efficiency values. The blind study demonstrated a distance uncertainty of less than 6 Å within the sensitivity range of FRET [111]. It could be shown that different experimental methods for determining the FRET efficiency, e.g. from fluorescence lifetimes, or intensities of both diffusing and surface-immobilized molecules, all gave reproducible and consistent FRET efficiencies and resulting distances [111]. Although various factors involved in the conversion of FRET efficiency values into accurate distances remain topics of active discussions (since a distance comparison also involves structural modelling [198, 267, 310]), the results presented by Hellenkamp et al. strongly support the idea that standardized smFRET measurements will be useful for structural biology in the very near future, i.e., the use of FRET-derived distances and distance constraints for the protein database [216].

Here, we take the next step by assessing whether the established procedures translate to biomacromolecules undergoing conformational dynamics and, in particular, proteins. Compared to double-stranded DNA, proteins are much more challenging systems to study due to their instability (problematic storage, potential aggregation), sensitivity to the biochemical environment and experimental conditions, temperature, stochastic labeling of the sample, necessity for biochemical characterization after labeling and the existence of conformational motions. So far, however, there is no generally accepted procedure in the smFRET field to detect and analyze conformational dynamics for proteins and to interpret the distance information obtained from smFRET measurements in a structural context [10, 11]. The key questions to be addressed are: 1) how consistently can smFRET histograms (and derived distances) be determined in different labs for proteins with stochastic labeling, 2) how reliably can smFRET measurements detect and quantify structural dynamics in biomacromolecules and 3) how accurately

can distances be determined for proteins and dynamic systems? Addressing these questions will promote smFRET to become recognized as a pivotal tool in the developing field of integrative structural biology [311] due to its unique ability to provide correlated information on structure and dynamics.

In a blind study involving 18 labs, we investigated how reliably smFRET histograms can be measured for proteins and whether structural dynamics can be detected and quantified. As model systems, we chose the maltose-binding protein (MalE) of *E. coli* and the human U2 auxiliary factor 65 (U2AF) that display conformational dynamics on different timescales. Protein mutants were prepared according to standard procedures in the field, i.e. by stochastically labeling protein double-cysteine mutants to monitor intramolecular distances. The comparison study confirmed the reproducibility of FRET efficiency histograms and the ability of smFRET to detect and quantify conformational dynamics on different timescales ranging from seconds to microseconds. In this challenging setting, we demonstrate reproducible FRET efficiency values with uncertainties of less than ± 0.06 corresponding to a distance precision of ≤ 0.2 nm and an accuracy ≤ 0.6 nm. The reliable detection of conformational dynamics in realistic protein samples in combination with high precision and accuracy supports the established experimental and data analysis routines for determining FRET efficiencies and distances.

6.1.1 Introduction of the model protein systems

We identified two protein model systems for our study, MalE of *E. coli* and the human U2 auxiliary factor (U2AF) protein from the pre-mRNA splicing machinery. Both proteins show conformational motions on different timescales that is altered in the presence of a ligand. In addition, both systems require special attention in the process of performing smFRET experiments due to possible sample instability or aggregate formation, which are both problematic for long-term storage and shipping and also the need to perform a functional test of activity once the protein is labelled.

MalE. Our first target was the periplasmic maltose-binding protein MalE of *E. coli*. This robust model protein, which is widely used in biochemistry and molecular biology [312], is a component of the ATP binding cassette transporter MalFGK2 of *E. coli* [195, 196]. MalE exhibits the classical type II periplasmic-binding protein (PBP) fold [240, 313] composed of two rigid-body lobes connected by a flexible 2-segment hinge (Figure 6.1.1A). The latter allows allosterically-driven conformational changes upon substrate binding (i.e. maltose). For smFRET investigations, double-cysteine mutants of MalE were prepared as described in Section 6.5 (and Supplementary Note 6.7.1) following our previous work [107]. We chose three different mutants (Figure 6.1.1B) to follow the conformation change between the two domains. For MalE K29C-S352C the interdyer distance decreases upon maltose binding, while for MalE D87C-A186C the distance increases. The third mutant MalE A134C-A186C shows no distance change upon substrate binding. The mutants of MalE were stochastically labeled at the given positions with the FRET pair Alexa Fluor 546 and Alexa Fluor 647 to monitor ligand-induced conformational changes and related distances (Figure 6.1.1B) by different implementations of

smFRET, i.e., μ s-ALEX [14, 15] or PIE [16] (Supplementary Figure 6.6.1). Stochastic labeling of the two cysteines results in a mix of donor-acceptor species bearing the two dyes at either cysteine.

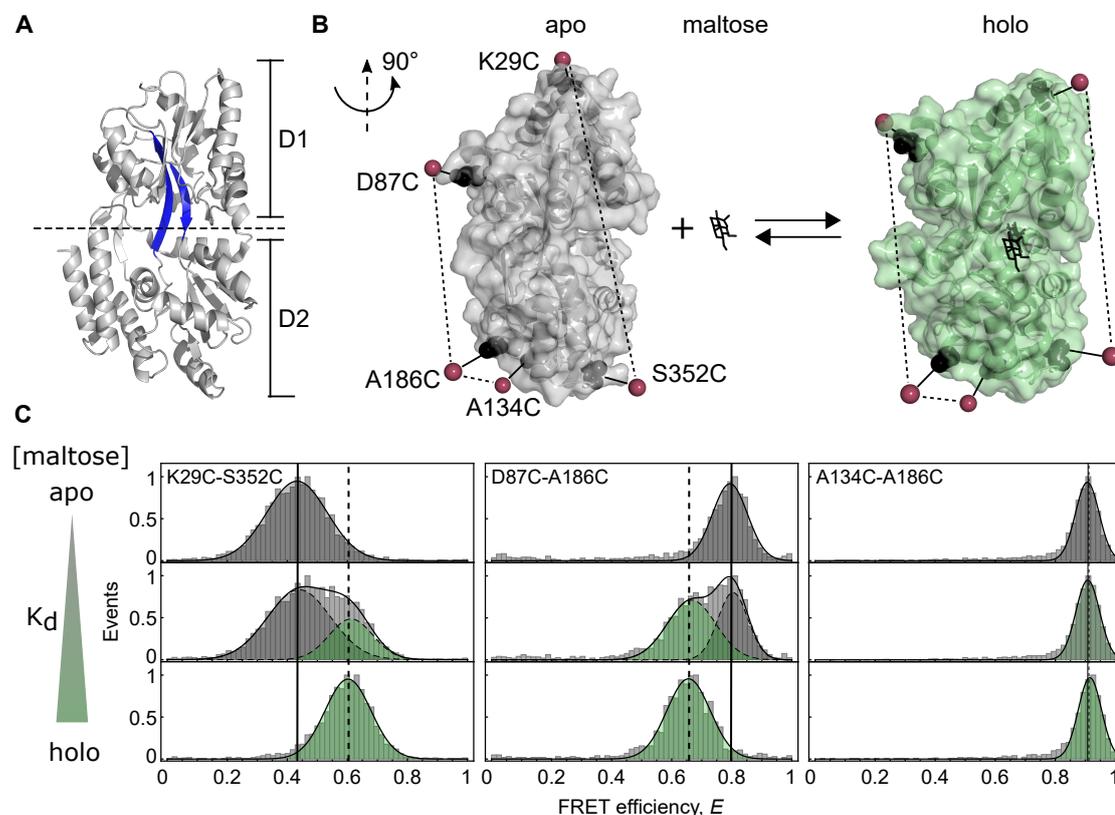


Figure 6.1.1: Experimental design of MalE as a protein model system for sm-FRET studies: **A** Crystal structure of MalE in its ligand-free apo state (PDB 1OMP) with domains D1 and D2 linked by flexible beta-sheets (blue). **B** The crystal structure of MalE (rotated by 90° as compared to A in the apo (grey, PDB 1OMP) and holo (green, PDB 1ANF) states with marked mutations K29C-S352C, D87C-A186C, and A134C-A186C. The estimated mean position of the fluorophores (from an AV3 model [206]) are shown as red spheres. **C** FRET efficiency E histogram for MalE mutant K29C-S352C (left), mutant D87C-A186C (middle), and mutant A134C-A186C (right) in the presence of 0 (top), 1 μ M (middle), and 1 mM maltose (bottom).

Protein stability and functionality was verified by affinity measurements using microscale thermophoresis, MST [208]. All preparations, i.e., MalE-wildtype, unlabeled cysteine mutants and fluorophore-labelled variants, showed an affinity between ~ 1 -2 μ M (Supplementary Figure 6.6.2) consistent with previously published K_d -values for wild type MalE [234, 314]. The stability of the sample was verified by fluorescence correlation spectroscopy (Supplementary Figure 6.6.3), which excluded the presence of larger aggregates in the samples and confirms that MalE is a robust and functional model

protein for a comparison study. Proper ligand binding in all labelled mutants, including MalE A134C-A186C, was verified by MST experiments (Supplementary Figure 6.6.2).

Results of μ s-ALEX experiments are shown in (Figure 6.1.1C), where the impact of different maltose concentrations on the conformational state of MalE are visualized. The mutant K29C-S352C showed a low-FRET apo- and a high-FRET holo-state, which is related to the compaction of the protein upon ligand binding. The MalE mutant D87C-A186C showed the inverse behavior with a high-FRET apo- and a low-FRET holo-state (Figure 6.1.1C). Labelling of MalE within one of the two lobes (A134C-A186C) showed no apparent change of FRET efficiency E , indicating that the selected positions are static or move in the same manner. FRET also directly monitors the biochemical function of this mutant suggesting that it is well preserved, since the addition of 1 μ M maltose, a concentration close to the K_d -value [315], shows equal probabilities to occupy either conformational state. This is true for both the K29C-S352C and D87C-A186C mutants. For the “neutral” A134C-A186C mutant, the interprobe distance is consequently not altered upon ligand binding. It is evident from a comparison of the distances of the C_α atoms (black) or the mean dye positions (red, Figure 6.1.1B) and the smFRET results that the relative distance changes were qualitatively consistent with the predictions from the crystal structures. The interdye distances in the chosen mutants thus cover a large part of the dynamic range of smFRET (Figure 6.1.1B/C) and are additionally modulated by ligand maltose.

Finally, the ruler character of FRET and independence of the FRET efficiency values from orientation factors were verified for all mutants by steady-state and time-resolved anisotropy experiments to check whether all combined residual anisotropies are below 0.22 (Supplementary Figure 6.6.4 and Supplementary Table 6.6.1/ 6.6.2), using the criteria for orientation independent measurements established before [305, 308].

U2AF65. For the second protein system, we chose U2AF65, the large subunit of the U2 auxiliary factor (U2AF) of the pre-mRNA splicing machinery (spliceosome) as it is known to undergo fast dynamics in the absence of RNA and becomes quasi static when a strong poly-pyrimidine tract (Py-tract) is bound to the protein. It is involved in identification of the 3' splice site of the pre-mRNA [316]. The RNA recognition motifs 1 and 2 (RRM1,2) of U2AF65 represent the minimal RNA binding regions (Figure 6.1.2A) [317], they are connected by a flexible linker and bind single-stranded Py-tract RNA with high affinity (K_d of ~ 0.3 μ M for U9 RNA, the ligand used in this study). The two motifs fluctuate between ensembles of detached and compacted conformations (Figure 6.1.2B, left panel) and stabilize in an open conformation upon binding of RNA containing a poly-pyrimidine tract (Figure 6.1.2B, right panel) [318].

For these investigations, we chose the already published double cysteine mutant (L187C and G326C) of the minimal RRM1,2 construct, where we have previously verified that the used mutations and the fluorophore labeling does not affect the protein function and affinity to the Py-tract [104, 319]. The construct contains one cysteine on each RNA recognition motif, which were labelled stochastically with the donor and acceptor dye pair ATTO532-ATTO643. Protein preparation and labeling were carried out as de-

scribed previously [104, 319] and summarized in the methods section (Section 6.5). In the apo-state, a single broad FRET peak (Figure 6.1.2C) is observed with a peak near 75% FRET efficiency. Under closer examination, this peak is actually a dynamically averaged peak as the system fluctuates between multiple open and closed conformations. In the holo-state, an open configuration is stabilized with a FRET efficiency near 44%. In contrast to what we published earlier, this batch of protein contains a significant fraction of protein in the apo-configuration despite the high excess of ligand (Supplemental Figure 6.6.5). The fast dynamics, the RNA ligand (which is sensitive to degradation) and fraction of non-functional protein makes it an extremely challenging test case for protein-based smFRET studies.

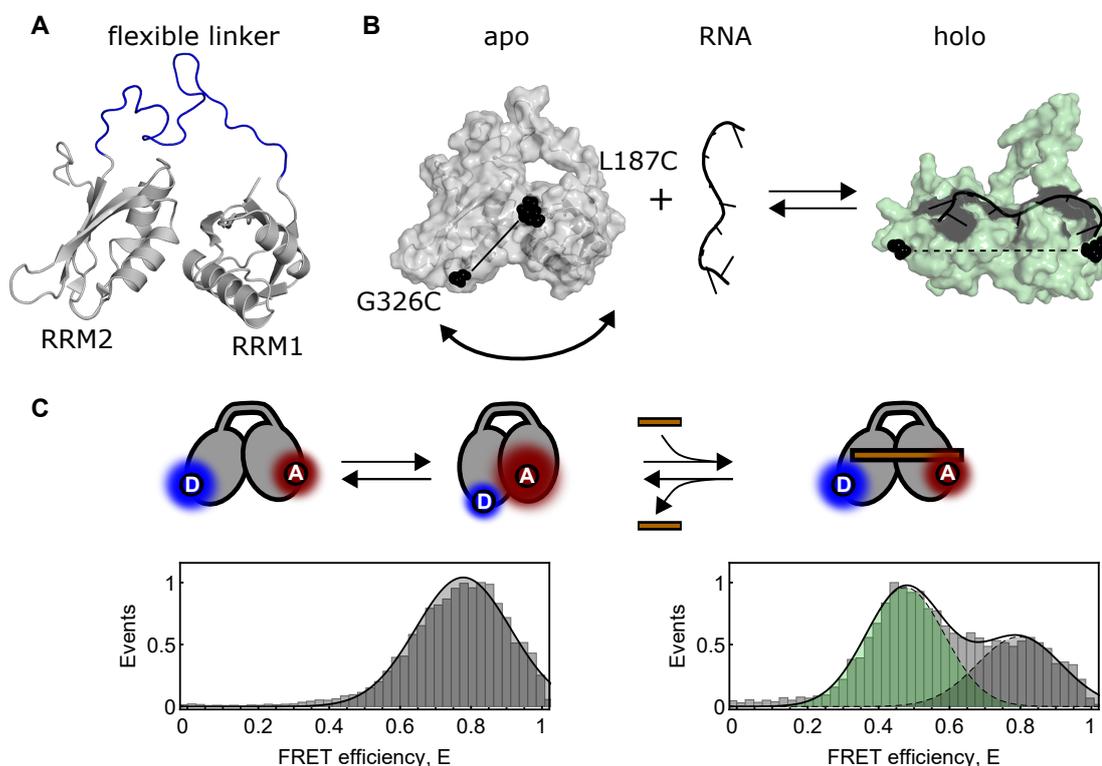


Figure 6.1.2: Experimental system U2AF(RRM1,2) **A** Ribbon representation of the NMR structure of U2AF(RRM1,2) [318]. The two tandem motifs are depicted as RRM1 and RRM2 and the flexible coiled coil linker connecting the motifs is highlighted in blue (PDB 2YH0). **B** RRM1,2 NMR structure in the apo-state (grey, PDB 2YH0) and the holo-state (green, 2YH1). Positions used for introducing cysteine mutations for labeling at L187 in RRM1 and G326 in RRM2 are depicted as black spheres. U8 RNA ligand is shown as sticks. **C** SmFRET efficiency histogram of RRM1,2 in the apo (left) and holo conformation (right). 5 μ M of U9 RNA was used to obtain the holo-state FRET histogram for RRM1,2. Note the decrease in FRET efficiency after binding of RNA to the RRM1,2. Gaussians fits to the different subpopulations are shown in green for the 0.44 FRET efficiency conformation and in grey for the 0.75 FRET efficiency conformation.

Due to the complexity of the dynamics (a detailed description of the dynamics is given in the Supporting Information, Supplementary Figure 6.6.6, Supplementary Note 6.7.2), we concentrate here on the comparison of the measured smFRET histograms and the model-independent relaxation times of the dynamics.

6.2 Results

6.2.1 Precision and reproducibility of smFRET values on proteins

In this blind comparison study, 18 labs performed smFRET experiments on diffusing molecules for three MalE mutants (Figure 6.1.1C). For the comparison, the proteins were labelled with the Alexa Fluor 546-Alexa Fluor 647 (donor-acceptor) pair and studied on a confocal microscope setup using either ALEX or PIE. We adapted the data analysis routine similar to that of Hellenkamp et al. [111] to convert apparent FRET efficiency values E_{app} into setup-independent FRET efficiency E values. Briefly, for each mutant, the photon counts in the donor and acceptor channels were extracted for the apo- and holo-state, i.e., in the absence and presence of 1 mM maltose (Figure 6.1.1). Subsequently, accurate FRET values were determined by correcting the measured values for background, donor leakage, acceptor direct excitation, and differences in donor and acceptor quantum yields and detection efficiencies (Supplementary Figure 6.6.7 and Supplementary Table 6.6.3/6.6.4) [110]. For this study, we asked the individual laboratories to use an averaged γ -value for all experiments conducted in the respective lab. We provide a comparison between the averaged and individual γ -corrections in Supplementary Fig. S7c and will come back to this point in the discussion. For the sake of comparison, the mean FRET efficiencies, E , were determined by the different labs using the peak of a 2D Gaussian fit to the E-S histogram. The mean FRET efficiency value and standard deviation determined from the results of the various labs for all MalE mutants in the absence and presence of its ligand maltose are summarized in Figure 6.2.1. The results from all labs showed an excellent agreement and reproducibility. All experiments combined reveal standard deviations of the FRET efficiency E values of less than ± 0.06 for apo- and holo-MalE. All labs observed the expected maltose-induced conformational change for the front-side and back-side mutants (compare Figure 6.1.1C and 6.2.1) with a precision similar to that found for static dsDNA [111], indicating that the samples did not degrade during transport at 4 °C (see Section 6.5.2). The absolute FRET efficiency values vary significantly more than the relative FRET efficiency change upon ligand binding indicating that the major variation arises from systematic correction errors (Figure 6.2.1B).

For U2AF65, the results of the smFRET solution measurements are shown in Figure 6.2.2. Due to the complexity of the dynamics, a subset of eight groups measured the second protein. The overall trends are consistent. All groups show a broad single FRET peak in the apo-state that shifts to a more extended and open conformation in the holo-state [318, 319], another example of ligand-induced conformational changes in proteins. To get a first approximation of the consistency of the results, we fit the spFRET

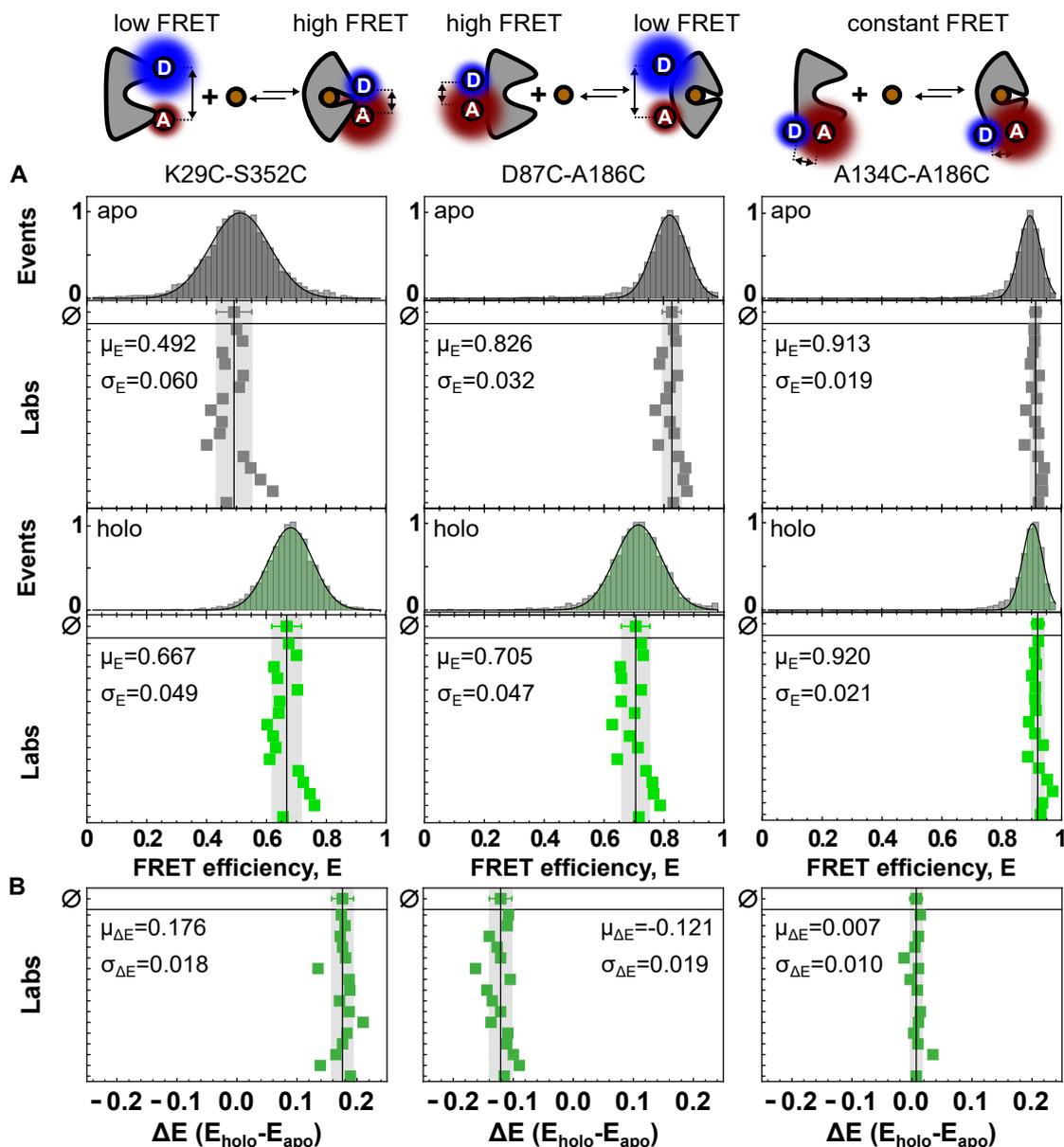


Figure 6.2.1: Determination of the precision of FRET efficiency values in different MalE mutants. **A** FRET efficiency E values of MalE in the apo state (without maltose, top) and in the holo state (in the presence of 1 mM maltose, bottom) for mutations K29C-S352C (left), D87C-A186C (middle), and A134C-A186C (right). The combined data from all labs (upper data point) is illustrated by the mean (solid line) with standard deviation (grey area). **B** FRET efficiency difference between apo and holo state in **A** shows significantly reduced variation.

histograms to a single Gaussian for the apo-state and to two Gaussians for the holo state. The peak values for the data from the different labs are plotted in Figure 6.2.2A. A more detailed comparison in Figure 6.2.2B shows the plot of smFRET histograms from the individual laboratories as well as a combined FRET efficiency distribution using all data, which is illustrated by the mean (solid line) and standard deviation (pale area) for both states. The spread in the results is broader in the case than for the measurements on MalE. The standard deviation of ± 0.05 was found for apo whereas for holo state they were ± 0.05 for high FRET and ± 0.04 for low FRET efficiency. On the one hand, more discrepancies between the measurements from the different labs are not unexpected as the dynamic biological systems is much more challenging and more sensitive to the experimental conditions. The dynamics in the apo-state is temperature dependent (Supplemental Figure 6.6.5) as well as the K_d of the protein-RNA interaction may shift with temperature. To create “real-world” like conditions, we did not put any constraints on what “room temperature” should be nor did we inform the different laboratories that the ligand was RNA. On the other hand, we were unsatisfied on the obtained consistency. To gain insights to the causes of the discrepancies between the laboratories, we collected the raw data from the different laboratories and had one person analyze the data using an agreed upon procedure (Supplementary Note 6.7.3). When comparing measurements on the apo protein, we found that the detection correction factor γ played a very prevalent role in describing the deviations between measurements (Supplemental Figure 6.6.8). Due to the dynamics for the apo protein, it is not possible to use the fluorescence lifetime for determination of the γ -factor nor does the apparent single population provided a good measure for the γ -factor [18, 64]. In this case, it was best to combine both apo and holo measurements and fit the S vs. E plot as described by Lee et al. [110]. We have to assume that there are no changes in the quantum yield of the fluorophores upon binding of the RNA. For measurement on the holo system, the peaks of the two observable states agree well (when using the appropriate γ -factor) but the amplitudes of the peaks vary. Many factors can influence the amplitude of the two peaks. These factors include temperature, buffer components such as salt concentration or stabilizers (e.g. BSA), protein activity, ligand degradation (Supplemental Figure 6.6.9). Although we cannot determine exactly why there is a discrepancy with one of the labs, we did observe an increase in a higher than normal stoichiometry population for these measurements. This population is present in the original sample preparation, but to a lower extent. When removing this population from analysis, the data are again more consistent (Supplemental Figure 6.6.9).

6.2.2 Discovery and quantification of conformational dynamics in proteins via smFRET

As dynamics are prevalent when investigating proteins, as a next step we asked the various groups in this blind study to evaluate whether the protein systems they studied were static or dynamic and which method they used to come to this conclusion. The results are given in Table 6.2.1. In the end, one is interested in whether the protein exhibits conformational fluctuations. As FRET reports on the energy transfer between the fluorophores,

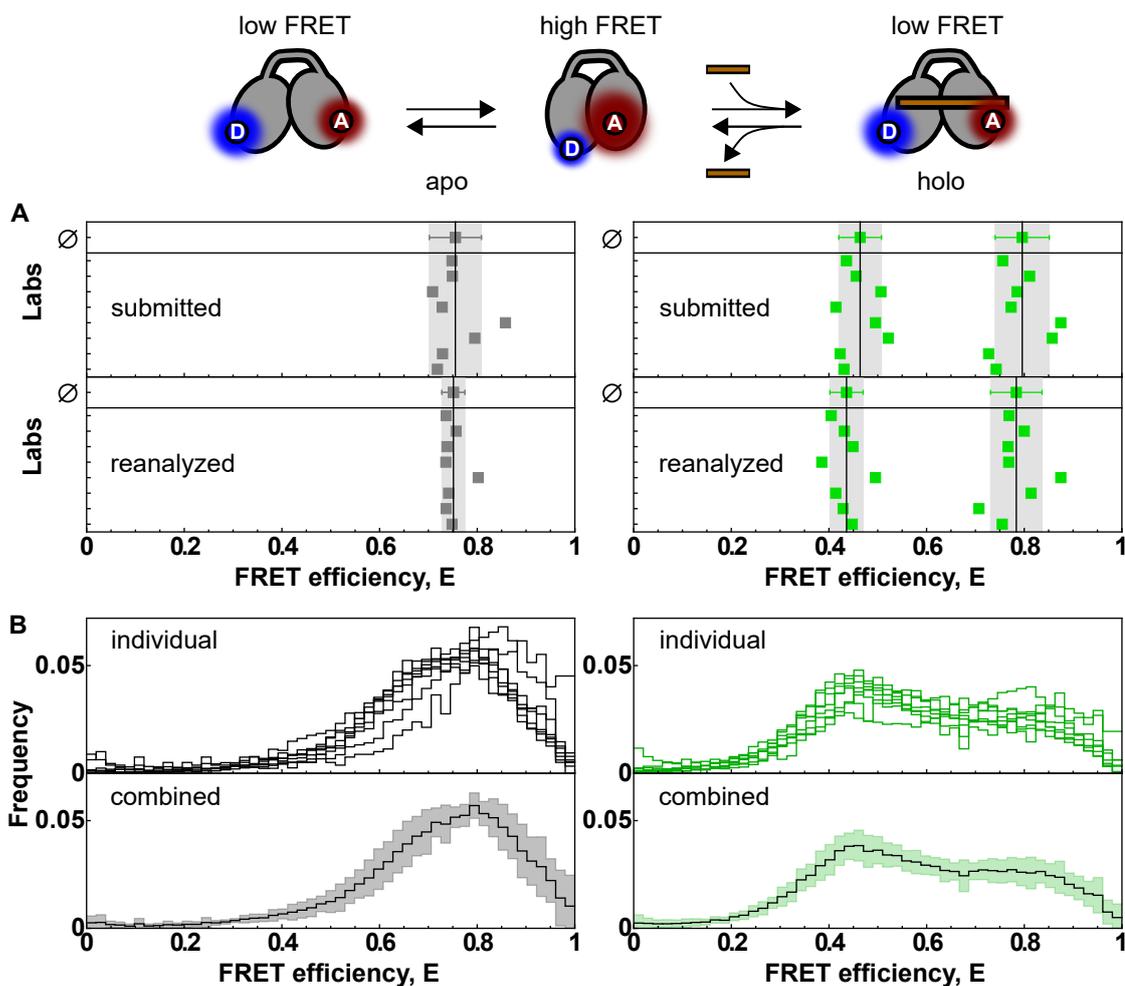


Figure 6.2.2: A comparison of FRET efficiency histograms from 8 different laboratories for U2AF65 (RRM1,2) **A** The apo-state (in grey), represented in the dashed square, undergoes fast exchange between an ensemble of detached structures. A slower exchange occurs between the dynamic detached ensemble and a compacted conformation (PDB 2YHO). The holo-state (PDB 2YH1), shown with a bound U8 RNA ligand in sticks, assumes a compacted open conformation. Mean FRET efficiency values of U2AF65 from single Gaussian fits to data without RNA ligand reported from 8 labs (top, left) and after re-analysis of the collected data (bottom, left). The mean value from all data sets is 0.75 ± 0.05 , shown above with the corresponding standard deviation in grey. After reanalysis by a single person, the agreement improved to 0.75 ± 0.02 . SmFRET efficiency histogram comparisons of RRM1,2 in the holo-state with $5 \mu\text{M}$ of U9 RNA (right). A 1-D FRET efficiency histogram was fit with two Gaussians to roughly approximate the FRET efficiency of the different subpopulations of bound and unbound proteins (top, right) with little improvement after reanalysis (bottom, right). **B** Individual FRET efficiency histograms reported by participating 8 labs for apo (top, left) and holo (top, right) and the combined data from all the 8 reporting labs illustrated via the mean (solid line) with standard deviation (pale area) in the bottom.

the linker dynamics of the fluorophores as well as photophysical contributions can also contribute to a “dynamic” FRET signal. We see a quantitative analysis of protein dynamics using FRET as one of the major challenges facing the FRET community but it also one of the strengths that such dynamics can be quantified on the single-molecule level using FRET. There are multiple approaches for detecting submillisecond dynamics in burst analysis data including the burst-variance analysis (BVA) [65], FRET-2CDE [320], dynamic PDA [64], or plots of FRET efficiency E versus fluorescence-weighted average donor lifetime $\langle \tau_{D(A)} \rangle_F$ (E-tau plots) [321]. Here, we focus on the use of BVA and E-tau plots. Both techniques visualize dynamic FRET behavior by comparing the burst properties against theoretical expectations (black solid lines in Figure 6.2.3). For BVA, the standard deviation of the individual bursts is compared to the theoretical shot-noise limits. Due to the dynamics of the protein, the variance of the FRET signal within a burst is higher than expected from shot noise, which becomes visible as a deviation from the shot-noise semi-circle. For the E-tau plots, the observed FRET efficiency determined via intensity is a species weighted average and the position along the axis depends on the fraction of time spent in the respective states whereas the fluorescence lifetime is a photon-weighted average (as only a single lifetime can be determined from the lifetime data) and is weighted towards the lifetime of the lower FRET state as the majority of photons are emitted from the donor in the low FRET efficiency state. Hence, the data are shifted to the right of the static FRET line (shown in black). For dynamics between two and three distinct states, a theoretical FRET lines can be calculated and added to the plots [64]. We previously showed that MalE exhibits slow ligand-driven dynamics on the sub-second timescale with the interconversion of high- and low-FRET states (ref. [107] and Supplementary Figure 6.6.10). Here, we investigated whether the apo- and holo-state of MalE is undergoing dynamics faster than or on the timescale of the diffusion time. Both techniques reveal that MalE has the conformation of MalE is quasi-static in both the apo- and holo-structures on the millisecond timescale (Supplementary 6.6.11/6.6.12). This is indicated by the center position of all data sets on or close to the static lines for both approaches (Figure 6.2.3A/B). Interestingly, minor deviations from the static FRET line appear when other dye combinations are used (Supplementary Note 6.7.4, Supplementary Figure 6.6.13). Three of the participating groups came to the conclusion that MalE is (to some extent) dynamic, although this statement needs to be qualified. Proteins are dynamic systems and the attached dyes with flexible linkers are also dynamic. Some groups see a slight deviation of the FRET data from what would be expected for a static system. However, it is on the limit of what can be detected using the given techniques and is still not clear whether this is an indication of conformational dynamics, linker-dye dynamics, or measurement artefacts and will be the topic of future collaborative work. All groups are in agreement that, on the millisecond timescale, the proteins are about as static as a protein can be.

In contrast, all groups clearly found U2AF65 to be dynamic (Table 6.2.1), although this conclusion is not obtainable from the 1D FRET histogram alone. Both the ligand-free apo state and RNA-bound holo state show pronounced deviations from the expected behavior for static molecules both in the BVA and E-tau plots (Figure 6.2.3C/D). A

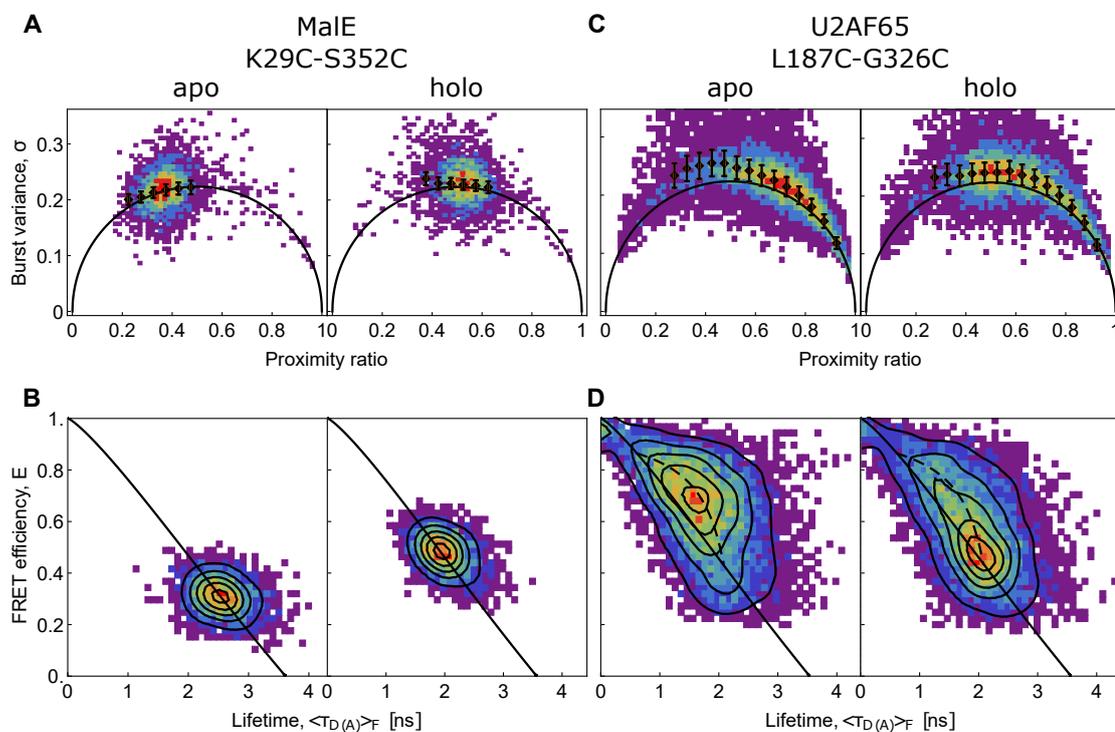


Figure 6.2.3: Discovery and characterization of conformational dynamics in MalE and U2AF65 on the sub-millisecond timescale **A** The Burst variance analysis of the MalE mutant K29C-S352C labelled with ATTO532-ATTO643 without maltose (left) and with 1mM maltose (right). The burst variance analysis is based on a photon binning of 5 photons for bursts with stoichiometry between 0.4 and 0.7. **B** The FRET efficiency E versus fluorescence-weighted average donor lifetime $\langle \tau_{D(A)} \rangle_F$ of the same measurement as in **A**. The donor only population was excluded for clarity in the plot. The E-Tau plot shows that, under both ligand concentrations, the FRET species lie on the static FRET line (black). **C** The burst variance analysis similar to **A** with the sample U2AF65 mutant L187C-G326C labelled with Atto532-Atto643 without RNA ligand (left) and with 5 μ M of U9 RNA (right). **D** The E-Tau plot similar to **B** with same data as plotted in **C**. The BVA and E-Tau plots in **C** and **D** show clear deviation from the static FRET line (black) as indicated with a dynamic FRET line (dashed, black).

Lab	Method	Sub-ms dynamics in MalE (sample 1/2/3)	Sub-ms dynamics in U2AF65 (apo / holo)
#1	BVA + E-Tau	no / no / no	yes / yes
#2	BVA + E-Tau	no / no / no	yes / yes
#3	BVA + E-Tau	no / no / no	yes / yes
#4	n/a	n/a	n/a
#5	E-Tau (fFCS)	no / no / -	yes / yes
#6	BVA	no / - / yes	yes / yes
#7	BVA	no / no / no	yes / yes
#8	n/a	n/a	n/a
#9	BVA	no / no / no	n/a
#10	E-Tau (fFCS, PDA)	yes / yes / yes	yes / yes
#11	BVA + E-Tau	no / no / no	yes / yes
#12	BVA	no / no / no	n/a
#13	fFCS	no / no / yes	n/a
#14	BVA + E-Tau	no / no / no	n/a
#15	BVA + E-Tau	no / no / no	n/a
#16	BVA	no / no / no	n/a
#17	n/a	n/a	n/a
#18	n/a	n/a	n/a

Table 6.2.1: Overview of dynamic statement for MalE and U2AF65. MalE sample 1: 29C/352C, MalE sample 2: 87C/186C; MalE sample 3: 134C/186C. “-“ = no statement, “n/a”=not applicable due to experimental limitations (setup equipment & established evaluation routines).

number of groups also applied a FRET FCS or filtered FCS analysis of the burst analysis data to obtain model-free insights into the rates of the kinetic transitions (Supplementary Figure 6.6.6). The results are given in Supplementary Table 6.6.5. Two relaxation rates were detected for the five groups that performed the analysis. As discussed in the Supplementary Note 6.7.2, the actual dynamics is complex and a more quantitative analysis depends on the model chosen to evaluate the data. Also, whether the holo state is really static or dynamic is difficult to determine. The individual RRM s have affinities of $\sim \mu\text{M}$, meaning that occasional rapid fluctuations away from the open state may happen, which is difficult to distinguish from the temporary sticking of the dyes. The fraction of underlying apo-protein in the measurement also makes it difficult to analyze the holo-dynamics in great detail. As mentioned above, the U2AF65 is a very challenging test case for measuring protein dynamics. However, the clear detection of dynamics and consistent estimations on the timescales for the dynamics considering the complexity of the system is very encouraging and demonstrates that smFRET is a reliable tool for investigating dynamic systems. For a truly comparative study on the quantification of the dynamics, the participants agreed that it would be better to start with a well-defined dynamic system where conformational changes occur on different timescales in proteins [322, 323] or oligonucleotide structures [112, 320, 324], which will be the subject of a future study.

The discovery and characterization of conformational dynamics via smFRET has sub-

stantial implications for the final interpretation of donor-acceptor distance values in comparison to structural models of the system. In summary, the above comparison allowed us to benchmark the ability of the smFRET experiment to (i) reliably discover or exclude conformational dynamics and (ii) to assess the timescales of interconversion. From the smFRET histograms, accurate distances can be derived as discussed previously [111].

6.3 Discussion

The presented results of our blind study of 18 labs clearly demonstrate that smFRET can provide accurate distances of conformational states and reliable information on dynamics in proteins. Despite potential problems for proteins samples, such as aggregation or degradation during storage or shipping (stochastic labeling and heterogenous dye environments), we could achieve a similar reproducibility as for stable oligonucleotide structures [111]. The high level of consistency for qualitative detection of sub-millisecond dynamics in U2AF65 and exclusion thereof for MalE shows that the community is well positioned to deal with dynamic protein systems. For a truly comparative study on the quantification of the dynamics, i.e., rate constants, the participants agreed that it would be better to start with a well-defined dynamic system where conformational changes occur on different timescales in proteins [111, 322, 323] or oligonucleotide structures [112, 320, 324], which will be the subject of a future study. All experiments were performed using established experimental procedures and analyzed with freely-available data analysis routines [59, 60, 103, 198, 206], indicating that the presented experiments and the conclusions drawn are accessible to groups with similar technical expertise.

This demonstrates that smFRET is a robust technique to visualize conformational states, changes and interconversion dynamics in a reproducible fashion and the retrieved information can be used to quantitatively determine distances.

Though the results are very encouraging, we learned a lot about the difficulties and potential pitfalls that can occur when performing smFRET experiments on proteins during this study. In the next sections, we will highlight the most challenging steps and unresolved problems that require careful attention when performing smFRET experiments to assess structure and dynamics of proteins.

Obtaining instrument-independent FRET-efficiency values. The correction of FRET-efficiency values for systematic errors introduced by the apparatus will directly be carried over to the conversion of apparent- into fully corrected FRET efficiency E values. These systematic errors are largely dominated by the γ correction factor when using ratiometric determination of E (see Supplementary Figure 6.6.14). This factor becomes especially important at small or large distances, i.e., high or low FRET efficiencies, and introduces systematic errors in both E and the derived distances. Thus, obtaining reliable γ factors is crucial for reliable FRET experiments. Unfortunately, a global γ factor only provides accurate corrections in cases where the fluorophore environment is (almost) identical between the different protein conformations and mutants studied. We

think that the work by Lee et al. [110], Kudryavtsev et al. [18] and Hellenkamp et al. [111] lays a solid foundation for this important step in quantitative use of smFRET. Both implementations of solution-based smFRET (used in this manuscript) μ sALEX and nsALEX/PIE, see Supplementary Figure 6.6.1, are capable of detecting sub-millisecond conformational fluctuations of biomolecules and to yield accurate FRET values. The advantage of μ sALEX is that it is less expensive to implement than nsALEX/PIE and that the excitation is less challenging to the fluorophores in terms of photophysical processes (e.g., bleaching or blinking) [199]. For certain acceptor fluorophores, the lower peak laser intensities used avoid pushing the fluorophores into the triplet state where it is no longer able to FRET. It is also easier to adjust the duty cycle of excitation such that more time is spent collecting photons for calculation of the FRET efficiency as for monitoring the photophysical properties of the acceptor. nsALEX/PIE have the advantage that they provide, in addition to the stoichiometry information available from μ sALEX, the fluorescence lifetime information of the donor, the acceptor upon activation via FRET and acceptor fluorophores with direct excitation [18]. With the lifetime information, it is possible to perform the E- τ plots or look for changes in acceptor lifetime due to protein or ligand interactions with the dye. Also, when a sub-optimal acceptor fluorophore is chosen such that the population of the triplet state of the acceptor increases due to the pulsed excitation, the γ -factor determined directly from the data incorporates this effect and a quantitative FRET efficiency is still obtained. In addition, the faster alternation timescale of nsALEX/PIE, one has better time-resolution for fluorescence correlation spectroscopy analyses and it ensures that quickly diffusing molecules are still in similar regions of the PSF during the measurement, decreasing potential spread in the stoichiometry distribution.

smFRET for quantification of structural dynamics. All groups were able to detect fast dynamics in the U2AF65 systems and could agree that MaleE shows no dynamics on the timescale of diffusion, while fast dynamics, if present, are at or below the detection limit. Both μ sALEX and nsALEX/PIE can detect the millisecond dynamics via BVA. In addition, nsALEX/PIE can utilize the lifetime information to identify dynamics. For all methods, it is easier to detect the dynamics when the number of FRET states is low (two or three) and when different conformational states have clearly different FRET efficiencies. To map the “detectability” of a dynamic exchange, we analyze the maximum dynamic shift can be observed for a two-state system for the BVA and E- τ methods as a function of the FRET efficiency E_1 of the first state and the contrast between the states ΔE . Interestingly, the E- τ method is more sensitive to dynamics at high FRET efficiencies, while BVA performs better at low FRET efficiencies. If dynamics were present in our samples, the small dynamic shifts would place them on the edge of what is detectable by single-molecule FRET. The precise detection limit depends hereby on the quality and statistics of the measurement that determines how accurately the position of a population can be determined in the two-dimensional plots. The accurate determination of dynamic shifts for the E- τ method further requires a proper calibration of the static FRET line, which has to be corrected for the linker dynamics, and a proper

correction of the data using a correct γ -factor. For BVA, on the other hand, the dynamic shift is reduced if dynamics are fast ($\lesssim 100$ μ s).

smFRET for structural biology. In general, converting FRET efficiency E values into distances (or distance distributions) is a challenging and error-prone procedure, for which there is not yet a clear consensus. As a result, there is not yet a standardized protocol for this essential step. A key problem is the determination of the Förster radius for dyes attached to proteins (see Supplementary Note 6.7.6), since the dyes can encounter very specific environments which may influence their mobility and/or photophysical properties, and consequently, the Förster radius. Moreover, the presence of conformational sub-populations, in which dyes experience different surroundings, makes this even more challenging. For each conformational state, information on the donor fluorescence spectrum, donor quantum yield, acceptor absorption spectrum, the refractive index of the donor-acceptor intervening medium and their averaged relative dipole orientation needs to be obtained. Each of these parameters is difficult to assess, because bulk approaches are currently used to determine them, and these may overlook changing properties of the involved sub-populations. Particularly challenging are the refractive index and the dye orientation factor (κ^2), since there are no direct experimental approaches established to determine their values. The refractive index is in general not determined experimentally, but estimated to be 1.4 as an intermediate value between the values of 1.33 for water and 1.5 for proteins [170, 325] that keeps the error to a minimum. To determine κ^2 , an assumption is made that the rotational relaxation of both the donor and acceptor dye occurs within the donor fluorescence lifetime. For this special case, an average $\kappa^2 \approx 2/3$ can be assumed. This assumption, however, does not hold for large organic dyes or dyes with very short lifetimes, since rotational correlation times (determined indirectly from fluorescence anisotropy decays) will be similar to the fluorescence lifetime [305]. The distance-error resulting from an incorrect κ^2 can be estimated [305], and some studies have shown the profound effect the actual κ^2 values can have on the FRET-derived distance estimation [326]. Furthermore, the conversion of a mean FRET efficiency into distance information has to consider the use of mean values as non-physical parameters vs. more physical meaningful distance distributions. In smFRET experiments, it is usually not a single distance that determines the mean transfer efficiency $\langle E \rangle$ but it is rather a distribution of distances over which there is no clear consensus on how this should be done. The measured distance from accurate FRET yields the separation of the two dyes. The organic fluorophores used in smFRET are often far from satisfying the criterion of point dipoles, especially in reference to the distances they are intended to probe. The conversion from the inter-dye distance to the inter-residue distance is non-trivial and depends on multiple parameters. To a certain extent, the recently developed nano-positioning system (NPS) provides a reasonable solution for this conversion and future developments of the NPS procedure may make it possible to produce the accessible volume of the dyes with more physical and non-uniform distributions [267, 310]. In our analysis, we identified the apo-state of Male K29C-S352C as an example, where a larger deviation from the expected distance value

was observed (Figure 6.2.3D). This deviation can have different reasons: The anisotropy values for this mutant indicate restricted motion, since r_{inf} is above the acceptable upper limit of ~ 0.22 (see above, Supplementary Table 6.6.1/ 6.6.2). In this regime, distance uncertainties of $\pm 10\%$ are caused by R_0 due to a non-ideal orientation factor [305]. However, the observed distance-deviation is still well below what is expected assuming static averaging [308, 327]. Alternatively, it might be explained by FRET-dynamics, which are likely caused by photophysical effects or conformational exchange of MalE between apo and holo state occurring without addition of ligand. The method of multiparameter fluorescence detection (MFD) [18, 206] experimentally addresses several of the abovementioned caveats and enables three-dimensional protein structure determination. Nevertheless, the dyes, which are large hydrophobic systems of conjugated rings, can perturb and affect the overall structure of the macromolecule they are attached to. Lastly, the differently-labelled macromolecules may show altered biochemical activities, thermodynamic stabilities, or structural features, which may lead to problems in three-dimensional structure convergence. Therefore, either at least two different dye-pairs should be used to determine a distance or a network of distances has to be measured to exclude outliers. This discussion shows that there are indeed various factors and aspects on the use of smFRET for structural biology that require attention and future work. Our encouraging results demonstrate, however, very clearly that smFRET is a reliable tool for discovery and quantification of structural dynamics in proteins while reliably and accurately determining distances in this challenging setting. This study brings smFRET a big step closer to become an integral part of a combined structural biology toolkit.

6.4 Acknowledgements and Author Contributions

Acknowledgements. Work in the lab of T.C. was financed by an ERC Starting Grant (ERC-StG 638536—SM-IMPORT), the Center of Nanoscience Munich (CeNS), Deutsche Forschungsgemeinschaft within GRK2062 (project C03) and SFB863 (project A13), LMUexcellent, the Center for integrated protein science Munich (CiPSM), and an Alexander von Humboldt postdoctoral fellowship (to N.Z.). Research in the labs was financed by the following sources: ERC grant agreement no. 681891 (to T.H.)². We finally thank D. Griffith for reading and commenting on the manuscript.

Author contributions. C.G. and T.C. initiated the study. C.G., G.A., M.P., A.B., D.C.L., C.A.M.S. and T.C. designed research. D.C.L., C.A.M.S. and T.C. supervised the project. C.G. performed FCS experiments on MalE variants, (time-resolved) anisotropy experiments, accessible volume calculation for MalE, R_0 -determination and performed error simulations including theory development. R.M. designed, cloned and purified all MalE variants. H.S.K. and M.S. provided U2AF65. C.G. and G.A. performed labeling of MalE and U2AF65 variants for shipment to participating labs. G.A. reperformed the analysis on the provided raw data for U2AF65 from all laboratories. G.M. per-

²Incomplete list at the moment of thesis submission

formed MST experiments. M.d.B. performed confocal scanning experiments for surface-immobilized MalE variants. B.A., N.C., T.D.C., D.G., T.H., A.H., J.H., V.H., C.G.H., Th.H., C.J., D.K., H.-S.K., A.K., G.K., K.K., E.M., J.M., J.M., R.Q., N.R., M.S., M.S., J.S., T.S., A.S., J., P.T., S.W., and N. Z. performed experiments and analyzed smFRET data. C.G., M.P. and G.A. consolidated data collection of participating labs and C.G., G.A., A.B. and M.P. created figures for the manuscript. C.G., G.A., A.B., C.A.M.S., D.C.L. and T.C. wrote the manuscript in consultation with all authors.

6.5 Methods

6.5.0.1 Sample preparation

MalE. The MalE wildtype plasmid was obtained and modified as described previously [107]. In this process, the sequence encoding the signal peptide (amino acid 1-26) was removed and the remaining MalE sequence sub-cloned with an in-frame His-tag in the pET20b vector (Novagen, EMD Millipore). MalE variants with cysteine mutations (indicated by residue numbers throughout the text) were constructed by QuickChange II Site-Directed Mutagenesis with the PfuUltra High-Fidelity DNA polymerase (Agilent Technologies). All primers used are given in prior section Supplementary Note 6.7.1. MalE mutants were over-expressed in BL21 (DE3) pLysS cells (F- ompT lon hsdSB (rB-mB-) gal dcm (DE3) pLysS (CamR)). Cells harboring expression plasmids were grown at 37°C until the OD600 was 0.6-0.8, and MalE expression was induced with 0.25 mM Isopropyl- β -D-thiogalactopyranosid (IPTG). To purify the proteins, the cells were lysed with an ultrasonic homogenizer (Branson Digital Sonifier: 0.5 s on/0.5 s off for 10 min in total, amplitude min. 25 %) in washing buffer 1 (50 mM Tris-HCl pH 8, 1 M KCl, 10% glycerol, 10 mM imidazole, 1 mM dithiothreitol (DTT)) containing 2 mM phenyl-methylsulfonyl fluoride (PMSF). The soluble protein fraction was obtained by sequential centrifugation (3040 xg, 40 min, 4°C, followed by 263,000 xg, 60 min, 4°C) and the supernatant was loaded on an Ni Sepharose 6 Fast Flow resin (GE Healthcare). Bound proteins were washed sequentially with 10 column volumes (CV) of washing buffer 1 and washing buffer 2 (50 mM Tris-HCl pH 8, 50 mM KCl, 10% glycerol, 20 mM imidazole, 1 mM DTT). Elution was done with 10 ml of elution buffer (50 mM Tris-HCl pH 8, 50 mM KCl, 10% glycerol, 250 mM imidazole, 1 mM DTT). Protein fractions were desalted (50 mM Tris-HCl pH 8, 50 mM KCl, 1 mM DTT), dialyzed (50 mM Tris-HCl pH 8, 50 mM KCl, 50% glycerol; 1 mM DTT), and stored at -20°C.

RRM1,2. Human RRM1,2 L187C-G326C mutant (U2AF65-148-342) was provided by Hyun-Seo Kang and Michael Sattler, Technical University Munich at Garching. It was purified as described in Mackereth et al. [318].

6.5.1 Protein labeling

MalE. Proteins were labelled as described previously [102]. In brief, cysteine positions were chosen based on the open and closed x-ray crystal structures of MalE (1OMP,

1ANF) such that the full dynamic range between 3.5 and 8 nm was covered. The cysteines were stochastic labelled with the maleimide derivative of the dyes Alexa Fluor 546 and Alexa Fluor 647 (ThermoFischer Scientific). His-tagged proteins were incubated in 1 mM DTT to keep all cysteine residues in a reduced state and subsequently immobilized on a Ni Sepharose 6 Fast Flow resin (GE Healthcare). The resin was incubated 2-4 h at 4°C with 25 nmol of each fluorophore dissolved in labelling buffer 1 (50 mM Tris-HCl pH 7.4-8.0, 50 mM KCl, 5% glycerol) and subsequently washed sequentially with 1 CV labelling buffer 1 and 2 (50 mM Tris-HCl pH 7.4-8.0, 150 mM KCl, 50% glycerol) to remove unbound fluorophores. Bound proteins were eluted with 500 ml of elution buffer (50 mM Tris-HCl pH 8, 50 mM KCl, 5% glycerol, 500 mM imidazole) The labelled protein was further purified by size-exclusion chromatography (ÄKTA pure, Superdex 200 Increase 10/300 GL, GE Healthcare) to eliminate remaining fluorophores and remove soluble aggregates. For all proteins, labelling efficiency was higher than 70% and donor-acceptor pairing at least 20%.

RRM1,2. Maleimide conjugates of Alexa Fluor 488-Alexa Fluor 647, Alexa Fluor 546-Alexa Fluor 647 and ATTO532-ATTO643 fluorophore pairs were coupled to cysteines of RRM1,2 L187C-G326C mutant stochastically as performed previously in Voithenberg L.V. et al. [170].

6.5.2 Sample handling

Male. The labelled proteins were stored in 50 mM Tris-HCl pH 7.4, 50 mM KCl with 1 mg/ml bovine serum albumin (BSA) at 4°C for less than 7 days. The samples were shipped on ice in a cooling box with overnight shipping to avoid unnecessary freezing and thawing. The samples were at the order of 10 to 100 nM concentration and dilution buffer for apo and holo measurement were provided. ALEX experiments were carried out by diluting the labelled proteins to concentrations of ≈ 50 pM in 50 mM Tris-HCl pH 7.4, 50 mM KCl supplemented with the ligand maltose. The coverslip was passivated with 1 mg/ml BSA in PBS buffer before adding the sample. The measurements were performed without any photo-stabilizer to keep the measurements as simple as possible to avoid any error (e.g. photo-stabilizer can go bad; slightly different concentrations change quantum yield; etc.).

RRM1,2. Labeled protein was measured at ~ 40 -100 pM on custom-build confocal set-up with multi-parameter fluorescence and pulsed-interleaved excitation [18]. Measurements were performed in 20 mM Potassium phosphate buffer pH 6.5, 50 mM NaCl. Purchased U9 RNA (Biomers.net GmbH, Ulm, Germany, IBA Solutions for Life Sciences, Göttingen, Germany) was dissolved in RNA-free water and added directly in the solution at a concentration of 5 μ M for holo measurements.

6.5.3 Data analysis

Data acquisition, burst extraction and intensity based corrections was performed for confocal measurements similar to the procedure as described in section 3.4 (see also Hellenkamp et al. [111]) on different ALEX or PIE setups in the individual labs (see SI Figure 6.6.1).

MFD analysis of fast conformational exchange. Two-dimensional histograms of the FRET efficiency E and donor fluorescence lifetime $\langle \tau_{D(A)} \rangle_F$ (Fig. 6.2.3 and 6.6.13) were created for single molecule measurements using MFD in combination with Pulsed-Interleaved Excitation (PIE). MFD, introduced by Eggeling et al. [17] combines spectral and polarized detection with picosecond pulsed lasers and Time Correlated Single Photon Counting (TCSPC), allowing the simultaneous detection of intensity, lifetime, anisotropy and spectral range of the fluorescence signal of single molecules. Using PIE, it becomes additionally possible to monitor the acceptor dye [18]. Selection of the bursts was performed using a countrate treshold based filter of 5 kHz [17]. Static FRET lines were calculated using following equation $E_{\text{static}} = 1 - \tau_{D(A)}/\tau_{fl}$ modified for linker dynamics [64]. Interdye distance was distributed normally with $\sigma = 6 \text{ \AA}$. Deviations of FRET populations from the static FRET line can indicate potential fast ($< 1 \text{ ms}$) conformational exchange. For more accurate analysis MFD is combined with a calibration-free lifetime-based analysis using the slope of the fluorescence decay [150]. Time-resolved FRET analysis accurately resolves the distance heterogeneities by revealing multiple components in the decay of the curve and recovers their specific species fractions and FRET rate constants. Dynamics are thus detected from the presence of multiple components in the sub-ensemble decay of a single FRET population.

6.6 Supplementary Figures and Tables

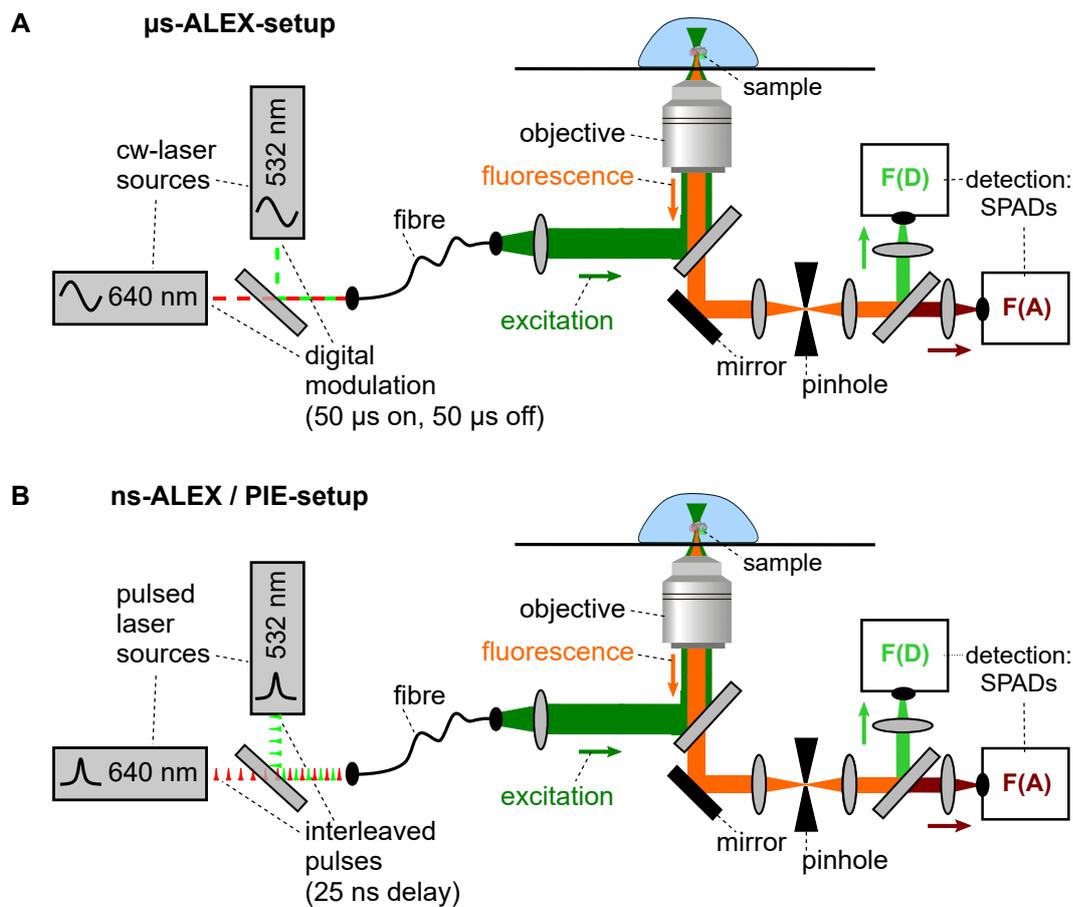


Figure 6.6.1: Scheme of experimental setup. **A** Scheme of a confocal microscope setup used for the acquisition of diffusion-based smFRET data in alternating-laser excitation (ALEX) mode; 532 and 640 nm cw-laser sources alternately excite the sample for 50 μ s through a microscope objective, F(D) and F(A) indicate the donor and acceptor detection channels, respectively. **B** Scheme of a confocal microscope setup used for the acquisition of diffusion-based smFRET data in pulsed interleaved excitation (PIE) mode; 532 and 640 nm pulsed laser sources alternately excite the sample with picosecond-pulses delayed by 25 ns.

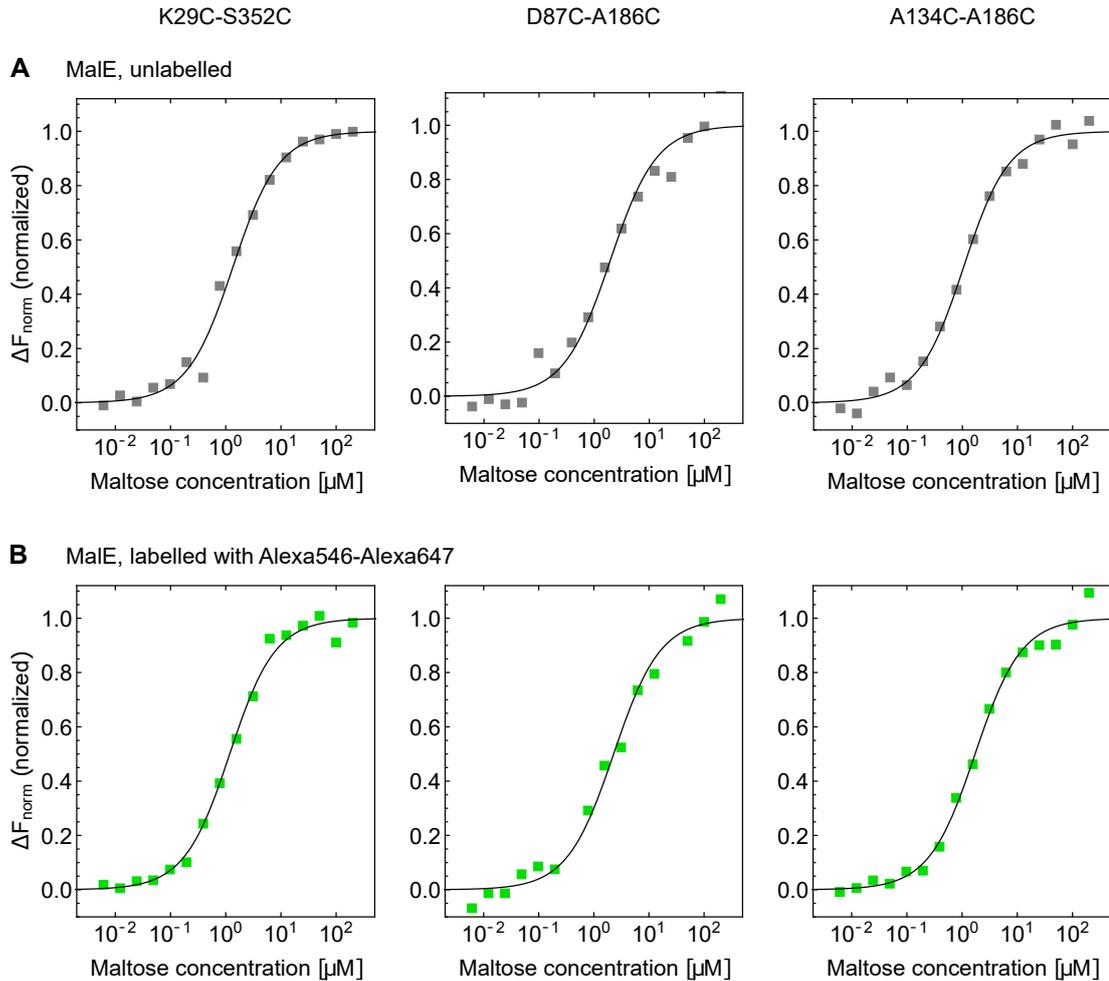


Figure 6.6.2: Binding affinity measurement. **A** Binding affinities are measured with microscale thermophoresis (Monolith NT.LabelFree, Nanotemper), where the ratio of fluorescence before and after heating $\Delta F_{norm} = F_{cold}/F_{hot}$ is recorded at different maltose concentrations [208]. Data points show ΔF_{norm} normalized to minimal and maximal for unlabelled mutants K29C-S352C (left), D87C-A186C (middle), and A134C-A186C (right). The curves are fitted with a standard model for receptor-ligand kinetics $\Delta F_{norm} = \left(K_d + c_P + c_{malt} - \sqrt{((K_d + c_P + c_{malt})^2 - 4c_P c_{malt})} \right) / (2c_P)$, where K_d is the dissociation constant, c_P the protein concentration set to $0.25 \mu\text{M}$ in the experiment, and c_{malt} the maltose concentration (see Section 3.4.2). The fits to the binding model (solid line) give K_d -values of $1.2 \mu\text{M}$ (left), $1.8 \mu\text{M}$ (middle), and $0.9 \mu\text{M}$ (right), respectively. **B** Similar measurements to **A** with same mutants labelled with Alexa546 and Alexa647 give K_d -values of $0.9 \mu\text{M}$ (left), $0.6 \mu\text{M}$ (middle), and $1.6 \mu\text{M}$ (right), respectively.

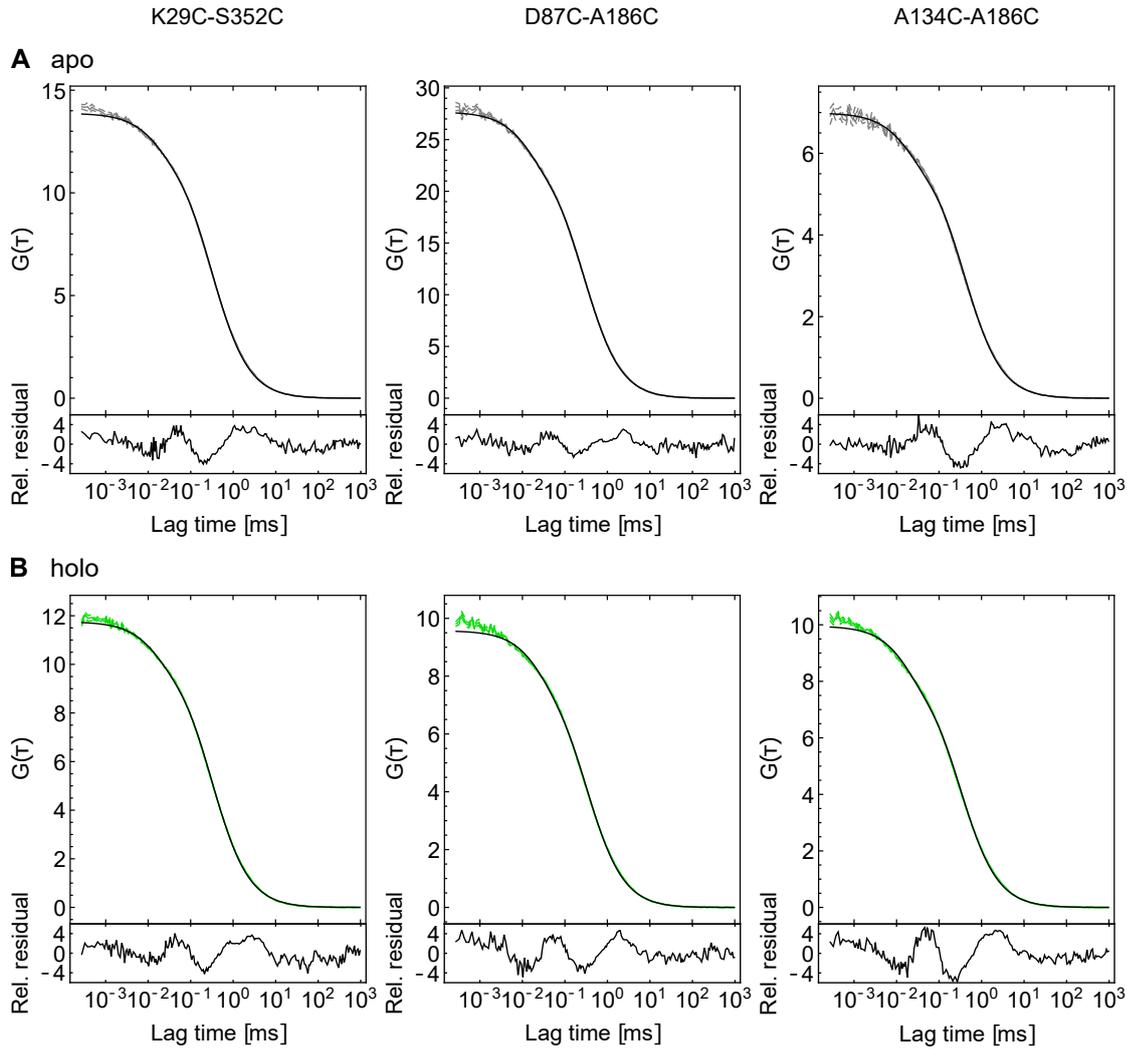


Figure 6.6.3: FCS diffusion time. Fluorescence Correlation Spectroscopy (FCS) measurement for mutants K29C-S352C (left), D87C-A186C (middle), and A134C-A186C (right) in **A** apo state and **B** holo state. The two orthogonally oriented polarizations in the donor detection channel were correlated to remove the detector afterpulsing effect at short lag times (green line). All correlation curves were fitted with a standard model including a triplet fraction (black line): $G(t) = (1 + T(e^{-t/\tau_{trip}} - 1)) / \left(\left(1 + \frac{t}{\tau_{diff}}\right) \sqrt{1 + \frac{t}{\kappa^2 \tau_{diff}}} \right)$, where T is the triplet fraction, τ_{trip} the triplet lifetime, κ^2 elongation factor of the confocal volume, and τ_{diff} the diffusion time [328]. The confocal instrument was calibrated using free Alexa 546 dye with diffusion coefficient $341 \mu\text{m}^2\text{s}^{-1}$ (literature value) [329], for which a diffusion time of $95 \pm 10 \mu\text{s}$ was found. The overall diffusion time for the six measurements is determined at $325 \pm 40 \mu\text{s}$. Moreover, none of the correlation curves show any indication of the presence of protein aggregates.

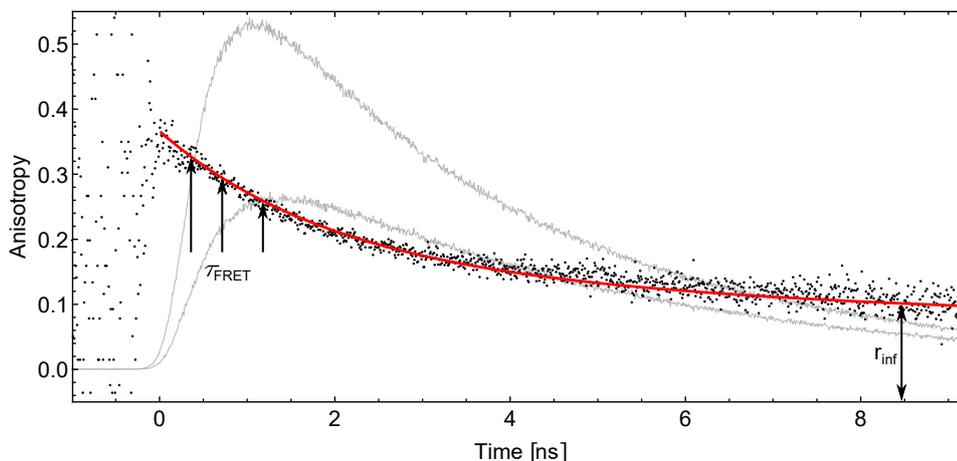


Figure 6.6.4: Time-resolved anisotropy decay. Exemplary time-resolved-anisotropy decay (MalE 186C, black dots) with double-exponential fit (red line) and polarization decays (light gray lines, scaled). Different lifetimes of the FRET mutants are marked (arrow) as well as infinite anisotropy r_{inf} [308].

Free dye	Steady-state anisotropy (532~nm)	Residual anisotropy (532~nm)	Steady-state anisotropy (635~nm)	Residual anisotropy (635~nm)
Alexa 546	0.035±0.003	0.01±0.02	-	
Alexa647	-		0.120±0.007	0.02±0.02
DNA-Standards				
8 base-pairs	0.114±0.003		0.184±0.012	
33 base-pairs	0.134±0.002		0.159±0.011	
Donor-only strand	0.134±0.003		-	
Acceptor-only strand	-		0.172±0.010	
Protein single mutants				
K29C, apo	0.285±0.017	0.21±0.03	0.198±0.015	0.15±0.03
K29C, holo	0.280±0.017	0.21±0.03	0.199±0.018	0.14±0.04
D87C, apo	0.231±0.012	0.13±0.03	0.217±0.016	0.11±0.03
D87C, holo	0.225±0.005	0.13±0.03	0.229±0.017	0.13±0.04
A134C, apo	0.290±0.016	0.21±0.02	0.215±0.019	0.15±0.02
A134C, holo	0.281±0.007	0.21±0.02	0.216±0.003	0.15±0.03
A186C, apo	0.176±0.018	0.10±0.02	0.186±0.014	0.07±0.04
A186C, holo	0.161±0.010	0.09±0.02	0.186±0.016	0.07±0.04
S352C, apo	0.247±0.007	0.16±0.02	0.272±0.015	0.19±0.03
S352C holo	0.243±0.010	0.15±0.02	0.263±0.002	0.17±0.03

Table 6.6.1: Steady-state and residual time-resolved anisotropy. Steady state anisotropy and residual time-resolved anisotropy of Alexa546 and Alexa647 for free dyes and single protein mutants of FRET pairs with steady-state anisotropy of the same fluorophores labelled to double-stranded DNA for comparison.

Protein samples	Donor (Alexa546) – acceptor (Alexa647)	Acceptor (Alexa647) – donor (Alexa546)
K29C-S352C, apo	0.20	0.15
K29C-S352C, holo	0.19	0.14
D87C-A186C, apo	0.10	0.10
D87C-A186C, holo	0.10	0.11
A134C-A186C, apo	0.12	0.12
A134C-A186C, holo	0.12	0.12

Table 6.6.2: Residual anisotropy for protein samples with stochastic labelling. Combined residual anisotropy $r_{inf} = \sqrt{r_{inf}^D r_{inf}^A}$ for protein samples with Alexa546 at first position / Alexa647 at second position and vice versa [308].

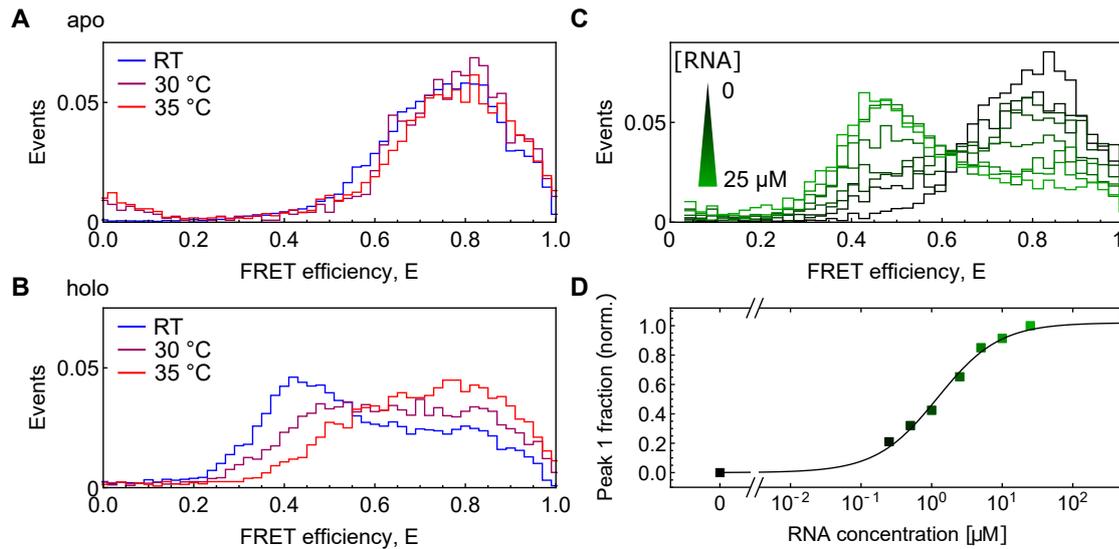


Figure 6.6.5: Temperature and RNA concentration dependency of U2AF65. **A/B** FRET histogram comparison of apo (A) and holo (B) measurements at room temperature (blue), 30°C (purple) and 35°C (red). **C** FRET histogram for U9 RNA titration measurements with of U2AF65 with increasing concentration (low to high RNA concentration shown as black to light green). Starting with apo shown in black). **D** Area under Peak 1 of the FRET histograms (0.1-0.6 FRET efficiency) from C were plotted against the increasing U9 RNA concentration to estimate the K_d . For area normalization, apo measurement was set to zero and holo measurement with 25 μM was set to 1. After fitting the data with the standard model for receptor-ligand kinetics as described in Supplementary Figure 6.6.2 was used and affinity to U9 RNA was estimated to be $\sim 1.2 \mu\text{M}$.

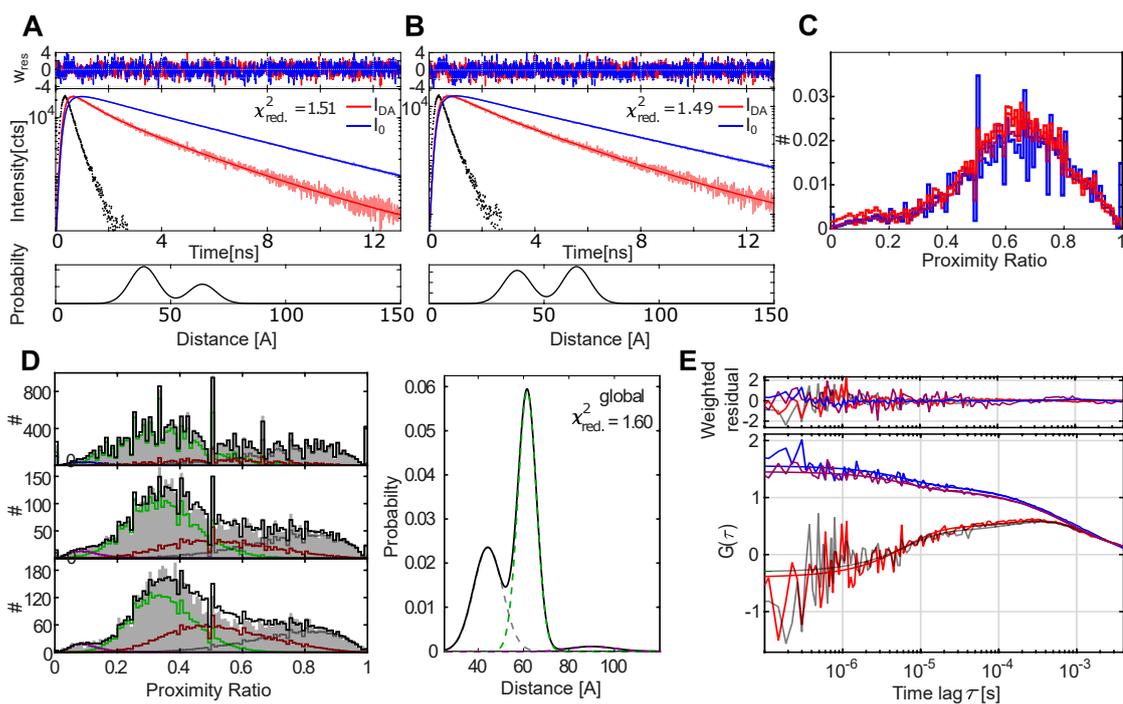


Figure 6.6: Quantification dynamics in U2AF65. **A/B** Two-Gaussian distance distribution fit for donor fluorescence lifetime for apo-state of U2AF65 shown in A, for holo-state shown in B. **C** Proximity ratio histograms binned with 0.5 ms, 1 ms and 1.5 ms for apo state. **D** A dynamic PDA fit with a two-state kinetic model for holo-state. **E** filtered FCS correlation curves of apo-state. Subspecies 1 and 2 are the burst showing highest and the lowest FRET values. Blue and red curves are depicting the cross-correlation function of subspecies 1 and 2, while green and magenta are autocorrelation functions.

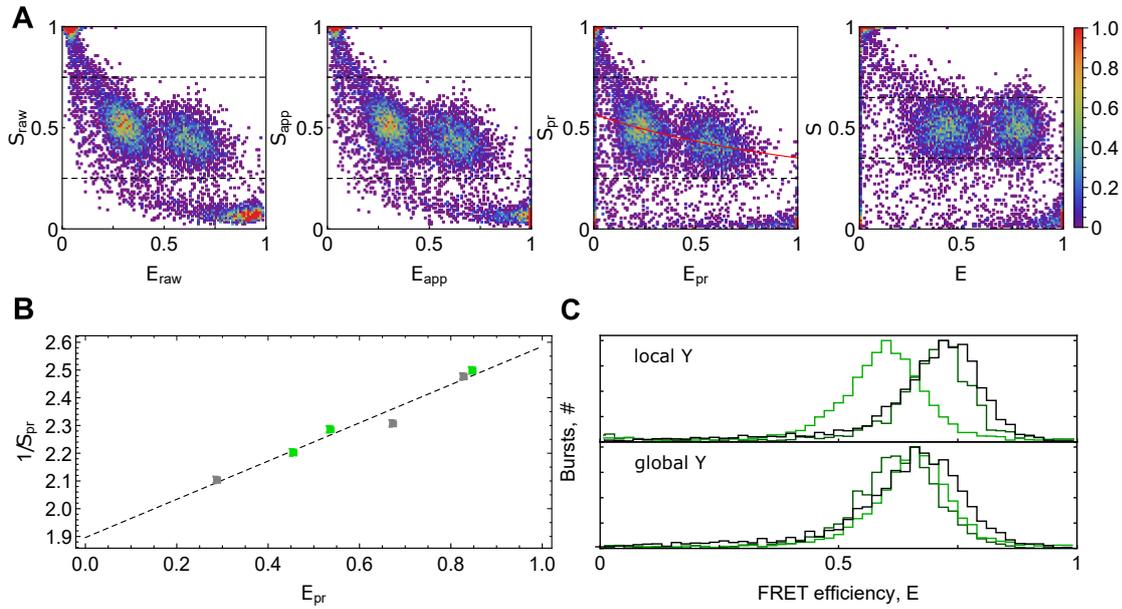


Figure 6.6.7: Correction procedure with global γ . **A** ES-diagrams of correction steps; from left to right: raw data, background corrected apparent FRET efficiency, crosstalk corrected proximity FRET efficiency E_{pr} with fitted γ curve (red), and γ corrected FRET efficiency with Alexa546-Alexa647 at two selected and merged data sets (K29C-S352C apo and D87C-A186C apo). **B** Proximity efficiency E_{pr} and stoichiometry S_{pr} of all MaleE mutants in apo (gray dots) and holo state (green dots) with linear fit (dashed line) for global γ correction. **C** Comparison of FRET efficiency histograms of MaleE mutant D87C-A186C with 1 mM maltose shows an extreme example of deviation between local (top) and global γ (bottom) calculation for some selected data sets. The data are based on three individual measurements of different samples. Local γ is calculated only based on the proximity efficiency E_{pr} and stoichiometry S_{pr} of the apo and holo state of this mutant, which leads to a strong deviation of the histograms, whereas the global γ calculation considers all mutants. The global fit gives a more robust γ value, such that the histograms match each other.

Lab	E_{raw}	α	δ	γ	β	E
1	0.56	0.05	0.07	0.51	1.70	0.66
2	0.42	0.04	0.06	0.31	2.65	0.66
3	0.54	0.04	0.14	0.34	1.56	0.72
4	0.43	0.03	0.08	0.23	2.38	0.72
5	0.57	0.08	0.12	0.54	1.59	0.66
6	0.59	0.07	0.12	0.63	0.64	0.70
7	0.47	0.04	0.11	0.48	1.53	0.63
8	0.57	0.06	0.10	0.47	1.84	0.69
9	0.46	0.03	0.11	0.31	1.88	0.72
10	0.54	0.05	0.32	0.34	0.5	0.70
11	0.43	0.03	0.07	0.32	2.37	0.64
12	0.47	0.04	0.07	0.25	1.99	0.74
13	0.58	0.06	0.32	0.34	0.60	0.76
14	0.41	0.04	0.09	0.23	2.00	0.77
15	0.57	0.05	0.05	0.33	1.42	0.79
16	0.60	0.05	0.14	0.46	1.31	0.72
17 *	0.18	0.01	0.10	0.09	4.86	0.59
18 **	0.59	0.18				

* The measurements were performed in a regime with $\gamma < 0.1$, where the error in gamma is significantly increased. Data are not considered for mean and standard deviation of measurements.

** Due to measurement problems, the data could not be corrected for direct excitation (δ) and gamma. Data are not considered in the evaluation of mean and standard deviation.

Table 6.6.3: FRET efficiency correction factors for Alexa546-Alexa647. Overview of all correction factors for protein 1 (MalE) and exemplarily change of FRET efficiency for D87C-a186C, holo.

Lab	E_{raw}	α	δ	γ
1	0.04	0.05	0.80	0.64
2	0.02	0.06	0.59	0.78
3	0.03	0.09	0.83	1.05
4 *				
5	0.04	0.06	1.24	0.62
6				
7	0.03	0.05	0.73	0.91
8 *				
9 *				
10	0.06	0.23	1.1	-
11	0.05	0.02	0.64	-
12 *				
13 *				
14	0.05	0.09	0.64	
15 *				
16 *				
17 *				
18 *				

* The groups did not measure the U2AF65 protein.

Table 6.6.4: FRET efficiency correction factors for Atto532-Atto643. Overview of all correction factors for protein 2 (U2AF).

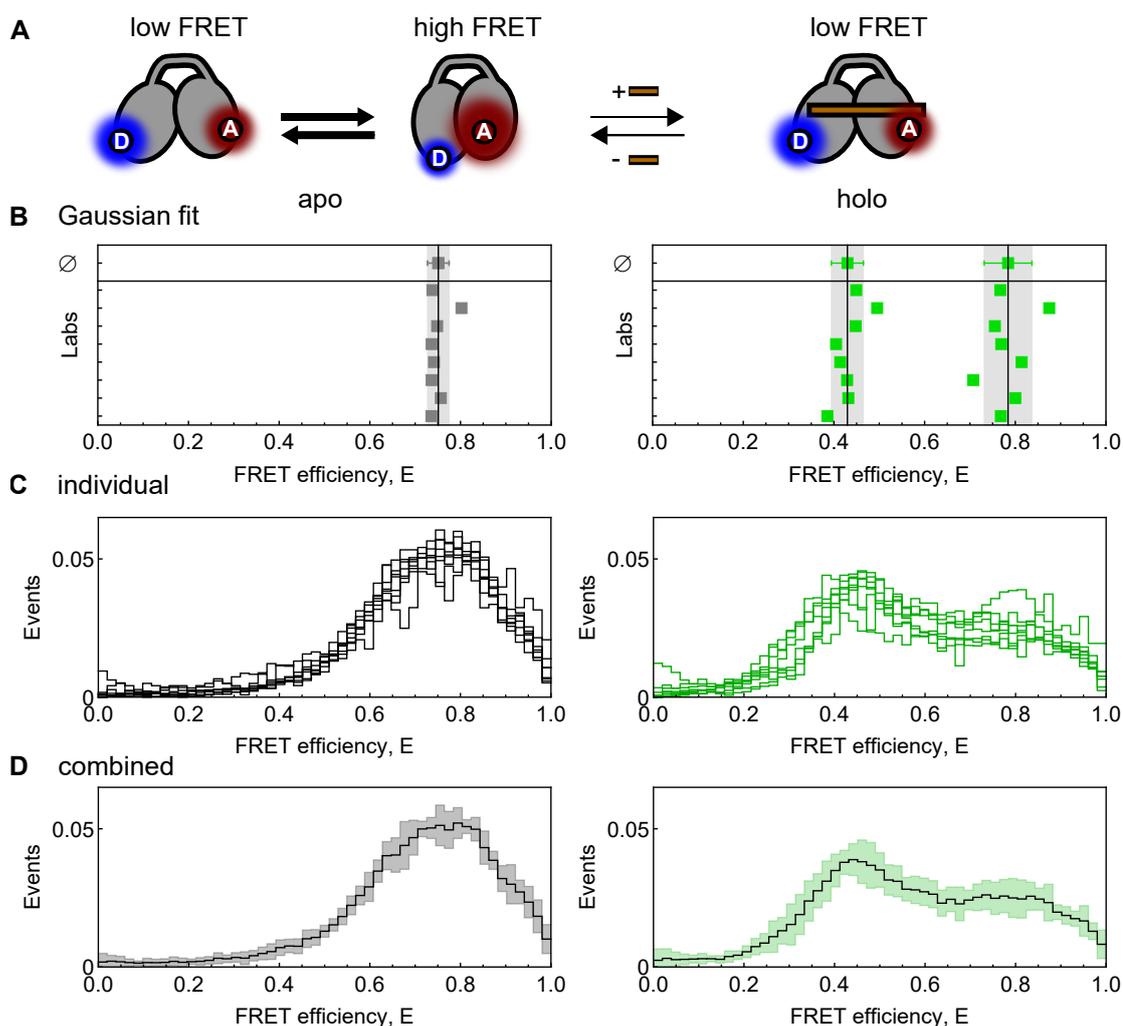


Figure 6.6.8: Determination of precision and FRET histogram comparison for U2AF65 after analyzing the collected data from 8 different laboratories. **A** Cartoon showing the U2AF65 conformational dynamics. In the absence of its RNA ligand, it fluctuates between open and closed state and becomes open after binding to the ligand. **B** Extracted mean values for the FRET histograms after analyzing and using a correct γ -factor are plotted with a mean value (upper data point) from all data sets with a corresponding standard deviation is shown in grey for both apo (left) and holo (right) states for U2AF65. **C/D** Individual FRET histograms for apo (left) and holo (right) conditions are shown in **C** while their mean (solid line) with a standard deviation (pale area) is shown in **D**.

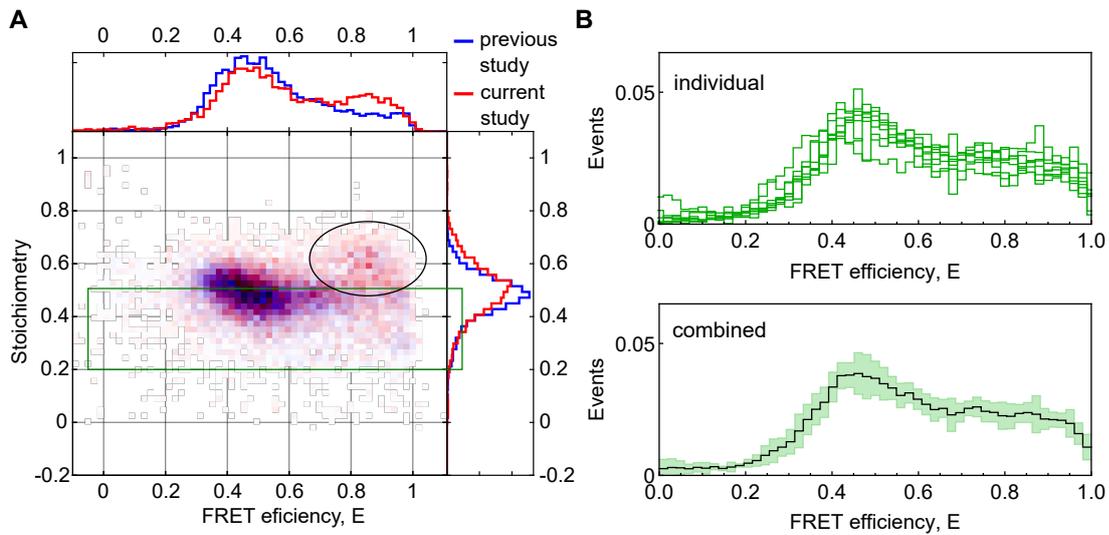


Figure 6.6.9: Identification of impurities present in the sample and its effect on the overall data consistency for holo state of U2AF65. **A** A 2D-Histogram plot of stoichiometry versus FRET efficiency for holo state was compared with previous preparation (blue) and the preparation used in the current study (red). Impurities are evident in sample preparation used in the current study with a stoichiometry of ~ 0.6 with a FRET efficiency of ~ 0.8 (highlighted with a circle). Selected bursts used for further analysis with a stoichiometry between 0.2-0.5 is highlighted **B** Compared FRET efficiency histogram, individual (top) and combined with a mean (solid line) and a standard deviation (pale) after removing the impurities by using only burst with a stoichiometry between 0.2-0.5 to build a histogram measured by different labs.

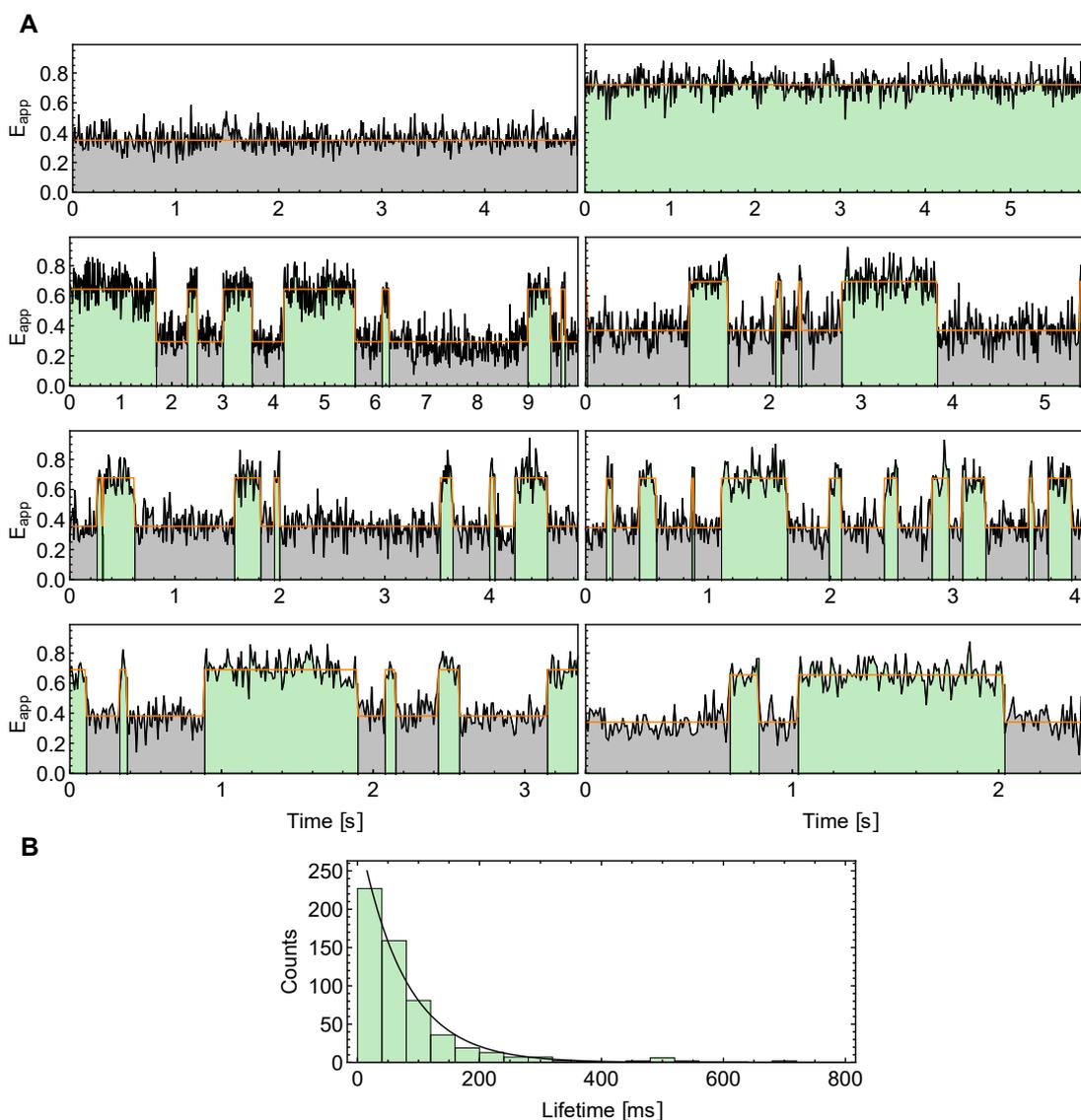


Figure 6.6.10: Ligand-induced slow conformational dynamics of MalE switching between apo- and holo states. **A** Time traces of immobilized MalE 36C/352C labelled with Alexa555-Alexa647. The samples were measured in a scanning confocal microscope as described in [107]. FRET states and lifetimes are extracted from a fitted two state Hidden-Markov-Model as described in [330]. The traces show ligand-induced interconversion of states on the >100 millisecond time-scale, but shot-noise limited fluctuations on the ms-timescale. **B** Dwell time histogram of holo state with exponential fit (solid line) shows a mean lifetime of $\tau_{dwell}=75$ ms.

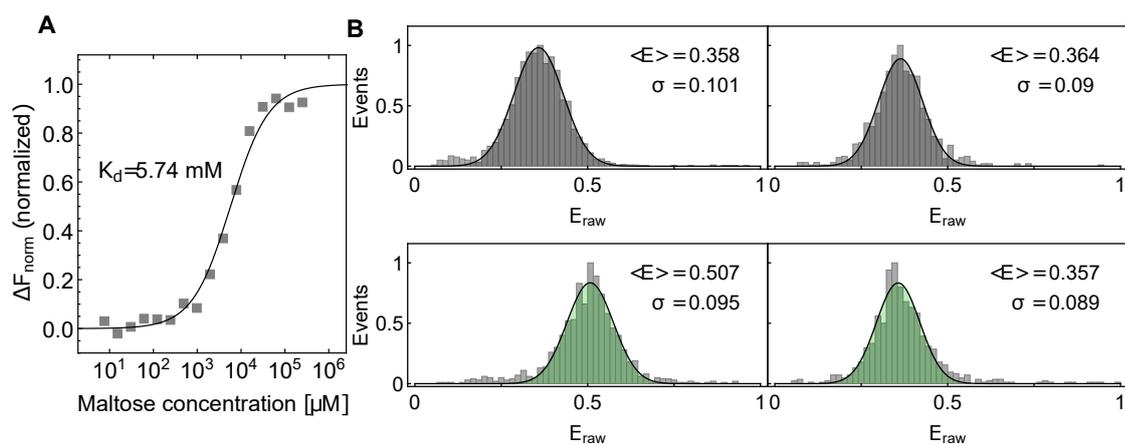


Figure 6.6.11: Unaffected fluorophore properties due to millimolar maltose. **A** Binding affinities are measured with microscale thermophoresis of MalE mutant K29C-S352C-D65A as described in Figure S3. **B** Raw FRET efficiency E_{raw} histogram for MalE mutant K29C-S352C labelled with Alexa546-Alexa647 without maltose (top, left), and with added 1 mM maltose (bottom, left) compared to same measurement with MalE mutant K29C-S352C-D65A (right) not showing any effect on the FRET efficiency.

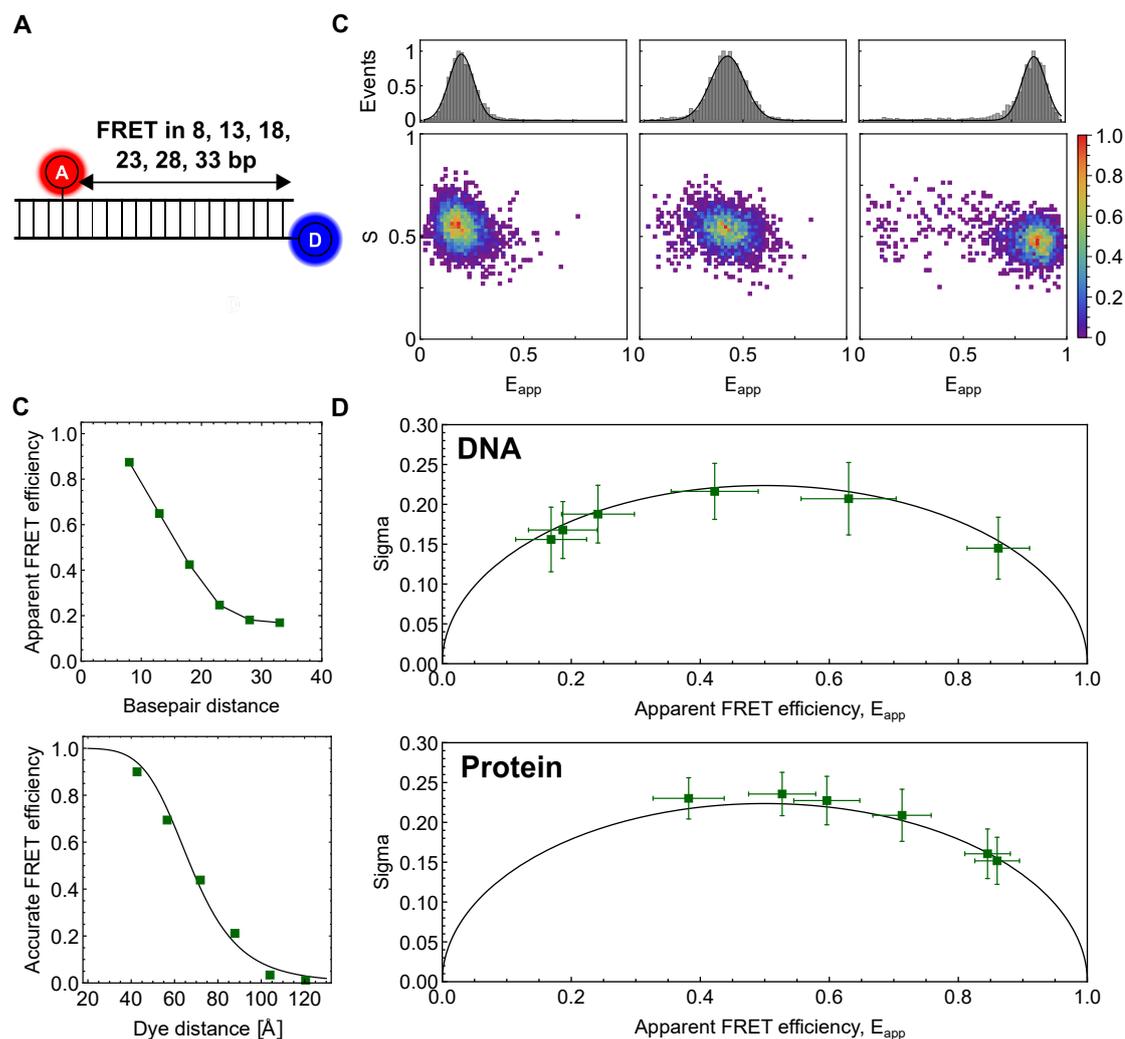


Figure 6.6.12: Shot-noise-limited measurements on static DNA and quasi-static protein. **A** Scheme of a dsDNA template containing a donor positioned at the 5'-end and an acceptor at 8, 13, 18, 23, 28 and 33 basepair distance from donor fluorophore. **B** Apparent FRET efficiency E_{app} depending on basepair distance (top) and corrected accurate FRET efficiency depending on modeled dye distance⁷ (bottom) for Alexa546-Alexa647. **C** Exemplary E-histograms (top) and ES-diagrams (bottom) for Alexa546-Alexa647 with 28 (left), 18 (center), and 8 (right) basepair distance. **D** Burst variance analysis for static DNA structure with Alexa546-Alexa647 (top) and all MalE mutants of this study in apo and holo states with Alexa546-Alexa647 shows shot noise limited signal⁸. The burst variance analysis is based on a photon binning of 5 photons. Results for separate measurements are combined into one graph.

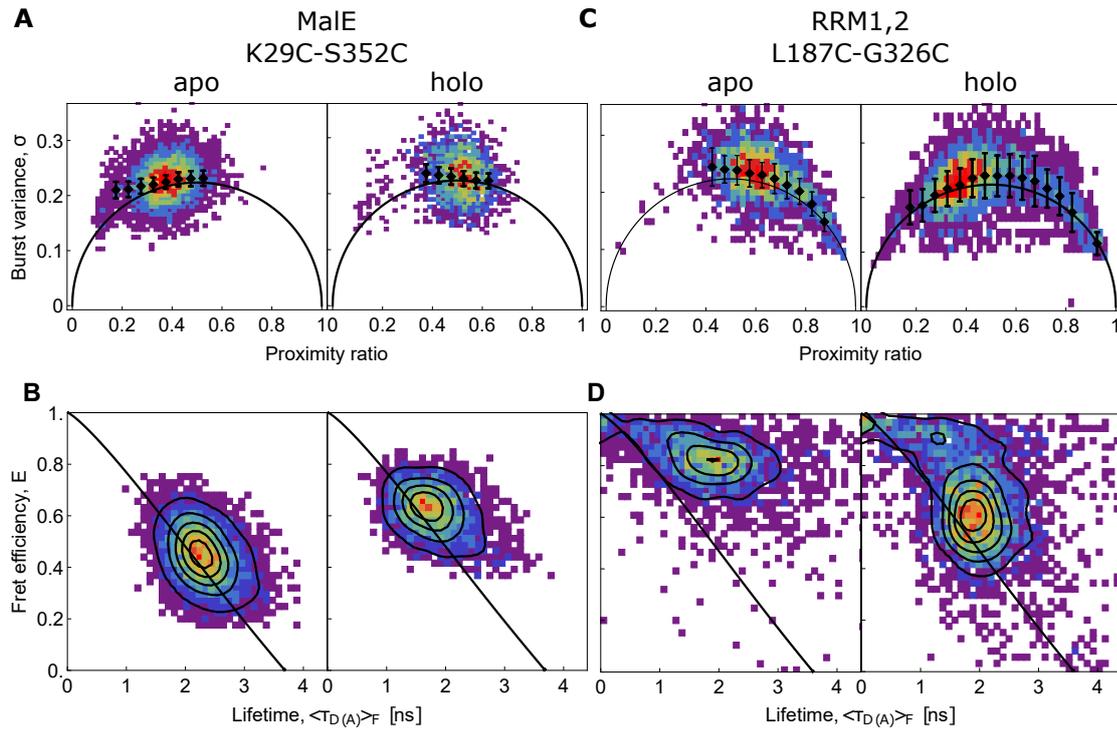


Figure 6.6.13: Characterization of conformational dynamics in MalE and U2AF65 on the sub-millisecond timescale. **A** Burst variance analysis of MalE mutant K29C-S352C labelled with Alexa546-Alexa647 without maltose (left) and with 1mM maltose (right). The burst variance analysis is based on a photon binning of 5 photons for bursts with stoichiometry between 0.4 and 0.7. **B** FRET efficiency E versus fluorescence-weighted average donor lifetime $\langle \tau_{D(A)} \rangle_F$ of same measurement as in A. The donor only population was excluded for clarity of the plot. The MFD plot shows that under both ligand concentrations the FRET species are slightly shifted from the static FRET line (black). **C** Burst variance analysis similar to A with sample U2AF65 mutant L187C-G326C labelled with Alexa546-Alexa647 without RNA ligand (left) and with 5 μ M of U9 RNA (right). **D** FRET efficiency E versus fluorescence-weighted average donor lifetime $\langle \tau_{D(A)} \rangle_F$ similar to B with same measurement as in C. The BVA and MFD plots in C and D show clear deviation from the static FRET line (black) as indicated with a dynamic FRET line (dashed, black).

Lab	Method	Apo	Holo
2	Dynamic PDA (R1 38 Å, R2 56 Å)	-	$k_{12} = 0.63 \text{ ms}^{-1}$ $k_{21} = 0.83 \text{ ms}^{-1}$ ($\tau_R = 680 \mu\text{s}$)
7	Dynamic PDA (R1 38 Å, R2 56 Å)	-	$k_{12} = 0.47 \text{ ms}^{-1}$ $k_{21} = 0.15 \text{ ms}^{-1}$ ($\tau_R = 1.6 \text{ ms}$)
10	Filtered-FCS	$\tau_{R1} = 320 \mu\text{s}$ $\tau_{R2} = 6 \mu\text{s}$	$\tau_{R1} = 320 \mu\text{s}$ $\tau_{R2} = 6 \mu\text{s}$
	FRET-FCS	$\tau_{R1} = 32 \mu\text{s}$ $\tau_{R2} = 2 \mu\text{s}$	$\tau_{R1} = 321 \mu\text{s}$ $\tau_{R2} = 10 \mu\text{s}$
	Dynamic PDA (R1 37 Å, R2 60 Å)	$k_{12} = 2.3 \text{ ms}^{-1}$ $k_{21} = 0.39 \text{ ms}^{-1}$ ($\tau_R = 370 \mu\text{s}$)	$k_{12} = 2.3 \text{ ms}^{-1}$ $k_{21} = 0.39 \text{ ms}^{-1}$ ($\tau_R = 370 \mu\text{s}$)
11	Filtered-FCS	$\tau_{R1} = 90 - 200 \mu\text{s}$ $\tau_{R2} = 3 - 10 \mu\text{s}$	-
	Dynamic PDA (R1 38 Å, R2 59 Å)	$k_{12} = 0.43 \text{ ms}^{-1}$ $k_{21} = 0.07 \text{ ms}^{-1}$ ($\tau_R = 2 \text{ ms}$)	$k_{12} = 0.52 \text{ ms}^{-1}$ $k_{21} = 0.14 \text{ ms}^{-1}$ ($\tau_R = 1.42 \text{ ms}$)
14	Filtered-FCS	$\tau_{R1} = 370 \mu\text{s}$ $\tau_{R2} = 12 \mu\text{s}$	$\tau_{R1} = 617 \mu\text{s}$ $\tau_{R2} = 15 \mu\text{s}$
	Dynamics PDA (R1 37 Å, R2 60 Å)	$k_{12} = 2.3 \text{ ms}^{-1}$ $k_{21} = 4.6 \text{ ms}^{-1}$ ($\tau_R = 144 \mu\text{s}$)	$k_{12} = 0.47 \text{ ms}^{-1}$ $k_{21} = 0.15 \text{ ms}^{-1}$ ($\tau_R = 1.6 \text{ ms}$)

Table 6.6.5: Results from dynamic investigation on U2AF65. Time scales and rates from groups derived by dynamic PDA, filtered-FCS, and FRET-FCS.

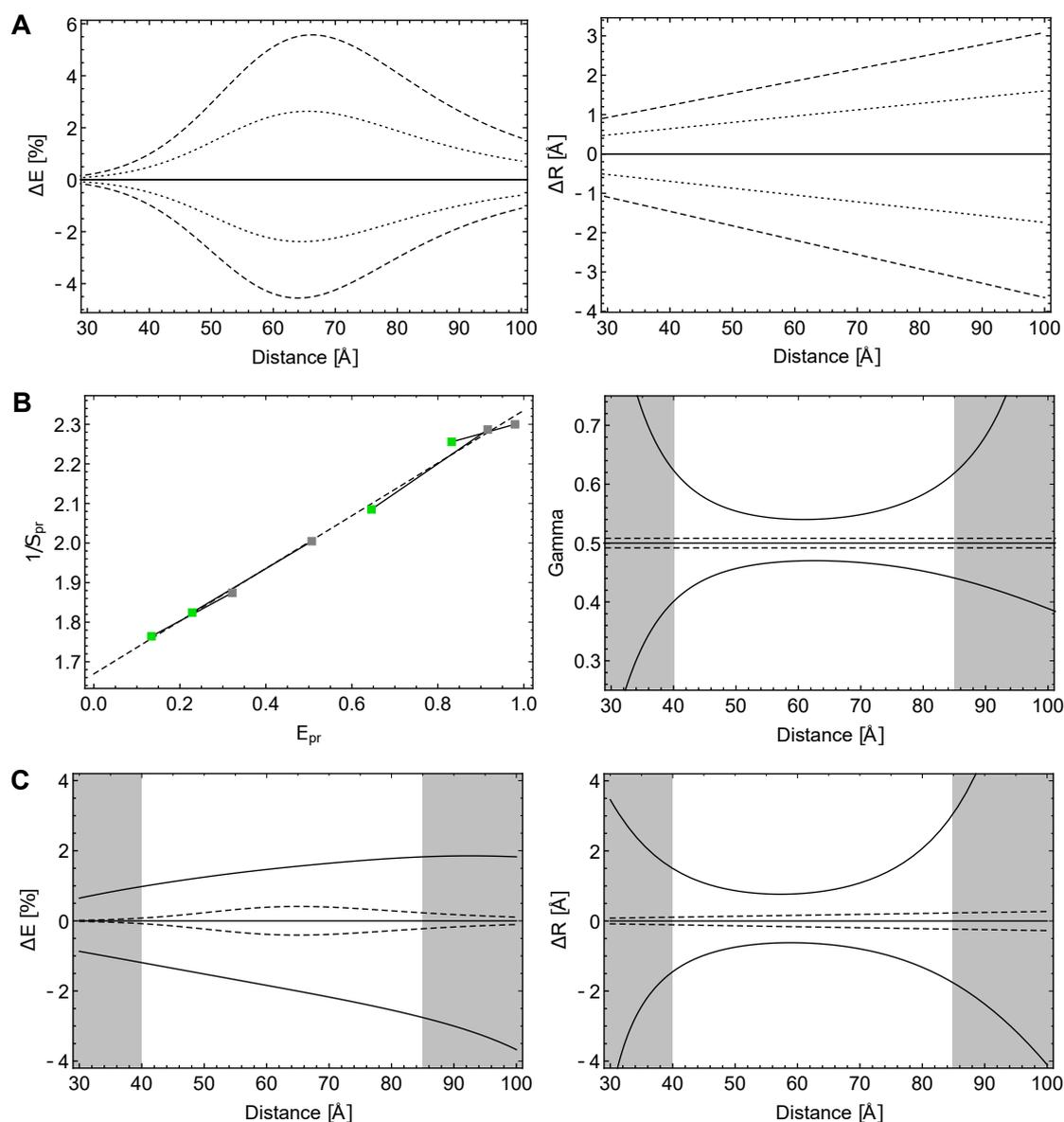


Figure 6.6.14: Measurement and correction errors (simulated data). **A** Error simulation in FRET efficiency (left) and distance (right) for deviations of -0.1, -0.05, 0.05, and 0.1 from the theoretical γ -value of 0.5. **B** Illustration of γ -determination with global fit (dashed line) and local fits (solid lines) for exemplary data. The theoretical sample data points range from 35 to 75 Å with a conformational change of 8 Å (left). Comparison of uncertainty in γ for global γ (dashed black) and local γ (solid black) based on error contributions of uncertainties in proximity FRET efficiency E_{pr} and proximity stoichiometry S_{pr} determination of $\pm 0.25\%$ for a theoretical γ -value of 0.5. The gray areas indicate the regions < 40 Å ($< 0.6 R_0$) and > 85 Å ($> 1.3 R_0$) with a γ uncertainty resulting in a significant distance uncertainty ($\gtrsim 2$ Å). **C** The uncertainty of γ in C results in uncertainty of FRET efficiency (left) and a following distance uncertainty (right) for the different errors (same line style as in B). For details on the error simulations, see Supplementary Note 6.7.3.

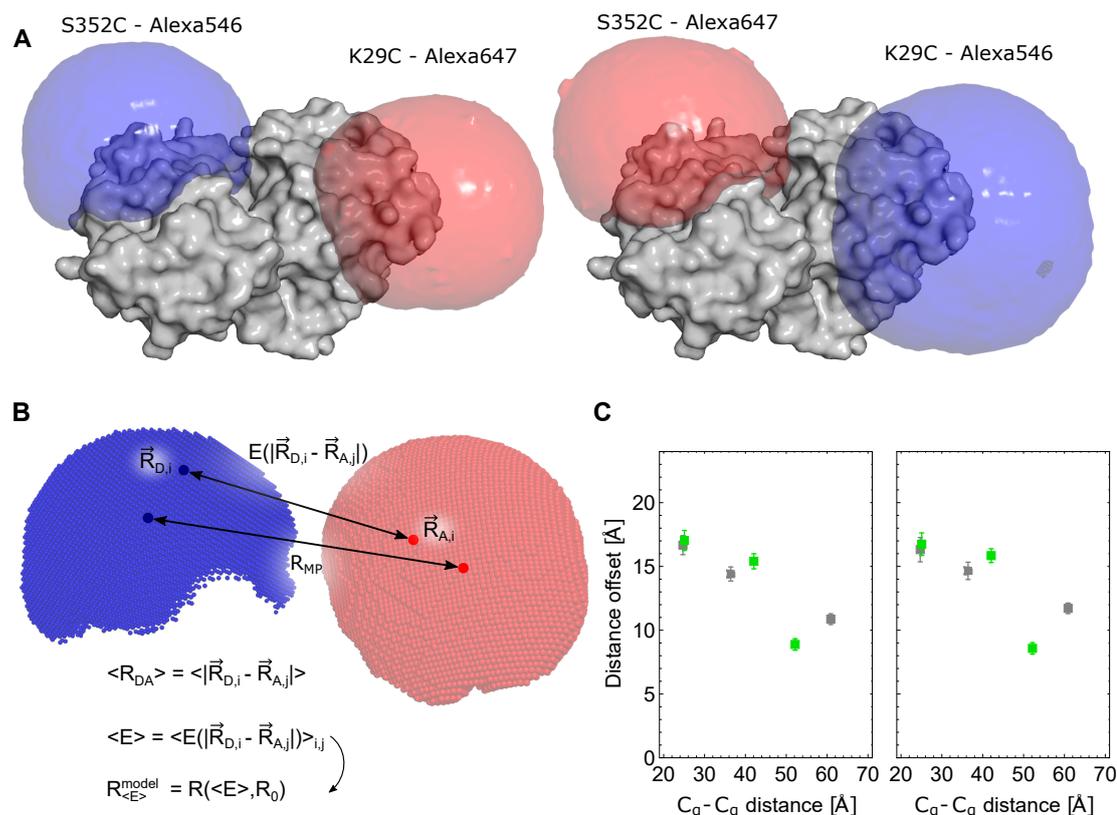


Figure 6.6.15: Accessible volume simulations. **A** Simulation of accessible volume for Alexa546 and Alexa647 at labelling positions 29 and 532 with dye combinations donor at position 352 (left), acceptor at position 29 and reverse (right). The C_α - C_α -distance is 60.9 Å, FRET efficiency averaged model distances are 72.6 ± 0.4 and 71.8 ± 0.4 Å for the two labelling combinations. **B** Comparison of different models for the dye distance, where R_{MP} is the distance between the mean positions of both accessible volumes, $\langle R_{DA} \rangle$ is an averaged distance of all possible combinations between donor and acceptor accessible volume, and $R_{\langle E \rangle}^{\text{model}}$ the FRET efficiency averaged model distance, which considers the non-linearity of the FRET efficiency E by averaging the efficiency over all possible combinations and calculating back the distance. **C** Comparison of FRET efficiency averaged model distance offset to the C_α - C_α -distance for the fluorophore pairs in apo (gray) and holo (green) state, where donor is at the lower residue number and acceptor at the higher residue number of the mutant (left) and vice versa (right). Differences between both dye combinations are less than 10 Å for all samples.

6.7 Supplementary Notes

6.7.1 Supplementary Note 1: Primer design and sequences

Primers for the MalE mutants were chosen as follows:

1	GTCGGTAAG
1 K I E E G K L V I W I N G D K G Y N G L A E V G K	
1 AAAATCGAAGAAGGTAAACTGGTAATCTGGATTAACGGCGATAAAGGCTATAACGGTCTCGCTGAAGTCGGTAAG	
76 AAATTCGAGWRMGATACCGG 3' Lys29Cys	
26 K F E K D T G I K V T V E H P D K L E E K F P Q V	
76 AAATTCGAGAAAGATACCGGAATTAAGTCACCGTTGAGCATCCGGATAAACTGGAAGAGAAAATCCACAGGTT	
51 A A T G D G P D I I F W A H D R F G G Y A Q S G L	
151 GCGGCAACTGGCGATGGCCCTGACATTATCTTCTGGGCACACGACCGCTTTGGTGGCTACGCTCAATCTGGCCTG	
226 CCGGACAAAGCGTTCCAGKRCAAGCTGTATCCG 3' Asp87Cys	
76 L A E I T P D K A F Q D K L Y P F T W D A V R Y N	
226 TTGGCTGAAATCACCCCGGACAAAGCGTTCCAGGACAAGCTGTATCCGTTTACCTGGGATGCCGTACGTTACAAC	
101 G K L I A Y P I A V E A L S L I Y N K D L L P N P	
301 GGCAAGCTGATTGCTTACCCGATCGCTGTTGAAGCGTTATCGCTGATTTATAACAAAGATCTGCTGCCGAACCCG	
376 GAAGAGATCCCGKSSCTGGATAAAGAAC 3' Ala134Cys	
126 P K T W E E I P A L D K E L K A K G K S A L M F N	
376 CCAAAAACCTGGGAAGAGATCCCGGCGTGGATAAAGAAGCTGAAAGCGAAAGGTAAGAGCGCGCTGATGTTCAAC	
151 L Q E P Y F T W P L I A A D G G Y A F K Y E N G K	
451 CTGCAAGAACCGTACTTCACCTGGCCGCTGATTGCTGCTGACGGGGTTATGCGTTCAAGTATGAAAACGGCAAG	
526 GTGGATAACKSYGGCGGAAAGCG 3' Ala186Cys	
176 Y D I K D V G V D N A G A K A G L T F L V D L I K	
526 TACGACATTAAGACGTGGGCGTGGATAACGCTGGCGGAAAGCGGGTCTGACCTTCTGGTTGACCTGATTAATA	
201 N K H M N A D T D Y S I A E A A F N K G E T A M T	
601 AACAAACACATGAATGCAGACACCGATTACTCCATCGCAGAAGCTGCCTTTAATAAAGGCGAAACAGCGATGACC	
226 I N G P W A W S N I D T S K V N Y G V T V L P T F	
676 ATCAACGGCCCGTGGGCATGGTCCAACATCGACACCAGCAAAGTGAATTATGGTGAACGGTACTGCCGACCTTC	
251 K G Q P S K P F V G V L S A G I N A A S P N K E L	
751 AAGGTCACCAATCCAAACCGTTCTGGCGTGTGAGCGCAGGTATTAACGCCCGCAGTCCGAACAAAGAGCTG	

```
276 A K E F L E N Y L L T D E G L E A V N K D K P L G
826 GCGAAAGAGTTCTCGAAAACCTATCTGCTGACTGATGAAGGTCTGGAAGCGGTTAATAAAGACAAACCGCTGGGT
```

```
301 A V A L K S Y E E E L A K D P R I A A T M E N A Q
901 GCCGTAGCGCTGAAGTCTTACGAGGAAGAGTTGGCGAAAGATCCACGTATTGCCGCCACCATGGAAAACGCCAG
```

```
976 GATCAACGCC
326 K G E I M P N I P Q M S A F W Y A V R T A V I N A
976 AAAGGTGAAATCATGCCGAACATCCCGCAGATGTCCGCTTTCTGGTATGCCGTGCGTACTGCGGTGATCAACGCC
```

```
1051 GCCWCGGTCGTCAG 3' Ser352Cys
```

```
351 A S G R Q T V D E A L K D A Q T
1051 GCCAGCGGTCGTCAGACTGTCGATGAAGCCCTGAAAGACGCGCAGACT
```

6.7.2 Supplementary Note 2: Dynamics of U2AF65

In a final step in characterizing the conformational dynamics, the Lamb group quantified the conformational dynamics of U2AF65 (Supplementary Figure 6.6.6). Before determining the kinetic rates for the actual conformational transitions, we defined the conformational states for U2AF65 transitions by fitting the donor fluorescence lifetime with global Gaussian distribution fitting for both apo and RNA-bound conditions for inter-dye distances. Donor fluorescence lifetime is an absolute readout which does not rely on dynamic timescales that are significantly slower than the fluorescence lifetime and can detect a mixture of different conformational states with defined lifetimes within a single molecule burst [64]. Clearly, a single-Gaussian model was not sufficient to explain the donor lifetime decay and inter-dye distance. After fitting the donor signal with bi-exponential decay, we could get two discrete donor lifetimes of ~ 0.5 ns and ~ 2.8 ns for apo and ~ 0.5 ns and ~ 2.2 ns for holo states for U2AF65 from two-Gaussian distance distribution model. These two lifetimes obtained could potentially correspond to ligand-free closed and RNA-bound open state with ~ 38 Å and ~ 59 -65 Å inter-dye distances respectively (Figure 6.1.2/6.2.2, Supplementary Figure 6.6.6A).

Qualitative BVA and E-tau plots gave a hint for sub-millisecond dynamics (Figure 6.2.3) which led us to use Photon distribution analysis (PDA) method. PDA allows to delineate the conformational heterogeneity over sub-millisecond transitions from the width of the FRET histograms beyond shot-noise with a use of raw photon counts [62, 331]. We binned raw photon data for both the apo and holo conditions with 0.5 ms, 1 ms, and 1.5 ms bin size for all the bursts. The resulting histogram hinted us that the apo-state has microsecond dynamics which cannot be visible in differently sized binned data while holo-state has relatively slow dynamics over milliseconds with a visible difference in the histograms (Supplementary Figure 6.6.6B/C). The apo state can be assumed as an ensemble of fast mixing of conformations between the closed and open states. Hence, we applied again applied FCS which offers the access to the fast fluorescence fluctuations to extract these fast dynamics. We used filtered-FCS (fFCS) as it increases the contrast of the FCS signal by weighing the correlation signals depending on the

difference in the lifetime, color and anisotropy of fluorescence fluctuations between the bursts showing the lowest and highest FRET values [332]. Relaxation timescales of ~ 90 - $200 \mu\text{s}$ was obtained with fFCS for U2AF65 apo-state with different repeats of the measurement (Supplementary Figure 6.6.6D). Comparatively slow dynamics of holo-state was quantified with dynamic PDA allowing to describe the FRET histograms analytically with a kinetic rate for distance distributions [64]. Two-state kinetic model was applied to perform the global analysis of binned histograms for different times (0.5 ms, 1 ms, 1.5 ms) for holo-state dynamics to increase the robustness of the fit. Highly dynamic RNA-free apo state converts slowly into RNA-bound holo state (closed to open, k_{12}) with a rate of 0.52 ms^{-1} and holo state converts into apo state (open to closed, k_{21}) with a rate of 0.14 ms^{-1} (Supplementary Figure 6.6.6B/C). This gives a relaxation time of 1.42 ms for these transitions.

This elaborate method and analysis was not available to most participants of this study. Yet, five labs could contribute to a dynamic quantification of U2AF65 (Supplementary Table 6.6.5). These all correctly reported a two-state behavior of the U2AF65 system. A comparison of kinetic rates based on further evaluation with (dPDA, fFCS) for U2AF65 from different labs with good consistent (Supplementary Table 6.6.5). Specifically, calculated the relaxation times for apo state with fFCS were consistent with an approximately two-fold variation across labs (~ 200 , ~ 320 and $\sim 370 \mu\text{s}$ from three different labs) considering different dye pairs and repeats of measurements. A dynamic PDA analysis for holo-state kinetic rate estimations were not fully consistent most likely due to the variation seen in the FRET histograms (maybe due to different measurement temperature conditions can affect the ligand binding than the inherent apo-state dynamics) which might have resulted in the difference in the quantification of the rate as the peak amplitudes in the histograms are very sensitive to fit the kinetic rate model. Three amongst five groups provided the kinetic rates for holo-state with only $\sim 20\%$ variation of the reported relaxation times (equals to $1/(k_{12} + k_{21})$) of 1.25 ms, 1.42 and 1.6 ms (see Supplementary Table 6.6.5).

6.7.3 Supplementary Note 3: Evaluation procedure for U2AF65

Analysis of the collected raw data of U2AF65 from different labs was performed with PAM software [103]. First, a burst search was performed using an all photon burst search with a threshold of 50-100 photons per sliding time window of $500 \mu\text{s}$ depending on the dataset. For one set of measurements, a lower threshold of 20 photons per $500 \mu\text{s}$ time window was necessary. To remove blinking and bleaching events an ALEX-2CDE filter with a lower limit of 5 and an upper limit of 25 was used depending on the data which may differ depending on the excitation intensities and sample concentrations used for the measurements [320]. After burst selection, background subtraction and correction for cross-talk and direct excitation were performed as discussed in the data analysis section. To determine the detection correction factor, we used the approach of Lee et al. [110] (i.e. fitting a line to $1/\text{SPR}$ vs EPR) as the sample is dynamic and a lifetime approach was not possible. The apo configuration shows a single, dynamically averaged static so that we had to combine data from both the holo and apo measurements. We verified that

there was no significant change in quantum yield of the donor and acceptor fluorophores in the absence and presence of ligand by measuring the fluorescence lifetime of the donor-only species and the acceptor lifetime with direct excitation. However, we did observe an additional subpopulation where the acceptor was slightly quenched, leading to a shift in the stoichiometry value (Supplementary Figure 6.6.9A). This population needs to be avoided when calculating the γ -value. The average values of SPR and EPR for the three populations (apo, holo – low FRET, holo – high FRET) were determined from the peak values of a 2D-Gaussian fit in ES-histograms for the respective populations. From these peak values, a straight line was fit to the three data points of $1/\text{SPR}$ vs EPR. When a single person evaluated the data, the consistence between laboratories was further improved (Supplementary Figure 6.6.9B). Part of the discrepancy came from the fact that all labs used a global γ approach but did not compensate for the presence of the second subpopulation with slight acceptor quenching.

6.7.4 Supplementary Note 4: FRET with different fluorophores

These artifacts are likely caused by photoisomerization processes in Alexa647 resulting in blinking and acceptor-intensity saturation for higher FRET efficiencies, i.e., FRET-dependent acceptor saturation [328, 333] and suggest that rhodamine dyes ATTO532/643 (which do not photoisomerize) can serve as references for such cases.

6.7.5 Supplementary Note 5: Error simulation

The γ correction factor requires at least two different measurements to be derived from a linear fit to the center values of a population ($1/S_{pr}$ versus proximity FRET efficiency E_{pr}) [110]. In principle, local variations of the dye environment might change the ratio of fluorescence quantum yields, and thus, the γ factor [2]. This fact supports a “local γ ” correction of each individual mutant using apo and holo state. However, the differences in E_{pr} for apo and holo is often rather small (especially for sample 3 in this study), which makes the linear fit very error-prone. In contrary, a “global γ ” correction based on all measurement points along the complete sensitive FRET efficiency can improve the fitting quality significantly. However, this approach can become extremely wrong, when the quantum efficiencies vary very much for different mutants due to different dye environments at the protein surface. For the selected mutants, the variation between different mutants is expected to be reasonably small (checked by quantum yield and lifetime measurements).

Theory. The error of the global γ compared to a local γ value due to bare measurement uncertainties was simulated based on a Monte Carlo approach. Distance dependent errors due to measurement uncertainties in proximity efficiency ΔE_{pr} and stoichiometry ΔS_{pr} are calculated using

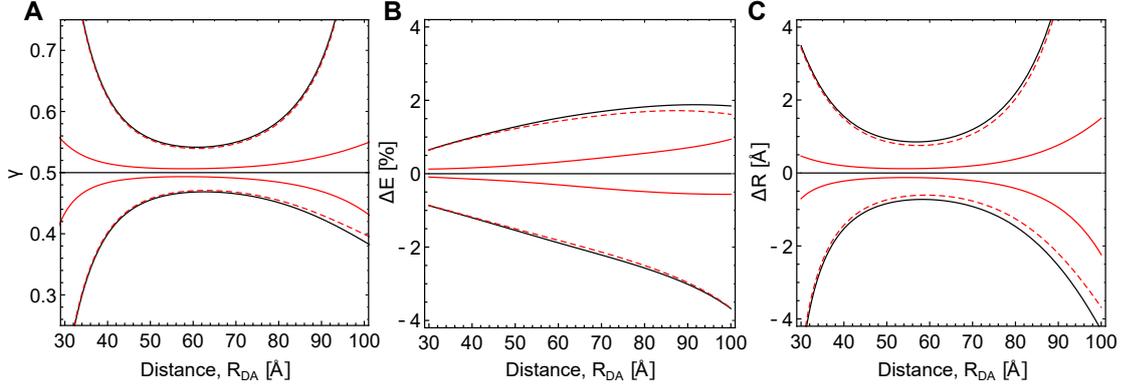


Figure 6.7.1: Error simulation of local γ determination. **A** Error contributions in γ from uncertainty in stoichiometry ΔS_{pr} (dashed red) and uncertainty in efficiency ΔE_{pr} (solid red) combined to the complete uncertainty (black). **B** Simulated error in in E (black) resulting from contributions of ΔS_{pr} (dashed red) and ΔE_{pr} (solid red). **C** Errors of B converted into absolute distance errors.

$$\begin{aligned}
 E(R_{DA}, R_0) &= \frac{1}{1 + \left(\frac{R_{DA}}{R_0}\right)^6} \\
 E_{pr}(E, \gamma) &= \frac{\gamma E}{1 + (\gamma - 1)E} \\
 S_{pr}(E_{pr}, \gamma) &= \frac{1}{2E_{pr} \frac{1-\gamma}{1+\gamma} + \frac{1+3\gamma}{1+\gamma}}
 \end{aligned} \quad (6.7.1)$$

Here, $S_{pr}(0.5, \gamma) = 0.5$ is assumed for the calculation. The error contributions in the simulation are separated into the two components for the local γ deviation

$$\begin{aligned}
 \Delta\gamma(E) &= \sqrt{\Delta\gamma^2(\Delta E_{pr}^1) + \Delta\gamma^2(\Delta E_{pr}^2)} \\
 \Delta\gamma(S) &= \sqrt{\Delta\gamma^2(\Delta S_{pr}^1) + \Delta\gamma^2(\Delta S_{pr}^2)}
 \end{aligned} \quad (6.7.2)$$

comparing the contribution of the error in S and E for the precision of γ , FRET efficiency E, and distance R_{DA} . In our measurement, the indices 1 and 2 stand for the two measurements of apo and holo state. For the error estimation, a Monte Carlo simulation with 10,000 repetitions was performed, where ΔE_{pr} and ΔS_{pr} are randomly chosen from a normal distribution (distribution width according to measurement error). The deviation $\Delta\gamma(\Delta E_{pr}^1, \dots, \Delta E_{pr}^N, \Delta S_{pr}^1, \dots, \Delta S_{pr}^N)$ from the theoretical value γ was calculated based on error propagation.

Model assumptions. In the error simulation of Figure 6.7.1, a comparison of local and global γ determination was based on the assumptions in Table 6.7.1.

Distance change between apo and holo (for all mutants):	8 Å
Measurement error of E_{pr} and S_{pr} (ΔE_{pr} and ΔS_{pr}):	± 0.0025
Förster radius:	65 Å
Theoretical γ :	0.5

Table 6.7.1: Model assumptions for error simulation.

The local γ error (and the resulting errors in E and R) is calculated for all average distances ($d_{\text{avg}} = (d_{\text{holo}} - d_{\text{apo}})/2$) from 30 to 100 Å for measurement errors in E_{pr} and S_{pr} individually. The global γ error is based on 4 mutants with d_{avg} ranging from 35 to 75 Å. The resulting error in γ is converted into the errors in E and R_{DA} along the complete range from 20 to 80 Å to compare it with the local errors.

The simulations (Figure 6.7.1A-C) show that the measurement uncertainty in stoichiometry ΔS_{pr} (dashed red) is the major contribution to the complete uncertainty (black), while the uncertainty in efficiency ΔE_{pr} (solid red) is negligible for distances larger 32.5 Å ($0.5 R_0$). The simulation results emphasize the importance of stable laser powers during the complete measurement time, since the stoichiometry is already significantly affected by a variation of 1% of the ratio of green and red laser power.

6.7.6 Supplementary Note 6: Förster radius determination

Fluorescence quantum yield of the donor dye Alexa 546 covalently bound to the protein was determined in reference to Rhodamin 6G.

Figure 6.7.2 shows the absorption spectra (A) and emission spectra at 510 nm excitation wavelength (B) are measured at multiple concentrations in the range of 0.5-1.0 μM for the sub-stoichiometrically labelled donor only samples and compared to Rhodamin 6G in water. A linear fit to the integrated fluorescence $\int_0^\infty F_D(\lambda) d\lambda$ dependent on the absorbance at 510 nm gives two slopes m_{A1546} and m_{R6G} (c). Fluorescence quantum yield of the donor Alexa546 is calculated from the slopes of the linear fit as

$$\Phi_{(F,D)} = \frac{m_{\text{A1546}}}{m_{\text{R6G}}} \Phi_{(F,R6G)}, \quad (6.7.3)$$

where $\Phi_{(F,R6G)} = 91 \pm 2\%$ is taken from literature [89]. The quantum efficiency of Alexa 546 bound to the protein was found to be $72 \pm 4\%$.

Overlap integral. The overlap integral J is retrieved from the emission spectrum $F_D(\lambda)$ of the donor only sample and a normalized absorption spectrum $\bar{\varepsilon}_A(\lambda)$ scaled to the literature extinction coefficient $\varepsilon_A(\lambda) = \varepsilon_{A,\text{max}} \bar{\varepsilon}_A(\lambda)$ according to

$$J = \frac{\int_0^\infty F_D(\lambda) \varepsilon_A(\lambda) \lambda^4 d\lambda}{\int_0^\infty F_D(\lambda) d\lambda}. \quad (6.7.4)$$

The overlap integral (gray area) is illustrated in Figure 6.7.3 resulting from the donor emission (green) and acceptor absorption (black) spectra.

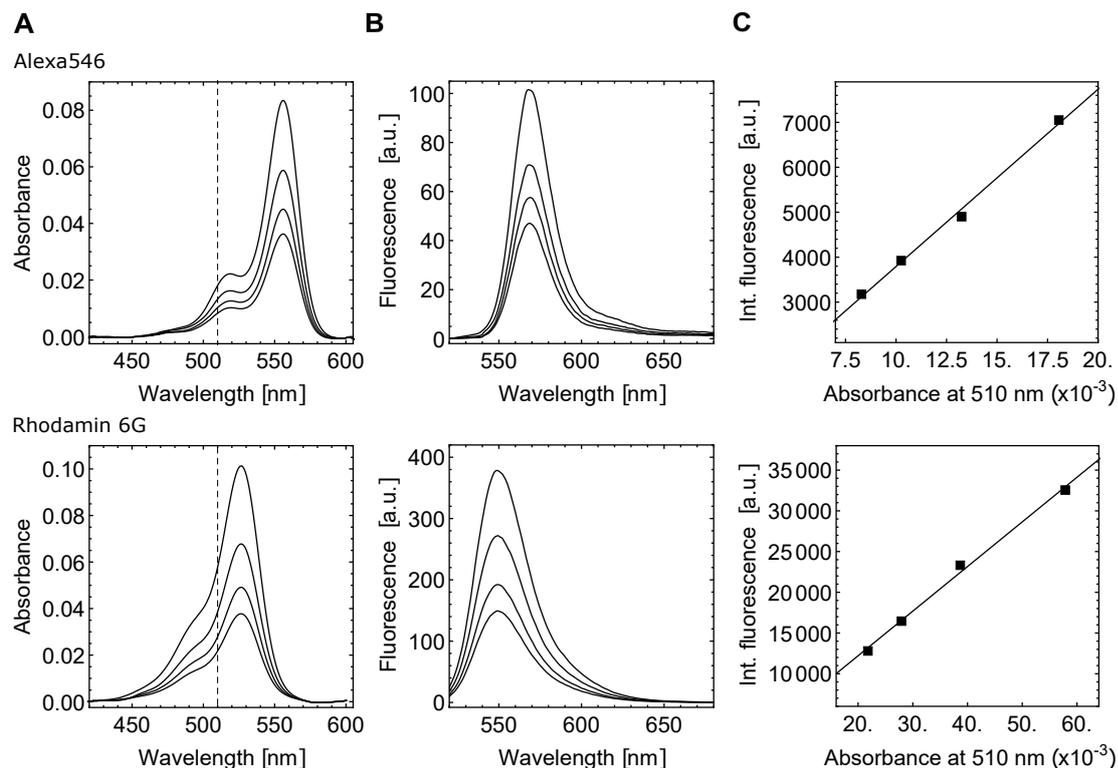


Figure 6.7.2: Fluorescence quantum yield determination exemplarily for K29C-S352C. (a) Absorption spectrum of Alexa546 (top) and Rhodamin 6G (bottom) at 0.5, 0.625, 0.75 and 1 μM concentration. (b) Emission spectrum of Alexa546 (top) and Rhodamin 6G (bottom) at 0.5, 0.625, 0.75 and 1 μM concentration excited at 510 nm. (c) Integrated fluorescence of Alexa546 (top) and Rhodamin 6G (bottom) plotted against absorbance at 510 nm at 0.5, 0.625, 0.75 and 1 μM concentration and linear fit to the data points (line).

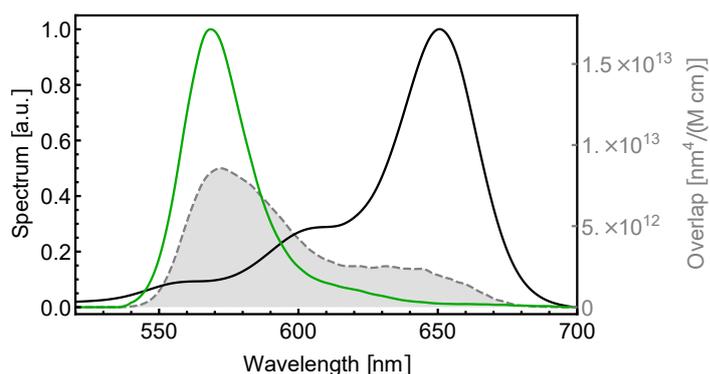


Figure 6.7.3: Alexa546-Alexa647 overlap integral. Emission spectrum of Alexa546 (green) and absorption spectrum (black) with overlap (gray) shown for mutant K29C-S352C.

Calculation. The Förster radius was calculated to be $R_0 = 65 \pm 3 \text{ \AA}$ with a fluorescence quantum yield $\Phi_{(F,D)} = 72 \pm 4\%$, an overlap integral $J = 7.0 \pm 0.1 \times 10^{15} \text{ nm}^4 / (\text{M} \cdot \text{cm})$ considering uncertainties in κ^2 and n of 10% and 5%, respectively.

7 | Labelizer: systematic selection of labeling sites for fluorescent dyes on proteins

In biophysical applications and imaging, the covalent attachment of fluorophores to biomolecules is often based on introduction of cysteines or unnatural amino-acids. In general, a manual inspection of the protein structure is used to identify suitable labeling sites based on some sensible ‘rules of thumb’ for typical assay modalities (e.g., distance measurements via Förster resonance energy transfer, FRET). Here, we present a general, systematic, and quantitative strategy to identify optimal residues for protein labeling using a naïve Bayes classifier. Based on a literature screening and bioinformatics analysis of >100 proteins with >400 successfully labeled residues, we identified a set of four parameters, which we combined into a label score to rank residues for potential labeling success. We show with the collected literature data and complementary experiments the predictive power of this labeling score and extend the method to systematically select residue pairs for Förster resonance energy transfer experiments and accurate distance measurements. We developed a python package called “labelizer”, which includes the presented analysis of the pdb-structure, label score calculation, and FRET assay scoring to make the analysis available to a large community of researchers. For straightforward use, we further provide a webserver (<http://labelizer.net>) with a user-friendly interface to apply the selection algorithms on new protein systems. ¹

7.1 Introduction

Microscopy and spectroscopy techniques are ubiquitously used both in fundamental research and a large range of applications and assays to investigate the structure, interactions, and dynamics of macromolecules and their complexes *in vitro* and *in vivo*. Typically they require labeling of the macromolecules of interest with specific functional probes, such as EPR spin labels or fluorescent dyes [21]. While fluorescent proteins are the first choice for imaging applications in live cells [334–336], synthetic organic fluorophores are often used for high sensitivity applications including single-molecule detection [2, 337, 338] and super-resolution microscopy [9, 33, 117]. A common strategy for the (covalent) attachment of functional probes, such as synthetic organic fluorophores, EPR spin probes, nanoparticles, reactive surfaces and others, to proteins is via reactive

¹This chapter is part of the publication *Gebhardt et al.* (in preparation). For details of the individual contribution see the “Author Contributions” statement in section 7.4.

linker moieties [21, 22].

A range of labeling strategies via reactive groups exist, each with unique advantages and disadvantages. Coupling to lysine residues is possible via N-Hydroxysuccinimide (NHS)-ester amino-coupling, yet this approach suffers from low specificity, since no direct selection of labeling sites is possible due to the high abundance of lysines in protein [22]. Alternatively, His-tag or N-terminal labeling allows selective attachment of functional probes to a protein, yet this route provides little flexibility on the location [22]. As an alternative, peptide tags can facilitate covalent (e.g., CLIP, SNAP, Halo, etc.) or enzymatic probe attachment (AP-BirA, LPXTG-SortaseA, etc.) at any desired location, yet the size of the tags limits applications and can impact the protein [339]. The most widely used strategy for site-specific labeling of proteins is to introduce cysteine residues and to label their sulfhydryl-moiety via a maleimide-conjugate of the probe. Cysteine labeling enables the attachment of functional probes site-specifically [22, 70] and often with minimal effects on protein structure and function. Alternatively, incorporation of unnatural amino-acids (UAAs) can be used, in particular if (too many) native cysteines are present in a protein or to label live cells [71–76]. Cysteine and UAA labeling have become the methods of choice for spectroscopic and microscopic studies of proteins, including the characterization of structural and functional dynamics of proteins, e.g., via single-molecule Förster resonance energy transfer (smFRET) [74, 158, 262] or pulsed electron-electron double resonance spectroscopy (PELDOR or DEER) [223, 252, 261, 340]. Thereby, a fundamental challenge in biophysical and biochemical assay design relates to the selection of optimal labeling sites for the functional probe to be introduced. While both methods share many requirements for the labeling sites [209, 227, 230], we focus on application of smFRET and other spectroscopic techniques and concentrate on the design of FRET protein variants for synthetic organic fluorophore labeling from hereon.

Currently, labeling sites are typically selected based on manual inspection of the protein structure. Often, a lengthy trial and error process is necessary to identify labeling sites that are suitable for the assay, e.g. for FRET measurements that require a proper inter-fluorophore distance between both two fluorophores at residues that are not essential for protein structure or function [152, 153, 155, 308, 341–346]. Frequently encountered problems in selecting a labeling site (Figure 7.1.1A) range from fluorophore influence on protein properties including altered biochemical function of the labeled protein (Figure 7.1.1A, “Protein Properties”), low labeling efficiency (“Assay Quality”), or unwanted dye-protein interactions (“Spatial Orientation”), and finally unpredictable or unfavourable photophysical properties of the dyes at the chosen site (“Fluorescence Properties”). Suitable residues for labeling must not only enable specific and efficient attachment of fluorophores, but also avoid the problem shown in Figure 7.1.1A.

So far, the selection is – in best cases – based on some sensible rules of thumb [104], such as (i) the selection of amino acids with low conservations scores [74], (ii) the exclusion of regions on the protein surface that are related to function or present potent fluorescence quenchers such as tryptophane [104, 207, 347], (iii) distance restrictions for FRET [207, 348] or consideration of a minimally required relative surface area of the amino acid to

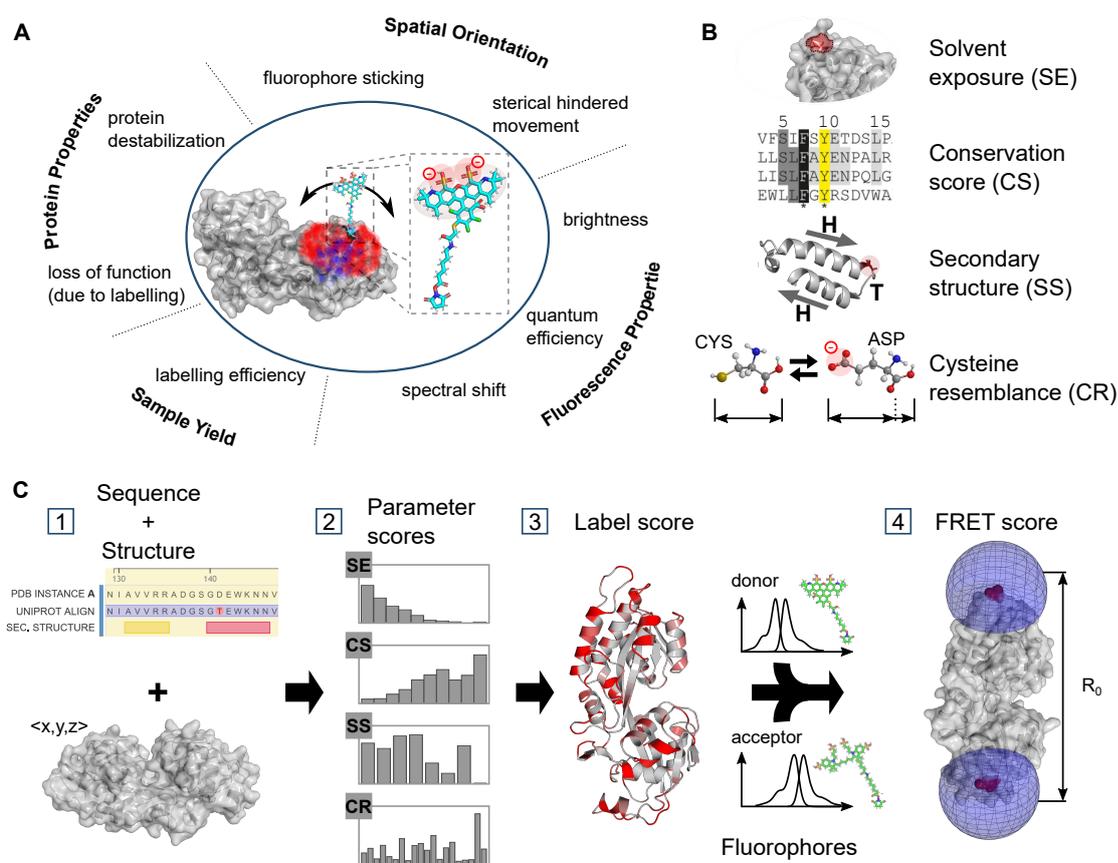


Figure 7.1.1: Labelizer workflow to compute label and FRET scores. **A** Overview of protein-fluorophore interactions that can impact the quality and success of fluorescent assays and are avoided by the labelizer analysis. **B** Parameters categories obtained from protein structures and databases used for the label scores analysis. **C** Workflow for identifying suitable labeling sites and selecting residue pairs for FRET experiments.

be altered towards a labeling position [349].

Here, we introduce a label score and an automated analysis pipeline to predict the best label sites based on structural and sequence information about a protein. To approach this selection procedure systematically, we compute a quantitative label score based on a selected parameter set that allows ranking protein residues according to their suitability as a label-site. It is important to stress that suitability for labeling is much more than high label efficiency, but a complex network of parameters (see Figure 7.1.1A).

Based on physic-chemical intuition, reports in the literature, and the ability to readily compute parameters from sequence and 3D structure information alone, we identified four broad parameters categories, which are (i) solvent exposure, (ii) residue conservation / conservation score (CS), (iii) secondary structure, and (iv) amino acid identity / cysteine resemblance (CR) as illustrated in Figure 7.1.1B. For each category, multiple specific features were defined (see Supplementary Tables 7.6.1-7.6.4), for a total of 28

parameters. To test which of the parameters are predictive of labelability and which combination might yield the most accurate predictions, we assembled a large database from a pool of available publications that report successful labeling of proteins (see methods and Supplementary Data). We introduce a general strategy based on a naïve Bayes classifier [350, 351] to rank residues for labeling and we find that parameters from all four categories help to predict labelability, but that parameters within categories are often highly correlated. Therefore, we settled on a final set of four parameters, one from each category, to predict labelability of residues (Figure 7.1.1C, step 3).

The label score can be calculated independently of the choice of label (fluorophore, EPR probe, beads, surfaces etc.), yet we here focus on the use of the label score for attachment of fluorescent dyes to proteins.

Finally, we extended our analysis to score pairs of residues for FRET assays, where the inter-dye distance needs to be close to the Förster radius, where FRET shows the highest sensitivity (Figure 7.1.1C, step 4). Therefore, we score different residue pairs according to both their labeling scores and simulated distances to obtain an optimal FRET assay expressed by a quantitative FRET score. The FRET score allows a quantitative comparison of different protein variants containing two cysteine-residues each to identify the best candidates for a successful FRET experiment. We finally show with the collected literature data and complementary experiments the predictive power of the label and FRET scores. So far, we identified >40 single cysteine labeling sites and designed ~20 FRET mutants with the software in this and other projects (see previous chapters). While our study focuses on protein labeling with fluorophores, it can be expanded to related applications, such as EPR-distance measurements, since both methods share similar requirements in regard to the residue selection [209, 227, 230].

To make the analysis routine available to a large community of researchers, we developed a standalone python package called “labelizer”, which includes the presented analysis of the pdb-structure, label score calculation, and FRET assay scoring. This allows researchers to build on our findings and adapt the code for their specific needs. For straightforward use, we also provide a webserver (<http://labelizer.net>) with a user-friendly interface to apply our analysis approach without any programming effort. While we hope for widespread use of the analysis routine and webserver, we would also like to build an open-access labeling database, where users are asked via email after some time to give feedback on the labeling efficiency and the overall assay success. This feedback will allow us to continuously improve and refine the labelizer algorithm and with that establish a systematic and quantitative design route towards identifying optimal sites for protein labeling.

7.2 Results

Database of successfully labeled residues. As a basis for a label-site selection tool, we created a database of proteins that have been successfully labeled with via introduced cysteines or UAAs by screening a large set (>1000) of peer-reviewed papers and preprints that were available before December 2020. To be included into the database of success-

fully labeled residues, the proteins and their labeled residues had to satisfy the following criteria: i) the labeled protein had a structure deposited in the PDB; ii) labeling was done by site-specific mutations of the protein using cysteines or UAAs; iii) the sites were labeled covalently with an organic fluorophore². From these proteins, we collected information on the labeled residue (chain, number), the type of mutation used for labeling (cysteine or UAA), the conducted assay (e.g., single fluorophore assays, smFRET assay with two labels, imaging, bulk FRET, etc.), and the type of label. We then gathered additional information on the protein, i.e., its oligomeric state (monomers, dimer, complexes), if only a homology model was available (due to a missing structure for the protein under investigation), and if it was a soluble or membrane protein. Overall, we identified >100 different proteins from >100 publications with >400 successfully labeled residues (see Supplementary Data). An overview of the data and summary statistics are presented in Supplementary Figure 7.7.1.

We used a standardized preprocessing routine (see Supplementary Note 7.6.1) to extract all relevant residues from the pdb-files of the proteins in the database. The final data set from 104 pdb structures contains 43357 residues, 407 of which are considered successfully labeled (the other residues are considered unknown). For all residues in our database, we computed characteristic parameter, which could be assigned to one of the four major categories (Figure 7.1.1B): (i) conservation of the residue (CS) (ii) solvent exposure of the residue (SE), (iii) secondary structure (SS), and (iv), amino acid similarity of the exchanged residues, which we abbreviate as cysteine resemblance (CR) (see Supplementary Note 7.6.1 and Table 7.7.2). The parameters were either directly extracted from the residues in question, i.e., amino acid type, mass, charge and size or calculated with the help of freely available software to compute the conservation score (ConSurf [352, 353]), solvent exposure (DSSP [354], HSE [355], MSMS [356]), and secondary structure (DSSP [354]). Altogether, we extracted and/or calculated 28 characteristic parameters.

Bayesian approach for the selection of predictive parameters. To identify suitable residues for labeling, we want to compute $P(l|s)$, the conditional probability that the residue can be labeled given a parameter value s . By Bayes' law

$$P(l|s) = \frac{P(s|l)}{P(s)}P(l), \quad (7.2.1)$$

$P(s)$ is the probability distribution of the parameter values s over all residues, whether or not they can be labeled, while $P(s|l)$ is the probability distribution of the parameter values s given that the residue can be labeled. Finally, $P(l)$ is the *a priori* probability that a residue can be labeled. While $P(s)$ and $P(s|l)$ can be readily computed from our database of labeled protein structures, $P(l)$ is harder to estimate, since the literature is biased towards reporting successful attempts of labeling. Since $P(l)$ only scales the final

²Note that we also included some spin labels or biotin-linked fluorophores, yet these represent <5% of all labels in the database (see Supplementary Figure 7.7.1).

probability, we use a simplified parameter score

$$PS(s) = \frac{P(s|l)}{P(s)} \quad (7.2.2)$$

instead of $P(l|s)$ to assess the suitability of residues for labeling. $PS(s)$, abbreviated PS later, is the odds ratio for a given parameter score to occur in a labeled residue compared to randomly selected residues. To compute PS , we calculated $P(s)$ from all 43357 residues of the 112 chains of the database and $P(s|l)$ for the ~400 labeled residues (see example in Figure 7.2.1A and Supplementary Figure 7.7.2/7.7.3).

To test whether the set of residues in our database is representative of protein structures overall, we compared $P(s)$ of our database to the distributions of a set of representative protein structures (PDBselect, November 2017) [357, 358]. Therefore, we randomly selected 300 chains (out of 4184 chains) from the PDBselect database and performed the identical analysis with those pdb files as was done with proteins in our database. We find only small deviations between the two $P(s)$ estimates, suggesting that the selection of labeled proteins is representative of the pdb content overall (Supplementary Figure 7.7.2). One notable difference is that cysteines are much less abundant (by 50%) in the database of labeled proteins compared to the PDB, suggesting that cysteine insertion and labeling is easier or at least more common for proteins with fewer native cysteines (Supplementary Figure 7.7.2). Although we also included residues that were labeled via unnatural amino acid incorporation, our database also suggests that the cysteine labeling is still the predominant strategy for protein labeling, since it makes up ~90% of all labeled residues in our database (see Supplementary Figure 7.7.1D).

To evaluate which parameters are predictive of labelability, we computed parameter score distributions for 28 parameters from all four categories from our database (Figure 2). For each PS distribution, we calculated the mean-square deviation from an equal distribution, the Gini coefficient, and the Shannon entropy (see Methods and Supplementary Table S6). We find that the PS distributions for many parameters clearly deviate from an equal distribution and contain significant information (low Shannon entropy) e.g., seen in #1: relative surface area (Wilke), #4: half-sphere exposure (10 Å), #16: variant length in homologues (see Supplementary Figure S3). Other parameters contain barely any information such as #17: cysteine in homologues (yes/no) or #27: amino acid charge (Supplementary Figure S3). Interestingly, it is rather unimportant for a selected residue if a cysteine is found in one of the homologue proteins at the same position or not if the residues later appears to be labeled well (see parameter #17, Supplementary Figure S3). One could have expected that those residues are easily mutated to cysteines, and therefore, significantly enhanced in our scoring.

Having established the predictive power of single parameters, we need to select which combinations should be used for predictions. We first calculated the correlation between all parameters to judge their statistical independence (Figure B). Since we deal with categorical data (e. g. secondary structure) and numerical data (e. g. relative surface area), we used Pearson correlation, interclass correlation and Cramer's V for the combinations of numeric-numeric, categorical-numeric, categorical-categorical values, respectively (see

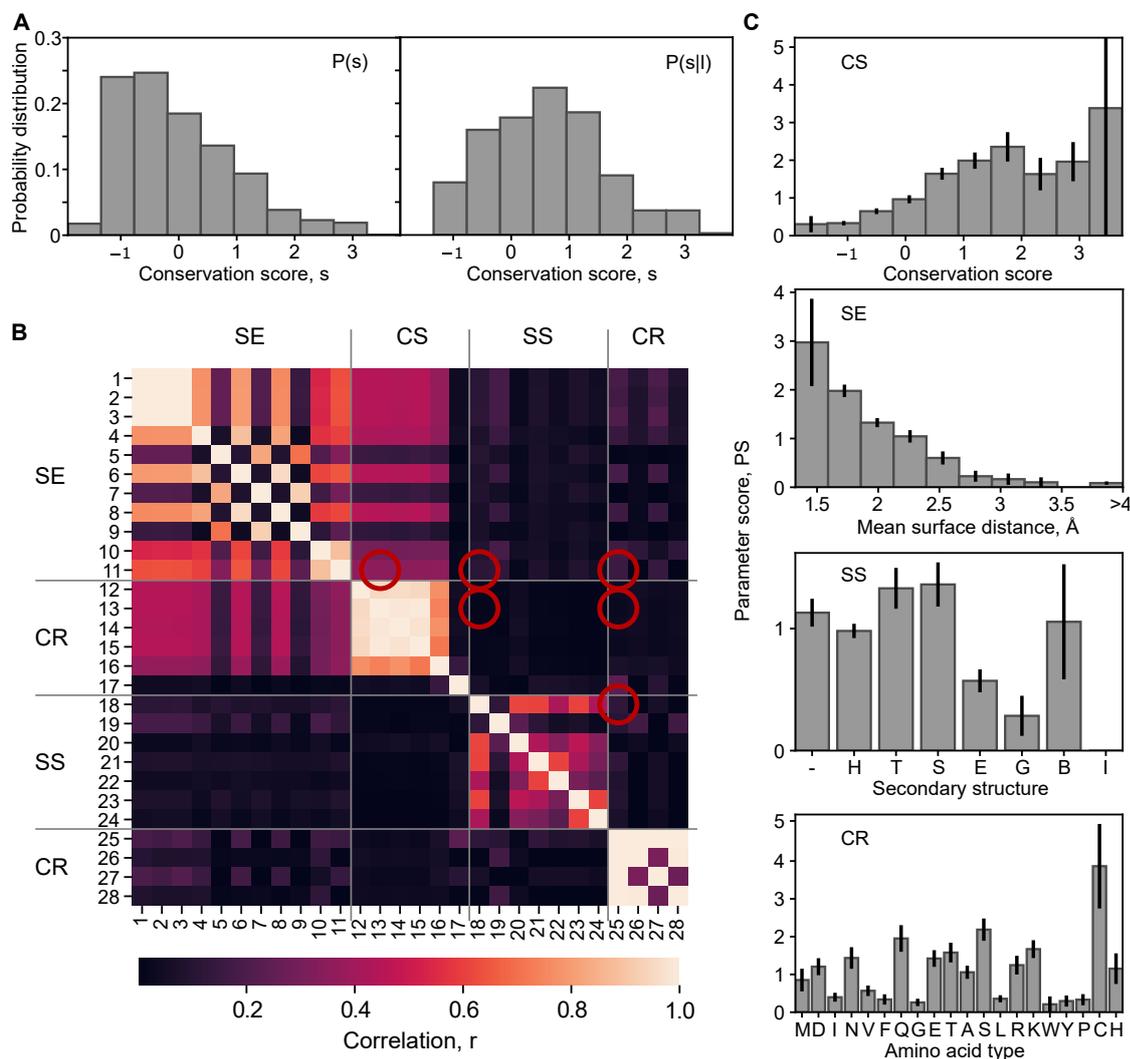


Figure 7.2.1: Parameter score analysis. **A** Probability distribution $P(s)$ for parameter #13 (conservation score from ConSurf, see Supplementary Table 7.6.2) for all analyzed residues (left) and for the successfully labeled residues $P(sl)$ (right). **B** Correlations between all parameters were calculated based on Pearson correlation (numeric-numeric), interclass correlation (categorical-numeric), or Cramer's V (categorical-categorical). The cross-correlations of the final parameter selection for the label algorithm are marked (red circles). **C** The parameter score distributions $PS = P(sl)/P(s)$ is shown for the selected default parameter selection. The upper row shows the resulting parameter score from the distribution in A. For the other categories, the parameters are: solvent exposure (#11, mean surface distance, Supplementary Table 7.6.1), secondary structure (#18, secondary structure from DSSP, Supplementary Table 7.6.3), and cysteine resemblance (#25, amino acid identity, Supplementary Table 7.6.4). Error bars were determined from counting uncertainties (standard deviation).

methods for details). We formed sets of four parameters and used a correlation measure (2-norm of all paired correlations, see Methods) to calculate a joined correlation estimator for all combinations of parameters (Supplementary Figure 7.7.5). Whereas this combined correlation measure shows rather high values (>1) for most combinations of two or more parameters within the same categories CS, SE, CR, and SS, the correlation of combinations of parameters from different categories is smaller (<0.5). This effect is independent of whether parameters with high or low predictive power (MSD) are combined (see methods and Supplementary Figure 7.7.5). The overall low correlation in the second case justifies our categorization and their consideration as independent variables, if we restrict our selection to one parameter per category. The strong correlation within one category also suggests that the actual choice of the concrete parameter is not critical, i.e., any of the parameters can account for the properties of the respective category.

The joined label score predicts potential labeling sites. To combine parameters scores into a final assessment of the labelability of a given residue, we introduce a joined label score, LS . By standard probability theory different parameters s_i can be combined by

$$P(l | \bigcap_{i=1}^n s_i) = \prod_{i=1}^n P(l | s_i) = \frac{\prod_{i=1}^n P(s_i | l) P(l)}{\prod_{i=1}^n P(s_i)} \quad (7.2.3)$$

under the assumption that the parameters are independent. This naïve Bayes classification [350, 351] is known to give good predictions for low and moderately correlated parameters [359–362], which is the case for our parameter set (Figure 7.2.1B). In general, any residual correlation pushes the calculated probability values towards the extremes of 0 and 1 [362]. However, we again use parameter scores as comparative figures without the meaning of probabilities and combine the PS_i into the joined label score by taking the geometric mean:

$$LS = \sqrt[n]{PS_1 \cdot \dots \cdot PS_n} \quad (7.2.4)$$

An important question was which concrete parameters and PS scores should be selected for inclusion to the LS . We included one parameter from each of the four categories CS, SE, SR, and SS, for which concrete values were mapped onto the structure of the phosphate binding protein PBP (Figure 7.2.2A). For a rationale selection of parameters, we strived (i) to maximize the range of values for LS , (ii) to maximize the enhancement/suppression level of LS of the successfully labeled residues in the database for high/low LS scores and (iii) to maximize the statistical significance level of LS values of random residues compared to LS values of the labeled residues in the database.

Based on these criteria, we were able to identify several parameter sets with predictive power but also many combinations without significant information. LS values based on parameters sets such as #5 (HSE, second half, 10 Å), #17 (cysteine in variety), #24 (secondary structure two positions ahead), #26 (amino acid mass) scored low in our selection criteria and do not show a significant prediction power (Figure 7.2.2C and Supplementary Figure 7.7.6). Here, the probability distribution of all residues, $P(LS)$, is

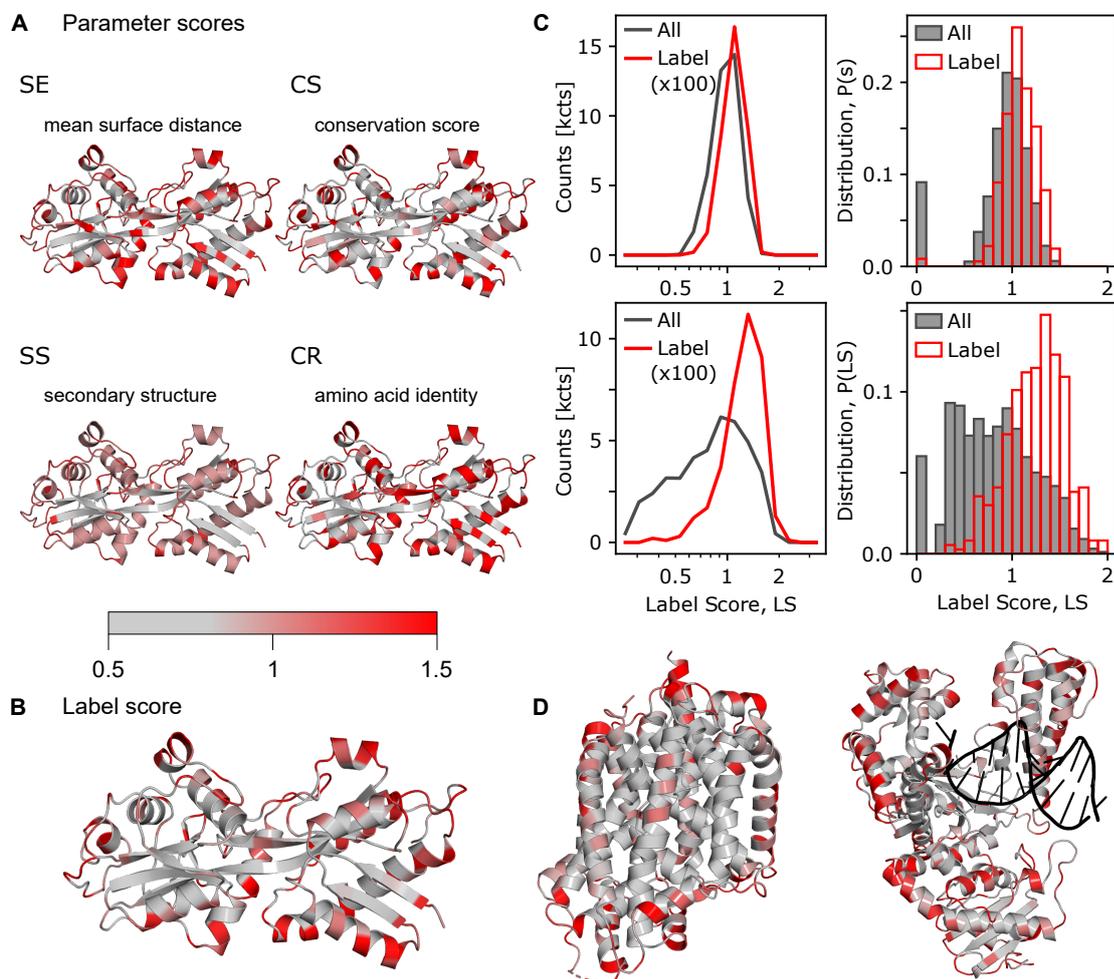


Figure 7.2.2: Parameter and label scores visualization. **A** Visualization of the four selected parameter scores from the four categories, which are used as default settings in our webserver for an exemplary protein (PBP of *E. coli* [363, 364], pdb:1OIB). **B** Visualization of the label score based on the parameter scores in **A** with PBP. **C** Label score histogram of all residues (gray) and labeled residues (red) in our database (left) and corresponding probability distribution of the label scores (right). **D** Further examples of the LS values in the membrane protein (left, LeuT of in *A. aeolicus* [365, 366], pdb:2A65) and a DNA-binding protein (right, DNA polymerase I of *B. stearothermophilus* [367, 368] with DNA template, pdb:1L3U) visualize the universality of the approach.

almost identical to the distribution of the labeled residues $P(LS|l)$. Other parameter sets show highly distinct label score distributions between all residues and labeled residues, e.g., as shown in Figure 3C (bottom). In the end, we decided for an optimal set of the following parameters: mean surface distance (SE, #11), conservation score (CS, #13), secondary structure of the labeled residue (SS, #18), and the mutated amino acid (CR, #25) as our default parameter selection for LS calculation in this manuscript and for the associated webserver. We choose this set out of all well-performing combinations, because of the intuitive nature of all selected parameters. In the labelizer python package the user can, however, select any parameter combination, since there is no right or wrong choice of parameters sets, only the predictive power of the resulting LS values will differ. Importantly, the choice can be extended to an arbitrary number of different categories and parameters (see discussion).

We find that the label scores LS of all residues are centered around 1 with values ranging from 0.2 to 2 (except 5% failed calculations with $LS = 0$, see Figure 7.2.2C). The ratio of LS distribution of successfully labeled residues in the database and all label scores shows that high label scores (>1.5) are significantly enhanced by a factor of $\sim 3-4$ for the labeled residues, whereas low label scores (<0.5) are suppressed by a factor of ~ 10 (see Supplementary Figure 7.7.6). These results show that the defined label score is a suitable measure to rank and compare residues for their suitability for labeling. We stress again that this is not a direct relation to a high labeling efficiency, but the LS value considers different parameters (Figure 7.1.1B). To visualize the LS values, we mapped the calculated LS values for three different types of proteins (soluble protein, membrane protein and DNA binding protein) onto the crystal structure (Figure 7.2.2B/D). We see that only a small proportion of around 10% of all residues is rated very good for labeling ($LS > 1.5$).

Experimental validation of the predictive power the label score. To experimentally validate our approach, we collected a data set of around 20 labeled mutants with single-cysteine labeling sites from our lab (68 measurements in total), which were selected for experiments and could all be labeled. No direct correlation between the label score LS and the degree of labeling $DOC = (\# \text{ of fluorophores})/(\# \text{ of proteins})$ (see methods for details) is seen (see Figure 7.2.3). Yet, we observe that the average LS value is ~ 1.4 and almost 90% of all proteins have LS values >1 . We focused on realistic mutants and did not include measurements of deeply buried residues with low label scores $LS < 0.5$, which would obviously exhibit low label efficiencies. Further, what needs to be considered in interpretation of Figure 7.2.3 is that it might be possible to label protein residues with low LS values, yet the function/structure etc. of these proteins might be hampered. Based on this, we see that the LS score is not a parameter that only allows to select residues for high labeling efficiency, but consider diverse aspects and to retain functionality as introduced earlier.

Python package and webserver. The analysis routine described above was implemented in a python package with multiple features, which can be modified and ex-

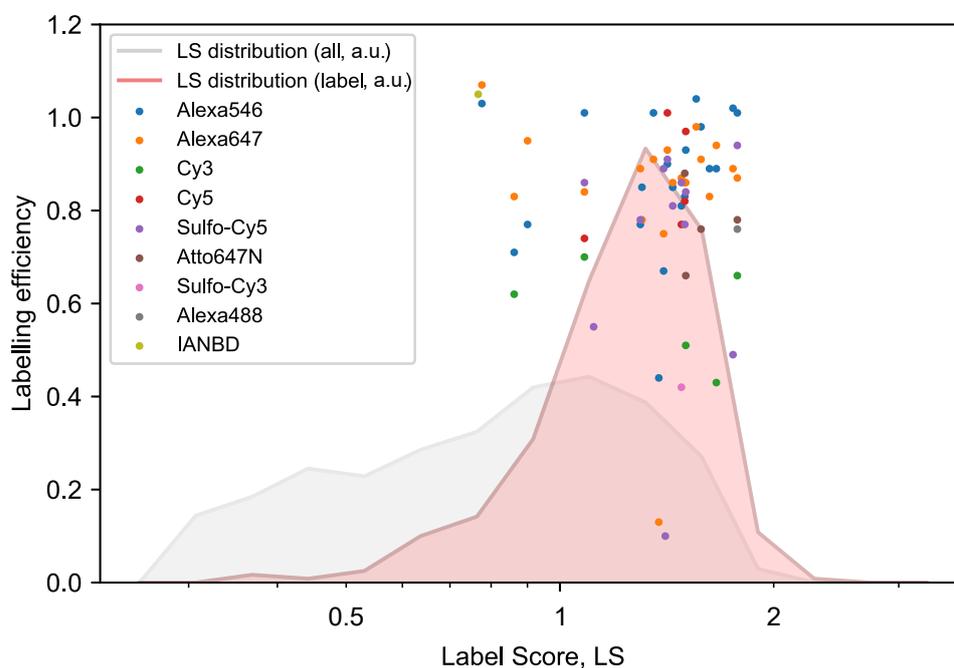


Figure 7.2.3: Experimental validation of label score. Label efficiency of 22 single mutants in MalE were labeled with 9 different dyes (68 measurements in total). The distribution of label scores of all residues (gray) and the successfully labeled residues (red) are included as orientation (arbitrary units). We do not observe a correlation of label efficiency and label score for the selected mutants with in general high label scores $LS > 1$.

tended to the researchers' special needs. Since this requires programming knowledge, an in-depth study of the documentation and additional existing software is required (DSSP [354], MSMS [356], FPS [206]), we provide the key functionalities as a webserver with an intuitive and user-friendly interface under <http://labelizer.net>. The webserver supports the label score calculation and its use for FRET experiments (see below) with the default parameters and the mostly used fluorophores. For this purpose, pdb-files can be loaded automatically and preprocessed from the pdb-database. We further retrieve conservation scores directly from the ConSurf server [352, 353] without the need of uploading any information (except when modified or user-specific pdb files should be used). The webserver visualizes the different scores in an interactive 3D structure viewer and provides a table with filter options for customized restrictions upon residue selection. Furthermore, human-readable result files (csv, json) are provided for subsequent analysis.

Extension of the LS score for FRET experiments. While efficient labeling of single protein residues is critical for a large range of applications, additional challenges

arise for FRET experiments, where two labeled residues with a suitable inter-fluorophore distance are required for a functional assay. To expand our labelability prediction to FRET experiments, we combine the label score LS with an additional parameter for the rationale design of FRET experiments. The central idea is to select residue pairs for FRET experiments that i) can be labeled well (considering LS), ii) are at a distance close to the Förster radius of the dyes used (for maximum sensitivity) and iii) allow for detection of conformational motion.

Criteria i/ii are relevant to the case where one protein structure is available, and a residue pair is wanted with a distance close to the Förster radius of the dye pair. In this scenario, the researcher can use combinations of residues in different domains of the protein for maximal sensitivity. We define the FRET score of a residue pair $\{i, j\}$ for a single available protein structure as

$$FS_s = \sqrt{LS_i LS_j} \cdot \left(1 - 2 \left| \frac{1}{2} - E_{i,j} \right| \right), \quad (7.2.5)$$

with the label scores LS of two residues i and j in the protein structure with corresponding predicted FRET efficiency $E_{i,j}$ (see Supplementary Note 7.6.2 for details on the FRET efficiency prediction). If two (interconverting) structures of a protein are available, one might want to find most informative FRET pairs with the largest possible shifts in FRET efficiency E . This scenario is encountered when ligand binding, protein-protein interactions or other macromolecular interactions are studied and distinct structures of the same protein, e.g., ligand-free and ligand-bound, are available. We define the FRET score of a residue pair $\{i, j\}$ for two available structures A and B of the same protein as

$$FS_d = \sqrt{LS_i^A LS_i^B} \sqrt{LS_j^A LS_j^B} \cdot |E_{i,j}^A - E_{i,j}^B|, \quad (7.2.6)$$

with the label scores LS of two residues i and j in two protein structures A and B with their corresponding FRET efficiencies $E_{i,j}^A$ and $E_{i,j}^B$, respectively.

A crucial step for calculation of the two FRET-scores is the ability to calculate accurate inter-dye distances (Figure 7.2.4A). The labelizer package supports three models for the fluorophore distance calculation. An approximate approach is to use the C_α or C_β distances between two residues as a rough estimate for the fluorophore distance and their expected approximate FRET efficiency [349]. However, these distances differ >10 Å from the actual fluorophore positions, in particular due to the flexible nature of the 10-20 Å long linkers used for fluorophore attachment [283] (Figure 7.2.4B).

Thus, it is necessary to simulate the accessible volumes (AVs) and resulting fluorophore distance distributions for a given choice of labeled residues (Figure 7.2.4A). While molecular dynamics simulations have been successfully used to compute fluorophore position distributions [181, 369, 370], they are computationally expensive and, therefore, impractical as a screening tool. An alternative is the coarse-grained simulation of FRET-restrained positioning and screening (FPS) for AV simulations, which is faster and an accepted way to obtain inter-fluorophore distances in the smFRET community [206, 371]. Here, all positions on a grid are examined to decide whether it can be

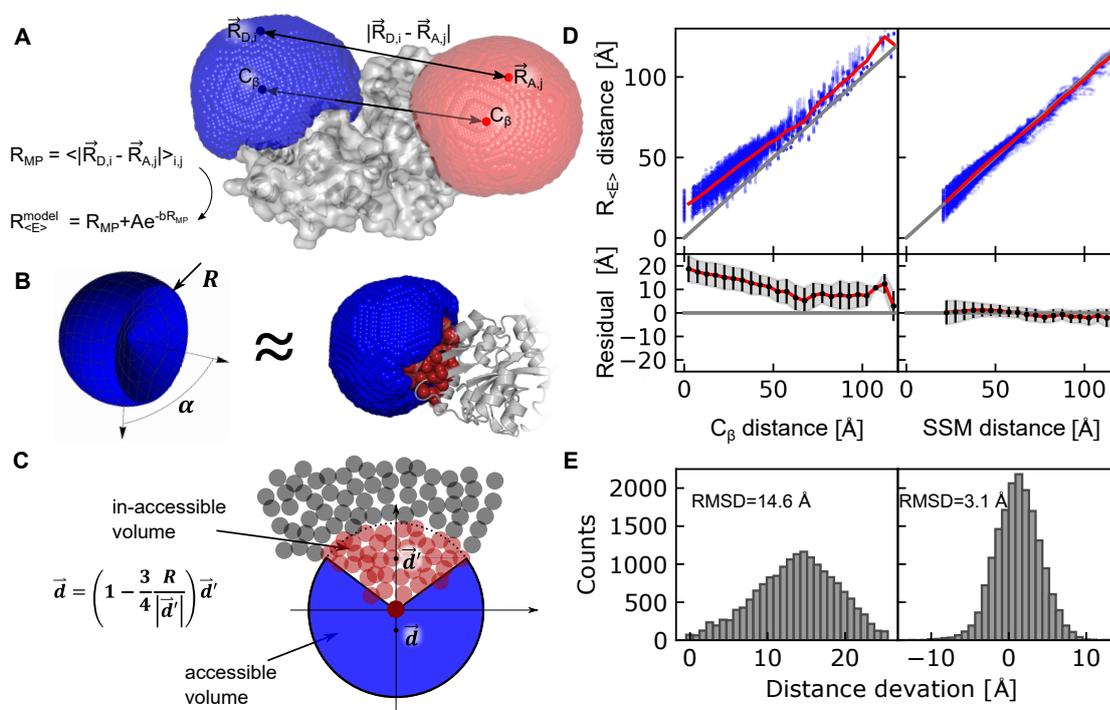


Figure 7.2.4: Accurate prediction of inter-dye distances on proteins. **A** Distance estimation with FPS computes a grid-based accessible volume to compute the mean-position of the fluorophore $\langle \vec{R}_F \rangle$, the averaged inter-fluorophore distance $R_{MP} = \langle |\vec{R}_{D,i} - \vec{R}_{A,j}| \rangle_{i,j}$, and the efficiency weighted average fluorophore distance $R_{\langle E \rangle}$. **B** Approximation of the accessible volume with a spherical sector. The spherical sector (left) is defined by the radius R (linker length of the fluorophore) plus an opening angle α and approximates the accessible volume simulated with FPS software (right, blue volume) [206]. The red spheres represent the protein atoms within radius R from the C_β atom. **C** Illustration of the determination of the mean position in 2D. The circles represent the atoms of the protein. The inaccessible volumes are the atoms within a radius R (pale red circle) to the C_β atom (dark red circle). **D** Comparison of C_β distances and modeled distances with the introduced spherical sector approximation compared to FPS-derived distances in 10 selected pdb structures with 35 different fluorophore parameters ($N=32116$). **E** Histogram of the distance offsets from C for the distances in the range $40 \text{ \AA} < R_{\langle E \rangle} < 75 \text{ \AA}$ ($N=17359$).

occupied by a fluorophore of distinct size and linker length. Comparing the calculated C_α or C_β distance of the residues with FRET-averaged distances $R_{\langle E \rangle}$ from AV simulations reveals deviations of 10 to 15 Å (RMSD, see Figure 7.2.4B and Supplementary Figure 7.7.8A). Since a good match between experimental values and AV calculated ones was demonstrated, FPS is a good approximation of the measured distances [111, 160, 164, 308, 372, 373], yet the computation involved is still rather slow for screening purposes where >10.000 residue pairs should be considered.

Therefore, we here introduce a faster distance estimation based on a spherical sector model (SSM) for dye-accessible and dye-inaccessible volumes for screening purposes, which is 100 to 1000 times faster than currently available simulations. Our algorithm relies on an approximation of the accessible volume by a spherical sector of angle α and radius R representing the linker length of the fluorophore (see Figure 7.2.4C). The atoms of the protein within radius R from the attachment site (C_β atom) define an inaccessible site (C_β atom) define an inaccessible volume (see Figure 7.2.4B/C, pale red spheres). We find a direct relation between the center of mass of these atoms \vec{d}' (inaccessible volume) and the center of mass of the accessible volume \vec{d} as

$$\vec{d} = \left(1 - \frac{3}{4} \frac{R}{|\vec{d}'|}\right) \vec{d}'. \quad (7.2.7)$$

We include a small correction ε (~ 0.5 Å for typical fluorophores) to the linker length $\tilde{R} = R + \varepsilon$ in this formula to compensate for the size of the fluorophore core (see Supplementary Note 7.6.2, Supplementary Figure 7.7.7) and we use an estimation to convert the distance of the mean positions to FRET-averaged distances (see Supplementary Note 7.6.2, Supplementary Figure 7.7.8). For this, we performed distance simulations for 100 donor-acceptor pairs in 10 different protein structures, where we altered the linker length and the dye dimension with 35 variations, i.e., 35.000 distance simulations in total. This spherical sector approach generates comparable results as the FPS method with a deviation of ± 3 Å (RMSD, Supplementary Figure 7.7.8) by adding this small correction factor ε to the linker length R in equation 7.2.7 and using the transfer function to compute $R_{\langle E \rangle}$ from R_{MP} (Figure 7.2.4E)³, which is on the order of the intrinsic distance precision of FRET [111]. Our method can screen >10.000 residue pairs within seconds and with that provide a first selection of suitable FRET-labeling positions. A subsequent reanalysis can then be done based on FPS [206] for the most promising variants identified. The mean-position distances are converted to FRET-averaged distances with an exponential correction factor at small distances (see methods and Supplementary Figure 7.7.8). Therefore, the spherical sector method allows to screen >10.000 FRET-pairs within seconds on a single CPU with <1 ms calculation time per residue-pair (see Supplementary Table 7.7.4).

Our standard strategy for the FRET assay design is the calculation of the FRET Score FS in the “labelizer” package (default settings) by first calculating the distances of the two spherical sectors for all pairs and refine only the best few hundred FRET

³The correction compensates the gap between protein atoms and the accessible volume due to finite collision radii of protein atoms and the fluorophore core (see Figure 7.7.7).

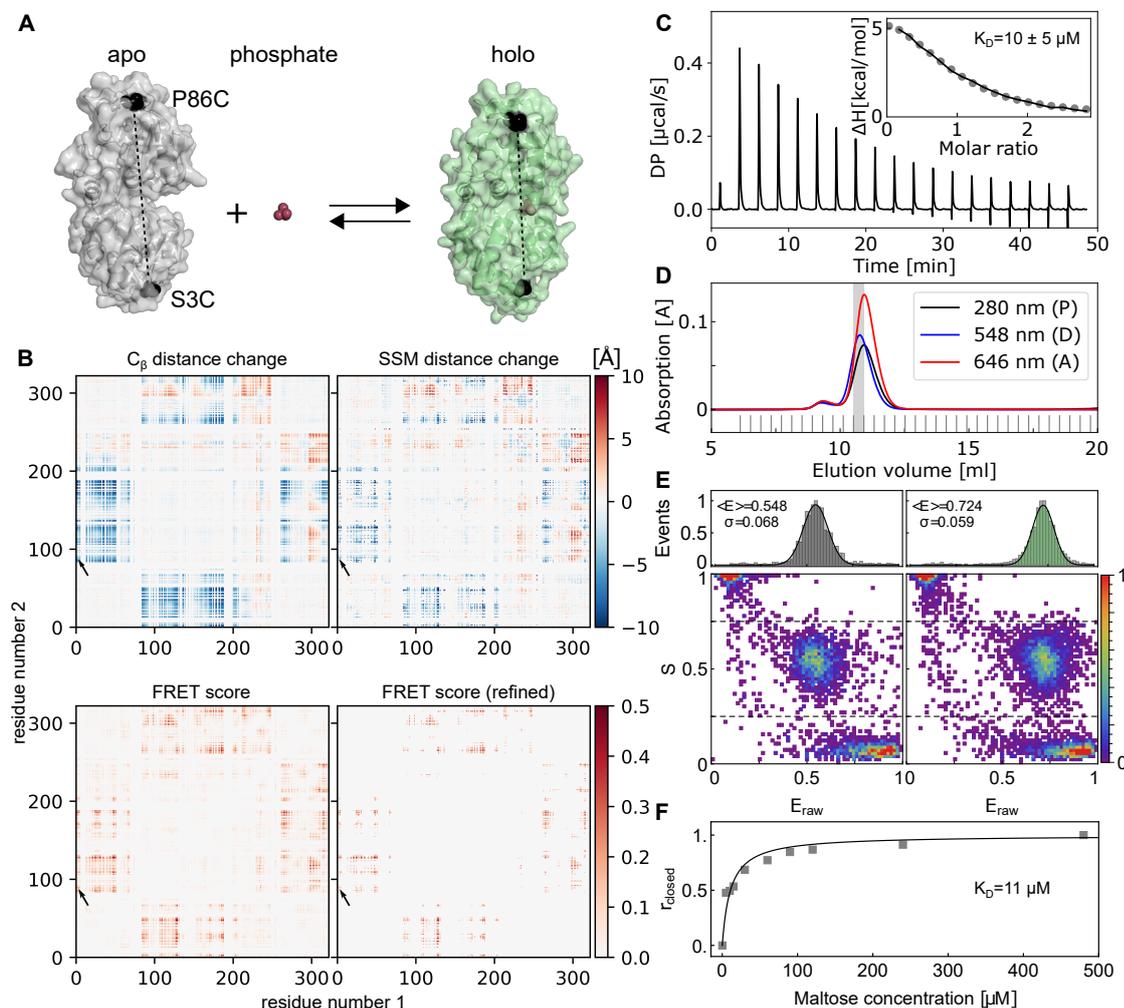


Figure 7.2.5: FRET mutant selection and experimental demonstration. **A** Crystal structures of PBP in the apo (grey, PDB-ID: 1OIB) and holo (green, PDB-ID: 1PBP) states with mutations S3C and P86C indicated in black. **B** Maps illustrating the distance change and associated FRET score for all pairs of mutants in PBP. The selected mutation 3-86 is marked with an arrow. The distance change of the attachment atom (C_β , top-left) and the distance change based on the simulated spherical sector model (top-right) show a clear pattern of correlated movements. The calculated FRET score map of all pairs with average label score $LS > 1$ shows only a few spots ($\sim 4\%$) with promising FRET scores $FS > 0.2$ (bottom-left). The selection of the 1000 pairs with highest FRET score and a refinement with FPS software (bottom-right) shows only minor variation compared to the screening map (bottom-left) for the analyzed data points. **C** ITC measurement of the mutant 3-86 reveals a dissociation constant of $K_D = 10 \pm 5 \mu\text{M}$. **D** Size exclusion chromatogram of the mutant in C labeled with LD555 and LD655. **E** ES-FRET histograms of PBP from labeling in C in apo state (left) and with $480 \mu\text{M}$ phosphate (see methods for details). **F** Calculated binding curve from FRET measurements.

pairs with the FPS AV-simulation [206]. Alternatively, our python package allows to manually select to calculate the C_β distances (low accuracy) or the FPS-derived derived distances (long runtime) for all residues, if desired.

Experimental test of the FRET score methodology. We validate our strategy to design optimal FRET pairs for the phosphate binding protein, which undergoes a ligand-induced transition from an open ligand-free apo to a ligand-bound closed holo state (Figure 7.2.5A). We used the labelizer routine to design optimal FRET mutants for the dye pair LD555 and LD655 on PBP for an interprobe distance that is approximately equal to their Förster radius of 5.2 nm and has the largest amplification of the resulting FRET changes for addition of inorganic phosphate. This was done by comparison and labelizer analysis of the apo (pdb: 1OIB) and holo state crystal structures (pdb: 1PBP) as shown in the distance-difference maps based on the different distance measures C_β and SSM and the calculated FRET scores before and after refinement (see Figure 7.2.5B).

We identified the pair 3-86 as suitable due to the FRET scores. The mutant shows wildtype-like affinity without and with labels (Figure 7.2.5C/F) shown in ITC and FRET titration experiments. The labeling procedure was done according to standard protocols in the lab and gave a combined yield of 57% and 44% for both cysteines with LD555 and LD655, respectively, using absorbance measurements in size exclusion chromatography (SEC) (Figure 7.2.5D). The success of the prediction procedure is seen in the resulting high quality smFRET histograms (Figure 7.2.5E).

7.3 Discussion and conclusion

Here, we present a general strategy to identify optimal residues for protein labeling using a naïve Bayes classifier. Based on a literature screening and bioinformatics analysis of >100 proteins with >400 successfully labeled residues, we identified a set of four parameters, which we combined into a label score to rank residues for potential labeling success. We show with the collected literature data and complementary experiments the predictive power of this labeling score and extend the method to systematically select residue pairs for Förster resonance energy transfer experiments and accurate distance measurements. To make the analysis available to a large community of researchers, we developed a publicly available python package called “labelizer”, which includes the presented analysis of the pdb-structure, label score calculation, and FRET assay scoring. For straightforward use, we provide a webserver with a user-friendly interface to apply the selection algorithms and contribute to an open-access database. We hope to provide other scientists in the FRET- and fluorescence community with a tool that enables them to design and justify the residue selection with an un-biased and universal measure.

The labelizer project in the current state is planned to be the starting point for a continuously improving analytical prediction tool for fluorescence assay design (and others). Therefore, we plan to extend the amount of training data with a feedback loop, where researcher can provide information on successfully and unsuccessfully labeled

residues via a form on our website. Especially, unsuccessful results are of particular interest, since those results are barely found in literature, where mainly successful results are published. Therefore, we call the researcher to use the labelizer and to provide feedback for ongoing improvement.

For future improvement and extension of the database and the algorithm, we consider to revise the available PS values by an extended database, where particularly residues with low chance for labeling will be an important new class of information. We also plan to combine different parameter scores to improve the prediction, which might happen within one category, e. g., via simultaneous use of HSE and RSA (see Supplementary Note 3.4.2) to include direction and better surface exposure or between categories, e.g., solvent exposure and cysteine resemblance. We can also see once a much larger dataset of labeled and non-labeled residues was available, an application of machine learning procedures (e.g. supporting vector machine or neural networks) should be possible. This might also be the bases to establish specific databases for cases other than cysteine labeling to meet e.g., very specific requirements for UAA or EPR-labels.

A next interesting perspective is the extend to include more parameters (fluorophore dependent) with a potential differentiation of labelability based on the selected fluorophores related to specific charge environment on the protein, tryptophan proximity. Further normal mode analysis (e.g. NMSim webserver [374, 375]), mutation specific energy analysis (e.g. SDM [376, 377]), or tailored MD-simulations [159] are options to consider which FRET-residue pairs are most promising for analysis of conformational motion when only one available protein structure is available. Finally, the concept of FRET scores could be extended towards other fluorescence assay types related to fluorophore quenching, protein-induced fluorescence enhancement and many others.

7.4 Acknowledgements and Author Contributions

Acknowledgments This work was financed by an ERC Starting Grant (ERC-StG 638536 - SM-IMPORT to T.C.), Deutsche Forschungsgemeinschaft within GRK2062 (project C03) and SFB863 (projects A11 and A13), LMUexcellent, the Center for integrated protein science Munich (CiPSM) and the Center for Nanoscience (CeNS). C.G. acknowledges a PhD fellowship from the Studienstiftung des deutschen Volkes. We thank all members of the Cordes lab for active testing of the labelizer procedure and webserver. We thank Oliver Brix and Rebecca Mächtel for providing information on successfully labeled protein residues, Markus Springer for help with data acquisition for MaleE, and Leonor Correia for help with data acquisition for PBP.

Author contributions C.G. and T.C. conceived and designed the study. C.G. performed research, data analysis and software implementation. C.G. and P.B. implemented the webserver. D.A.G. provided reagents. J.L. provided new analytical tools. J.L. and T.C. supervised the study. C.G., J.L. and T.C. discussed and interpreted the results and wrote the manuscript.

7.5 Materials and Methods

7.5.1 Database generation and evaluation

To identify parameters with predictive power for the possibility to label residues in proteins, we created a dataset based on a non-automated screening of more than 1000 publications published or preprinted which were available on or before December 2020 with a focus on the field of single-molecule microscopy and single-molecule FRET. The papers were screened to identify proteins and residues that were labeled successfully with a fluorophore and that satisfied the following criteria: (i) the proteins had a structure available in the PDB-database (with PDB identification code); (ii) the protein was labeled via site-specific mutagenesis and introduction of cysteines or UAAs; (iii) the protein was successfully labeled synthetic organic fluorophores (or spin labels) and used preferentially single-molecule assays. In order to increase the number of database entries, we complemented our search whenever some information was missing. Typical cases were missing PDB identification codes or residue numbers. In this case, the required information was obtained from other referenced papers (often) of the same research group.

For each successfully labeled protein variant, which fulfilled the aforementioned criteria, the following information was collected⁴:

- Protein (PDB identification code)
- Soluble or membrane protein
- Stoichiometry (monomers, dimer, complexes)
- Homology model (true/false)
- Labeled residue (chain and residue number)
- Mutation (cysteine or UAA)
- Assay type (smFRET, imaging, bulk-FRET, other)
- Name of labeled fluorophores
- Research group
- Publication reference

The final database with information on those positions in proteins that were successfully labeled had 406 successfully labeled residues in 112 different chains in 104 different protein structures (Supplementary Data). As comparison, we used a representative set of proteins (PDBselect, November 2017) [357, 358] as a random reference database to check how representative the analyzed pdb structures are. Therefore, we randomly selected 300 chains (out of 4184 chains) from the PDBselect database and performed the identical analysis with those pdb files.

⁴Additional notes were gathered to account for issues such as: (i) dimer and polymer protein structures, which were crystallization artefacts and needed to be deleted for structural analysis; (ii) missing residues in protein structure, i.e., when parts of the protein were not resolved completely; (iii) we identified inconsistencies or missing information

Parameter frequency calculation. For every extracted parameter, the relative frequency defines a parameter score

$$PS = \frac{P(s|l)}{P(s)} \quad (7.5.1)$$

where $P(s)$ is the probability distribution of the score s (calculated from the 112 chains of the database) and $P(s|l)$ is the probability distribution of the score given that the residue was labeled (calculated from the 406 successfully labeled residues).

The error bars σ_{sl} and σ_s for $P(s|l)$ and $P(s)$, respectively, were determined from Poissonian counting statistics as $\sigma_{sl} = \sqrt{P(s|l)/n}$ and $\sigma_s = \sqrt{P(s)/n}$ with n being the total number of evaluated residues. The error bar σ_{PS} for PS follows from standard error propagation rules:

$$\sigma_{PS} = \sqrt{\frac{\sigma_{sl}^2}{P(s|l)^2} + \frac{\sigma_s^2}{P(s)^2}} PS. \quad (7.5.2)$$

Parameter information analysis. To evaluate the amount of information a single parameter score inheres, we used three measures to estimate the deviation from a equal distribution, which corresponds to the case of zero information.

We used standard Pearson correlation for a pair of numeric parameters

$$MSD(PS) = \frac{\sum_{i=1}^n (PS(i) - 1)^2}{n}, \quad (7.5.3)$$

with n the number of bins/categories.

We used standard Pearson correlation for a pair of numeric parameters

$$gini(PS) = \frac{\frac{n-1}{2} \sum_{i=1}^n PS(i) - \sum_{i=2}^n \sum_{j=1}^{i-1} PS(j)}{\frac{n}{2} \sum_{i=1}^n PS(i)}, \quad (7.5.4)$$

with n the number of bins/categories.

We used an adapted Shannon entropy accounting for the number of bins/categories as

$$H(PS) = \frac{-\sum_{i=1}^n \widetilde{PS}(i) \cdot \ln(\widetilde{PS}(i))}{\ln(n)}, \quad (7.5.5)$$

with a normalized parameter score $\widetilde{PS}(i) = PS(i) / (\sum_{j=1}^n PS(j))$ and n the number of bins/categories.

Parameter cross-correlation. To evaluate the mutual statistical dependence of all calculated parameters, we use three different types of correlation coefficients, depending on the datatypes of the parameters:

(i) We used standard Pearson correlation for a pair of numeric parameters

$$r_{nn} = \frac{\sum_{i=1}^n (x_i - \bar{x}_i)(y_i - \bar{y}_i)}{\sqrt{\sum_{i=1}^n (x_i - \bar{x}_i)^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y}_i)^2}} \quad (7.5.6)$$

with n different residues with parameter scores x_i , y_i and corresponding mean values $\bar{x} = 1/n \sum_{i=1}^n x_i$ (and \bar{y} accordingly) [378].

We used the interclass correlation for a pair of a categorical parameter and a numeric parameter [379]. The n data points are grouped in k categories c_i with $i \in \{1, 2, \dots, k\}$ of length n_i .

$$r_{CN} = \frac{MST - MSE}{MST + (n_0 - 1)MSE}, \quad (7.5.7)$$

with

$$\begin{aligned} MST &= \frac{\sum_{i=1}^k n_i \sum_{j=1}^{n_i} (\bar{x}_i - \bar{x})^2}{k-1}, \\ MSE &= \frac{\sum_{i=1}^k \sum_{j=1}^{n_i} (x_{i,j} - \bar{x})^2}{n-k}, \\ n_0 &= \frac{n - \sum_{i=1}^k n_i^2/n}{k-1} \end{aligned}, \quad (7.5.8)$$

where \bar{x}_i is the mean of category i , \bar{x} the mean of all data, $x_{i,j}$ the j^{th} numeric value in category c_i , and $(n_0 - 1)$ the averaged interclass degree of freedom [379].

We used the Cramer's V for a pair of a categorical parameters [380]. The data are grouped in the two categories c_i with $i \in \{1, 2, \dots, k\}$ and d_j with $j \in \{1, 2, \dots, l\}$.

$$r_{cc} = \sqrt{\frac{\chi^2}{n \cdot (\min(k, l) - 1)}}, \quad (7.5.9)$$

with

$$\chi^2 = \sum_{i=1}^k \sum_{j=1}^l \frac{(n_{i,j} - \tilde{n}_{i,j})^2}{\tilde{n}_{i,j}}, \quad (7.5.10)$$

where $\tilde{n}_{i,j} = \left(\sum_{j'=1}^l n_{i,j'} \right) \left(\sum_{i'=1}^k n_{i',j} \right) / n$, n total number of residues and $n_{i,j}$ number of residues of class c_i and d_j . The cross-correlation was calculated for every combination of the 28 extracted parameters to identify dependencies as shown in Figure 7.2.1B.

Parameter selection criteria. The selection of a suitable parameter set is based on two criteria. First, a joined correlation for any combination of parameters is calculated as

$$r_{set} = \sqrt{\sum_{i=1}^n r_{ij}}, \quad (7.5.11)$$

with r_{ij} the correlation of parameter i with j and n the number of selected parameters (in our case 4). Secondly, we used three measures to characterize our parameter sets: (i) We calculate the t value of the calculated label scores as

$$t = \frac{\mu_l - \mu_{all}}{\sqrt{SEM_l^2 + SEM_{all}^2}}, \quad (7.5.12)$$

with the mean values μ_l , μ_{all} and standard error of mean SEM_l , SEM_{all} of the labeled/all residues, respectively. (ii) The dynamic range was calculated as the standard deviation of the logarithmic values $\sigma(\log(LS_{all}))$. (iii) The suppression/enhancement of the labeling score of labeled residues for small/large values was calculated from the slope of a linear least square fit to the logarithm of the label score LS and the label score distribution of labeled residues and all residues. The data are binned into logarithmic bins with bin intervals $[1.5^i, 1.5^{i+1}]$ for $i \in \{-12, \dots, 11\}$ and fitted to the function

$$\log\left(\frac{P(LS|l)}{P(LS)}\right) = m \log(LS) + \log(c) \quad (7.5.13)$$

where LS is the label score and $P(LS)/P(LS|l)$ the probability distributions of the label score of all and the labeled residues. The slope m is used as analysis parameter from the fitted values m , c .

7.5.2 Protein production and labeling

In the current study we used single cysteine variants of MaleE (Figure 7.2.3) that were obtained and fluorophore-labeled as described previously [102, 107]. The cysteine positions for fluorophore attachment were chosen, since they are part of residue pairs that are used in the laboratory for smFRET studies of the maltose transporter.

Label efficiencies for single and double cysteine mutants were determined as described in section 3.1.

7.5.3 smFRET spectroscopy

ALEX experiments of phosphate binding protein (PBP) were carried out as described previously [113]. PBP was studied by diluting the labeled proteins to concentrations of ≈ 50 pM in 50 mM Tris-HCl pH 7.4, 50 mM KCl supplemented with the ligand phosphate as described in the text and figures.

E-histograms of double-labeled FRET species with LD555 and LD6555 were extracted by selecting $0.25 < S < 0.75$. E-histograms of the open state without ligand (apo) and closed state with saturation of the ligand (holo) were fitted as described in section 3.4.2.

7.6 Supplementary Notes

7.6.1 Supplementary Note 1: Database parameter evaluation

Data preprocessing. The 104 PDB files of the database and the comparison PDB files were downloaded from the protein databank and preprocessed to unify the data structures. Therefore, all hetero atom entries (HETATM), anisotropy entries (ANISOU), and connection entries (CONNECT), as well as all the meta-information (REMARK) were removed from the pdb-files [19, 381]. Chains of polymeric protein assemblies in crystals were deleted if these were bare crystallization artifacts and do not occur in natural environments. The conservation score was calculated for all 112 chains containing the labeled

residues (and the reference database) with the default settings (see Supplement Table 7.7.1) [352, 353, 382]. Failed conservation score calculations (e.g. if too few homologue structures are available) were ignored for further analysis.

PDB data processing. The pdb files are parsed and processed with the “Bio.PDB” module [383] of the “biopython” package [384].

Parameter extraction. 28 parameters were calculated or extracted from third party software and assigned to the four categories (i) solvent exposure, (ii) residue conservation, (iii) cysteine resemblance, and (iv) secondary structure.

#	Parameter name	Library / webserver	Extracted value	Data type
1	N_SE1_RSA_Wilke	DSSP[354]	Relative surface area with amino acid surface areas according to Wilke[385]	float
2	N_SE2_RSA_Sander	DSSP[354]	Relative surface area with amino acid surface areas according to Sander[386]	float
3	N_SE3_RSA_Miller	DSSP[354]	Relative surface area with amino acid surface areas according to Miller[387]	float
4	I_SE4_HSE1_10A	Bio.PDB[383], HSE[355]	Number of atoms in half-sphere 1 within 10 Å	integer
5	I_SE5_HSE2_10A	Bio.PDB[383], HSE[355]	Number of atoms in half-sphere 2 within 10 Å	integer
6	I_SE7_HSE1_13A	Bio.PDB[383], HSE[355]	Number of atoms in half-sphere 1 within 13 Å	integer
7	I_SE8_HSE2_13A	Bio.PDB[383], HSE[355]	Number of atoms in half-sphere 2 within 13 Å	integer
8	I_SE10_HSE1_16A	Bio.PDB[383], HSE[355]	Number of atoms in half-sphere 1 within 16 Å	integer
9	I_SE11_HSE2_16A	Bio.PDB[383], HSE[355]	Number of atoms in half-sphere 2 within 16 Å	integer
10	N_SE13_CB_SURFACE_DIST	Bio.PDB[383], MSMS[356]	Distance of the C-beta atom to the protein surface	float
11	N_SE14_MEAN_SURFACE_DIST	Bio.PDB[383], MSMS[356]	Mean distance of all atoms to the protein surface	float

Table 7.6.1: Solvent exposure related values were extracted using the third party algorithms (i) “Define Secondary Structure of Proteins” (DSSP) to calculate relative surface accessibility[354], (ii) “Half-Sphere-Exposure” (HSE) to calculate the number of C-alpha atoms in the half-spheres defined by the $C_\alpha - C_\beta$ vector[355], and (iii) “Michel Sanner’s Molecular Surface” (MSMS) to calculate the protein surface and the residue depth of the atoms[356].

#	Parameter name	Library / webserver	Extracted value	Data type
12	I_CS1_Color	ConSurf[352, 353]	Color representation of conservation score (binned conservation score with upper and lower boundary)	integer
13	N_CS2_Score	ConSurf[352, 353]	Conservation score	float
14	N_CS3_Lower_Score	ConSurf[352, 353]	Lower value of confidence interval of conservation score	float
15	N_CS4_Upper_Score	ConSurf[352, 353]	Upper value of confidence interval of conservation score	float
16	I_CS5_Variety_Length	ConSurf[352, 353]	Number of different amino acids in homologues	integer
17	C_CS6_Cys_In_Variety	ConSurf[352, 353]	Boolean: true if cysteine is in amino acid list of homologues; false else	categorical

Table 7.6.2: Parameters related to residue conservation are extracted from the grades-file of the consurf server [352, 353]. Settings for the ConSurf analysis are listed in Table 7.7.1.

#	Parameter name	Library / webserver	Extracted value	Data type
18	C_SS1_SS	DSSP[354]	Secondary structure	categorical
19	N_SS2_Phi	DSSP[354]	Backbone torsion angle (n-1)-n	float
20	N_SS3_Psi	DSSP[354]	Backbone torsion angle n-(n+1)	float
21	C_SS4_SS-1	DSSP[354]	Secondary structure of predecessor residue	categorical
22	C_SS4_SS-2	DSSP[354]	Secondary structure two positions before	categorical
23	C_SS4_SS+1	DSSP[354]	Secondary structure of successor residue	categorical
24	C_SS4_SS+2	DSSP[354]	Secondary structure two positions after	categorical

Table 7.6.3: Secondary structure related values were extracted using the third party algorithms “Define Secondary Structure of Proteins” (DSSP) to calculate the secondary structure of the residue of interest and its adjacent residues as well as the backbone torsion angles [354].

#	Parameter name	Library / webserver	Extracted value	Data type
25	C_CR1_Name	-	Amino acid (name)	categorical
26	N_CR2_Mass	-	Mass of amino acid [u]	float
27	C_CR3_Charge	-	Charge of amino acid in buffer solution at pH=7.4 [e]	categorical
28	I_CR4_N_Sidechain	-	Number of sidechain atoms (without H-atoms)	integer

Table 7.6.4: Cysteine resemblance related values were taken from the amino acids structures to either compare the individual amino acids or group the amino acids by charge and size/mass.

Overall, 43357 and 29898 residues from the database and reference dataset are considered in the calculations, respectively. Failed parameter calculations were ignored for further analysis. Therefore, the number of calculated values varies for the 28 parameters (failure rate <10% for all parameters in the database and reference database, see Supplementary Table 7.7.2 for exact numbers).

7.6.2 Supplementary Note 2: Förster resonance energy transfer

Förster radius calculation. Spectral information, quantum yield, and extinction coefficients are taken from the database <https://www.fpbases.org/spectra/> and provided with the labelizer-package for the most commonly used fluorophores. The Förster radius R_0 is given by

$$R_0 = \left(\frac{9 \ln(10)}{128 \pi^5 N_A} \frac{\kappa^2}{n} Q_D \frac{\int_0^\infty F_D(\lambda) \varepsilon_{A_{max}} \varepsilon_A(\lambda) \lambda^4 d\lambda}{\int_0^\infty F_D(\lambda) d\lambda} \right)^{\frac{1}{6}}, \quad (7.6.1)$$

whereby Q_D is the donor quantum yield, F_D the normalized donor emission spectrum, ε_A the normalized acceptor emission spectrum, and $\varepsilon_{A_{max}}$ the acceptor extinction coefficient.

The following values are set fix to theoretical values:

Orientation factor κ^2 : 2/3
 Averaged refractive index n : 1.33 in buffer / 1.4 for protein (ref. [170])

Distance screening. Center of mass of a sphere with radius R cut with a cone of angle α in the z-dimension (spherical sector):

$$\vec{d} = \frac{\int_0^{\pi-\alpha} d\theta \int_0^{2\pi} d\phi \int_0^R dr r^2 \sin(\theta) \cdot \vec{r}}{\int_0^{\pi-\alpha} d\theta \int_0^{2\pi} d\phi \int_0^R dr r^2 \sin(\theta)} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \frac{3}{8} R (1 - \cos(\alpha)). \quad (7.6.2)$$

We assume the origin to be at the attachment site (C_β atom) of the fluorophore and model the accessible volume of the dye with a cut sphere of angle α and the non-accessible

space with $\pi - \alpha$. We approximate the center of mass of the non-accessible volume \vec{d}' with the center of mass of all N atom positions \vec{r}_i within the protein closer than R to the attachment point (see Figure 7.2.4B and Supplementary Figure 7.7.7A):

$$\vec{d}' = \frac{1}{N} \sum_i \vec{r}_i \quad \forall i \text{ with } |\vec{r}_i| < R. \quad (7.6.3)$$

The direct relation between the center of mass of atoms in the protein \vec{d}' and the center of mass of the fluorophores accessible-volume \vec{d} is given by:

$$\vec{d} = \left(\frac{\vec{d}'}{R} - \frac{3}{4} \frac{\vec{d}'}{|\vec{d}'|} \right) R. \quad (7.6.4)$$

We add an empirically determined correction factor based on simulations with 35 different fluorophore parameters on 100 residue pairs in 10 different protein structures to account for the finite size of atoms and fluorophores, which leads to a gap between protein atoms and accessible volume (see Supplementary Figure 7.7.7B).

The offset for the protein surface is added as a small addition to the fluorophore linker

$$\vec{d} = \left(\frac{\vec{d}'}{R} - \frac{3}{4} \frac{\vec{d}'}{|\vec{d}'|} \right) (R + \varepsilon) \quad (7.6.5)$$

and reads as

$$\varepsilon = \max(R_A, 2 \cdot \min(R_1, R_2, R_3) - R_A) + 0.014 \cdot R - 0.0059 \cdot R^2. \quad (7.6.6)$$

This correction can reproduce the simulated mean positions \vec{R}_{MP} with a root mean square deviation of $\pm 2.7 \text{ \AA}$ (see Supplementary Figure 7.7.7F) and mean position distances $R_{MP} = |\vec{R}_{MP,1} - \vec{R}_{MP,2}|$ with a root mean square deviation of $\pm 2.1 \text{ \AA}$ (see Supplementary Figure 7.7.8C).

We approximated the relation between the mean positions of the accessible volumes to the measured FRET averaged distances as

$$R_{\langle E \rangle} = R_{MP} + A e^{-b R_{MP}}, \quad (7.6.7)$$

whereby the second term accounts for FRET-efficiency weighted averaging effects at small distances (see Supplementary Figure 7.7.8B). The values A and b are determined as $A = 20.6 \text{ \AA}$ and $b = 0.037 \text{ 1/\AA}$ from a fit to the 35000 simulated distances within the ten selected protein structures, which is similar to the reported relations in ref. [283] and [111] for DNA. With this relationship, the simulated distance with FPS is reproduced up to a deviation of $\pm 3.4 \text{ \AA}$ ($\pm 3.1 \text{ \AA}$ for distances between 40 and 75 \AA).

Based on the corrected FRET values, the (screening) FRET-efficiency is calculated as

$$E_{\text{screen}} = \frac{1}{1 + \left(\frac{R_{\langle E \rangle}}{R_0} \right)^6}. \quad (7.6.8)$$

Distance refinement. A refinement is calculated based on the N highest FRET scores (with default N=300) using the available FPS simulation software [206].

7.7 Supplementary Information

Parameter	Value
DNA_AA	AA
NMR	no
PDB_yes_no	yes
MSA_yes_no	no
Homolog_search_algorithm	HMMER
ITERATIONS	1
E_VALUE	0.0001
proteins_DB	UNIREF90
user_select_seq	no
MAX_NUM_HOMOL	150
best_uniform_sequences	uniform
MAX_REDUNDANCY	95
MIN_IDENTITY	35
MSAprogram	MAFFT
ALGORITHM	Bayes
SUB_MATRIX	BEST

Table 7.7.1: ConSurf-server settings. Overview of all user parameters set for the conservations score calculation on <https://consurf.tau.ac.il/> (accessed January 24th, 2021).

#	Name	Number of all analyzed residues	Number of labeled residues	MSD	Gini coeff.	Shannon entropy
1	N_SE1_RSA_Wilke	40056	385	0.951	0.303	0.931
2	N_SE2_RSA_Sander	40056	385	0.866	0.323	0.923
3	N_SE3_RSA_Miller	40056	385	0.948	0.330	0.922
4	I_SE4_HSE1_10A	40056	385	0.674	0.663	0.687
5	I_SE5_HSE2_10A	40056	385	0.259	0.421	0.851
6	I_SE6_HSE1_13A	40056	385	0.922	0.625	0.712
7	I_SE7_HSE2_13A	40056	385	0.266	0.385	0.861
8	I_SE8_HSE1_16A	40056	385	0.972	0.596	0.744
9	I_SE9_HSE2_16A	40056	385	0.206	0.347	0.895
10	N_SE10_CB_SURFACE_DIST	40056	385	0.210	0.585	0.721
11	N_SE11_MEAN_SURFACE_DIST	40056	385	0.883	0.578	0.744
12	I_CS1_Color	39409	376	0.322	0.271	0.940
13	N_CS2_Score	39409	376	0.856	0.339	0.914
14	N_CS3_Lower_Score	39409	376	0.869	0.469	0.819
15	N_CS4_Upper_Score	39409	376	0.530	0.343	0.900
16	I_CS5_Variety_Length	39409	376	0.677	0.354	0.911
17	C_CS6_Cys_In_Variety	39409	376	0.006	0.038	0.996
18	C_SS1_SS	41502	390	0.218	0.307	0.894
19	N_SS2_Phi	41502	390	0.448	0.370	0.869
20	N_SS3_Psi	41502	390	0.231	0.339	0.883
21	C_SS4_SS-1	41227	385	0.258	0.383	0.861
22	C_SS5_SS-2	41018	382	0.175	0.250	0.919
23	C_SS6_SS+1	41227	388	0.218	0.316	0.900
24	C_SS7_SS+2	41021	385	0.207	0.259	0.917
25	C_CR1_Name	43357	396	0.753	0.401	0.909
26	N_CR2_Mass	43357	396	0.361	0.367	0.905
27	C_CR3_Charge	43357	396	0.156	0.162	0.956
28	I_CR4_N_Sidechain	43357	396	0.413	0.368	0.902

Table 7.7.2: Parameter overview and statistics. The table summarizes the number of analyzed residues (complete database, labeled residues) and the statistical evaluation of the parameter scores with respect to mean-square deviation (MSD), gini coefficient, and adapted Shannon entropy (see methods for details).

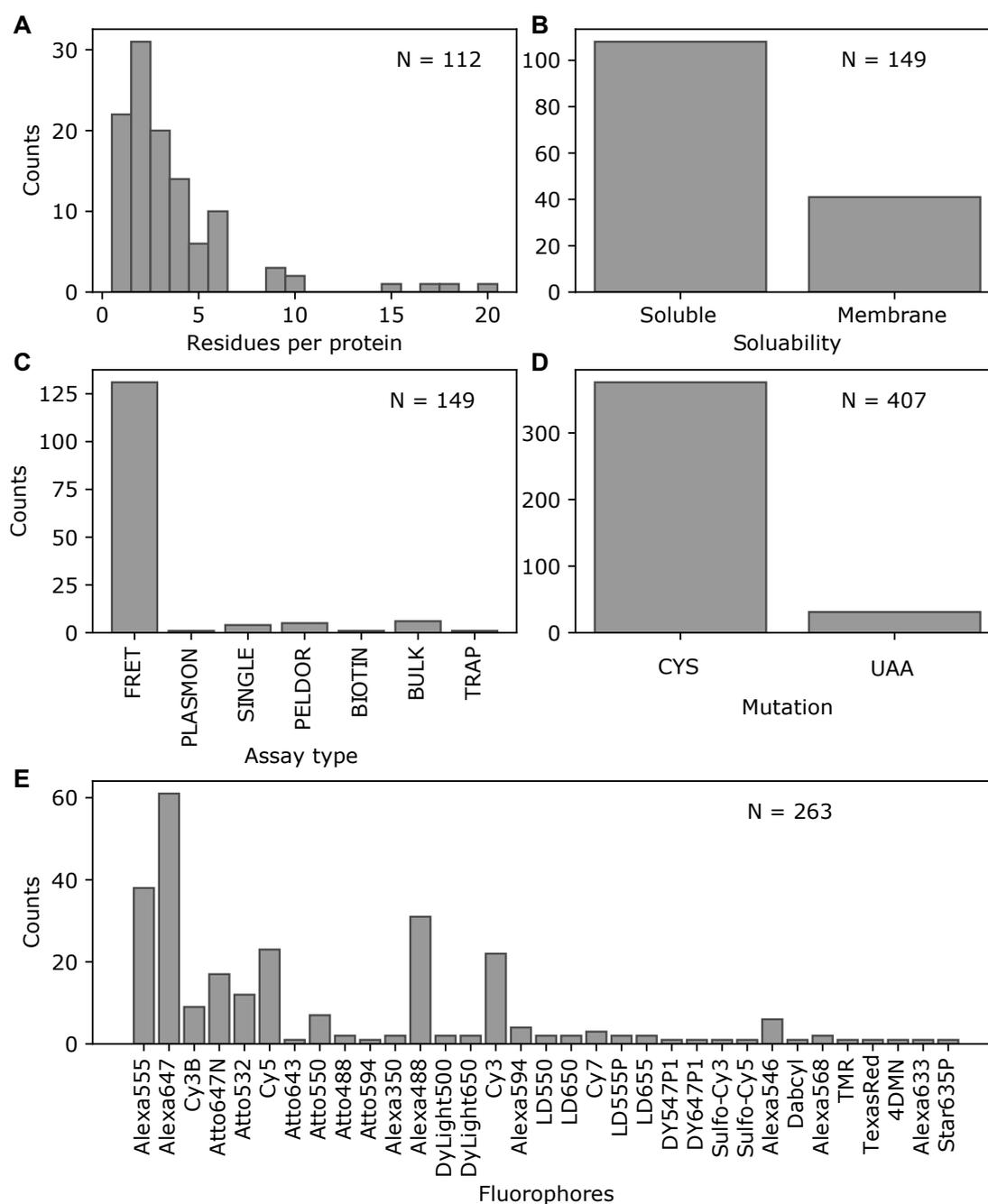


Figure 7.7.1: Labeling database statistics. **A** Number of labeled residues per chain in the database with $N=112$ different protein chains. **B** Comparison of published protein systems with soluble and membrane proteins ($N=149$ published protein systems, multiple occurrence possible). **C** Statistics of the different assay types used for the labeling database ($N=149$ published protein systems, multiple occurrence possible). Around 90% of the assays are single-molecule FRET assays (FRET), the others are bulk FRET (BULK) or single fluorophore labeled (SINGLE) assays, spin labels (PELDOR), gold labels (PLASMON), biotin labels (BIOTIN) and linker labels for optical traps (TRAP). **D** Statistics on the labeling residues (cysteine or unnatural amino acid, $N=407$ residues). **E** Statistics on the fluorophores used in the publications ($N=263$ occurrences in the publications).

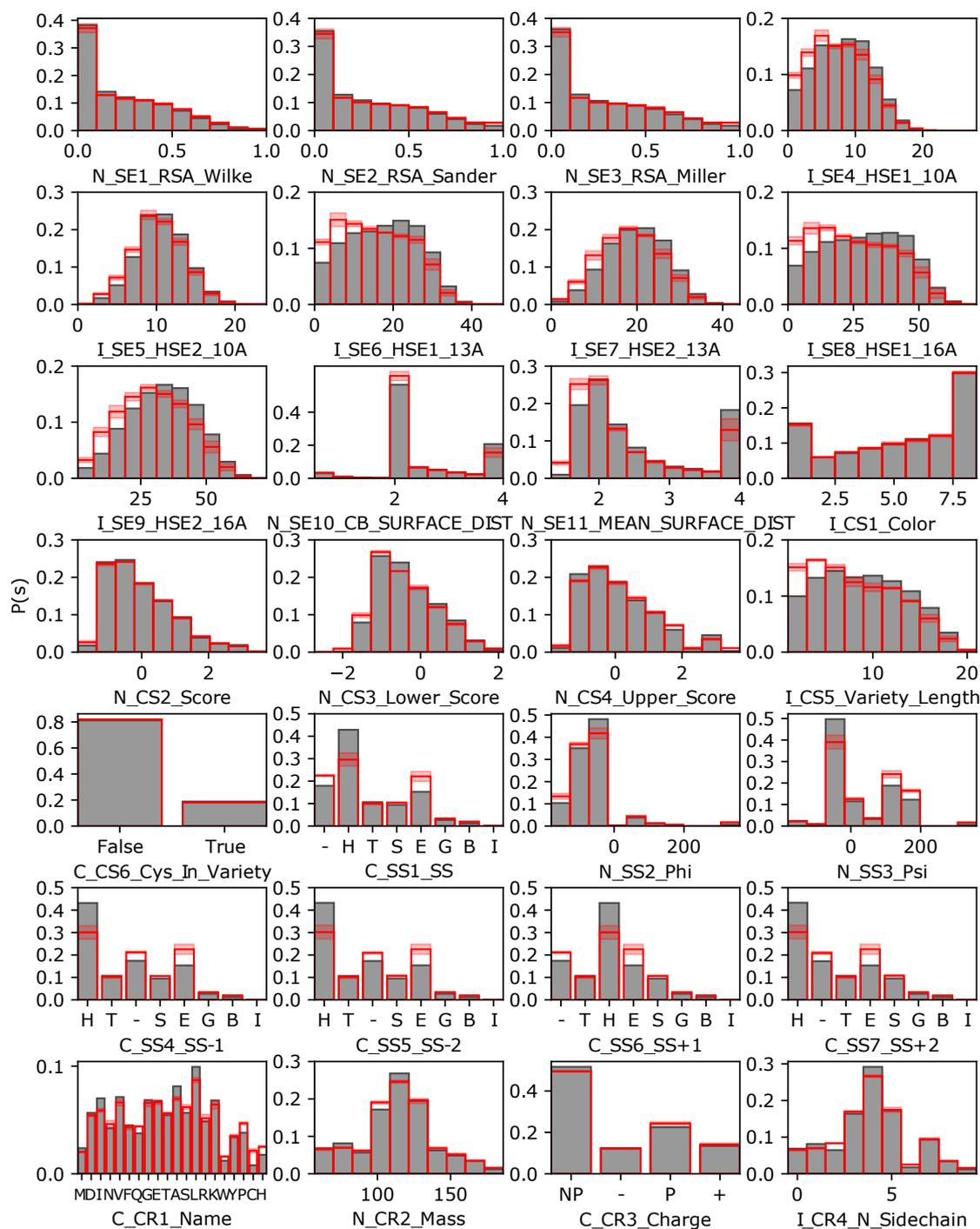


Figure 7.7.2: Parameter distribution comparison Probability distributions $P(s)$ of all 28 scores in the labeling database (gray). The values are compared to a randomly selected representative reference dataset (red line: mean values, pale area: standard deviation of triplicates) based on the pdbselect dataset [357, 358] (see methods for details). The x-axis label specifies the parameter: first part for the type of data (I: integer, N: numeric, C: categorical), second part for the parameter group (SE: solvent exposure, CS: conservation score, SS: secondary structure, CR: cysteine resemblance), and the rest for a reasonable name (see Table 7.7.2 for details).

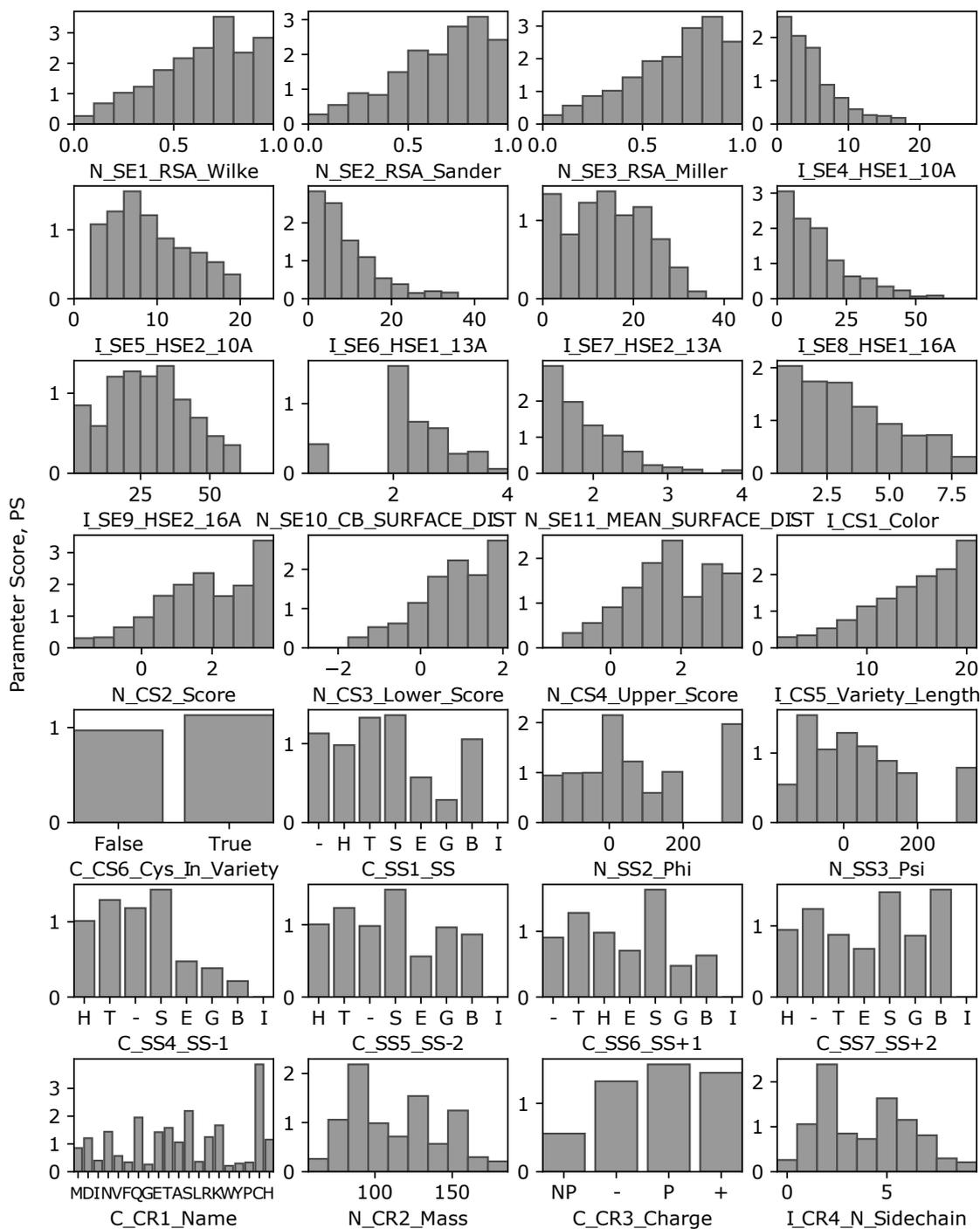


Figure 7.7.3: Conditional frequencies of parameters. Conditional frequency distributions $PS = P(sl)/P(s)$ defining the parameter scores of all 28 parameters in the labeling database. The x-axis label specifies the parameter: first part for the type of data (I: integer, N: numeric, C: categorical), second part for the parameter group (SE: solvent exposure, CS: conservation score, SS: secondary structure, CR: cysteine resemblance), and the rest for a reasonable name (see Table 7.7.2 for details).

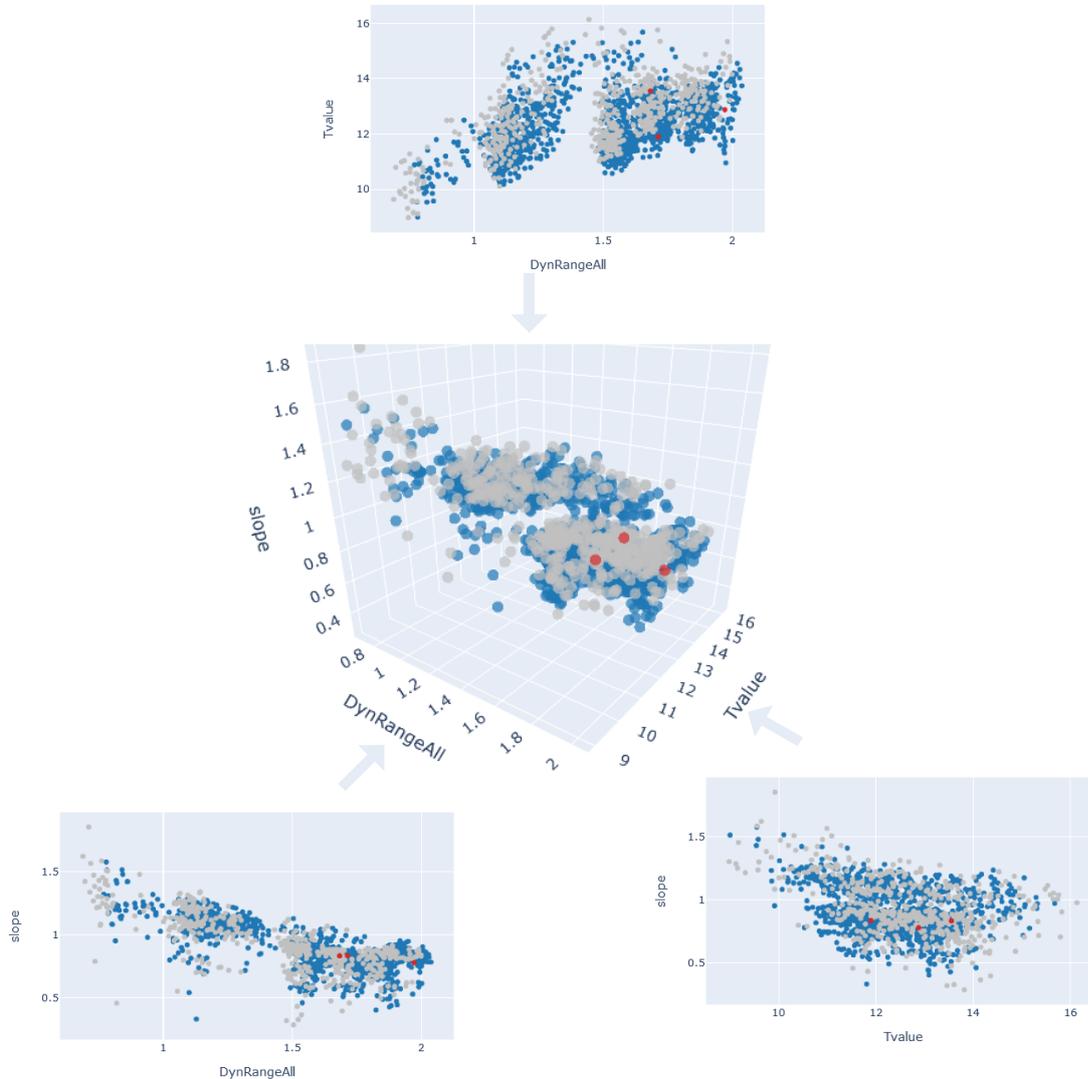


Figure 7.7.4: Assessment of parameter sets. All possible parameter sets (combination of each parameter in a group with all combinations from other groups) was assessed based on the dynamic range of the label scores, the statistical significance and the suppression/enhancement of the label scores (see parameter selection criteria in methods). The marked positions are the favorite parameter sets (SE11, CS6, SS1, CR2) (default settings), (SE1, CS6, SS1, CR2), and (SE4, CS6, SS1, CR2) (see Table 7.7.2).

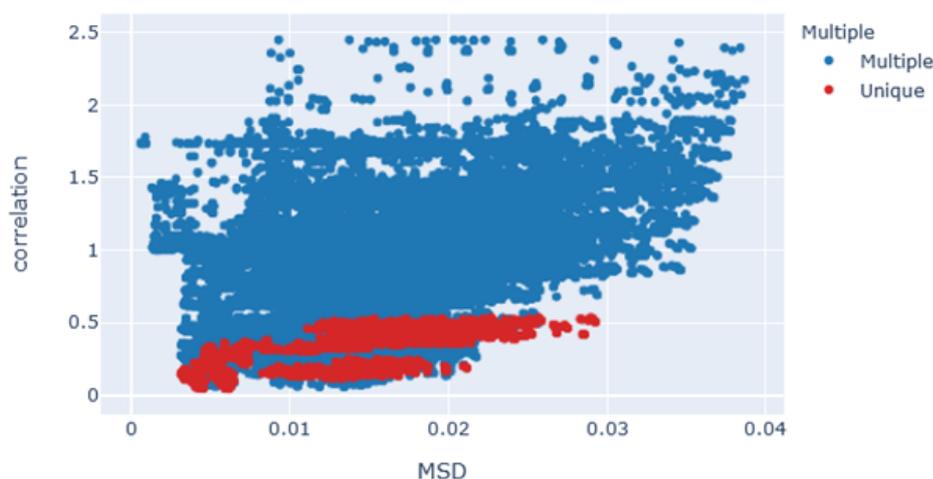


Figure 7.7.5: Correlation measure and averaged mean square deviation (MSD).

The plot shows the geometric mean of the MSD value of the selected parameters (mean square deviation from equal contribution, see methods) versus the correlation measure (2-norm of all correlations) for a parameter set. All parameters are combined with each other to sets of 4, whereby points with multiple (2 or more) parameters from the same group (e.g. solvent exposure) are marked blue. Combinations with one parameter from each group are marked red.

LabelLib.dyeDensityAV3

Parameter	Value
discStep	0.8 (coarse-grained: 1.2)
linkerLength	varied
linkerDiameter	4.5
dyeRadii	varied

LabelLib. meanEfficiency

Parameter	Value
R0	57.5
nsamples	100000 (coarse-grained: 10000)

Table 7.7.3: FPS settings. Overview of all user parameters set for the runtime analysis of the distance refinement simulation with the FPS software [206]. The default settings for the labelizer package and webserver use discStep = 0.8 and nsamples = 100000 (all other parameters depend on the selected fluorophore pair).

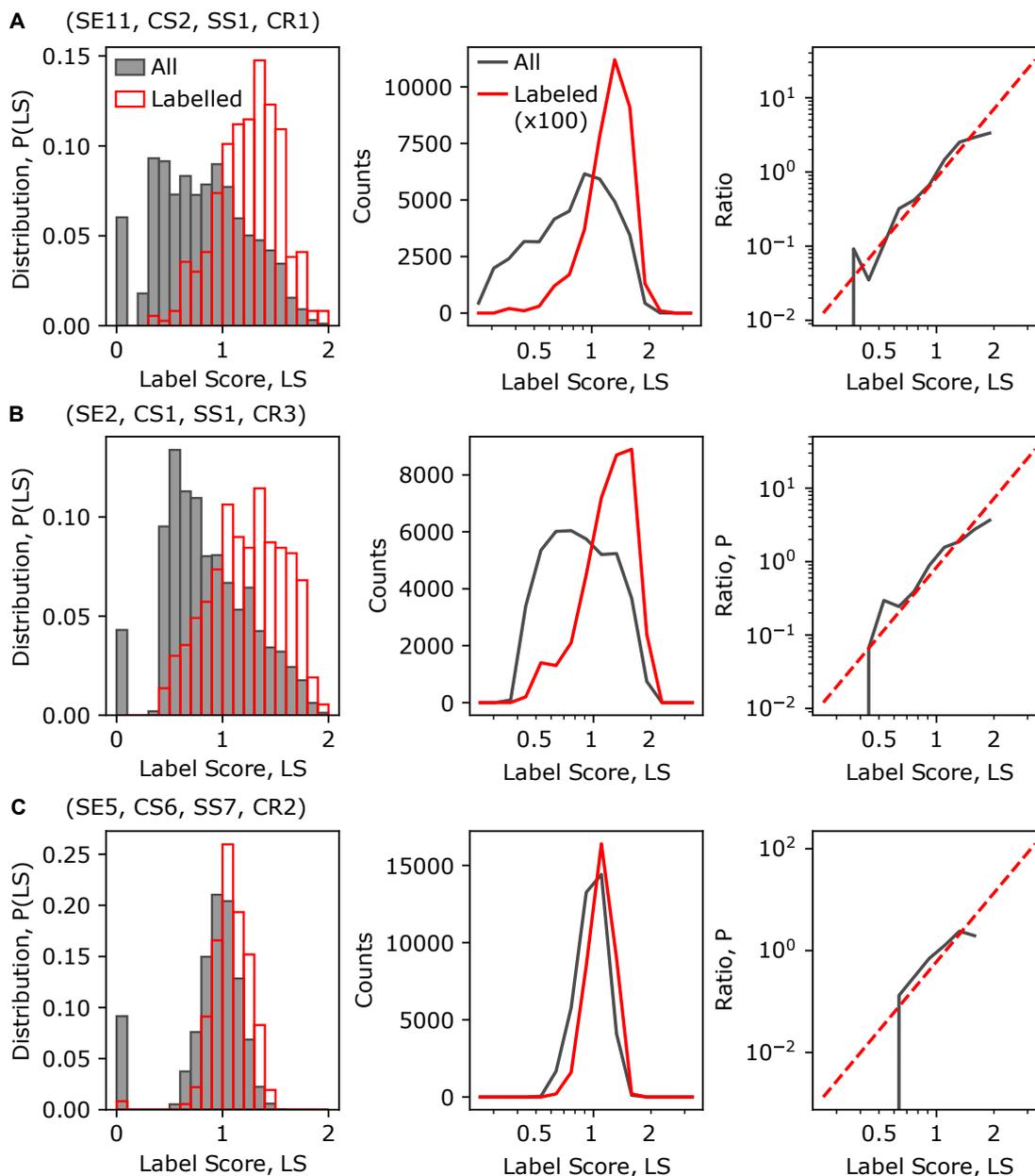


Figure 7.7.6: Label score evaluation for different parameter sets. **A** Label score probability distribution of all residues (gray) and labeled residues (red) in our database (left) and the a histogram with logarithmic scale of the label scores (middle) for the selected quadruple (SE11, CS6, SS1, CR2) (default settings, see Table 7.7.2). The ratio of the probability distribution of labeled and all residues (gray) is fitted with a linear dependency (red, dashed) in the log-log-plot (right). **B** Same evaluation as in **A** for another suitable parameter selection of the quadruple (SE2,CS1, SS1, CR3). **C** Same evaluation as in **A** for a parameter set with poor prediction power (SE5, CS6, SS7, CR2).

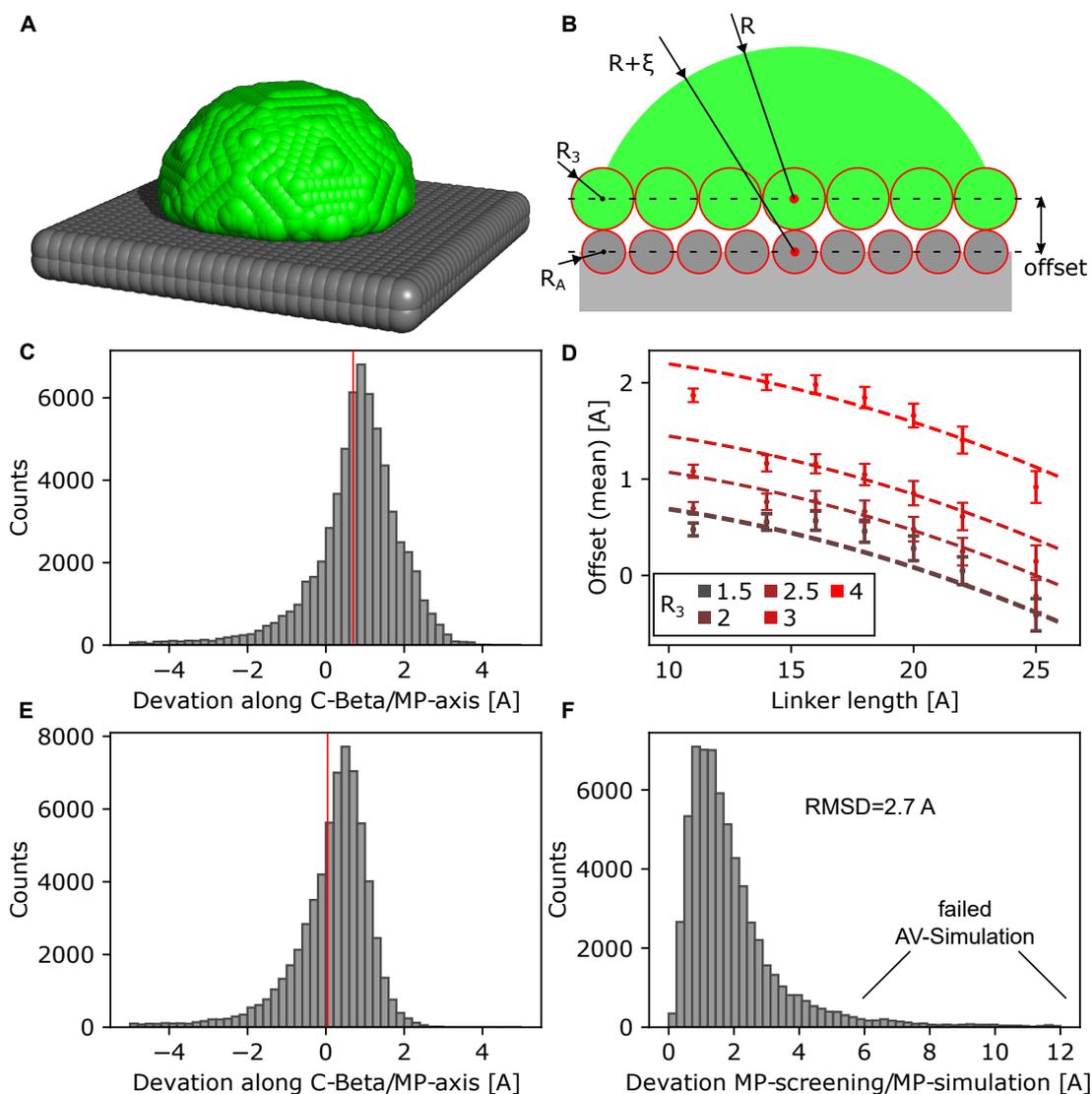


Figure 7.7.7: Correction parameter for mean position of accessible volume.

A Simplified accessible volume (AV) simulation (green) in an idealized system of a planar array of atoms (gray). **B** Correction factor ϵ considers the offset between the atom coordinates (lower dashed line) and the AV coordinates (upper dashed line). ϵ corrects for this gap between accessible surface (green) and inaccessible surface (gray) under the consideration of the linker length (R , corresponds to the AV radius), the atom radius R_A and the smallest fluorophore radius R_3 of the ellipsoidal approximation [113, 206, 283]. **C** Deviation between simulated mean position of AV (FPS software [206]) and estimated mean position (SSM approach) with indicated mean value (red line). **D** Mean offset from C for different linker lengths R and fluorophore radii R_3 is shown with errorbars (standard error of the mean from simulations). The estimation of the offset in C is fitted globally with the correction factor $\epsilon = \max(R_A, 2 \min(R_1, R_2, R_3) - R_A) + 0.014 \cdot R - 0.0059 \cdot R^2$ (dashed lines). **E** Deviation between simulated mean position of accessible volume (FPS software) and estimated mean position (SSM approach) including the correction factor ϵ (mean value: red line). **F** Distance between corrected SSM approach and AV mean position from FPS results in a deviation of 2.7 Å (RMSD). The large deviations (>6 Å) result from failed FPS runs (due to interfering atoms close to the linker attachment site).

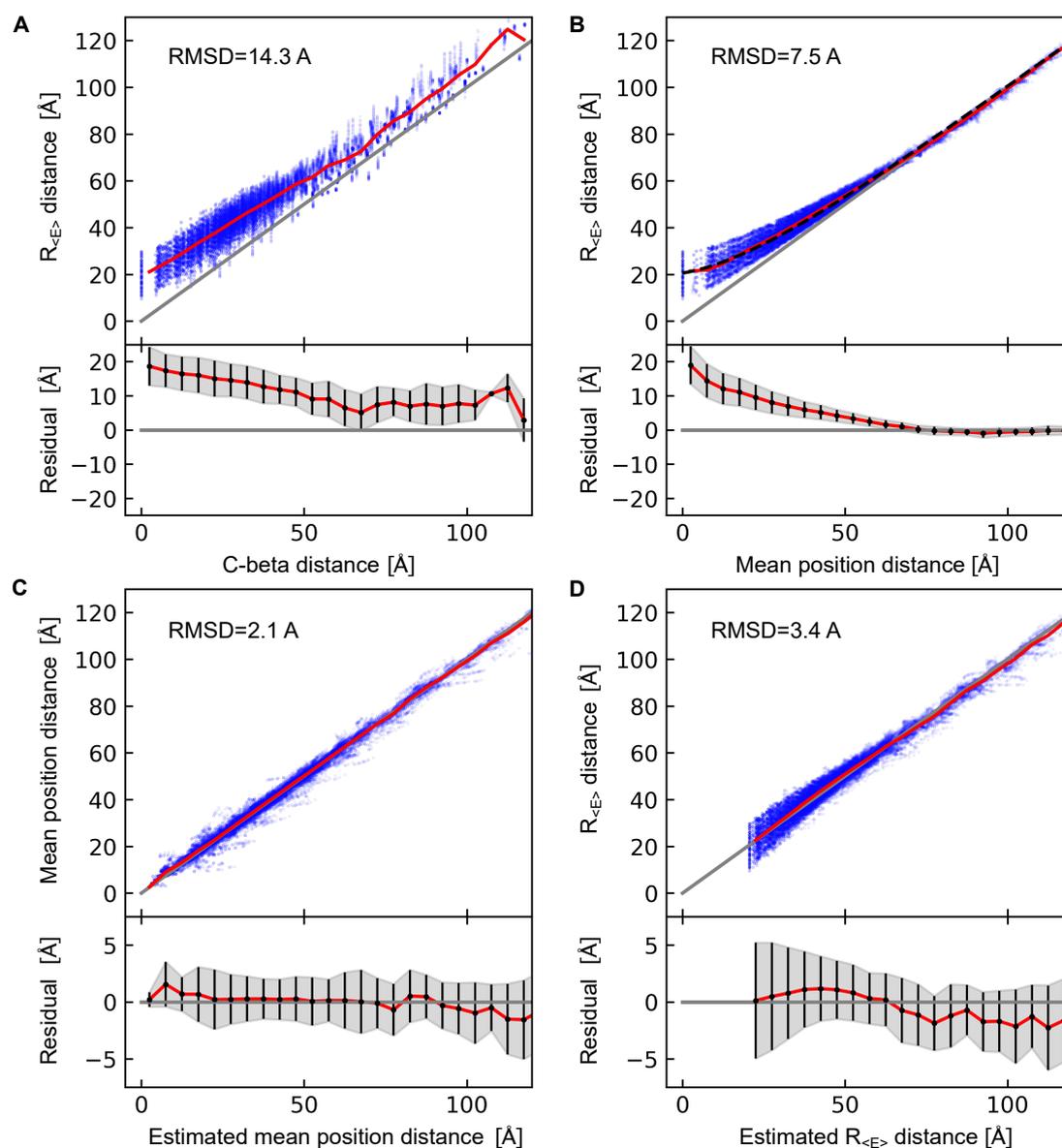


Figure 7.7.8: Correction parameter for distance simulation. **A** C-beta distances are plotted against simulated distances with FPS-software [206] (blue datapoints) with mean values (red line). The bottom axis shows the mean residual (red line) and the standard deviation interval (error bars / gray area) on binned data from the top. **B** Mean dye position distances R_{MP} (center of mass of the AV-simulation) are plotted against the FRET-averaged distances (blue datapoints) with mean values (red line). The mean values are fitted to the curve $R_{MP} + Ae^{-bR_{MP}}$ with $A=20.6$ Å and $b=0.037$ $1/\text{Å}$ (black dashed line). **C** Mean dye position estimations based on the spherical sector calculation (SSM) approach are plotted against mean dye position distances R_{MP} from FPS-simulation software. **D** Mean dye position distances from the SSM-estimation are converted to FRET-averaged distances with the correction factors from B and plotted against the simulated $R_{<E>}$ from FPS-simulation.

PDB	Molar mass [u]	SSM time per DA-pair [ms]	Coarse-grained FPS time per DA-pair [ms]	FPS time per DA-pair [ms]
3L6G	28,830	0.81±0.04	103±5	315±11
2KHO	65,650	0.84±0.01	166±5	419±8
2CG9	188,730	0.94±0.01	303±12	647±10
4B1O	831,160	4.31±0.09	1113±36	2087±46
172L	18,730	0.67±0.01	77±2	249±10
2A65	59,750	0.93±0.04	169±5	474±9
1WDN	25,130	0.78±0.01	100±3	315±4
5XPD	33,310	0.67±0.02	94±3	282±10
1P7B	74,510	0.80±0.01	156±12	398±14
1HKA	17,970	0.81±0.04	81±2	259±11

Table 7.7.4: Spherical sector vs. FPS runtime comparison. Calculation time overview of the fast screening method (spherical sector calculation) and a coarse-grained distance refinement simulation (FPS software [206]) with 3500 distance pairs per pdb-file (100 distances, 35 different dye parameter) and a refined FPS simulation with 1400 distance pairs per pdb-file (40 distances, 35 different dye parameter).

8 | Conclusion

Advances in microscopy and spectroscopy techniques [24, 388] and the possibility of site-specific covalent attachment of performant organic fluorophores to biomolecules [21, 22] enabled the broad usage of these highly sensitive biophysical applications in the life sciences but also in biomedical assays to investigate structure, interactions, and dynamics of macromolecules. Especially, smFRET measurements have become a standard tool to obtain structural, conformational, or kinetic information on the biological system under physiological conditions [12].

In this work, I contributed to further develop the technique of smFRET towards a reliable and systematic tool for the study of biochemical processes. In symmetrically structured, homogeneously charged, and mostly static DNA structures, smFRET has proven to be an accurate measurement tool [111]. However in protein systems, the standard evaluation of single-molecule burst analysis experiments is often limited by environmental variations in fluorophore lifetime, anisotropy, mobility, electrostatic interaction with the protein surface, as well as unwanted side-effects to the protein such as destabilization or steric interference. This work has treated multiple aspects of the current challenges by a systematic investigation of the interplay of the named aspects, thereby facilitating the overall experimental workflow and data analysis routines.

An in-depth knowledge of the chemical structure of the fluorophores and their fluorescent properties is essential for high-quality smFRET measurements. The first part of the thesis addresses this aspect by a molecular analysis of the cyanine derivatives AF555, AF647, Alexa Fluor 555, and Alexa Fluor 647 with fluorescent, mass, and NMR spectroscopy complemented by molecular dynamics simulations. Further, I have shown that seemingly minor changes in the chemical structure of a fluorophore, e.g. the attachment site of negatively charged sulfo-groups or the length of their linker, can have a significant influence on the mobility of a fluorophore. I could observe a huge impact on the quantum efficiency and lifetime for the investigated environmentally sensitive cyanine derivatives, which highly influences the accuracy of FRET-derived distances.

The second result chapter showed how to circumvent methodological weaknesses by the use of the two complementary methods smFRET and EPR spectroscopy. It could be shown that both methods share a reasonable agreement on the determined distances with an average ± 5 Å spread between them. This was shown to be partly attributed to the distinct labels with different linker lengths leading to a systematic offset. Additional differences due to label-protein interactions were identified and investigated in detail. The results show that the selection of labeling site is important for both methods. Further, the choice of hydrophobic or hydrophilic fluorophores in FRET or the addition of cryoprotectant in EPR can highly influence the data quality and the accuracy of the

measurements. We found that a combination of PELDOR/DEER and smFRET reduces the measurement errors and provides highly complementary and synergistic insights into different aspects of conformational states of macromolecules.

The multi-laboratory benchmark study in this thesis addresses another key challenge in smFRET experiments: the disentanglement of measurement errors, variation in the correction and evaluation procedure and sample-intrinsic deviations from modeled distances. I provided the same protein sample with predefined measurement instructions to 19 laboratories to investigate two protein systems exhibiting different types of dynamic behavior. The collected and reanalyzed results showed deviations in FRET efficiency measurements similar to what was determined using DNA standards, which could be further improved by a standardized analysis routine. The outcome of the study is that a robust determination of the detection-correction factor is indispensable for the calculation of setup-independent accurate FRET efficiencies. It was demonstrated that smFRET allows to detect and characterize conformational dynamics in proteins, which can be disentangled from dye quenching, blinking, photobleaching, and sticking. The presented results underline the importance of reliable measurement routines and powerful methods to predict or prevent label-protein interactions, destabilization of the protein, and a mutant selection in the sensitive distance range of the donor-acceptor pair.

The last chapter present a systematic approach to combine the literature data on protein labeling of the last 15 years of smFRET measurements with general parameters derived from databases to score and predict optimal labeling sites using basic concepts of probability theory. I developed a standalone python package and a webserver with a user-friendly interface for straightforward use to enable all researchers in the field of single-molecule spectroscopy to apply the selection algorithms. The theoretical work was validated with a broad screening of label efficiency, lifetime, and anisotropy measurements of single cysteine mutants and designed various FRET mutants for the scope of this thesis as well as for the research of other scientists in my research group. The work represents the first step towards an open-access database of fluorescent protein labeling for single-molecule spectroscopy assays to continuously improve the prediction and enable other researchers to access and reanalyze the edited and unified database.

In conclusion, the analysis of the fluorophore environment on the protein surface and the investigation of charge-dependent fluorophore-protein interaction increases the control of the central photophysical parameters. Comparison and benchmark smFRET measurements show high reproducibility and accuracy under favorable conditions and deviations can be attributed to non-ideal labeling sites or fluorophore selections. Though, the combination of complementary measurement techniques with the aid of computational methods provides a promising platform to refine experimental assays and minimize measurement artifacts, which opens a wide field of biophysical applications of smFRET for structural modeling and dynamic analysis under physiological conditions. Considering recent advances in computational biophysics, particularly the continuous improvement of molecular dynamics simulations, it is even desirable to use coarse-grained MD simulations for assay design purposes and refine the prediction power of the *labelizer* framework. One could envision to move from experiment-driven MD simulations (attempt of simu-

lating the measured results) to simulation-driven experiments (predicting and avoiding fluorescent signal disturbance and protein malfunction) for cleaner measurement results and clearer findings. This way, the presented results could be an important building block for next-generation single-molecule experiments opening new research opportunities for quantitative observations of complex, fundamental biochemical processes such as drug delivery, DNA maintenance and repair, transcription, translation, membrane transport, and molecular diagnostics.

Nomenclature

Abbreviations

ALEX	alternating-laser excitation microscopy
APBS	all photon burst search
BVA	burst variance analysis
DCBS	dual channel burst search
EPR	electron paramagnetic resonance
FRET	Förster resonance energy transfer
HOMO	highest occupied molecular orbital
LUMO	lowest unoccupied molecular orbital
MD	molecular dynamics
MFD	multifluorescent detection
PALM	photo-activated localization microscopy
PDA	photon distribution analysis
PELDOR	pulsed electron-electron double resonance
PIE	pulsed-interleaved excitation microscopy
PIFE	protein induced fluorescence enhancement
PSF	point-spread function
smFRET	single-molecule FRET
STED	stimulated emission depletion microscopy
STORM	stochastic optical reconstruction microscopy
TCSPC	time-correlated single photon counting

Bibliography

- [1] G. G. Stokes, “On the Change of Refrangibility of Light”, *Philosophical Transactions of the Royal Society of London* **142**, 463–562 (1852).
- [2] J. R. Lakowicz, “Principles of Fluorescence Spectroscopy”, in (Springer US, Boston, MA, 2006), pp. 205–235.
- [3] F. J. Duarte and L. W. Hillman, *Dye Laser Principles* (Elsevier Science, 1990), p. 472.
- [4] B. Liphardt, B. Liphardt and W. Lüttke, “Laser dyes with intramolecular triplet quenching”, *Optics Communications* **38**, 207–210 (1981).
- [5] D. B. Schmolze, C. Standley, K. E. Fogarty and A. H. Fischer, “Advances in Microscopy Techniques”, *Archives of Pathology & Laboratory Medicine* **135**, 255–263 (2011).
- [6] P. Tinnefeld, C. Eggeling and S. W. Hell, *Far-Field Optical Nanoscopy* (Springer Berlin Heidelberg, 2015), p. 335.
- [7] J. H. M. van der Velde et al., “A simple and versatile design concept for fluorophore derivatives with intramolecular photostabilization”, *Nature Communications* **7**, 10144 (2016).
- [8] A. A. Jazi, E. Ploetz, M. Arizki, B. Dhandayuthapani, I. Waclawska, R. Krämer, C. Ziegler and T. Cordes, “Caging and Photoactivation in Single-Molecule Förster Resonance Energy Transfer Experiments”, *Biochemistry* **56**, 2031–2041 (2017).
- [9] L. Möckl, D. C. Lamb and C. Bräuchle, “Super-resolved fluorescence microscopy: Nobel Prize in Chemistry 2014 for Eric Betzig, Stefan Hell, and William E. Moerner.”, *Angewandte Chemie (International ed. in English)* **53**, 13972–13977 (2014).
- [10] T. Foerster, “Zwischenmolekulare Energiewanderung und Fluoreszenz”, *Annalen der Physik* **437**, 55–75 (1948).
- [11] L. Stryer and R. P. Haugland, “Energy transfer: a spectroscopic ruler.”, *Proceedings of the National Academy of Sciences of the United States of America* **58**, 719–726 (1967).
- [12] E. Lerner, T. Cordes, A. Ingargiola, Y. Alhadid, S. Chung, X. Michalet and S. Weiss, “Toward dynamic structural biology: Two decades of single-molecule Förster resonance energy transfer”, *Science (New York, N.Y.)* **359**, 6373 (2018).

- [13] T. Ha, T. Enderle, D. F. Ogletree, D. S. Chemla, P. R. Selvin and S. Weiss, "Probing the interaction between two single molecules: fluorescence resonance energy transfer between a single donor and a single acceptor.", *Proceedings of the National Academy of Sciences of the United States of America* **93**, 6264–6268 (1996).
- [14] A. N. Kapanidis, N. K. Lee, T. A. Laurence, S. Doose, E. Margeat and S. Weiss, "Fluorescence-aided molecule sorting: Analysis of structure and interactions by alternating-laser excitation of single molecules", *Proceedings of the National Academy of Sciences of the United States of America* **101**, 8936–8941 (2004).
- [15] A. N. Kapanidis, T. A. Laurence, N. K. Lee, E. Margeat, X. Kong and S. Weiss, "Alternating-Laser Excitation of Single Molecules", *Accounts of Chemical Research* **38**, 523–533 (2005).
- [16] B. K. Müller, E. Zaychikov, C. Bräuchle and D. C. Lamb, "Pulsed interleaved excitation", *Biophysical journal* **89**, 3508–3522 (2005).
- [17] C. Eggeling, S. Berger, L. Brand, J. R. Fries, J. Schaffer, A. Volkmer and C. A. M. Seidel, "Data registration and selective single-molecule analysis using multi-parameter fluorescence detection", *Journal of Biotechnology* **86**, 163–180 (2001).
- [18] V. Kudryavtsev, M. Sikor, S. Kalinin, D. Mokranjac, C. A. M. Seidel and D. C. Lamb, "Combining MFD and PIE for Accurate Single-Pair Förster Resonance Energy Transfer Measurements", *ChemPhysChem* **13**, 1060–1078 (2012).
- [19] F. C. Bernstein, T. F. Koetzle, G. J. B. Williams, E. F. Meyer, M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi and M. Tasumi, "The protein data bank: A computer-based archival file for macromolecular structures", *Journal of Molecular Biology* **112**, 535–542 (1977).
- [20] 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* **28**, 235–242 (2000).
- [21] M. S. T. Gonçalves, "Fluorescent Labeling of Biomolecules with Organic Probes", *Chemical Reviews* **109**, 190–212 (2009).
- [22] C. P. Toseland, "Fluorescent labeling and modification of proteins", *Journal of chemical biology* **6**, 85–95 (2013).
- [23] M. E. Sanborn, B. K. Connolly, K. Gurunathan and M. Levitus, "Fluorescence Properties and Photophysics of the Sulfoindocyanine Cy3 Linked Covalently to DNA", *The Journal of Physical Chemistry B* **111**, 11064–11074 (2007).
- [24] M. Sauer, J. Hofkens and J. Enderlein, "Basic Principles of Fluorescence Spectroscopy", in *Handbook of fluorescence spectroscopy and imaging* (Wiley Online Library, 2011), pp. 1–30.
- [25] J. J. Sakurai and J. Napolitano, *Modern Quantum Mechanics*, 2nd ed. (Cambridge University Press, Cambridge, 2017).

-
- [26] F. Perrin, "Polarisation de la lumière de fluorescence. Vie moyenne des molécules dans l'état excité", *Journal de Physique et le Radium* **7**, 390–401 (1926).
- [27] G. Weber, "Rotational Brownian motion and polarization of the fluorescence of solutions.", *Advances in protein chemistry* **8**, 415–459 (1953).
- [28] M. D. Barkley, A. A. Kowalczyk and L. Brand, "Fluorescence decay studies of anisotropic rotations of small molecules", *The Journal of Chemical Physics* **75**, 3581–3593 (1981).
- [29] T. Cordes, A. Maiser, C. Steinhauer, L. Schermelleh and P. Tinnefeld, "Mechanisms and advancement of antifading agents for fluorescence microscopy and single-molecule spectroscopy", *Physical Chemistry Chemical Physics* **13**, 6699–6709 (2011).
- [30] J. Dominik Spiegel, S. Fulle, M. Kleinschmidt, H. Gohlke and C. M. Marian, "Failure of the IDA in FRET Systems at Close Inter-Dye Distances Is Moderated by Frequent Low κ^2 Values", *Journal of Physical Chemistry B* **120**, 8845–8862 (2016).
- [31] T. Cordes, J. Vogelsang and P. Tinnefeld, "On the Mechanism of Trolox as Antifading and Antibleaching Reagent", *Journal of the American Chemical Society* **131**, 5018–5019 (2009).
- [32] J. Vogelsang, R. Kasper, C. Steinhauer, B. Person, M. Heilemann, M. Sauer and P. Tinnefeld, "A Reducing and Oxidizing System Minimizes Photobleaching and Blinking of Fluorescent Dyes", *Angewandte Chemie International Edition* **47**, 5465–5469 (2008).
- [33] T. Ha and P. Tinnefeld, "Photophysics of Fluorescence Probes for Single Molecule Biophysics and Super-Resolution Imaging", *Annual review of physical chemistry* **63**, 595–617 (2012).
- [34] J. H. M. van der Velde, E. Ploetz, M. Hiermaier, J. Oelerich, J. W. de Vries, G. Roelfes and T. Cordes, "Mechanism of Intramolecular Photostabilization in Self-Healing Cyanine Fluorophores", *ChemPhysChem* **14**, 4084–4093 (2013).
- [35] J. H. M. van der Velde, J. J. Uusitalo, L.-J. Ugen, E. M. Warszawik, A. Herrmann, S. J. Marrink and T. Cordes, "Intramolecular photostabilization via triplet-state quenching: design principles to make organic fluorophores 'self-healing'", *Faraday Discuss.* **184**, 221–235 (2015).
- [36] M. Levitus and S. Ranjit, "Cyanine dyes in biophysical research: The photophysics of polymethine fluorescent dyes in biomolecular environments", *Quarterly Reviews of Biophysics* **44**, 123–151 (2011).
- [37] E. M. S. Stennett, M. A. Ciuba, S. Lin and M. Levitus, "Demystifying PIFE: The Photophysics behind the Protein-Induced Fluorescence Enhancement Phenomenon in Cy3", *Journal of Physical Chemistry Letters* **6**, 1819–1823 (2015).
- [38] H. Hwang and S. Myong, "Protein induced fluorescence enhancement (PIFE) for probing protein-nucleic acid interactions", *Chem. Soc. Rev.* **43**, 1221–1229 (2014).

- [39] N. Scholz, A. Jadhav, M. Shreykar, T. Behnke, N. Nirmalanathan, U. Resch-Genger and N. Sekar, “Coumarin-Rhodamine Hybrids – Novel Probes for the Optical Measurement of Viscosity and Polarity”, *Journal of Fluorescence* **27**, 1949–1956 (2017).
- [40] S. Ludwanowski, A. Samanta, S. Loescher, C. Barner-Kowollik and A. Walther, “A Modular Fluorescent Probe for Viscosity and Polarity Sensing in DNA Hybrid Messtructures”, *Advanced Science* **8**, 2003740 (2021).
- [41] N. Marmé, J. P. Knemeyer, M. Sauer and J. Wolfrum, “Inter- and Intramolecular Fluorescence Quenching of Organic Dyes by Tryptophan”, *Bioconjugate Chemistry* **14**, 1133–1139 (2003).
- [42] M. Möller and A. Denicola, “Protein tryptophan accessibility studied by fluorescence quenching”, *Biochemistry and Molecular Biology Education* **30**, 175–178 (2002).
- [43] A. W. Varnes, R. B. Dodson and E. L. Wehry, “Interactions of transition-metal ions with photoexcited states of flavines. Fluorescence quenching studies”, *Journal of the American Chemical Society* **94**, 946–950 (1972).
- [44] D. Dexter, “A Theory of Sensitized Luminescence in Solids”, *J. Chem. Phys.* **21**, 836–850 (1953).
- [45] S. Doose, H. Neuweiler and M. Sauer, “Fluorescence quenching by photoinduced electron transfer: A reporter for conformational dynamics of macromolecules”, *ChemPhysChem* **10**, 1389–1398 (2009).
- [46] R. M. Clegg, *Fluorescence Imaging Spectroscopy and Microscopy*, edited by X. F. Wang and B. Herman, *Chemical Analysis: A Series of Monographs on Analytical Chemistry and Its Applications* (John Wiley & Sons Inc., 1996), pp. 179–251.
- [47] H. Yang, “The Orientation Factor in Single-Molecule Förster-Type Resonance Energy Transfer, with Examples for Conformational Transitions in Proteins”, *Israel Journal of Chemistry* **49**, 313–321 (2009).
- [48] B. Corry and D. Jayatilaka, “Simulation of structure, orientation, and energy transfer between AlexaFluor molecules attached to MscL”, *Biophysical journal* **95**, 2711–2721 (2008).
- [49] C. J. R. Sheppard, M. Gu and M. Roy, “Signal-to-noise ratio in confocal microscope systems”, *Journal of Microscopy* **168**, 209–218 (1992).
- [50] M. Marvin, *Microscopy apparatus (patent US3013467)*, 1961.
- [51] L. Novotny and B. Hecht, “Chapter 3: Propagation and focusing of optical field”, in *Principles of nano-optics*, 2nd ed. (Cambridge University Press, Cambridge, 2012).
- [52] N. Naredi-Rainer, J. Prescher, A. Hartschuh and D. C. Lamb, “Chapter 5: Confocal Microscopy”, in *Biomedical imaging fluorescence applications in biotechnology and life sciences handbook of fluorescence spectroscopy and imaging*, edited by U. Kubitschek (WILEY-VCH Verlag, 2013), pp. 175–214.

-
- [53] T. Wilson et al., *Confocal microscopy* (Academic Press: New York, 1990), p. 426.
- [54] J. Enderlein, D. L. Robbins, W. P. Ambrose, P. M. Goodwin and R. A. Keller, “Statistics of Single-Molecule Detection”, *The Journal of Physical Chemistry B* **101**, 3626–3632 (1997).
- [55] J. Schaffer, A. Volkmer, C. Eggeling, V. Subramaniam, G. Striker and C. A. M. Seidel, “Identification of Single Molecules in Aqueous Solution by Time-Resolved Fluorescence Anisotropy”, *The Journal of Physical Chemistry A* **103**, 331–336 (1999).
- [56] C. Eggeling, J. R. Fries, L. Brand, R. Günther and C. A. M. Seidel, “Monitoring conformational dynamics of a single molecule by selective fluorescence spectroscopy”, *Proceedings of the National Academy of Sciences* **95**, 1556 LP –1561 (1998).
- [57] J. R. Fries, L. Brand, C. Eggeling, M. Köllner and C. A. M. Seidel, “Quantitative Identification of Different Single Molecules by Selective Time-Resolved Confocal Fluorescence Spectroscopy”, *The Journal of Physical Chemistry A* **102**, 6601–6613 (1998).
- [58] E. Nir, X. Michalet, K. M. Hamadani, T. A. Laurence, D. Neuhauser, Y. Kovchegov and S. Weiss, “Shot-noise limited single-molecule FRET histograms: comparison between theory and experiments”, *The journal of physical chemistry. B* **110**, 22103–22124 (2006).
- [59] A. Ingargiola, E. Lerner, S. Chung, S. Weiss and X. Michalet, “FRETbursts: An Open Source Toolkit for Analysis of Freely-Diffusing Single-Molecule FRET”, *PLoS ONE* **11**, edited by V. E. Degtyar, e0160716 (2016).
- [60] B. Ambrose et al., “The smfBox is an open-source platform for single-molecule FRET”, *Nature Communications* **11**, 5641 (2020).
- [61] I. Gopich and A. Szabo, “Theory of photon statistics in single-molecule Förster resonance energy transfer”, *The Journal of Chemical Physics* **122**, 14707 (2005).
- [62] M. Antonik, S. Felekyan, A. Gaiduk and C. A. M. Seidel, “Separating Structural Heterogeneities from Stochastic Variations in Fluorescence Resonance Energy Transfer Distributions via Photon Distribution Analysis”, *The Journal of Physical Chemistry B* **110**, 6970–6978 (2006).
- [63] Y. Santoso, J. P. Torella and A. N. Kapanidis, “Characterizing Single-Molecule FRET Dynamics with Probability Distribution Analysis”, *ChemPhysChem* **11**, 2209–2219 (2010).
- [64] S. Kalinin, A. Valeri, M. Antonik, S. Felekyan and C. A. M. Seidel, “Detection of Structural Dynamics by FRET: A Photon Distribution and Fluorescence Lifetime Analysis of Systems with Multiple States”, *The Journal of Physical Chemistry B* **114**, 7983–7995 (2010).

- [65] J. P. Torella, S. J. Holden, Y. Santoso, J. Hohlbein and A. N. Kapanidis, “Identifying molecular dynamics in single-molecule fret experiments with burst variance analysis”, *Biophysical Journal* **100**, 1568–1577 (2011).
- [66] N. C. Robb et al., “The transcription bubble of the RNA polymerase-promoter open complex exhibits conformational heterogeneity and millisecond-scale dynamics: implications for transcription start-site selection”, *Journal of molecular biology* **425**, 875–885 (2013).
- [67] R. Rigler and E. Elson, *Fluorescence Correlation Spectroscopy: Theory and Applications* (Springer Science & Business Media, 2012), p. 487.
- [68] D. P. Nair, M. Podgórski, S. Chatani, T. Gong, W. Xi, C. R. Fenoli and C. N. Bowman, “The Thiol-Michael Addition Click Reaction: A Powerful and Widely Used Tool in Materials Chemistry”, *Chemistry of Materials* **26**, 724–744 (2014).
- [69] B. H. Northrop, S. H. Frayne and U. Choudhary, “Thiol-maleimide "click" chemistry: evaluating the influence of solvent, initiator, and thiol on the reaction mechanism, kinetics, and selectivity”, *Polymer Chemistry* **6**, 3415–3430 (2015).
- [70] Y. Kim, S. O. Ho, N. R. Gassman, Y. Korlann, E. V. Landorf, F. R. Collart and S. Weiss, “Efficient Site-Specific Labeling of Proteins via Cysteines”, *Bioconjugate Chemistry* **19**, 786–791 (2008).
- [71] R. B. Quast, F. Fatemi, M. Kranendonk, E. Margeat and G. Truan, “Accurate Determination of Human CPR Conformational Equilibrium by smFRET Using Dual Orthogonal Noncanonical Amino Acid Labeling”, *ChemBioChem* **20**, 659–666 (2019).
- [72] S. Schulz, K. Kramm, F. Werner and D. Grohmann, “Fluorescently labeled recombinant RNAP system to probe archaeal transcription initiation”, *Methods* **86**, 10–18 (2015).
- [73] J. Kim, M.-H. Seo, S. Lee, K. Cho, A. Yang, K. Woo, H.-S. Kim and H.-S. Park, “Simple and efficient strategy for site-specific dual labeling of proteins for single-molecule fluorescence resonance energy transfer analysis.”, *Analytical chemistry* **85**, 1468–1474 (2013).
- [74] T. C. Lee, M. Kang, C. H. Kim, P. G. Schultz, E. Chapman and A. A. Deniz, “Dual Unnatural Amino Acid Incorporation and Click-Chemistry Labeling to Enable Single-Molecule FRET Studies of p97 Folding”, *Chembiochem : a European journal of chemical biology* **17**, 981–984 (2016).
- [75] P. Laxman, S. Ansari, K. Gaus and J. Goyette, “The Benefits of Unnatural Amino Acid Incorporation as Protein Labels for Single Molecule Localization Microscopy”, *Frontiers in chemistry* **9**, 641355 (2021).
- [76] H. Tian, T. P. Sakmar and T. Huber, “Site-Specific Labeling of Genetically Encoded Azido Groups for Multicolor, Single-Molecule Fluorescence Imaging of GPCR”, *Methods in Cell Biology* **117**, 267–303 (2013).

-
- [77] G. J. Gopal and A. Kumar, “Strategies for the Production of Recombinant Protein in *Escherichia coli*”, *The Protein Journal* **32**, 419–425 (2013).
- [78] J. L. Hartley, “Cloning technologies for protein expression and purification.”, *Current opinion in biotechnology* **17**, 359–366 (2006).
- [79] S. A. Doyle, *High Throughput Protein Expression and Purification*, edited by S. A. Doyle, 1st ed. (Humana Press, 2009), p. 322.
- [80] J. Kaur, A. Kumar and J. Kaur, “Strategies for optimization of heterologous protein expression in *E. coli*: Roadblocks and reinforcements”, *International Journal of Biological Macromolecules* **106**, 803–822 (2018).
- [81] G. T. Hermanson, “The Reactions of Bioconjugation”, in *Bioconjugate techniques* (Academic Press: New York, 2013), pp. 229–258.
- [82] F. A. Quijcho, J. C. Spurlino and L. E. Rodseth, “Extensive features of tight oligosaccharide binding revealed in high-resolution structures of the maltodextrin transport/chemosensory receptor”, *Structure* **5**, 997–1015 (1997).
- [83] C. N. Pace, F. Vajdos, L. Fee, G. Grimsley and T. Gray, “How to measure and predict the molar absorption coefficient of a protein.”, *Protein Science : A Publication of the Protein Society* **4**, 2411–2423 (1995).
- [84] A. M. Brouwer, “Standards for photoluminescence quantum yield measurements in solution (IUPAC Technical Report)*”, *Pure Appl. Chem.* **83**, 2213–2228 (2011).
- [85] C. Würth, M. Grabolle, J. Pauli, M. Spieles and U. Resch-Genger, “Relative and absolute determination of fluorescence quantum yields of transparent samples”, *Nature Protocols* **8**, 1535–1550 (2013).
- [86] A. Alessi, M. Salvalaggio and G. Ruzzon, “Rhodamine 800 as reference substance for fluorescence quantum yield measurements in deep red emission range”, *Journal of Luminescence* **134**, 385–389 (2013).
- [87] S. S. Henrikus, K. Tassis, L. Zhang, J. H. M. van der Velde, C. Gebhardt, A. Herrmann, G. Jung and T. Cordes, “Characterization of Fluorescent Proteins with Intramolecular Photostabilization*.”, *ChemBioChem* **22**, 3283–3291 (2021).
- [88] D. Magde, R. Wong and P. G. Seybold, “Fluorescence Quantum Yields and Their Relation to Lifetimes of Rhodamine 6G and Fluorescein in Nine Solvents: Improved Absolute Standards for Quantum Yields”, *Photochemistry and Photobiology* **75**, 327–334 (2002).
- [89] K. Rurack, *Standardization and Quality Assurance in Fluorescence Measurements I* (2008).
- [90] P. Kapusta, *Time-resolved fluorescence anisotropy measurements made simple*, tech. rep. (2002), pp. 1–9.
- [91] D. V. O’Connor, W. R. Ware and J. C. Andre, “Deconvolution of fluorescence decay curves. A critical comparison of techniques”, *The Journal of Physical Chemistry* **83**, 1333–1343 (1979).

- [92] T. A. Smith, M. Irwanto, D. J. Haines, K. P. Ghiggino and D. P. Millar, “Time-resolved fluorescence anisotropy measurements of the adsorption of Rhodamine-B and a labelled polyelectrolyte onto colloidal silica”, *Colloid and Polymer Science* **276**, 1032–1037 (1998).
- [93] T. A. Smith and K. P. Ghiggino, “A review of the analysis of complex time-resolved fluorescence anisotropy data”, *Methods and Applications in Fluorescence* **3**, 22001 (2015).
- [94] G. Cosa, K.-S. Focsaneanu, J. R. N. McLean, J. P. McNamee and J. C. Scaiano, “Photophysical Properties of Fluorescent DNA-dyes Bound to Single- and Double-stranded DNA in Aqueous Buffered Solution”, *Photochemistry and Photobiology* **73**, 585 (2001).
- [95] H. Zipper, H. Brunner, J. Bernhagen and F. Vitzthum, “Investigations on DNA intercalation and surface binding by SYBR Green I, its structure determination and methodological implications”, *Nucleic Acids Research* **32**, 10.1093/nar/gnh101 (2004).
- [96] B. Santos-Carballal, M. J. Swamy, B. M. Moerschbacher and F. M. Goycoolea, “SYBR Gold Fluorescence Quenching is a Sensitive Probe of Chitosan-microRNA Interactions”, *Journal of Fluorescence* **26**, 37–42 (2016).
- [97] K. Uno et al., “Key Structural Elements of Unsymmetrical Cyanine Dyes for Highly Sensitive Fluorescence Turn-On DNA Probes”, *Chemistry - An Asian Journal* **12**, 233–238 (2017).
- [98] P. J. Kolbeck, W. Vanderlinden, G. Gemmecker, C. Gebhardt, M. Lehmann, A. Lak, T. Nicolaus, T. Cordes and J. Lipfert, “Molecular structure, DNA binding mode, photophysical properties and recommendations for use of SYBR Gold”, *Nucleic Acids Research* **49**, 5143–5158 (2021).
- [99] G. G. Belford, R. L. Belford and G. Weber, “Dynamics of fluorescence polarization in macromolecules”, *Proceedings of the National Academy of Sciences of the United States of America* **69**, 1392–1393 (1972).
- [100] A. Kowski, “Fluorescence Anisotropy: Theory and Applications of Rotational Depolarization”, *Critical Reviews in Analytical Chemistry* **23**, 459–529 (1993).
- [101] M. Roelfs, “A novel algorithm for reduced data acquisition times in diffusion-based single-molecule FRET microscopy”, BSc Thesis (University of Groningen, 2014), p. 42.
- [102] G. Gouridis, G. K. Schuurman-Wolters, E. Ploetz, F. Husada, R. Vietrov, M. de Boer, T. Cordes and B. Poolman, “Conformational dynamics in substrate-binding domains influences transport in the ABC importer GlnPQ”, *Nat Struct Mol Biol* **22**, 57–64 (2015).
- [103] W. Schrimpf, A. Barth, J. Hendrix and D. C. Lamb, “PAM: A Framework for Integrated Analysis of Imaging, Single-Molecule, and Ensemble Fluorescence Data”, *Biophysical Journal* **114**, 1518–1528 (2018).

-
- [104] C. Sánchez-Rico, L. Voith von Voithenberg, L. Warner, D. C. Lamb and M. Sattler, “Effects of Fluorophore Attachment on Protein Conformation and Dynamics Studied by spFRET and NMR Spectroscopy”, *Chemistry (Weinheim an der Bergstrasse, Germany)* **23**, 14267–14277 (2017).
- [105] S. Sharma, K. Chakraborty, B. K. Müller, N. Astola, Y.-C. Tang, D. C. Lamb, M. Hayer-Hartl and F. U. Hartl, “Monitoring Protein Conformation along the Pathway of Chaperonin-Assisted Folding”, *Cell* **133**, 142–153 (2008).
- [106] F. Husada, G. Gouridis, R. Vietrov, G. K. Schuurman-Wolters, E. Ploetz, M. de Boer, B. Poolman and T. Cordes, “Watching conformational dynamics of ABC transporters with single-molecule tools”, *Biochemical Society Transactions* **43** (2015).
- [107] M. de Boer et al., “Conformational and dynamic plasticity in substrate-binding proteins underlies selective transport in ABC importers”, *eLife* **8**, edited by R. Aldrich, B. Chanda, B. Chanda and M. P. Goldschen-Ohm, e44652 (2019).
- [108] M. F. Peter, C. Gebhardt, J. Glaenger, N. Schneberger, M. de Boer, G. H. Thomas, T. Cordes and G. Hagelueken, “Triggering Closure of a Sialic Acid TRAP Transporter Substrate Binding Protein through Binding of Natural or Artificial Substrates”, *Journal of Molecular Biology* **433**, 166756 (2021).
- [109] E. C. Hulme and M. A. Trevethick, “Ligand binding assays at equilibrium: validation and interpretation”, *British Journal of Pharmacology* **161**, 1219–1237 (2010).
- [110] N. K. Lee, A. N. Kapanidis, Y. Wang, X. Michalet, J. Mukhopadhyay, R. H. Ebright and S. Weiss, “Accurate FRET Measurements within Single Diffusing Biomolecules Using Alternating-Laser Excitation”, *Biophysical Journal* **88**, 2939–2953 (2005).
- [111] B. Hellenkamp et al., “Precision and accuracy of single-molecule FRET measurements - a multi-laboratory benchmark study”, *Nature Methods* **15**, 669–676 (2018).
- [112] R. Tsukanov, T. E. Tomov, Y. Berger, M. Liber and E. Nir, “Conformational Dynamics of DNA Hairpins at Millisecond Resolution Obtained from Analysis of Single-Molecule FRET Histograms”, *The Journal of Physical Chemistry B* **117**, 16105–16109 (2013).
- [113] C. Gebhardt, M. Lehmann, M. M. Reif, M. Zacharias, G. Gemmecker and T. Cordes, “Molecular and Spectroscopic Characterization of Green and Red Cyanine Fluorophores from the Alexa Fluor and AF Series**”, *ChemPhysChem* **22**, 1566–1583 (2021).
- [114] V. M. Corman et al., “Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR”, *Eurosurveillance* **25**, <https://doi.org/10.2807/1560-7917.ES.2020.25.3.2000045> (2020).
- [115] S. W. Hell, “Far-Field Optical Nanoscopy”, *Science* **316**, 1153 LP –1158 (2007).
- [116] S. W. Hell, “Microscopy and its focal switch.”, *Nature methods* **6**, 24–32 (2009).

- [117] S. W. Hell et al., “The 2015 super-resolution microscopy roadmap”, *Journal of Physics D: Applied Physics* **48**, 443001 (2015).
- [118] S. W. Hell and J. Wichmann, “Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy”, *Optics Letters* **19**, 780–782 (1994).
- [119] E. Betzig, G. H. Patterson, R. Sougrat, O. W. Lindwasser, S. Olenych, J. S. Bonifacino, M. W. Davidson, J. Lippincott-Schwartz and H. F. Hess, “Imaging intracellular fluorescent proteins at nanometer resolution.”, *Science (New York, N.Y.)* **313**, 1642–1645 (2006).
- [120] M. Heilemann, S. van de Linde, M. Schüttpelz, R. Kasper, B. Seefeldt, A. Mukherjee, P. Tinnefeld and M. Sauer, “Subdiffraction-Resolution Fluorescence Imaging with Conventional Fluorescent Probes”, *Angewandte Chemie International Edition* **47**, 6172–6176 (2008).
- [121] T. Cordes et al., “Far-Field Nanoscopy with Conventional Fluorophores: Photostability, Photophysics, and Transient Binding”, in *Far-field optical nanoscopy. springer series on fluorescence (methods and applications)* (2015), pp. 215–242.
- [122] R. B. Altman et al., “Cyanine fluorophore derivatives with enhanced photostability.”, *Nature methods* **9**, 68–71 (2011).
- [123] R. B. Altman, Q. Zheng, Z. Zhou, D. S. Terry, J. D. Warren and S. C. Blanchard, “Enhanced photostability of cyanine fluorophores across the visible spectrum”, *Nature methods* **9**, 428–429 (2012).
- [124] P. Tinnefeld and T. Cordes, “‘Self-healing’ dyes: intramolecular stabilization of organic fluorophores”, *Nature Methods* **9**, 426–427 (2012).
- [125] S.-N. Uno et al., “A spontaneously blinking fluorophore based on intramolecular spirocyclization for live-cell super-resolution imaging.”, *Nature chemistry* **6**, 681–689 (2014).
- [126] D. Brox, A. Kiel, S. J. Wörner, M. Pernpointner, P. Comba, B. Martin and D.-P. Herten, “Ensemble and Single-Molecule Studies on Fluorescence Quenching in Transition Metal Bipyridine-Complexes”, *PLOS ONE* **8**, e58049 (2013).
- [127] R. Dave, D. S. Terry, J. B. Munro and S. C. Blanchard, “Mitigating Unwanted Photophysical Processes for Improved Single-Molecule Fluorescence Imaging”, *Biophysical Journal* **96**, 2371–2381 (2009).
- [128] I. Rasnik, S. A. McKinney and T. Ha, “Nonblinking and long-lasting single-molecule fluorescence imaging.”, *Nature methods* **3**, 891–893 (2006).
- [129] E. A. Lemke, Y. Gambin, V. Vandelinder, E. M. Brustad, H.-W. Liu, P. G. Schultz, A. Groisman and A. A. Deniz, “Microfluidic Device for Single-Molecule Experiments with Enhanced Photostability”, *Journal of the American Chemical Society* **131**, 13610–13612 (2009).

-
- [130] C. E. Aitken, R. A. Marshall and J. D. Puglisi, “An Oxygen Scavenging System for Improvement of Dye Stability in Single-Molecule Fluorescence Experiments()”, *Biophysical Journal* **94**, 1826–1835 (2008).
- [131] G. T. Dempsey, M. Bates, W. E. Kowtoniuk, D. R. Liu, R. Y. Tsien and X. Zhuang, “Photoswitching Mechanism of Cyanine Dyes”, *Journal of the American Chemical Society* **131**, 18192–18193 (2009).
- [132] Thermo Fisher Scientific, “1.3 Alexa Fluor Dyes Spanning the Visible and Infrared Spectrum”, in *The molecular probes handbook. a guide to fluorescent probes and labeling technologies*, edited by I. Johnson and M. T. Spence, 11th (2009) Chap. 1.
- [133] C. Eggeling et al., “Direct observation of the nanoscale dynamics of membrane lipids in a living cell”, *Nature* **457**, 1159–1162 (2009).
- [134] T. Kottke, S. van de Linde, M. Sauer, S. Kakorin and M. Heilemann, “Identification of the Product of Photoswitching of an Oxazine Fluorophore Using Fourier Transform Infrared Difference Spectroscopy”, *The Journal of Physical Chemistry Letters* **1**, 3156–3159 (2010).
- [135] O. Söderberg et al., “Direct observation of individual endogenous protein complexes in situ by proximity ligation”, *Nature Methods* **3**, 995–1000 (2006).
- [136] W. Yang, J. Gelles and S. M. Musser, “Imaging of single-molecule translocation through nuclear pore complexes”, *Proceedings of the National Academy of Sciences of the United States of America* **101**, 12887 LP –12892 (2004).
- [137] H. Kim, S. B. Robinson and K. G. Csaky, “Investigating the Movement of Intravitreal Human Serum Albumin Nanoparticles in the Vitreous and Retina”, *Pharmaceutical Research* **26**, 329–337 (2009).
- [138] F. Léveillé, F. El gaamouch, E. Gouix, M. Lecocq, D. Lobner, O. Nicole and A. Buisson, “Neuronal viability is controlled by a functional relation between synaptic and extrasynaptic NMDA receptors”, *The FASEB Journal* **22**, 4258–4271 (2008).
- [139] J.-R. Lin, M. Fallahi-Sichani and P. K. Sorger, “Highly multiplexed imaging of single cells using a high-throughput cyclic immunofluorescence method”, *Nature Communications* **6**, 8390 (2015).
- [140] R. Yarani, T. Shiraishi and P. E. Nielsen, “Effective photo-enhancement of cellular activity of fluorophore-octaarginine antisense PNA conjugates correlates with singlet oxygen formation, endosomal escape and chromophore lipophilicity”, *Scientific Reports* **8**, 638 (2018).
- [141] A. Esteban, M. W. Popp, V. K. Vyas, K. Strijbis, H. L. Ploegh and G. R. Fink, “Fungal recognition is mediated by the association of dectin-1 and galectin-3 in macrophages”, *Proceedings of the National Academy of Sciences* **108**, 14270 LP –14275 (2011).

- [142] A. H. Clayton, F. Walker, S. G. Orchard, C. Henderson, D. Fuchs, J. Rothacker, E. C. Nice and A. W. Burgess, “Ligand-induced dimer-tetramer transition during the activation of the cell surface epidermal growth factor receptor-A multidimensional microscopy analysis”, *Journal of Biological Chemistry* **280**, 30392–30399 (2005).
- [143] A. P. French, S. Mills, R. Swarup, M. J. Bennett and T. P. Pridmore, “Colocalization of fluorescent markers in confocal microscope images of plant cells”, *Nature Protocols* **3**, 619–628 (2008).
- [144] C. Wingren and C. A. K. Borrebaeck, “Antibody microarray analysis of directly labelled complex proteomes”, *Current Opinion in Biotechnology* **19**, 55–61 (2008).
- [145] R. H. Kimura, A. M. Levin, F. V. Cochran and J. R. Cochran, “Engineered cystine knot peptides that bind $\alpha v\beta 3$, $\alpha v\beta 5$, and $\alpha 5\beta 1$ integrins with low-nanomolar affinity”, *Proteins: Structure, Function, and Bioinformatics* **77**, 359–369 (2009).
- [146] B. P. Ziemba, G. H. Swisher, G. Masson, J. E. Burke, R. L. Williams and J. J. Falke, “Regulation of a Coupled MARCKS-PI3K Lipid Kinase Circuit by Calmodulin: Single-Molecule Analysis of a Membrane-Bound Signaling Module”, *Biochemistry* **55**, 6395–6405 (2016).
- [147] J. L. Ballard, V. K. Peeva, C. J. S. DeSilva, J. L. Lynch and N. R. Swanson, “Comparison of Alexa Fluor® and CyDye for practical DNA microarray use”, *Molecular Biotechnology* **36**, 175–183 (2007).
- [148] E. Cauchon, J.-P. Falgoutyret, A. Auger and R. A. Melnyk, “A High-Throughput Continuous Assay for Screening and Characterization of Inhibitors of HIV Reverse-Transcriptase DNA Polymerase Activity”, *Journal of Biomolecular Screening* **16**, 518–524 (2011).
- [149] T. Vöpel, C. S. Hengstenberg, T.-O. Peulen, Y. Ajaj, C. A. M. Seidel, C. Herrmann and J. P. Klare, “Triphosphate Induced Dimerization of Human Guanylate Binding Protein 1 Involves Association of the C-Terminal Helices: A Joint Double Electron-Electron Resonance and FRET Study”, *Biochemistry* **53**, 4590–4600 (2014).
- [150] 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* **121**, 8211–8241 (2017).
- [151] A. Zander, P. Holzmeister, D. Klose, P. Tinnefeld and D. Grohmann, “Single-molecule FRET supports the two-state model of argonaute action”, *RNA Biology* **11**, 45–56 (2014).
- [152] F. Husada, K. Bountra, K. Tassis, M. de Boer, M. Romano, S. Rebuffat, K. Beis and T. Cordes, “Conformational dynamics of the ABC transporter McjD seen by single-molecule FRET”, *The EMBO Journal* **37**, e100056 (2018).

-
- [153] Y. Feng et al., “Conformational Dynamics of apo-GlnBP Revealed by Experimental and Computational Analysis”, *Angewandte Chemie International Edition* **55**, 13990–13994 (2016).
- [154] V. C. DeRocco, T. Anderson, J. Piehler, D. A. Erie and K. Wening, “Four-color single-molecule fluorescence with noncovalent dye labeling to monitor dynamic multimolecular complexes”, *BioTechniques* **49**, 807–816 (2010).
- [155] J. J. McCann, L. Zheng, S. Chiantia and M. E. Bowen, “Domain Orientation in the N-Terminal PDZ Tandem from PSD-95 Is Maintained in the Full-Length Protein”, *Structure* **19**, 810–820 (2011).
- [156] W. Ma, K. D. Whitley, Y. R. Chemla, Z. Luthey-Schulten and K. Schulten, “Free-energy simulations reveal molecular mechanism for functional switch of a DNA helicase”, *eLife* **7**, edited by Y. Shan, e34186 (2018).
- [157] M. Vrljic, P. Strop, J. A. Ernst, R. B. Sutton, S. Chu and A. T. Brunger, “Molecular mechanism of the synaptotagmin-SNARE interaction in Ca²⁺-triggered vesicle fusion”, *Nature Structural & Molecular Biology* **17**, 325–331 (2010).
- [158] A. Gust, A. Zander, A. Gietl, P. Holzmeister, S. Schulz, B. Lalkens, P. Tinnefeld and D. Grohmann, “A starting point for fluorescence-based single-molecule measurements in biomolecular research”, *Molecules* **19**, 15824–15865 (2014).
- [159] T. Graen, M. Hoeffling and H. Grubmüller, “AMBER-DYES: Characterization of Charge Fluctuations and Force Field Parameterization of Fluorescent Dyes for Molecular Dynamics Simulations”, *Journal of Chemical Theory and Computation* **10**, 5505–5512 (2014).
- [160] H. Höfig, M. Gabba, S. Poblete, D. Kempe and J. Fitter, “Inter-Dye Distance Distributions Studied by a Combination of Single-Molecule FRET-Filtered Lifetime Measurements and a Weighted Accessible Volume (wAV) Algorithm”, *Molecules* **19**, 19269–19291 (2014).
- [161] Y. Dai, M. Seeger, J. Weng, S. Song, W. Wang and Y.-W. Tan, “Lipid Regulated Intramolecular Conformational Dynamics of SNARE-Protein Ykt6”, *Scientific Reports* **6**, 30282 (2016).
- [162] L. D. Hughes, R. J. Rawle and S. G. Boxer, “Choose Your Label Wisely: Water-Soluble Fluorophores Often Interact with Lipid Bilayers”, *PLOS ONE* **9**, e87649 (2014).
- [163] F. Rashid, V.-S. Raducanu, M. S. Zaher, M. Tehseen, S. Habuchi and S. M. Hamdan, “Initial state of DNA-Dye complex sets the stage for protein induced fluorescence modulation”, *Nature Communications* **10**, 2104 (2019).
- [164] A. Schmidt, N. Altincekic, H. Gustmann, J. Wachtveitl and M. Hengesbach, “The Protein Microenvironment Governs the Suitability of Labeling Sites for Single-Molecule Spectroscopy of RNP Complexes”, *ACS Chemical Biology* **13**, 2472–2483 (2018).

- [165] L. C. Zanetti-Domingues, C. J. Tynan, D. J. Rolfe, D. T. Clarke and M. Martin-Fernandez, “Hydrophobic Fluorescent Probes Introduce Artifacts into Single Molecule Tracking Experiments Due to Non-Specific Binding”, *PLoS ONE* **8**, 10.1371/journal.pone.0074200 (2013).
- [166] J. A. Hanson, K. Duderstadt, L. P. Watkins, S. Bhattacharyya, J. Brokaw, J. W. Chu and H. Yang, “Illuminating the mechanistic roles of enzyme conformational dynamics”, *Proceedings of the National Academy of Sciences of the United States of America* **104**, 18055–18060 (2007).
- [167] Fluoroprobes, “AFDye 647 Maleimide”, date accessed: 2020-10-09, <https://fluoroprobes.com/product/af-647-maleimide/>.
- [168] G. Gouridis, B. Hertzert, K. Kiosze-Becker, M. de Boer, H. Heinemann, E. Nürenberg-Goloub, T. Cordes and R. Tampé, “ABCE1 Controls Ribosome Recycling by an Asymmetric Dynamic Conformational Equilibrium”, *Cell Reports* **28**, 723–734.e6 (2019).
- [169] G. Y. Wiederschain, “The Molecular Probes handbook. A guide to fluorescent probes and labeling technologies”, *Biochemistry (Moscow)* **76**, 1276 (2011).
- [170] L. Voith von Voithenberg and D. C. Lamb, “Single Pair Förster Resonance Energy Transfer: A Versatile Tool To Investigate Protein Conformational Dynamics”, *BioEssays* **40**, 1700078 (2018).
- [171] Jena Bioscience, “AF647 Maleimide”, date accessed: 2020-10-09, <https://www.jenabioscience.com/probes-epigenetics/reactive-components/reactive-fluorescent-dyes/maleimides/apc-009-af647-maleimide>.
- [172] S. S. Vogel, C. Thaler, P. S. Blank and S. V. Koushik, “Time-Resolved Fluorescence Anisotropy”, in *Flim microscopy in biology and medicine* (2009).
- [173] A. M. Castillo, L. Patiny and J. Wist, “Fast and accurate algorithm for the simulation of NMR spectra of large spin systems.”, *Journal of magnetic resonance (San Diego, Calif. : 1997)* **209**, 123–130 (2011).
- [174] A. J. Sharff, L. E. Rodseth, J. C. Spurlino and F. A. Quiocho, “Crystallographic evidence of a large ligand-induced hinge-twist motion between the two domains of the maltodextrin binding protein involved in active transport and chemotaxis”, *Biochemistry* **31**, 10657–10663 (1992).
- [175] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindahl, “GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers”, *SoftwareX* **1-2**, 19–25 (2015).
- [176] L. Schrödinger, “The PyMOL Molecular Graphics System, Version 2.4”, Nov. 2020.
- [177] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg and C. Simmerling, “Comparison of multiple Amber force fields and development of improved protein backbone parameters”, *Proteins* **65**, 712–725 (2006).

-
- [178] A. Jakalian, B. L. Bush, D. B. Jack and C. I. Bayly, “Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method”, *Journal of Computational Chemistry* **21**, 132–146 (2000).
- [179] J. Wang, W. Wang, P. A. Kollman and D. A. Case, “Automatic atom type and bond type perception in molecular mechanical calculations”, *Journal of Molecular Graphics and Modelling* **25**, 247–260 (2006).
- [180] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, “Development and testing of a general amber force field”, *Journal of Computational Chemistry* **25**, 1157–1174 (2004).
- [181] R. B. Best, H. Hofmann, D. Nettels and B. Schuler, “Quantitative Interpretation of FRET Experiments via Molecular Simulation: Force Field and Validation”, *Biophysical Journal* **108**, 2721–2731 (2015).
- [182] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, “Comparison of simple potential functions for simulating liquid water”, *The Journal of Chemical Physics* **79**, 926–935 (1983).
- [183] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola and J. R. Haak, “Molecular dynamics with coupling to an external bath”, *The Journal of Chemical Physics* **81**, 3684–3690 (1984).
- [184] B. Hess, H. Bekker, H. J. C. Berendsen and J. G. E. M. Fraaije, “LINCS: A linear constraint solver for molecular simulations”, *Journal of Computational Chemistry* **18**, 1463–1472 (1997).
- [185] J. E. Jones and S. Chapman, “On the determination of molecular fields. – I. From the variation of the viscosity of a gas with temperature”, *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* **106**, 441–462 (1924).
- [186] J. A. Barker and R. O. Watts, “Monte Carlo studies of the dielectric properties of water-like models”, *Molecular Physics* **26**, 789–792 (1973).
- [187] G. F. Schröder, U. Alexiev and H. Grubmüller, “Simulation of fluorescence anisotropy experiments: probing protein dynamics”, *Biophysical journal* **89**, 3757–3770 (2005).
- [188] N. Panchuk-Voloshina, R. P. Haugland, J. Bishop-Stewart, M. K. Bhalgat, P. J. Millard, F. Mao, W.-Y. Leung and R. P. Haugland, “Alexa Dyes, a Series of New Fluorescent Dyes that Yield Exceptionally Bright, Photostable Conjugates”, *Journal of Histochemistry & Cytochemistry* **47**, 1179–1188 (1999).
- [189] S. C. Blanchard, H. D. Kim, R. L. Gonzalez, J. D. Puglisi and S. Chu, “tRNA dynamics on the ribosome during translation”, *Proceedings of the National Academy of Sciences of the United States of America* **101**, 12893–12898 (2004).
- [190] Jena Bioscience, “AF555 Maleimide”, date accessed: 2020-10-09, <https://www.jenabioscience.com/probes-epigenetics/reactive-components/reactive-fluorescent-dyes/maleimides/apc-007-af555-maleimide>.

- [191] Fluoroprobes, “AFDye 555 Maleimide”, date accessed: 2020-10-09, <https://fluoroprobes.com/product/alexa-fluor-555-maleimide/>.
- [192] Thermo Fisher Scientific, “Alexa Fluor Dyes - Across the Spectrum”, <https://www.thermofisher.com/de/de/home/brands/molecular-probes/key-molecular-probes-products/alexa-fluor/alexa-fluor-dyes-across-the-spectrum.html>.
- [193] Thermo Fisher Scientific, “Alexa Fluor 555 C2 Maleimid”, date accessed: 2020-10-27, <https://www.thermofisher.com/order/catalog/product/A20346#/A20346>.
- [194] Thermo Fisher Scientific, “Alexa Fluor 647 C2 Maleimid”, date accessed: 2020-10-27, <https://www.thermofisher.com/order/catalog/product/A20347#/A20347>.
- [195] A. L. Davidson, E. Dassa, C. Orelle and J. Chen, “Structure, function, and evolution of bacterial ATP-binding cassette systems.”, *Microbiology and molecular biology reviews : MMBR* **72**, 317–64 (2008).
- [196] R. Mächtel, A. Narducci, D. A. Griffith, T. Cordes and C. Orelle, “An integrated transport mechanism of the maltose ABC importer”, *Research in Microbiology*, <https://doi.org/10.1016/j.resmic.2019.09.004> (2019).
- [197] E. Dassa, “Natural history of ABC systems: not only transporters.”, *Essays in biochemistry* **50**, 19–42 (2011).
- [198] A. Muschielok, J. Andrecka, A. Jawhari, F. Brückner, P. Cramer and J. Michaelis, “A nano-positioning system for macromolecular structural analysis”, *Nature Methods* **5**, 965 (2008).
- [199] X. Kong, E. Nir, K. Hamadani and S. Weiss, “Photobleaching pathways in single-molecule FRET experiments”, *Journal of the American Chemical Society* **129**, 4643–4654 (2007).
- [200] H. Hwang, H. Kim and S. Myong, “Protein induced fluorescence enhancement as a single molecule assay with short distance sensitivity”, *Proceedings of the National Academy of Sciences* **108**, 7414 LP –7418 (2011).
- [201] E. Ploetz, E. Lerner, F. Husada, M. Roelfs, S. Chung, J. Hohlbein, S. Weiss and T. Cordes, “Förster resonance energy transfer and protein-induced fluorescence enhancement as synergetic multi-scale molecular rulers.”, *Scientific Reports* **6**, 33257 (2016).
- [202] E. Lerner, E. Ploetz, J. Hohlbein, T. Cordes and S. Weiss, “A Quantitative Theoretical Framework For Protein-Induced Fluorescence Enhancement - Förster-Type Resonance Energy Transfer (PIFE-FRET)”, *The Journal of Physical Chemistry. B* **120**, 6401–6410 (2016).
- [203] B. J. Harvey, C. Perez and M. Levitus, “DNA sequence-dependent enhancement of Cy3 fluorescence”, *Photochemical & Photobiological Sciences* **8**, 1105–1110 (2009).

-
- [204] M. Levitus, R. M. Negri and P. F. Aramendia, “Rotational Relaxation of Carbocyanines. Comparative Study with the Isomerization Dynamics”, *The Journal of Physical Chemistry* **99**, 14231–14239 (1995).
- [205] Y. Gidi, M. Götte and G. Cosa, “Conformational Changes Spanning Angstroms to Nanometers via a Combined Protein-Induced Fluorescence Enhancement - Förster Resonance Energy Transfer Method”, *The Journal of Physical Chemistry B* **121**, 2039–2048 (2017).
- [206] 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* **9**, 1218–1225 (2012).
- [207] M. Dimura, T. O. Peulen, C. A. Hanke, A. Prakash, H. Gohlke and C. A. M. Seidel, “Quantitative FRET studies and integrative modeling unravel the structure and dynamics of biomolecular systems”, *Current Opinion in Structural Biology* **40**, 163–185 (2016).
- [208] M. Jerabek-Willemsen, T. André, R. Wanner, H. M. Roth, S. Duhr, P. Baaske and D. Breitsprecher, “MicroScale Thermophoresis: Interaction analysis and beyond”, *Journal of Molecular Structure* **1077**, 101–113 (2014).
- [209] M. F. Peter, C. Gebhardt, R. Mächtel, J. Glaenger, G. H. Thomas, T. Cordes and G. Hagelueken, “Cross-validation of distance measurements in proteins by PELDOR/DEER and single-molecule FRET”, *bioRxiv*, 2020.11.23.394080 (2020).
- [210] K. P. Locher, “Mechanistic diversity in ATP-binding cassette (ABC) transporters”, *Nat Struct Mol Biol* **23**, 487–493 (2016).
- [211] C. Thomas and R. Tampé, “Structural and Mechanistic Principles of ABC Transporters.”, *Annual review of biochemistry* **89**, 605–636 (2020).
- [212] M. Huse and J. Kuriyan, “The conformational plasticity of protein kinases.”, *Cell* **109**, 275–282 (2002).
- [213] E. J. Loveridge, E. M. Behiry, J. Guo and R. K. Allemann, “Evidence that a ‘dynamic knockout’ in *Escherichia coli* dihydrofolate reductase does not affect the chemical step of catalysis.”, *Nature chemistry* **4**, 292–297 (2012).
- [214] S. Hofmann et al., “Conformation space of a heterodimeric ABC exporter under turnover conditions.”, *Nature* **571**, 580–583 (2019).
- [215] A. B. Ward, A. Sali and I. A. Wilson, “Biochemistry. Integrative structural biology.”, *Science* **339**, 913–915 (2013).
- [216] A. Sali et al., “Outcome of the First wwPDB Hybrid/Integrative Methods Task Force Workshop”, *Structure* **23**, 1156–1167 (2018).
- [217] M. P. Rout and A. Sali, “Principles for Integrative Structural Biology Studies.”, *Cell* **177**, 1384–1403 (2019).

- [218] S. K. Burley, G. Kurisu, J. L. Markley, H. Nakamura, S. Velankar, H. M. Berman, A. Sali, T. Schwede and J. Trehwella, *PDB-Dev: a Prototype System for Depositing Integrative/Hybrid Structural Models*. eng, Sept. 2017.
- [219] T. E. Wales and J. R. Engen, “Hydrogen exchange mass spectrometry for the analysis of protein dynamics.”, *Mass spectrometry reviews* **25**, 158–170 (2006).
- [220] J. M. A. Bullock, N. Sen, K. Thalassinos and M. Topf, “Modeling Protein Complexes Using Restraints from Crosslinking Mass Spectrometry.”, *Structure* **26**, 1015–1024.e2 (2018).
- [221] J. Hohlbein, T. D. Craggs and T. Cordes, “Alternating-laser excitation: single-molecule FRET and beyond”, *Chem. Soc. Rev.* **43**, 1156–1171 (2014).
- [222] A. T. Tuukkanen, A. Spilotros and D. I. Svergun, “Progress in small-angle scattering from biological solutions at high-brilliance synchrotrons.”, *IUCrJ* **4**, 518–528 (2017).
- [223] A. D. Milov, A. G. Maryasov and Y. D. Tsvetkov, “Pulsed electron double resonance (PELDOR) and its applications in free-radicals research”, *Applied Magnetic Resonance* **15**, 107–143 (1998).
- [224] G. Jeschke, “MMM: A toolbox for integrative structure modeling”, *Protein science: a publication of the Protein Society* **27**, 76–85 (2018).
- [225] E. Lerner et al., “FRET-based dynamic structural biology: Challenges, perspectives and an appeal for open-science practices”, *eLife* **10**, edited by O. Boudker, e60416 (2021).
- [226] D. Klose, J. P. Klare, D. Grohmann, C. W. M. Kay, F. Werner and H.-J. Steinhoff, “Simulation vs. reality: a comparison of in silico distance predictions with DEER and FRET measurements”, *PloS one* **7**, e39492–e39492 (2012).
- [227] D. Grohmann, D. Klose, J. P. Klare, C. W. Kay, H. J. Steinhoff and F. Werner, “RNA-binding to archaeal RNA polymerase subunits F/E: A DEER and FRET study”, *Journal of the American Chemical Society* **132**, 5954–5955 (2010).
- [228] E. Boura, B. Różycki, D. Z. Herrick, H. S. Chung, J. Vecer, W. A. Eaton, D. S. Cafiso, G. Hummer and J. H. Hurley, “Solution structure of the ESCRT-I complex by small-angle X-ray scattering, EPR, and FRET spectroscopy”, *Proceedings of the National Academy of Sciences* **108**, 9437 LP –9442 (2011).
- [229] E. Boura, B. Różycki, H. S. Chung, D. Z. Herrick, B. Canagarajah, D. S. Cafiso, W. A. Eaton, G. Hummer and J. H. Hurley, “Solution structure of the ESCRT-I and -II supercomplex: implications for membrane budding and scission.”, *Structure (London, England : 1993)* **20**, 874–886 (2012).
- [230] H. Sanabria et al., “Resolving dynamics and function of transient states in single enzyme molecules”, *Nature Communications* **11**, 1231 (2020).

-
- [231] B. M. Stadtmueller, M. D. Bridges, K.-M. Dam, M. T. Lerch, K. E. Huey-Tubman, W. L. Hubbell and P. J. Bjorkman, “DEER Spectroscopy Measurements Reveal Multiple Conformations of HIV-1 SOSIP Envelopes that Show Similarities with Envelopes on Native Virions.”, *Immunity* **49**, 235–246.e4 (2018).
- [232] A. Müller, E. Severi, C. Mulligan, A. G. Watts, D. J. Kelly, K. S. Wilson, A. J. Wilkinson and G. H. Thomas, “Conservation of structure and mechanism in primary and secondary transporters exemplified by SiaP, a sialic acid binding virulence factor from *Haemophilus influenzae*.”, *The Journal of biological chemistry* **281**, 22212–22222 (2006).
- [233] C. Mulligan, M. Fischer and G. H. Thomas, “Tripartite ATP-independent periplasmic (TRAP) transporters in bacteria and archaea.”, *FEMS microbiology reviews* **35**, 68–86 (2011).
- [234] J. a. Hall, A. K. Ganesan, J. Chen, H. Nikaido, H. J. B. Chem, H. P. Natl and A. Sci, “Two Modes of Ligand Binding in Maltose-binding Protein of *Escherichia coli*”, *J Biol Chem.* **272**, 17615–17622 (1997).
- [235] C. Tang, C. D. Schwieters and G. M. Clore, “Open-to-closed transition in apo maltose-binding protein observed by paramagnetic NMR”, *Nature* **449**, 1078 (2007).
- [236] F. Fulyani, G. K. Schuurman-Wolters, A. V. Zagar, A. Guskov, D.-J. Slotboom and B. Poolman, “Functional diversity of tandem substrate-binding domains in ABC transporters from pathogenic bacteria.”, *Structure (London, England : 1993)* **21**, 1879–1888 (2013).
- [237] G. K. Schuurman-Wolters and B. Poolman, “Substrate specificity and ionic regulation of GlnPQ from *Lactococcus lactis*. An ATP-binding cassette transporter with four extracytoplasmic substrate-binding domains.”, *The Journal of biological chemistry* **280**, 23785–23790 (2005).
- [238] S. J. Juris, A. E. Rudolph, D. Huddler, K. Orth and J. E. Dixon, “A distinctive role for the *Yersinia* protein kinase: actin binding, kinase activation, and cytoskeleton disruption.”, *Proceedings of the National Academy of Sciences of the United States of America* **97**, 9431–9436 (2000).
- [239] C. Trasad et al., “*Yersinia* protein kinase YopO is activated by a novel G-actin binding process.”, *The Journal of biological chemistry* **282**, 2268–2277 (2007).
- [240] R. P.-A. Berntsson, S. H. J. Smits, L. Schmitt, D.-J. Slotboom and B. Poolman, “A structural classification of substrate-binding proteins.”, *FEBS letters* **584**, 2606–17 (2010).
- [241] G. N. Phillips, V. K. Mahajan, A. K. Siu and F. A. Quiocho, “Structure of L-arabinose-binding protein from *Escherichia coli* at 5 Å resolution and preliminary results at 3.5 Å”, *Proceedings of the National Academy of Sciences* **73**, 2186 LP–2190 (1976).

- [242] F. A. Quijoch and N. K. Vyas, “Novel stereospecificity of the L-arabinose-binding protein”, *Nature* **310**, 381–386 (1984).
- [243] G. Prehna, M. I. Ivanov, J. B. Bliska and C. E. Stebbins, “Yersinia virulence depends on mimicry of host Rho-family nucleotide dissociation inhibitors.”, *Cell* **126**, 869–880 (2006).
- [244] W. L. Lee, J. M. Grimes and R. C. Robinson, “Yersinia effector YopO uses actin as bait to phosphorylate proteins that regulate actin polymerization.”, *Nature structural & molecular biology* **22**, 248–255 (2015).
- [245] M. F. Peter, A. T. Tuukkanen, C. A. Heubach, A. Selsam, F. G. Duthie, D. I. Svergun, O. Schiemann and G. Hagelueken, “Studying Conformational Changes of the Yersinia Type-III-Secretion Effector YopO in Solution by Integrative Structural Biology.”, *Structure (London, England : 1993)* **27**, 1416–1426.e3 (2019).
- [246] G. Jeschke, “DEER Distance Measurements on Proteins”, *Annual Review of Physical Chemistry* **63**, 419–446 (2012).
- [247] O. Schiemann and T. F. Prisner, “Long-range distance determinations in biomacromolecules by EPR spectroscopy.”, *Quarterly reviews of biophysics* **40**, 1–53 (2007).
- [248] J. P. Klare and H.-J. Steinhoff, “Spin labeling EPR.”, *Photosynthesis research* **102**, 377–390 (2009).
- [249] G. Jeschke, “Dipolar Spectroscopy - Double-Resonance Methods”, *eMagRes, Major Reference Works* **5**, 1459–1476 (2016).
- [250] M. Srivastava and J. H. Freed, “Singular Value Decomposition Method To Determine Distance Distributions in Pulsed Dipolar Electron Spin Resonance: II. Estimating Uncertainty”, *The Journal of Physical Chemistry A* **123**, 359–370 (2019).
- [251] M. Gränz, N. Erlenbach, P. Spindler, D. B. Gophane, L. S. Stelzl, S. T. Sigurdsson and T. F. Prisner, “Dynamics of Nucleic Acids at Room Temperature Revealed by Pulsed EPR Spectroscopy”, *Angewandte Chemie International Edition* **57**, 10540–10543 (2018).
- [252] M. J. Schmidt, J. Borbas, M. Drescher and D. Summerer, “A Genetically Encoded Spin Label for Electron Paramagnetic Resonance Distance Measurements”, *Journal of the American Chemical Society* **136**, 1238–1241 (2014).
- [253] R. Igarashi, T. Sakai, H. Hara, T. Tenno, T. Tanaka, H. Tochio and M. Shirakawa, “Distance determination in proteins inside *Xenopus laevis* oocytes by double electron-electron resonance experiments.”, *Journal of the American Chemical Society* **132**, 8228–8229 (2010).
- [254] T. Hett et al., “Spatiotemporal Resolution of Conformational Changes in Biomolecules by Combining Pulsed Electron-Electron Double Resonance Spectroscopy with Microsecond Freeze-Hyperquenching”, *Journal of the American Chemical Society* **143**, 6981–6989 (2021).

-
- [255] M. Isaksson, P. Hägglöf, P. Håkansson, T. Ny and L. B.-Å. Johansson, “Extended Förster theory for determining intraprotein distances: 2. An accurate analysis of fluorescence depolarisation experiments”, *Physical Chemistry Chemical Physics* **9**, 3914–3922 (2007).
- [256] Y.-W. Chiang, P. P. Borbat and J. H. Freed, “The determination of pair distance distributions by pulsed ESR using Tikhonov regularization.”, *Journal of magnetic resonance (San Diego, Calif. : 1997)* **172**, 279–295 (2005).
- [257] G. Jeschke, V. Chechik, P. Ionita, A. Godt, H. Zimmermann, J. Banham, C. R. Timmel, D. Hilger and H. Jung, “DeerAnalysis2006 – a comprehensive software package for analyzing pulsed ELDOR data”, *Applied Magnetic Resonance* **30**, 473–498 (2006).
- [258] S. G. Worswick, J. A. Spencer, G. Jeschke and I. Kuprov, “Deep neural network processing of DEER data.”, *Science advances* **4**, eaat5218 (2018).
- [259] G. Jeschke, M. Pannier, A. Godt and H. Spiess, “Dipolar spectroscopy and spin alignment in electron paramagnetic resonance”, *Chemical Physics Letters* **331**, 243–252 (2000).
- [260] L. Galazzo, G. Meier, M. H. Timachi, C. A. J. Hutter, M. A. Seeger and E. Bordignon, “Spin-labeled nanobodies as protein conformational reporters for electron paramagnetic resonance in cellular membranes.”, *Proceedings of the National Academy of Sciences of the United States of America* **117**, 2441–2448 (2020).
- [261] M. R. Fleissner et al., “Site-directed spin labeling of a genetically encoded unnatural amino acid”, *Proceedings of the National Academy of Sciences* **106**, 21637 LP –21642 (2009).
- [262] R. Roy, S. Hohng and T. Ha, “A practical guide to single-molecule FRET”, *Nature Methods* **5**, 507 (2008).
- [263] T. C. Lee, M. Kang, C. H. Kim, P. G. Schultz, E. Chapman and A. A. Deniz, “Dual Unnatural Amino Acid Incorporation and Click-Chemistry Labeling to Enable Single-Molecule FRET Studies of p97 Folding”, *Chembiochem : a European journal of chemical biology* **17**, 981–984 (2016).
- [264] H. El Mkami and D. G. Norman, “EPR distance measurements in deuterated proteins”, in *Electron paramagnetic resonance investigations of biological systems by using spin labels, spin probes, and intrinsic metal ions, part b*, edited by P. Z. Qin and K. Warncke, *Methods in enzymology* (Academic Press, 2015), pp. 125–152.
- [265] T. Schmidt, M. A. Wälti, J. L. Baber, E. J. Hustedt and G. M. Clore, “Long Distance Measurements up to 160 Å in the GroEL Tetradecamer Using Q-Band DEER EPR Spectroscopy.”, *Angewandte Chemie (International ed. in English)* **55**, 15905–15909 (2016).

- [266] G. Krainer, A. Hartmann and M. Schlierf, “farFRET: Extending the Range in Single-Molecule FRET Experiments beyond 10 nm”, *Nano Letters* **15**, 5826–5829 (2015).
- [267] M. Beckers, F. Drechsler, T. Eilert, J. Nagy and J. Michaelis, “Quantitative structural information from single-molecule FRET”, *Faraday Discussions* **184**, 117–129 (2015).
- [268] L. J. Berliner, J. Grunwald, H. O. Hankovszky and K. Hideg, “A novel reversible thiol-specific spin label: papain active site labeling and inhibition.”, *Analytical biochemistry* **119**, 450–455 (1982).
- [269] H. Yagi, D. Banerjee, B. Graham, T. Huber, D. Goldfarb and G. Otting, “Gadolinium Tagging for High-Precision Measurements of 6 nm Distances in Protein Assemblies by EPR”, *Journal of the American Chemical Society* **133**, 10418–10421 (2011).
- [270] G. W. Reginsson, N. C. Kunjir, S. T. Sigurdsson and O. Schiemann, “Trityl Radicals: Spin Labels for Nanometer-Distance Measurements”, *Chemistry - A European Journal* **18**, 13580–13584 (2012).
- [271] O. Krumkacheva and E. Bagryanskaya, “Trityl radicals as spin labels”, in *Electron paramagnetic resonance*, Vol. 25, edited by V. Chechik and D. M. Murphy, SPR - Electron Paramagnetic Resonance (The Royal Society of Chemistry, 2017), pp. 35–60.
- [272] L. V. Kulik, S. A. Dzuba, I. A. Grigoryev and Y. D. Tsvetkov, “Electron dipole-dipole interaction in ESEEM of nitroxide biradicals.”, *Chem. Phys. Lett.* **343** (2001).
- [273] G. Hagelueken, R. Ward, J. H. Naismith and O. Schiemann, “MtsslWizard: In Silico Spin-Labeling and Generation of Distance Distributions in PyMOL”, *Applied Magnetic Resonance* **42**, 377–391 (2012).
- [274] J. Glaenger, M. F. Peter, G. H. Thomas and G. Hagelueken, “PELDOR Spectroscopy Reveals Two Defined States of a Sialic Acid TRAP Transporter SBP in Solution”, *Biophysical Journal* **112**, 109–120 (2017).
- [275] J. W. Johnston, N. P. Coussens, S. Allen, J. C. D. Houtman, K. H. Turner, A. Zaleski, S. Ramaswamy, B. W. Gibson and M. A. Apicella, “Characterization of the N-acetyl-5-neuraminic acid-binding site of the extracytoplasmic solute receptor (SiaP) of nontypeable *Haemophilus influenzae* strain 2019.”, *The Journal of biological chemistry* **283**, 855–865 (2008).
- [276] G. Hagelueken, D. Abdullin and O. Schiemann, *MtsslSuite: Probing biomolecular conformation by spin-labeling studies*, 1st ed., Vol. 563 (Elsevier Inc., 2015), pp. 595–622.
- [277] M. de Boer, G. Gouridis, Y. A. Muthahari and T. Cordes, “Single-Molecule Observation of Ligand Binding and Conformational Changes in FeuA”, *Biophysical Journal* **117**, 1642–1654 (2019).

-
- [278] J. E. D. Lillington, J. E. Lovett, S. Johnson, P. Roversi, C. R. Timmel and S. M. Lea, “Shigella flexneri Spa15 crystal structure verified in solution by double electron electron resonance.”, *Journal of molecular biology* **405**, 427–435 (2011).
- [279] A. Collauto, H. A. DeBerg, R. Kaufmann, W. N. Zagotta, S. Stoll and D. Goldfarb, “Rates and equilibrium constants of the ligand-induced conformational transition of an HCN ion channel protein domain determined by DEER spectroscopy”, *Physical Chemistry Chemical Physics* **19**, 15324–15334 (2017).
- [280] M. M. Hatmal, Y. Li, B. G. Hegde, P. B. Hegde, C. C. Jao, R. Langen and I. S. Haworth, “Computer modeling of nitroxide spin labels on proteins”, *Biopolymers* **97**, 35–44 (2012).
- [281] G. Jeschke, “Conformational dynamics and distribution of nitroxide spin labels.”, *Progress in nuclear magnetic resonance spectroscopy* **72**, 42–60 (2013).
- [282] N. S. Alexander, R. A. Stein, H. A. Koteiche, K. W. Kaufmann, H. S. McHaourab and J. Meiler, “RosettaEPR: rotamer library for spin label structure and dynamics.”, *PloS one* **8**, e72851 (2013).
- [283] S. Sindbert, S. Kalinin, H. Nguyen, A. Kienzler, L. Clima, W. Bannwarth, B. Appel, S. Müller and C. A. M. Seidel, “Accurate Distance Determination of Nucleic Acids via Förster Resonance Energy Transfer: Implications of Dye Linker Length and Rigidity”, *Journal of the American Chemical Society* **133**, 2463–2480 (2011).
- [284] U. B. Choi, P. Strop, M. Vrljic, S. Chu, A. T. Brunger and K. R. Weninger, “Single-molecule FRET-derived model of the synaptotagmin 1-SNARE fusion complex”, *Nature Structural & Molecular Biology* **17**, 318–324 (2010).
- [285] T. Bordenave et al., “Synthesis and in Vitro and in Vivo Evaluation of MMP-12 Selective Optical Probes.”, *Bioconjugate chemistry* **27**, 2407–2417 (2016).
- [286] Y. NejatyJahromy and E. Schubert, “Demystifying EPR: A Rookie Guide to the Application of Electron Paramagnetic Resonance Spectroscopy on Biomolecules”, *en, Progress in Biological Sciences* **4**, 133–152 (2014).
- [287] O. Schiemann, P. Cekan, D. Margraf, T. F. Prisner and S. T. Sigurdsson, “Relative Orientation of Rigid Nitroxides by PELDOR: Beyond Distance Measurements in Nucleic Acids”, *Angewandte Chemie International Edition* **48**, 3292–3295 (2009).
- [288] A. Godt, M. Schulte, H. Zimmermann and G. Jeschke, “How flexible are poly(paraphenyleneethynylene)s?”, *Angewandte Chemie (International ed. in English)* **45**, 7560–7564 (2006).
- [289] B. Endeward, J. A. Butterwick, R. MacKinnon and T. F. Prisner, “Pulsed electron-electron double-resonance determination of spin-label distances and orientations on the tetrameric potassium ion channel KcsA.”, *Journal of the American Chemical Society* **131**, 15246–15250 (2009).

- [290] M. R. Fleissner, M. D. Bridges, E. K. Brooks, D. Cascio, T. Kálai, K. Hideg and W. L. Hubbell, “Structure and dynamics of a conformationally constrained nitroxide side chain and applications in EPR spectroscopy”, *Proceedings of the National Academy of Sciences* **108**, 16241 LP –16246 (2011).
- [291] N. Fleck, C. A. Heubach, T. Hett, F. R. Haege, P. P. Bawol, H. Baltruschat and O. Schiemann, “SLIM: A Short-Linked, Highly Redox-Stable Trityl Label for High-Sensitivity In-Cell EPR Distance Measurements.”, *Angewandte Chemie (International ed. in English)* **59**, 9767–9772 (2020).
- [292] A. Potapov, H. Yagi, T. Huber, S. Jergic, N. E. Dixon, G. Otting and D. Goldfarb, “Nanometer-Scale Distance Measurements in Proteins Using Gd³⁺ Spin Labeling”, *Journal of the American Chemical Society* **132**, 9040–9048 (2010).
- [293] T. F. Cunningham, M. R. Putterman, A. Desai, W. S. Horne and S. Saxena, “The double-histidine Cu²⁺-binding motif: a highly rigid, site-specific spin probe for electron spin resonance distance measurements.”, *Angewandte Chemie (International ed. in English)* **54**, 6330–6334 (2015).
- [294] S. Spicher, D. Abdullin, S. Grimme and O. Schiemann, “Modeling of spin-spin distance distributions for nitroxide labeled biomacromolecules”, *Physical Chemistry Chemical Physics* **22**, 24282–24290 (2020).
- [295] S. Spicher and S. Grimme, “Robust Atomistic Modeling of Materials, Organometallic, and Biochemical Systems”, *Angewandte Chemie International Edition* **59**, 15665–15673 (2020).
- [296] T. Schmidt, J. Jeon, Y. Okuno, S. C. Chiliveri and G. M. Clore, “Submillisecond Freezing Permits Cryoprotectant-Free EPR Double Electron-Electron Resonance Spectroscopy.”, *Chemphyschem : a European journal of chemical physics and physical chemistry* **21**, 1224–1229 (2020).
- [297] V. Vagenende, M. G. S. Yap and B. L. Trout, “Mechanisms of Protein Stabilization and Prevention of Protein Aggregation by Glycerol”, *Biochemistry* **48**, 11084–11096 (2009).
- [298] L. Vera and E. A. Stura, “Strategies for Protein Cryocrystallography”, *Crystal Growth & Design* **14**, 427–435 (2014).
- [299] E. R. Georgieva, A. S. Roy, V. M. Grigoryants, P. P. Borbat, K. A. Earle, C. P. Scholes and J. H. Freed, “Effect of freezing conditions on distances and their distributions derived from Double Electron Electron Resonance (DEER): a study of doubly-spin-labeled T4 lysozyme.”, *Journal of magnetic resonance (San Diego, Calif. : 1997)* **216**, 69–77 (2012).
- [300] K. Ackermann, J. L. Wort and B. E. Bode, “Nanomolar Pulse Dipolar EPR Spectroscopy in Proteins: Cu^{II}-Cu^{II} and Nitroxide-Nitroxide Cases”, *The Journal of Physical Chemistry B* **125**, 5358–5364 (2021).

-
- [301] J. Glaenger, M. F. Peter and G. Hugelueken, “Studying structure and function of membrane proteins with PELDOR/DEER spectroscopy - The crystallographers’ perspective.”, *Methods* (San Diego, Calif.) **147**, 163–175 (2018).
- [302] O. Schiemann et al., “Benchmark Test and Guidelines for DEER/PELDOR Experiments on Nitroxide-Labeled Biomolecules.”, *Journal of the American Chemical Society* **143**, 17875–17890 (2021).
- [303] L. Fábregas Ibáñez, G. Jeschke and S. Stoll, “DeerLab: a comprehensive software package for analyzing dipolar electron paramagnetic resonance spectroscopy data”, *Magn. Reson.* **1**, 209–224 (2020).
- [304] T. H. Edwards and S. Stoll, “A Bayesian approach to quantifying uncertainty from experimental noise in DEER spectroscopy”, *Journal of Magnetic Resonance* **270**, 87–97 (2016).
- [305] E. Haas, E. Katchalski-Katzir and I. Z. Steinberg, “Effect of the orientation of donor and acceptor on the probability of energy transfer involving electronic transitions of mixed polarization”, *Biochemistry* **17**, 5064–5070 (1978).
- [306] P. D. Harris, A. Narducci, C. Gebhardt, T. Cordes, S. Weiss and E. Lerner, “Multi-parameter photon-by-photon hidden Markov modeling”, *bioRxiv*, 2021.04.08.439035 (2021).
- [307] D. Goldfarb, “Pulse EPR in biological systems â Beyond the expert’s courtyard”, *Journal of Magnetic Resonance* **306**, 102–108 (2019).
- [308] B. Hellenkamp, P. Wortmann, F. Kandzia, M. Zacharias and T. Hugel, “Multi-domain structure and correlated dynamics determined by self-consistent FRET networks”, *Nature methods* **14**, 174–180 (2017).
- [309] T. D. Craggs and A. N. Kapanidis, “Six steps closer to FRET-driven structural biology”, *Nature Methods* **9**, 1157 (2012).
- [310] 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* **115**, 11927–11937 (2011).
- [311] E. Karaca and A. M. J. J. Bonvin, “Advances in integrative modeling of biomolecular complexes”, *Methods* **59**, 372–381 (2013).
- [312] A. Malik, “Protein fusion tags for efficient expression and purification of recombinant proteins in the periplasmic space of *E. coli*”, *3 Biotech* **6**, 44 (2016).
- [313] K. Fukami-Kobayashi, Y. Tateno and K. Nishikawa, “Domain dislocation: a change of core structure in periplasmic binding proteins in their evolutionary history”, *Journal of Molecular Biology* **286**, 279–290 (1999).
- [314] E. Kim, S. Lee, A. Jeon, J. M. Choi, H. S. Lee, S. Hohng and H. S. Kim, “A single-molecule dissection of ligand binding to a protein with intrinsic dynamics”, *Nature Chemical Biology* **9**, 313–318 (2013).

- [315] N. Hiroshi, “Maltose transport system of *Escherichia coli*: An ABC-type transporter”, *FEBS Letters* **346**, 55–58 (1994).
- [316] H. Banerjee, A. Rahn, W. Davis and R. Singh, “Sex lethal and U2 small nuclear ribonucleoprotein auxiliary factor (U2AF65) recognize polypyrimidine tracts using multiple modes of binding”, *RNA (New York, N.Y.)* **9**, 88–99 (2003).
- [317] E. A. Sickmier, K. E. Frato, H. Shen, S. R. Paranawithana, M. R. Green and C. L. Kielkopf, “Structural Basis for Polypyrimidine Tract Recognition by the Essential Pre-mRNA Splicing Factor U2AF65”, *Molecular Cell* **23**, 49–59 (2006).
- [318] C. D. Mackereth, T. Madl, S. Bonnal, B. Simon, K. Zanier, A. Gasch, V. Rybin, J. Valcárcel and M. Sattler, “Multi-domain conformational selection underlies pre-mRNA splicing regulation by U2AF”, *Nature* **475**, 408 (2011).
- [319] L. Voith von Voithenberg, C. Sánchez-Rico, H.-S. Kang, T. Madl, K. Zanier, A. Barth, L. R. Warner, M. Sattler and D. C. Lamb, “Recognition of the 3’ splice site RNA by the U2AF heterodimer involves a dynamic population shift”, *Proceedings of the National Academy of Sciences* **113**, E7169 LP –E7175 (2016).
- [320] T. E. Tomov, R. Tsukanov, R. Masoud, M. Liber, N. Plavner and E. Nir, “Disentangling subpopulations in single-molecule FRET and ALEX experiments with photon distribution analysis”, *Biophysical Journal* **102**, 1163–1173 (2012).
- [321] A. Gansen, A. Valeri, F. Hauger, S. Felekyan, S. Kalinin, K. Tóth, J. Langowski and C. A. M. Seidel, “Nucleosome disassembly intermediates characterized by single-molecule FRET”, *Proceedings of the National Academy of Sciences* **106**, 15308 LP –15313 (2009).
- [322] A. Hartmann, G. Krainer, S. Keller and M. Schlierf, “Quantification of Millisecond Protein-Folding Dynamics in Membrane-Mimetic Environments by Single-Molecule Förster Resonance Energy Transfer Spectroscopy”, *Analytical Chemistry* **87**, 11224–11232 (2015).
- [323] A. Talavera et al., “Phosphorylation decelerates conformational dynamics in bacterial translation elongation factors”, *Science advances* **4**, eaap9714–eaap9714 (2018).
- [324] A. Ingargiola, S. Weiss and E. Lerner, “Monte Carlo Diffusion-Enhanced Photon Inference: Distance Distributions and Conformational Dynamics in Single-Molecule FRET”, *Journal of Physical Chemistry B*, 10.1021/acs.jpcc.8b07608 (2018).
- [325] R. M. Clegg, “Fluorescence resonance energy transfer and nucleic acids”, in *Dna structures part a: synthesis and physical analysis of dna*, Vol. 211, edited by J. E. D. David M.J. Lilley (Academic Press, 1992), pp. 353–388.
- [326] B. Corry, D. Jayatilaka, B. Martinac and P. Rigby, “Determination of the orientational distribution and orientation factor for transfer between membrane-bound fluorophores using a confocal microscope”, *Biophysical journal* **91**, 1032–1045 (2006).

-
- [327] R. E. Dale, J. Eisinger and W. E. Blumberg, “The orientational freedom of molecular probes. The orientation factor in intramolecular energy transfer”, *Biophysical journal* **26**, 161–193 (1979).
- [328] J. Widengren and P. Schwille, “Characterization of Photoinduced Isomerization and Back-Isomerization of the Cyanine Dye Cy5 by Fluorescence Correlation Spectroscopy”, *The Journal of Physical Chemistry A* **104**, 6416–6428 (2000).
- [329] Z. Petrášek and P. Schwille, “Precise Measurement of Diffusion Coefficients using Scanning Fluorescence Correlation Spectroscopy”, *Biophysical Journal* **94**, 1437–1448 (2008).
- [330] S. A. McKinney, C. Joo and T. Ha, “Analysis of single-molecule FRET trajectories using hidden Markov modeling”, *Biophysical journal* **91**, 1941–1951 (2006).
- [331] S. Kalinin, S. Felekyan, M. Antonik and C. A. M. Seidel, “Probability Distribution Analysis of Single-Molecule Fluorescence Anisotropy and Resonance Energy Transfer”, *The Journal of Physical Chemistry B* **111**, 10253–10262 (2007).
- [332] S. Felekyan, S. Kalinin, H. Sanabria, A. Valeri and C. A. M. Seidel, “Filtered FCS: species auto- and cross-correlation functions highlight binding and dynamics in biomolecules”, *Chemphyschem : a European journal of chemical physics and physical chemistry* **13**, 1036–1053 (2012).
- [333] Z. Huang, D. Ji, S. Wang, A. Xia, F. Koberling, M. Patting and R. Erdmann, “Spectral identification of specific photophysics of Cy5 by means of ensemble and single molecule measurements”, *Journal of Physical Chemistry A* **110**, 45–50 (2006).
- [334] R. D. Goldman and D. L. Spector, *Live cell imaging : a laboratory manual* (Cold Spring Harbor, N.Y., 2005), p. 631.
- [335] E. A. Specht, E. Braselmann and A. E. Palmer, “A Critical and Comparative Review of Fluorescent Tools for Live-Cell Imaging”, *Annual Review of Physiology* **79**, 93–117 (2017).
- [336] K. Nienhaus and G. Ulrich Nienhaus, “Fluorescent proteins for live-cell imaging with super-resolution”, *Chem. Soc. Rev.* **43**, 1088–1106 (2014).
- [337] H. Yokota, “Fluorescence microscopy for visualizing single-molecule protein dynamics”, *Biochimica et Biophysica Acta (BBA) - General Subjects* **1864**, 129362 (2020).
- [338] Z. Farka, M. J. Mickert, M. Pastucha, Z. Mikusova, P. Skladal and H. H. Gorris, “Advances in Optical Single-Molecule Detection: En Route to Supersensitive Bioaffinity Assays”, *Angewandte Chemie International Edition* **59**, 10746–10773 (2020).
- [339] J. Liu and Z. Cui, “Fluorescent Labeling of Proteins of Interest in Live Cells: Beyond Fluorescent Proteins”, *Bioconjugate Chemistry* **31**, 1587–1595 (2020).
- [340] G. Jeschke, “The contribution of modern EPR to structural biology.”, *Emerging topics in life sciences* **2**, 9–18 (2018).

- [341] C. Sheppard et al., “Repression of RNA polymerase by the archaeo-viral regulator ORF145/RIP”, *Nature Communications* **7**, 1–13 (2016).
- [342] E. E. Sadler, A. N. Kapanidis and S. J. Tucker, “Solution-Based Single-Molecule FRET Studies of K⁺ Channel Gating in a Lipid Bilayer”, *Biophysical Journal* **110**, 2663–2670 (2016).
- [343] T. D. Craggs et al., “Substrate conformational dynamics facilitate structure-specific recognition of gapped DNA by DNA polymerase”, *Nucleic acids research* **47**, 10788–10800 (2019).
- [344] G. A. Fitzgerald, D. S. Terry, A. L. Warren, M. Quick, J. A. Javitch and S. C. Blanchard, “Quantifying secondary transport at single-molecule resolution”, *Nature* **575**, 528–534 (2019).
- [345] K. Mapa, M. Sikor, V. Kudryavtsev, K. Waegemann, S. Kalinin, C. A. M. Seidel, W. Neupert, D. C. Lamb and D. Mokranjac, “The Conformational Dynamics of the Mitochondrial Hsp70 Chaperone”, *Molecular Cell* **38**, 89–100 (2010).
- [346] R. K. Vishwakarma, A. M. Cao, Z. Morichaud, A. S. Perumal, E. Margeat and K. Brodolin, “Single-molecule analysis reveals the mechanism of transcription activation in *M. tuberculosis*”, *Science Advances* **4**, 1–8 (2018).
- [347] K. Bavishi, D. Li, S. Eiersholt, E. N. Hooley, T. C. Petersen, B. L. Møller, N. S. Hatzakis and T. Laursen, “Direct observation of multiple conformational states in Cytochrome P450 oxidoreductase and their modulation by membrane environment and ionic strength”, *Scientific Reports* **8**, 1–9 (2018).
- [348] I. S. Yanez Orozco et al., “Identifying weak interdomain interactions that stabilize the supertertiary structure of the N-terminal tandem PDZ domains of PSD-95”, *Nature Communications* **9**, 10.1038/s41467-018-06133-0 (2018).
- [349] H. R. Sikkema and B. Poolman, “In silico method for selecting residue pairs for single-molecule microscopy and spectroscopy”, *Scientific Reports* **11**, 5756 (2021).
- [350] T. M. Mitchell, *Machine Learning - Chapter 6: Bayesian Learning*, 1st ed. (McGraw-Hill, Inc., USA, 1997).
- [351] D. D. Lewis, “Naive (Bayes) at forty: The independence assumption in information retrieval BT - Machine Learning: ECML-98”, in, edited by C. Nédellec and C. Rouveirol (1998), pp. 4–15.
- [352] F. Glaser, T. Pupko, I. Paz, R. E. Bell, D. Bechor and E. Martz, “ConSurf: identification of functional regions in proteins by surface mapping of phylogenetic information”, *Bioinformatics* **19**, 163–164 (2003).
- [353] H. Ashkenazy, S. Abadi, E. Martz, O. Chay, I. Mayrose, T. Pupko and N. Ben-Tal, “ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules.”, *Nucleic acids research* **44**, W344–50 (2016).
- [354] W. Kabsch and C. Sander, “Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features.”, *Biopolymers* **22**, 2577–2637 (1983).

-
- [355] T. Hamelryck, “An amino acid has two sides: A new 2D measure provides a different view of solvent exposure”, *Proteins: Structure, Function and Genetics* **59**, 38–48 (2005).
- [356] M. F. Sanner, A. J. Olson and J.-C. Spohner, “Reduced surface: An efficient way to compute molecular surfaces”, *Biopolymers* **38**, 305–320 (1996).
- [357] U. Hobohm, M. Scharf, R. Schneider and C. Sander, “Selection of representative protein data sets.”, *Protein science : a publication of the Protein Society* **1**, 409–417 (1992).
- [358] S. Griep and U. Hobohm, “PDBselect 1992-2009 and PDBfilter-select”, *Nucleic Acids Research* **38**, D318–D319 (2010).
- [359] D. J. Hand and K. Yu, “Idiot’s Bayes: Not So Stupid after All?”, *International Statistical Review / Revue Internationale de Statistique* **69**, 385–398 (2001).
- [360] I. Rish, “An empirical study of the naive Bayes classifier”, in *Ijcai 2001 workshop on empirical methods in artificial intelligence*, Vol. 3, 22 (2001), pp. 41–46.
- [361] P. Domingos, “On the Optimality of the Simple Bayesian Classifier under Zero-One Loss”, *Machine Learning* **29**, 103–130 (1997).
- [362] A. Niculescu-mizil and R. Caruana, “Predicting Good Probabilities With Supervised Learning”, in *Proceedings of the 22nd international conference on machine learning* (Association for Computing Machinery, 2005).
- [363] N. Yao, P. S. Ledvina, A. Choudhary and F. A. Quioco, “Modulation of a Salt Link Does Not Affect Binding of Phosphate to Its Specific Active Transport Receptor”, *Biochemistry* **35**, 2079–2085 (1996).
- [364] Y. Yang, W. Ballent and B. K. Mayer, “High-affinity phosphate-binding protein (PBP) for phosphorous recovery: proof of concept using recombinant *Escherichia coli*.”, *FEMS microbiology letters* **363**, 10.1093/femsle/fnw240 (2016).
- [365] A. Yamashita, S. K. Singh, T. Kawate, Y. Jin and E. Gouaux, “Crystal structure of a bacterial homologue of Na⁺/Cl⁻-dependent neurotransmitter transporters.”, *Nature* **437**, 215–223 (2005).
- [366] Y. Zhao, D. Terry, L. Shi, H. Weinstein, S. C. Blanchard and J. A. Javitch, “Single-molecule dynamics of gating in a neurotransmitter transporter homologue.”, *Nature* **465**, 188–193 (2010).
- [367] S. J. Johnson, J. S. Taylor and L. S. Beese, “Processive DNA synthesis observed in a polymerase crystal suggests a mechanism for the prevention of frameshift mutations.”, *Proceedings of the National Academy of Sciences of the United States of America* **100**, 3895–3900 (2003).
- [368] J. Hohlbein, L. Aigrain, T. D. Craggs, O. Bermek, O. Potapova, P. Shoolizadeh, N. D. F. Grindley, C. M. Joyce and A. N. Kapanidis, “Conformational landscapes of DNA polymerase I and mutator derivatives establish fidelity checkpoints for nucleotide insertion”, *Nature Communications* **4**, 2131 (2013).

- [369] M. Hoefling and H. Grubmüller, “In silico FRET from simulated dye dynamics”, *Computer Physics Communications* **184**, 841–852 (2013).
- [370] I. Reinartz et al., “Simulation of FRET dyes allows quantitative comparison against experimental data”, *The Journal of Chemical Physics* **148**, 123321 (2018).
- [371] T. Dörfler, T. Eilert, C. Röcker, J. Nagy and J. Michaelis, “Structural Information from Single-molecule FRET Experiments Using the Fast Nano-positioning System”, *Journal of Visualized Experiments : JoVE*, 54782 (2017).
- [372] M. Cerminara, A. Schöne, I. Ritter, M. Gabba and J. Fitter, “Mapping Multiple Distances in a Multidomain Protein for the Identification of Folding Intermediates”, *Biophysical Journal* **118**, 688–697 (2020).
- [373] C. Fijen, M. M. Mahmoud, M. Kronenberg, R. Kaup, M. Fontana, J. B. Towle-Weicksel, J. B. Sweasy and J. Hohlbein, “Using single-molecule FRET to probe the nucleotide-dependent conformational landscape of polymerase β -DNA complexes”, *The Journal of biological chemistry* **295**, 9012–9020 (2020).
- [374] A. Ahmed, F. Rippmann, G. Barnickel and H. Gohlke, “A Normal Mode-Based Geometric Simulation Approach for Exploring Biologically Relevant Conformational Transitions in Proteins”, *Journal of Chemical Information and Modeling* **51**, 1604–1622 (2011).
- [375] D. M. Krüger, A. Ahmed and H. Gohlke, “NMSim web server: integrated approach for normal mode-based geometric simulations of biologically relevant conformational transitions in proteins”, *Nucleic acids research* **40**, W310–W316 (2012).
- [376] C. L. Worth, R. Preissner and T. L. Blundell, “SDM – a server for predicting effects of mutations on protein stability and malfunction”, *Nucleic Acids Research* **39**, W215–W222 (2011).
- [377] A. P. Pandurangan, B. Ochoa-Montaña, D. B. Ascher and T. L. Blundell, “SDM: a server for predicting effects of mutations on protein stability”, *Nucleic Acids Research* **45**, W229–W235 (2017).
- [378] D. A. Freedman, “Chapter 2 - The Regression Line”, in *Statistical models : theory and practice* (Cambridge University Press, New York, 2009), pp. 18–28.
- [379] G. Shieh, “A comparison of two indices for the intraclass correlation coefficient”, *Behavior Research Methods* **44**, 1212–1223 (2012).
- [380] J. Cohen, “Chapter 7 - Chi-Square Tests for Goodness of Fit and Contingency Tables”, in *Statistical power analysis for the behavioral sciences* (Academic Press, 1977), pp. 215–271.
- [381] WwPDB, *Protein Data Bank Contents Guide:Atomic Coordinate Entry Format Description Version 3.30*, tech. rep. (2012).
- [382] H. Ashkenazy, E. Erez, E. Martz, T. Pupko and N. Ben-Tal, “ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids.”, *Nucleic acids research* **38**, W529–33 (2010).

-
- [383] T. Hamelryck and B. Manderick, “PDB file parser and structure class implemented in Python”, *Bioinformatics* **19**, 2308–2310 (2003).
- [384] P. J. A. Cock et al., “Biopython: freely available Python tools for computational molecular biology and bioinformatics”, *Bioinformatics* **25**, 1422–1423 (2009).
- [385] M. Z. Tien, A. G. Meyer, D. K. Sydykova, S. J. Spielman and C. O. Wilke, “Maximum Allowed Solvent Accessibilities of Residues in Proteins”, *PLOS ONE* **8**, e80635 (2013).
- [386] B. Rost and C. Sander, “Conservation and prediction of solvent accessibility in protein families”, *Proteins: Structure, Function, and Bioinformatics* **20**, 216–226 (1994).
- [387] S. Miller, J. Janin, A. M. Lesk and C. Chothia, “Interior and surface of monomeric proteins”, *Journal of Molecular Biology* **196**, 641–656 (1987).
- [388] W. E. Moerner, Y. Shechtman and Q. Wang, “Single-molecule spectroscopy and imaging over the decades”, *Faraday Discussions* **184**, 9–36 (2015).
- [389] R. Mächtel, C. Gebhardt and T. Cordes, “Konformationsbewegungen von aktiven Membrantransportern”, *BIOspektrum* **24**, 495–497 (2018).
- [390] B. Safaric, E. Chacin, M. J. Scherr, L. Rajappa, C. Gebhardt, C. F. Kurat, T. Cordes and K. E. Duderstadt, “The fork protection complex recruits FACT to reorganize nucleosomes during replication.”, *Nucleic acids research* **50**, 1317–1334 (2022).
- [391] P. D. Harris, A. Narducci, C. Gebhardt, T. Cordes, S. Weiss and E. Lerner, “Multi-parameter photon-by-photon hidden Markov modeling”, *Nature Communications* **13**, 1000 (2022).

Publications and conference contributions

Key publications of this thesis

- *Molecular and Spectroscopic Characterization of Green and Red Cyanine Fluorophores from the Alexa Fluor and AF Series*
Gebhardt, C., Lehmann, M., Reif, M. M., Zacharias, M., Gemmecker, G., Cordes, T.
ChemPhysChem **22**, 1566-1583 (2021) [113]
- *Cross-validation of distance measurements in proteins by PELDOR/DEER and single-molecule FRET*
Peter, M. F.*, **Gebhardt, C.***, Mächtel, R., Glaenger, Thomas, G.H., Cordes, T., & Hagelueken, G.
(under revision, *Nat. Commun.*)
- *How reliable and accurate is single-molecule FRET for characterization of structural dynamics and distances in proteins?*
Gebhardt, C.*, Agam, G.*, Popara, M.*, Mächtel, R., Folz, J., Ambrose, B., Chamachi, N., Craggs, T.D., de Boer, M., Grohmann, D., Ha, T., Hartmann, A., Hendrix, J., Hirschfeld, V., Hübner, C.G., Hugel, T., Jackers, C., Kammerer, D., Kang, H.-S., Achilles Kapanidis, A., Krainer, G., Kramm, K., Lerner, E., Margeat, E., Martens, K., Michaelis, J., Mitra, J., Moya Muñoz, G.G., Quast, R., Robb, N., Sattler, M., Schlierf, M., Schneider, J., Schröder, T., Sefer, A., Thurn, J., Tinnefeld, P., van Noort, J., Weiss, S., Zijlstra, N., Barth, A., Seidel, C.A.M., Lamb, D.C., Cordes, T.
(in preparation)
- *Labelizer: systematic selection of labeling sites for fluorescent dyes and functional probes on proteins*
Gebhardt, C., Bawidamann, P., Glaenger, Griffith, D.A., Lipfert, J., Cordes, T.
(in preparation)

* these authors contributed equally to the paper

Further publications

- *Precision and accuracy of single-molecule FRET measurements—a multi-laboratory benchmark study*
Hellenkamp, B., Schmid, S., Doroshenko, O., Opanasyuk, O., Kühnemuth, R., Rezaei Adariani, S., Ambrose, B., Aznauryan, M., Barth, A., Birkedal, V., Bowen, M. E., Chen, H., Cordes, T., Eilert, T., Fijen, C., **Gebhardt, C.**, Götz, M., Gouridis, G., Gratton, E., ... Hugel, T.
Nat. Methods **15**, 669–676 (2018) [111]
- *Konformationsbewegungen von aktiven Membrantransportern*
Mächtel, R., **Gebhardt, C.** & Cordes, T.
BIOspektrum **24**, 495–497 (2018) [389]
- *Triggering Closure of a Sialic Acid TRAP Transporter Substrate Binding Protein through Binding of Natural or Artificial Substrates*
Peter, M. F., **Gebhardt, C.**, Glaenzer, J., Schneberger, N., de Boer, M., Thomas, G. H., Cordes, T., & Hugelueken, G.
J. Mol. Biol. **433**, 166756 (2021) [108]
- *Characterization of fluorescent proteins with intramolecular photostabilization*
Henrikus, S. S., Tassis, K., Zhang, L., van der Velde J. H. M., **Gebhardt C.**, Herrmann, A., Jung, G., & Cordes, T.
ChemBioChem. **22**, 3283– 3291 (2021) [87]
- *Molecular structure, DNA binding mode, photophysical properties and recommendations for use of SYBR Gold*
Kolbeck, P.J., Vanderlinden, W., Gemmecker, G., **Gebhardt, C.**, Lehmann, M., Lak, A., Nicolaus T., Cordes, T., und Lipfert, J.
Nucleic Acid Res. **49**, 5143-5158 (2021) [98]
- *The Fork Protection Complex Recruits FACT to Reorganize Nucleosomes During Replication*
Safaric, B., Chacin, E., Scherr, M., Rajappa, L., **Gebhardt, C.**, Kurat, C., Cordes, T., Duderstadt, K.
Nucleic Acid Res. **50**, 1317-1334 (2022) [390]
- *Multi-parameter photon-by-photon hidden Markov modeling*
Harris, P. D., Narducci, A., **Gebhardt, C.**, Cordes, T., Weiss, S., Lerner, E.
Nat. Commun. **13**, 1000 (2022) [391]
- *Inferring kinetic rate constants from single-molecule FRET trajectories – a blinded benchmark of kinetic analysis tools*
Götz, M., Barth, A., Bohr, S.S.-R., Börner, R., Chen, J., Cordes, T., Erie, D.A., **Gebhardt, C.**, Hadzic, M.C.A.S., Hamilton, G., Hatzakis, N.S., Hugel, T., Kiskey, L., Lamb, D.C., Lannoy, C., Mahn, C., Mudiyansele, D.D., ... Schmid, S.
(under revision, *Nat. Commun.*)

Conference contributions

- *Dynamic structural biology of membrane transporters: tools and applications of single-molecule microscopy*
CeNS/SFB1032 Workshop, Poster
Venice, Italy, September 19, 2017
- *Alternating Laser Excitation for Solution-Based Single-Molecule FRET*
COST Membrane Transporter Conference, Talk
Lisbon, Portugal, November 2, 2017
- *Conformation and dynamics of substrate binding proteins*
SFB 863 Winterschool, Talk
Antholz, Italy, February 18, 2018
- *How accurate is single-molecule Förster-resonance energy transfer for dynamic protein structures?*
Synthetic Biology III, Poster
Landshut, Germany, July 23, 2018
- *How accurate is single-molecule Förster-resonance energy transfer for dynamic protein structures?*
Biochemical Society “Structure and mechanism of membrane proteins”, Poster
Birmingham, UK, August 2, 2018
- *How accurate is single-molecule Förster-resonance energy transfer for dynamic protein structures?*
SFB 863 Workshop “Biomolecular Mechanics”, Poster
Hohenkammer, Germany, November 26, 2018
- *Angstrom precision distance measurements in dynamic protein structures with FRET*
DGFB Tagung "Hünfeld 2019" Poster
Hünfeld, Germany, February 15, 2019
- *Angstrom precision distance measurements in dynamic protein structures with single-molecule Förster-resonance energy transfer*
DPG-Frühjahrstagung 2019, Talk
Regensburg, Germany, April 4, 2019
- *Watching Proteins Dancing – how to observe dynamics of single molecules*
CeNS Science Rocks Vortragsserie, Talk
Munich, Germany, July 4, 2019
- *Where to mutate proteins and where to label fluorophores for single-molecules FRET?*
International Conference on Molecular Systems Engineering 2019, Poster
Nijmegen, The Netherlands, August 18, 2019

-
- *Angstrom precision distance measurements in dynamic protein structures with FRET*
CeNS/CRC235 Workshop "Evolving Nanosciences", Poster
Venice, Italy, September 24, 2019
 - *Watching single proteins dancing: Angstrom precision distance measurements in dynamic protein structures with single-molecule Förster-resonance energy transfer*
<interact> 2020, Talk
Munich, Germany, February 28, 2020

Danksagung

Ich möchte an dieser Stelle noch all jenen meinen Dank aussprechen, die mich während der viereinhalb Jahre der Promotion unterstützt und begleitet habend und damit letztlich zu deren Gelingen dieser Arbeit beigetragen haben.

Ganz besonders möchte ich meinem Betreuer Prof. Thorben Cordes danken, der mich in meinem breitgefächerten Forschungsgebiet zwischen Physik, Chemie und Biologie mit seinen fachübergreifenden Kenntnissen und seinen Ideen unterstützte. Insbesondere möchte ich mich auch dafür bedanken, dass er meine Arbeit immer aktiv gefördert hat, z. B. indem er mich bei Stipendienbewerbungen unterstützte oder Kollaborationen außerhalb der Gruppe initiierte, und es mir ermöglicht hat an einer Vielzahl von Konferenzen teilzunehmen. Weiter bin dankbar, dass ich die Richtung meiner Dissertation selbst aktiv mitgestalten konnte und während der Zeit sehr eigenständig den Fokus richten konnte. Des Weiteren weiß ich zu schätzen, dass er stets an einem guten, freundschaftlichen Miteinander in der Gruppe interessiert ist und dies durch die gemeinsamen Abendessen oder den Wanderausflügen der Gruppe unterstützt hat.

Mein Dank gilt ebenso Prof. Jan Lipfert dafür, dass er die Betreuung dieser Arbeit von Seiten der Physikfakultät übernommen hat und mich mit Rat und Tat unterstützt hat. Insbesondere möchte ich mich für die vielen inspirierenden Diskussionen und die vielen guten Ideen im Rahmen unserer gemeinsamen Projekte bedanken.

Ich möchte mich weiter bei all meinen Kollegen für die Zusammenarbeit bedanken, die sowohl direkt oder auch indirekt an der Verwirklichung dieser Arbeit beteiligt waren. Dazu zählt Rebecca Mächtel, die den Großteil meiner Mutanten gemacht. Danke auch an Michael Isselstein und Oliver Brix für den Austausch in physikalischen Fragestellungen und die gute Zusammenarbeit bei der Lehre. Further I want to thank Alessandra Narducci who was always supportive for any smaller or larger request. I also want to thank Douglas Griffith and Niels Zijlstra for their help and their advice.

Ich durfte im Laufe dieser Arbeit einige Studenten betreuen, was ich immer wieder gerne gemacht habe. Danke für die Zusammenarbeit an Jonathan Schneider, Dominik Kammerer, Sebastian Bibinger. Insbesondere gilt hier mein Dank Gabriel Moya, Markus Springer und Pascal Bawidaman, die mit Ihrer Arbeit auch signifikant zu meiner Promotion beigetragen haben.

Weiter möchte ich allen Kollaborationspartnern danken, mit denen ich bei unterschiedlichen Projekten zusammengearbeitet habe. Hier werde ich insbesondere die Zeit mit Barbara Šafarić trotz der vielen Rückschläge in guter Erinnerung behalten.

Des Weiteren möchte ich Susanne Mentz und Marija Ram danken, dass sie unser Labor während den letzten vier Jahren stets am Laufen gehalten haben und durch ihre Tätigkeit die wissenschaftliche Arbeit sehr erleichtert haben.

Im Laufe der Jahre habe ich mir mit Michael Isselstein, Alessandra Narducci und Rebecca Mächtel, sowie mit Niels Zijlstra und Peng Zhou ein Büro geteilt. Ich möchte mich für das angenehme Miteinander, die vielen Gespräche und die unterhaltsamen Momente bedanken. Insbesondere die Zeit in unserem Viererbüro am Anfang wird mir in guter Erinnerung bleiben.

At this point I would also like to thank all international group members for the interesting conversations, where I learned a lot about other cultures. Thank you for the nice discussions. Here, a special thanks goes to Zhongyin Han.

Zu guter Letzt möchte ich den Menschen danken, die mich in den vergangenen Jahren bei allem unterstützt haben immer für mich da waren. Da sind insbesondere meine Eltern zu nennen, die mich, was immer auch kam, bedingungslos auf meinem Werdegang unterstützt haben. Mein ganz besonderer Dank gilt meiner Frau Julia, die mir in den vergangenen Monaten so viel abgenommen hat, die mich auch in schwierigen Phasen immer wieder aufgemuntert hat und mein ganzes Leben erfüllend macht.

Vielen lieben Dank euch allen!